Evidence-based oncology:

the use of methodologically complex systematic reviews to inform cancer research and clinical practice

Inaugural Dissertation

zur

Erlangung des Doktorgrades
philosophiae doctor (PhD) in Health Sciences
der Medizinischen Fakultät
der Universität zu Köln

vorgelegt von

Angela Aldin

aus Bagdad

A&A Digital Print Center, Bonn

2023

Betreuerin: Frau Prof. Dr. Nicole Skoetz

Gutachter: Herr Prof. Dr. Matthias Schmidt

Herr Prof. Dr. Achim Tresch

Datum der mündlichen Prüfung: 22.06.2023

ACKNOWLEDGEMENTS

I want to express my sincere gratitude to my supervisor, Prof. Dr. Nicole Skoetz, who gave me the opportunity to pursue a PhD in the first place. Thank you for everything that you taught me during the five years that I have worked in your team. Thank you for your guidance, support, encouragement, and for helping me grow professionally and personally.

I deeply thank my tutors, Prof. Dr. Matthias Schmidt, and Prof. Dr. Achim Tresch, for their continuous support throughout the years, and for believing in me and my projects.

I thank my former team at Cochrane Haematology / Working group for Evidence-Based Medicine (former Evidence-Based Oncology) at the University Hospital of Cologne, with whom I had the privilege to work with during these five years. With you, I met as ambitious, inspiring, intelligent, friendly, supportive, and kind-hearted colleagues as one could wish for. Thank you for all the professional support. Thank you for all the laughter we had together, and the memories that we created together - from the cosy Christmas markets of Cologne to the breath-taking beaches of Chile.

I am deeply grateful to all my co-authors on both systematic reviews of this cumulative dissertation, who supported me in realising these projects from the beginning until the end. Thank you for sharing your expertise and knowledge with me throughout the processes.

I thank all staff at Cochrane who were involved in the peer-review process and publication of both systematic reviews within this dissertation. I particularly thank the staff at the Cochrane Central Editorial Service for managing the editorial processes of my reviews. I thank all external peer-reviewers who provided feedback on both reviews.

I thank Annika and Digo ("Glimmies"), my former colleagues who have become my dear friends. You two are so incredibly inspiring and I have learned so much from you. I thank you for your constant support throughout the years, for your encouragement and for believing in me.

A big thank you to my two best friends – my sisters, my chosen family - Natascha and Yasmin. I am the luckiest person because I have the greatest two best friends that one could wish for in this world. Thank you for always believing in me, supporting me, for washing away my doubts and always providing me with joy and happiness when I need it the most.

I thank you Philipp, my partner, my love. I cannot put into words how much strength you have given me during the past years. Thank you for supporting me, for believing in my life choices and giving me the time and space that I needed to complete this dissertation and to pursue new professional opportunities.

I dedicate this dissertation to my parents, Naimet and Maria. As your child, it is easy to say "I would not be here without you"; but really, I would not be HERE in this position, writing these words, without you. Thank you for all your hard work, for migrating from country to country to give me and my brother a better and brighter future, and more opportunities to make the most of life. Thank you for giving me the opportunity - and paying for it when you had so little - to pursue an academic career, and for believing in me. Thank you for everything you have done and sacrificed that allowed me to sit here today and write this dissertation.

"It is surely a great criticism of our profession that we have not organised a critical summary, by speciality and subspeciality, adapted periodically, of all randomised controlled trials."

(Archie Cochrane, 1979)

ABSTRACT

Background

Systematic reviews are produced to inform health research and clinical practice, e.g., by identifying research gaps and by formulating recommendations in clinical practice guidelines. Standardised methodology exists for the conduct of systematic reviews of interventions. To answer clinically diverse research questions, new methods are constantly being developed for the systematic synthesis of results from different types of studies. Moreover, constant monitoring of newly available evidence, particularly in clinical areas that are rapidly evolving, is important to ensure the currency of systematic reviews.

Objective

The primary objective of this cumulative dissertation was to conduct systematic reviews using new and complex systematic review methods, and to contribute to the further development and refinement of these methods. Secondary objective was to conduct clinically relevant systematic reviews to provide meaningful evidence that can inform clinical practice and health care in oncology.

Methods

Two clinically relevant systematic reviews using novel and complex methodological approaches were conducted:

Systematic review I: A systematic review with network meta-analysis and an adapted *living* approach to evaluate and compare the benefits and harms of first-line therapies for adults with advanced renal cell carcinoma.

Systematic review II: A systematic review with meta-analysis of prognostic factor studies to explore the interim positron emission tomography (PET) scan result as a prognostic factor in adults with newly diagnosed Hodgkin lymphoma.

Results

Methodological results

Systematic review I: The evidence for the currently recommended treatments and important comparisons in this review stem from direct evidence from one trial per comparison only. This is due to the great lack of head-to-head comparisons of the many treatment options available. Statistical validation of the homogeneity and consistency assumptions was not possible for every network meta-analysis, so the validity of estimates is largely based on the transitivity assumption. When a strong evidence base is missing, the results of a network meta-analysis, including the ranking of treatments, should be interpreted with caution. The adapted *living* approach, where monthly update searches were conducted during the conduct of the review, was an appropriate method to maintain the currency of the evidence in such a rapidly evolving treatment landscape.

Systematic review II: The greatest methodological challenges identified in synthesising evidence from prognostic factor studies were that, firstly, searching for prognosis studies is challenging due to insufficient indexing and missing search filters that are specific and sensitive enough to identify prognostic factor studies. Secondly, extracting and analysing outcome results was particularly difficult due to incomplete reporting of important data in the, usually retrospective, studies. Thirdly, available methods for the quality assessments had to be adapted to fit to the review question. Lastly, methods for the certainty assessment of the evidence from prognosis studies had to be developed during the conduct of the review as there was no official guidance at that time.

The challenges encountered during the conduct of both reviews were discussed and resolved through the involvement of methodological and clinical experts as co-authors.

Clinical results

Systematic Review I: Combinations of novel therapies (e.g., a checkpoint inhibitor with a tyrosine kinase inhibitor) appear to be superior to monotherapy with sunitinib (a tyrosine kinase inhibitor) as first-line therapy in terms of survival for adults with advanced renal cell carcinoma. However, these novel treatments may cause more (serious) side effects. Moreover, the question on the potential impact of these novel treatments on the quality of life of affected individuals remains unanswered.

Systematic Review II: Evidence was found on the prognostic ability of the interim PET-scan result to predict survival in adults with Hodgkin lymphoma. It successfully distinguishes between PET-negative people, who have a better outcome prognosis, and PET-positive people, who have a worse outcome prognosis.

Conclusion

Future methodological research needs to further address these different challenges, for example the challenges one encounters when trying to search for and identify prognostic factor studies, or the limitations one encounters when underlying assumptions of a network meta-analysis cannot be verified. When evidence from such methodologically complex systematic reviews shall be used to inform clinical practice guidelines and, thereby, health care decision making, all involved stakeholders need to be aware of the methodological complexity and limitations behind the evidence produced.

GERMAN ABSTRACT

Hintergrund

Systematische Reviews werden erstellt, um die Gesundheitsforschung und klinische Praxis zu informieren, z. B. durch die Ermittlung von Forschungslücken und die Formulierung von Handlungsempfehlungen in Leitlinien. Für die Durchführung von systematischen Reviews zu Interventionen existieren standardisierte Methoden. Um klinisch verschiedene Forschungsfragen zu beantworten, wird stetig an neuen Methoden für die systematische Synthese von Ergebnissen aus unterschiedlichen Studientypen gearbeitet. Darüber hinaus ist eine ständige Überwachung neu verfügbarer Evidenz, insbesondere in klinischen Bereichen, die sich rapide weiterentwickeln, wichtig, um die Aktualität von systematischen Reviews zu erhalten.

Ziel

Das primäre Ziel dieser kumulativen Dissertation bestand darin, systematische Reviews unter Verwendung neuer und komplexer Methoden zu erstellen und damit zur Weiterentwicklung und Verfeinerung dieser Methoden beizutragen. Das sekundäre Ziel bestand darin, klinisch relevante systematische Reviews zu erstellen, die wichtige Erkenntnisse für die klinische Praxis und die Gesundheitsversorgung in der Onkologie liefern.

Methoden

Es wurden zwei klinisch relevante systematische Reviews unter Verwendung von neuen und komplexen methodischen Ansätzen erstellt:

Systematisches Review I: Ein systematisches Review mit Netzwerk-Metaanalyse und einem adaptierten *living* Ansatz zur Untersuchung von Nutzen und Schaden von Erstlinientherapien für Erwachsene mit fortgeschrittenem Nierenzellkarzinom.

Systematisches Review II: Ein systematisches Review mit Meta-Analyse von prognostischen Faktor Studien, um das Interim Positronen-Emissions-Tomographie (PET) Scan Resultat als prognostischen Faktor bei Erwachsenen mit neu diagnostiziertem Hodgkin Lymphom zu untersuchen.

Ergebnisse

Methodische Ergebnisse

Systematisches Review I: Die Evidenz für die derzeit empfohlenen Therapien in diesem Review stammt aus nur einer Studie pro Vergleich. Dies ist eine Folge des großen Mangels von direkten Vergleichen der vielen verfügbaren Behandlungsoptionen. Daher war die statistische Validierung der Homogenitäts- und Konsistenzannahmen nicht für jede Netzwerk-Metaanalyse möglich, so dass die Gültigkeit der Effektschätzer weitestgehend auf der Transitivitätsannahme beruht. In

Ermangelung einer soliden Evidenzbasis sollten die Ergebnisse einer Netzwerk-Meta-Analyse, einschließlich der Rangfolge der Behandlungen, vorsichtig interpretiert werden. Der adaptierte *living* Ansatz, bei dem während der Durchführung des Reviews monatliche Updatesuchen durchgeführt wurden, war eine geeignete Methode, um die Aktualität der Evidenz in einer sich so rapide entwickelnden Behandlungslandschaft zu gewährleisten.

Systematisches Review II: Die größten methodischen Herausforderungen, die bei der Evidenzsynthese von prognostischen Faktorstudien identifiziert wurden, waren erstens, die Suche nach Prognosestudien aufgrund unzureichender Indexierung und fehlender Suchfilter, die spezifisch und sensitiv genug sind, um prognostische Faktorstudien zu identifizieren. Zweitens war es besonders schwierig, Studienergebnisse zu extrahieren und zu analysieren, da wichtige Daten in den meist retrospektiven Studien unvollständig berichtet wurden. Drittens mussten die verfügbaren Methoden für die Qualitätsbewertung an die Fragestellung des Reviews angepasst werden. Schließlich mussten Methoden für die Bewertung des Vertrauens in die Evidenz aus Prognosestudien während der Review-Erstellung entwickelt werden, da es zu diesem Zeitpunkt keine offizielle Anleitung gab.

Die Herausforderungen, die bei der Durchführung von beiden systematischen Reviews aufgetreten sind, wurden diskutiert und durch die Einbeziehung von methodischen und klinischen Expertinnen und Experten als Co-Autorinnen und Co-Autoren bewältigt.

Klinische Ergebnisse

Systematisches Review I: Kombinationen von neuartigen Therapien (z.B. ein Checkpoint-Inhibitor mit einem Tyrosinkinase-Inhibitor) scheinen einer Monotherapie mit Sunitinib (ein Tyrosinkinase-Inhibitor) als Erstlinientherapie im Hinblick auf das Überleben von Erwachsenen mit fortgeschrittenem Nierenzellkarzinom überlegen zu sein. Allerdings können diese neuartigen Therapien und Kombinationen mehr (schwere) unerwünschte Nebenwirkungen mit sich bringen. Des Weiteren bleibt die Frage über die möglichen Auswirkungen dieser Therapien auf die Lebensqualität von Betroffenen unbeantwortet.

Systematisches Review II: Das Interim PET-Scan Resultat ist ein geeigneter prognostischer Faktor zur Vorhersage der Überlebenswahrscheinlichkeit von Erwachsenen mit Hodgkin Lymphom. Das Interim PET unterscheidet erfolgreich PET-negative Personen, die eine bessere Prognose haben, von PET-positiven Personen, die eine schlechtere Prognose haben.

Schlussfolgerung

Zukünftig sollte die Methodenforschung die genannten Herausforderungen weiter angehen, z. B. solche, auf die man bei der Suche und Identifizierung von prognostischen Faktorstudien stößt, oder jene, wenn die zugrunde liegenden Annahmen einer Netzwerk-Metaanalyse nicht untersucht oder validiert werden

können. Wenn die Erkenntnisse aus methodisch komplexen systematischen Reviews zur Erstellung von Empfehlungen in Leitlinien, und damit zur Entscheidungsfindung im Gesundheitswesen herangezogen werden sollen, sollten sich alle Beteiligten der methodischen Komplexität sowie den Grenzen der gewonnenen Erkenntnisse bewusst sein.

CONTENTS

LIST OF ABBREVIATIONS

Abbreviation	Definition
AE	Adverse Event
ASCO	American Society of Clinical Oncology
AVE	Avelumab
AXI	Axitinib
BMBF	Bundesministerium für Bildung und Forschung
CAB	Cabozantinib
CENTRAL	Cochrane Central Register of Controlled Trials
CI	Confidence Interval
CSR	Clinical Study Report
CT	Computed Tomography
CTCAE	Common Terminology Criteria for Adverse Events
DS	Deauville Score
ESMO	European Society of Medical Oncology
FDG	Fluorodeoxyglucose
GCO	Global Cancer Observatory
GGPO	German Guideline Program in Oncology
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HDI	Human Development Index
HL	Hodgkin Lymphoma
HR	Hazard Ratio
IMDC	International Metastatic RCC Database Consortium
IOM	Institute of Medicine
IPI	Ipilimumab
LEN	Lenvatinib
MSKCC	Memorial Sloan Kettering Cancer Center
NCCN	National Comprehensive Cancer Network

NIV	Nivolumab
OS	Overall Survival
PAZ	Pazopanib
PEM	Pembrolizumab
PET	Positron Emission Tomography
PET-CT	Positron Emission Tomography with Computed Tomography
PFS	Progression-Free Survival
PICO	Population, Intervention, Comparator, Outcome
PICOTS	Population, Index prognostic factor, Comparator prognostic factor, Outcome, Timing, Setting
PROBAST	Prediction Model Risk of Bias Assessment Tool
PROGRESS	Prognosis Research Strategy
PROSPERO	International Prospective Register of Systematic Reviews
QoL	Quality of Life
QUIPS	Quality in Prognostic Studies
RCC	Renal Cell Carcinoma
R-classification	Residual Tumour Classification
RCT	Randomised Controlled Trial
RoB	Risk of Bias
RR	Risk Ratio
SAE	Serious Adverse Events
SoF	Summary of Findings
SUN	Sunitinib
SR	Systematic review
TFST	Time to initiation of the First Subsequent anticancer Therapy
TNM	TNM Classification of Malignant Tumours (T = tumour, N = lymph node, M = metastasis)
WHO	World Health Organization
FDA	US Food and Drug Administration

LIST OF FIGURES

Figures

All figures have been self-created using the Whimsical online software and Cochrane's ReviewManager (Version 5.4.1).

Visual abstracts

Visual abstracts have been self-created using the CANVA online software.

CHAPTER 1

INTRODUCTION

Introduction

Evidence-based medicine has been defined as *"the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients"* (1). The key components of evidence-based medicine are the clinical evidence derived from systematic research on the one hand, and the individual clinical expertise on the other hand (1, 2). Important to note is that one never precludes the other. In fact, while clinical research is crucial for assessing, for example, the effectiveness and safety of a therapeutic substance, the use of this substance may be appropriate for one patient but inappropriate for the next. Therefore, clinical expertise acquired from year-long experience in clinical practice is necessary for considering the individual patient's characteristics, preferences and rights when taking a health care decision (1, 2). Ultimately, it is the treating physician's clinical expertise, combined with the patient's individual wishes and preferences as well as the available clinical research evidence that together shall be used to take an informed health care decision.

Keeping up with the growing body of evidence in many clinical areas can be challenging. In oncology, the number of clinical trials continuously increases. Most clinical trials are carried out to compare the efficacy of different therapeutic or non-therapeutic interventions for the treatment of certain illnesses. Most often, the comparative efficacy of novel therapies and standard therapies (i.e., therapies that are approved and most recommended) is assessed. While nowadays many different treatments for different types of cancers are available, clinical research in oncology increasingly focusses on the development of even more effective therapies while also trying to minimise their harmful effects. Ultimately, this leads to a growing availability of treatment alternatives that patients, their caregivers and treating physicians as well as health systems can choose from. However, choosing a treatment option requires informed decision-making, but it is not possible for one person alone to review all literature on all the different available treatments to take an informed decision on the possibly most effective one.

This is where systematic reviews come into play. The aim of a systematic review is to combine and evaluate the evidence from primary research studies to answer a specific, explicit, and clinically relevant research question (3). Systematic reviews are systematic, transparent, and replicable in their methodology (2, 4), and they are important for evidence-based decision-making in health care (3). Systematic reviews even go beyond the provision of information about the effectiveness of a treatment or a health care service; they also highlight research gaps where evidence is lacking and research questions remain unanswered (4). Systematic reviews have become indispensable for health research and practice. The systematic synthesis of results from primary research studies combined with the critical appraisal of the quality of these studies can shed light on important clinical questions, and enable increased and fast access to relevant evidence for health professionals, patients and their caregivers, and health policy makers (2). Moreover, systematic reviews have become essential for developing and updating clinical practice guidelines (5). Over the past 35 years, the

number of published systematic reviews and meta-analyses has increased enormously. Between 1986 and 2015, 266,782 references on PubMed[1] were tagged as *systematic review* and 58,611 references were tagged as *meta-analysis*[2] (6). While wrong indexing of references has been a problem (6), these numbers still show the value and importance that has been attributed to the summary of evidence in form of systematic reviews (with or without meta-analysis).

In 1993, Cochrane[3] (formerly the Cochrane Collaboration) was established, a British not-for-profit organisation, which constitutes a global network of different stakeholders who aim *"to produce credible information that is free from commercial sponsorship and other conflicts of interest"* (3). Under the maxim *"Trusted evidence. Informed decisions. Better health."* (3), health and social care professionals, health researchers and methodologists, patients and patient representatives as well as other people interested in health matters have made it their mission to produce systematic reviews that are clinically relevant, up to date and methodologically of high quality, whilst also being accessible and informative to the general population (3). Ultimately, the evidence derived from these systematic reviews shall be used to inform evidence-based clinical practice guidelines as well as health policies (3). At focus during the development of clinically relevant and up to date systematic reviews, published in the Cochrane Library[4], lies also the use and constant development and refinement of rigorous systematic review methodology (3). Internationally, Cochrane reviews have been recognised as the gold standard for producing high-quality and trusted systematic reviews (7, 8).

The primary objective of this cumulative dissertation was to conduct systematic reviews using new and complex systematic review methods to contribute to the development and refinement of these methods. The secondary objective was to conduct clinically relevant systematic reviews to provide meaningful evidence that may inform clinical practice in oncology. To achieve these objectives, two methodologically complex Cochrane systematic reviews were conducted.

[1] PubMed, of the US National Library of Medicine, is a database of references and abstracts on life sciences and biomedical topics, primarily accessing references from Medline[5] (pubmed.ncbi.nlm.nih.gov/).
[2] A meta-analysis is the statistical pooling and analysis of the results from multiple individual primary studies.
[3] Cochrane's official website: www.cochrane.org
[4] The Cochrane Library (www.cochranelibrary.com) is a collection of databases in medicine and other health sciences. Furthermore, Cochrane systematic reviews and meta-analyses are published in the Cochrane Library.

CHAPTER 2

BACKGROUND

Background

Systematic reviews of interventions

The PICO

A systematic review is based on a simple framework – the **PICO** – which is the acronym for:

- **P**opulation,
- **I**ntervention,
- **C**omparison, and
- **O**utcome (3).

It is the basis of a review and defines the scope of its research question (3). The PICO contains information about the population of interest that will be studied (e.g., people with a specific condition); the intervention(s) of interest (usually a novel therapy or a range of different therapies, e.g., novel therapeutic substances); the comparator of interest (an alternative intervention, e.g., current standard therapy for the specific condition of interest); and lastly, the outcome(s) that will be studied (e.g., overall survival [5], progression-free survival [6]). Ultimately, the PICO helps authors of intervention reviews in formulating a specific and clear research question, upon which basis the inclusion and exclusion criteria for studies to be or not to be included in the review can be determined. The research question of a standard systematic review of interventions can be formulated, for example, as such: *"How effective is drug W (the **I**ntervention) compared to drug X (the **C**omparator) in improving Y (the **O**utcome) in people with indication Z (the **P**opulation)?"*. Ideally, the PICO and the review should be formulated and conducted, respectively, by a team of review authors with different expertise, e.g., one or more co-authors who have clinical expertise in the health topic of interest, and one or more co-authors with methodological expertise in conducting systematic reviews. Moreover, different stakeholders such as people affected by the condition of interest should be involved in the review process (3).

The *systematic* in a systematic review

A *systematic* review requires the methods used to be formulated a priori, to be explicit and reproducible (2); readers of a well carried-out systematic review should be transparently informed about the methods used. For every systematic review in planning, there are fundamental steps in the development process that should be followed. In doing so, authors of systematic reviews can, in good conscience, present a review that has been conducted using gold standards and, thereby, provide reliable results about the effects of an intervention (2). Therefore, authors of a future review should register the title of their proposed review and publish a protocol (within the Cochrane Library for a Cochrane systematic review in planning) that fully stipulates

[5] Overall survival (OS) is a direct measure of clinical benefit and universally accepted in clinical research and practice. It is defined as the duration of patient survival after initiation of treatment.
[6] Progression-free survival (PFS), also a universally accepted clinical endpoint, defined as the time from treatment initiation until disease progression or relapse.

the methods that will be used to conduct the review; or, for reviews planned to be published in another scientific journal, prospectively register the review on the international Prospective Register of Systematic Reviews[7] (PROSPERO). Registering a planned review is not only relevant for being transparent about its methodology, but also for ensuring the individuality of one's own work by minimising the risk for unplanned duplication of the same work. It has been a known problem that much research in form of systematic reviews is repetitive as different investigators have examined the same research question(s), arguing that their work is different from others' because it uses other methods or examines different outcomes (6).

Figure 1 broadly outlines the key steps in the methodology of a systematic review. Future review authors should use the Cochrane Handbook for Systematic Reviews of Interventions (3) (irrespective of whether the review is planned to be published in the Cochrane Library, or in another scientific journal), where the methodology for the conduct of different types of systematic reviews is described and explained in detail. As shown in Figure 1, the key steps include, firstly, the definition of the research question that involves formulating the PICO. Based on the PICO, the inclusion criteria for studies to be included in the review, more specifically the types of studies (e.g., randomised controlled trials [RCTs]), types of interventions (e.g., a novel therapeutic intervention), types of comparators (e.g., a placebo) and types of participants (e.g., adults with a specific condition) are set. Concurrently, exclusion criteria for studies not suitable for answering the research question are also defined. Secondly, a systematic electronic search for relevant literature (that is, primary research studies to be included in the review) in medical databases such as CENTRAL[8] and MEDLINE[9] needs to be conducted by an information specialist[10]. Information specialists develop search strategies[11] that are adapted to the PICO of a review and to each database where studies will be searched for. Thirdly, using the inclusion and exclusion criteria, the results of the literature search are screened using the following steps: 1) title and abstract screening of references, a rather broad first step in selecting potential studies that broadly fit the research question, while excluding studies that do not address the PICO in any way, followed by 2) full-text screening, a stringent and final selection of studies for the review, by obtaining and reading the full-text article of each previously included reference and applying the inclusion and exclusion criteria to decide upon final inclusion or exclusion of a study into the review. Fourthly, extracting relevant data from the included studies, including study characteristics and statistical results for the outcomes of interest that were assessed in the study. Fifthly, using tools such as the Risk of Bias 2.0[12] (9) tool, the methodological quality of each study included in the review needs to be assessed.

[7] PROSPERO is an open access online registry where authors can register different types of reviews, thereby providing an overview of systematic reviews in development (www.crd.york.ac.uk/prospero/).
[8] The Cochrane Central Register of Controlled Trials (CENTRAL) contains journal articles of randomised and quasi-randomised controlled trials (www.cochranelibrary.com/central/about-central).
[9] MEDLINE is the National Library of Medicine's (NLM) primary bibliographic database containing journal articles in life sciences and biomedicine (www.nlm.nih.gov/medline/medline_overview.html).
[10] An information specialist is a person trained in designing and running search strategies in electronic databases.
[11] To systematically search for relevant studies, a sensitive search strategy including search filters (when available) needs to be developed by experienced medical and health care librarians or information specialists.
[12] Risk of bias 2.0 is a revised Cochrane tool for assessing the methodological quality of RCTs.

There are different tools appropriate for different types of studies to assess their risk of bias. Sixthly, synthesise the results extracted on the outcomes of interest, if possible, quantitatively in a meta-analysis. A statistical pooling of results in form of a meta-analysis is not always feasible or recommended (for example, in case of high heterogeneity between studies or uncertainty in the published data). If pooling of individual study results is not possible, study results should be reported and synthesised narratively. Lastly, the certainty in the evidence needs to be assessed by using the GRADE[13] approach. This is particularly important when the evidence from the systematic review shall be used for developing recommendations within clinical practice guidelines; this is described in the sub-chapter *How systematic reviews can inform clinical practice* on page 23.

Upon completion of these steps, write the review and report its results.

Figure 1. Key steps in the conduct of a systematic review

To minimise bias in the review process and to avoid random errors, most steps of the review process should be conducted by two review authors independently. For example, for step five (the risk of bias assessment), two review authors should, firstly, independently assess the risk of bias in each study included in the review. Secondly, the review authors should compare their individual assessments. Identical judgements can then be summarised; conflicting judgements should be discussed and, if necessary, a third review author involved in the discussion to reach a consensus and final judgement.

Pairwise meta-analysis and network meta-analysis

Pairwise meta-analysis

A systematic review summarises the evidence from primary research studies. Ideally, this will include a quantitative, statistical synthesis in a *pairwise* meta-analysis. The greatest benefit from a meta-analysis is that when there have been many individual studies conducted to assess the same intervention, and if these studies provide

[13] The Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach is used to grade the certainty in the evidence. It is mostly used in Cochrane reviews as well as in clinical guidelines to assess the strength of the recommendations for clinical practice.

conflicting evidence on its effectiveness, combining all individual results into one meta-analysis can provide insightful results and help reach a conclusion.

A meta-analysis is conducted by, firstly, extracting relevant data on the outcomes of interest (e.g., extracting the number of people randomised, the number of participants experiencing the event of interest, and the corresponding effect estimate). Secondly, the results from the studies on the outcome of interest are pooled in a meta-analysis, provided that the studies are similar enough (e.g., in terms of participants and interventions). Moreover, the interventions and comparators need to be comparable across studies that are combined in the same meta-analysis. A pairwise meta-analysis aims to compare one intervention to another intervention (the latter is usually called the *comparator*). In Figure 2, this is illustrated using a simple example. Results from multiple studies where participants received the same intervention (here: ibuprofen) are compared to participants who have received the same comparator (here: diclofenac); the outcome of interest being adverse events[14]. To further ensure similarity and comparability between studies, review authors should consider additional study and participant characteristics. For example, did the participants in the included studies suffer from the same condition? Did they all suffer from a chronic, or an acute condition? Did participants receive additional medication, for example for co-morbidity? Were administration routes, frequencies of administration and drug doses similar in participants receiving the same intervention and participants receiving the same comparator? Did participants receive the drugs for the same length of time?

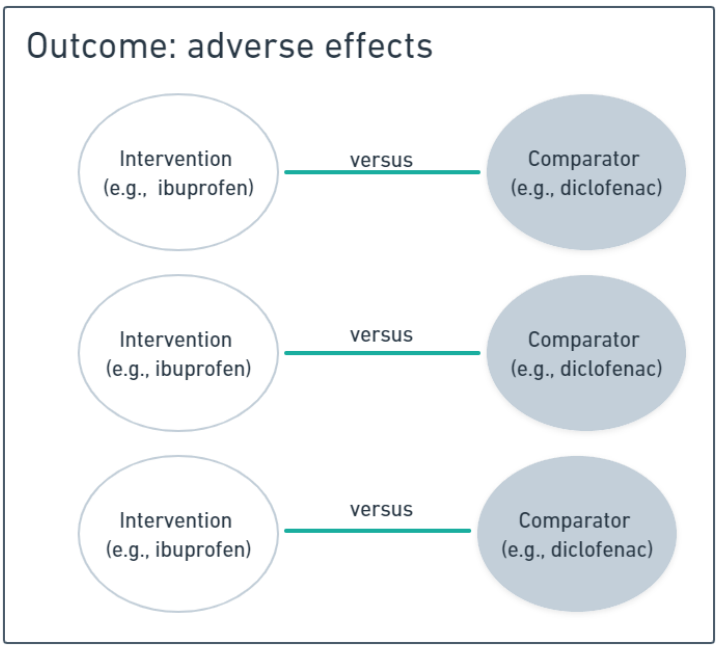


Figure 2. Illustration of a simple pairwise meta-analysis

[14] Adverse event: an unexpected event (e.g., fatigue, or nausea) that occurs during treatment that may or may not be related to the treatment itself. An adverse event can be mild, moderate, or severe.

In Figure 3, the same example is illustrated in a forest plot[15] using fictional data to demonstrate how a simple pairwise meta-analysis of ibuprofen versus diclofenac. In this fictional example, adults suffering from a condition causing pain were investigated in the studies included in this analysis (in the figure named as studies 1, 2 and 3). In all three RCTs, participants were randomised to receive either ibuprofen (e.g., 2x1600mg/day) or diclofenac (e.g., 2x50mg/day) to reduce pain; the outcome of interest being adverse effects. Across studies, participants received the intervention or comparator drug twice daily for a duration of 10 to 15 days, at the same dosage, and the drugs were administered orally. The results of this fictional meta-analysis are presented as a risk ratio[16].

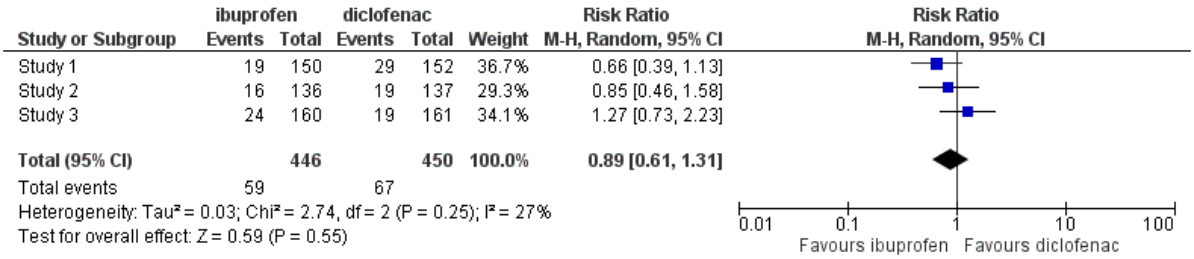


Figure 3. Example of a pairwise meta-analysis (data are fictional)

In this fictional pairwise meta-analysis, three RCTs with a total of 896 participants were included. Across studies, 59 events (i.e., adverse events) in total occurred across the intervention groups (participants receiving ibuprofen) and 67 events in total occurred across the comparator groups (participants receiving diclofenac). The individual results of study 1 and study 2 show that less adverse events occurred in participants who received ibuprofen, whereas in study 3, less adverse events occurred in participants who received diclofenac. This is also reflected in the individual effect estimates (the risk ratios) going into different directions in studies 1 and 2 versus study 3. The overall result of this meta-analysis that pooled the individual effect estimates shows a risk ratio of 0.89, with a 95% confidence interval[17] of 0.61 to 1.31. This means that participants receiving ibuprofen have a 11% lower risk to experience adverse events. However, the range of the confidence interval is relatively wide and includes the value "1", which in statistical terms means that there is no *significant* difference between the ibuprofen and diclofenac. This is, again, reflected in the individual study results that show contradictory results for ibuprofen versus diclofenac. Lastly, there is moderate heterogeneity between studies, which is reflected in the I^2 [18] of 27%. It could be that there are important clinical differences between participants. For example, some participants may suffer from a chronic condition, whilst others experience an acute condition causing pain. In addition, it could be that

[15] A forest plot is a visual illustration of a meta-analysis.
[16] A risk ratio (RR) is a measure of the risk of a certain event to occur in one group (the intervention group) compared to the risk of the same event to occur in another group (the control group). A RR that equals 1 means that the risk is the same between the two groups. A RR of < 1 indicates a decreased risk for the intervention group compared to the control group. A RR of > 1 indicates a decreased risk for the control group.
[17] A 95% confidence interval (CI) represents a range of values where there is a 95% chance that the true effect will lie within this range.
[18] The I^2 is a statistic that reflects the variation (i.e., heterogeneity) across the pooled studies in percentages.

some participants received other medications (in cases of comorbidity), too. Moreover, some participants in one study may have received the drugs for a duration of 10 days, whilst participants of the other studies may have received the drugs for up to 15 days. Hence, it is important to look out for clinical or methodological differences between studies to minimise heterogeneity. If there is too much variation between studies, it might be inappropriate to combine them in the same meta-analysis.

Network meta-analysis

In the previous sub-chapter, it was described how a pairwise meta-analysis works, where one experimental intervention (intervention A) is compared to one comparator intervention (intervention B), and the overall difference in effects between these two interventions using evidence from multiple similar and comparable studies is estimated. Imagining that there is a third intervention (intervention C) that in another study was compared to intervention B, one could, again, conduct pairwise meta-analysis and calculate the effects of intervention A versus intervention B in one analysis, and the effects of intervention B versus intervention C in a second analysis. However, there might be no such evidence from studies on the comparative effectiveness of intervention A versus intervention C. In addition, even if there would be evidence available for all three comparisons (A versus B, B versus C, and A versus C), with so many interventions, the question arises which intervention is the best regarding the outcome of interest: A, B, or C? Individual meta-analysis of each comparison cannot answer this question. To answer this question, all available interventions need to be combined in one single analysis to assess the comparative effectiveness of these competing interventions. This aim is what constitutes a *network* meta-analysis (10).

Network meta-analysis is a method that allows to combine multiple interventions into one analysis and thereby, compare all interventions using what is called *direct* and *indirect* evidence (10-13). Moreover, it is called a *network* meta-analysis because a network of all available interventions is created (11). There are two advantages in using a network meta-analysis compared to a pairwise meta-analysis. Firstly, clinicians and patients are usually not only interested in how one therapeutic drug performs in comparison to another therapeutic drug, but they are also interested in which therapeutic drug works best (i.e., is the most effective and tolerable) when there are multiple different drugs to choose from for treatment (11). Secondly, not all interventions have been compared to one another in clinical trials; this scenario is illustrated in Figure 4. The green lines illustrate that there is direct evidence from one or more studies on the comparisons A versus B and B versus C, while the dashed red line indicates that the comparison A versus C has not been investigated in any study. This means that it is not possible to conduct a pairwise meta-analysis for intervention A versus intervention C as there are simply no studies that investigated this comparison. This is where a network meta-analysis comes into play.

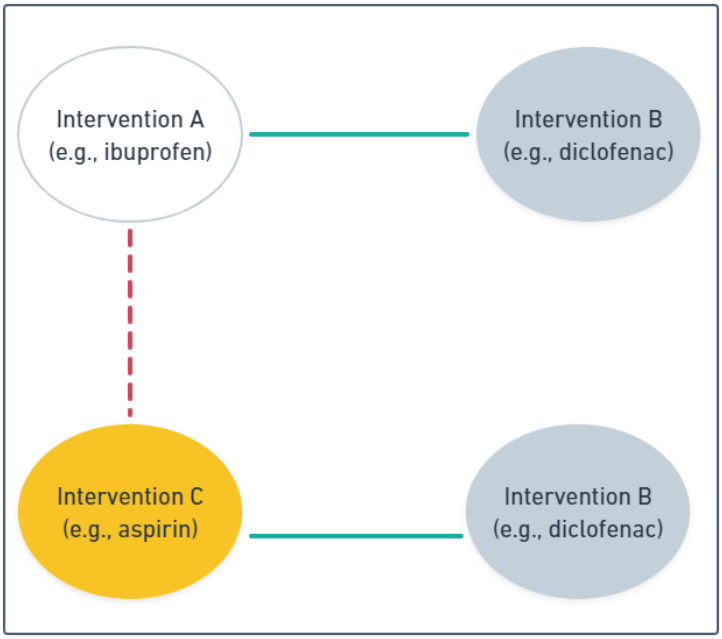


Figure 4. Illustration of a scenario where there is direct evidence for the comparison intervention A versus intervention B, and intervention B versus intervention C. There is no direct evidence for a comparison between intervention A and intervention C

A network meta-analysis can provide information about comparisons of interventions that have not been investigated in a clinical study by using *indirect* evidence (12). So-called indirect evidence is created by estimating the relative effect of intervention A versus intervention C by using the direct evidence that was obtained from the studies comparing intervention A versus intervention B and intervention B versus intervention C. Simply put, the effect estimates from the direct comparisons of A versus B are subtracted from the direct comparisons of B versus C to obtain the indirect treatment effect of A versus C (12). To compare two interventions to each other for which there is no direct evidence available, they need to have one comparator in common. In Figure 5, the common comparator of A and C is B; this is illustrated by the blue (and green) lines.

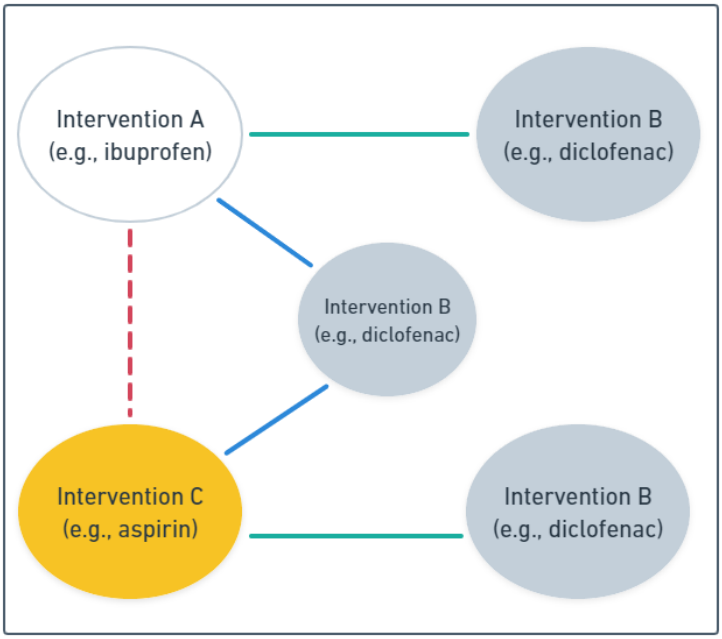


Figure 5. Illustration of a scenario where intervention A can be compared to intervention C by the common comparator intervention B

There may be a great number of different drugs or interventions available for a certain condition. Again, clinicians and patients are faced with deciding about the best treatment option. In Figure 6, a network of interventions is shown where there is direct evidence (green lines) for some comparisons, and indirect evidence (red dashed lines) for other comparisons. In this network, all interventions are compared to each other, one way and/or another by using direct and/or indirect evidence. For example, while there is no direct evidence to compare intervention A with intervention D, one could compare these two indirectly by using the evidence from the direct routes of A versus B and B versus D, with intervention B being the common comparator. The same can be seen in the indirect comparison of B versus E: from B versus D to D versus E, the common comparator being intervention D. Such comparisons, or routes, are called *simple indirect evidence*, as there is only one intermediate treatment (13). However, there is also *compound indirect evidence*, which is based on more than one intermediate treatment (13). Such an example would be a comparison of intervention A versus intervention E. These two can be compared by using the route of A versus B, B versus D, and D versus E; here, there are two intermediate treatments, namely B and D.

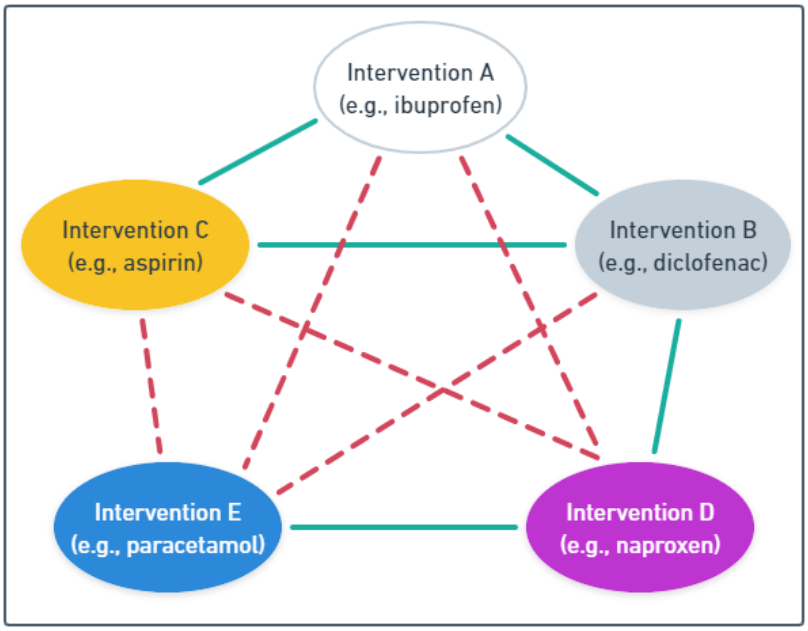


Figure 6. Example of a network with several direct (green) and indirect (red) comparisons

Ideally, a network contains so-called *mixed evidence* (13) where there is both direct and indirect evidence available for intervention comparisons, thereby creating a *closed loop* (see Figure 7). When mixed evidence is available, network meta-analysis can simultaneously analyse all direct and indirect evidence available for all comparisons in a network (13). Hence, depending on the overall evidence base, the results from network meta-analysis may be derived from direct evidence alone, indirect evidence alone, or from mixed evidence (13).

Criteria for conducting a network meta-analysis

To conduct a valid network meta-analysis (12), the following three *assumptions* need to be met beforehand, of which the first two also apply to a pairwise meta-analysis:

- ✓ *Transitivity:* the assumption that the studies in the network are similar enough regarding clinical and methodological factors (12). Firstly, the common comparator in comparisons such as intervention A versus intervention B and intervention A versus intervention C, where intervention A is the common comparator, should be similar (for example, in terms of administration route or dosage of a substance). Secondly, other effect modifiers such as the characteristics of the participants (for example the sex, age, and clinical disease stage) should be similar. Transitivity can be assessed by visually inspecting the distribution of these effect modifiers across comparisons and studies (12).
- ✓ *Homogeneity*: the assumption that there is no heterogeneity in the results between trials included in the pairwise comparisons of a network (12). This can be assessed statistically by using the I^2 statistic as well as by visually inspecting the forest plot.

- ✓ *Consistency*: the assumption that there is agreement between the direct and indirect evidence about the comparisons under study (12, 14). Inconsistency occurs when the available direct evidence and the estimated indirect evidence disagree in their effect estimates of a comparison. Consistency can be checked in so-called *global* and *local* approaches (11). In the global approach, the consistency of the effect estimates of the different comparisons (from direct and indirect evidence) across the entire network is checked using statistical measures. In the local approach, consistency within a closed loop (see Figure 7) in a network is checked by examining whether the effect estimates of direct and indirect comparisons agree (11). In a closed loop, each intervention has been compared directly with both other interventions (here: A with B, A with C and B with C) (14). Thus, each direct comparison can be supplemented by an indirect comparison.

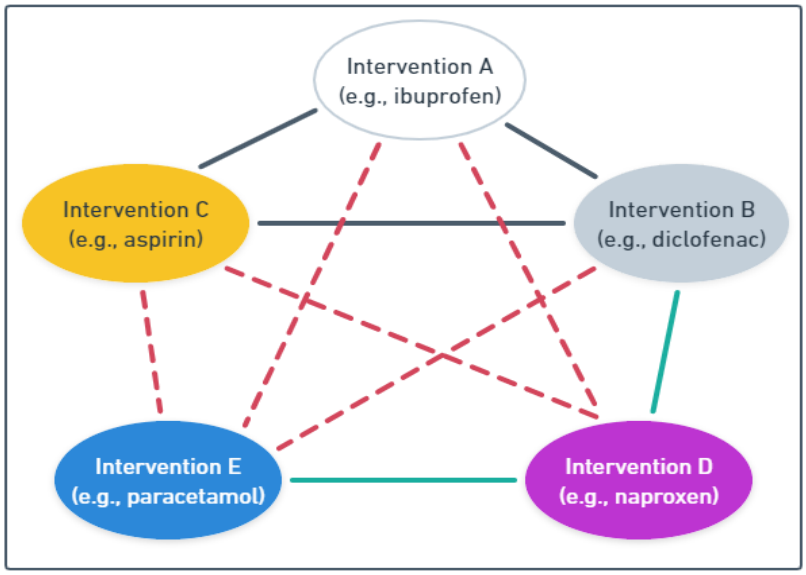


Figure 7. A network with a closed loop (A - B - C)

The concept of *living* systematic reviews

A systematic review may be updated at some point after its publication. According to the methods suggested in the Cochrane Handbook (15), published systematic reviews should be assessed over time to identify whether an update of the review is needed. This can be decided by determining the (continuing) relevance of the review question. In addition, there may be new evidence (i.e., new study results) that could be incorporated and that could have an impact on the review and change its findings (15). However, updating a review particularly when years have passed after its initial publication is time and resource constraining. To review authors, it is often associated with a high workload comparable to the conduct of a new review from the beginning (16).

Therefore, *living* systematic reviews (LSR) have become a new relevant approach around the systematic review development. The aim of conducting a living systematic review is to frequently update the review to incorporate newly identified evidence as soon as it becomes available (16, 17). This is particularly important in clinical areas where new research evidence emerges rapidly and may influence treatment recommendations and/or health policies (16, 17). It can also be important in areas where evidence is sparse or of low certainty, and new evidence is continuously searched for to update a review and inform clinical practice as soon as possible (16). Living systematic reviews differ from regular systematic review updates as the latter are not usually planned for and may be conducted years after the initial review has been published, leaving a large gap between the old and the new evidence (16). Upon planning a living systematic review, however, review authors are determined and committed to frequently update the review, and the frequency of updates and the methods to do so are pre-defined (16, 18). A living systematic review can be updated, for example, during a 12-month period after the publication of the baseline systematic review (i.e., the first version of the systematic review). To decide whether or not a living systematic review approach should be applied to a review, Simmonds and colleagues (16) propose three criteria for consideration:

- the systematic review is important to evidence-based health decision-making;
- the certainty in the existing evidence is (very) low and new evidence may change the findings of a review;
- the research field in the topic area of interest is rapidly evolving and continuously producing new evidence (16).

To conduct a living systematic review, the standard methodology of systematic reviews (see Figure 1 on p. 10) will be followed and repeated, and any form of systematic review can be a living review (16). During the development of the protocol for a new living systematic review, review authors need to take decisions about the living systematic approach and clearly outline how frequently new studies (i.e., new evidence) will be searched for (15). For example, monthly literature searches in databases and trial registries could be conducted, while grey literature may be searched for bi-monthly. With every literature search, the process of screening and selecting eligible studies is repeated. However, Simmonds and colleagues (16) argue

that newly identified evidence does not necessarily need to be incorporated into a review immediately. In fact, inclusion into the review may be deferred until it is obvious that the new study results will change the reviews' conclusions (16). If new studies are identified but results not immediately incorporated into the review, readers of a review should also be transparently informed about the pragmatic decisions that have been taken to defer inclusion of the newly identified studies (16). Once review authors decide to incorporate the newly identified studies into the review, the standard systematic review development process is repeated (i.e., data extraction, risk of bias assessment, data analysis, and so forth; see Figure 1 on p. 10) (16). Simmonds and colleagues further argue that, even if no new studies are identified, readers of a review should be informed about the date of the last search and that no new evidence has been identified (18).

Systematic reviews of prognosis research studies

Prognosis in health sciences

The term *prognosis* can be described as foreseeing that a certain event, outcome, or condition will occur or develop in the future (19, 20). Probably the most common form of prognosis is the weather forecast. Meteorologists use several *prognostic factors* such as the current temperature, humidity level, wind speed and wind direction, that in combination form a *prognostic model* that allows foreseeing how the weather will develop, for example, during the upcoming seven days. Accordingly, the weather forecast is an aid to us when taking decisions about our attire and future activities.

In medicine and health sciences, prognosis is concerned with the probability that an individual will develop a certain health outcome (e.g., a certain type of cancer), or experience a health-related event (e.g., death), after a certain period. Prognosis can be made in healthy individuals who may be at risk to fall ill or develop a certain health outcome. Likewise, a prognosis about the occurrence of a particular event can be made for individuals who are already ill. Hence, different health-related outcomes can be addressed in prognosis research: the risk of developing a certain disease, the risk of, for example, death due to an already present disease, or the risk of disease progression or relapse (19, 20).

Health-related prognosis can be best studied in people who have not been treated (for a certain illness or health condition) before, as this allows for studying the natural course of this disease or condition. Alternatively, it can be studied in people who have received the same treatment regimen but without treatment modification during the treatment (19, 20). Ultimately, prognosis can be best studied in a cohort presenting similar characteristics under the same conditions during the study period, so that the occurrence of the specific outcome or event predicted can be attributed to the prognostic factor(s) under study (21).

Types of prognosis research studies

To better understand the different ways of prognosis in health sciences, the different types of prognosis research need to be explained. There are four types of prognosis research studies.

Overall prognosis research

In overall prognosis research, also termed *fundamental* prognosis research (19), the following question is raised: "*What is the likely outcome of people with a particular health condition under current clinical practice?*" (19). In this type of prognosis research, prognosis is studied in a cohort of people with similar characteristics, for example all presenting the same condition (e.g., men with prostate cancer), and by using knowledge about current medical care standards for this specific condition. For example, one could prognosticate the overall survival of people with this specific health condition of interest (19). An example of a question about overall prognosis could be: *"What is the average five-year survival of men with prostate cancer?"*

Prognostic factor research

In prognostic factor research, the following question is raised: *"Which factor can be associated with a specific health outcome for people with a specific health condition?"*. This type of prognosis does not simply provide an average estimate about the likelihood (or risk) of an outcome to occur. It rather aims to provide a more specific prognosis using a specific prognostic factor that may be associated with a specific health event or outcome. A prognostic factor can be a clinical factor, such as a characteristic of the disease or health condition under study. For example, in men with prostate cancer, an important prognostic factor is the tumour volume, which is associated with survival prognosis (22). It can also be a non-clinical factor, a patient demographic such as age or sex (19, 20).

Prognostic model research

In prognostic model research, the following question is raised: *"What is the individual risk for a certain health outcome, based on a combination of different prognostic factors?"* (21). This type of prognosis research makes use of a combination of prognostic factors that previously, in prognostic factor research, have been identified to have some prognostic value (23). The added value of a prognostic model, in comparison to an individual prognostic factor, is that the model can provide a more accurate estimate on the likelihood of an individual to develop an outcome or to experience a certain event. In addition, prognostic models can be adapted individually by using multiple factors specific for an individual (e.g., age, sex and the disease stage) and thereby provide an individual prognosis for each person affected (23). Using the example of prostate cancer, clinically relevant prognostic factors are the Gleason score (a grading system), the disease stage of the cancer according to the TNM Classification of Malignant Tumours (TNM), and the residual tumour (R) classification (24). A prognostic model may then be adapted individually by combining these known prognostic factors and assessing them for every individual.

Stratified medicine research

Lastly, stratified medicine is the application of prognostic information to take an informed therapeutic decision for an individual with a specific characteristic, or a group of individuals that have a specific characteristic in common (25). In other words, treatment regimens are adapted to the individual prognosis of a person affected to reach the best health outcome possible. This can be achieved by using prognostic information, such as an established prognostic model containing different prognostic factors that altogether can prognosticate a certain health event or outcome. The aim of stratified medicine is to provide the best adapted treatment for an individual patient to prevent the occurrence of a specific health event or outcome. Hence, prognosis research is fundamental for stratified medicine as it can aid clinicians and patients in deciding on the most appropriate treatment to achieve the best health outcome possible (25).

Conducting a systematic review of prognostic factor studies

The steps that are outlined in Figure 1 (p. 10) can be applied one to one to a systematic review of prognostic factor studies, with one exception. In systematic reviews of prognostic factor studies, the *PICOTS* is formulated. The acronym **PICOTS** stands for:

- **P**opulation,
- **I**ndex prognostic factor,
- **C**omparator prognostic factor,
- **O**utcome,
- **T**iming, and
- **S**etting (26).

The difference to the PICO for a systematic review of interventions (p. 8) are the **I**, **C**, **T,** and **S**. Instead of an intervention, one investigates an index (I) prognostic factor, which is compared to a comparator (C) prognostic factor. In addition, there is a timing (T) component, which is essential to prognostic factor research. Firstly, it must be determined at which time point the prognostic factor under study will be measured. Secondly, it must be determined for which point in time the outcomes are to be predicted by the prognostic factor(s). Lastly, the setting (S) of where the index prognostic factor will be used has to be determined (26).

Moreover, there are tools specifically developed for the conduct of systematic reviews of prognosis studies. For example, to assess the risk of bias in prognostic factor studies, the recommended tool is the Quality In Prognosis Studies (QUIPS) tool (27). Generally, future review authors of systematic reviews of prognosis studies should use the PROGRESS series (19) as a helpful guide.

As for meta-analysis, the pairwise meta-analysis approach (p. 10) is primarily used to pool and analyse outcome data from prognostic factor studies.

How systematic reviews can inform clinical practice guidelines

In vast developing clinical areas, where different treatment options become increasingly available, it is essential to assess the comparative effectiveness, including the benefits and harms, of the many alternatives (4). It becomes increasingly challenging to patients and health care practitioners to find the most effective and appropriate treatment option. While a systematic review summarises all available evidence on the same intervention or health care service, it can only provide evidence about the effectiveness of the specific intervention or health care service under specific and ideal circumstances (4). From a clinical practitioner's point of view, however, the evidence is of most relevance when it is applicable to the real clinical world, too (4). Moreover, while systematic reviews can provide information about effectiveness, they do not intend to provide recommendations for or against the use of a health care service in clinical practice (4). Instead, the findings from a systematic review can be used to inform *recommendations* in clinical practice guidelines (4, 28).

Clinical practice guidelines are developed to summarise all available evidence on a specific health condition, assess the quality of the evidence and ultimately, formulate recommendations for the diagnosis, treatment and follow-up of that condition in an understandable and practical way to aid health care decision-making in daily clinical practice (4, 29). They are systematically developed guidelines that entail specific recommendations for patients and their treating physicians on the use of different health care services (e.g., a diagnostic tool or a specific treatment) in a specific clinical situation or circumstance (29). However, they do not intend to replace individual judgements that are dependent on individual characteristics or circumstances, but they can, as the word itself indicates, *guide* health care decisions, for example when it comes to choosing the most appropriate and effective health care service when there are many alternatives to choose from (4, 28). The U.S. Institute of Medicine (IOM) defines clinical practice guidelines as *"statements that include recommendations intended to optimise patient care that are informed by a systematic review of evidence and an assessment of the benefits and harms of alternative care options"* (5). The development of clinical practice guidelines is initiated by professional societies, who gather all important stakeholders to develop a clinical practice guideline (4). This involves a multidisciplinary panel of different stakeholders, including clinical experts, patients and patient representatives, and methodologists experienced in the relevant research methods needed to evaluate the available evidence and translate it into recommendations (4, 29, 30).

In the area of oncology, there are different international societies that develop clinical practice guidelines, such as the American Society of Clinical Oncology (ASCO), the European Society of Medical Oncology (ESMO), the National Comprehensive Cancer Network (NCCN), or the German Guideline Program in Oncology (GGPO), to name a few. The guidelines developed within these societies mainly differ in their methodology in that some societies produce more evidence-based guidelines than others (see next section Recommendations

Most importantly, the guidelines are only applicable in the respective countries. For example, the GGPO creates guidelines that are solely used in Germany. It can be the case that some treatments that are approved and therefore recommended in the U.S., may not be approved by the relevant authorities in Germany. Hence, the recommendations in these different international guidelines can differ to some extent, even though addressing the same clinical topic.

Recommendations

Ideally, a high-quality clinical practice guideline entails recommendations that are based on the available evidence (so-called evidence-based recommendations), summarised and synthesised in high-quality systematic reviews (4, 5, 28, 30). If no systematic reviews are available, the guideline panel can apply systematic methods to synthesise evidence from individual studies that address the underlying research question(s), and thus, evaluate the identified evidence using systematic review methods (30). However, not every available health care service has been assessed, or every research question addressed, in medical research. In cases where no evidence is available, or when the few available evidence is of low or very low quality, recommendations are formulated based primarily on the consensus of the involved clinical experts and their expert opinions (so-called consensus-based recommendations) (31). Important to note is that both forms of recommendations require group consensus in the final development of the recommendations, irrespective of whether there is an evidence-base or not (31). Hence, it is crucial that the guideline panel, consisting of clinical and methodological experts who formulate recommendations, is unbiased and free from any conflicts of interest (4, 32).

There are two underlying criteria to each recommendation that is being developed: the certainty in the underlying evidence, and the strength of the recommendation (4). The certainty in the evidence reflects the validity of the studies identified to answer the underlying research question, and the certainty that the health outcome(s) of interest addressed in that specific question will be reached (4). Rating the certainty in the evidence is commonly done using the GRADE approach, as in systematic reviews (30). The strength of a recommendation depends on the level of certainty in the underlying evidence, the outweighing of benefits and harms of the health care service in question, its clinical effect(s), patients' preference(s) as well as its applicability and feasibility in clinical practice (4, 30). The strength of a recommendation also clearly indicates how important it is to adhere to the recommendation (4). Recommendations may advise for or against the application of a health care service (33).

Cancer – a public health burden

To date, cancer is one of the leading causes of death worldwide (34). In 2019, the World Health Organization (WHO) ranked cancer as the first or second leading cause of death in 61% of countries, and the third or fourth leading cause of death in 12% of countries, in people under the age of 70 years (34). In 2020, there were around 19.3 million new cancer cases in males and females, and around 9.9 million deaths across both sexes worldwide, according to Globocan[19] (34, 35). Worldwide, males had a 19% higher incidence of cancer, and 43% higher death rates compared to females. However, these numbers vary in different regions of the world, and depend on the distribution of the different cancer sites (34). As for the most common cancer sites in 2020, female breast cancer was the most common cancer site, followed by lung cancer, as each had an incidence of more than two million new cases (breast cancer making up 11.7%, and lung cancer making up 11.4% of all cancer cases) (34). Lung cancer remains the leading cause of death among all cancers, with 1.8 million new deaths in 2020 (34). This is particularly the case for males, whereas for females, breast cancer is the leading cause of mortality (21). Across both sexes, the highest incidence in cancer cases was observed in Asia, followed by Europe, Northern America, Latin America and the Caribbean, Africa, and lastly, the Oceania (35).

Different social and economic factors contribute to this increase, combined with a growing and ageing population (34). Statistics show that incidence rates for cancer increase particularly when the Human Development Index[20] (HDI) level in a country increases, for both males and females (34). Moreover, cancer mortality in males is about two-fold higher in countries with a higher HDI level compared to countries with a lower HDI. However, little variation across HDI levels was observed for females in terms of cancer mortality (34). To address this growing burden and to inform diagnosis, treatment, and follow-up of cancer (in fact, any illness or health condition), trusted clinical practice guidelines are required.

[19] Globocan is a visualisation of cancer statistics by the Global Cancer Observatory (GCO), a web-based platform for global cancer statistics (https://gco.iarc.fr/).

[20] The Human Development Index (HDI) is a metric that measures three dimensions of human development: 1) a long and healthy life (e.g., measured by life expectancy at birth), 2) knowledge (e.g., measured by (mean) years of schooling) and 3) a decent standard of living (e.g., measured by the Gross National Income (GN) per capita) (https://hdr.undp.org/data-center/human-development-index#/indicies/HDI).

CHAPTER 3

OBJECTIVES

Objectives

The primary objective of this cumulative dissertation was to conduct systematic reviews using new and complex systematic review methods to contribute to the development and refinement of these methods. The secondary objective was to conduct clinically relevant systematic reviews to provide meaningful evidence that may inform clinical practice in oncology.

This cumulative dissertation includes two Cochrane systematic reviews:

Systematic review I: First-line therapy for adults with advanced renal cell carcinoma: a systematic review and network meta-analysis

The primary aim of this systematic review with network meta-analysis was to evaluate and compare the benefits and harms of first-line therapies for adults with advanced renal cell carcinoma, and to produce a clinically relevant ranking of therapies (36). The secondary aims were to maintain the currency of the evidence by conducting continuous update searches, using a living systematic review approach, and to incorporate data from clinical study reports. In the main part of this dissertation, a visual abstract of the review is provided on p. 30, and a written summary (including text excerpts from the review) with main results is provided on p. 31. The original manuscript of the systematic review as published in the Cochrane Library can be found in Appendix i (Systematic review I).

Systematic review II: Interim-PET scan results for prognosis in adults with Hodgkin lymphoma: a systematic review and meta-analysis of prognostic factor studies

The aim of this systematic review with meta-analysis was to determine whether in previously untreated adults with Hodgkin lymphoma receiving first-line therapy, the interim positron emission tomography scan result can distinguish between those with a poor prognosis and those with a better prognosis, and thereby predict survival outcomes in each group (37). A visual abstract of the review is provided on p. 37, and a written summary (including text excerpts from the review) with main results is provided on p. 38. The original manuscript of the systematic review as published in the Cochrane Library can be found in Appendix ii (Systematic review II).

CHAPTER 4

DISSERTATION PROJECTS

Dissertation projects

Systematic review I: First-line therapy for adults with advanced renal cell carcinoma: a systematic review and network meta-analysis

In the main body of this cumulative dissertation, a visual abstract (p. 30) and an extensive summary (p. 31) of the review *First-line therapy for adults with advanced renal cell carcinoma: a systematic review and network meta-analysis* (36) are provided. The summary entails sentences and text excerpts from the original manuscript, which can be found in Appendix i (Systematic review I).

Citation of the published article:

Aldin A, Besiroglu B, Adams A, Monsef I, Piechotta V, Tomlinson E, Hornbach C, Dressen N, Goldkuhle M, Maisch P, Dahm P, Heidenreich A, Skoetz N. First-line therapy for adults with advanced renal cell carcinoma: a systematic review and network meta-analysis. Cochrane Database Syst Rev. 2023 May 4;5(5):CD013798. doi: 10.1002/14651858.CD013798.pub2. PMID: 37146227; PMCID: PMC10158799.

Visual abstract

First-line therapy for adults with advanced renal cell carcinoma

Systematic review with network meta-analysis

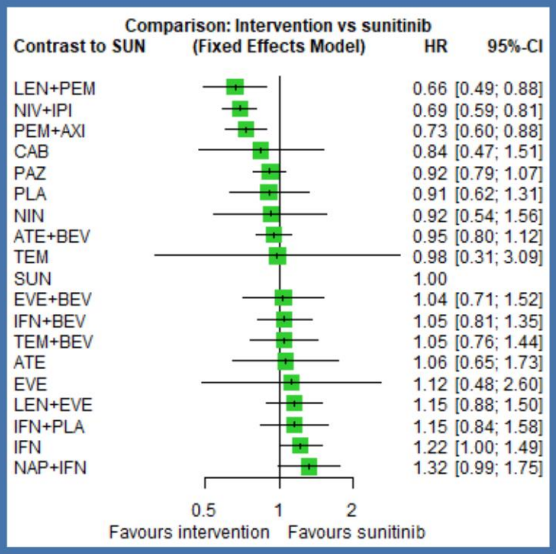
N=36 randomised controlled trials

N=15,117 adults with advanced renal cell carcinoma

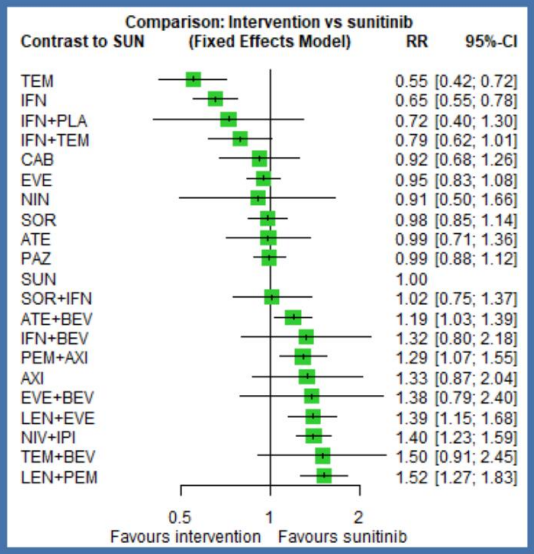
First-line therapies (immunotherapy, targeted therapy)

Overall survival, quality of life, serious adverse events

Overall survival



Serious adverse events



Quality of life

Trial	time point of measurement (in months)	Intervention (N analysed)	Intervention post mean score (SD)	Comparator (N analysed)	Comparator post mean score (SD)
Scale: FKSI-DRS (score range 0-36; higher scores represent better QoL)					
NCT01108455	40	EVE (N=33)	26.6 (6.85)	SUN (N=47)	26.6 (6.13)
NCT00098657/NCT00083889	28.5	SUN (N=53)	29.44 (4.210)	IFN (N=351)	29.22 (7.694)
NCT00920816	24.6	AXI (N=72)	26.556 (5.487)	SOR (N=95)	26.786 (5.982)
Scale: EQ-5D-VAS (score range 0-100; higher scores represent better QoL)					
NCT00098657/NCT00083889	28.5	SUN (N=54)	76.85 (16.863)	IFN (N=352)	75.44 (25.060)
NCT00920816	24.6	AXI (N=71)	67.254 (19.495)	SOR (N=94)	67.048 (22.570)
Scale: FACT-G (score range 0-108; higher scores represent better QoL)					
NCT00098657/NCT00083889	28.5	SUN (N=52)	84.62 (16.257)	IFN (N=351)	79.54 (26.109)
Scale: FACIT-F (score range 0-52; higher scores represent better QoL)					
NCT00720941	38	PAZ (N=2)	38.5 (13.59)	SUN (N=2)	29.5 (0.71)

Summary and main results

This summary entails sentences and text excerpts from the original manuscript (see Appendix i (Systematic review I)).

Background

Description of the condition

In the U.S., kidney cancer accounts for 5% of all cancers in men and 3% of all cancers in women (38); in comparison, it occurs for 3.5% of men and 2.4% of women in Germany (39). The most common type of kidney cancer is renal cell carcinoma (RCC) (40). With a 2:1 ratio, RCC develops predominantly in men and commonly after the 60th year of life (41). Besides male gender and age, additional risk factors include an increased body mass index and active as well as passive smoking (39, 41-43). Comorbidities associated with an increased risk for developing RCC include hypertension, a history of kidney stones, or type 2 diabetes, amongst others (39, 41-43).

The most common subtype of RCC is the clear cell type (75%), followed by the papillary type (10%), and the chromophobe type (5%) (44, 45). Comparing all, the clear cell type is associated with the worst prognosis (44, 45). Individuals with advanced RCC are categorised into favourable, intermediate, or poor risk groups. These are the risk groups as defined by the International Metastatic RCC Database Consortium (IMDC) and the Memorial Sloan Kettering Cancer Center (MSKCC).

Description of the intervention

Before 2005, treatment options for advanced RCC were limited to immunotherapies such as the cytokine therapies interferon-alpha and interleukin-L. However, these are associated with many adverse events, and with partial or complete remission rates of approximately 12%, they benefit only a small percentage of participants (46). Nowadays, targeted therapies such as tyrosine kinase inhibitors, angiogenesis inhibitors, and immune checkpoint inhibitors have emerged as an effective alternative, and the benefit of standard targeted approaches, such as sunitinib or temsirolimus, over cytokine therapies regarding mortality, quality of life, and adverse events in advanced RCC has been indicated (47). Multiple drugs such as sunitinib, sorafenib, bevacizumab, nivolumab, pazopanib, axitinib, cabozantinib, and everolimus have been approved by the US Food and Drug Administration (FDA), mostly for second-line therapy, but several of them have been approved for first-line treatment as well.

Objectives

The primary objective of this systematic review with network meta-analysis was to evaluate and compare the benefits and harms of first-line therapies for adults with advanced RCC, and to produce a clinically relevant ranking of therapies.

The secondary objectives were to maintain the currency of the evidence and conduct continuous update searches by using a living systematic review approach as well as to incorporate data available from clinical study reports (CSRs).

Methods

Criteria for considering studies for this review

Population	Intervention	Comparator	Outcome(s)
- Adults (≥18) with a confirmed diagnosis of advanced RCC - Individuals without previous systemic anticancer therapy	Targeted therapy - Tyrosine kinase inhibitor (e.g., pazopanib, axitinib) - mTOR inhibitor (e.g., temsirolimus, everolimus) - Angiogenesis inhibitor (e.g., bevacizumab, lenvatinib) Immunotherapy - Checkpoint inhibitor (e.g., nivolumab, ipilimumab) - Interferon - Interleukin	Sunitinib	- Overall survival (OS) - Quality of life (QoL) - Serious adverse events (SAEs) - Progression-free survival (PFS) - Adverse events (AEs) - Number of participants who discontinued treatment due to an AE - Time to initiation of first subsequent anticancer therapy (TFST)

Methods for the conduct of this systematic review

In accordance with the methods recommended in the Cochrane Handbook for Systematic Reviews of Interventions (3), the following steps were performed to conduct this systematic review with network meta-analysis. More details on study selection, data extraction, dealing with missing data, assessment of heterogeneity, data synthesis (including direct and indirect comparisons in the network meta-analysis) and other information can be found in the full manuscript.

I. Electronic literature searches in relevant databases (i.e., CENTRAL, MEDLINE, Embase); conference proceedings (American Society of Clinical Oncology, European Society of Clinical Oncology); trial registries (ISRCTN, EU Clinical Trials Register, Clinicaltrials.gov, WHO ICTRP). The searches were conducted by an experienced information specialist. This step was followed by screening of search results and study selection based on the inclusion and exclusion criteria of the review.

   Living systematic review considerations

   A baseline search was conducted in February and October 2020. Starting from December 2020, monthly update searches until April 2021 were conducted. Together with the clinical experts on this review, it was decided to stop the update searches in April 2021 to finalise data extraction and risk of bias assessments. One final update search was conducted in February 2022, as searches for intervention reviews should not be older than 12 months at publication of the review.

II. Searching for Clinical Study Reports (CSR) parallel to the identified studies. CSR were searched for on the following data platforms: the European

Medicine Agency (EMA) clinical data platform (clinicaldata.ema.europa.eu/web/cdp/home), the Yale University Open Data Access (YODA) platform (yoda.yale.edu/), the Clinical Data Study Request (CSDR) platform (clinicalstudydatarequest.com), and the Vivli platform (vivli.org).

III. Data extraction (i.e., study characteristics, outcome data) using a data extraction form developed specifically for this review.
IV. Risk of bias assessment, using the Risk of Bias 2.0 tool (9).
V. Data analyses, conducted by a statistician experienced in network meta-analysis.
VI. Assessment of the certainty in the evidence (i.e. GRADE) using the methods for network meta-analysis as proposed by Salanti et al (2014) (13).
VII. Creation of the Summary of Findings (SoF) tables and writing the text of the review.

Except for the electronic searches (conducted by an experienced information specialist) and the statistical analyses (conducted by an experiences statistician), all steps in the review development process were conducted independently by at least two review authors. Discrepancies were resolved through discussion and by involving a third author.

Main results

Thirty-six RCTs with 15,177 female and male adult participants with advanced RCC from 53 countries were included. More information on the included studies and participants are reported in the manuscript. The results of the network meta-analyses presented in this summary are the main results for the primary outcomes in the combined risk groups. Separate results for the different risk groups as well as results for secondary outcomes and subgroup and sensitivity analyses can be found in the full manuscript.

Overall survival (OS)

For the combined risk groups, data from 21 trials with a total of 10,304 participants were analysed. The network was not fully connected and consisted of three subnetworks (see Visual abstract); subnetwork 1 included SUN as the main comparator. Moderate between-study heterogeneity was observed (Q=1.81, df=1, p=0.18; I^2=44.6%, Tau^2=0.0284). It was found that LEN+PEM may improve OS (HR 0.66, 95% CI 0.42 to 1.03, low certainty). The combinations NIV+IPI (HR 0.69, 95% CI 0.69 to 1.00, moderate certainty, P-score: 0.83) and PEM+AXI (HR 0.73, 95% CI 0.50 to 1.07, moderate certainty, P-score: 0.78) probably improve OS when compared to SUN alone (P-score: 0.47), respectively. It is uncertain whether CAB alone improves OS (HR 0.84, 95% CI 0.43 to 1.64, very low certainty, P-score: 0.63) when compared to SUN alone, and there is probably little or no difference in OS between PAZ alone (HR 0.91, 95% CI 0.64 to 1.32, moderate certainty, P-score: 0.57) and SUN alone. Comparison data was not available for AVE+AXI and NIV+CAB versus SUN, respectively (see Visual abstract).

Quality of life (QoL)

Analysing data was not feasible for QoL, so results were reported narratively. Results for the different time points are reported for every scale where data was extractable for. The time point of main interest in this review was QoL at the end of treatment (see Visual abstract). Regarding the long-term results (results at the end of treatment), all trials were overall judged to have a 'high risk of bias' mainly due to the outcome assessors' awareness of the assigned interventions, which is owed to the nature of self-reported questionnaires and due to the trials' study design (open-label, non-masked trials) as well as due to the high number of participants without outcome data at the end of treatment. In most comparisons including SUN, across all scales, participants in the experimental groups seemed to achieve a higher score in the post-intervention assessments compared to participants in the comparator arm (see Visual abstract).

One RCT measured QoL using FACIT-F (score range 0-52; higher scores mean better QoL) and reported that the mean post-score was 9.00 points higher (9.86 lower to 27.86 higher, very low certainty) with PAZ than with SUN. Comparison data was not available for PEM+AXI, AVE+AXI, NIV+CAB, LEN+PEM, NIV+IPI, and CAB alone versus SUN, respectively.

Serious adverse events (SAE)

Serious adverse events were not consistently reported across trials. To be able to meta-analyse results, SAEs were considered when the number of participants with at least one SAE was reported; cumulated events were not considered. Serious adverse events were assessed in 22 trials (18 two-arm trials, four three-arm trials), for a total of 10,709 participants. The network was fully connected, and substantial heterogeneity was observed in the network (Q_{total}=15.40, df=6, p=0.017; Q_{within}=3.44, df=1, p=0.064; $Q_{between}$=11.96, df=5, p=0.035; I^2=61.0%, Tau^2=0.0256). PEM+AXI probably increase slightly the risk for SAEs (RR 1.29, 95% CI 0.90 to 1.85, moderate certainty, P-score: 0.31), when compared to SUN alone (P-score: 0.59). The combinations LEN+PEM (RR 1.52, 95% CI 1.06 to 2.19, moderate certainty, P-score: 0.17) and NIV+IPI (RR 1.40, 95% CI 1.00 to 1.97, moderate certainty, P-score: 0.23) probably increase the risk for SAEs when compared to SUN alone, respectively. It is uncertain whether CAB alone reduces or increases the risk for SAE (RR 0.92, 95% CI 0.60 to 1.43, very low certainty, P-score: 0.65) when compared to SUN alone, and there is probably little or no difference in the risk for SAEs between PAZ alone (RR 0.99, 95% CI 0.75 to 1.31, moderate certainty, P-score: 0.59) and SUN alone. Comparison data was not available for AVE+AXI and NIV+CAB versus SUN, respectively (see Visual abstract).

Discussion

The results of this review mostly apply to people with clear cell RCC, as most trials in this review included participants with the clear cell type, whereas other carcinomas (non-clear cell or other subtypes) were underrepresented. Regarding the general heterogeneity between trials, moderate heterogeneity within the network for the outcome OS and substantial heterogeneity for the outcome SAE in the analyses of the

combined risk groups was observed. This heterogeneity probably originates from the slight differences in the included trials regarding some effect modifiers such as histology type, some participants having received prior nephrectomy or radiotherapy, differences in the sites of metastases, and the combination of all risk groups in the analyses presented here.

Conclusion

Implications for practice

Particularly the evidence for currently recommended treatment options for the different risk groups stems from evidence of one trial only; hence, the results of this review should be interpreted with caution. Furthermore, before a decision is met about a treatment option, the results of all outcomes should be taken into consideration, meaning benefits and harms should be contrasted with one another.

Implications for research

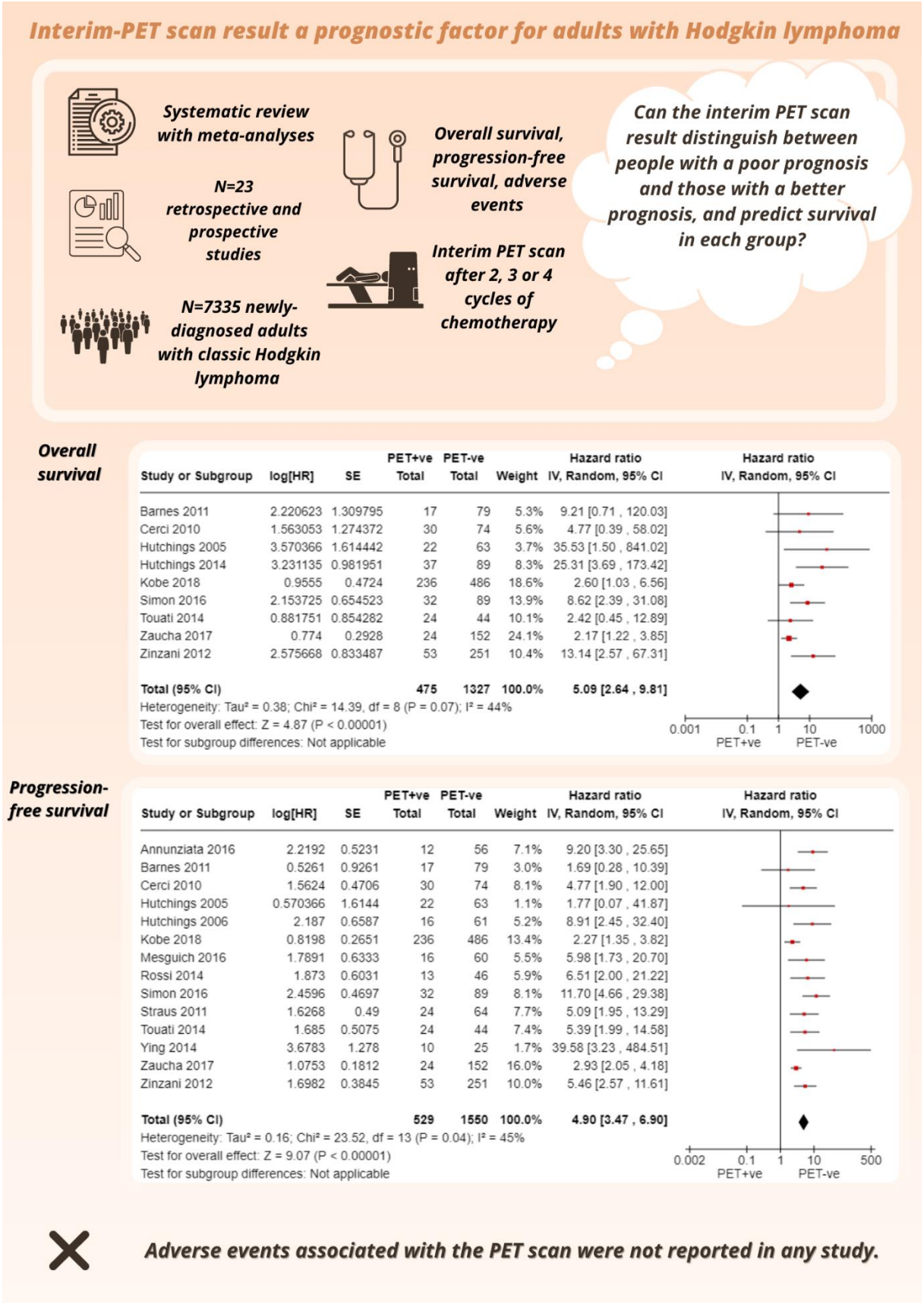
For most intervention comparisons in this review, there was direct evidence from one trial only. Furthermore, most interventions were compared to sunitinib only. More direct evidence from head-to-head comparisons between all different available interventions is needed to identify the most effective and safe treatment options for the different risk groups of advanced RCC. Furthermore, the effect of these interventions on the quality of life for people with advanced RCC needs to be extensively researched.

Systematic review II: Interim-PET scan results for prognosis in adults with Hodgkin lymphoma: a systematic review and meta-analysis of prognostic factor studies

In the main body of this cumulative dissertation, a visual abstract (p. 37) and an extensive summary (p. 38) of the review *Interim-PET scan results for prognosis in adults with Hodgkin lymphoma: a systematic review and meta-analysis of prognostic factor studies* (37) are provided. The summary entails sentences and text excerpts from the original manuscript, which can be found in Appendix ii (Systematic review II).

Citation of the published article:
Aldin A, Umlauff L, Estcourt LJ, Collins G, Moons KG, Engert A, Kobe C, von Tresckow B, Haque M, Foroutan F, Kreuzberger N, Trivella M, Skoetz N. Interim PET-results for prognosis in adults with Hodgkin lymphoma: a systematic review and meta-analysis of prognostic factor studies. Cochrane Database of Systematic Reviews 2020, Issue 1. Art. No.: CD012643. DOI: 10.1002/14651858.CD012643.pub3.

Visual abstract



Summary and main results

This summary entails sentences and text excerpts from the original manuscript (see Appendix ii (Systematic review II)).

Background

Description of the condition

Hodgkin lymphoma (HL) is a cancer of the lymph nodes and the lymphoid system with possible involvement of other organs such as the liver, lung, bones, or the bone marrow (48). It is a comparatively rare disease, but one of the most common haematological malignancies in young adults (49). With cure rates of 90%, it has become curable for most individuals.

Hodgkin lymphoma is classified into early favourable, early unfavourable, and advanced stage (50, 51). In Europe, the early favourable-stage group usually comprises Ann Arbor stages I and II without risk factors. The early unfavourable-stage group includes individuals with Ann Arbor stages I or II and one or more risk factors. Most individuals with stages IIB, III or IV disease are included in the advanced-stage risk group (52).

Description of the index (prognostic) factor

The prognostic factor of interest in this review was [18F]-fluorodeoxy-D-glucose (FDG)-positron-emission-tomography (PET). The PET is an imaging tool for identifying the state of FDG-avid tumours. It is used to monitor a tumour's metabolic activity, disease stage, and its progression. Therefore, it has become a standard imaging tool for various cancers (53), and is used more and more for staging, prognosis, treatment planning and response evaluation in individuals with HL (54-61). Furthermore, it is widely accepted to use the PET in combination with a computed tomography (CT), known as PET-CT (62), as it is argued to provide clearer imaging and a more accurate measurement of nodal size (55).

To assess the grade of uptake and report the PET scan result, it is generally recommended to use a five-point scale (59). Common criteria used to assess the PET scan result are the 5-PS Deauville criteria (59): scores 1-3 indicate PET-negativity, while scores 4-5 indicate PET-positivity (62).

The PET (and PET-CT) scan has been primarily used for the pre-treatment assessment to determine the stage of the disease of an individual and, thereby, to decide on the appropriate treatment regimen (55, 59). However, it has been argued that the PET scan should also be conducted during active treatment, namely an *interim PET* conducted after a few cycles of treatment (i.e., chemotherapy) have been administered (54, 59). The result of the interim PET scan (i.e., PET-positive, or PET-negative) is believed to be a good predictor of outcome, aiding the distinction between individuals with a poor prognosis from those with a better prognosis. Henceforth, the prognostic factor is referred to as *interim PET*.

Objective

The aim of this systematic review with meta-analyses was to determine whether in previously untreated adults with HL receiving first-line therapy, interim PET scan

results can distinguish between those with a poor prognosis and those with a better prognosis, and thereby predict survival outcomes in each group.

Methods

Criteria for considering studies for this review

Population	Index (prognostic) factor	Comparator	Outcome(s)	Timing	Setting
- People with classic HL, at any stage of the disease - Newly diagnosed individuals undergoing first-line therapy - Adults, as defined in the studies	Interim PET scan result	Not applicable to this review	- Overall survival (OS) - Progression-free survival (PFS) - PET-associated adverse events (AEs) The outcome should be measured after a minimum follow-up of 12 months.	Interim PET scan should be conducted during chemotherapy (after one, two, three or four cycles of chemotherapy)	Hospital/ treatment centre

Methods for the conduct of the systematic review

The following steps to conduct this systematic review were performed in accordance with the methods recommended in the Cochrane Handbook for Systematic Reviews of Interventions (3).

I. Electronic literature searches in relevant databases (i.e., CENTRAL, MEDLINE, Embase); conference proceedings (American Society of Haematology; European Haematology Association; International Symposium on Hodgkin Lymphoma); and one trial registry (Clinicaltrials.gov). The searches were conducted by an experienced information specialist. This step was followed by screening of search results and study selection based on the inclusion and exclusion criteria of the review.
II. Data extraction (i.e., study characteristics, outcome data) using a data extraction form developed specifically for this review.
III. Risk of bias assessment, using the Quality in Prognostic Studies (QUIPS) tool (27).
IV. Data analyses, conducted by a statistician experienced in survival analyses.
V. Assessment of the certainty in the evidence (i.e., GRADE) using the methods proposed by the GRADE Prognosis Working Group (63).
VI. Creation of the Summary of Findings (SoF) tables and writing the text of the review.

Except for the electronic searches (conducted by an experienced information specialist) and the statistical analyses (conducted by an experiences statistician), all

steps in the review development process were conducted independently by at least two review authors. Discrepancies were resolved through discussion and by involving a third author.

Main results

Twenty-three retrospective and prospective studies were included into this review, with a total of 7,335 female and male adult participants with newly diagnosed Hodgkin lymphoma (all disease stages). More information on the included studies and participants are reported in the full manuscript. The results presented here are the main results from univariable analyses of the outcomes of interest. Results from subgroup and sensitivity analyses as well as narratively reported results from multivariable analyses are reported in the main text of the manuscript.

Overall survival (OS)

Nine studies (eight observational studies and one randomised controlled trial (RCT)) with 1,802 participants were included in the meta-analysis for OS. There were 475 interim PET-positive and 1,327 interim PET-negative participants. Results from the meta-analysis showed a clear advantage in OS for participants with a negative interim PET scan compared to participants with a positive interim PET scan (hazard ratio (HR) 5.09, 95% confidence interval (CI) 2.64 to 9.81, I2 = 44%, moderate certainty in the evidence) (see Visual abstract).

Progression-free survival (PFS)

Fourteen studies (12 observational studies and two RCTs) with 2,079 participants were included in the meta-analysis for PFS. There were 529 interim PET-positive and 1,550 interim PET-negative participants. Results from the meta-analysis showed a clear advantage in PFS for participants with a negative interim PET scan compared to participants with a positive interim PET scan (HR 4.90, 95% CI 3.47, 6.90, I^2 = 45%, very low certainty in the evidence) (see Visual abstract).

Adverse events (AEs)

None of the included studies reported AEs associated with the PET-scan.

Discussion

The results mostly apply to adults newly diagnosed with classic HL who receive a PET scan in combination with a CT (PET-CT) after two cycles of chemotherapy (i.e., PET-2). Most participants in the included studies received chemotherapy with ABVD (Adriamycin/Doxorubicin, Bleomycin, Vinblastine and Dacarbazine), which is the standard treatment regimen for early-stage disease. However, as therapy regimen differ between participants according to their disease stage and other clinical or individual characteristics, results should always be interpreted with caution for different patient groups. Hence, this naturally restrains the applicability of the evidence for all people with classic HL.

Most studies used the Deauville five-point scale (DS 1 - 5) for the evaluation of the PET scans. However, different cut-off values were used for PET-positivity. Most studies considered scores one to three (DS 1-3) for PET-negativity, and scores four to

five (DS 4-5) for PET-positivity. In some studies, however, DS3 was also considered and tested for PET-positivity. Results from these studies should be interpreted with caution, as using a score of ≥3 can have an important impact on the results and possibly introduce bias (64).

Conclusion

Implications for practice

The results from univariable meta-analyses in this review provided moderate certainty evidence that the interim PET-scan result predicts OS, and very low certainty evidence that it predicts PFS in people with HL. The evidence on the ability of interim PET scan results to distinguish between individuals with a poor prognosis (i.e., PET-positive) and individuals with a good prognosis (i.e., PET-negative) can aid decision-making for clinicians and diagnosed individuals, and the evidence may inform clinical practice guidelines for individuals with HL.

Implications for research

Thus far, the prognostic value of the interim PET scan result has mostly been assessed in univariable analyses, where its prognostic ability of determining survival outcomes in individuals with HL has been shown. However, it is important to assess the independent prognostic value of the PET scan result against other established prognostic factors for HL, such as age, sex, B symptoms or other relevant clinical and individual factors in multivariable analyses as well. In such analyses, the independent prognostic ability of the PET scan result, as well as its incremental value on top of other prognostic factors, can be assessed (20).

CHAPTER 5

DISCUSSION AND CONCLUSION

Discussion

The primary objective of this cumulative dissertation was to conduct systematic reviews using new and complex systematic review methods, and to contribute to the development and adaptation of these methods. Secondary objective was to conduct clinically relevant systematic reviews to provide meaningful evidence that may inform clinical practice in oncology.

To achieve these objectives, two Cochrane systematic reviews with different complex methodological approaches were conducted and published in the Cochrane Library. For the first systematic review (36) in this cumulative dissertation, a new approach for living systematic reviews was proposed and applied, where during the conduct of the review new evidence was continuously searched for in a systematic manner, and newly emerged studies and study results were embedded into the review immediately as they became available. Further peculiarities of this review included complex statistical methods (network meta-analysis) and the use of the new Risk of Bias 2.0 tool. To conduct a clinically relevant systematic review, the benefits and harms of first-line therapies for adults with advanced renal cell carcinoma (RCC) were assessed. The second systematic review (37) within this cumulative dissertation was a systematic review of prognostic factor studies. This review is an exemplar review as it was the first systematic review with meta-analysis of prognostic factor studies published in the Cochrane Library. Retrospective and prospective prognostic factor studies were included in this review, and the methodological quality of the included studies was critically appraised using the Quality in Prognostic Factor Studies (QUIPS) (27) tool. The GRADE assessment for rating the certainty in the evidence was conducted in close collaboration with the GRADE Prognosis Working Group. To conduct a clinically relevant systematic review, the interim positron emission tomography (PET) scan result (i.e., PET-positive, or PET-negative) was explored as a prognostic factor in adults with Hodgkin lymphoma (HL).

Implications for systematic review methodology

In the following sections, implications for the methodology of systematic reviews with network meta-analyses, systematic reviews with a *living* approach, and for systematic reviews of prognosis studies are discussed. These implications were derived from the challenges encountered and the experiences collected during the conduct of the two systematic reviews (36, 37) within this cumulative dissertation.

Implications for systematic reviews with network meta-analyses

Checking and validating assumptions (homogeneity, consistency, transitivity)

There were two main challenges to the present network meta-analysis. Firstly, there was a great lack of available direct, head-to-head comparisons of the many different

treatment options from primary studies. Secondly, most direct evidence stemmed from only one source of evidence (i.e., only one trial) per comparison. Hence, the expected additional benefit of the network meta-analytic approach is limited in this systematic review. This challenged the methodology in this review through the fact that the assumptions that need to be met for a network meta-analysis could not be checked or validated for some analyses.

Homogeneity

The homogeneity assumption could not be tested and validated statistically for every analysis in the present network meta-analysis. Heterogeneity statistics to validate the assumption on homogeneity in a network meta-analysis (p. 16) could not be calculated for seven analyses (in three outcomes) because in the respective networks, each pairwise comparison was reported by a single trial only. In other words, there was a great lack of primary studies that evaluated the same comparisons, and that could have been incorporated into the evidence base had they been available. For two analyses in the present network meta-analyses, where the direct evidence stemmed from at least two trials, substantial heterogeneity (using I^2) in the pairwise comparisons was observed. This was probably because of some differences in clinical effect modifiers (e.g., risk groups or histology types) in these analyses. However, for most analyses in this review, heterogeneity statistics were calculated and little to no heterogeneity was observed. It is important that heterogeneity in pairwise comparisons is low, as otherwise high statistical heterogeneity can affect the confidence in the results of the network meta-analysis (11). In cases where heterogeneity is found, subgroup analyses for important effect modifiers can be appropriate to see whether they are the source of this heterogeneity (65). Therefore, in the present network meta-analysis, subgroup analyses were initially planned, for example for different histology types, sex, sites of metastases, and others. Unfortunately, subgroup analyses particularly of clinical characteristics was not possible due to a great lack of reporting of such in the primary studies (36). However, to minimise heterogeneity, analyses were conducted separately for the different risk groups of RCC whenever possible (36).

Consistency

The assumption on consistency (p. 17) could also not be checked statistically for every outcome. Consistency can be checked within a *closed loop* of direct evidence (see Figure 7 on p. 17) by examining whether the effect estimates of direct and indirect comparisons agree. This would have been possible if each intervention had been compared directly with (at least) two other interventions, thereby creating a closed loop where each direct comparison can be supplemented by an indirect comparison. In the present network meta-analysis, closed loops were available only in six analyses. There were small signs for inconsistency for one analysis. For all other analyses, no or negligible signs for inconsistency were observed. Nevertheless, a great lack of closed loops remains, but these are particularly important for the added value

(or benefit) of network meta-analysis over conventional meta-analysis (14). When closed loops are not available and the effect estimates for most comparisons are based only on indirect estimates, it is particularly important that the transitivity assumption holds for the comparator treatment (in this case SUN) (10).

Transitivity

The validity of the present network meta-analysis is largely based on the transitivity assumption (p. 16). The distribution of important effect modifiers was examined during data extraction as it is important that potential effect modifiers do not vary substantially across studies; otherwise, a quantitative analysis using the network meta-analytic approach would not be recommended (11). The studies included in the present network meta-analysis (36) were similar regarding clinical and methodological characteristics and it was assumed that the transitivity assumption holds. A methodological characteristic that was accounted for was study design (i.e., all trials being RCTs). In addition, for cross over RCTs, only outcome data from the first period (i.e., before cross over) was collected and considered for analyses to ensure that data was comparable to data from parallel-group RCTs without planned cross over. From a clinical point of view, transitivity was also given because all interventions were administered via the same administration route across all trials, and most interventions were administered at the same doses. Particularly the main comparator in this review (SUN) was administered via the same route and at the same dose in all trials that included that comparator (36). This is particularly important because the relative effects for many comparisons are based only on indirect estimates, where SUN was the common comparator (10). As for the included participants, all had advanced RCC, and most participants had ≥ 2 metastatic sites. All participants were around the age of 60 years, and both sexes (males and females) were included in each trial. All trials explored first-line treatment, and 80% of trials included only treatment-naive participants. For the remaining trials, data was extracted for the treatment-naive population whenever possible to ensure that data was comparable. Eighteen trials included only people with clear cell carcinoma, and 14 trials mostly included people with clear cell carcinoma, whereas the remaining four trials included non-clear cell carcinomas. In all trials but one, participants had previously received a nephrectomy and in most trials, prior radiotherapy was previously administered. Lastly, regarding the risk groups, separate analyses for the different risk groups according to the different criteria (IMDC or MSKCC) were conducted whenever possible in order for data to be even more comparable (36).

A network meta-analysis is considered valid when transitivity can be assumed, meaning that there are no systematic differences in the comparisons, except for the treatments themselves that are being compared (11). It should be possible that participants can be, hypothetically, randomised to any of the treatments included in the network meta-analysis (11). Besides checking and validating the assumptions for a valid network meta-analysis, it is key that the network meta-analysis is accurately planned and conducted, and that its findings are transparently reported (65).

Reporting and analysis of 'adverse events'

For the network meta-analysis in this cumulative dissertation (36), outcome data for *adverse events* was analysed for only 50% of the included trials; hence, data from the other half of included trials was not analysable due to inconsistent reporting of safety data. It is crucial that in clinical trials, data on adverse events is collected in a systematic manner (66, 67). It should entail definitions of the adverse events as well as additional information such as on severity, timing of occurrence, mode of data collection, planned analysis and reasons for the collection of specific adverse events (66, 67). For the present network meta-analysis, a focus was laid on data that was collected on the number of participants who experienced at least one adverse event; instead of the number of events that occurred in total during the trial. Likewise, when specific adverse events were reported (e.g., diarrhoea), data was collected on the number of participants who experienced that specific adverse event at least once, instead of data on cumulated events (i.e., when participants experienced diarrhoea several times but at different severity levels) as this would have led to double counting of participants. However, not all studies reported the number of participants with at least one adverse event, and in some studies, it was simply unclear which number was reported (i.e., the number of participants with at least one event or the number of occurrences of an event). In addition, when participants of a trial experience the same adverse event more than once, the highest severity grade experienced for that specific event should be reported. In this review, some studies reported only on severity grades 3 and 4 (which were of interested in this review), while others reported all severity grades combined. However, important to note is that for severity grading, all studies used the Common Terminology Criteria for Adverse Events (CTCAE). Moreover, for this network meta-analysis, data was collected on adverse events that were categorised as all-cause events. Instead, some trials reported adverse events that were categorised as treatment-related by the principal investigators. However, adverse events cannot be strictly traced back to the treatment and may have other causes as well; a causal relationship to a treatment may be possible but does not necessarily need to be the case (67). Some studies did not clearly report whether all-cause or treatment-related events are being reported; hence, these data were also not comparable. As for timing of outcome measurement, the adverse events reported in the studies occurred during active treatment. However, in most included trials, continuous therapy was provided, while in other studies, therapy was provided for a fixed period. In such cases, the exact time points of occurrence of adverse events most likely varied between trials. Lastly, it was also found that varying terminology was used to report on this outcome, such as adverse events, adverse effects, toxic effects, or safety.

Phillips and colleagues (67) conducted a review to explore analysis and reporting of adverse events in RCTs. Their findings were similar to the findings in the present review in that data collection, analysis and reporting of adverse events in clinical trials is inconsistent and often lacked important and valuable information (e.g., method of data collection) (67). Due to the above mentioned reasons, a consequence for the present network meta-analysis was that data from half of the included studies

could not be used for network meta-analysis of adverse events, leaving a huge gap in the evidence base for this outcome (36). Hence, making firm conclusions about the safety profiles of the treatments is difficult and findings should be interpreted cautiously (67). For future clinical trials, it is recommended that the extension of the CONSORT[21] statement for reporting of adverse events in clinical trials is used by clinical study investigators (66, 67).

Measurement and reporting of 'quality of life'

Patient-reported outcomes, such as *quality of life*, are of utmost importance because they provide information about direct treatment benefit and harm from the patients' perspectives (68). They are particularly important for patients and clinicians when making treatment choices, thereby outweighing the benefits and harms of a treatment. In the present systematic review with network meta-analysis, it was found that uniform outcome measurement and reporting of this outcome is needed (36). In this review, 22 trials were included that measured quality of life. Across these trials, 25 different scales or subscales were used to measure different constructs related to a person's quality of life (e.g., pain or physical function) (68). Due to this high number, scales for the measurement of quality of life were prioritised together with the co-authors on this review who have clinical expertise in RCC, based on clinical relevance in daily clinical practice. Creating a prioritisation or even hierarchy of patient-reported outcome measurements can be a helpful method for selecting from multiple available measurements (68). Thereby, for this review, five scales were prioritised that particularly measured constructs related to kidney cancer or cancer in general, such as kidney cancer symptoms, physical, social, emotional, and functional well-being, or fatigue. Based on this prioritisation, 15 trials were identified that used the scales of interest for this review. Ultimately, data was extracted only from seven trials, as for the remaining eight trials, extracting results was not possible for different reasons. For example, some studies reported that the outcome was measured with a specific scale, but results could not be found anywhere (e.g., in the publication or trial registry entry). Other methodologists have also observed that studies often use several patient-reported outcome measurements within a study to measure similar constructs. However, this has then led to selective outcome reporting where study authors reported only a subset of all outcome measurements initially used, based on their results (68). Another reason was that most data had to be estimated from graphs, creating an important insecurity in the data. While data was ultimately extracted from seven trials, neither network meta-analyses nor pairwise meta-analyses were possible for this outcome as there was a great lack of comparisons that entail at least one common comparator, so individual study results could not be pooled. Moreover, important variation was identified regarding the timing of measurement. All in all, these identified issues made pooling of results

[21] The Consolidated Standards of Reporting Trials (CONSORT) statement is an evidence-based set of recommendations for reporting of randomised controlled trials (https://www.consort-statement.org/)

impossible or not feasible for this outcome. The lack of and variation in reporting data for the outcome quality of life has left a huge gap in the evidence base for this outcome in this review. The few evidence that was found was reported narratively (36). Future systematic review authors who aim to synthesise evidence on a patient-reported outcome should be aware of these issues. It is recommended to establish a core set of outcome measurements for the review topic, ideally at protocol stage (68). Moreover, future systematic review authors should consider contacting principal study investigators to request outcome data, particularly when reporting bias is suspected.

Implications for living systematic reviews

In this cumulative dissertation, a different approach to a living systematic review (p. 18) is proposed, where monthly update searches for relevant studies are already conducted during the conduct of the baseline systematic review. As mentioned, living systematic reviews are particularly important in clinical areas where new research (i.e., clinical studies) is continuously conducted, and new study findings become rapidly available. For example, in the research field of advanced renal cell carcinoma (RCC), which was the clinical topic of the systematic review within this cumulative dissertation (p. 29), new clinical trials are continuously conducted to assess the effectiveness of the many available treatments for treating RCC. Hence, for this review, a baseline search for relevant studies was conducted in all relevant medical databases. This was then followed by monthly update searches in the same databases, for a period of six months. The length of the search period and the date of the final update search were discussed and set with clinical experts from the field. One possible way to decide about a date for a final update search is by consulting clinical experts who are aware of the current study landscape and relevant studies (or study results) that may become available soon (i.e., within a few months). It may be feasible or recommended to await these results to include them in the review. In addition, relevant conferences, where new study results are usually first presented in public, could also be awaited. The results from every monthly update search can then be screened using standard systematic review methods. For this systematic review, with every update search, newly published clinical trials were found and included, and updated results from trials that were already included in the review have become available as well. On the one hand, some studies provided updated outcome results for a longer follow-up (usually the follow-up time pre-specified in the trial protocol), as previously published results were interim results. On the other hand, some studies have been just completed, for example, at the time of the first literature search for the review, and results were not yet available at that time. With every update search, it was found that new or updated study results have become available. Hence, the continuous update searches allowed to find these and incorporate them into the review. Using this approach, it was possible to continuously update the review by incorporating the newest and most recent study results. With every update search,

the data extraction and risk of bias assessments were also updated. One final update search was conducted ten months later, because for systematic reviews conducted within Cochrane, literature searches for intervention reviews should not be older than 12 months at publication of the review. The analyses were performed after the last update search, and after data extraction and risk of bias assessments were finalised (36).

Implications for systematic reviews of prognostic factor studies

Searching for prognostic factor studies

The first challenge encountered was in the identification and retrieval of prognostic factor studies (37). There were no search filters (69) that were specific and sensitive enough to identify prognostic factor studies. Hence, the search strategy used for the present prognostic factor review (37) was also not specific and retrieved a high number of studies to be screened. Several factors contributed to the difficulty of identifying prognostic factor studies: a broad spectrum of different types of studies, usually non-interventional and non-randomised studies produce prognostic factor studies, although randomised controlled trials (RCTs) can also provide some prognostic information; prognostic factor studies are usually not named as such; and indexing of relevant terms is seldomly done, particularly due to an inconsistent use of terminology (70, 71). Hence, it is difficult to identify prognostic factor studies in electronic medical databases (70, 71). During the conduct of this systematic review (37), it was the stringent screening and reading of all articles identified by the non-specific search to identify prognostic factor studies where the interim PET scan result was explored as a prognostic factor in people with Hodgkin lymphoma. To date, although some search filters have been developed to identify prognostic factor studies, their sensitivity (i.e., the proportion of relevant articles retrieved (72)) is still low (70). As mentioned, prognosis studies most often are retrospective studies, which have not been named as such and were not indexed as such in medical databases, making it difficult to identify them. Moreover, this lack for using consistent prognosis terminology and key words makes it even more difficult for search specialists to develop an appropriate search strategy with a search filter that includes relevant search terms on prognosis to achieve high sensitivity (70, 71).

Extracting data from prognostic factor studies

The second challenge encountered was that at the time of the conduct of the review, there was no guidance available for extracting data from prognostic factor studies. Therefore, a data extraction form specific to the PICOTS and objective of the prognostic factor review was developed by the doctoral student, and in consultation with the methodological co-authors on the review (37). The form was then pilot tested by extracting data from a small sample of the included studies. The data extraction form was further refined during several internal discussions between the doctoral student and the co-authors about required adaptations. After several

amendments of the form, two teams of two review authors independently extracted all relevant data from all studies included in the review.

Critically appraising prognostic factor studies

The third challenge encountered was in the critical appraisal of the methodological quality of studies. The doctoral student and co-authors used the recommended Quality of Prognostic Factor Studies (QUIPS) tool (27) to assess the risk of bias in the included studies. While this was a helpful tool, the doctoral student and co-authors made some amendments to it (37). Firstly, a fourth rating, namely an 'unclear' rating that was not an option in the tool, was added as a fourth possible rating. Although this can lead to potential biases in the assessment, this rating was only used when relevant information was evidently missing, thereby making it difficult to make a fair and transparent judgement for the respective study and risk of bias domain. The doctoral student and co-authors felt that rating a domain as a 'high risk of bias' in cases where data was missing for unknown reasons would be inappropriate. Nevertheless, the doctoral student and co-authors advise against the use of the rating 'unclear' as a default option (37). Secondly, the QUIPS tool included a domain named *confounding*. The doctoral student and the prognosis-experienced co-authors on the review renamed this domain to *other prognostic factors (covariates)*. This was done to highlight the important distinction between confounding (the preferred term when seeking estimates of the causal effect of a specific etiologic factor) and adjusting for other important prognostic factors, namely covariates (advocated when seeking the independent prognostic ability of index prognostic factors) (37). For example, in the context of this prognostic factor review, the disease stage is a key factor that is considered together with the interim PET scan result when decisions about treatment adaptation are made in daily clinical practice (37, 73). Hence, the doctoral student and the co-authors assessed studies that only included participants within one disease stage (e.g., only early stages or only advanced stages of Hodgkin lymphoma) to be at 'low risk of bias', as such patient sampling can be considered as accounting for disease stage as another prognostic factor (37). Likewise, studies that included participants within all disease stages, but offered adjusted results including disease stage as another prognostic factor, were also assessed to be at 'low risk of bias' (37). However, studies with participants of all disease stages, not accounting for disease stage, were assessed to be at 'high risk of bias' in this domain (37).

Meta-analyses of prognostic factor studies

The fourth challenge encountered was in the pooling and analyses of study results due to missing information and data from the included studies (37). After extracting data from all included studies, the doctoral student contacted ten study authors (i.e., the principal investigators) to request additional data. In some instances, relevant data such as effect estimates, sample sizes, number of events, the log rank p-value and/or confidence intervals (CI), which are needed for statistical pooling of results,

were not reported in the studies. If these could not be retrieved by contacting the study authors, the experienced statistician, who was a co-author on the review, re-calculated and estimated data by using an in-house calculator based on published methods for recovering survival data, whenever possible (74-76). Only through these extensive methodological approaches was it possible to conduct meaningful meta-analyses in this review.

Assessing the certainty in the evidence from prognostic factor studies

The last challenge encountered was in assessing the certainty in the evidence. During the conduct of the review, there was no official guidance on rating the certainty in the evidence from prognostic factor studies. In the present prognostic factor review, a general approach that has been proposed for prognosis studies by the GRADE working group was applied, suggesting that the starting point is one of high certainty for observational studies (63). The GRADE assessment was conducted in close collaboration with the GRADE Prognosis Working Group, particularly as three co-authors on this review (FF, NK, NS), including the doctoral student at the time of the review development, are members of this working group.

Implications for health research

Head-to-head comparisons of first-line therapies for adults with advanced renal cell carcinoma

The research field on first-line therapies for adults with advanced renal cell carcinoma (RCC) is a rapidly evolving field due to the continuously changing treatment landscape that includes newer combinations of targeted therapies (i.e., tyrosine kinase inhibitors or angiogenesis inhibitors) and immunotherapies (i.e., checkpoint inhibitors). Hence, 19 ongoing trials were identified in the systematic review (as of the 9th of February 2022, date of the last update search). However, for those substances and combinations that are currently recommended across the different risk groups, such as LEN+PEM, NIV+CAB, NIV+IPI, AVE+AXI, PEM+AXI, CAZ alone or PAZ alone, direct evidence from one trial only, respectively, was identified. In addition, in all trials, these substances were all compared to sunitinib (SUN) alone. What is needed, however, is more direct evidence from head-to-head comparisons between all available substances and combinations to assess their comparative effectiveness and, thereby, identify the most effective and safe treatment options for the individual risk groups of advanced RCC. Due to this lack of direct evidence, the additional benefit from the network meta-analytic approach in this review is limited.

Prognostic models for adults with Hodgkin lymphoma

The results from the systematic review on the interim PET scan result as a prognostic factor have shown the prognostic ability of the interim PET scan in univariable

analyses for determining survival outcomes in people with Hodgkin lymphoma. However, using one single factor is usually not sufficient to give a satisfactory prediction about the occurrence of an event or outcome (20, 37). Therefore, clinicians make use of multiple factors to provide an accurate prognosis on the course of a disease (20, 37). There are additional established prognostic factors for Hodgkin lymphoma, including age, gender, B symptoms, Ann Arbor disease stage, bulky disease, albumin level, anaemia, and white blood cell count, amongst others (77-79). In further research, multivariable analyses assessing the interim PET scan result and additional prognostic factors are needed to explore the independent prognostic value and the incremental value of the interim PET scan result on top of those additional factors (20, 37). Furthermore, to optimise prognosis, prognostic models can be built that consist of multiple prognostic factors that have been proven to be predictive of outcome in adults with Hodgkin lymphoma. Such models are built for risk adaptation and treatment stratification for people who present those specific factors included in a prognostic model for a specific disease, and thereby enables more individualised disease monitoring and treatment guidance. Using a combination of factors, rather than one factor only, allows for a more individual and accurate estimate for a patient's risk to experience a certain health event or health outcome within a specific period (20, 23, 37).

Implications for clinical practice

Informing recommendations on the treatment of advanced renal cell carcinoma

The results of this review may inform treatment recommendations in clinical practice guidelines for advanced RCC. Currently, for the first-line treatment setting, the clinical practice guideline by the National Comprehensive Cancer Network (NCCN) (80), the European Association of Urology (EAU) (81) and the German guideline (82) suggest PEM+AXI as a treatment option across all risk groups (i.e., favourable-, intermediate- or poor risk groups) (36). For the favourable risk group specifically, the NCCN, ESMO and EAU guidelines also list LEN + PEM or NIV + CAB as options (36, 80, 81, 83). For intermediate- or poor risk groups, other options can be NIV + CAB, LEN + PEM or NIV+IPI (36, 80-83). In addition, the German guideline and the NCCN also suggest avelumab + axitinib (AVE + AXI) across all risk groups (36, 80, 82). In situations where immune checkpoint inhibitors cannot be administered or tolerated, targeted therapy is another option (36). This can include PAZ alone for IMDC favourable or intermediate + poor risk groups (81), and CAB or SUN for intermediate- and poor-risk groups (36, 81). The NCCN guideline recommends CAB, PAZ or SUN across all risk groups as possible options (36, 80). The German guideline recommends BEV+IFN, PAZ, SUN or TIV for the favourable risk group; TIV, SUN, PAZ, CAB, or BEV+IFN for the intermediate risk group; CAB, SUN, or alternatively PAZ or temsirolimus (TEM) for the poor risk group (36, 82).

Updated recommendations for the use of the interim PET scan

In 2021, the Federal Joint Committee (German: Gemeinsamer Bundesausschuss) in Germany has published an updated decision on the use and reimbursement of the PET (and combination of PET-CT) scan in adults with Hodgkin lymphoma (84). The Federal Joint Committee has recognised the PET scan as an accepted procedure for initial, interim and re-staging purposes of the disease during first- and second-line therapy. Excluded from this, however, is the use of the PET scan in routine follow-up of patients without a reasonable indication for relapse (84).

Long before this decision was taken, the German clinical practice guideline for the diagnosis, treatment, and follow-up for adults with Hodgkin lymphoma (85) had already formulated recommendations for the use of the PET scan during treatment for interim and re-staging purposes. Different study results have contributed to the formulation of such recommendations, including the results of the present systematic review on the interim PET-scan result as a prognostic factor in adults with Hodgkin lymphoma in this cumulative dissertation (37). In addition, recommendations for treatment adaptation based on the interim PET scan result (i.e., PET-positive, or PET-negative) were also formulated to achieve maximum efficacy of the treatment. However, in former versions of the guideline, each recommendation for the use of the PET scan for interim and re-staging purposes also included a CAVE warning that the PET scan was not included in the catalogue of services provided by the statutory health insurance; meaning a cost coverage was not guaranteed. However, after publication of the updated decision by the Federal Joint Committee to allow for reimbursement of the PET procedure, the recommendations within the clinical practice guideline affected by this decision were immediately updated by the guideline committee. Hence, the CAVE warnings were removed from the recommendations in the updated version of the guideline (85). Basis for the updated decision by the Federal Joint Committee were multiple study results by the German Hodgkin Study Group[22] that have proven the added value of the PET scan for initial, interim and re-staging purposes (84), particularly on the advantages of treatment adaptation based on the PET scan result.

Strengths and limitations of this cumulative dissertation

A key strength of this cumulative dissertation is that it followed high-quality methodological and reporting standards for systematic reviews as proposed in the Cochrane handbook (3). Where methods were not available or standardised yet, such as for synthesising evidence from prognostic factor studies, this dissertation contributed to the development and refinement of these methods by involving methodological and clinical experts as co-authors on both reviews to cumulate knowledge and experience, and to ensure methodological rigour during the conduct of both reviews. As a result, the methodological findings from these reviews have informed and may continue to inform the development of new methodological

[22] The German Hodgkin Study Group is located at the Department I of Internal Medicine of the University Hospital of Cologne and conducts clinical trials on the therapy of Hodgkin lymphoma. Official website: https://en.ghsg.org/

guidance. Each review included extensive discussions with methodological and clinical experts (as co-authors) in person, via teleconferences or e-mail exchange during different steps of the review development processes to discuss methodological and clinical issues, and to find suitable and appropriate solutions to any issues that arose during the review development processes. Moreover, during the conduct of each review, each step in the review process was conducted independently by at least two review authors, and through involvement of a third author, whenever necessary, to reach a final consensus.

Challenges and limitations that were encountered during the conduct of both reviews were thoroughly examined in the discussion of this cumulative dissertation. A key limitation to be highlighted again is that of the systematic review with network meta-analysis regarding the statistical validation of important assumptions, as reported in the chapter *Implications for systematic reviews with network meta-analyses* on page 43. The lack of evidence for direct comparisons of the various treatment options has led to a lack of closed loop testing for consistency, and to missing heterogeneity statistics for some outcomes. When a strong evidence base is missing, the results of network meta-analyses, including the ranking of treatments, should be interpreted with caution as they may be biased or misleading, given this lack of evidence.

Conclusion

This cumulative dissertation provided important and valuable methodological insights and implications for future systematic review authors through the conduct of systematic reviews with three methodologically highly complex concepts: network meta-analysis, living systematic review, and prognostic factor research. Moreover, it provided evidence on two clinically important research questions, namely on the benefits and harms of the available first-line treatment options for adults with advanced renal cell carcinoma, and on the use of the interim PET scan result as a prognostic factor for survival in adults with Hodgkin lymphoma. The challenges encountered during the conduct of both reviews were discussed and resolved through the involvement of methodological and clinical experts in the respective fields as co-authors. Future methodological research needs to further assess and address these different challenges, for example (but not limited to) the challenges one encounters in searching and identifying prognostic factor studies, in repeatedly searching for and incorporating new study results, or the limitations one encounters when conducting network meta-analysis (e.g., validating the consistency assumption). When evidence from such methodologically complex systematic reviews shall be used to inform clinical practice guidelines and, thereby, health care decision making, all involved stakeholders need to be aware of the methodological complexity and limitations behind the evidence produced.

REFERENCES

1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. Bmj. 1996;312(7023):71-2.
2. Gopalakrishnan S, Ganeshkumar P. Systematic Reviews and Meta-analysis: Understanding the Best Evidence in Primary Healthcare. J Family Med Prim Care. 2013;2(1):9-14.
3. Higgins JPT TJ, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated February 2022). Available from www.training.cochrane.org/handbook: Cochrane; 2022.
4. Jill Eden BW, Barbara McNeil, Harold Sox - Institute of Medicine. Knowing What Works in Health Care: A Roadmap for the Nation. Washington, DC: The National Academies Press; 2008. 278 p.
5. Institute of Medicine Committee on Standards for Developing Trustworthy Clinical Practice G. Clinical Practice Guidelines We Can Trust. In: Graham R, Mancher M, Miller Wolman D, Greenfield S, Steinberg E, editors. Clinical Practice Guidelines We Can Trust. Washington (DC): National Academies Press (US); 2011. p. 1-5.
6. Ioannidis JP. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. The Milbank quarterly. 2016;94(3):485-514.
7. Anthony J. Cochrane Library: an improved online platform to guide health decision-making across the world [updated August 8, 2018. Available from: https://www.cochrane.org/news/cochrane-library-improved-online-platform-guide-health-decision-making-across-world#:~:text=Cochrane%20Reviews%20are%20recognized%20internationally,databases%20within%20the%20Cochrane%20Library.
8. Useem J, Brennan A, LaValley M, Vickery M, Ameli O, Reinen N, et al. Systematic Differences between Cochrane and Non-Cochrane Meta-Analyses on the Same Topic: A Matched Pair Analysis. PLoS One. 2015;10(12):e0144980.
9. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. 2019;366:l4898.
10. Georgia Salanti DC, Anna Chaimani, Julian Higgins. Network meta-analysis. In: Adrian Levy SG, Constantine Gatsonis, Boris Sobolev, Ewout van Ginneken, Reinhard Buss, editor. Health Services Evaluation. Vancouver: Springer; 2019.
11. Rouse B, Chaimani A, Li T. Network meta-analysis: an introduction for clinicians. Internal and emergency medicine. 2017;12(1):103-11.
12. Chaimani A CD, Li T, Higgins JPT, Salanti G. Chapter 11: Undertaking network meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated February 2022). Available from www.training.cochrane.org/handbook. Cochrane; 2022.
13. Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JPT. Evaluating the Quality of Evidence from a Network Meta-Analysis. PLOS ONE. 2014;9(7):e99682.
14. Wilson DB, Tanner-Smith E, Mavridis D. Network meta-analysis. 2016;12(1):i-11.
15. Cumpston M CJ. Chapter IV: Updating a review. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated February 2022). Available from www.training.cochrane.org/handbook: Cochrane; 2022.
16. Simmonds M, Elliott JH, Synnot A, Turner T. Living Systematic Reviews. Methods in molecular biology (Clifton, NJ). 2022;2345:121-34.
17. Elliott JH, Synnot A, Turner T, Simmonds M, Akl EA, McDonald S, et al. Living systematic review: 1. Introduction-the why, what, when, and how. Journal of Clincal Epidemiology. 2017;91:23-30.

18. Joanne Brooker AS, Steve McDonald, Julian Elliott, Tari Turnerith, Rebecca Hodder, Laura Weeks, Juliane Ried, Harriet MacLehose, Elie Akl, Lara Kahale, John Hilton,Ella Flemyng, Toby Lasserson, James Thomas, Nicole Skoetz, Justin Clark, Robin Featherstone, AnnaNoel-Storr, Nicole Martin, Rachel Marshall, Mark Simmonds, Matthew Bagg, Paolo Fusar-Poli, Ambrish Singh, Roger Chou and the Living Evidence Network. Guidance for the production and publication of Cochrane living systematic reviews: Cochrane Reviews living mode. 2019.
19. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, et al. Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes. 2013;346:e5595.
20. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? 2009;338:b375.
21. Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. PLoS medicine. 2013;10(2):e1001380.
22. Buhmeida A, Pyrhönen S, Laato M, Collan Y. Prognostic factors in prostate cancer. Diagn Pathol. 2006;1:4-.
23. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS medicine. 2013;10(2):e1001381.
24. Leitlinienprogramm Onkologie (Deutsche Krebsgesellschaft DK, AWMF). S3-Leitlinie Prostatakarzinom, Langversion 6.2, 2021, AWMF Registernummer: 043/022OL, http://www.leitlinienprogramm-onkologie.de/leitlinien/prostatakarzinom/ (accessed on: 17.08.2022). 2022.
25. Hingorani AD, Windt DAvd, Riley RD, Abrams K, Moons KGM, Steyerberg EW, et al. Prognosis research strategy (PROGRESS) 4: Stratified medicine research. BMJ : British Medical Journal. 2013;346:e5793.
26. Riley RD, Moons KGM, Snell KIE, Ensor J, Hooft L, Altman DG, et al. A guide to systematic review and meta-analysis of prognostic factor studies. BMJ. 2019;364:k4597.
27. Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. Annals of internal medicine. 2013;158(4):280-6.
28. Wolf JS, Jr., Hubbard H, Faraday MM, Forrest JB. Clinical practice guidelines to inform evidence-based clinical practice. World journal of urology. 2011;29(3):303-9.
29. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. Journal of clinical epidemiology. 2011;64(4):383-94.
30. Qaseem A, Forland F, Macbeth F, Ollenschläger G, Phillips S, van der Wees P. Guidelines International Network: toward international standards for clinical practice guidelines. Annals of internal medicine. 2012;156(7):525-31.
31. Djulbegovic B, Guyatt G. Evidence vs Consensus in Clinical Practice Guidelines. Jama. 2019;322(8):725-6.
32. Tetreault L, Nater A, Garwood P, Badhiwala JH, Wilson JR, Fehlings MG. Development and Implementation of Clinical Practice Guidelines: An Update and Synthesis of the Literature With a Focus in Application to Spinal Conditions. 2019;9(1_suppl):53S-64S.
33. Andrews J, Guyatt G, Oxman AD, Alderson P, Dahm P, Falck-Ytter Y, et al. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. Journal of clinical epidemiology. 2013;66(7):719-25.
34. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. 2021;71(3):209-49.
35. Global-Cancer-Observatory. All cancers fact sheet. In: Cancer IAfRo, editor.: World Health Organization; 2020.
36. Aldin A, Besiroglu B, Adams A, Monsef I, Piechotta V, Tomlinson E, et al. First-line therapy for adults with advanced renal cell carcinoma: a systematic review and network meta-analysis. The Cochrane database of systematic reviews. 2023;5(5):CD013798.

37. Aldin A, Umlauff L, Estcourt LJ, Collins G, Moons KG, Engert A, et al. Interim PET-results for prognosis in adults with Hodgkin lymphoma: a systematic review and meta-analysis of prognostic factor studies. The Cochrane database of systematic reviews. 2020;1(1):CD012643.
38. American-Cancer-Society. Cancer Facts & Figures 2022. Atlanta: American Cancer Society; 2022.
39. Robert-Koch-Institute. Cancer in Germany 2017/2018. Robert Koch Institute; 2021.
40. American-Society-of-Clinical-Oncology. Kidney Cancer: Statistics.
41. Rini BI, Campbell SC, Escudier B. Renal cell carcinoma. Lancet (London, England). 2009;373(9669):1119-32.
42. Capitanio U, Bensalah K, Bex A, Boorjian SA, Bray F, Coleman J, et al. Epidemiology of Renal Cell Carcinoma. European Urology. 2019;75(1):74-84.
43. Scelo G, Larose TL. Epidemiology and Risk Factors for Kidney Cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2018;36(36):Jco2018791905.
44. Lopez-Beltran A, Carrasco JC, Cheng L, Scarpelli M, Kirkali Z, Montironi R. 2009 update on the classification of renal epithelial tumors in adults. International journal of urology : official journal of the Japanese Urological Association. 2009;16(5):432-43.
45. Warren AY, Harrison D. WHO/ISUP classification, grading and pathological staging of renal cell carcinoma: standards and controversies. World journal of urology. 2018;36(12):1913-26.
46. Coppin C, Porzsolt F, Autenrieth M, Kumpf J, Coldman A, Wilt T. Immunotherapy for advanced renal cell cancer. Cochrane Database of Systematic Reviews. 2004(3).
47. Unverzagt S, Moldenhauer I, Nothacker M, Roßmeißl D, Hadjinicolaou AV, Peinemann F, et al. Immunotherapy for metastatic renal cell carcinoma. Cochrane Database of Systematic Reviews. 2017(5).
48. Lister TA, Crowther D, Sutcliffe SB, Glatstein E, Canellos GP, Young RC, et al. Report of a committee convened to discuss the evaluation and staging of patients with Hodgkin's disease: Cotswolds meeting. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 1989;7(11):1630-6.
49. Howlader N, Noone AM, Krapcho M, Garshell J, Neyman N, Altekruse SF, et al. SEER Cancer Statistics Review, 1975-2010. National Cancer Institute Bethesda, MD http://seercancergov/csr/1975_2010/ based on November 2012 SEER data submission, posted to the SEER web site, April 2013 (accessed 23 June 2015).
50. Engert A, Franklin J, Eich HT, Brillant C, Sehlen S, Cartoni C, et al. Two cycles of doxorubicin, bleomycin, vinblastine, and dacarbazine plus extended-field radiotherapy is superior to radiotherapy alone in early favorable Hodgkin's lymphoma: final results of the GHSG HD7 trial. Journal of Clinical Oncology. 2007;25(23):3495-502.
51. Klimm B, Engert A, Diehl V. First-line treatment of Hodgkin's lymphoma. Current Hematology Reports. 2005;4(1):15-22.
52. Engert A, Schiller P, Josting A, Herrmann R, Koch P, Sieber M, et al. Involved-field radiotherapy is equally effective and less toxic compared with extended-field radiotherapy after four cycles of chemotherapy in patients with early-stage unfavorable Hodgkin's lymphoma: results of the HD8 trial of the German Hodgkin's Lymphoma Study Group. Journal of Clinical Oncology. 2003;21(19):3601-8.
53. Boellaard R, O'Doherty MJ, Weber WA, Mottaghy FM, Lonsdale MN, Stroobants SG, et al. FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. European Journal of Nuclear Medicine and Molecular Imaging. 2010;37(1):181-200.
54. Barrington SF, Kluge R. FDG PET for therapy monitoring in Hodgkin and non-Hodgkin lymphomas. European Journal of Nuclear Medicine and Molecular Imaging. 2017;44(1):97-110.
55. Cheson BD, Fisher RI, Barrington SF, Cavalli F, Schwartz LH, Zucca E, et al. Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification. Journal of Clinical Oncology. 2014;32(27):3059-68.
56. Fitzgerald TJ. Optimisation of adaptive therapy for advanced Hodgkin lymphoma. Lancet Oncology. 2019;20(2):167-8.

57. Kobe C, Dietlein M, Kriz J, Furth C, Fuchs M, Borchmann P, et al. The role of PET in Hodgkin's lymphoma and its impact on radiation oncology. Expert Review of Anticancer Therapy. 2010;10(9):1419-28.
58. Markova J, Kobe C, Skopalova M, Klaskova K, Dedeckova K, Plutschow A, et al. FDG-PET for assessment of early treatment response after four cycles of chemotherapy in patients with advanced-stage Hodgkin's lymphoma has a high negative predictive value. Annals of Oncology. 2009;20(7):1270-4.
59. Meignan M, Gallamini A, Meignan M, Gallamini A, Haioun C. Report on the First International Workshop on Interim-PET-Scan in Lymphoma. Leukemia and Lymphoma. 2009;50(8):1257-60.
60. Radford J, Illidge T, Counsell N, Hancock B, Pettengell R, Johnson P, et al. Results of a trial of PET-directed therapy for early-stage Hodgkin's lymphoma. New England Journal of Medicine. 2015;372(1):1598-607.
61. Specht L. FDG-PET scan and treatment planning for early stage Hodgkin lymphoma. Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology. 2007;85(2):176-7.
62. Barrington SF, Mikhaeel NG, Kostakoglu L, Meignan M, Hutchings M, Mueller SP, et al. Role of imaging in the staging and response assessment of lymphoma: consensus of the International Conference on Malignant Lymphomas Imaging Working Group. Journal of Clinical Oncology. 2014;32(27):3048-58.
63. Iorio A, Spencer FA, Falavigna M, Alba C, Lang E, Burnand B, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. Bmj. 2015;350:h870.
64. Kobe C, Goergen H, Baues C, Kuhnert G, Voltin CA, Zijlstra J, et al. Outcome-based interpretation of early interim PET in advanced-stage Hodgkin lymphoma. Blood. 2018;132(21):2273-9.
65. Guelimi R, Metelli S, Sbidian E, van Zuuren EJ, Flohr C, Leonardi-Bee J, et al. Network meta-analysis: methodological points for readers, authors and reviewers. 2022;186(6):917-8.
66. Ioannidis JP, Evans SJ, Gøtzsche PC, O'Neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Annals of internal medicine. 2004;141(10):781-8.
67. Phillips R, Hazell L, Sauzet O, Cornelius V. Analysis and reporting of adverse events in randomised controlled trials: a review. 2019;9(2):e024537.
68. Johnston BC PD, Devji T, Maxwell LJ, Bingham III CO, Beaton D, Boers M, Briel M, Busse JW, Carrasco-Labra A, Christensen R, da Costa BR, El Dib R, Lyddiatt A, Ostelo RW, Shea B, Singh J, Terwee CB, Williamson PR, Gagnier JJ, Tugwell P, Guyatt GH. Chapter 18: Patient-reported outcomes. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated February 2022). Cochrane; 2022.
69. Lefebvre C GJ, Briscoe S, Featherstone R, Littlewood A, Marshall C, Metzendorf M-I, Noel-Storr A, Paynter R, Rader T, Thomas J, Wieland LS. Chapter 4: Searching for and selecting studies. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated February 2022). Cochrane, 2022. Available from www.training.cochrane.org/handbook.
70. Boulos L, Ogilvie R, Hayden JA. Search methods for prognostic factor systematic reviews: a methodologic investigation. Journal of the Medical Library Association : JMLA. 2021;109(1):23-32.
71. Stallings E, Gaetano-Gil A, Alvarez-Diaz N, Solà I, López-Alcalde J, Molano D, et al. Development and evaluation of a search filter to identify prognostic factor studies in Ovid MEDLINE. BMC Medical Research Methodology. 2022;22(1):107.
72. Zhang L, Ajiferuke I, Sampson M. Optimizing search strategies to identify randomized controlled trials in MEDLINE. BMC Medical Research Methodology. 2006;6(1):23.
73. Bröckelmann PJ, Eichenauer DA, Jakob T, Follmann M, Engert A, Skoetz N. Hodgkin Lymphoma in Adults. Dtsch Arztebl International. 2018;115(31-32):535-40.

74. Altman D. Practical Statistics for Medical Research. 1 ed. London: Chapman & Hall; 1991.
75. Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. Statistics in medicine. 1998;17(24):2815-34.
76. Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. Trials. 2007;8(1):16.
77. Cuccaro A, Bartolomei F, Cupelli E, Galli E, Giachelia M, Hohaus S. Prognostic factors in hodgkin lymphoma. Mediterranean journal of hematology and infectious diseases. 2014;6(1):e2014053.
78. Josting A. Prognostic factors in Hodgkin lymphoma. Expert Review of Hematology. 2010;3(5):583-92.
79. Kılıçkap S, Barışta I, Ulger S, Celik I, Selek U, Yıldız F, et al. Clinical Features and Prognostic Factors of Hodgkin's Lymphoma: A Single Center Experience. Balkan medical journal. 2013;30(2):178-85.
80. Kumar SK, Callander NS, Adekola K, Anderson LD, Baljevic M, Campagnaro E, et al. Kidney Cancer, Version 3.2022, NCCN Clinical Practice Guidelines in Oncology. Journal of the National Comprehensive Cancer Network : JNCCN. 2022;20(1):67-81.
81. Ljungberg B, Albiges L, Abu-Ghanem Y, Bedke J, Capitanio U, Dabestani S, et al. European Association of Urology Guidelines on Renal Cell Carcinoma: The 2022 Update. European Urology. 2022;82(4):399-410.
82. Leitlinienprogramm Onkologie (Deutsche Krebsgesellschaft DK, AWMF). Diagnostik, Therapie und Nachsorge des Nierenzellkarzinoms, Langversion 3.0, 2021, AWMF-Registernummer: 043/017OL, https://www.leitlinienprogramm-onkologie.de/leitlinien/nierenzellkarzinom/ (accessed on 27.10.2022). 2021.
83. Powles T, Albiges L, Bex A, Grünwald V, Porta C, Procopio G, et al. ESMO Clinical Practice Guideline update on the use of immunotherapy in early stage and advanced renal cell carcinoma. Annals of oncology : official journal of the European Society for Medical Oncology. 2021;32(12):1511-9.
84. GemeinsamerBundesausschuss. Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung Richtlinie Methoden vertragsärztliche Versorgung: Positronenemissionstomographie; Positronenemissionstomographie/Computertomographie beim Hodgkin-Lymphom des Erwachsenen. Available from: https://www.g-ba.de/beschluesse/4943/.
85. Leitlinienprogramm Onkologie (Deutsche Krebsgesellschaft DK, AWMF). Diagnostik, Therapie und Nachsorge des Hodgkin Lymphoms bei erwachsenen Patienten, Langversion 3.1, 2022; AWMF Registernummer: 018/029 OL, https://www.leitlinienprogramm-onkologie.de/leitlinien/hodgkin-lymphom/ (accessed on 18.08.2022). 2022.

FINANCIAL SUPPORT

The two published systematic reviews in this cumulative dissertation were each financially supported with a research grant by the Federal Ministry of Education and Research of Germany (German: Bundesministerium für Bildung und Forschung, abbreviated BMBF), which was paid to the institution.

- BMBF grant number for the manuscript titled *"First-line therapy for adults with advanced renal cell carcinoma: a systematic review and network meta-analysis":* 01KG1901
- BMBF grant number for the manuscript titled *"Interim PET-results for prognosis in adults with Hodgkin lymphoma: a systematic review and meta-analysis of prognostic factor studies"*: 01KG1709

Conflicts of interest disclosure

No conflicts of interest to declare.

APPENDICES

Appendix I: Scientific contributions of authors

Scientific contribution to review 1: The doctoral student Angela Aldin contributed to the review process as follows: screening the results of the systematic literature search; study selection; development of a data extraction form; data extraction; quality assessment of the studies (risk of bias); assessment of the certainty in the evidence (GRADE); interpretation and writing up of the results; writing up of the publication; and coordination of the entire project. The co-authors provided support with methodological or clinical expertise (for a detailed list of individual contributions, see following table).

SR I	Author	Tasks
"First-line therapy for adults with advanced renal cell carcinoma: a systematic review and network meta-analysis"	Angela Aldin	– Screening and selection of studies – Development of data extraction form – Data extraction – 'Risk of bias' assessment – GRADE assessment – Interpretation and writing of results, including creation of 'Summary of findings' tables – Writing and drafting the publication – Managing the communication with and between authors (coordination of the project)
	Burcu Besiroglu	– Screening and selection of studies – Data extraction – 'Risk of bias' assessment – Assisted in writing (i.e., description of studies and bias assessments)
	Anne Adams	– Statistical analysis – Proofread and commented on the review draft
	Ina Monsef	– Designed the search strategies and conducted all searches – Proofread and commented on the review draft
	Vanessa Piechotta	– Screening and selection of studies – Provided methodological expertise on network meta-analyses – Assisted in GRADE assessment – Proofread and commented on the review draft
	Eve Tomlinson	– Data extraction – Risk of bias assessment – Proofread and commented on the review draft
	Carolin Hornbach	– Searching for Clinical Study Reports (CSR) – Data extraction (i.e., characteristics of included studies) – Proofread and commented on the review draft
	Nadine Dressen	– Data extraction – Risk of bias assessment – Proofread and commented on the review draft
	Marius Goldkuhle	– Writing the protocol for this review – Screening and selection of studies – Searching for CSR

		– Provided methodological expertise – Proofread and commented on the review draft
	Philipp Maisch	– Provided clinical expertise – Proofread and commented on the review draft
	Philipp Dahm	– Provided clinical and methodological expertise – Proofread and commented on the review draft
	Axel Heidenreich	– Provided clinical expertise
	Nicole Skoetz	– Provided clinical and methodological expertise – Proofread and commented on the review draft

Scientific contribution to review 2: The doctoral student Angela Aldin contributed to the review process as follows: screening of systematic literature search results; study selection; development of a data extraction form; data extraction; quality assessment of studies (risk of bias); assessment of the certainty in the evidence (GRADE); interpretation and writing up of the results; writing of the publication; and coordination of the entire project. In addition, the doctoral student has presented the results of the project at international and national congresses in the form of oral presentations and posters. The co-authors have supported with methodological or clinical expertise (for a detailed list of individual contributions, see following table).

SR II	Author	Tasks
***"Interim PET-results for prognosis in adults with Hodgkin lymphoma: a systematic review and meta-analysis of prognostic factor studies"* (37)**	Angela Aldin	– Screening and selection of studies – Development of data extraction form – Data extraction – 'Risk of bias' assessment – GRADE assessment – Interpretation and writing of results, including creation of 'Summary of findings' tables – Writing and drafting the publication – Managing the communication with and between authors (coordination of the project) – Presentation of results at conferences
	Lisa Umlauff	– 'Risk of bias' assessment – Characteristics of included and excluded studies (in text and tables) – Abstract and Plain language summary – Proofread and commented on the draft
	Karel Moons	– Methodological input on reviews of prognosis studies
	Lise J. Estcourt	– Screening and selection of studies – Data extraction – 'Risk of bias 'assessment – Clinical and methodological input
	Andreas Engert	– Medical and content input, particularly on the clinical comparability of studies and subgroup analyses
	Carsten Kobe	– Nuclear medical input on PET-CT
	Bastian von Tresckow	– Clinical input, particularly on the clinical comparability of studies and subgroup analyses
	Gary Collins	– Methodological input on reviews of prognostic studies
	Madhuri Haque	– Screening and selection of studies
	Farid Faroutan	– Input on risk of bias and GRADE assessments of prognostic factor studies
	Nina Kreuzberger	– 'Risk of bias' assessment – Proofread and commented on the review draft
	Marialena Trivella	– Screening and selection of studies – Data extraction – Risk of bias assessment – Statistical analysis

		– Proofread and commented on the review draft
	Nicole Skoetz	– Protocol development – Screening and selection of studies – Data extraction – Risk of bias assessment – GRADE assessment – Proofread and commented on the review draft

Appendix II: Original publications

Appendix i (Systematic review I): *"First-line therapy for adults with advanced renal cell carcinoma: a systematic review and network meta-analysis"*

Appendix ii (Systematic review II): *"Interim PET-results for prognosis in adults with Hodgkin lymphoma: a systematic review and meta-analysis of prognostic factor studies"*

Appendix i (Systematic review I)

Aldin A, Besiroglu B, Adams A, Monsef I, Piechotta V, Tomlinson E, Hornbach C, Dressen N, Goldkuhle M, Maisch P, Dahm P, Heidenreich A, Skoetz N. First-line therapy for adults with advanced renal cell carcinoma: a systematic review and network meta-analysis. Cochrane Database Syst Rev. 2023 May 4;5(5):CD013798. doi: 10.1002/14651858.CD013798.pub2. PMID: 37146227; PMCID: PMC10158799.

Cochrane Library

Cochrane Database of Systematic Reviews

First-line therapy for adults with advanced renal cell carcinoma: a systematic review and network meta-analysis (Review)

Aldin A, Besiroglu B, Adams A, Monsef I, Piechotta V, Tomlinson E, Hornbach C, Dressen N, Goldkuhle M, Maisch P, Dahm P, Heidenreich A, Skoetz N

Aldin A, Besiroglu B, Adams A, Monsef I, Piechotta V, Tomlinson E, Hornbach C, Dressen N, Goldkuhle M, Maisch P, Dahm P, Heidenreich A, Skoetz N.
First-line therapy for adults with advanced renal cell carcinoma: a systematic review and network meta-analysis.
Cochrane Database of Systematic Reviews 2023, Issue 5. Art. No.: CD013798.
DOI: 10.1002/14651858.CD013798.pub2.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Intervention Review]

First-line therapy for adults with advanced renal cell carcinoma: a systematic review and network meta-analysis

Angela Aldin[1], Burcu Besiroglu[1], Anne Adams[2], Ina Monsef[1], Vanessa Piechotta[1], Eve Tomlinson[3], Carolin Hornbach[1], Nadine Dressen[1], Marius Goldkuhle[1], Philipp Maisch[4], Philipp Dahm[5], Axel Heidenreich[6], Nicole Skoetz[1]

[1]Cochrane Haematology, Department I of Internal Medicine, Center for Integrated Oncology Aachen Bonn Cologne Duesseldorf, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. [2]Institute of Medical Statistics and Computational Biology, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. [3]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. [4]Department of Urology, University of Ulm, Ulm, Germany. [5]Urology Section, Minneapolis VA Health Care System, Minneapolis, Minnesota, USA. [6]Department of Urology, Uro-oncology, Special Urological and Robot-assisted Surgery, University Hospital of Cologne, Cologne, Germany

Contact: Angela Aldin, angela.aldin@uk-koeln.de.

Editorial group: Cochrane Urology Group.
Publication status and date: New, published in Issue 5, 2023.

Citation: Aldin A, Besiroglu B, Adams A, Monsef I, Piechotta V, Tomlinson E, Hornbach C, Dressen N, Goldkuhle M, Maisch P, Dahm P, Heidenreich A, Skoetz N. First-line therapy for adults with advanced renal cell carcinoma: a systematic review and network meta-analysis. *Cochrane Database of Systematic Reviews* 2023, Issue 5. Art. No.: CD013798. DOI: 10.1002/14651858.CD013798.pub2.

ABSTRACT

Background

Since the approval of tyrosine kinase inhibitors, angiogenesis inhibitors and immune checkpoint inhibitors, the treatment landscape for advanced renal cell carcinoma (RCC) has changed fundamentally. Today, combined therapies from different drug categories have a firm place in a complex first-line therapy. Due to the large number of drugs available, it is necessary to identify the most effective therapies, whilst considering their side effects and impact on quality of life (QoL).

Objectives

To evaluate and compare the benefits and harms of first-line therapies for adults with advanced RCC, and to produce a clinically relevant ranking of therapies. Secondary objectives were to maintain the currency of the evidence by conducting continuous update searches, using a living systematic review approach, and to incorporate data from clinical study reports (CSRs).

Search methods

We searched CENTRAL, MEDLINE, Embase, conference proceedings and relevant trial registries up until 9 February 2022. We searched several data platforms to identify CSRs.

Selection criteria

We included randomised controlled trials (RCTs) evaluating at least one targeted therapy or immunotherapy for first-line treatment of adults with advanced RCC. We excluded trials evaluating only interleukin-2 versus interferon-alpha as well as trials with an adjuvant treatment setting. We also excluded trials with adults who received prior systemic anticancer therapy if more than 10% of participants were previously treated, or if data for untreated participants were not separately extractable.

Data collection and analysis

All necessary review steps (i.e. screening and study selection, data extraction, risk of bias and certainty assessments) were conducted independently by at least two review authors. Our outcomes were overall survival (OS), QoL, serious adverse events (SAEs), progression-free survival (PFS), adverse events (AEs), the number of participants who discontinued study treatment due to an AE, and the time to

initiation of first subsequent therapy. Where possible, analyses were conducted for the different risk groups (favourable, intermediate, poor) according to the International Metastatic Renal-Cell Carcinoma Database Consortium Score (IMDC) or the Memorial Sloan Kettering Cancer Center (MSKCC) criteria. Our main comparator was sunitinib (SUN). A hazard ratio (HR) or risk ratio (RR) lower than 1.0 is in favour of the experimental arm.

Main results

We included 36 RCTs and 15,177 participants (11,061 males and 4116 females). Risk of bias was predominantly judged as being 'high' or 'some concerns' across most trials and outcomes. This was mainly due to a lack of information about the randomisation process, the blinding of outcome assessors, and methods for outcome measurements and analyses. Additionally, study protocols and statistical analysis plans were rarely available.

Here we present the results for our primary outcomes OS, QoL, and SAEs, and for all risk groups combined for contemporary treatments: pembrolizumab + axitinib (PEM+AXI), avelumab + axitinib (AVE+AXI), nivolumab + cabozantinib (NIV+CAB), lenvatinib + pembrolizumab (LEN+PEM), nivolumab + ipilimumab (NIV+IPI), CAB, and pazopanib (PAZ). Results per risk group and results for our secondary outcomes are reported in the summary of findings tables and in the full text of this review. The evidence on other treatments and comparisons can also be found in the full text.

Overall survival (OS)

Across risk groups, PEM+AXI (HR 0.73, 95% confidence interval (CI) 0.50 to 1.07, moderate certainty) and NIV+IPI (HR 0.69, 95% CI 0.69 to 1.00, moderate certainty) probably improve OS, compared to SUN, respectively. LEN+PEM may improve OS (HR 0.66, 95% CI 0.42 to 1.03, low certainty), compared to SUN. There is probably little or no difference in OS between PAZ and SUN (HR 0.91, 95% CI 0.64 to 1.32, moderate certainty), and we are uncertain whether CAB improves OS when compared to SUN (HR 0.84, 95% CI 0.43 to 1.64, very low certainty). The median survival is 28 months when treated with SUN. Survival may improve to 43 months with LEN+PEM, and probably improves to: 41 months with NIV+IPI, 39 months with PEM+AXI, and 31 months with PAZ. We are uncertain whether survival improves to 34 months with CAB. Comparison data were not available for AVE+AXI and NIV+CAB.

Quality of life (QoL)

One RCT measured QoL using FACIT-F (score range 0 to 52; higher scores mean better QoL) and reported that the mean post-score was 9.00 points higher (9.86 lower to 27.86 higher, very low certainty) with PAZ than with SUN. Comparison data were not available for PEM+AXI, AVE+AXI, NIV+CAB, LEN+PEM, NIV+IPI, and CAB.

Serious adverse events (SAEs)

Across risk groups, PEM+AXI probably increases slightly the risk for SAEs (RR 1.29, 95% CI 0.90 to 1.85, moderate certainty) compared to SUN. LEN+PEM (RR 1.52, 95% CI 1.06 to 2.19, moderate certainty) and NIV+IPI (RR 1.40, 95% CI 1.00 to 1.97, moderate certainty) probably increase the risk for SAEs, compared to SUN, respectively. There is probably little or no difference in the risk for SAEs between PAZ and SUN (RR 0.99, 95% CI 0.75 to 1.31, moderate certainty). We are uncertain whether CAB reduces or increases the risk for SAEs (RR 0.92, 95% CI 0.60 to 1.43, very low certainty) when compared to SUN. People have a mean risk of 40% for experiencing SAEs when treated with SUN. The risk increases probably to: 61% with LEN+PEM, 57% with NIV+IPI, and 52% with PEM+AXI. It probably remains at 40% with PAZ. We are uncertain whether the risk reduces to 37% with CAB. Comparison data were not available for AVE+AXI and NIV+CAB.

Authors' conclusions

Findings concerning the main treatments of interest comes from direct evidence of one trial only, thus results should be interpreted with caution. More trials are needed where these interventions and combinations are compared head-to-head, rather than just to SUN. Moreover, assessing the effect of immunotherapies and targeted therapies on different subgroups is essential and studies should focus on assessing and reporting relevant subgroup data. The evidence in this review mostly applies to advanced clear cell RCC.

PLAIN LANGUAGE SUMMARY

Initial treatment for adults with advanced kidney cancer (renal cell carcinoma)

Abbreviations

• renal cell carcinoma (RCC)

• avelumab (AVE)

• axitinib (AXI)

• cabozantinib (CAB)

• ipilimumab (IPI)

• lenvatinib (LEN))

• nivolumab (NIV)

• pazopanib (PAZ)

• pembrolizumab (PEM)

• sunitinib (SUN)

Key messages

• When making treatment decisions, it is important to think about whether drugs lengthen life, and whether they decrease or increase harmful side effects.

• The findings in this review apply mostly to advanced renal cell carcinoma (RCC) with a clear cell component.

What is advanced RCC, and how is it treated?

RCC is a type of kidney cancer. It is more common in older people and in men than in women. This is because age (≥60 years) and male sex put people at higher risk of getting it. Other risk factors include body weight, smoking, a history of kidney stones and high blood pressure. More than half of people with RCC discover they have it from routine health check-ups, because many do not have symptoms in the early stages. When symptoms appear, they can impact people's quality of life and day-to-day activities. Before 2005, drugs for treatment of advanced RCC were few and treatments caused many side effects. Now, there are new types of drugs: immunotherapy (use people's own immune system to find and destroy cancer cells), or targeted therapy (interferes with molecules that are responsible for helping cancer cells to grow, divide, and spread). Combinations of these drugs are used for therapy. With these drugs, people may live longer, with a good quality of life and fewer or milder side effects. These drugs are evaluated in clinical studies with people with RCC.

What did we want to find out?

We wanted to use the most up-to-date information from clinical studies to measure the benefits and harms of different treatments for people with advanced RCC. We also wanted to learn if the drugs worked better for some people than others.

What did we do?

We searched for studies that explored different drugs that are immunotherapies or targeted therapies. We examined these in adults (≥18 years) with advanced RCC who receive their first therapy. We compared these drugs to the drug SUN, which is a widely used targeted drug and a commonly used comparator drug in studies. We used a standardised process to assess the quality of the findings and our certainty in them. We rated our certainty in the findings based on factors such as study methods, the number of participants in them, and the precision of study results.

What did we find?

We found 36 studies with 4116 women and 11,061 men, around 60 years of age, with advanced RCC. Most people had ≥2 metastatic sites. We found 22 drugs and 17 combinations of drugs that were measured in the studies. We also performed analyses for different risk groups of advanced RCC. We present and discuss our results for the different risk groups, drugs and combinations in the main text of this review, plus further outcomes. Below we present our main results for our primary outcomes, when all risk groups are combined. We focus on selected drugs (and combinations) (PEM+AXI, AVE+AXI, NIV+CAB, LEN+PEM, NIV+IPI, CAB alone, PAZ alone) that are currently recommended in international guidelines for the treatment of advanced RCC. We report their impact on survival, quality of life and serious side effects.

How long do people live?

People live an average of 28 months when treated with SUN. In comparison, people may live an average of 43 months with LEN+PEM, probably 41 months with NIV+IPI, probably 39 months with PEM+AXI, and probably 31 months with PAZ alone. We are uncertain whether people live an average of 34 months with CAB alone. We do not have information for AVE+AXI and NIV+CAB.

How do people rate their quality of life?

People who receive PAZ alone reported a higher level of quality of life than people who receive SUN, but we are uncertain about the findings. We do not have information for PEM+AXI, AVE+AXI, NIV+CAB, LEN+PEM, NIV+IPI or CAB alone.

What is people's risk for serious side effects?

People who receive SUN have an average risk of 40% for experiencing serious side effects. In comparison, the average risk is probably: 61% with LEN+PEM, 57% with NIV+IPI, 52% with PEM+AXI, and 40% with PAZ. We are uncertain whether the risk is on average 37% with CAB alone. We do not have information for AVE+AXI and NIV+CAB.

What are the limitations of the evidence?

More studies are needed where these new drugs (and combinations) are not only compared to SUN alone, but also to each other. We lack information on the comparative benefits and harms of these drugs in different people, e.g. when comparing men with women, or different histology types of RCC (e.g. clear cell type, papillary type, sarcomatoid type).

How up to date is this evidence?

We conducted our last search for studies in February 2022 and incorporated the most recent study results into this review.

Summary of findings 1. Summary of findings table for all risk groups combined

First-line therapy for adults with advanced renal cell carcinoma

Population: people with a confirmed diagnosis of advanced renal cell carcinoma (combined risk groups) without previous systemic anticancer therapy

Setting: outpatient

Interventions: pembrolizumab + axitinib (PEM+AXI), avelumab + axitinib (AVE+AXI), nivolumab + cabozantinib (NIV+CAB), lenvatinib + pembrolizumab (LEN+PEM), nivolumab + ipilimumab (NIV+IPI), pazopanib (PAZ), cabozantinib (CAB)

Comparator: sunitinib (SUN)

Effect estimates (hazard ratio (HR) or risk ratio (RR) < 1 favours intervention) and 95% confidence intervals (CI). Main comparator is SUN[1]

Outcomes	№ of participants (trials) in the network	Intervention	Relative effect (95% CI) of the network meta-analyses	Anticipated absolute effects (95% CI)		Certainty of the evidence (GRADE)	Interpretation of findings
				Risk with SUN[1,2,3]	Risk with intervention[4]		
Overall survival (OS) - Network (subnet 1) included 19 pairwise comparisons - Median follow-up across trials[5]: 32.2 months - Median OS with SUN across trials[2] in this network: 28.7 months	9705 (17 RCTs)	**PEM + AXI**	HR 0.73 (0.50 to 1.07)[6]	28.7 months	39.3 months (26.8 to 57.4)	⊕⊕⊕⊖ moderate[a]	PEM+AXI probably improve OS, when compared to SUN.
		AVE + AXI	n.a.[7]		-	-	-
		NIV + CAB	n.a.[7]		-	-	-
		LEN + PEM	HR 0.66 (0.42 to 1.03)[6]		43.5 months (27.9 to 68.3)	⊕⊕⊖⊖ low[a, b]	LEN+PEM may improve OS, when compared to SUN.
		NIV+IPI	HR 0.69 (0.69 to 1.00)[6]		41.6 months (28.7 to 41.6)	⊕⊕⊕⊖ moderate[c]	NIV + IPI probably improve OS, when compared to SUN.
		CAB	HR 0.84 (0.43 to 1.64)[6]		34.2 months (17.5 to 66.7)	⊕⊖⊖⊖ very low[d, e]	We are uncertain whether CAB improves OS, when compared to SUN.

		PAZ	HR 0.91 (0.64 to 1.32)[6]		31.5 months (21.7 to 44.8)	⊕⊕⊕⊖ moderate[f]	There is probably little or no difference in OS between PAZ and SUN.
Quality of life (QoL) We reported this outcome narratively in this review. Here, long-term results (i.e., at the end of treatment) are presented. In the comparison PAZ versus SUN, QoL was measured using FACIT-F (score range 0-52; higher scores represent better QoL).	-	**PEM + AXI**	n.a.[7]	-	-	-	-
		AVE + AXI	n.a.[7]	-	-	-	-
		NIV + CAB	n.a.[7]	-	-	-	-
		LEN + PEM	n.a.[7]	-	-	-	-
		NIV+IPI	n.a.[7]	-	-	-	-
		CAB	n.a.[7]	-	-	-	-
		PAZ	-	The mean post-score of the control group was 29.5.	One RCT (N = 2) reported that the mean post-score of the intervention group was 9.00 points higher (9.86 lower to 27.86 higher) than that of the control group.	⊕⊖⊖⊖ very low [g, h]	We are uncertain whether PAZ compared to SUN improves quality of life.
Serious adverse events (SAEs) - Network included 31 pairwise comparisons - Mean risk with SUN across trials[3] included in this network: 40.7%	10,709 (22 RCTs)	**PEM + AXI**	RR 1.29 (0.90 to 1.85)[6]	40.7%	52.5% (36.6 to 75.3)	⊕⊕⊕⊖ moderate[f]	PEM+AXI probably increase slightly the risk for SAEs, when compared to SUN.
		AVE + AXI	n.a.[7]		-	-	-
		NIV + CAB	n.a.[7]		-	-	-
		LEN + PEM	RR 1.52 (1.06 to 2.19)		61.9% (43.1 to 89.1)	⊕⊕⊕⊖ moderate[b]	LEN+PEM probably increase the risk for SAEs, when compared to SUN.
		NIV+IPI	RR 1.40 (1.00 to 1.97)[6]		57% (40.7 to 80.2)	⊕⊕⊕⊖ moderate[b]	NIV+IPI probably increase the risk for SAEs, when compared to SUN.

		CAB	RR 0.92 (0.60 to 1.43)[6]		37.4% (24.4 to 58.2)	⊕⊝⊝⊝ very low[b, i]	We are uncertain whether CAB reduces or increases the risk for SAE, when compared to SUN.
		PAZ	RR 0.99 (0.75 to 1.31 [6]		40.3% (30.5 to 53.3)	⊕⊕⊕⊝ moderate[f]	There is probably little or no difference in the risk for SAEs between PAZ and SUN.
Progression-free survival (PFS) - Network (subnet 1) included 27 pairwise comparisons - Median follow-up across trials[5]: 9.1 months - Median PFS with SUN across trials[2] in this network: 7.9 months	11,737 (25 RCTs)	**PEM + AXI**	HR 0.68 (0.52 to 0.89)[6]	9.2 months	13.5 months (10.3 to 17.7)	⊕⊕⊕⊝ moderate[b]	PEM+AXI probably improve slightly PFS, when compared to SUN.
		AVE + AXI	n.a.[7]		-	-	-
		NIV + CAB	n.a.[7]		-	-	-
		LEN + PEM	HR 0.39 (0.29 to 0.53)[6]		23.6 months (17.3 to 31.7)	⊕⊕⊕⊝ moderate[b]	LEN+PEM probably improve PFS, when compared to SUN.
		NIV+IPI	HR 0.89 (0.68 to 1.16)[6]		10.3 months (7.9 to 13.5)	⊕⊕⊝⊝ low[b, f]	There may be little or no difference between NIV+IPI and SUN in improving PFS.
		CAB	HR 0.54 (0.37 to 0.76)[8]		17.0 months (12.1 to 24.9)	⊕⊕⊝⊝ low[b, d]	CAB may improve PFS, when compared to SUN.
		PAZ	HR 1.05 (0.81 to 1.36)[6]		8.8 months (6.8 to 11.3)	⊕⊕⊕⊝ moderate[f]	There probably is little or no difference in PFS between PAZ and SUN.
Adverse events (AEs) (grade 3 or 4) - Network included 19 pairwise comparisons	6909 participants (13 RCTs)	**PEM + AXI**	n.a.[7]	70.6%	-	-	-
		AVE + AXI	RR 1.00 (0.92 to 1.08)[6]		70.6% (64.9 to 76.2)	⊕⊕⊕⊝ moderate[b]	There probably is little or no difference in the risk for AEs between AVE+AXI and SUN.

- Mean risk with SUN across trials[3] in this network: 70.6%		**NIV + CAB**	RR 1.07 (0.97 to 1.17)[6]		75.5% (68.5 to 82.6)	⊕⊕⊕⊖ moderate[b]	There probably is little or no difference in the risk for AEs between NIV+CAB and SUN.
		LEN + PEM	RR 1.15 (1.06 to 1.25)[6]		81.2% (74.8 to 88.2)	⊕⊕⊕⊖ moderate[b]	LEN+PEM probably increase slightly the risk for AEs (grade 3 or 4), when compared to SUN.
		NIV+IPI	n.a.[7]		-	-	-
		CAB	RR 1.04 (0.83 to 1.31)[6]		73.4% (58.6 to 92.5)	⊕⊖⊖⊖ very low[b, j]	We are uncertain whether CAB reduces or increases the risk for AEs, when compared to SUN.
		PAZ	RR 1.02 (0.96 to 1.09)[6]		72% (67.7 to 76.9)	⊕⊕⊕⊖ moderate[b]	There probably is little or no difference in the risk for AEs between PAZ and SUN.
Time to initiation of first subsequent therapy This outcome was not reported as a time-to-event outcome. Instead, authors of the trials reported the number of participants who received subsequent anticancer therapy after discontinuation of trial treatment.	861 (1 RCT)	**PEM + AXI**	RR 0.72 (0.64 to 0.81)[6]	65%[3]	46.8% (41.6 to 52.6)	⊕⊕⊖⊖ low[k]	PEM+AXI may reduce the risk for subsequent therapy, when compared to SUN.
	886 (1 RCT)	**AVE + AXI**	RR 0.61 (0.52 to 0.72)[6]	51%[3]	31.1% (26.5 to 36.7)	⊕⊕⊖⊖ low[k]	AVE+AXI may reduce the risk for subsequent therapy, when compared to SUN.
	651 (1 RCT)	**NIV + CAB**	RR 0.57 (0.44 to 0.75)[6]	33%[3]	18.8% (14.5 to 24.7)	⊕⊖⊖⊖ very low[b,k]	We are uncertain whether NIV+CAB reduce the risk for subsequent therapy, when compared to SUN.
	712 (1 RCT)	**LEN + PEM**	RR 0.57 (0.48 to 0.68)[6]	60%[3]	34.2% (28.8 to 40.8)	⊕⊖⊖⊖ very low[b, k]	We are uncertain whether LEN +PEM reduce the risk for subsequent therapy, when compared to SUN.
	1096 (1 RCT)	**NIV+IPI**	RR 0.86 (0.79 to 0.94)[6]	70%	60.2% (55.3 to 65.8)	⊕⊕⊖⊖ low[k]	NIV+IPI may reduce the risk for subsequent therapy, when compared to SUN.
	151	**CAB**	RR 0.93	64%	59.5%	⊕⊖⊖⊖	We are uncertain whether CAB reduces or increases the risk for sub-

	(1 RCT)		(0.74 to 1.16)[6]		(47.4 to 74.2)	very low[b, k, j]	sequent therapy, when compared to SUN.
-	**PAZ**		n.a.[7]		-	-	-

CI: confidence interval; **HR:** hazard ratio.

GRADE Working Group grades of evidence
High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.
Moderate certainty: we are moderately confident in the effect estimate; the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low certainty: our confidence in the effect estimate is limited; the true effect may be substantially different from the estimate of the effect.
Very low certainty: we have very little confidence in the effect estimate; the true effect is likely to be substantially different from the estimate of effect.

[1] Basis for the assumed risks
[2] The risk of SUN for OS and PFS was obtained from the included trials in the networks, respectively, and estimated by calculating the mean of all available medians for SUN
[3] Mean risk for AEs and SAEs, respectively, was estimated by dividing the total events under SUN-therapy by the total of participants treated with SUN across all trials in the network. For TFST, the risk for SUN was calculated using the number of events / number of participants for SUN in the respective trial.
[4] Methods of calculating the assumed risks in the intervention group:
- For OS and PFS: The median survival in the intervention group was calculated using the methods by Tierney 2007: Corresponding median survival in the intervention group (in months) = comparator group median survival time (in months) divided by the HR. Upper and lower confidence limits for the corresponding intervention risk were obtained by replacing HRs by their upper and lower confidence limits, respectively.
- For AEs and SAEs: The assumed risk in the intervention group was calculated with the formula available in the Cochrane Handbook. For the meta-analytic RR and assumed comparator risk (ACR) the corresponding intervention risk is obtained per 1000: 1000 x ACR x RR. Upper and lower confidence limits for the corresponding intervention risk were obtained by replacing RRs by their upper and lower confidence limits, respectively.
[5] Median follow-up across trials in the networks for OS and PFS, respectively, was estimated by calculating the mean of all available medians
[6] Only direct evidence from one trial.
[7] Not applicable, comparison not available.
[8] Only direct evidence from two trials.
[a] Downgraded by 1 level for imprecision because of a wide CI and upper CI limit suggests no difference between interventions.
[b] Downgraded by 1 level for study limitations because the one trial contributing all direct evidence is at high risk of bias.
[c] Downgraded by 1 level for imprecision because upper CI limit suggests no difference between interventions.
[d] downgraded by 1 level for indirectness because in one trial, 7% of the total study population received previous systemic therapy.
[e] Downgraded by 2 levels for imprecision because of a very wide CI that includes values that favour either of the compared treatments, and evidence stems from only one trial with 90 participants.
[f] Downgraded by 1 level for imprecision because of a wide CI that favours either of the compared treatments.
[g] Downgraded by 2 levels for study limitations due to a high risk of bias.
[h] Downgraded by 2 levels for imprecision because of a very wide CI that includes values that favour either of the compared treatments, and evidence stems from only one trial with four participants analysed.
[i] Downgraded by 2 levels for imprecision because of a very wide CI that includes values that favour either of the compared treatments, and evidence stems from only one trial with 157 participants.
[j] Downgraded by 2 levels for imprecision because of a wide CI that includes values that favour either of the interventions, and the evidence stems from only one trial with 157 participants.
[k] Downgraded by 2 levels for indirectness due to indirect measurement of outcome of interest.

Summary of findings 2. Summary of findings table for the favourable risk groups (according to IMDC and MSKCC)

First-line therapy for adults with advanced renal cell carcinoma

Population: people with a confirmed diagnosis of advanced renal cell carcinoma (RCC) and a favourable risk according to the International Metastatic RCC Database Consortium (IMDC) and Memorial Sloan-Kettering Cancer Center (MSKCC) risk models

Setting: outpatient

Interventions: pembrolizumab + axitinib (PEM+AXI), avelumab + axitinib (AVE+AXI), nivolumab + cabozantinib (NIV+CAB), lenvatinib + pembrolizumab (LEN+PEM), nivolumab + ipilimumab (NIV+IPI), cabozantinib (CAB), pazopanib (PAZ)

Comparator: sunitinib (SUN)

Effect estimate (hazard ratio (HR) < 1 favours intervention) and 95% confidence intervals (CI). Main comparator is SUN[1]

Outcomes	№ of participants (trials) in the network	Intervention	Relative effect (95% CI) of the network meta-analyses	Anticipated absolute effects (95% CI)		Certainty of the evidence (GRADE)	Interpretation of findings
				Risk with SUN[1,2]	Risk with intervention[3]		
IMDC risk group							
Overall survival (OS) - Network (subnet 1) included 5 pairwise comparisons - Median follow-up across trials[4]: 35 months - Median OS with SUN could not be estimated from data of the included trials in this network. We used the reported median survival from mdalc[5] for IMDC favourable risk groups	933 (4 RCTs)	**PEM + AXI**	n.a.[7]	43.2[5] months	-	-	-
		AVE + AXI	HR 0.66 (0.36 to 1.22)[6]		65.4 months (35.4 to 120.0)	⊕⊕⊖⊖ low[a, b]	AVE+AXI may improve OS, when compared to SUN.
		NIV + CAB	HR 0.94 (0.46 to 1.92)[6]		45.9 months (22.5 to 93.9)	⊕⊖⊖⊖ very low[a, c]	We are uncertain whether NIV+CAB improve or decrease OS, when compared to SUN.
		LEN + PEM	HR 1.15 (0.55 to 2.40)[6]		37.7 months (18.0 to 78.5)	⊕⊕⊖⊖ low[d]	There may be little or no difference in OS between LEN+PEM and SUN.

		NIV+IPI	HR 0.93 (0.62 to 1.40)[6]		46.4 months (30.8 to 69.7)	⊕⊕⊕⊖ moderate[b]	There probably is little or no difference in OS between NIV+IPI and SUN.
		CAB	n.a.[7]		-	-	-
		PAZ	n.a.[7]		-	-	-
Serious adverse events	Subgroup data not available.						
Quality of life	Subgroup data not available.						
Progression-free survival (PFS)	933	**PEM + AXI**	n.a.[7]	20.9 months	-	-	-
- Network (subnet 1) included 5 pair-wise comparisons	(4 RCTs)	**AVE + AXI**	HR 0.71 (0.49 to 1.02)[6]		29.4 months (20.5 to 42.6)	⊕⊕⊖⊖ low[a, e]	AVE+AXI may improve PFS, when compared to SUN.
- Median follow-up across trials[4]: 35 months		**NIV + CAB**	HR 0.58 (0.36 to 0.93)[6]		36.0 months (22.5 to 58.0)	⊕⊕⊖⊖ low[a,f]	NIV+CAB may improve PFS, when compared to SUN.
- Median PFS with SUN across trials[2] in this network: 20.9 months		**LEN + PEM**	HR 0.41 (0.28 to 0.61)[6]		51.0 months (34.3 to 74.6)	⊕⊕⊖⊖ low [a,f]	LEN+PEM may improve PFS, compared to SUN.
		NIV+IPI	HR 1.84 (1.29 to 2.62)[6]		11.3 months (7.8 to 16.2)	⊕⊕⊕⊖ moderate[a]	NIV+IPI probably reduce PFS, when compared to SUN.
		CAB	n.a.[7]		-	-	-
		PAZ	n.a.[7]		-	-	-
Adverse events (grade 3 to 4)	Subgroup data not available.						
Time to initiation of first subsequent therapy	Subgroup data not available.						
MSKCC risk group							
Overall survival (OS)	594	**PEM + AXI**	n.a.[7]	43.6 months	-	-	-

	(2 RCTs)						
- Network (subnet 1) included 3 pairwise comparisons		**AVE + AXI**	n.a.[7]		-	-	-
- Median follow-up across trials[4]: 26.6 months		**NIV + CAB**	n.a.[7]		-	-	-
- Median OS with SUN across trials[2] in this network: 43.6 months		**LEN + PEM**	HR 0.86 (0.38 to 1.93)[6]		50.7 months (22.6 to 114.7)	⊕⊖⊖⊖ very low[a, d]	We are uncertain whether LEN+PEM improve OS, when compared to SUN.
		NIV+IPI	n.a.[7]		-	-	-
		CAB	n.a.[7]		-	-	-
		PAZ	HR 0.88 (0.63 to 1.21)[6]		49.5 months (36.0 to 69.2)	⊕⊕⊖⊖ low[a, b]	There may be little or no difference between PAZ and SUN.
Serious adverse events (SAEs)	Subgroup data not available.						
Quality of life (QoL)	Subgroup data not available.						
Progression-free survival (PFS)	784 (6 RCTs)	**PEM + AXI**	n.a.[7]	13.7 months	-	-	-
		AVE + AXI	n.a.[7]		-	-	-
- Network (subnet 1) included 7 pairwise comparisons		**NIV + CAB**	n.a.[7]		-	-	-
- Median follow-up across trials[4]: 25 months		**LEN + PEM**	HR 0.36 (0.11 to 1.23)[6]		38.0 months (11.1 to 124.5)	⊕⊖⊖⊖ very low[a, d]	We are uncertain whether LEN+PEM improve PFS, when compared to SUN.
- Median PFS with SUN across trials[2] in this network: 13.7 months		**NIV+IPi**	n.a.[7]		-	-	-
		CAB	n.a.[7]		-	-	-
		PAZ	n.a.[7]		-	-	-

Adverse events (grade 3 to 4)	Subgroup data not available.
Time to initiation of first subsequent therapy	Subgroup data not available.

CI: confidence interval; **HR:** hazard ratio.

GRADE Working Group grades of evidence
High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.
Moderate certainty: we are moderately confident in the effect estimate; the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low certainty: our confidence in the effect estimate is limited; the true effect may be substantially different from the estimate of the effect.
Very low certainty: we have very little confidence in the effect estimate; the true effect is likely to be substantially different from the estimate of effect.

[1] Basis for the assumed risks
[2] The risk of SUN for OS and PFS was obtained from the included trials in the networks, respectively, and estimated by calculating the mean of all available medians for SUN
[3] Method of calculating the assumed risks in the intervention group for survival outcomes: The median survival in the intervention group was calculated using the methods by Tierney 2007: Corresponding median survival in the intervention group (in months) = comparator group median survival time (in months) divided by the HR. Upper and lower confidence limits for the corresponding intervention risk were obtained by replacing HRs by their upper and lower confidence limits, respectively.
[4] Median follow-up across trials in the networks for OS and PFS, respectively, was estimated by calculating the mean of all available medians
[5] Median OS with SUN could not be estimated from data of the included in this network. We used the reported median survival from mdalc for IMDC favourable risk groups, which is comparable to MSKCC favourable risk groups under SUN therapy
[6] Only direct evidence from one trial.
[7] Not applicable, comparison not available.
[a] Downgraded by 1 level for study limitations because the one trial contributing all direct evidence is at high risk of bias.
[b] Downgraded by 1 level for imprecision because of a wide CI that includes values that favour either of the compared treatments.
[c] Downgraded by 2 levels for imprecision because of a very wide CI that includes values that favour either of the compared treatments, and evidence stems from only one trial with 146 participants.
[d] Downgraded by 2 levels for imprecision because of a very wide CI that includes values that favour either of the compared treatments.
[e] Downgraded by 1 level for imprecision because of a wide CI and upper CI limit suggests no difference.
[f] Downgraded by 1 level for imprecision because evidence stems from only one trial with < 150 participants.

Summary of findings 3. Summary of findings for the intermediate and poor risk groups (according to IMDC and MSKCC)

First-line therapy for adults with advanced renal cell carcinoma
Population: people with a confirmed diagnosis of advanced renal cell carcinoma (RCC) and an intermediate or poor risk according to the International Metastatic RCC Database Consortium (IMDC) and Memorial Sloan -Kettering Cancer Center (MSKCC) risk models **Setting:** outpatient

Interventions: pembrolizumab + axitinib (PEM+AXI), avelumab + axitinib (AVE+AXI), nivolumab + ipilimumab (NIV+IPI), nivolumab + cabozantinib (NIV+CAB), lenvatinib + pembrolizumab (LEN+PEM), cabozantinib (CAB), pazopanib (PAZ)

Comparator: sunitinib (SUN)

Effect estimate (hazard ratio (HR) < 1 favours intervention) and 95% confidence intervals (CI). Main comparator is SUN[1]

Outcomes	№ of participants (trials) in the network	Intervention	Relative effect (95% CI) of the network meta-analyses	Anticipated absolute effects (95% CI)		Certainty of the evidence (GRADE)	Interpretation of findings
				Risk with SUN[1,2]	Risk with intervention[3]		
IMDC risk groups							
Overall survival (OS) - Network (subnet 1) included 10 pairwise comparisons - Median follow-up across trials[4]: 35.1 months - Median OS with SUN across trials[2] in this network: 23.9 months	2908 (5 RCTs)	**PEM + AXI**	n.a.[5]	23.9 months	-	-	-
		AVE + AXI	HR 0.73 (0.48 to 1.11)[6]		32.7 months (21.5 to 49.8)	⊕⊕⊖⊖ low[a, b]	AVE+AXI may improve OS, when compared to SUN.
		NIV + CAB	HR 0.60 (0.37 to 0.96)[6]		39.8 months (24.9 to 64.6)	⊕⊕⊕⊖ moderate[a]	NIV+CAB probably improve OS, when compared to SUN.
		LEN + PEM	HR 0.55 (0.33 to 0.91)[6]		43.4 months (26.3 to 72.4)	⊕⊕⊕⊖ moderate[a]	LEN+PEM probably improve OS, when compared to SUN.
		NIV + IPI	HR 0.65 (0.38 to 1.10)[6]		36.8 months (21.7 to 62.9)	⊕⊕⊕⊖ moderate[b]	NIV+IPI probably improve OS, when compared to SUN.
		CAB	HR 0.80 (0.42 to 1.52)[6]		29.8 months (15.7 to 56.9)	⊕⊖⊖⊖ very low[a, c]	CAB may improve slightly OS, when compared to SUN.
		PAZ	n.a.[5]		-	-	-
Quality of life	Subgroup data not available.						
Serious adverse events	Subgroup data not available.						

Progression-free survival (PFS) - Network (subnet 1) included 11 pairwise comparisons - Median follow-up across trials[4]: 34.5 months - Median PFS with SUN across trials[2] in this network: 6.0 months	2908 (5 RCTs)	**PEM + AXI**	n.a.[5]	6.0 months	-	-	-
		AVE + AXI	HR 0.60 (0.43 to 0.84)[6]		10.0 months (7.1 to 13.9)	⊕⊕⊕⊖ moderate[a]	AVE+AXI probably improve PFS, when compared to SUN.
		NIV + CAB	HR 0.48 (0.34 to 0.69)[6]		12.5 months (8.7 to 17.6)	⊕⊕⊕⊖ moderate[a]	NIV+CAB probably improve PFS, when compared to SUN.
		LEN + PEM	HR 0.36 (0.24 to 0.54)[6]		16.6 months (11.1 to 25.0)	⊕⊕⊕⊖ moderate[a]	LEN+PEM probably improve PFS, when compared to SUN.
		NIV + IPI	HR 0.74 (0.49 to 1.11)[6]		8.1 months (5.4 to 12.2)	⊕⊕⊖⊖ low[a, b]	There may be little or no difference in PFS between NIV+IPI and SUN.
		CAB	HR 0.46 (0.27 to 0.79)[6]		13.0 months (7.6 to 22.2)	⊕⊕⊕⊖ moderate[d]	CAB probably improves PFS, when compared to SUN.
		PAZ	n.a.[5]		-	-	-
Adverse events (grade 3 or 4)	Subgroup data not available.						
Time to initiation of first subsequent therapy	Subgroup data not available.						
MSKCC risk groups							
Overall survival (OS) - Network included 15 pairwise comparisons - Median follow-up across trials[4]: 36.4 months	3937 (7 RCTs)	**PEM + AXI**	n.a.[5]	18.2 months	-	-	-
		AVE + AXI	n.a.[5]		-	-	-
		NIV + CAb	n.a.[5]		-	-	-
		LEN + PEM	HR 0.63 (0.46 to 0.86)[6]		28.9 months (21.2 to 39.6)	⊕⊕⊕⊖ moderate[a]	LEN+PEM probably improve OS, when compared to SUN.

- Median OS with SUN across trials[2] in this network: 18.2 months		**NIV + IPI**	n.a.[5]		-	-	-
		CAB	n.a.[5]		-	-	-
		PAZ	HR 0.89 (0.75 to 1.06)[6]		20.4 months (17.2 to 24.3)	⊕⊕⊖⊖ low[a, b]	There may be little or no difference in OS between PAZ and SUN.
Quality of life (QoL)	Subgroup data not available.						
Serious adverse events	Subgroup data not available.						
Progression-free survival (PFS)	1522 (5 RCTs)	**PEM + AXI**	n.a.[5]	5.4 months	-	-	-
		AVE + AXI	n.a.[5]		-	-	-
- Network (subnet 1) included 10 pairwise comparisons		**NIV + CAB**	n.a.[5]		-	-	-
		LEN + PEM	HR 0.33 (0.17 to 0.62)[6]		16.4 months (8.7 to 31.8)	⊕⊕⊕⊖ moderate[a]	LEN+PEM probably improve PFS, when compared to SUN.
- Median follow-up across trials[4]: 25 months		**NIV + IPI**	n.a.[5]		-	-	-
		CAB	n.a.[5]		-	-	-
- Median PFS with SUN across trials[2] in this network: 5.4 months		**PAZ**	n.a.[5]		-	-	-
Adverse events (grade 3 or 4)	Subgroup data not available.						
Time to initiation of first subsequent therapy	Subgroup data not available.						

CI: confidence interval; **HR:** hazard ratio.

GRADE Working Group grades of evidence
High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.
Moderate certainty: we are moderately confident in the effect estimate; the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low certainty: our confidence in the effect estimate is limited; the true effect may be substantially different from the estimate of the effect.
Very low certainty: we have very little confidence in the effect estimate; the true effect is likely to be substantially different from the estimate of effect.

[1] Basis for the assumed risks.
[2] The risk of SUN for OS and PFS was obtained from the included trials in the networks, respectively, and estimated by calculating the mean of all available medians for SUN.
[3] Method of calculating the assumed risks in the intervention group for survival outcomes: The median survival in the intervention group was calculated using the methods by Tierney 2007: Corresponding median survival in the intervention group (in months) = comparator group median survival time (in months) divided by the HR. Upper and lower confidence limits for the corresponding intervention risk were obtained by replacing HRs by their upper and lower confidence limits, respectively.
[4] Median follow-up across trials in the networks for OS and PFS, respectively, was estimated by calculating the mean of all available medians.
[5] Not applicable, comparison not available.
[6] Only direct evidence from only one trial.
[a] Downgraded by 1 level for study limitations because the one trial contributing all direct evidence is at high risk of bias.
[b] Downgraded by 1 level for imprecision because of a wide CI that favours either of the compared treatments.
[c] Downgraded by 2 levels for imprecision because of a very wide CI that includes values that favour either of the compared treatments, and evidence stems from only one trial with 157 participants.
[d] Downgraded by 1 level for imprecision because the evidence stems from only one trial with 157 participants.

BACKGROUND

Description of the condition

In 2020, it was estimated that 431,288 people were diagnosed with kidney cancer worldwide (ASCO 2022). The most common type of kidney cancer is renal cell carcinoma (RCC) (ASCO 2021). In the USA for example, kidney cancers account for 5% of all cancers in men and 3% of cancers in women (American Cancer Society 2022). It is estimated that in 2022, 79,000 new cases of kidney cancer (including the renal pelvis) will be diagnosed in the USA (50,290 estimated new cases in men and 28,710 estimated new cases in women) and that 13,920 people will die from this disease (American Cancer Society 2022; Siegel 2022). Males are twice as likely to be diagnosed with kidney cancer (with a lifetime risk for developing kidney cancer being 2.02%), as compared to females (with a lifetime risk for developing kidney cancer being 1.03%) (American Cancer Society 2022). The number of deaths in the USA in 2022 is estimated to be 13,920: 8,960 for men and 4,960 for women (American Cancer Society 2022; Siegel 2022). The five-year relative survival rates of all stages (i.e. local, regional, distant) are estimated at 76% (American Cancer Society 2022). For Germany, the Robert Koch Institute reported a kidney cancer incidence of 14,830 new cases in the year 2018, with an incidence rate of 15.4% in men and 7.6% in women. The mortality rate due to kidney cancer was 4.5% for men and 1.9% for women (Robert Koch Institute 2021). Moreover, kidney cancer was the most frequent tumour site for 3.5% of men and 2.4% of women in Germany (Robert Koch Institute 2021). For 2022, the Robert Koch Institute predicts 14,500 new cases of kidney cancer (36% in women and 64% in men).

With a 2:1 ratio, RCC presents predominantly in men and commonly develops after the 60th year of life (Rini 2009). Besides gender and age, further risk factors include an increased body mass index (BMI) (i.e. increased body weight) and active as well as passive smoking (Capitanio 2019; Rini 2009; Scelo 2018; Robert Koch Institute 2021). Important co-morbidity associated with an increased risk for developing this type of kidney cancer include hypertension, a history of kidney stones, type 2 diabetes, increased use of certain analgesics such as non-aspirin non-steriodal anti-inflammatory drugs, and several chronic liver and kidney diseases (Capitanio 2019; Rini 2009; Robert Koch Institute 2021; Scelo 2018). Physical activity is associated with a decreased risk of RCC (Robert Koch Institute 2021). Other factors, which may be protectively related to a risk for developing RCC, are fruit and vegetable and moderate alcohol consumption (Capitanio 2019; Rini 2009).

Staging of RCC is performed in accordance with the Union International Cancer Control (UICC) tumour, node, and metastasis (TNM) classification system (UICC 2017). First, the TNM system is used for classifying the tumour, where T stands for tumour (i.e. size and extent of the tumour); N for nodes (i.e. whether the cancer has spread to nearby lymph nodes); M for metastasis (i.e. whether the cancer spread to other organs (e.g. bones, brain, lungs). Thus, each category provides detailed information about the cancer, and a number (i.e. 1, 2 or 3) is assigned to each category, with a higher number indicating a more advanced cancer. Second, by combining these three categories and assigning a number to each, the overall cancer stage is determined (so-called group staging). Stages I to III are considered to be local or locoregional disease (depending on the group staging according to the TNM system: stage I includes T1; stage II includes T2; stage III includes T3 or T1-T3, and N1), and stage IV, which involves tumour spread beyond the renal/Gerota's fascia and/or distant metastases, to be advanced disease (stage IV includes T4 or N2 or M1) (Brierley 2016; Escudier 2019). While the overall five-year survival rates are approximately 76% (American Cancer Society 2022), the rates decrease drastically to 71% amongst individuals with locoregional disease (stage II and III, i.e. when the cancer has spread outside the kidney to nearby tissue and/or nearby lymph nodes), and to 14% for those with metastatic disease (stage IV, i.e. has spread to distant parts of the body) (ASCO 2022). Around a third of those affected will present with advanced disease. Furthermore, every fourth patient receiving treatment for localised RCC (stage I) will relapse and eventually develop distant metastases (???Choueiri 2017b; Dabestani 2016; Sun 2011).

Renal cell carcinoma is characterised by a variety of subtypes, the most common of which amongst adults are the clear cell type (75%), the papillary type (10%), and the chromophobe type (5%) (Lopez-Beltran 2009; Warren 2018). Of these three subtypes, the clear cell type is associated with the worst prognosis (Lopez-Beltran 2009; Warren 2018). For clear cell and papillary RCC, grading with prognostic value is commonly done by the International Society of Urological Pathology (ISUP) tumour grading system, which is adopted by the World Health Organization (WHO) and, therefore, also considered the ISUP/WHO grading classification system (Delahunt 2019). The validity of the grading systems with regard to the correlation of grade and outcome has not been shown for other subtypes, but these systems can be applied for descriptive purposes (Delahunt 2019). The ISUP/WHO grading system includes four stages, with classification based on the nucleus of the tumour cell: tumour cell nucleoli is absent or not clearly visible and basophilic at 400× magnification (grade 1); tumour cell nucleoli is clearly visible and eosinophilic at 400× magnification and visible but not prominent at 100× magnification (grade 2); tumour cell nucleoli is clearly visible and eosinophilic at 100× magnification (grade 3); tumour showing extreme nuclear pleomorphism, tumour giant cells and/or the presence of any proportion of tumour showing sarcomatoid and/or rhabdoid dedifferentiation (grade 4) (Delahunt 2019).

Renal cell carcinomas present in both local symptoms, including haematuria or flank pain, and systemic symptoms evoked, inter alia, through metastases. The latter may include, for example, hypercalcaemia, hypertension, erythrocytosis (increased numbers of red blood cells), and fever (Rini 2009). Nevertheless, renal cell carcinomas primarily present asymptomatically, meaning that today over half of renal cell carcinomas are discovered incidentally (Escudier 2019). Once advanced, they are associated with many symptoms, reduced health-related quality of life, and fatigue in those affected, especially when the disease progresses (de Groot 2018). For example, in a qualitative survey 46% of 287 participants reported psychiatric symptoms such as depressive symptoms and post-traumatic stress disorder. Due to poor survival rates, advanced renal cell carcinoma puts an immense burden on healthcare systems (Thekdi 2015).

Individuals with advanced RCC are categorised into favourable, intermediate, or poor risk groups. These are the common risk groups as defined by the International Metastatic RCC Database Consortium (IMDC) and the Memorial Sloan Kettering Cancer Center (MSKCC). The IMDC model (also known as Heng's model) determines the risk group based on the presence of six clinical factors: <1 year from time of diagnosis to systemic treatment; Karnofsky performance status < 80%; haemoglobin < lower limit of

normal; corrected calcium > upper limit of normal; neutrophils > upper limit of normal; platelets > upper limit of normal). For every factor that applies, one point (+1) is added. The risk group is then based on the total sum of points appointed (i.e., favourable risk = 0 points, intermediate risk = 1 to 2 points, poor risk = 3 to 6 points) (www.mdcalc.com/). The MSKCC model (also known as the Motzer model) includes five clinical factors: time from diagnosis to systemic treatment <1 year; haemoglobin < lower limit of normal; calcium >10 mg/dL (>2.5 mmol/L); lactate dehydrogenase (LDH) > 1.5x upper limit of normal; Karnofsky performance status <80%). The risk group is also based on the total sum of points appointed (i.e., favourable risk = 0 points, intermediate risk = 1 to 2 points, poor risk = 3 to 5 points) (www.mdcalc.com/).

Description of the intervention

Before 2005, treatment options for advanced RCC were limited to immunotherapies such as the cytokine therapies interferon (IFN)-alpha and interleukin (IL)-L. These are associated with many adverse events and with partial or complete remission rates of approximately 12%, they benefit only a small percentage of participants (Coppin 2004). Nowadays, targeted therapies such as tyrosine kinase inhibitors, and immunotherapies, such as immune checkpoint inhibitors, have emerged as an effective alternative, and the benefit of standard approaches, such as sunitinib or temsirolimus, over cytokine therapies with regard to mortality, quality of life, and adverse events in advanced renal cell carcinoma has been indicated (Unverzagt 2017). Multiple drugs such as sunitinib, sorafenib, bevacizumab, nivolumab, pazopanib, axitinib, cabozantinib, and everolimus have therefore been approved by the US Food and Drug Administration (FDA), mostly for second-line therapy, but several of them have been approved for first-line treatment as well. However, further novel therapeutic options could be associated with increased toxicities, which require consideration within an organised framework (Qin 2018).

For the first-line treatment setting, the National Comprehensive Cancer Network (NCCN) (Motzer 2022), the European Association of Urology (EAU) (Ljungberg 2022), the European Society for Medical Oncology (ESMO) guideline (Powles 2021), and the German guideline (Leitlinienprogramm Onkologie) all recommend the combination of pembrolizumab + axitinib (PEM + AXI) as the treatment option across all risk groups (i.e. favourable-, intermediate- or poor risk) for first-line therapy of advanced clear cell RCC. In addition, for the favourable risk group, the guidelines by NCCN, ESMO and EAU also list the combinations lenvatinib + pembrolizumab (LEN + PEM) or nivolumab + cabozantinib (NIV + CAB) as additional options (Ljungberg 2022; Motzer 2022; Powles 2021). For the intermediate- or poor risk groups, additional options can also be NIV + CAB, LEN + PEM or nivolumab + ipilimumab (NIV + IPI) (Ljungberg 2022; Motzer 2022; Powles 2021). The German guideline also lists NIV+IPI as an additional option for the intermediate or poor risk groups (Leitlinienprogramm Onkologie). In addition, the German guideline and the NCCN also suggest avelumab + axitinib (AVE + AXI) across all risk groups (Leitlinienprogramm Onkologie; Motzer 2022). Recommendations are also provided for situations when immune checkpoint inhibitors cannot be administered or tolerated. In such cases, targeted therapy is another option: pazopanib (PAZ) for IMDC favourable or intermediate/poor risk groups (Ljungberg 2022), and additionally cabozantinib (CAB) or sunitinib (SUN) for intermediate-, and poor-risk groups (Ljungberg 2022). The NCCN guideline recommends CAB, PAZ or SUN across all risk groups as possible options (Motzer 2022). The German guideline recommends bevacizumab + interferon (BEV+IFN), PAZ, SUN or tivozanib (TIV) for the favourable risk group; TIV, SUN, PAZ, CAB, or alternatively BEV+IFN for the intermediate risk group; CAB, SUN, or alternatively PAZ or temsirolimus (TEM) for the poor risk group, in cases where checkpoint inhibitors cannot be administered or tolerated (Leitlinienprogramm Onkologie). It should be noted that the recommendations of the German guidelines and the EAU guidelines are specifically for the IMDC risk groups.

Due to the high cost of targeted drugs and novel immunotherapeutic agents in cancer care, the economic burden of treatment of advanced RCC is enormous. Swallow 2018 reported additional cost per month of overall survival of USD 49,000 for cabozantinib and USD 24,000 for nivolumab compared to everolimus. On the other hand, Edwards 2018 analysed data from more than 4000 relapsed participants and showed that everolimus is cost-effective compared to best supportive care, with an incremental cost-effectiveness ratio (ICER) of GBP 45,000 per quality-adjusted life-year (QALY), as it is likely to be considered an end-of-life treatment. They reported that cabozantinib compared to everolimus might not be cost-effective, with an ICER of GBP 126,000 per QALY. In their economic analysis, nivolumab performed even worse than cabozantinib, as it was more costly but less effective.

How the intervention might work

In immunotherapy, which has as its primary aim to enhance the response of the immune system to the tumour cells, the classic, non-specific immunotherapeutic agents interleukin-2 (IL-2)—and especially interferon-alpha (INF-a)—have largely been replaced by novel agents. More advanced immunotherapeutics such as nivolumab, atezolizumab, and ipilimumab target specific immune checkpoints. Together with its ligand 1 (PD-L1), the programmed cell death protein 1 (PD-1) inhibits the immune response, the release of cytokines, and the cytotoxic function of T-cell lymphocytes (Harshman 2014). This PD-1—PD-L1 pathway is used by most renal cell carcinoma tumour cells to avoid the immune system (Aguiar 2018; Choueiri 2017b;?? Harshman 2014). Nivolumab, a monoclonal antibody, directly targets and binds the PD-1 receptor, thus stimulating the immune response against cancer cells. Another monoclonal antibody, atezolizumab, targets the PD-1—PD-L1 pathway by binding PD-L1, which then further prevents interaction of the receptor and its ligand (Keir 2007). Besides the PD-1—PD-L1 pathway, the cytotoxic T -ymphocyte-associated antigen 4 (CTLA-4) pathway has gained relevance in the treatment of renal cell carcinoma. The monoclonal antibody ipilimumab targets the CTLA-4 receptor, which is responsible for the regulation of tumour-specific T cell lymphocytes, and stimulates the immune response by inhibiting the regulatory function of CTLA-4 (Aguiar 2018; Sanchez-Gastaldo 2017).

Besides immunotherapeutic approaches, targeted therapies, which are aimed directly at preventing the growth and/or spread of cancer cells by targeting specific proteins or genes, are today an integral component of the treatment of advanced renal cell carcinoma. An effective target for such approaches is the vascular endothelial growth factor (VEGF) pathway that affects tumour angiogenesis, growth, and survival (Aguiar 2018). The monoclonal antibody and angiogenesis inhibitor bevacizumab directly targets and neutralizes VEGF. Another common target specifically used

by tyrosine kinase inhibitors is the VEGF receptor (VEGFR). Its neutralization inhibits angiogenesis as well. Because most tyrosine kinase inhibitors do not focus on the VEGF pathway only, for example to overcome resistance of the tumour to VEGFR inhibition alone, many of them are considered multikinase inhibitors (Aguiar 2018; Sanchez-Gastaldo 2017). This group includes the agents sunitinib, sorafenib, pazopanib, axitinib, and cabozantinib (Sanchez-Gastaldo 2017). Another important target for targeted approaches in the treatment of renal cell carcinoma is the mechanistic target of rapamycin (mTOR) pathway, which triggers cell growths and division. More precisely, mTOR is itself part of a protein complex which performs important tasks in cell growth and proliferation and subsequently in tumour angiogenesis and survival (Sabatini 2006). Both temsirolimus and everolimus inhibit the function of mTOR, and by these means deactivate the associated protein complexes (Sanchez-Gastaldo 2017). Among the afore-outlined agents, combinations within and across groups and mechanisms involved are common. INF-a, for example, is used in combination with bevacizumab, and has shown lower mortality rates as well as reduced side effects compared to INF-a alone, whereas it has not shown a difference in combination with temsirolimus compared to temsirolimus alone (Unverzagt 2017).

Why it is important to do this review

Our preliminary searches of the literature identified a great number of trials, including many ongoing trials that will be completed within the next years. In fact, we are aware of at least 36 published randomised controlled trials (RCTs) involving more than 10,000 participants, as well as 19 ongoing trials that have been registered in trial registries. This highlights the importance of a living systematic review approach, which applies all the detailed methods recommended by Cochrane, and is updated and republished whenever new evidence relevant to the review is identified (Elliott 2014). Such systematic and continuous updates of the available evidence ensure that recent findings are rapidly integrated into the body of evidence to support recommendations given in guidelines and to contribute to an up-to-date and high-grade decision support for effective therapeutic strategies for the individual patient.

However, recommendations can be complicated when economic arguments are introduced into discussions on the best strategy, because the related costs differ enormously per treatment option. This dissent provides the rationale for a network meta-analytic approach to the existing evidence for all available first-line therapy regimens. Although we are aware of several recently conducted network meta-analyses, none of these have analyzed indirect comparisons of all evaluable treatment options.

Lastly, as a critically necessary innovation within Cochrane, we planned to integrate evidence identified from clinical study reports (CSRs) into our systematic review and favoured this new source of evidence, where available, over the journal publication of eligible trials. Furthermore, as publication bias might influence all subsequent analyses and conclusions, all potential relevant trial registries were searched in detail to detect each conducted trial evaluating eligible drugs.

OBJECTIVES

The primary objective of this systematic review with network meta-analysis (NMA) was to evaluate and compare the benefits and harms of first-line therapies for adults with advanced renal cell carcinoma (RCC), and to produce a clinically relevant ranking of therapies.

The secondary objectives were to maintain the currency of the evidence by conducting continuous update searches, using a living systematic review approach, and to incorporate data from clinical study reports (CSRs).

METHODS

Criteria for considering studies for this review

Types of studies

We included randomised controlled trials (RCTs), both parallel-group RCTs and cross-over RCTs, in this review. For cross-over trials, we only extracted data from the first treatment period. We excluded cluster-RCTs as these do not fit with the aim of this review as we are interested in treatment benefit and harm in individuals, rather than in group effects. We also excluded quasi-randomised trials.

Where a clinical study report (CSR) for an individual eligible trial was available, we extracted available data on trial design and trial results from the CSR instead of the respective journal publications.

There was no limitation on trial eligibility with respect to the length of follow-up in individual trials.

Types of participants

We included trials involving adult participants (18 years of age or older) with a confirmed diagnosis of advanced renal cell carcinoma and (RCC) without previous systemic anticancer therapy, irrespective of gender and ethnicity of participants. Because first-line therapy only relate to participants with metastatic renal cell carcinoma, only trials including participants with metastatic disease were eligible.

We also included trials with previously treated participants in the total trial population if results for the previously untreated participants were separately extractable. However, when sufficient subgroup data were unavailable for untreated participants, we still extracted results from the entire trial population if less than 10% of participants have received previous systemic anti-cancer treatment.

Types of interventions

We included trials evaluating at least one of the following therapeutics without restrictions on the dose, dosage form, frequency, or duration of treatment, for example as shown below.

- Targeted therapy
  - Tyrosine kinase inhibitor (e.g. sunitinib, sorafenib, pazopanib, axitinib, cabozantinib, savolitinib, anlotinib)
  - mTOR inhibitor (e.g. temsirolimus, everolimus)
  - Angiogenesis inhibitor (e.g. bevacizumab, levantinib)
- Immunotherapy
  - Checkpoint inhibitors (e.g. atezolizumab, avelumab, nivolumab, ipilimumab, pembrolizumab)
  - Interferon
  - Interleukin
- Placebo

- any combination of the above (e.g. nivolumab + ipilimumab, avelumab + axitinib, pembrolizumab + axitinib)

We included trials evaluating at least one targeted therapy or immunotherapy in at least one intervention arm to provide up-to-date results. We excluded trials evaluating these agents in an adjuvant setting. We also excluded trials that assessed the comparison of interleukin versus interferon only. Instead, we only included trials with interleukin and interferon when given in combination with another substance (e.g. interferon-alpha (IFN-a) + bevacizumab) or when compared to another substance (e.g. IFN-a versus sunitinib).

We analysed interventions for favourable-risk groups separately from interventions for intermediate- and poor-risk groups (intermediate- and poor-risk groups were combined). Moreover, we analysed risk groups according to IMDC and MSKCC criteria separately (see Differences between protocol and review). All interventions were analysed using direct and indirect comparisons. When no direct evidence from randomised trials was available, but the trials were considered sufficiently similar with respect to the participant population, indirect estimates of intervention effects were obtained by means of network calculations. In the protocol of this review, we pre-specified that different doses of the same drug will be combined to single drug categories if these would differ. However, most interventions were administered at the same dose across trials (see Table 1 in Results).

We included sunitinib as our main comparator as it is a widely used tyrosine kinase inhibitor and is often used as the comparator drug in trials. For the transitivity assumption to hold true, we assessed the administration routes, the dosage and the discontinuation rates of this comparator in each trial (Salanti 2012). In the protocol of this review we had pre-specified that we would create networks of trials with the same administration route and average dose if these would differ. However, in all included trials that assessed sunitinib, the drug was provided via the same administration route (oral) and the administration dose was 50mg/ day in all trials (see Table 1 in Results).

Types of outcome measures

We included all trials fulfilling the inclusion criteria defined above, irrespective of the reported outcomes. To inform this review and to ensure that we assess outcomes that are most relevant to adults with advanced renal cell carcinoma, during the protocol development of this review, patients and patient representatives were invited in a two-hour session to discuss relevant outcomes from their perspectives. The following outcomes and order of outcomes (i.e., primary and secondary outcomes) were prioritised together with the patients and patient representatives during the workshop.

Primary outcomes

- Overall survival (OS), defined as the time from random treatment assignment to death from any cause
- Quality of life (QoL), assessed with validated and reliable instruments
- Serious adverse events (SAEs)*, assessed as the number of participants with at least one event

We prioritised OS, QoL, and SAEs as our primary outcomes together with the participants and patient representatives, who regarded these outcomes as most relevant, and also because they are a direct measure of treatment benefit. Furthermore, OS can be considered the most robust endpoint as it does not require blinding.

*An adverse event that results in death or is life-threatening.

Secondary outcomes

- Progression-free survival (PFS), defined as the time interval from randomisation to the first confirmed disease progression, disease relapse, or death from any cause, or to the last time point of follow-up
- Adverse events (AEs), assessed as the number of participants with at least one event
- Number of participants who discontinued study treatment due to an AE

We included PFS as a secondary outcome as it is commonly used to assess stable disease.

With regard to AEs, we assessed severity grades 3 and 4** in the number of participants with at least one AE. We only extracted data on AEs that were labelled as 'all-cause' AEs; hence, we did not extract data when AEs were labelled as 'treatment-related'. In addition, we put a special focus on specific AEs that were regarded as most relevant by the participants and patient representatives. These included: hand-foot syndrome, fatigue, diarrhoea, vomiting, loss off appetite, weight loss, mucous membrane damage (generic term; we looked at mucosal inflammation and stomatitis separately), insomnia, and depression. We extracted data for these specific AEs separately.

In the protocol for this review, we had stated that we would, additionally, extract all individual AEs reported in the included studies, as well as their frequency of occurrence. However, this was not feasible (see Differences between protocol and review).

**Severity grading according to Common Terminology Criteria for Adverse Events (CTCAE). Trials usually report grades 3 and 4 together, and a severe AE (grade 3 or 4) does not necessarily need to be considered serious.

- Time to initiation of the first subsequent anticancer therapy (TFST), defined as the time from initiation of first-line chemotherapy until the start of subsequent therapy or death

Method and timing of outcome measurement

We analysed OS and PFS as time-to-event outcomes, and included results representing the longest follow-up time available. The outcome TFST was not reported as a time-to-event outcome in the included trials. Pooling of this outcome was not feasible, so we report results narratively.

For QoL, we initially accepted all validated instruments, and we would have calculated standardised mean differences (SMD) instead of mean difference (MD) when scales used between trials differed (see Measures of treatment effect). However, during the conduct of this review, we decided to prioritise scales for the assessment of this outcome because we initially identified a total of 25 scales and sub-scales across trials that were used to assess QoL. Due to this high heterogeneity, we decided to prioritise scales that are most clinically relevant and used in clinical daily practice. To prioritise QoL-scales, two review authors (AA, ET) first

created a list of all scales that were reported in the included trials, which was then provided to two co-authors with a clinical background (AH, PD), who ranked them by assigning them to either low priority, medium priority or high priority based on clinical relevance. Prioritisation was further guided by a third clinician (PM) on the author team and there was discussion amongst author team members (AA, AH, ET, PM, PD) via teleconference. Ultimately, the following scales were prioritised to extract data for QoL:

- the Functional Assessment of Cancer Therapy Kidney Cancer Symptom Index – Disease Related Symptoms (FKSI-DRS);
- the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-C30 (QLQ-C30);
- the EuroQol Visual Analogue Scale (EQ-VAS);
- the Functional Assessment of Cancer Therapy General (FACT-G);
- the Functional Assessment of Chronic Illness Therapy Fatigue (FACIT-F).

We grouped the measurement time points of QoL into those measured directly after initiation of treatment up to four weeks after initiation treatment, medium-term outcomes (1 month up to 12 months after initiation of treatment), and longer-term outcomes (over one year after initiation of treatment). Where available, we also extracted data at the end of treatment.

We included all other outcome categories for the observational periods reported in the CSRs or trial publications. We planned to include AEs and SAEs occurring during active treatment as well as long-term AEs and SAEs. However, we were not able to extract long-term AEs or SAEs, and we could also not group the timing of outcome measurements as we had pre-specified in the protocol, because in the publications of the trials it was not stated which time points were being reported. Hence, for AEs and SAEs, we extracted data for events that occurred during the time of treatment.

Outcomes to be included in GRADE summary of findings table

During the development of this protocol, participants and patient representatives were invited to share their opinions and perspectives regarding the most patient-relevant outcome measures to be included in this review. The most relevant outcome categories, to be included in summary of findings tables, were OS, QoL, SAEs, PFS, AEs, and TFST.

Search methods for identification of studies

We adapted all search strategies for electronic database searches and searching other sources from those suggested in the *Cochrane Handbook for Systematic Reviews of Interventions* and in accordance with the specified recommendations therein (Lefebvre 2019). We applied no language restriction in order to reduce language bias. All abstracts were available in English.

Electronic searches

Searching for clinical study reports

For this systematic review, the inclusion of trial design and results data from clinical study reports (CSRs) was preferred above the respective journal publications. The search method was initiated by the identification of the sponsors of the included clinical trials. This was done by referring to the clinicaltrials.gov platform (www.clinicaltrials.gov/). After identification of the respective sponsor, the possibility of a direct request for CSRs was checked. Furthermore, the availability of the CSRs on the manufacturer's platform was verified. To complement the search method, the following data platforms were enclosed for the search of the CSRs: the European Medicines Agency (EMA) 'clinical data platform' (clinicaldata.ema.europa.eu/web/cdp/home), the Yale University Open Data Access (YODA) platform (yoda.yale.edu/), the clinical data study request (CSDR) platform (clinicalstudydatarequest.com), and the Vivli platform (https://vivli.org). Initially, the FDA platform was intended to be included, however it was indicated to have insufficient data, as the platform remained in its pilot stage during the search process. The EMA 'clinical data platform' was searched for active substances of the included clinical trials. This search offered an overview of all available trials encompassing the respective active substances, and subsequently we screened the search for the trials included in this review and for available CSRs to these trials. The YODA platform allows utilising the NCT (i.e. the clinicaltrials.gov registry number) within the search process. This approach was exclusively performed for this particular platform. The CSDR platform was used to search and request for CSRs. The search process was done by searching for active substances of the included clinical trials. The CSDR platform only offers CSRs from its members; hence, requests are also only possible to be made if the sponsor of all included clinical trials is an official member of the platform. The pharmaceutical company Bayer is excluded from this particular case, as it is a member of the CSDR, however does not offer the opportunity to take in requests. Two types of requests were offered by the CSDR: 1. datasets that are not yet shared on the CSDR platform and 2. trial documents only. Almost all requests that were made throughout this search process included both types. In total, 21 requests were made, and 19 requests included both types. The final platform utilised for this search method was Vivli. The search process included searching for key terms such as "renal", "kidney", and the active substances, and complementary the NCT was used to find available CSR. One request on the Vivli platform was made.

Ultimately, we identified two CSRs to two trials (NCT00334282; NCT00720941) and one scientific summary result to one trial (NCT01064310) through the CSDR platform. The CSRs and the scientific result summary were used for data extraction and to inform risk of bias assessment.

Electronic database searches

We searched the following databases/sources to identify eligible trials.

- Databases of medical literature:
  - Cochrane Library, including the Central Register of Controlled Trials (CENTRAL), 2022 issue 02, see Appendix 1 and Appendix 2);
  - MEDLINE (Ovid, from 1946 up to 9 February 2022, see Appendix 3 and Appendix 4);
  - Embase (from 1974 up to 9 February 2022, see Appendix 5 and Appendix 6).
- Conference proceedings of annual meetings of the following societies (included in CENTRAL):
  - American Society of Clinical Oncology (ASCO);
  - European Society of Medical Oncology (ESMO).

As publication bias might influence all subsequent analyses and conclusions, we searched all potential relevant trial registries in

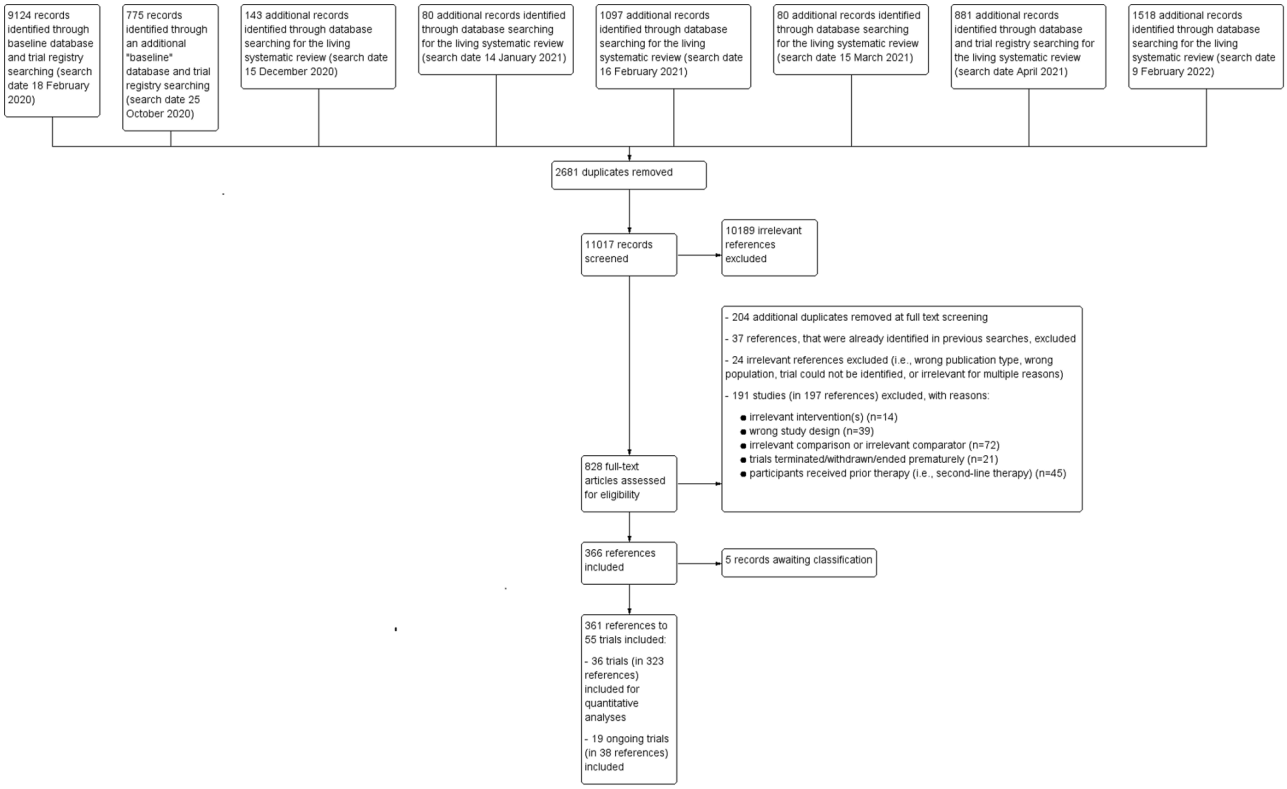
detail to detect ongoing as well as completed studies that have not yet been published. It is mandatory today for the type of studies eligible for inclusion in this review to provide results at least in the study registry (United States Congress 2007; World Medical Association). When results were not published elsewhere, data from the trial registries were extracted and analysed.

- Trial registries to identify ongoing trials and results of completed trials (up to 9 February 2022), see Appendix 7 and Appendix 8:
  - ISRCTN registry (www.isrctn.com);
  - EU Clinical Trials Register (www.clinicaltrialsregister.eu/ctr-search/search);
  - US National Institutes of Health Ongoing Trials Register ClinicalTrials.gov (www.clinicaltrials.gov/);
  - WHO ICTRP (https://www.who.int/clinical-trials-registry-platform).

Living systematic review considerations

We first conducted baseline review searches in February and October 2020. Starting from December 2020, after publication of the protocol for this review, we ran monthly update searches until April 2021 (Figure 1). Together with the clinical experts on this review, we decided to stop the update-searches in April 2021 in order to be able to finalise data extraction and risk of bias assessments. However, one final update search was conducted on 9 February 2022, as searches for intervention reviews should not be older than 12 months at publication.

Figure 1. Flow diagram



Search strategies for electronic databases were reviewed yearly to ensure that they reflected any terminology changes in the topic area, the databases, or the eligibility criteria of the review. In addition, our primary search strategy (see Appendix 4), developed by our Information Specialist, was peer-reviewed by another Information Specialist (see Acknowledgements). We searched trial registries every six months.

Searching other resources

If needed, we would have extended the electronic searches by handsearching the references of all identified trials and relevant review articles. However, all relevant trials and articles were identified by our electronic searches.

Living systematic review considerations

We planned to search additional sources only yearly, as novel RCTs in this field are included in study registers or databases and thus were identified by our electronic searches.

Data collection and analysis

Selection of studies

Three review authors (AAa, MG, VP) screened citations retrieved by the baseline searches. The following monthly/update searches were screened by two review authors (AAa, BB). All records were assessed immediately for eligibility by reading the abstracts using Covidence software (Covidence). In case of disagreement on the relevance of a citation, we obtained the full-text of the respective article for further review. We then eliminated all articles that did not

meet the eligibility criteria and obtained the full-text articles of the remaining articles. We proceeded similarly with the electronically and manually gathered registry entries as well as any reports identified from CSR databases. Subsequently, the full-text articles were screened. Both at title and abstract screening and at full-text screening, the four review authors (AA, BB, MG, VP) screened the references independently and any disagreements were resolved by discussion.

We documented the overall numbers of trials identified, excluded, and included at every stage of the search and screening of the literature in a PRISMA flow diagram (Figure 1).

We listed all eligible trials in the Characteristics of included studies section of the full review irrespective of whether measured outcome data were reported in a way that allows inclusion into a quantitative analysis. We recorded excluded trials in the Excluded studies section; trials that are ongoing with no results available in the Characteristics of ongoing studies section; and trials that are completed with no result data available, and where eligibility for inclusion was unclear, in the Studies awaiting classification section. We considered completed trials for which no results are available narratively in our publication bias judgements (see Assessment of reporting biases).

Clinical study report considerations

Besides our primary search for eligible and available CSR in the databases of pharmaceutical manufacturers, the EMA database, the FDA database, the YODA database and the CSDR, we searched specifically for additional reports on the trials identified by our searches. When a CSR that was linked to a primary trial publication could be retrieved, we preferred any data given in the report over the respective data from the clinical trial publication. For informational purposes, we would have reported, if found, discrepancies between the CSR and the clinical study publication in a separate table.

Living systematic review considerations

Two review authors (AA, BB) screened any new citations retrieved by the monthly searches immediately for eligibility by reading the abstracts and following all afore-outlined steps. With every update search, we documented overall numbers of additionally identified trials and references in an updated PRISMA flow diagram (Moher 2009) (Figure 1).

Data extraction and management

We performed data extraction in accordance with the guidelines proposed by Cochrane (Li 2019). In total, five review authors were involved in data extraction of outcome data and trial characteristics (AAa, BB, CH, ET, ND), and independently extracted data from CSRs and study publications using a standardised data extraction form. Each outcome was extracted by two review authors independently; extractions were then compared to detect and resolve any discrepancies.

We extracted the following items.

- **General information**: author, title, source, publication date, country, language, duplicate publications.
- **Quality assessment**: (see Assessment of risk of bias in included studies).
- **Trial characteristics**: trial design, aims, setting and dates, source of participants, inclusion/exclusion criteria, comparability of groups, subgroup analysis, statistical methods, treatment cross-overs, compliance with assigned treatment, length of follow-up, time point of randomisation.
- **Participant characteristics**: age, gender, number of participants recruited/allocated/evaluated, participants lost to follow-up, stage of disease, histologic type, site of metastases, concomitant therapy.
- **Interventions**: type, dosage, duration, and administration route of therapy; type, dosage, duration, and administration route of therapy in control arm; concomitant therapy; duration of follow-up.
- **Outcomes**: all outcomes mentioned above (including assessment of causality, relationship between intervention and adverse drug reaction, how severity or seriousness was measured).
- **Additional information**: sponsorship/funding for the trial, potential conflicts of interest, trial registry record information (e.g. NCT numbers).

For cross-over RCTs, we only extracted results from the first treatment period (i.e. before treatment cross-over).

Some of the above-mentioned characteristics (age, sex, histologic type, site of metastases (i.e. lung, bone, liver), administration route and dosage of substances) are potential effect modifiers for which we extracted data to check for validity of the transitivity assumption (see Assessment of heterogeneity).

We collated all reports of the same trial so that each trial, rather than each report, was the unit of interest. This applied to trial publications, conference abstracts, trial registry information and CSRs/scientific result summaries. For all studies, except three, we used published data only (e.g. published in full-text articles, abstracts or on trial registries). For the other three studies (NCT00720941; NCT00334282; NCT01064310), we used unpublished data from CSRs/scientific result summaries).

Assessment of risk of bias in included studies

We used the Risk of Bias 2.0 (RoB 2) tool to assess the risk of bias in the underlying trial results of the included RCTs (Sterne 2019). For cross-over RCTs, from which we extracted data from the first treatment period, we also used the standard RoB 2 tool for parallel-group RCTs, as suggested in Chapter 23 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2019a).

Of interest in this review was the effect of the assignment to the intervention (the intention-to-treat effect); hence, we performed all assessments with RoB 2 on this effect. We assessed the risk of bias of all trials that contributed results to the analyses of the outcomes overall survival (OS), : progression-free survival**(**PFS), adverse events (AEs) and serious adverse events (SAEs). As we initially assumed that analyses for quality of life (QoL) would also be feasible, we assessed the risk of bias for this outcome as well (time point: QoL at the end of treatment), although we ended up reporting this outcome narratively in this review. Furthermore, risk of bias was assessed for the total population (i.e. all risk groups combined) for the following outcomes: OS, PFS, AEs, SAEs and QoL. For OS and PFS, we additionally assessed the risk of bias for each risk group (i.e. favourable, intermediate or poor risk group

per International Metastatic RCC Database Consortium (IMDC) or Memorial Sloan Kettering Cancer Center (MSKCC)) separately.

In total, four review authors (AAa, BB, ET, ND) were involved in the risk of bias assessments. Two review authors independently assessed the risk of bias for a specific outcome result, and assessments were then compared to detect disagreements. When disagreements arose and the review authors were unable to reach a consensus by discussion, a third review author was consulted to reach a final decision. We assessed the following types of bias as outlined in Chapter 8 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2019), using the RoB 2 Excel tool (available at riskofbiasinfo.org):

- Bias arising from the randomisation process.
- Bias due to deviations from the intended interventions.
- Bias due to missing outcome data.
- Bias in measurement of the outcome.
- Bias in selection of the reported result.

To address these types of bias, we employed the signalling questions recommended in RoB 2 and made a judgement using the following options;

- 'yes': if there is firm evidence that the question is fulfilled in the trial (i.e. the trial is at low or high risk of bias for the given direction of the question);
- 'probably yes': a judgement has been made that the question is fulfilled in the trial (i.e. the trial is at low or high risk of bias for the given direction of the question);
- 'no': if there is firm evidence that the question is unfulfilled in the trial (i.e. the trial is at low or high risk of bias for the given direction of the question);
- 'probably no': a judgement has been made that the question is unfulfilled in the trial (i.e. the trial is at low or high risk of bias for the given direction of the question);
- 'no information' if the trial report provides insufficient information to permit a judgement.

We used the algorithms proposed by RoB 2 to assign each domain one of the following levels of bias.

- Low risk of bias
- Some concerns
- High risk of bias

Subsequently, we derived a 'Risk of bias' rating for each prespecified outcome in each trial in accordance with the following suggestion.

- 'Low risk of bias': the trial is judged to be at low risk of bias for all domains for this result.
- 'Some concerns': the trial is judged to raise some concerns in at least one domain for this result, but not to be at high risk of bias for any domain.
- 'High risk of bias': the trial is judged to be at high risk of bias in at least one domain for the result OR the trial is judged to have some concerns for multiple domains in such a way that substantially lowers our confidence in the results.

We stored and presented our consensus decisions for the signalling questions of RoB 2 in the appendices (Appendix 9; Appendix 10; Appendix 11; Appendix 12; Appendix 13). We used the online available visualisation software robvis to summarise and visually present our assessments. We created traffic light plots (domain-level judgements) and summary plots (distribution of judgements within each domain) for each outcome.

Measures of treatment effect

Relative treatment effect

We conducted all analyses on the effect of the randomised intervention (intention-to-treat effect).

To estimate effects in binary outcomes, we extracted the number of participants and events per arm and calculated risk ratios (RRs) with corresponding 95% confidence intervals (CIs) for each trial individually. However, this was not possible for one outcome reported in this review ('TFST') because, firstly, the definition of this outcome varied between trials. Furthermore, it was unclear which time points were being reported. Therapies for advanced renal cell carcinoma are usually long-term therapies, and people may stop therapy at very varying time points. Hence, this would have led to high heterogeneity and thus, pooling data were not feasible. Therefore, we decided to report results for this outcome narratively in a tabular form.

For time-to-event outcomes, we extracted hazard ratios (HRs) and their corresponding measures of statistical uncertainty to directly retrieve an HR and a corresponding 95% CI for each individual trial. If this information had not been included in individual trial reports, we would have used the methodology proposed by Parmar 1998 and Tierney 2007 to reconstruct HRs indirectly from the information given in the trial report. We considered the following hierarchy of direct and indirect reconstruction methods, according to which HR from individual trials is preferred (Tudur 2001).

1. Unadjusted direct estimates (e.g. log HR and variance).
2. Indirect calculation 1: log HR and CI.
3. Indirect calculation 2: log-rank P value and number of events.
4. Indirect calculation 3: estimating the log HR and variance from survival curves.

It is important to note here that the directly extracted HR as well as its different reconstruction methods produce either adjusted or unadjusted HR. For our calculations, we preferred unadjusted HR. If we would have needed to reconstruct HRs, we would have conducted sensitivity analyses to explore the effect of different reconstruction methods on our findings whenever necessary (see Sensitivity analysis). For two trials, however, we had to recalculate the CI to obtain a 95% CI, as one trial (NCT01108445) provided a 80% CI and the second trial (NCT00732914) provided a 90% CI.

For both binary and time-to-event outcomes, we clearly indicated the direction of the effect in the SoF table along with the individually reported outcomes (i.e. 'RR or HR smaller than 1.0 favours the intervention'), as this has led to confusion and flawed reporting of review results in the past (Skoetz 2019).

For continuous outcomes we had planned to use the mean difference (MD) when the same instruments were used for assessments; otherwise we would have calculated the

standardised mean difference (SMD) with corresponding 95% CIs. We would have interpreted an SMD of zero as equivalent effects between the experimental and the control intervention. Depending on whether an improvement in the outcome of interest was associated with a higher or lower score, an SMD greater or lower than zero would be associated with a positive effect of the experimental intervention over the control intervention. This would have applied to the outcome QoL.

Absolute treatment effect

In addition to the relative measures of treatment effect outlined above, we presented absolute effect measures (Skoetz 2020) for every network estimate. To keep the SoF simple, understandable and reader-friendly, we refrained from adding number needed to treat for an additional beneficial (NNTB) outcome/ number needed to treat for an additional harmful (NNTH) outcome to the SoF (see Differences between protocol and review).

Unit of analysis issues

Trials with multiple treatment groups

As recommended in Chapter 23 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2019a), for trials with multiple treatment groups, we planned to combine arms as long as they could be regarded as subtypes of the same intervention. When arms could not be pooled this way, we included multi-arm trials using a network meta-analysis approach that accounted for the within-trial correlation between the effect sizes by re-weighting all comparisons of each multi-arm trial (Rücker 2012; Rücker 2014). For pairwise meta-analyses, if conducted, we would have treated multi-arm trials as multiple independent comparisons and not combine these data in any analyses.

Cross-over trials

For cross-over RCTs, we only extracted results from the first treatment period, thereby treating these trials as parallel-group RCTs for network meta-analysis (Higgins 2019c).

Dealing with missing data

As suggested in Chapter 10 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Deeks 2022), we planned to take the following steps to deal with missing data.

When only percentages but no absolute number of events were reported for binary outcomes, we calculated numerators using percentages. When data were not reported numerically but were reported graphically, we tried to estimate missing data from figures.

If needed, we would have contacted the original investigators to request relevant missing data. If the number of participants evaluated for a given outcome would not have been reported, we would have used the number of participants randomised per treatment arm as the denominator. If estimates for mean and standard deviations had been missing, we would have calculated these statistics from reported data, using the approaches described in Chapter 6 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2019b). If standard deviations had been missing and if we would have not been able to calculate them from the reported data, we would have calculated values according to a validated imputation method (Furukawa 2006).

For binary and continuous outcomes, if needed, we would have performed sensitivity analyses based on assumptions to assess how robust the analysis results are to missing data (Guyatt 2017). We addressed the potential impact of missing data in the risk of bias assessment, the grading of the evidence, and the discussion. As there is currently no procedure available permitting the quantitative assessment of the sensitivity of time-to-event outcomes to issues of missing data, we used the visual inspection of survival curves to evaluate any potential bias introduced by missing outcome data on such outcomes. We regarded this in the risk of bias section of the review, our grading of the outcomes, and the discussion.

Assessment of heterogeneity

Assessment of clinical and methodological heterogeneity within treatment comparisons

We planned to evaluate the presence of clinical and methodological heterogeneity through subgroup and sensitivity analyses (see Subgroup analysis and investigation of heterogeneity).

Assessment of transitivity across treatment comparison

To assess the adequacy of the assumption of transitivity, we evaluated whether the included interventions were similar when they were evaluated in RCTs with different designs: for example, whether double-drug combinations are administered the same way in trials comparing them to other double-drug combinations and in those comparing double-drug combinations to triple-drug combinations. Further potential effect modifiers of interest were age, sex, histology type, site of metastases (i.e. lung, bone, liver), administration routes, and dosage of substances. We evaluated the transitivity assumption by visually assessing the distribution of these across the different pairwise comparisons and explored their potential influence in subgroup analyses, whenever possible (see Subgroup analysis and investigation of heterogeneity).

Assessment of statistical heterogeneity and inconsistency

A critical prerequisite for a valid network analytic approach to the available evidence is consistency of effects within the network. This refers to sufficient agreement amongst the direct and indirect effect on the same comparisons (White 2012; Puhan 2014). To evaluate the presence of heterogeneity and inconsistency in the entire network, we provided the generalised heterogeneity statistic Q_{total} and the generalised I^2 statistic, as described in (Schwarzer 2015). We used the decomp.design command in the R package netmeta for decomposition of the heterogeneity statistic into a Q statistic for assessing the heterogeneity between trials with the same design, and a Q statistic for assessing design inconsistency to identify the amount of heterogeneity/inconsistency within as well as between designs (R Core Team 2019; Rücker 2019).

To evaluate the presence of inconsistency locally, we compared direct and indirect treatment estimates of each treatment comparison. This served as a check for consistency of a network meta-analysis (Dias 2010). For this purpose, we used the netsplit command in the R package netmeta, which enables the splitting of the network evidence into direct and indirect contributions (R Core Team 2019; Rücker 2019). For each treatment comparison, we presented direct and indirect treatment estimates plus the network estimate using forest plots. In addition, for each comparison, we gave the Z value and P value of test for disagreement (direct

versus indirect). It should be noted that in a network of evidence there may be many loops, and with multiple testing there is an increased likelihood to find an inconsistent loop by chance. We were, therefore, cautious in deriving conclusions from this approach.

Furthermore, we created a net heat plot (Krahn 2013), a graphical tool for locating inconsistency in network meta-analysis.

We planned to explore possible sources of heterogeneity by performing prespecified sensitivity and subgroup analyses, whenever possible (Subgroup analysis and investigation of heterogeneity; Sensitivity analysis). In addition, we reviewed the evidence base and discussed the potential role of unmeasured effect modifiers in order to identify additional sources of heterogeneity.

We interpreted I^2 values according to Chapter 9 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Deeks 2022), as follows:

- 0% to 40% might not be important;
- 30% to 60% may represent moderate heterogeneity;
- 50% to 90% may represent substantial heterogeneity;
- 75% to 100% represents considerable heterogeneity.

We used the P value of the Chi^2 test only for describing the extent of heterogeneity and not for determining statistical significance. In addition, we reported Tau^2, the between-study variance in random-effects meta-analysis.

Assessment of reporting biases

We searched trial registries to identify completed trials that have not been published elsewhere in order to minimise or determine publication bias. In the protocol for this review, we pre-specified that for meta-analyses involving at least 10 trials, we would explore potential and small-study effects by generating a funnel plot and assessing it using a linear regression test (Egger 1997). We would have considered a P value of < 0.1 as significant for this test. However, in this review, we did not have direct comparisons that involved more than 10 trials.

As the identified evidence was sufficient, and a natural common comparator exists for the interventions in a single outcome, we planned to use a 'comparison-adjusted' funnel plot to support our judgements on potential for publication bias within the network meta-analysis (Chaimani 2012; Chaimani 2013). This type of funnel plot allows for the inclusion of all trials in a given network regardless of the respective interventions under trial. However, creating such a funnel-plot was not feasible for this review (see Differences between protocol and review). Hence, in accordance with the advice given in the *Cochrane Handbook*, our judgements on potential publication bias were primarily non-statistical (Chaimani 2019).

Data synthesis

We included all eligible trials in our analyses, but conducted sensitivity analyses according to risk of bias ratings (low bias/some concerns versus high risk of bias; see Sensitivity analysis) for our primary outcomes. We analysed interventions for favourable-risk groups separately from interventions for intermediate- and poor-risk groups. Furthermore, we analysed risk groups according to the MSKCC criteria separately from risk groups according to the IMDC criteria.

Direct comparison of interventions

If data had been insufficient to be combined in network meta-analyses (e.g. in the case of inconsistency), and the clinical and methodological characteristics of individual studies sufficiently homogeneous, we would have performed pairwise meta-analyses with an overall estimate, according to the recommendations provided in Chapter 10 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Deeks 2022). Had we conducted pairwise meta-analyses, we would have used R package meta for statistical analyses (R Core Team 2019; Schwarzer 2007). We would have used a random-effects model, and calculated corresponding 95% CIs for all analyses, as some heterogeneity in trial design, interventions, and outcome measurements can be expected.

To outline the available direct evidence (even when network meta-analyses were conducted), we provided forest plots for pairwise comparisons, but without giving an overall estimate.

Indirect and mixed comparison of interventions

We considered the data to be sufficiently similar to be combined and, therefore, performed network meta-analyses using the frequentist weighted least squared approach described by Rücker 2015. We used the R package netmeta for statistical analyses (R Core Team 2019; Rücker 2019), and we used a random-effects model, taking into account the correlated treatment effects in multi-arm trials. We assumed a common estimate for the heterogeneity variance across the different comparisons. We created and provided visual network plots (i.e., network graphs) for all analyses of our primary and secondary outcomes to evaluate the extent to which treatments are connected within a network, and also to visually present whenever networks were not fully connected. When a network was not fully connected and consisted of two or more sub-networks, we analysed each sub-network separately. Analyses were conducted whenever a sub-network included more than one trial. We provided visual network plots and forest plots with results for each sub-network analysis. In the network plots, any two treatments were connected by a line when there was at least one trial comparing the two treatments. The line width represents the number of trials within a comparison, while the plot width represents the number of participants within that comparison. For each comparison, we gave the estimated treatment effect along with its 95% CI. We graphically presented the results using forest plots, with sunitinib (SUN) as the reference treatment.

To evaluate the transitivity assumption, we visually assessed the distribution of important effect modifiers across the different pairwise comparisons and explored their potential influence in subgroup analyses, whenever possible (see Assessment of heterogeneity). To check for consistency, we compared direct and indirect treatment estimates of each treatment comparison (see Assessment of heterogeneity). Our assessment and judgement of potential publication bias was primarily non-statistical (see Assessment of reporting biases).

Relative treatment ranking

We obtained a ranking of treatment options using P-scores (Rücker 2015). P-scores allow ranking treatments on a continuous 0-to-1

scale in a frequentist network meta-analysis. We provided a ranking of treatments for the outcomes OS, SAEs, PFS, all-cause grade 3 or 4 AEs (including the individual AEs explored in this review), and for the outcome number of participants who discontinued treatment due to an AE. Data for the outcomes QoL and TFST were not analysed, hence, we could not calculate P-scores for these outcomes as we had initially planned.

Subgroup analysis and investigation of heterogeneity

We intended to perform subgroup analyses for our primary outcomes OS, SAE and QoL and for the following characteristics:

- age of participants (≤ 65 versus > 65);
- sex of participants (male versus female);
- histology type (clear cell type, papillary type, sarcomatoid type);
- nephrectomy (yes versus no);
- radiotherapy (yes versus no);
- follow-up times (< 5 years versus ≥ 5 years);
- site of metastases (lung, bone, liver);
- administration routes (oral versus intravenous);
- dosages (clinically relevant dose categories).

However, most subgroup analyses were not possible due to the distribution of these characteristics in the included trials, and a lack of reporting on subgroup data (for more details see Differences between protocol and review).

Sensitivity analysis

We planned to conduct sensitivity analyses for our primary outcomes OS, QoL, and SAEs.

To test the robustness of the results, we conducted fixed-effect network meta-analyses. As a post-hoc decision, this was also completed for the outcome PFS.

Furthermore, we conducted sensitivity analyses on quality components (overall low risk of bias or some concerns versus overall high risk of bias) and sensitivity analyses on whether the assumption of proportional hazards underlying the HR had been tested and was justified in primary trials.

We had also planned to explore the influence of trial design (blinded trials versus unblinded trials) and the influence of completed but not published trials. For time-to-event outcomes, we had planned to use sensitivity analyses to explore the robustness of our findings should variable techniques to reconstruct HR from primary trial reports be necessary. However, some of these pre-specified sensitivity analyses were not possible (for more details see Differences between protocol and review).

Methods for future updates

We planned to review the scope and methods approximately yearly, or more frequently if appropriate in light of potential changes in the topic area or the evidence included in the review (e.g. when additional comparisons, interventions, subgroups, or outcomes, or new review methods become available) (Garner 2016).

Summary of findings and assessment of the certainty of the evidence

Summary of findings tables

We include 'Summary of findings (SoF) tables to present the main findings of the review in a transparent and simple tabular format. In particular, we included key information concerning the certainty of the evidence, the magnitude of effect of the interventions examined, and the sum of available data on the outcomes. We reported the following outcomes and time points in the SoF table: OS and PFS (as time-to-event) at the longest follow-up available and AEs and SAEs that occurred during treatment. The outcomes QoL and TFST were not meta-analysed, so we reported them narratively in the SoF.

We initially planned to create two networks (one for the favourable-risk group and one combined for the intermediate- and poor-risk groups) and also to present one SoF table, respectively, for participants with a favourable risk and one combined for participants with an intermediate or poor risk. During the conduct of this review, however, we decided to additionally analyse and report results for the IMDC and MSKCC risk groups separately, as well as to provide a combined analysis (and SoF) of all risk groups combined (i.e. an overall analysis with the total trial populations); for more details, see Differences between protocol and review). Thus, in this review, we provided three SoF tables: one for the total trial population (all risk groups combined); one for the favourable risk group (separated by IMDC and MSKCC); and one for the intermediate and poor risk groups (separated by IMDC and MSKCC).

As SUN was our main comparator in this review, it was also chosen as the main comparator in all SoF tables. Moreover, for all SoF tables, we chose the clinically most relevant interventions that are currently recommended across all risk groups in four clinical practice guidelines (ESMO, EAU, NCCN and the German guideline; see Description of the intervention). This resulted in seven (combinations of) substances that we chose for the SoF tables: PEM+AXI, AVE+AXI, NIV+CAB, LEN+PEM, NIV+IPI, CAB alone, PAZ alone.

Certainty of the evidence

One review author (AA) independently rated the certainty of the evidence for each outcome. Another review author (VP) independently checked the assessments and then the two authors met to discuss and finalise the assessments. We used the GRADE approach to rank the certainty of the evidence and the guidelines provided in Chapter 15 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Schünemann 2019). More precisely, we used the GRADE network meta-analysis approach by Salanti 2014 to assess the certainty of the evidence and included our final judgements in the SoF tables.

The GRADE approach to assess the certainty of the body of evidence for each outcome of a network meta-analysis uses five domains: trial limitations (risk of bias of included trials, using the overall 'risk of bias' judgement as derived from the RoB 2 Excel tool), indirectness (relevance to the review question), inconsistency (looking at heterogeneity and incoherence), imprecision (e.g. confidence intervals), and publication bias (Chaimani 2019). GRADE ratings of the evidence are interpreted as follows.

- High: we are very confident that the true effect lies close to that of the effect estimate.
- Moderate: we are moderately confident in the effect estimate: the true effect is likely to be close to the effect estimate, but there is a possibility that it is substantially different.
- Low: our confidence in the effect estimate is limited: the true effect may be substantially different from the effect estimate.
- Very low: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the effect estimate.

The GRADE system uses the following criteria for assigning a rating to a body of evidence (Schünemann 2019).

- High: randomised trials or double-upgraded observational trials.
- Moderate: downgraded randomised trials or upgraded observational trials.
- Low: double-downgraded randomised trials or observational trials.
- Very low: triple-downgraded randomised trials, downgraded observational trials, or case series/case reports.

We downgraded the certainty of the evidence as follows.

- Serious (−1) or very serious (−2) limitation to trial quality.
- Important inconsistency (−1).
- Some (−1) or major (−2) uncertainty about directness.
- Imprecise or sparse data (−1) or very imprecise (i.e. very wide confidence interval) (-2).
- High probability of reporting bias (−1).

In the protocol for this review we stated that we will use GRADEpro GDT for the GRADE assessment; however, this was not feasible (see Differences between protocol and review).

Living systematic review considerations

At protocol stage, we proposed an approach for updating this review (see Differences between protocol and review). However, due to restricted funding an update of the review is currently not planned.

RESULTS

Description of studies

Results of the search

The overall numbers of trials screened, included and excluded, are documented in a PRISMA flow diagram (Figure 1).

We identified a total of 13,689 records through our living systematic review approach. We conducted baseline searches in February and in October 2020. Between December 2020 and April 2021, we conducted monthly database searches according to the living systematic review approach. Baseline searches and update searches in the trial registries were also conducted in February 2020, in October 2020 and in April 2021. We conducted one final update search in databases and trial registries in February 2022.

After removal of 2681 duplicates, we screened a total of 11,017 records. At title and abstract screening, we regarded 10,189 records as irrelevant and excluded these. As a result, we screened 828 full-text articles. At full-text screening, we identified another 204 duplicates, and 37 references that were already found in previous update searches, so we excluded these as well. Another 221 references were excluded at full text screening with reasons (see Characteristics of excluded studies). Hence, we included 366 relevant references. Thereof, five references (trials) are still awaiting classification (see Characteristics of studies awaiting classification). Ultimately, we included 36 trials (in 323 references) (see Characteristics of included studies) and 19 ongoing trials (in 38 references) (see Ongoing studies).

Clinical study reports

We identified clinical study reports for two trials (NCT00720941; NCT00334282) and one scientific result summary for one trial (NCT01064310) through the CSDR platform. We used these documents as our primary sources to extract relevant data and to inform our risk of bias assessments.

Included studies

We included a total of 55 trials that met our pre-specified inclusion criteria. Of these, 36 trials were included in quantitative analyses and narrative reporting (see Characteristics of included studies); 19 trials were classified as ongoing (see Characteristics of ongoing studies).

In these 36 included trials, a total of 15,177 participants (11,061 males; 4116 females) from 53 countries were included. Thirty-three trials were multi-centre trials. The median age of participants ranged from 55 to 68 years.

Design

All 36 trials were RCTs, out of which 32 were two-arm trials; the remaining four trials were three-arm trials (NCT00065468; NCT01984242; NCT02811861; NCT00619268). Six trials were cross-over RCTs (NCT00732914; NCT00903175; NCT01392183; NCT01481870; NCT01613846; NCT00117637) and we extracted data from the first period (before cross over) whenever possible. Most trials were open-label (non-blinded), except for one trial (NCT00081614), which was double-blinded (participants and investigators) and two trials (NCT00334282; NCT01064310), which were blinded quadruple (participants, care providers, investigators and outcome assessors). In one trial, blinding was not reported (Jonasch 2010). Three studies were placebo-controlled (NCT00334282; NCT00081614; NCT00738530).

Sample size

The smallest trial had a sample size of N = 22 and the largest trial had a sample size of N = 1110.

Locations

Most trials were multi-centre trials (33 multi-centre trials, three single-centre trials) and included participants from Europe, North- and South America, Asia, Australia, Africa and the Pacific region. 28 trials included participants from European Countries (NCT00065468; NCT00098657/NCT00083889; NCT00117637; NCT00334282; NCT00420888; NCT00609401; NCT00619268; NCT00631371; NCT00719264; NCT00720941; NCT00732914; NCT00738530; NCT00903175; NCT00920816; NCT00979966; NCT01024920; NCT01030783; NCT01064310;

NCT01108445; NCT01274273; NCT01613846; NCT01984242; NCT02231749; NCT02420821; NCT02684006; NCT02811861; NCT02853331; NCT03141177), 24 trials from northern America (Jonasch 2010; NCT00065468; NCT00072046; NCT00081614; NCT00098657/NCT00083889; NCT00117637; NCT00126594; NCT00631371; NCT00719264; NCT00720941; NCT00903175; NCT00920816; NCT01030783; NCT01108445; NCT01392183; NCT01835158; NCT01984242; NCT02231749; NCT02420821; NCT02684006; NCT02761057; NCT02811861; NCT02853331; NCT03141177), 15 trials from Asia (NCT00334282; NCT00631371; NCT00719264; NCT00720941; NCT00738530; NCT00903175; NCT00920816; NCT01030783; NCT01481870; NCT02231749; NCT02420821; NCT02684006; NCT02811861; NCT02853331; NCT03141177), 13 trials from southern America (NCT00065468; NCT00098657/NCT00083889; NCT00334282; NCT00631371; NCT00719264; NCT00903175; NCT00920816; NCT01030783; NCT02231749; NCT02420821; NCT02684006; NCT02853331; NCT03141177), 12 trials from Australia (NCT00065468; NCT00098657/NCT00083889; NCT00334282; NCT00631371; NCT00720941; NCT00738530; NCT00903175; NCT02231749; NCT02420821; NCT02684006; NCT02811861; NCT03141177), five trials from Africa (NCT00065468; NCT00334282; NCT00631371; NCT00719264; NCT00920816), and three trials from the Pacific region (NCT00065468; NCT00334282; NCT02684006).

Participants

All trials included participants with advanced renal cell carcinoma (RCC), and all participants were above the age of 18 years. All trials explored first-line treatment and most included treatment-naive participants (i.e. participants who have not received prior systemic anticancer treatment). Both treatment-naive and previously treated participants were included in seven trials. However, for two trials, where more than 10% of participants were previously treated, we were able to extract data for the treatment-naive participants only for some outcomes (NCT00334282; NCT01030783). For four trials, separate data for the treatment-naive participants was not extractable. However, in one trial, 7% of the trial population received previous systemic therapy (NCT02761057); in another trial, 4% of the trial population previously received therapy (NCT01392183); and in two trials, 3% of the trial population received prior therapy (NCT00732914; NCT00420888). This is less than our pre-defined threshold of 10%, meaning we included these trials in our analyses. The results of one trial were not included in any analyses because, although it was stated in the methods that participants who had not received prior systemic therapy were eligible, 90% of the trial population that was ultimately included have had some prior anticancer therapy, but without further details about what this therapy consisted of (NCT01064310).

Risk groups

In most trials, the total trial population included all risk groups (i.e. favourable, intermediate or poor risk groups), according to either IMDC or MSKCC criteria. In three trials, the total trial population included only intermediate and poor risk groups (NCT01392183; NCT00065468; NCT01835158) and in four trials, the total trial population included only favourable and intermediate risk groups (NCT00081614; NCT00420888; NCT01481870; NCT01064310).

Histology type

In 18 trials, only participants with clear cell renal cell carcinoma were included (NCT01030783; Jonasch 2010; NCT00072046; NCT00098657/NCT00083889; NCT00117637; NCT00126594; NCT00720941; NCT00903175; NCT00920816; NCT01024920; NCT01030783; NCT01274273; NCT01392183; NCT01481870; NCT02231749; NCT02684006; NCT02853331; NCT03141177). In 10 trials, most participants (80% to 90%) had clear cell carcinoma (NCT00065468; NCT00334282; NCT00609401; NCT00619268; NCT00631371; NCT01064310; NCT01984242; NCT00609401; NCT00719264; NCT01613846). In two trials, at least 50% of participants had clear cell carcinoma (NCT00081614; NCT00738530). Another two trials mostly included participants with clear cell carcinoma, of which some had sarcomatoid features (NCT02811861; NCT02420821). One trial included only participants with non-clear cell carcinoma (NCT01108445); another trial included both participants with clear cell or non-clear (papillary) carcinoma (NCT00420888), and in another trial, 79% of participants had non-clear cell (papillary) carcinoma (NCT00979966). One trial included several subtypes (NCT02761057).

Sites of metastases

In 26 trials, more than one metastatic site was reported in each trial, including the lung, lymph nodes, bones, liver and/or kidney (Jonasch 2010; NCT00072046; NCT00098657/NCT00083889; NCT00117637; NCT00334282; NCT00609401; NCT00619268; NCT00719264; NCT00720941; NCT00732914; NCT00738530; NCT00903175; NCT00920816; NCT01024920; NCT01030783; NCT01108445; NCT01274273; NCT01392183; NCT01481870; NCT01613846; NCT02231749; NCT02420821; NCT02761057; NCT02811861; NCT02853331; NCT03141177). In two trials, only participants with bone or brain metastases (NCT01835158) or metastases in the central nervous system (NCT00631371) were included. In the remaining eight trials, the sites of the metastases were not reported. However, these trials reported the following: 80% of participants had ≥ 2 metastatic sites (NCT00065468); participants with metastatic RCC were eligible and had one or two tumour sites (NCT02684006); only participants with metastatic disease were eligible (NCT00081614; NCT00126594); 86.5% had metastatic disease and 13.5% a locally advanced stage (NCT00979966); 73% had two metastatic sites, 26% had 1 or none (NCT01064310); participants with metastatic or unresectable locally advanced RCC were eligible (NCT00420888; NCT01984242).

Nephrectomy

In all trials but one (NCT00979966, information not provided) it was either reported that participants had previously received a nephrectomy (either full or partial), or prior nephrectomy was generally expected by the inclusion criteria of the trials, but without further information about how many participants actually have had a prior nephrectomy.

Radiotherapy

In eight trials, participants in both arms had received prior radiotherapy (NCT00117637; NCT00631371; NCT00720941; NCT00732914; NCT01064310; NCT02231749; NCT02853331; NCT03141177). In 11 trials, prior radiotherapy was generally allowed (based on the inclusion criteria of the trials), but had to be completed at least two, three or four weeks (depending

on the trial) prior to initiation of the first cycle of systemic treatment in the trial (NCT00065468; NCT00081614; NCT00126594; NCT00903175; NCT00920816; NCT00979966; NCT01024920; NCT01613846; NCT01984242; NCT02420821; NCT02811861). In the remaining 17 trials, no information about prior radiotherapy was provided.

Interventions

We identified 22 drugs and 17 different combinations in the included trials. In 16 trials, the substance sunitinib (main comparator in this review) was assessed in the comparator arm; in three trials, it was assessed in the experimental arm. In all 19 trials, sunitinib was administered via the same administration route (oral) and the same dose (50 mg/day). Discontinuation rates of sunitinib were high in most trials: in two trials, less than 20% of participants discontinued treatment; in four trials, 50% to 80% of participants discontinued treatment; and in the remaining 13 trials, between 80% to 100% of participants discontinued sunitinib treatment.

As for the other interventions, all were administered via the same administration route and most were also administered at the same dose (see Table 1). For more details per trial, see Characteristics of included studies.

Duration of therapy

In most trials, therapy was provided as continuous therapy, meaning therapy was continued as long as there was no disease progression, unacceptable toxicity (intolerable adverse events), clinical deterioration, loss of clinical benefit or withdrawal of consent. In three trials, participants were allowed to continue therapy despite disease progression if evidence of clinical benefit was observed according to the trial investigators (NCT02420821; NCT02684006; NCT00920816). In seven trials, therapy (either all or certain drugs) was provided for a fixed period: for 24 months (NCT00081614); for 18 months (NCT00420888); in one cross-over study, period 1 lasted for 10 weeks (NCT01064310); bevacizumab was administered for a maximum of one year (NCT01274273); pembrolizumab was administered for a maximum of 35 cycles in two trials (NCT02853331; NCT02811861); nivolumab was administered for a maximum of two years (NCT03141177). In five trials, specific information about the treatment duration was not provided (NCT00126594; NCT00719264; NCT00979966; NCT01613846; NCT01984242).

Table 1. Interventions in the included trials

Drug substance	Administration route	Dose	Combinations with other drugs in the included trials
Atezolizumab (ATE)	intravenous infusion	1200 mg	ATE+BEV
Avelumab (AVE)	intravenous infusion	10 mg	AVE+AXI PEM+AXI
Axitinib (AXI)	oral administration	5 mg	AVE+AXI
Bevacizumab (BEV)	intravenous infusion	10 mg	BEV+ERL IFN+BEV TEM+BEV ATE+BEV EVE+BEV
Cabozantinib (CAB)	oral administration	60 mg	NIV+CAB
Crizotinib (CRI)	oral administration	60 mg	-
Erlotinib (ERL)	oral administration	150 mg	BEV+ERL
Everolimus (EVE)	oral administration	5 mg or 10 mg	EVE+BEV LEN+EVE
Interferon-alpha (IFN)	subcutaneous injection	0.5 MIU; or 3 MIU; or 6 MIU; or 9 MIU	IFN+BEV SOR+IFN NAP+IFN

			IFN+TEM ILN+IFN ILN+IFN+BEV
Interleukin (ILN)	subcutaneous injection	2.4 MIU	SOR+ILN ILN+IFN
Ipilimumab (IPI)	intravenous infusion	1mg	NIV+IPI
Lenvatinib (LEN)	oral administration	18 mg; or 20 mg	LEN+PEM LEN+EVE
Naptumomab (NAP)	intravenous infusion	15 mg	NAP+IFN
Nintedanib (NIN)	oral administration	200 mg	-
Nivolumab (NIV)	intravenous infusion	3 mg or 240 mg	NIV+IPI NIV+CAB
Pazopanib (PAZ)	oral administration	800 mg	-
Pembrolizumab (PEM)	intravenous infusion	200 mg	PEM+AXI LEN+PEM
Savolitinib (SAV)	oral administration	600 mg	-
Sorafenib (SOR)	oral administration	400 mg	SOR+IFN SOR+ILN
Sunitinib (SUN) (main comparator in this review**)**	oral administration	50 mg	-
Temsirolimus (TEM)	intravenous infusion	15 mg or 25 mg	IFN+TEM TEM+BEV
Tivozanib (TIV)	oral administration	1.5 mg	-

Outcome Measures

Primary outcomes

Overall survival

Overall survival (OS) was reported in 32 out of 36 trials included in this review (Jonasch 2010; NCT00065468; NCT00072046; NCT00081614; NCT00098657/NCT00083889; NCT00334282; NCT00420888; NCT00609401; NCT00619268; NCT00631371; NCT00719264; NCT00720941; NCT00732914; NCT00738530; NCT00903175; NCT00920816; NCT00979966; NCT01024920; NCT01030783; NCT01108445; NCT01392183; NCT01481870; NCT01613846; NCT01835158; NCT01984242; NCT02231749; NCT02420821; NCT02684006; NCT02761057; NCT02811861; NCT02853331; NCT03141177). Trials reported outcome data on OS for the total trial population and/or for individual risk groups according to IMDC and/or MSKCC. Whenever possible, we extracted data for the individual risk groups.

Quality of life

As stated in the Methods, we prioritised scales for the assessment of this quality of life (QoL). The final prioritisation included the following scales: FKSI-DRS; EORTC-QLQ-C30; EQ-VAS; FACT-G; FACIT-F.

Quality of life was assessed in a total of 22 trials, out of which 15 trials assessed this outcome using at least one of our prioritised scales. However, we could only extract or estimate data from

seven trials (NCT00098657/NCT00083889; NCT00720941; NCT00920816; NCT01108445; NCT02231749; NCT03141177; NCT00903175). For the remaining eight trials, extracting the results was not possible for the following reasons.

- In one trial, QoL should have been assessed with FACT-G, but we could not find results anywhere (NCT01392183).
- In another trial, QoL should have been assessed with FACIT-F, but we could not find results anywhere (NCT01613846).
- Separate data for treatment-naive participants were not reported in two trials (NCT00334282; NCT01030783).
- In one trial, 'time to definitive deterioration' analyses were reported, which were not a focus of this review (NCT00719264).
- In another trial, 'time to first deterioration' analyses were reported, which were not a focus of this review (NCT02811861).
- For one trial, we did not extract data because of a discrepancy between methods and reported results (NCT01064310).
- For one trial, it was not possible to estimate data from the provided graphs (NCT00631371).

Where data extraction was possible, we extracted from a variety of different sources, including the full-text publications (NCT01108445; NCT02231749; NCT03141177; NCT00903175); the trial registry (clinicaltrials.gov) (NCT00920816); both the full-text publication and trial registry (NCT00098657/NCT00083889); the clinical study report (NCT00720941) and the scientific result summary (NCT01064310). For two of these trials, we tried to estimate data for the DRS-scale from the graphs (NCT03141177; NCT00903175). In NCT00903175, the EORTC-scale was also reported, but data could not be estimated. In NCT03141177, only TTD results for EQ-5D-VAS were reported.

We extracted (or estimated) data for the following scales.

- FKSI-DRS in five trials (NCT00098657/NCT00083889; NCT00920816; NCT01108445; NCT03141177; NCT00903175).
- EQ-5D-VAS in three trials (NCT00098657/NCT00083889; NCT00920816; NCT02231749).
- FACT-G in two trials NCT00098657/NCT00083889; NCT02231749).
- FACIT-F in one trial (NCT00720941).

It was not possible to extract or estimate data for EORTC-QLQ-C30 from any trial. Furthermore, QoL was assessed only in the total trial populations (all risk groups combined), meaning the outcome was not assessed in the individual risk groups separately.

Serious adverse events

Serious adverse events (SAEs) were reported in 27 out of 36 included trials (NCT01108445; NCT00738530; NCT00631371; NCT01835158; NCT02231749; NCT00720941;NCT00719264; NCT00903175;NCT01984242; NCT02420821; NCT00117637;NCT00619268; NCT00732914; NCT01613846; NCT00979966; NCT02853331; NCT02811861; NCT00065468; NCT00098657/NCT00083889; NCT00920816; NCT01024920; NCT00126594; NCT00334282; NCT01064310; NCT01030783; NCT02761057; NCT01392183). Serious adverse events were reported for the total population only; meaning they were not reported for the individual risk groups separately.

Evaluable data for SAEs was available for only 22 trials (NCT01108445; NCT00738530; NCT00631371; NCT01835158; NCT02231749; NCT00720941; NCT00719264; NCT00903175; NCT01984242; NCT02420821; NCT00117637; NCT00619268; NCT00732914; NCT01613846; NCT00979966; NCT02853331; NCT02811861; NCT00065468; NCT00098657/NCT00083889; NCT00920816; NCT01024920; NCT00126594). The remaining five trials were not evaluable due to the following reasons: only treatment-related SAEs were reported in one trial (NCT02761057); SAEs were not extractable for treatment-naive participants in two trials that included more than 10% of previously treated participants (NCT00334282; NCT01030783); data for SAEs that occurred during the first treatment period was not extractable for one cross-over trial (NCT01392183). Results of one trial were not presented because we were unsure whether participants were treatment-naive (discrepancy between methods and results in the trial) (NCT01064310). Nine trials did not report SAEs (NCT03141177; NCT00420888; NCT01481870; NCT00081614; Jonasch 2010; NCT00072046; NCT00609401; NCT01274273; NCT02684006).

If available, data for SAEs were preferably extracted from the trial registries (clinicaltrials.gov; clinicaltrialsregister.eu), where we extracted the number of participants with at least one SAE. Furthermore, we assumed that this was the most current data. It was not explicitly stated whether all-cause or treatment-related SAEs were reported, but we strongly assumed that all-cause SAEs were reported on the trial registries. This applied to 18 trials (NCT01108445; NCT00631371; NCT00738530; NCT01835158; NCT02231749; NCT00720941; NCT00065468; NCT00098657/NCT00083889; NCT00903175; NCT00719264; NCT00117637;NCT01984242; NCT00920816; NCT01024920; NCT02853331; NCT02811861; NCT00126594; NCT00979966). Only for four trials, data for all-cause SAEs in the number of participants with at least one SAE were extracted from the respective publications (NCT00619268; NCT01613846; NCT00732914; NCT02420821).

Secondary outcomes

Progression-free survival

Progression-free survival (PFS) was reported in 34 out of 36 trials included in this review (Jonasch 2010; NCT00065468; NCT00072046; NCT00081614; NCT00098657/NCT00083889; NCT00117637; NCT00334282; NCT00420888; NCT00609401; NCT00619268; NCT00631371; NCT00719264; NCT00720941; NCT00732914; NCT00738530; NCT00903175; NCT00920816; NCT00979966; NCT01024920; NCT01030783; NCT01108445; NCT01274273; NCT01392183; NCT01481870; NCT01613846; NCT01835158; NCT01984242; NCT02231749; NCT02420821; NCT02684006; NCT02761057; NCT02811861; NCT02853331; NCT03141177). Trials reported outcome data on PFS for the total trial population and/or for individual risk groups according to IMDC and/or MSKCC. Whenever possible, we extracted data for the individual risk groups.

Adverse events

Adverse events (AEs) were reported in all included trials (N = 36). However, evaluable data for all-cause grade 3 or 4 AEs was available for only 18 trials (NCT00065468; NCT00081614; NCT00719264; NCT00720941; NCT00732914; NCT01024920;

NCT01613846; NCT01835158; NCT01984242; NCT02420821; NCT02684006; NCT02811861; NCT03141177; NCT00738530; NCT00920816; NCT01030783; NCT01108445; NCT01274273). One of these trials did not report individual AEs, meaning we could only extract data for the total number of participants with at least one grade 3 or 4 AE for this trial (NCT01984242). Another five of these trials did not report the total number of participants with at least one grade 3 or 4 AE, meaning we could only extract data for individual grade 3 or 4 AEs. (NCT00738530; NCT00920816; NCT01030783; NCT01108445; NCT01274273). Moreover, AEs were reported for the total population only; meaning they were not reported for the individual risk groups separately.

Eleven of the 18 trials reported AEs of "grade 3 or 4" (NCT01835158; NCT00732914; NCT01613846; NCT01984242; NCT02420821; NCT00719264; NCT00081614; NCT00720941; NCT00065468; NCT01274273; NCT01108445). The remaining seven trials reported AEs of "grade 3 or higher" and we assumed that grade 5 was not included as grade 5 AEs should be regarded as serious adverse events (NCT03141177; NCT02684006; NCT02811861; NCT01024920; NCT01030783; NCT00920816; NCT00738530). Lastly, in two of these trials, it was not explicitly stated whether all-cause or treatment-related AEs were reported, but we assumed all-cause (NCT00065468; NCT00081614).

For the remaining 18 trials, this outcome was not evaluated in this review for the following reasons: only treatment-related AEs were reported in 10 trials (NCT00072046; NCT00098657/NCT00083889; NCT00117637; NCT00126594; NCT00420888; NCT00979966; NCT02231749; NCT02853331; NCT02761057; NCT00609401); the event rate of AEs was reported in three trials (NCT01392183; NCT00903175; NCT00631371); in one trial, it was unclear whether the event rate or the number of participants with one event was reported (Jonasch 2010); AEs were not extractable for treatment-naive participants in one trial that included more than 10% of previously treated participants (NCT00334282); only all-grade AEs were reported in one trial (NCT00619268); for one cross-over trial, it was unclear which treatment period where the AEs occurred was reported (NCT01481870). Lastly, results of one trial were not presented because we were unsure whether participants were treatment-naive (discrepancy between methods and results in the trial) (NCT01064310).

Reporting of individual grade 3 or 4 AEs was common. All individual AEs that were of special interest in this review were reported: **hand-food syndrome** was reported in 10 trials (NCT00720941; NCT00920816; NCT01024920; NCT01030783; NCT01108445; NCT01613846; NCT01835158; NCT02684006; NCT02811861; NCT03141177); **fatigue** in 14 trials (NCT00719264; NCT00720941; NCT00732914; NCT00738530; NCT00920816; NCT01024920; NCT01030783; NCT01108445; NCT01274273; NCT01613846; NCT01835158; NCT02684006; NCT02811861; NCT03141177); **diarrhoea** in 16 trials (NCT00081614; NCT00719264; NCT00720941; NCT00732914; NCT00738530; NCT00920816; NCT01024920; NCT01030783; NCT01108445; NCT01274273; NCT01613846; NCT01835158; NCT02684006; NCT02811861; NCT03141177; NCT00065468); **vomiting** in10 trials (NCT00720941; NCT00920816; NCT01024920; NCT01108445; NCT01274273; NCT01835158; NCT02684006; NCT02811861; NCT03141177; NCT00065468); **loss of appetite** in 11 trials (NCT00719264; NCT00720941; NCT00732914; NCT00920816; NCT01024920; NCT01108445; NCT01613846; NCT01835158; NCT02684006; NCT02811861; NCT03141177); **weight loss** in 12 trials (NCT00719264; NCT00720941; NCT00920816; NCT01024920; NCT01108445; NCT01274273; NCT01613846; NCT01835158; NCT02684006; NCT02811861; NCT03141177; NCT00065468); **insomnia** in two trials (NCT00720941; NCT01108445); **depression** in two trials (NCT00738530; NCT01274273). For **mucous membrane damage,** the following were reported: **mucosal inflammation** in four trials (NCT00720941; NCT01108445; NCT02684006; NCT03141177) and **stomatitis** in t12 trials (NCT00719264; NCT00720941; NCT00732914; NCT01024920; NCT01108445; NCT01274273; NCT01613846; NCT01835158; NCT02684006; NCT02811861; NCT03141177; NCT00065468).

Number of participants who discontinued study treatment due to an Adverse event

This outcome was reported in 34 out of 36 included trials. Results were reported for the total population only; meaning they were not reported for the individual risk groups separately.

However, data for this outcome were evaluable for only 30 trials (NCT01108445; NCT00732914; NCT01613846; NCT00920816; NCT01024920; NCT00979966; NCT00719264; NCT00903175; NCT01481870; NCT00631371; NCT01835158; NCT02231749; NCT00720941; NCT01984242; NCT00081614; NCT01030783; NCT00117637; Jonasch 2010; NCT00065468; NCT00072046; NCT00098657/NCT00083889; NCT00609401; NCT00619268; NCT01274273; NCT01392183; NCT02420821; NCT02684006; NCT02853331; NCT03141177; NCT02811861). Data from two trials were not evaluable due to the following reasons: discontinuations due to treatment-related AEs were reported in one trial (NCT02761057) and discontinuations due to AEs were not extractable for treatment-naive participants in another trial that included more than 10% of previously treated participants (NCT00334282). Results of one trial were not included in the analysis because we were unsure whether participants were treatment-naive (discrepancy between methods and results in the trial) (NCT01064310). Lastly, one trial did not report the data in a way in which it would evaluable (NCT00738530). Two trials did not report this outcome (NCT00420888; NCT00126594).

Time to initiation of first subsequent therapy

None of the included trials reported this outcome as a time-to-event outcome. However, 19 trials (NCT00081614; NCT00098657/NCT00083889; NCT00609401; NCT00619268; NCT00719264; NCT00732914; NCT00738530; NCT00920816; NCT01024920; NCT01108445; NCT01274273; NCT01835158; NCT01984242; NCT02231749; NCT02420821; NCT02684006; NCT02811861; NCT02853331; NCT03141177) reported the number of participants who received any subsequent anticancer therapy after discontinuing study treatment. As reporting between trials was heterogenous, for example in terms of definition of this outcome and the timing of reporting being unclear, we refrained from pooling data in quantitative analyses and reported the results narratively in tabular form instead.

Description of studies awaiting classification

We included five trials that still await classification. Of these, two trials are still active (but not recruiting) (NCT01217931; NCT03541902). Another two trials are completed but not yet

published (NCT01688973; NCT01829841). For the latter two trials we are awaiting results to see how many participants have received prior therapy, and whether results for treatment-naive may be reported separately, in order to be able to make a decision about inclusion into this review. Lastly, thus far one trial is only published as an abstract and does not yet provide enough information for us to be able to decide whether it is eligible for inclusion into this review (Liu 2017). For more details per trial, see Characteristics of studies awaiting classification.

Description of ongoing studies

We identified 19 ongoing trials that would be eligible for inclusion into this review once results are published. One of these trials is still ongoing (EUCTR2008-000928-71-IT); seven are active (but not recruiting) (NCT02210117; NCT03260894; NCT03729245; NCT03873402; NCT03937219; NCT02996110; NCT04540705); 10 are still recruiting (NCT03075423; NCT03592472; NCT03793166; NCT04090710; NCT04203901; NCT04394975; NCT04523272; NCT04736706; NCT05043090; UMIN 000012522); and one trial is not yet recruiting (NCT05096390). For more details per trial, see Characteristics of ongoing studies.

Excluded studies

After title and abstract screening, we excluded a total of 10,189 records that did not match our inclusion criteria (see Figure 1).

At full-text stage, we excluded 191 trials (in a total of 197 references) after detailed evaluation. The trials were excluded for the following reasons:

- Irrelevant intervention(s), for example interventions included a cancer vaccine, hormone therapy, adjuvant therapy or chemotherapy agents that are not relevant to this review (Aass 2005; Adler 1987; Bex 2017; Demirci 1999; Eisen 2019; EUCTR2008-002667-13-DE 2008; Euctr2015-002133-22-FR; Gruenwald 2020; Haas 2016; NCT02960906; NCT03829111; NCT00467025; Ravaud 2016; Richards 1977).
- Wrong study design, for example single-arm trials, dose-finding trials, cohort trials or non-randomised trials (Abdel 2018; Amin 2018; Amin 2018a; Barrios 2009; Bracarda 2007; Buckley 2019; Cirkel 2016; Cirkel 2017; Climent 2020; Collinson 2018; Colomba 2021; Conter 2013; Epaillard 2020; Euctr 2006-003429-95-ES; Feldman 2020; Feldman 2020a; Gedye 2021; Hutson 2006; Hutson 2021; ISRCTN95351638; Jeon 1999; Larkin 2019; Lee 2020; Lee 2021; McDermott 2020; Minasian 1993; NCT01408004; NCT01444807; NCT02127710; NCT00835978; NCT00100906; NCT03173560; Nosov 2010; Nosov 2012; Plimack 2015; Sternberg 2013; Taylor 2020; Taylor 2020a; Voss 2015).
- Irrelevant comparisons or irrelevant comparator (Atkins 1991; Atkins 1993; Atzpodien 1997; Atzpodien 1997a; Atzpodien 1999; Atzpodien 2001; Atzpodien 2004; Atzpodien 2006; Berg 1998; Boccardo 1998; Cole 2003; Collinson 2012; de Mulder 1991; Dexeus 1988; Dexeus 1989; Dubois 1997; Elhilali 2000; Escudier 2005; Euctr2007-002556-41-AT; Figlin 1998; Figlin 1999; Foon 1988; Fossa 1989; Fossa 1992; Gleave 1997; Gleave 1997a; Gleave 1998; Gore 2008; Gore 2010; Hainsworth 2015; Hainsworth 2016; Han 2002; Harima 1990; Henriksson 1998; Jayson 1998; JPRN-jRCTs031180024; Kinouchi 2004; Kinouchi 2006; Law 1995; Lindskog 2020; Lissoni 1993; Liu 2012; Lummen 1996; Madhusudan 2004; McDermott 2001; McDermott 2005; Mickisch 2001; Motzer 2001; Naglieri 1998; NCT00002737; NCT00005966; NCT00019539; NCT00027664; NCT00053820; NCT00416871; NCT01164228; Negrier 1996; Negrier 1997; Negrier 1998; Negrier 2000; Negrier 2006; Negrier 2007; Negrier 2008; Passalacqua 2010; Pyrhonen 1995; Pyrhonen 1996; Pyrhonen 1999; Rini 2011; Rini 2012; Rpcec 2017; Trump 2004; Verzoni 2018; Witte 1995).
- Trials were terminated (DRKS00010309 2016; Figlin 2014; Figlin 2014a; Figlin 2017; Figlin 2018; Figlin 2020; NCT00491738; NCT01673386; NCT03035630; Rexer 2017; Rodriguez-Vida 2020; Tannir 2016; Wood 2013), ended prematurely (Euctr2006-002851-33-AT; Euctr2006-005751-16-NL; Euctr2012-001730-33-ES; Euctr2018-001495-38-FR; NCT00873236; NCT02014636; NCT00709995) or were withdrawn (NCT01616186).
- Participants have previously received therapy, i.e. trials assessed second-line therapy (Beaumont 2009; Beaumont 2011; Cella 2016; Choueiri 2017; Choueiri 2020; Choueiri 2020a; Flaherty 2015; Gao 2017; Gao 2019; Ghiorghiu 2018; Jager 2005; JPRN-JapicCTI-122014; JPRN-UMIN000001995; McDermott 2013; Molina 2009; Mulders 2012; NCT00378703; NCT00073307; NCT01223027; NCT01664182; NCT01727089; NCT01727336; NCT01793636; NCT02667886; NCT02724020; NCT03092856; NCT03095040; NCT03501381; NCT03595124; NCT03095040; NCT04195750; NCT04300140; Pal 2015; Pal 2021a; Ravaud 2006; Szarek 2021; Thiam 2010; Twardowski 2015; Twardowski 2017; Voss 2019; Wright 2020; Yang 2002; Yang 2003; Zhou 2016; Zhou 2019).

Risk of bias in included studies

Detailed risk of bias assessments (domain judgements and support for judgements) can be found in the appendices (Appendix 9; Appendix 10; Appendix 11; Appendix 12; Appendix 13). Further details (including answers to the signalling questions of the RoB 2 tool) are available in a supplementary file (Aldin 2023).

Overall survival (OS)

Bias assessment of OS (domain judgements and support for judgements) is reported in Appendix 9 and visually presented in Figure 2 (traffic light plot) and Figure 53 in Appendix 14 (summary plot) for the total population; in Figure 3 (traffic light plot) and Figure 54 in Appendix 14 (summary plot) for the Memorial Sloan Kettering Cancer Center (MSKCC) risk groups; and in Figure 4 (traffic light plot) and Figure 55 in Appendix 14 (summary plot) for the International Metastatic RCC Database Consortium (IMDC) risk groups. This outcome was predominantly judged as 'some concerns' mainly due to missing study protocols and statistical analyses plans (SAPs). For the majority of the remaining trials, OS was judged as 'high risk of bias' due to the lack of information about missing outcome data, the randomisation process and allocation concealment. Risk of bias judgement differed between the total population and the risk groups for only one trial: whereas OS for the total population was judged as 'low risk of bias', OS per risk group was judged as 'high risk of bias' because this subgroup analysis was conducted as post-hoc analysis (NCT00720941).

Figure 2. Traffic light plot for OS for all risk groups combined

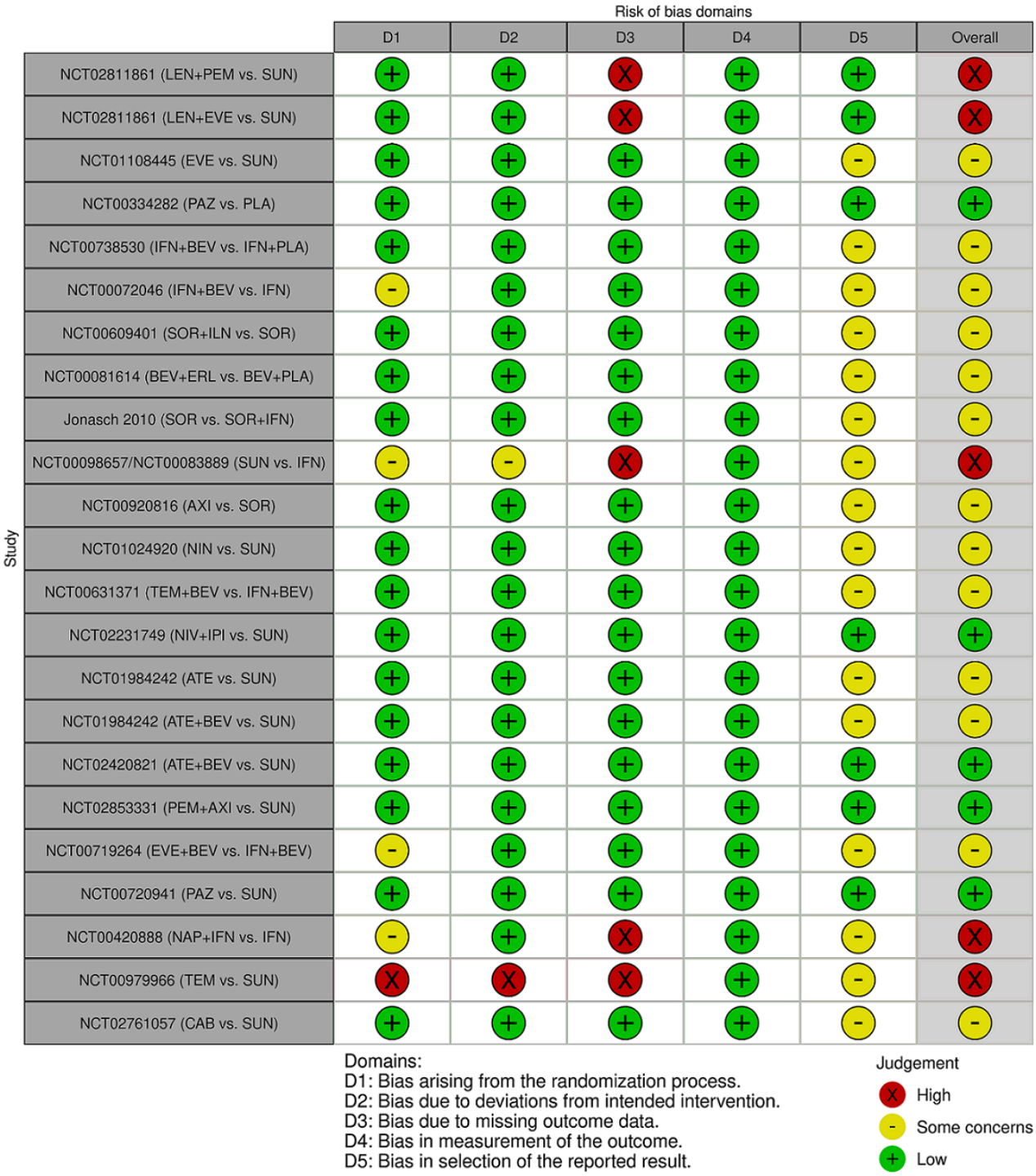


Figure 3. Traffic light plot for OS per MSKCC favourable, intermediate, poor risk

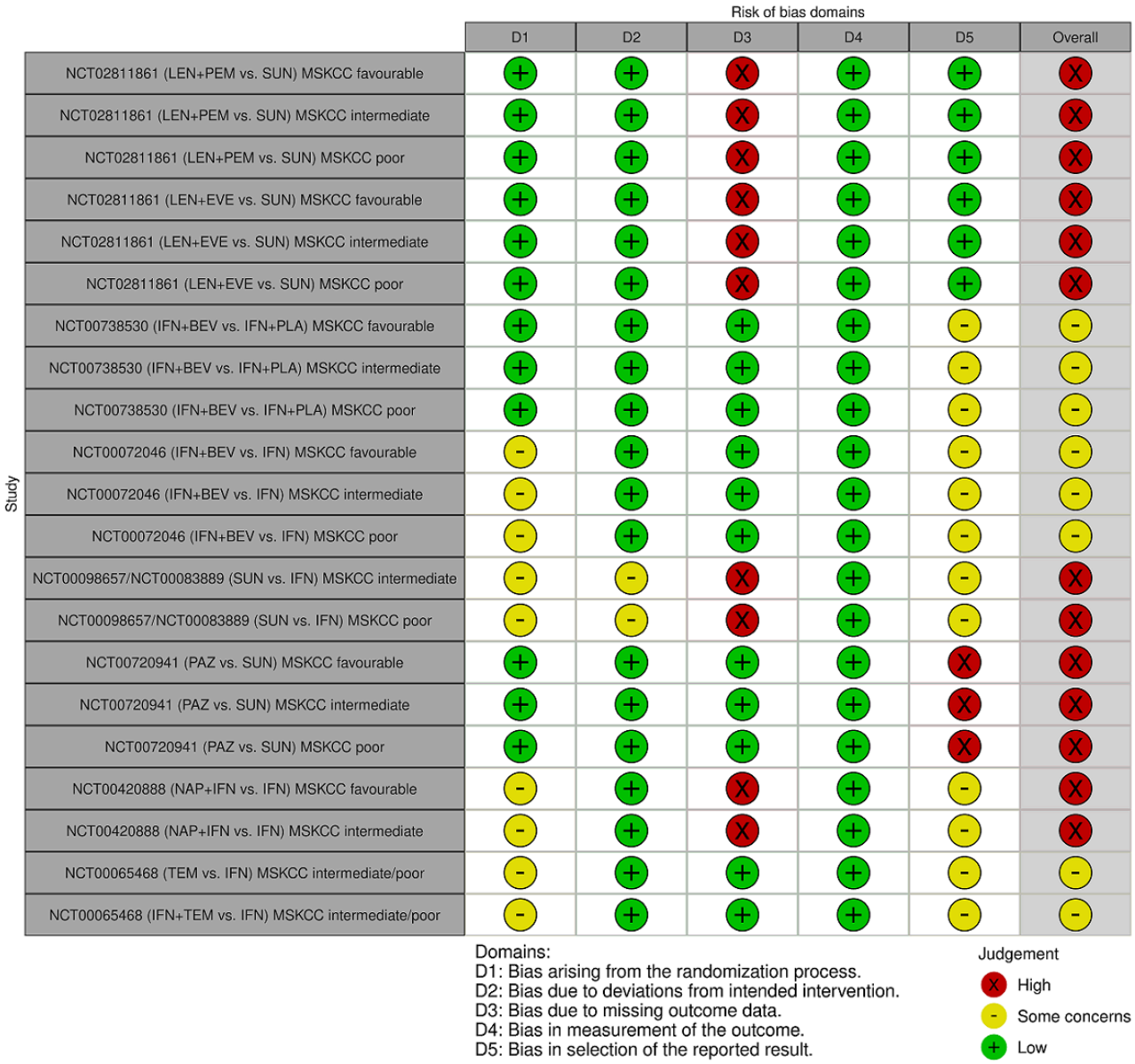
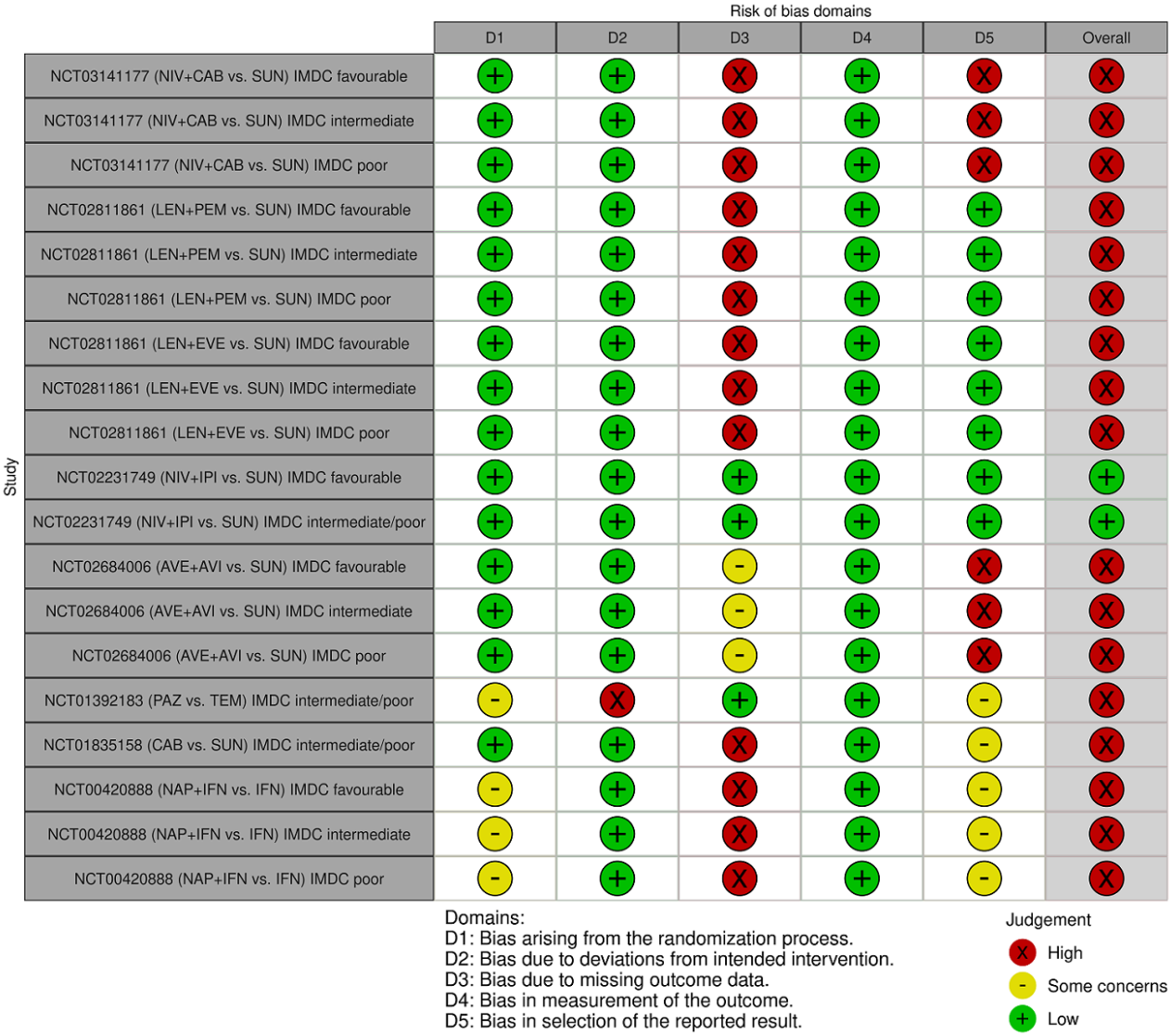


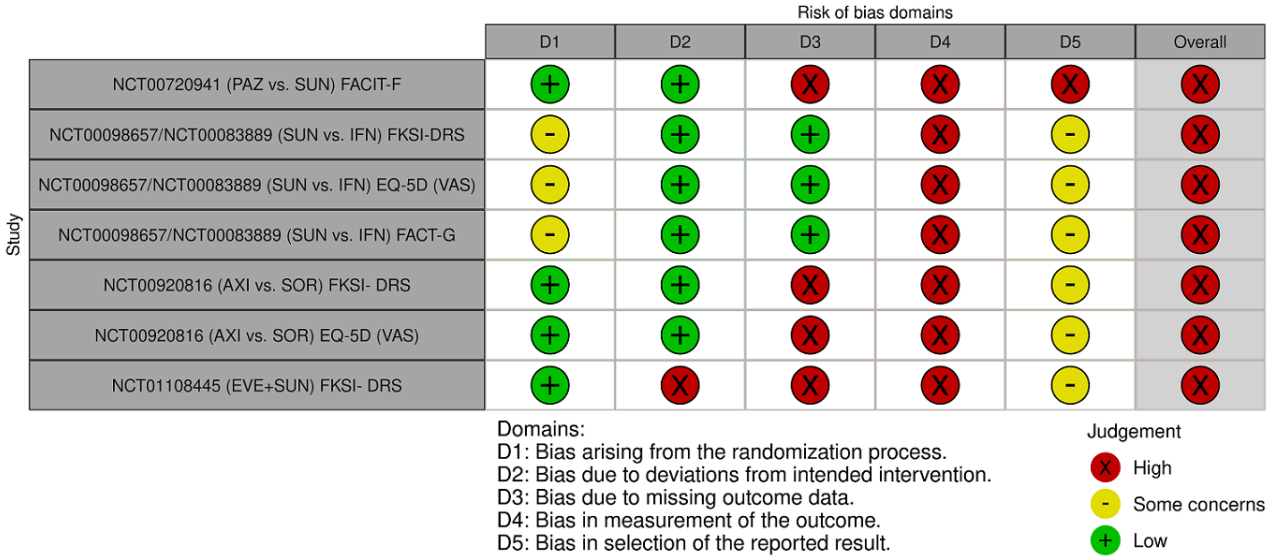
Figure 4. Traffic light plot for OS per IMDC favourable, intermediate, poor risk



Quality of life (QoL)

The outcome QoL is presented in Appendix 10 and visually presented in Figure 5 (traffic light plot) and Figure 56 in Appendix 14 (summary plot). It was also predominantly judged as 'high risk of bias' mainly due to the outcome assessors' awareness of the assigned interventions, which is owed to the nature of self-reported questionnaires and participants (the outcome assessors) not being blinded to the intervention received in open-label (non-blinded) trials, as well as due to the high number of participants without outcome data at the end of treatment (time point for which risk of bias was assessed).

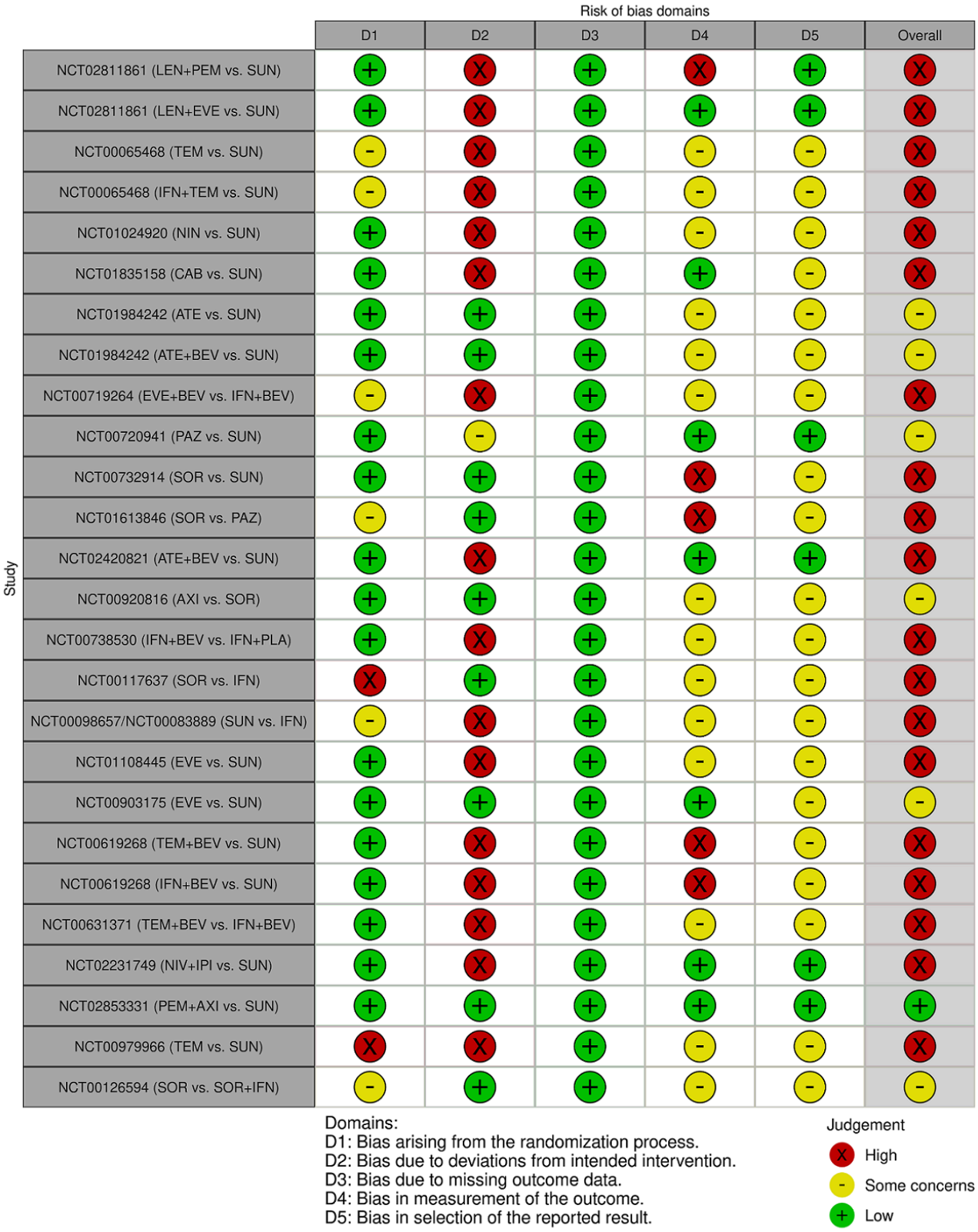
Figure 5. Traffic light plot for QoL for all risk groups combined at the end of treatment



Serious adverse events (SAEs)

The outcome SAEs is reported in Appendix 11 and visually presented in Figure 6 (traffic light plot) and Figure 57 in Appendix 14 (summary plot). This outcome was predominantly judged as 'high risk of bias' mainly due to the lack of information about method of analysis and method of outcome measurement. In few cases, risk of bias was judged as 'high risk' due to the lack of information about the randomisation process and allocation concealment.

Figure 6. Traffic light plot for SAEs for all risk groups combined



Progression-free survival (PFS)

Bias assessment of PFS is reported in Appendix 12 and visually presented in Figure 7 (traffic light plot) and Figure 58 in Appendix 14 (summary plot) for the total population; in Figure 8 (traffic light plot) and Figure 59 in Appendix 14 (summary plot) for the MSKCC risk groups; and in Figure 9 (traffic light plot) and Figure 60 in Appendix 14 (summary plot) for the IMDC risk groups. This outcome was predominantly judged as 'high risk of bias' mainly due to the lack of information about missing outcome data and allocation concealment as well as the outcome assessors' probable or evident awareness of the assigned interventions. There were no differences in the risk of bias judgement between the total population and the risk groups.

Figure 7. Traffic light plot for PFS for all risk groups combined

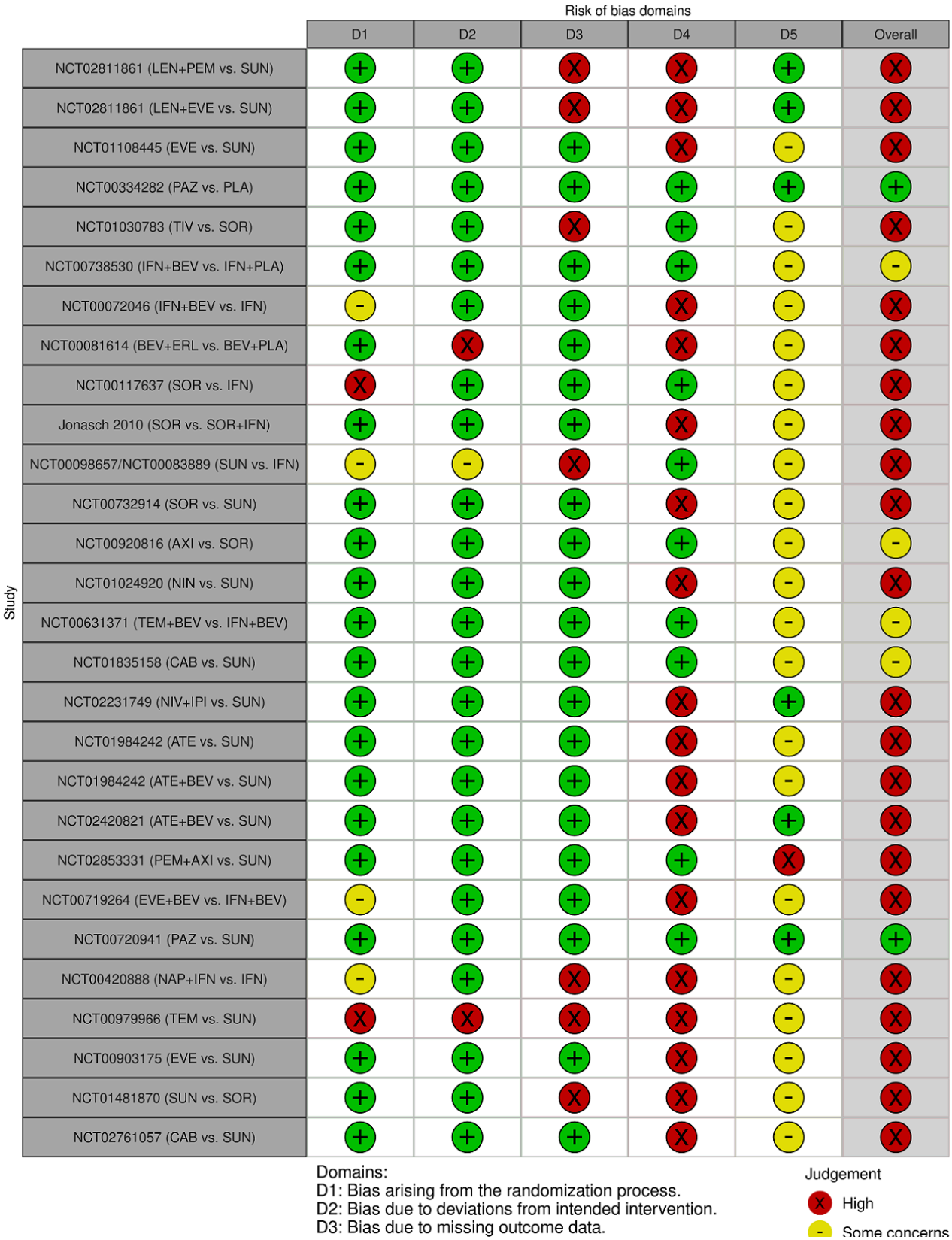


Figure 7. (Continued)

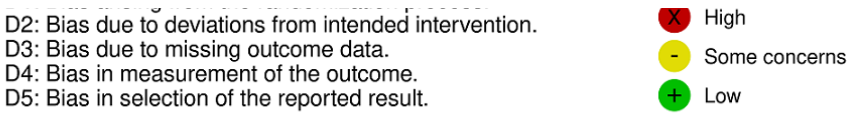


Figure 8. Traffic light plot for PFS per MSKCC favourable, intermediate, poor risk

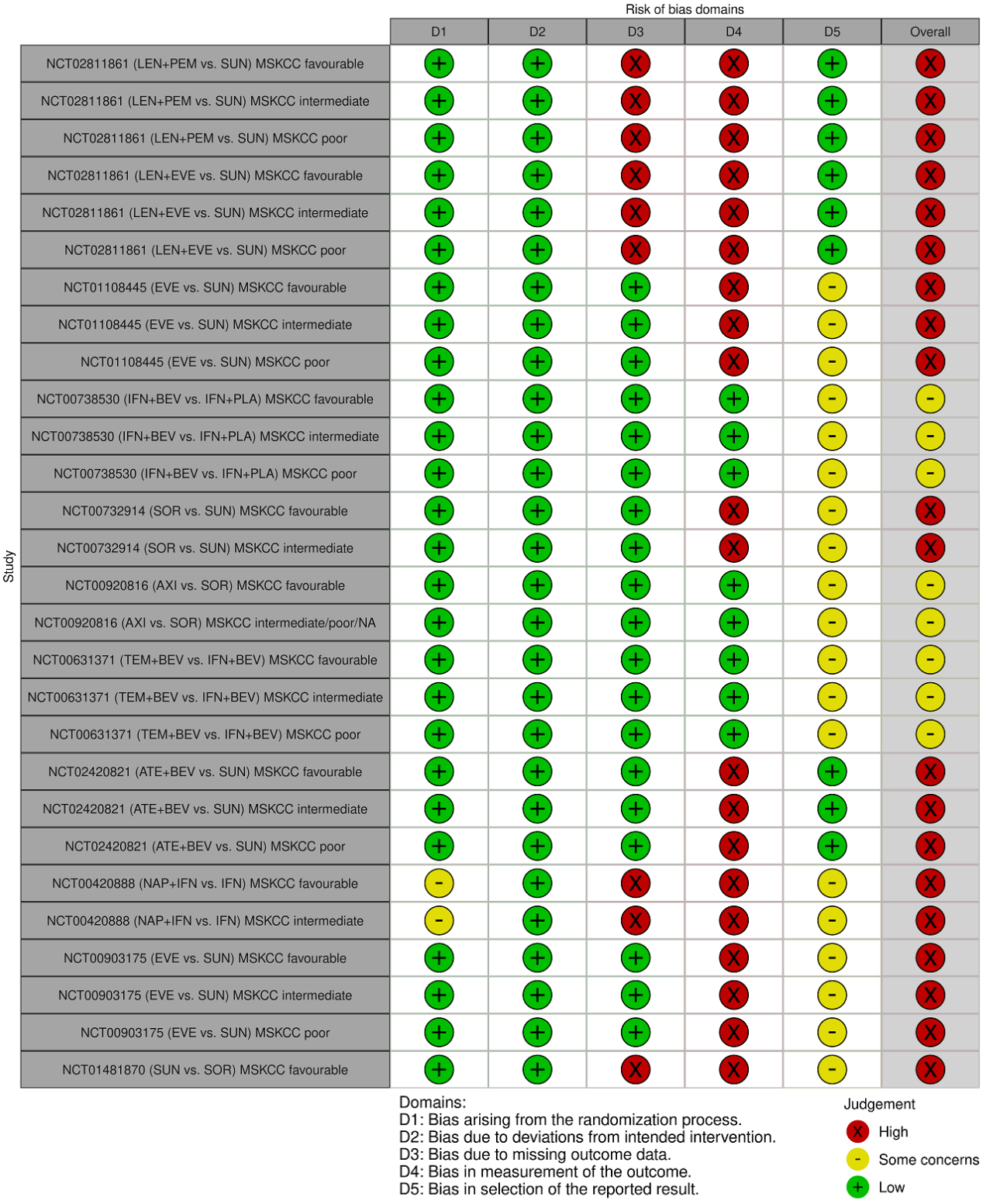
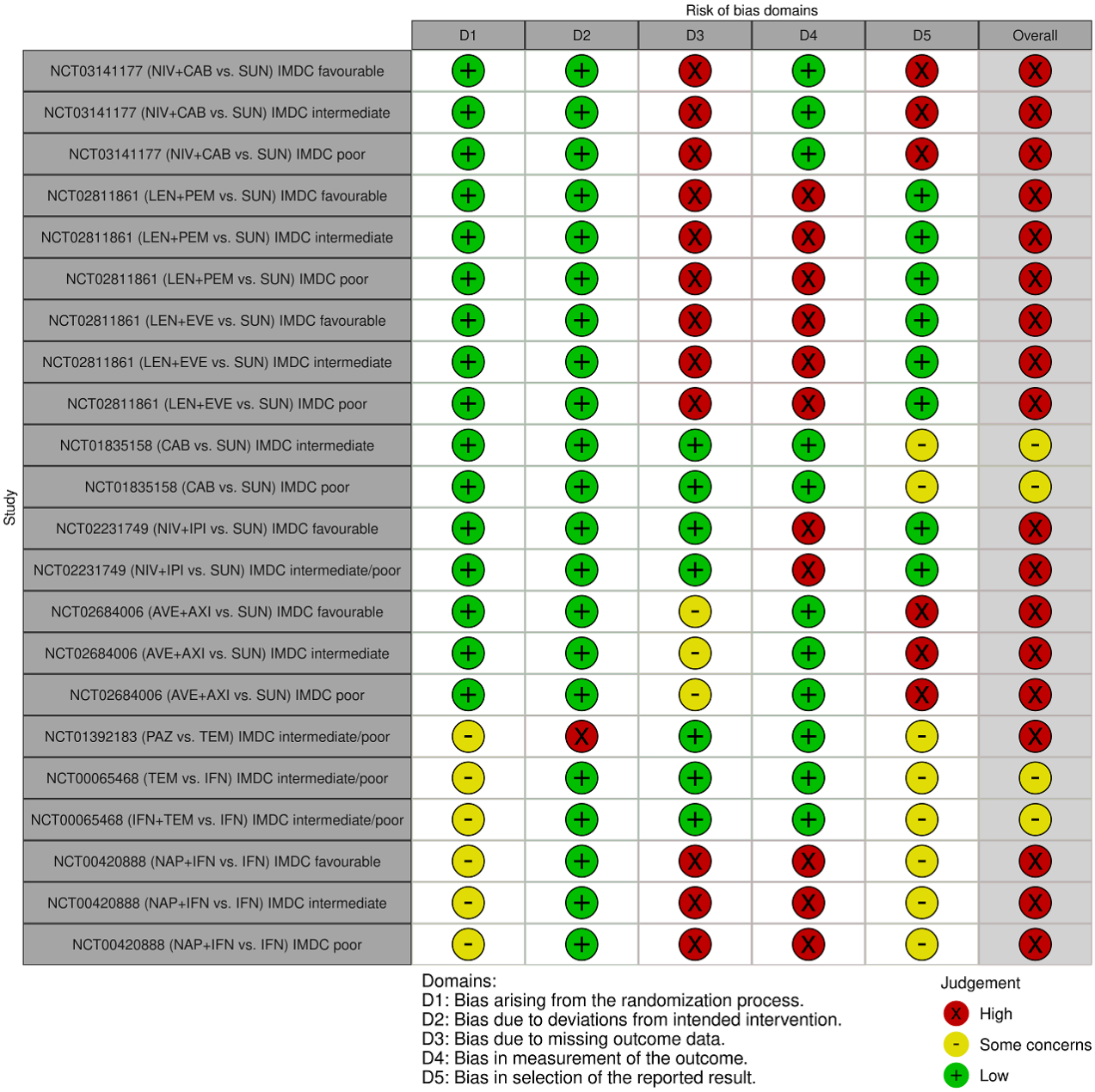


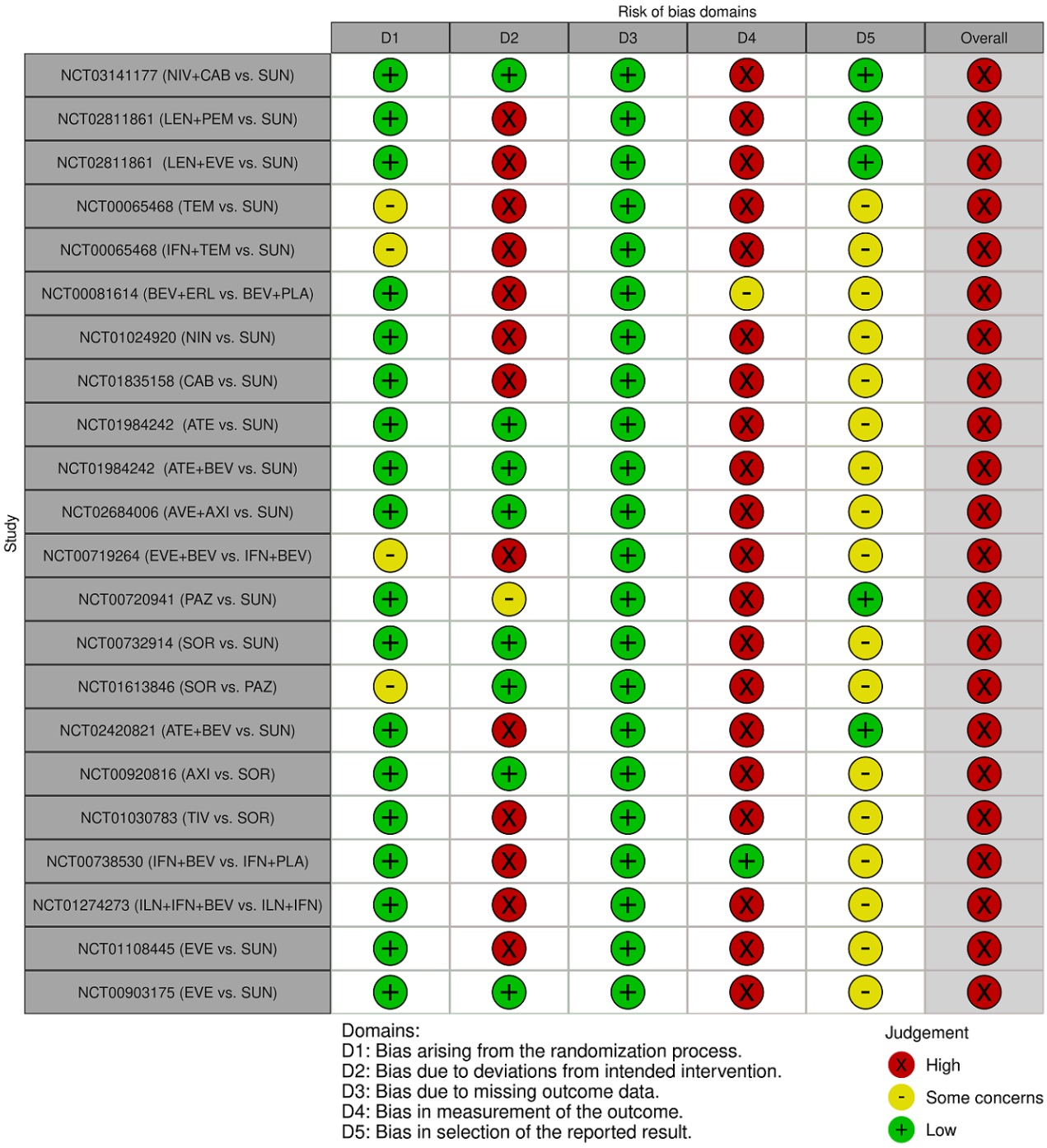
Figure 9. Traffic light plot for PFS per IMDC favourable, intermediate, poor risk



Adverse events (AEs)

Bias assessment of the outcome AEs is reported in Appendix 13 and visually presented in Figure 10 (traffic light plot) and Figure 61 in Appendix 14 (summary plot). This outcome was continuously judged as 'high risk of bias' mainly due to the outcome assessors' awareness of the assigned interventions as well as the lack of information about method of analysis and method of outcome measurement.

Figure 10. Traffic light plot for all-cause grade 3 or 4 AEs for all risk groups combined



Publication bias

We searched trial registries to identify completed trials that have not been published elsewhere in order to determine publication bias. Thereby, we identified 19 ongoing trials (see Characteristics of ongoing studies). Furthermore, we identified three trials that were completed but have not been published yet: one trial (NCT01688973) was completed in 2019, and we are awaiting publication of results in order to be able to make a decision about inclusion or exclusion of the trial in this review, as participants may have received up to one prior systemic therapy; the second trial (NCT01829841) was completed in 2018 and results are yet to be published; for the third trial (Liu 2017), we only found an abstract. We listed these three trials in the Studies awaiting classification.

Out of the 36 trials included in analyses for this review, for one trial (NCT00126594), we were able to extract data on the outcome adverse events, which were published in the 'Results' section on the trial registry (https://clinicaltrials.gov/). We only identified one publication related to the trial, in which retrospective analyses of

a subgroup of participants who were initially included in the RCT was conducted. However, these analyses were not of interest for our review, and we could not find a full-text publication of the RCT.

Allocation

All risk of bias assessments (domain judgements and support for judgements) can be found in the appendices (Appendix 9; Appendix 10; Appendix 11; Appendix 12; Appendix 13). Further details are available in a supplementary file (Aldin 2023).

Blinding

All risk of bias assessments (domain judgements and support for judgements) can be found in the appendices (Appendix 9; Appendix 10; Appendix 11; Appendix 12; Appendix 13). Further details are available in a supplementary file (Aldin 2023).

Incomplete outcome data

All risk of bias assessments (domain judgements and support for judgements) can be found in the appendices (Appendix 9; Appendix 10; Appendix 11; Appendix 12; Appendix 13). Further details are available in a supplementary file (Aldin 2023).

Selective reporting

All risk of bias assessments (domain judgements and support for judgements) can be found in the appendices (Appendix 9; Appendix 10; Appendix 11; Appendix 12; Appendix 13). Further details are available in a supplementary file (Aldin 2023).

Other potential sources of bias

All risk of bias assessments (domain judgements and support for judgements) can be found in the appendices (Appendix 9; Appendix 10; Appendix 11; Appendix 12; Appendix 13). Further details are available in a supplementary file (Aldin 2023).

Effects of interventions

See: **Summary of findings 1** Summary of findings table for all risk groups combined; **Summary of findings 2** Summary of findings table for the favourable risk groups (according to IMDC and MSKCC); **Summary of findings 3** Summary of findings for the intermediate and poor risk groups (according to IMDC and MSKCC)

Main findings

The main findings of this review are reported in the Summary of findings 1 for the combined risk groups, in the Summary of findings 2 for the favourable risk groups (and separately according to the International Metastatic RCC Database Consortium (IMDC) and the Memorial Sloan Kettering Cancer Center (MSKCC) criteria) and in the Summary of findings 3 for the intermediate and poor risk groups (and separately according to IMDC and MSKCC criteria). The main comparator in our review was SUN. For the SoF tables, we chose the clinically most relevant treatments that are currently recommended across all risk groups in four clinical practice guidelines (European Society for Medical Oncology (ESMO), European Association of Urology (EAU), National Comprehensive Cancer Network (NCCN) and the German guideline; see Description of the intervention): PEM+AXI, AVE+AXI, NIV+CAB, LEN+PEM, NIV +IPI, CAB alone, PAZ alone. For the description of results below, we also focused on these prioritised treatments. The results for the analyses of all available treatments and comparisons per outcome can be found in the figures 11 to 107 as well as in the additional tables 1 to 19. In the results section below, for each outcome, the corresponding figures and tables are linked.

Transitivity

The included trials were similar with regard to clinical and methodological characteristics, therefore we assumed that the transitivity assumption holds and conducted analyses for the outcomes OS, SAEs, PFS, AEs, and the number of participants who discontinued treatment due to an AE. All trials were RCTs and most were open-label (non-masked). For cross-over trials, we extracted data from the first period of treatment for results to be comparable. The same definitions for OS and PFS were used across trials. For analysing potential harms, we made sure that data were as comparable as possible (for more information see 'Outcome measures' in the section Included studies). All interventions were administered via the same administration route across all trials, and most interventions were administered at the same dose (see Table 1 in Included studies). Particularly our main comparator SUN was administered via the same route and at the same dosis in all trials that included SUN. Discontinuation rates of SUN were high: in 13 out of 19 trials in which SUN was administered, between 80% to 100% of participants who received SUN discontinued treatment.

All trials included participants with advanced and metastatic renal cell carcinoma (RCC), and participants in most trials had several metastatic sites. All participants were above the age of 18 years and both sexes (males and females) were included in all trials. The median age was approximately 60 years across trials. Furthermore, all trials explored first-line treatment and 80% of trials included only treatment-naive participants. For the remaining trials, we extracted data for the treatment-naive population whenever possible. Eighteen trials included only people with clear cell carcinoma and 14 trials mostly clear cell carcinoma, whereas the remaining four trials included non-clear cell carcinoma. In all trials but one, participants had previously received a nephrectomy and in most trials, prior radiotherapy was previously administered. Lastly, with regard to the risk groups, we conducted separate analyses for the different risk groups according to the different criteria (IMDC or MSKCC) whenever possible in order for results to be even more comparable.

Primary outcomes

Overall survival

Overall survival (OS) was reported in 32 trials (29 two-arm trials and three three-arm trials) (Jonasch 2010; NCT00065468; NCT00072046; NCT00081614; NCT00098657/NCT00083889; NCT00334282; NCT00420888; NCT00609401; NCT00619268; NCT00631371; NCT00719264; NCT00720941; NCT00732914; NCT00738530; NCT00903175; NCT00920816; NCT00979966; NCT01024920; NCT01030783; NCT01108445; NCT01392183; NCT01481870; NCT01613846; NCT01835158; NCT01984242; NCT02231749; NCT02420821; NCT02684006; NCT02761057; NCT02811861;NCT02853331; NCT03141177). However, evaluable data for OS was available for only 26 trials; the remaining six trials were not evaluable for this outcome for different reasons: one trial included more than 10% of previously treated participants, and separate data for treatment-naive participants was not reported for this outcome (NCT01030783); one trial did not report this outcome in a way that it would have been evaluable and estimating data were not possible (NCT00619268); four trials

were cross-over trials that did not report outcome data after the first period (NCT00732914; NCT01613846; NCT00903175; NCT01481870).

As for the 26 trials that were evaluable for this outcome, some provided data for the total population (i.e. all risk groups combined) and the different risk groups (according to MSKCC or IMDC criteria) separately, at the longest follow-up available. Other trials provided data either only for the total population or only for the different risk groups, at the longest follow-up available. With regard to the three three-arm trials, we did not combine the different arms but rather treated these as multiple independent comparisons.

Results for all risk groups combined

We analysed data on the combined risk groups (i.e. the total trial population) from 21 trials (Jonasch 2010; NCT00072046; NCT00081614; NCT00098657/NCT00083889; NCT00334282; NCT00420888; NCT00609401; NCT00631371; NCT00719264; NCT00720941; NCT00738530; NCT00920816; NCT00979966; NCT01024920; NCT01108445; NCT01984242; NCT02231749; NCT02420821; NCT02761057; NCT02811861; NCT02853331). Thereof, two three-arm trials were included, each presenting two pairwise comparisons (we did not have data for the third comparison). A total of 10,304 participants were included in the analyses. Figure 62 in Appendix 15 outlines the available direct evidence (23 pairwise comparisons). The network was not fully connected and consisted of three sub-networks (Figure 11). We conducted network meta-analysis for the sub-networks 1 and 2. Sub-network 3 contained only one trial, so no further analyses were conducted. Results for all network comparisons, including the ranking of treatments, are shown in Table 1 and Figure 12, per subnetwork.

Figure 11. Network graph for OS (all risk groups combined). Any two treatments are connected by a line when there is at least one trial comparing the two treatments.

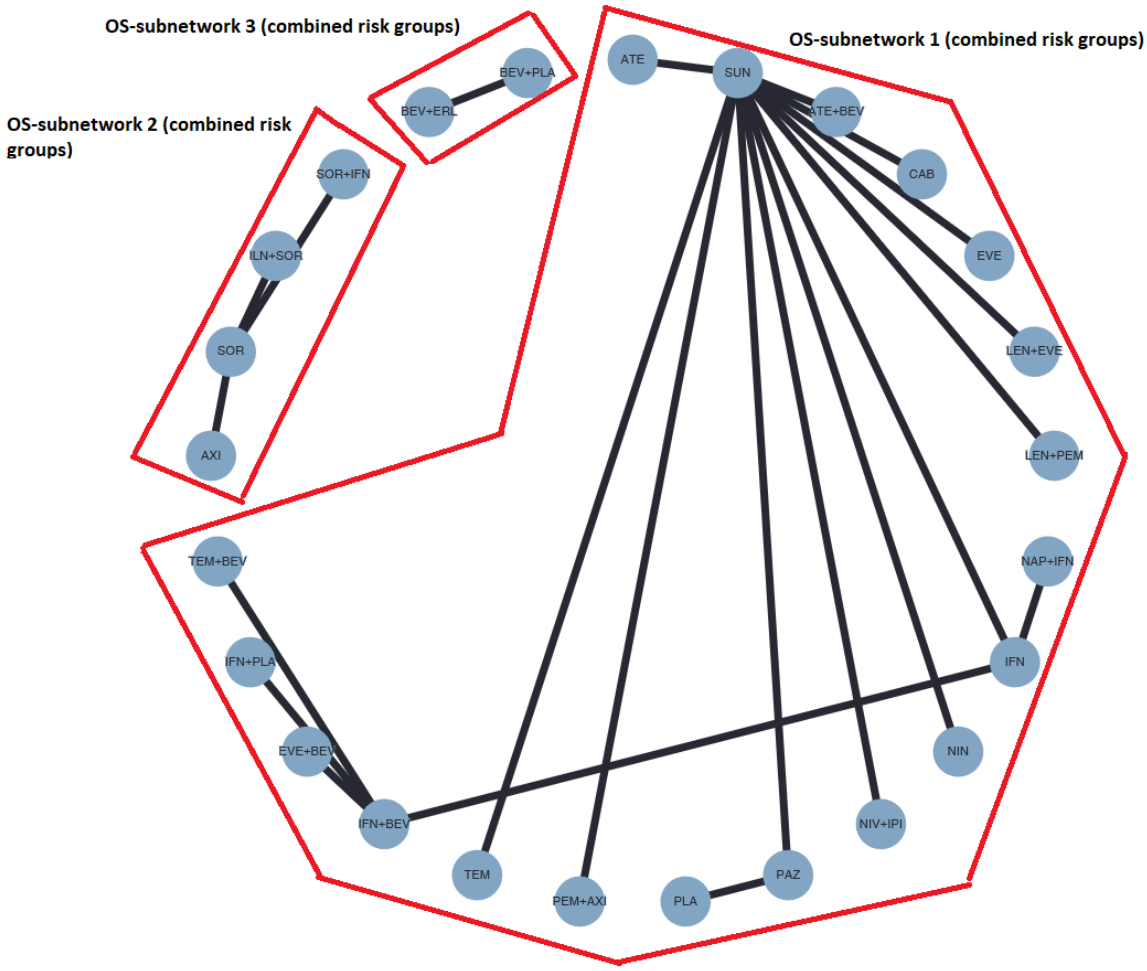
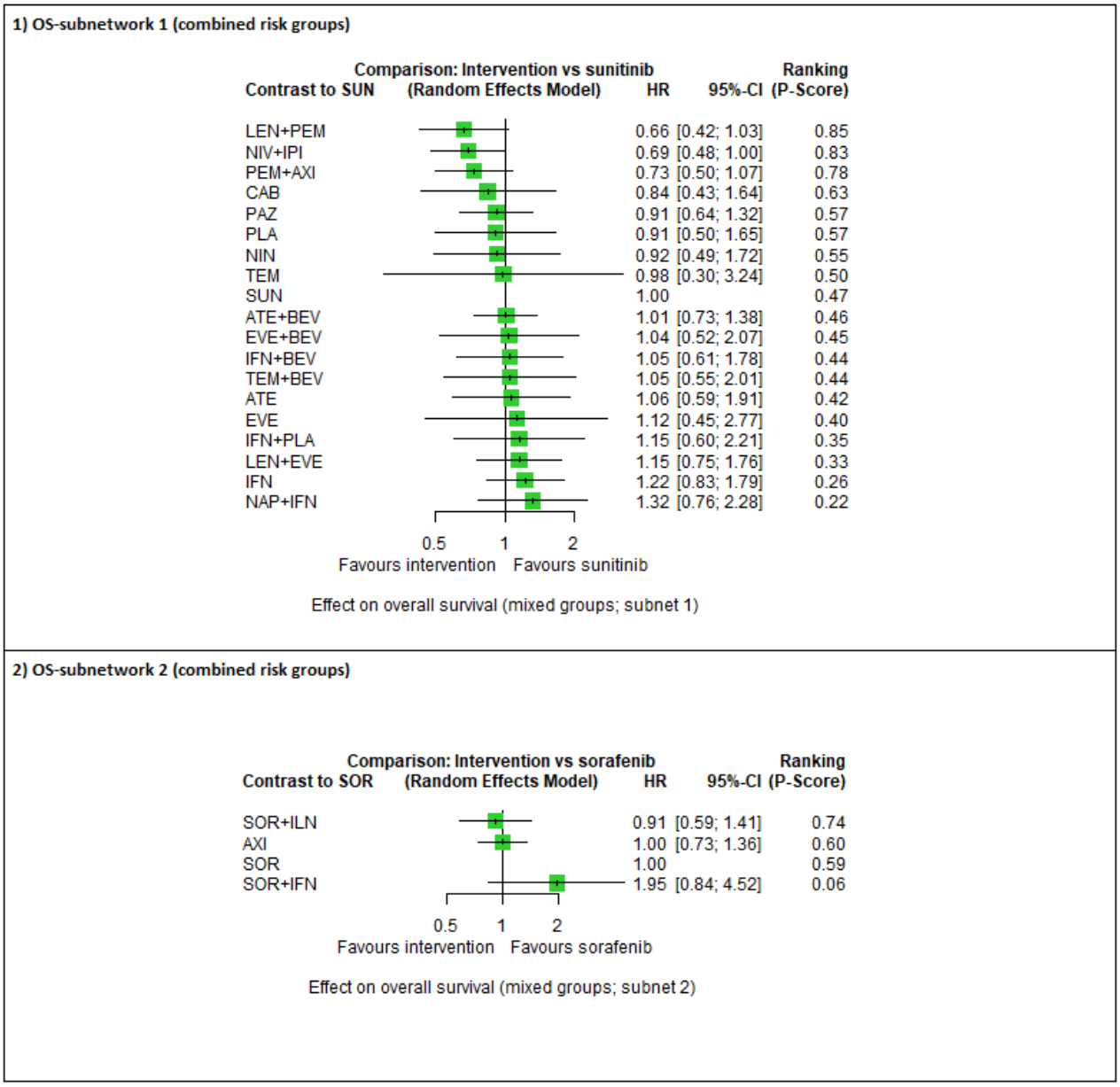


Figure 12. Forest plot for OS (all risk groups combined). 1) OS-subnetwork 1. Reference treatment: sunitinib (SUN); 2) OS-subnetwork 2. Reference treatment: sorafenib (SOR). Treatments are ordered by P-score (descending).



In sub-network 1, we observed moderate between-study heterogeneity (Q = 1.81, df = 1, P = 0.18; I^2 = 44.6%, Tau^2 = 0.0284). We found that LEN+PEM may improve OS (hazard ratio (HR) 0.66, 95% confidence interval (CI) 0.42 to 1.03), low certainty). The combinations NIV+IPI (HR 0.69, 95% CI 0.69 to 1.00, moderate certainty, P-score 0.83) and PEM+AXI (HR 0.73, 95% CI 0.50 to 1.07, moderate certainty, P-score 0.78) probably improve OS when compared to SUN alone (P 0.47), respectively. We are uncertain whether CAB alone improves OS (HR 0.84, 95% CI 0.43 to 1.64, very low certainty, P-score: 0.63) when compared to SUN alone, and there is probably little or no difference in OS between PAZ alone (HR 0.91, 95% CI 0.64 to 1.32, moderate certainty, P-score: 0.57) and SUN alone. We have no comparison data for AVE+AXI and NIV +CAB. In the ranking of treatments, LEN+PEM (P-score: 0.85) was the best treatment option, and NAP+IFN was the worst option (P-score: 0.22) (Figure 12). For this sub-network, the fixed-effect model yielded somewhat different results (see Sensitivity analysis).

In sub-network 2, each pairwise comparison was reported by a single trial only, so no heterogeneity statistics could be calculated. Here, SOR alone was the comparator treatment, and the ranking of treatments suggested that SOR+ILN (P-score: 0.74) was the best treatment option (Figure 12).

Results for MSKCC favourable risk group

We analysed data on the favourable risk group according to the MSKCC criteria from five trials (1175 participants) (NCT00072046; NCT00420888; NCT00720941; NCT00738530; NCT02811861). Figure 63 in Appendix 15 outlines the available direct evidence (six

pairwise comparisons). The network was not fully connected and consisted of two sub-networks (Figure 13). We conducted network meta-analysis for both networks. Results for all network comparisons, including the ranking of treatments, are shown in Table 2 and Figure 14. In both networks, each pairwise comparison was reported by a single trial only, so no heterogeneity statistics could be calculated.

Figure 13. Network graph for OS (MSKCC favourable risk group). Any two treatments are connected by a line when there is at least one trial comparing the two treatments. Line width: number of trials. Plot width: number of participants.

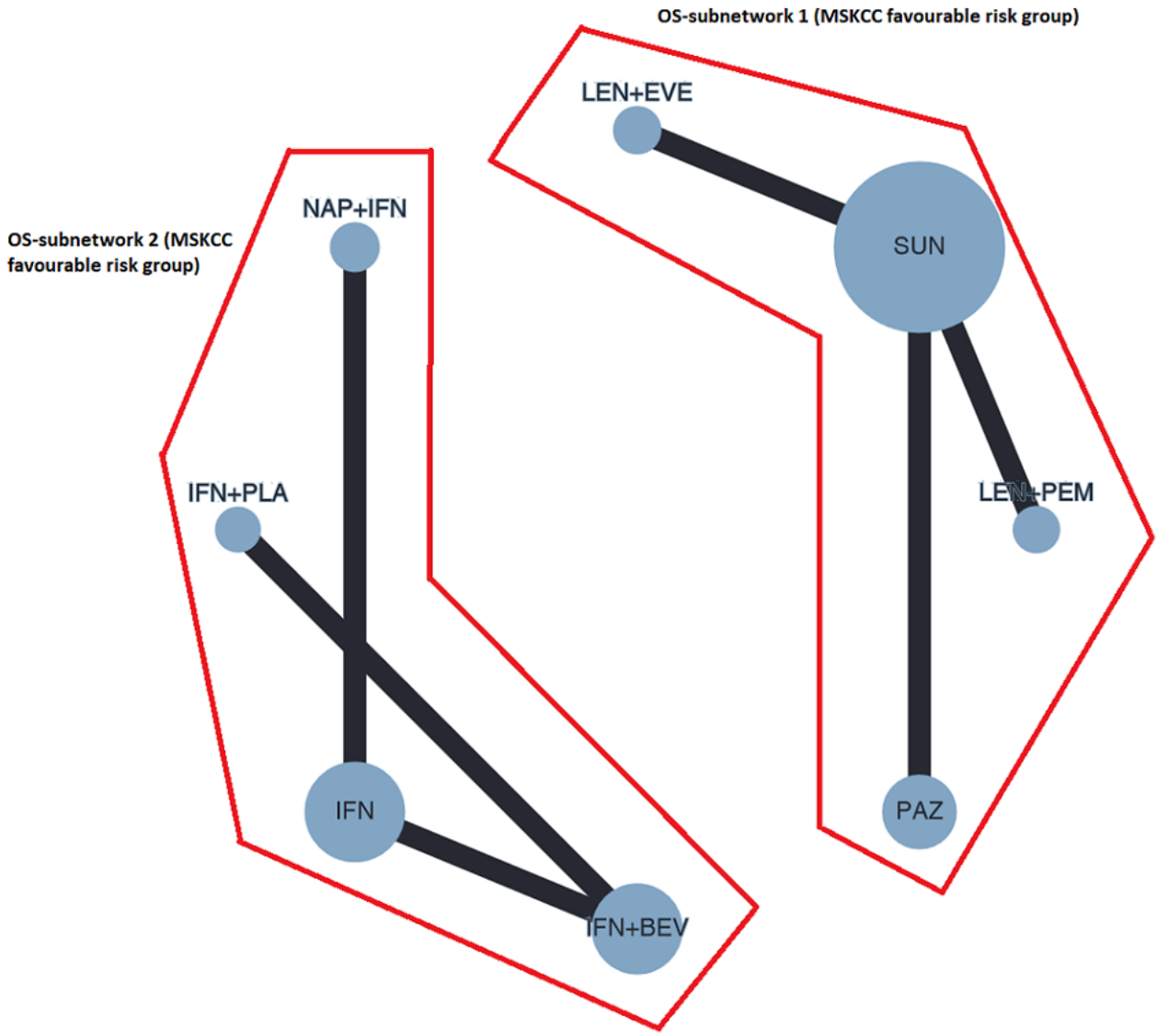
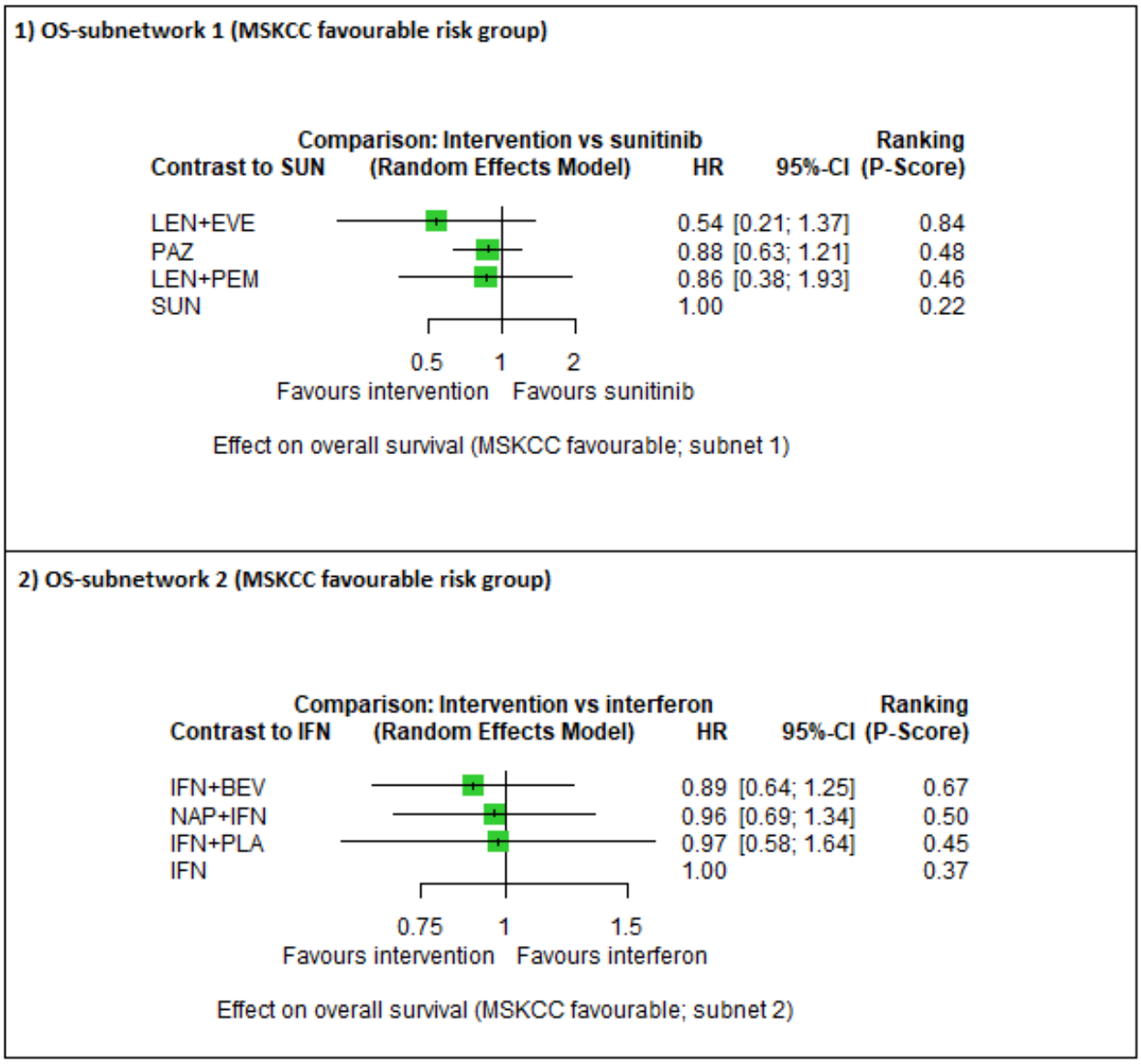


Figure 14. Forest plot for OS (MSKCC favourable risk group). 1) OS-subnetwork 1. Reference treatment: sunitinib (SUN); 2) OS-subnetwork 2. Reference treatment: interferon-alpha (IFN). Treatments are ordered by P-score (descending).



We are uncertain whether LEN+PEM improves OS (HR 0.86, 95% CI 0.38 to 1.93, very low certainty, P-score: 0.46) when compared to SUN alone (P-score: 0.22). There may be little or no difference in OS between PAZ alone (HR 0.88, 95% CI 0.63 to 1.21, low certainty, P-score: 0.48) and SUN alone. We have no comparison data for AVE +AXI, NIV+CAB, PEM+AXI, NIV+IPI and CAB alone. In the ranking of treatments, LEN+EVE was the best treatment option (P-score: 0.84) and SUN alone (P-score: 0.22) the worst option (Figure 14).

In sub-network 2, where IFN alone was the comparator treatment, the ranking of treatments suggested that IFN+BEV was the best treatment option (P-score: 0.67) and IFN alone the worst option (P-score: 0.37).

Results for IMDC favourable risk group

We analysed data on the favourable risk group according to the IMDC criteria from five trials (1007 participants) (NCT00420888; NCT02231749; NCT02684006; NCT02811861; NCT03141177). Figure 64 in Appendix 15 outlines the available direct evidence (six pairwise comparisons). The network was not fully connected and consisted of two sub-networks (Figure 15). We conducted network meta-analysis for subnetwork 1; subnetwork 2 contained only one trial, so no further analyses were conducted. Results for all network comparisons, including the ranking of treatments, are shown in Table 3 and Figure 16.

Figure 15. Network graph for OS (IMDC favourable risk group). Any two treatments are connected by a line when there is at least one trial comparing the two treatments. Line width: number of trials. Plot width: number of participants.

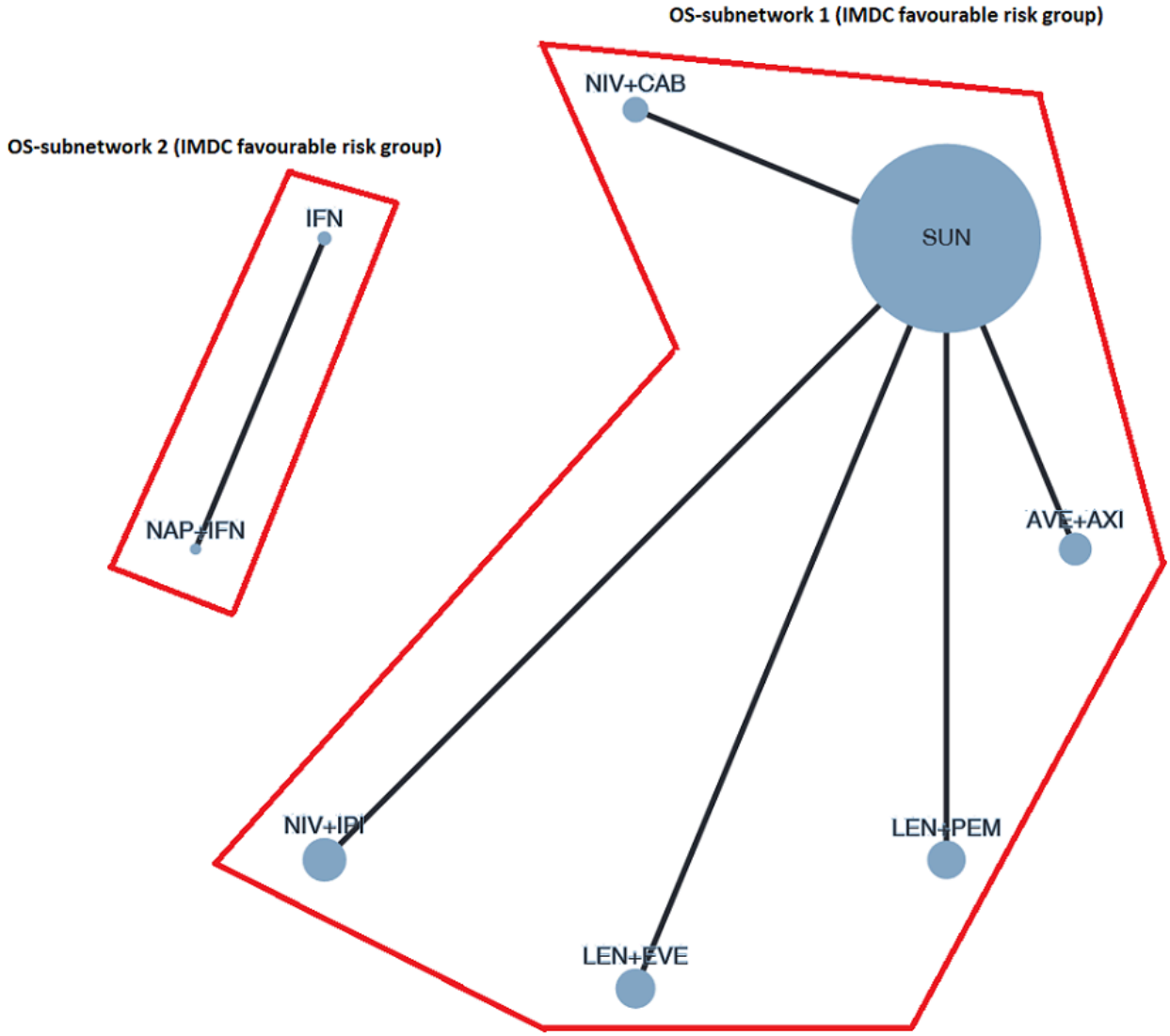
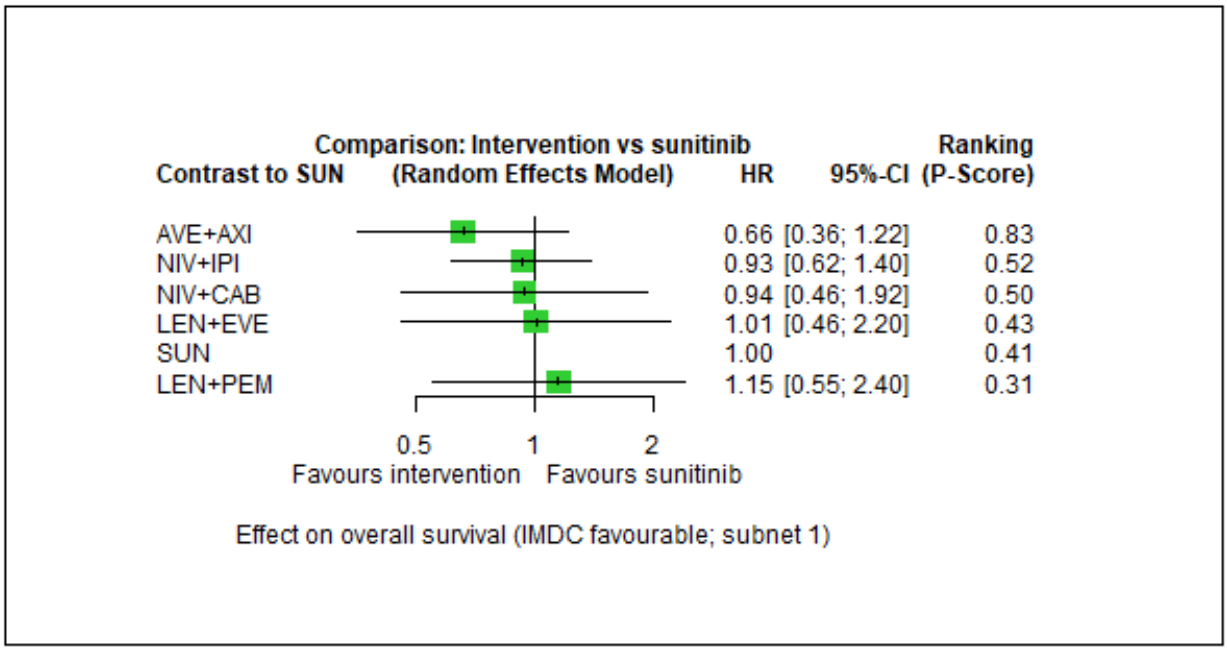


Figure 16. Forest plot for OS (IMDC favourable risk group). 1) OS-subnetwork 1. Reference treatment: sunitinib (SUN). Treatments are ordered by P-score (descending).



In sub-network 1, each pairwise comparison was reported by a single trial only, so no heterogeneity statistics could be calculated. We found that AVE+AXI may improve OS (HR 0.66, 95% CI 0.36 to 1.22, low certainty, P-score: 0.83) when compared to SUN alone (P-score: 0.41). There probably is little or no difference in OS between NIV+IPI (HR 0.93, 95% CI 0.62 to 1.40, moderate certainty, P-score: 0.52) and SUN alone, and there may be little or no difference in OS between LEN+PEM (HR 1.15, 95% CI 0.55 to 2.40, low certainty, P-score: 0.31) and SUN alone. We are uncertain whether NIV+CAB improves or decreases OS (HR 0.94, 95% CI 0.46 to 1.92, very low certainty in the evidence, P-score: 0.50) when compared to SUN alone. We have no comparison data for PEM+AXI, CAB alone and PAZ alone. In the ranking of treatments, AVE+AXI was the best treatment option (P-score: 0.83) and LEN+PEM was the worst option (P-score: 0.31) (Figure 16).

Results for MSKCC Intermediate and poor risk groups

We analysed data on the intermediate and poor risk groups according to the MSKCC criteria from seven trials (3937 participants) (NCT00065468; NCT00072046; NCT00098657/ NCT00083889; NCT00420888; NCT00720941; NCT00738530; NCT02811861). Figure 65 in Appendix 15 outlines the available direct evidence (15 pairwise comparisons). The network was fully connected (Figure 17). Results for all network comparisons, including the ranking of treatments, are shown in Table 4 and Figure 18. No heterogeneity (Q=1.45, df = 6, P = 0.96; I^2 = 0%, Tau^2 = 0.0) was detected in this network.

Figure 17. Network graph for OS (MSKCC intermediate and poor risk groups). Any two treatments are connected by a line when there is at least one trial comparing the two treatments. Line width: number of trials. Plot width: number of participants.

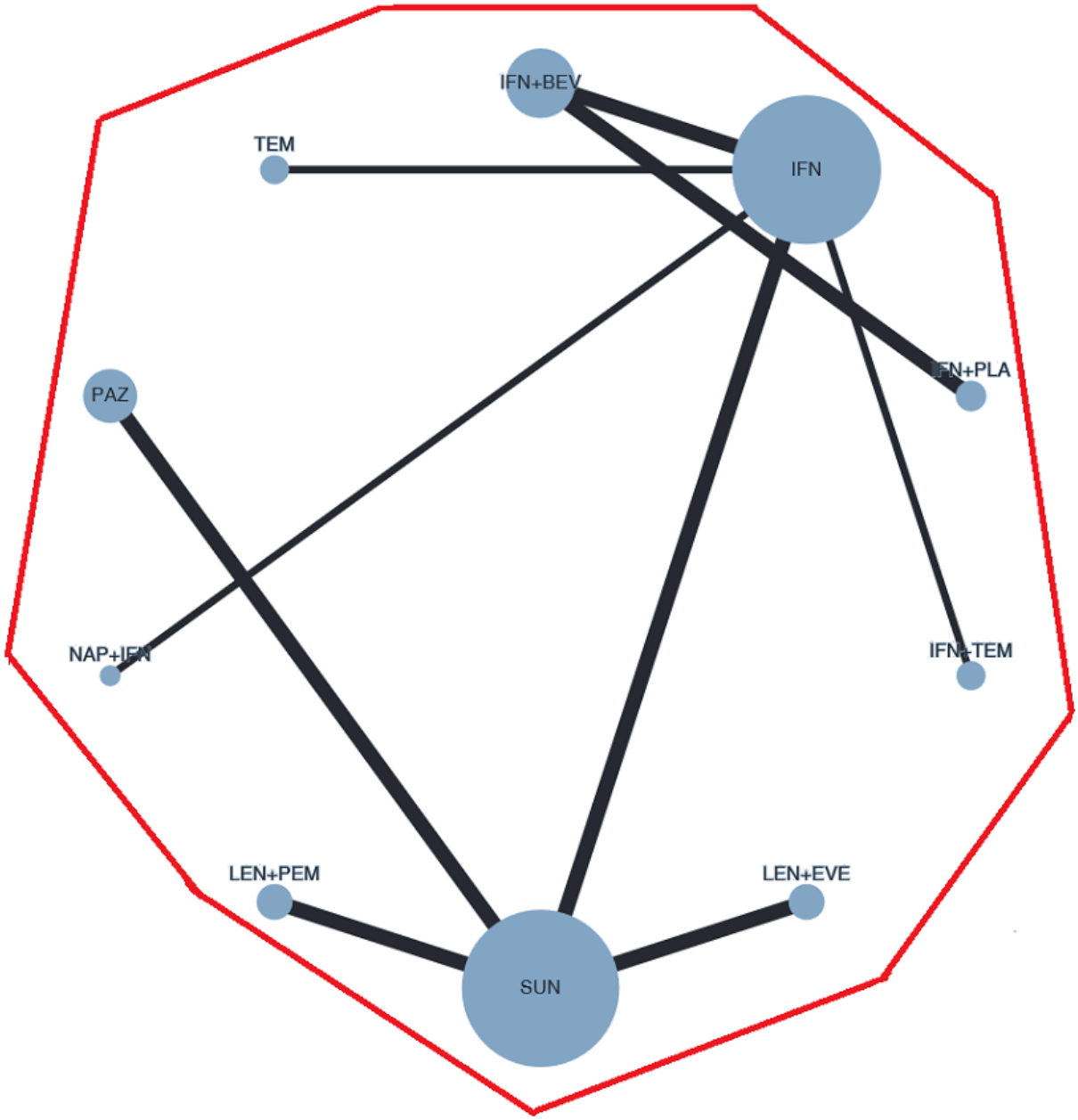
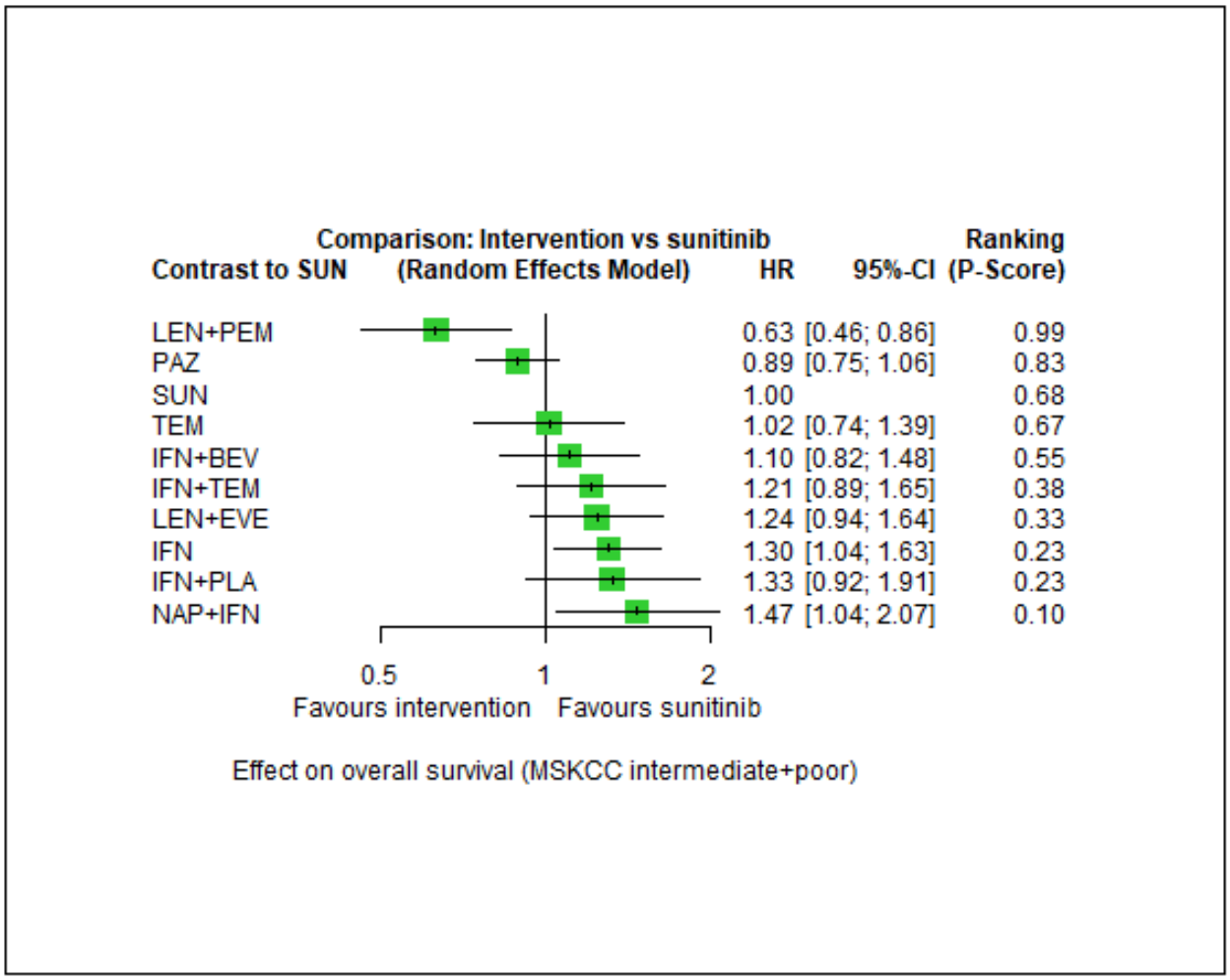


Figure 18. Forest plot for OS (MSKCC intermediate and poor risk groups). 1) OS-network. Reference treatment: sunitinib (SUN). Treatments are ordered by P-score (descending).



The combination LEN+PEM probably improves OS (HR 0.63, 95% CI 0.46 to 0.86, moderate certainty, P-score: 0.99), when compared to SUN alone (P-score: 0.68). There may be little or no difference in OS between PAZ alone (HR 0.89, 95% CI 0.75 to 1.06, low certainty, P-score: 0.83) and SUN alone. We have no comparison data for PEM +AXI, AVE+AXI, NIV+IPI, NIV+CAB, and CAB alone. In the ranking of treatments, LEN+PEM was the best treatment option (P-score 0.99) and NAP+IFN was the worst option (P-score: 0.10) (Figure 18).

Results for IMDC intermediate and poor risk groups

We analysed data on the intermediate and poor risk groups according to the IMDC criteria from seven trials (3416 participants) (NCT00420888; NCT01392183; NCT01835158; NCT02231749; NCT02684006; NCT02811861; NCT03141177). Figure 66 in Appendix 15 outlines the available direct evidence (13 pairwise comparisons). The network was not fully connected and consisted of three sub-networks (Figure 19). We conducted network meta-analysis for sub-networks 1 and 2; sub-network 3 contained only one trial, so no further analyses were conducted. Results for all network comparisons, including the ranking of treatments, are shown in Table 5 and Figure 20.

Figure 19. Network graph for OS (IMDC intermediate and poor risk groups). Any two treatments are connected by a line when there is at least one trial comparing the two treatments. Line width: number of trials. Plot width: number of participants.

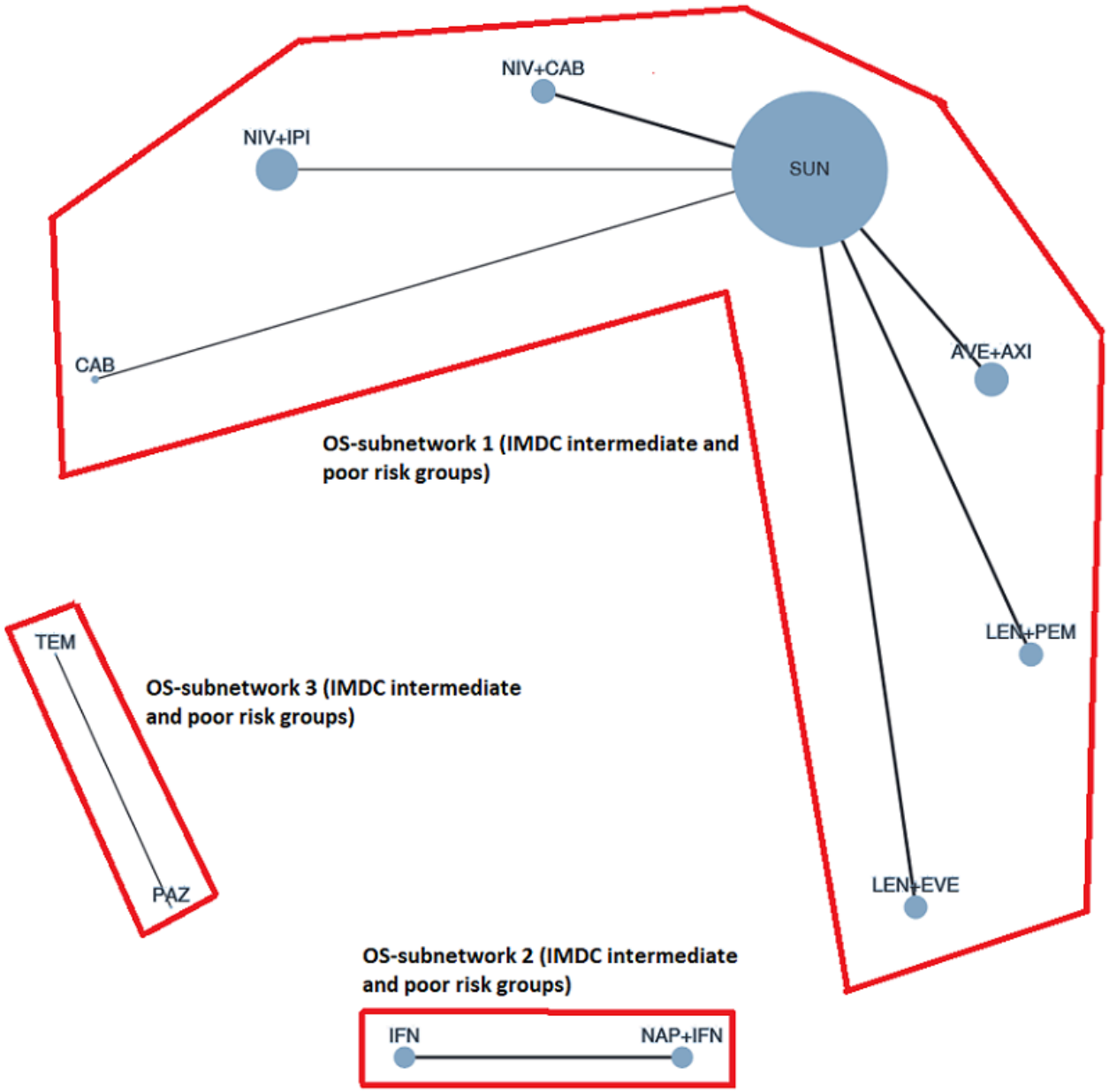
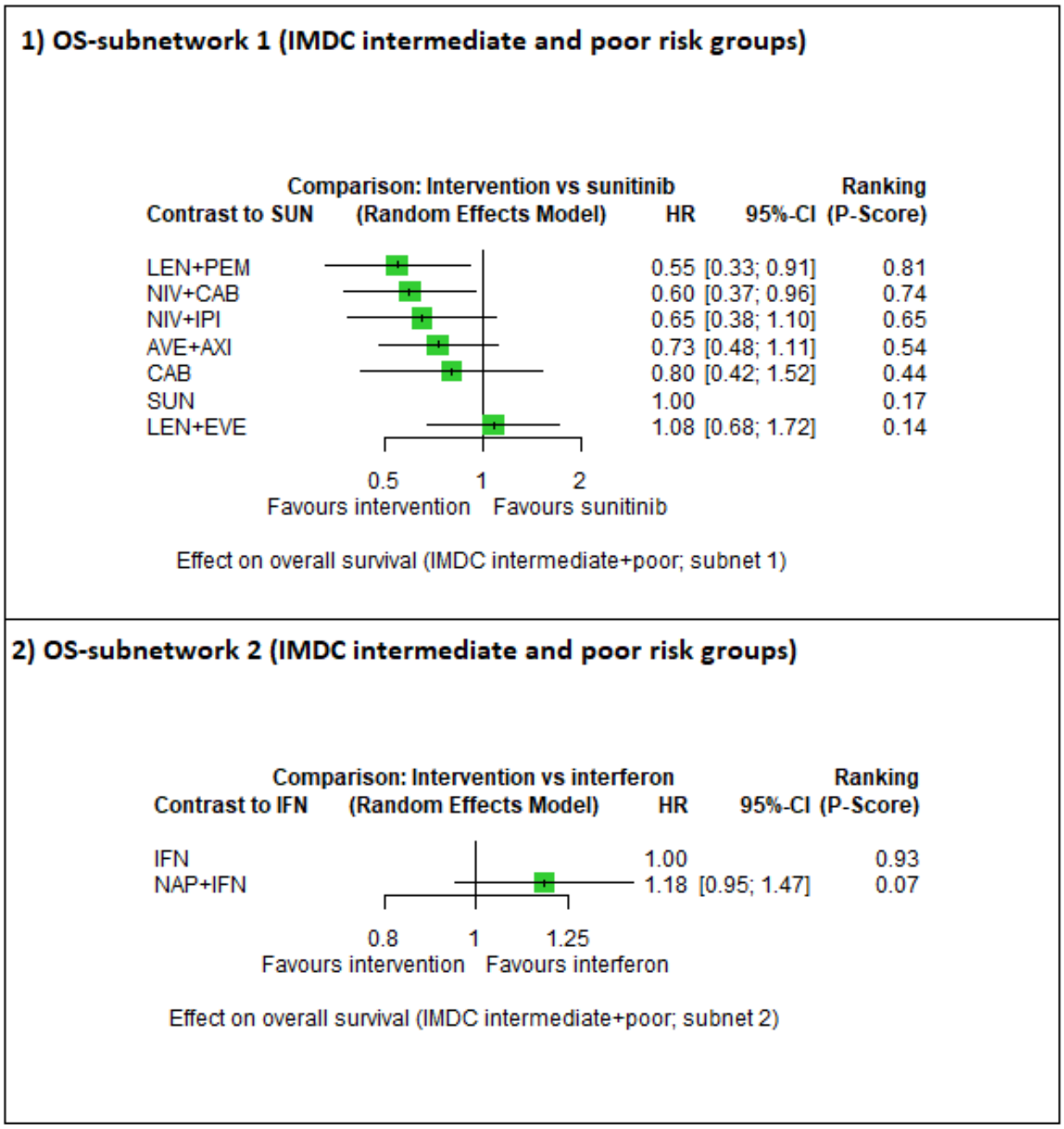


Figure 20. Forest plot for OS (IMDC intermediate and poor risk groups). 1) OS-subnetwork 1. Reference treatment: sunitinib (SUN); 2) OS-subnetwork 2. Reference treatment: interferon-alpha (IFN). Treatments are ordered by P-score (descending).



In sub-network 1, we observed moderate between-study heterogeneity (Q = 9.1, df=4, P = 0.059; I^2 = 56.1%, Tau^2 = 0.0635). The combinations LEN+PEM (HR 0.55, 95% CI 0.33 to 0.91, moderate certainty, P-score: 0.81), NIV+CAB (HR 0.60, 95% CI 0.37 to 0.96, moderate certainty, P-score: 0.74) and NIV+IPI (HR 0.65, 95% CI 0.38 to 1.10, moderate certainty, P-score: 0.65) probably improve OS when compared to SUN alone (P-score: 0.17), respectively. The combination AVE+AXI may improves OS (HR 0.73, 95% CI 0.48 to 1.11, low certainty, P-score: 0.54), and CAB alone may improves slightly OS (HR 0.80, 95% CI 0.42 to 1.50, low certainty, P-score: 0.44), when compared to SUN alone, respectively. We have no comparison data for PEM+AXI and PAZ alone. In the ranking of treatments, LEN+PEM (P-score: 0.81) was the best treatment option, and LEN+EVE was the worst option (P-score: 0.14) (Figure 20).

In sub-network 2, only one pairwise comparison by a singe trial only was reported, so no heterogeneity statistics could be calculated. Here, IFN alone was the comparator treatment, and the ranking of treatments suggested that IFN alone was the best treatment option (P-score: 0.93) and NAP+IFN the worst option (P-score: 0.07).

Quality of life

Pooling data were not feasible for the outcome quality of life (QoL), so we reported results in a tabular form. Results for the different time points are reported for every scale where data were extractable. The time point of main interest for this review was QoL at the end of treatment, which is reported below in Table 2. Results for other time points are reported in the additional tables: short-term results (one month after initiation of treatment) are reported in Table 6; mid-term results (six months after initiation of treatment) are reported in Table 7; mid-term results (12 months after initiation of treatment) are reported in Table 8; long-term results (approximately 24 months after initiation of treatment) are reported in Table 9; long-term results (at the end of treatment) are reported in Table 10.

Long-term results (at the end of treatment)

As pre-specified in the protocol of this review, we assessed the risk of bias for QoL at the end of treatment (see Table 10 for results, Risk of bias in included studies and Appendix 10). All trials were overall judged to have a 'high risk of bias' mainly due to the outcome assessors' awareness of the assigned interventions, which is owed to the nature of self-reported questionnaires and due to the trials' design (open-label, non-masked trials) as well as due to the high number of participants without outcome data at the end of treatment. In most comparisons including SUN, across all scales, participants in the experimental groups seemed to achieve a higher score in the post-intervention assessments compared to participants in the comparator arm.

One RCT measured QoL using FACIT-F (score range 0 to 52; higher scores mean better QoL) and reported that the mean post-score was 9.00 points higher (9.86 lower to 27.86 higher, very low certainty) with PAZ than with SUN. Comparison data were not available for PEM+AXI, AVE+AXI, NIV+CAB, LEN+PEM, NIV+IPI, and CAB alone.

Serious adverse events

Serious adverse events (SAEs) were not consistently reported across trials. To be able to meta-analyse results, we only considered SAEs when the number of participants with at least one SAE was reported. We did not consider cumulated events. Serious adverse events were assessed in 22 trials (NCT00065468; NCT00098657/NCT00083889; NCT00117637; NCT00126594; NCT00619268; NCT00631371; NCT00719264; NCT00720941; NCT00732914; NCT00738530; NCT00903175; NCT00920816; NCT00979966; NCT01024920; NCT01108445; NCT01613846; NCT01835158; NCT01984242; NCT02231749; NCT02420821; NCT02811861; NCT02853331) (18 two-arm trials, four three-arm trials), for a total of 10,709 participants. All trials provided data on SAE for all risk groups combined only. Figure 67 in Appendix 15 outlines the available direct evidence (31 comparisons). The network was fully connected (Figure 21). Results for all network comparisons, including the ranking of treatments, are shown in Table 11 and Figure 22. We observed substantial heterogeneity (Q_{total}= 15.40, df = 6, P = 0.017; Q_{within}=3.44, df=1, P = 0.064; $Q_{between}$=11.96, df=5, P = 0.035; I^2 = 61.0%, Tau^2 = 0.0256) in the network.

Figure 21. Network graph for SAEs (all risk groups combined). Any two treatments are connected by a line when there is at least one trial comparing the two treatments. Line width: number of trials. Plot width: number of participants.

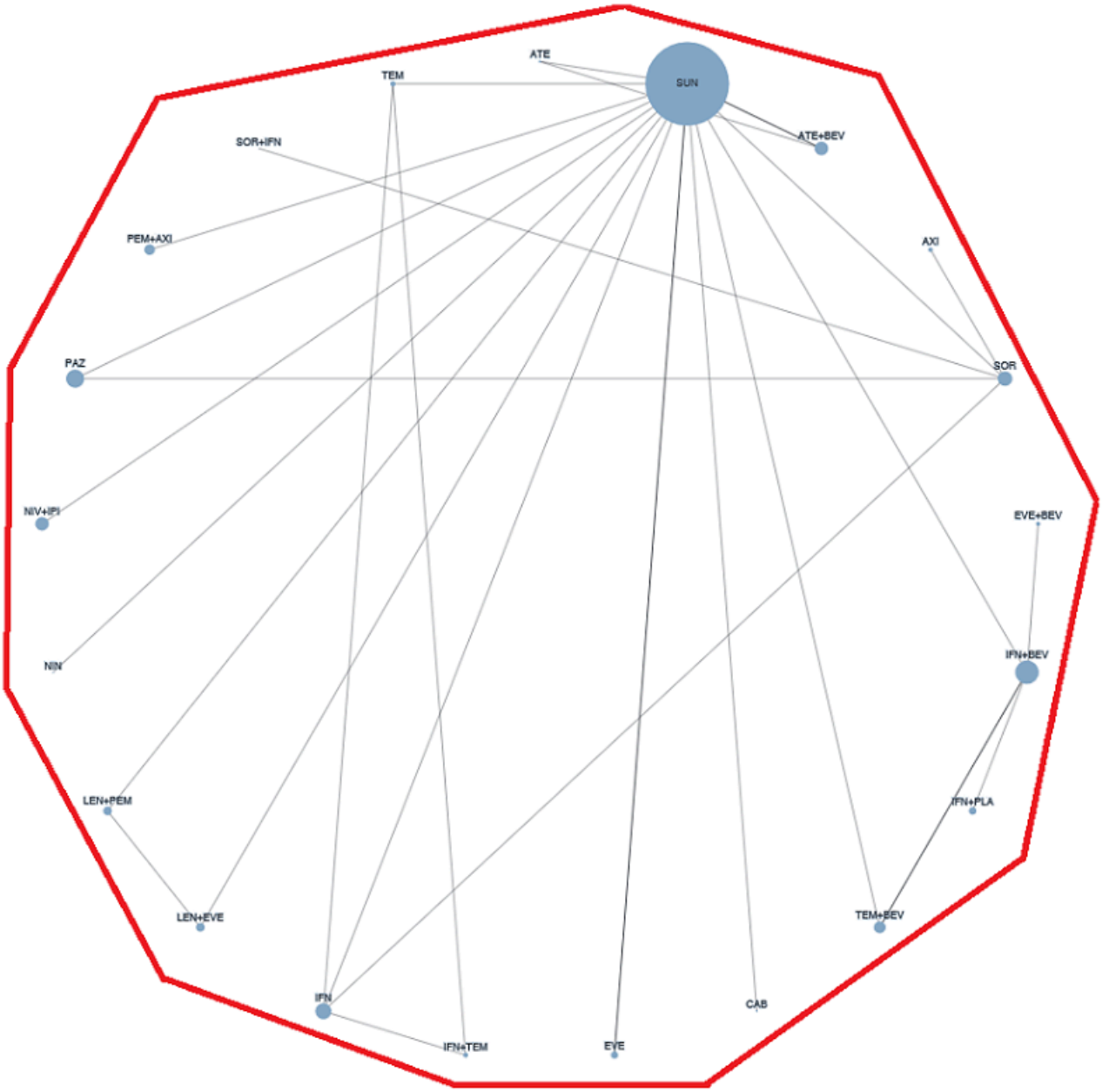
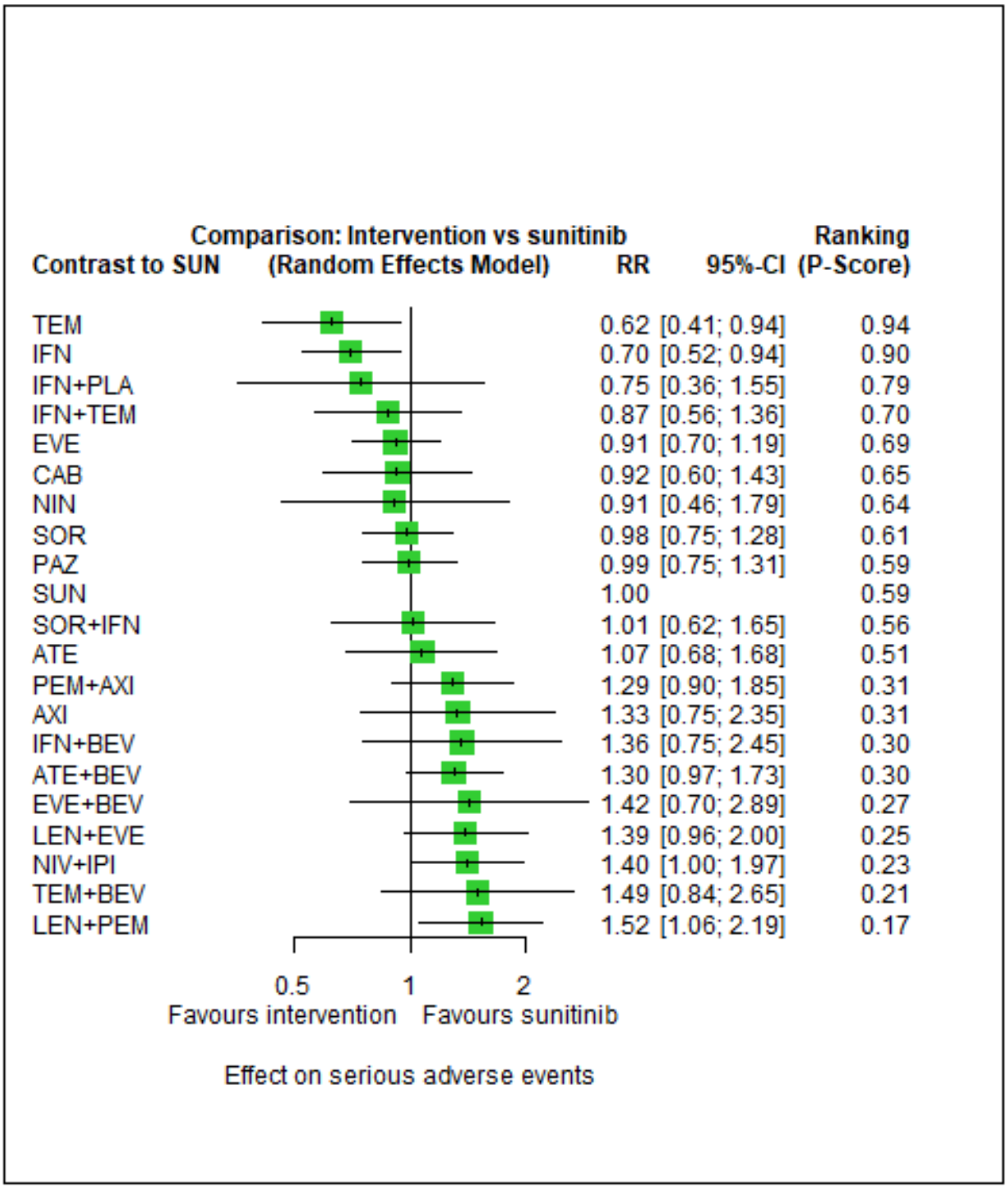


Figure 22. Forest plot for SAEs (all risk groups combined). Reference treatment: sunitinib (SUN). Treatments are ordered by P-score (descending).



PEM+AXI probably increase slightly the risk for SAEs (risk ratio (RR) 1.29, 95% CI 0.90 to 1.85, moderate certainty, P-score: 0.31), when compared to SUN alone (P-score: 0.59). The combinations LEN +PEM (RR 1.52, 95% CI 1.06 to 2.19, moderate certainty, P-score: 0.17) and NIV+IPI (RR 1.40, 95% CI 1.00 to 1.97, moderate certainty, P-score: 0.23) probably increase the risk for SAEs when compared to SUN alone, respectively. We are uncertain whether CAB alone reduces or increases the risk for SAE (RR 0.92, 95% CI 0.60 to 1.43, very low certainty, P-score: 0.65) when compared to SUN alone, and there is probably little or no difference in the risk for SAEs between PAZ alone (RR 0.99, 95% CI 0.75 to 1.31, moderate certainty, P-score: 0.59) and SUN alone. Comparison data were not available for

AVE+AXI and NIV+CAB. In the ranking of treatments, TEM alone (P-score: 0.94) was the best treatment option, and LEN+PEM the worst option (P-score: 0.17) (Figure 22). The fixed-effect model yielded somewhat different results (Sensitivity analysis).

There are closed loops in the network (Figure 21). Figure 68 in Appendix 15 depicts the forest plot of splitting direct and indirect evidence. There was no significant difference between direct and indirect estimates (data not shown). The net heat plot showed negligible signs for inconsistency (Figure 69 in Appendix 15).

Secondary outcomes

Progression-free survival

Progression-free survival (PFS) was reported in 34 trials (31 two-arm trials and three three-arm trials) (Jonasch 2010; NCT00065468; NCT00072046; NCT00081614; NCT00098657/NCT00083889; NCT00117637; NCT00334282; NCT00420888; NCT00609401; NCT00619268; NCT00631371; NCT00719264; NCT00720941; NCT00732914; NCT00738530; NCT00903175; NCT00920816; NCT00979966; NCT01024920; NCT01030783; NCT01108445; NCT01274273; NCT01392183; NCT01481870; NCT01613846; NCT01835158; NCT01984242; NCT02231749; NCT02420821; NCT02684006; NCT02761057; NCT02811861; NCT02853331; NCT03141177). However, evaluable data for PFS was available for only 30 trials; the remaining four trials were not evaluable for this outcome for different reasons: one trial was a cross-over trial that did not report outcome data after the first period (NCT01613846); three trials did not report this outcome in a way that it would have been evaluable and estimating data were not possible (NCT00609401; NCT00619268; NCT01274273).

As for the 30 trials that were evaluable for this outcome, some provided data for the total population (i.e. all risk groups combined) and the different risk groups (according to MSKCC and/or IMDC criteria) separately, at the longest follow-up available, while some trials provided data for either the total population or for the different risk groups only. With regard to the three three-arm trials, we did not combine the different arms but rather treated these as multiple independent comparisons.

Results for all risk groups combined

We analysed data on the combined risk groups from 26 trials (11,840 participants) (Jonasch 2010; NCT00072046; NCT00081614; NCT00098657/NCT00083889; NCT00117637; NCT00334282; NCT00420888; NCT00631371; NCT00719264; NCT00720941; NCT00732914; NCT00738530; NCT00903175; NCT00920816; NCT00979966; NCT01024920; NCT01030783; NCT01108445; NCT01481870; NCT01835158; NCT01984242; NCT02231749; NCT02420821; NCT02761057; NCT02811861; NCT02853331). Thereof, two three-arm trials were included, each presenting two pairwise comparisons (we did not have data for the third comparison). Figure 70 in Appendix 15 outlines the available direct evidence (28 pairwise comparisons). The network was not fully connected and consisted of two sub-networks (Figure 23). We conducted network meta-analysis for subnetwork 1; subnetwork 2 contained only one trial, so no further analyses were conducted. Results for all network comparisons, including the ranking of treatments, are shown in Table 12 and Figure 24.

Figure 23. Network graph for PFS (all risk groups combined). Any two treatments are connected by a line when there is at least one trial comparing the two treatments. The green lines highlight the one available closed loop.

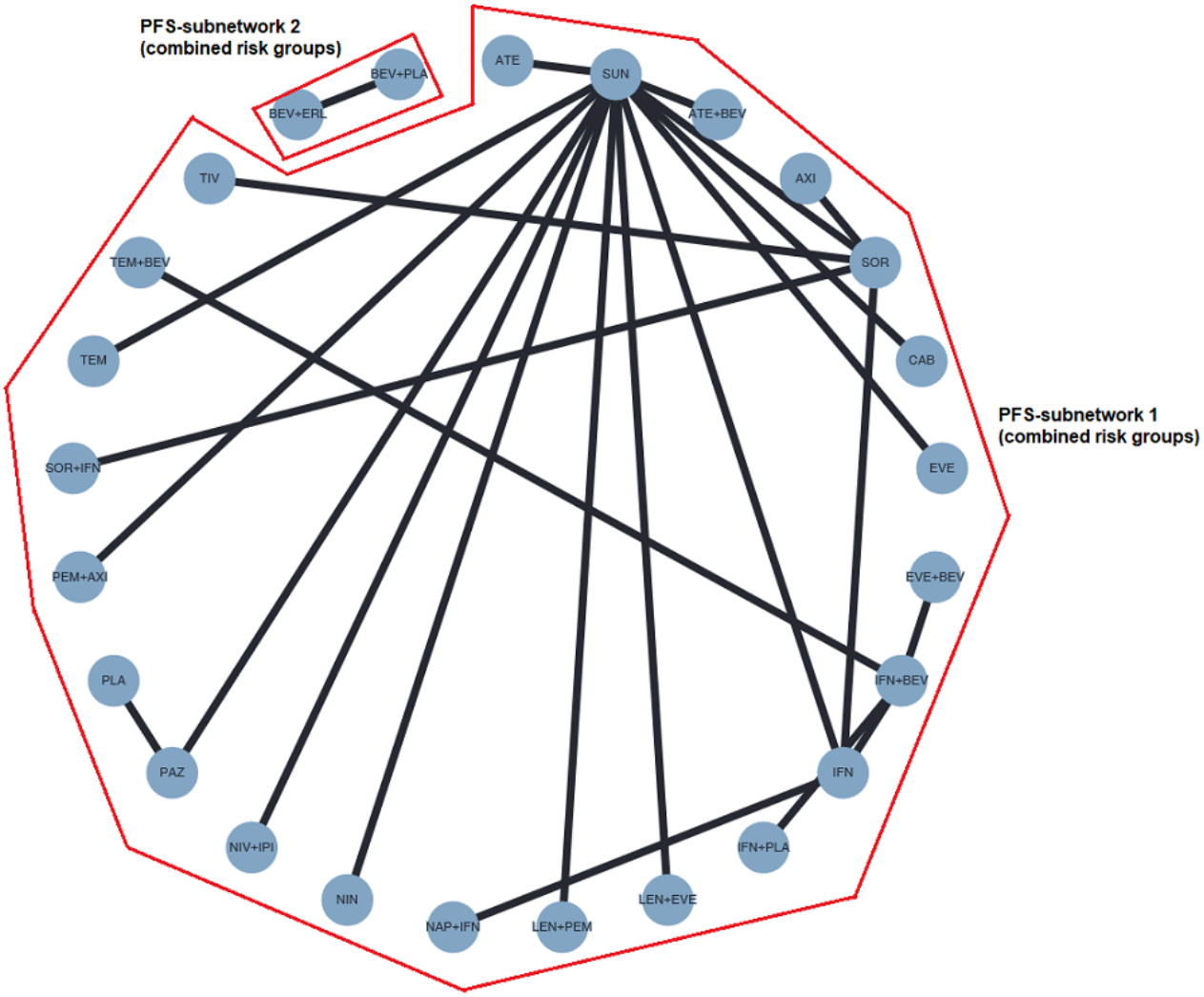
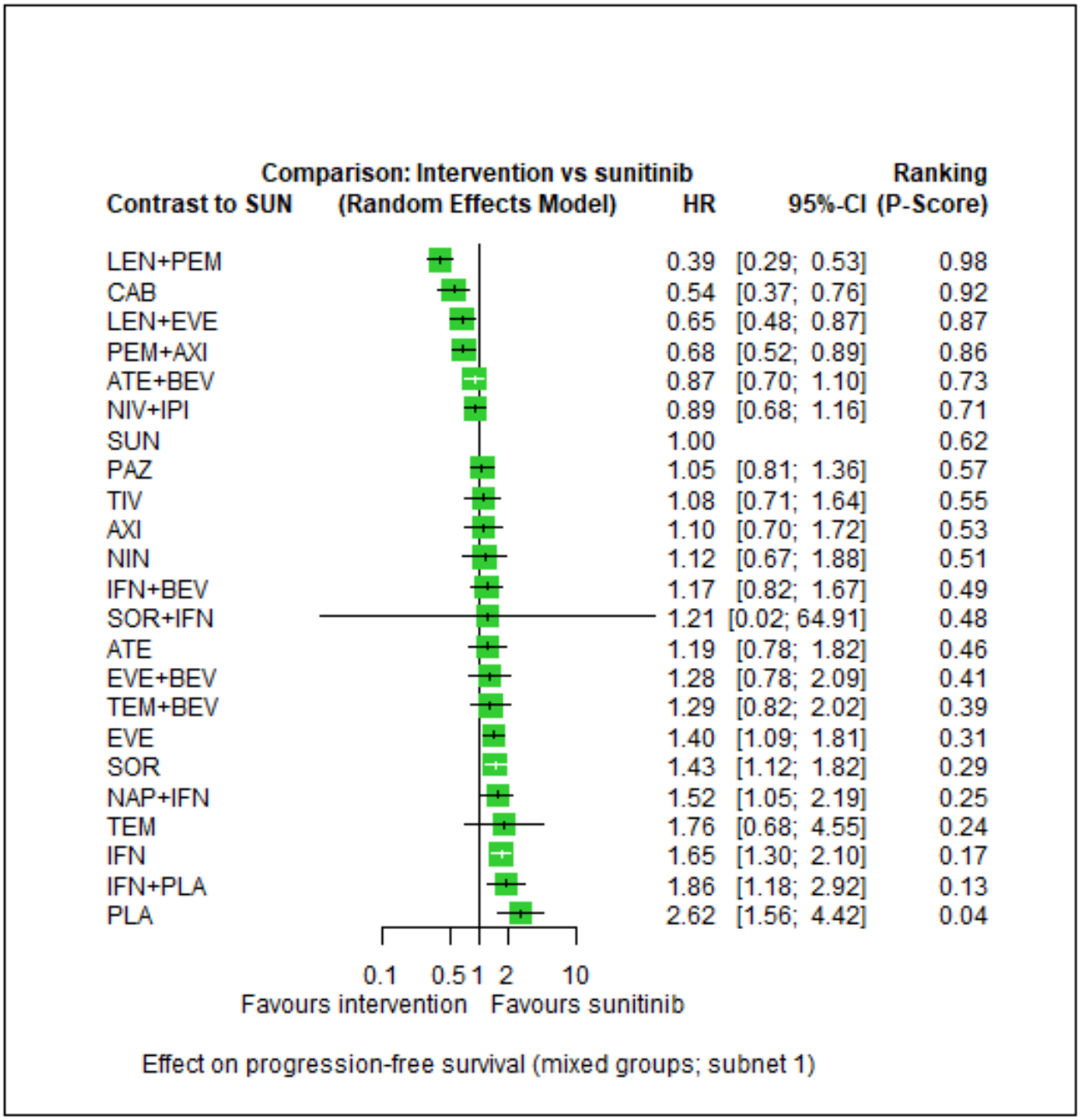


Figure 24. Forest plot for PFS (all risk groups combined). 1) PFS-subnetwork 1. Reference treatment: sunitinib (SUN). Treatments are ordered by P-score (descending).



In subnetwork 1, we observed little between-study heterogeneity (Q_{total} = 6.93, df = 5, P = 0.23; Q_{within} = 2.02, df = 4, P = 0.73; $Q_{between}$ = 4.91, df = 1, P = 0.027; I^2 = 27.9%, Tau^2 = 0.0155). The combination LEN+PEM (HR 0.39, 95% CI 0.29 to 0.53, moderate certainty, P-score: 0.98) probably improves PFS, and CAB alone may improve PFS (HR 0.54, 95% CI 0.37 to 0.76, low certainty, P-score: 0.92), when compared to SUN alone (P-score: 0.62), respectively. PEM+AXI probably improve slightly PFS (HR 0.68, 95% CI 0.52 to 0.89, moderate certainty, P-score: 0.86), when compared to SUN alone. There probably is little or no difference in PFS between PAZ alone (HR 1.05, 95% CI 0.81 to 1.36, moderate certainty, P-score: 0.57) and SUN alone, and there may be little or no difference in PFS between NIV+IPI (HR 0.89, 95% CI 0.68 to 1.16, low certainty, P-score: 0.71) and SUN alone. Comparison data were not available for AVE+AXI and NIV+CAB. In the ranking of treatments, LEN+PEM (P-score: 0.98) was the best treatment option, and IFN+PLA the worst option (0.13) (Figure 24).

As shown in Figure 23, there was one closed loop in the network. Figure 71 in Appendix 15 depicts the forest plot of splitting direct and indirect evidence. Results suggested that there is no difference between direct and indirect estimates (P = 0.083 (data not shown)).

The net heat plot showed small signs for inconsistency (Figure 72 in Appendix 15).

Results for MSKCC favourable risk group

We analysed data on the favourable risk group according to the MSKCC criteria from nine trials (1410 participants) (NCT00420888; NCT00631371; NCT00732914; NCT00738530; NCT00903175; NCT00920816; NCT01108445; NCT01481870; NCT02811861). Of these, one three-arm trial was included, presenting two pairwise comparisons (we did not have data for the third comparison). One additional trial (NCT00920816) did not report a confidence interval, and it was not possible to reconstruct it, so we excluded this trial from this analysis. Figure 73 in Appendix 15 outlines the available direct evidence (10 pairwise comparisons). The network was not fully connected and consisted of three sub-networks (Figure 25). We conducted network meta-analysis for the sub-networks 1 and 2; subnetwork 3 contained only one trial, so no further analyses were conducted. Results for all network comparisons, including the ranking of treatments, are shown in Table 13 and Figure 26, per subnetwork.

Figure 25. Network graph for PFS (MSKCC favourable risk group). Any two treatments are connected by a line when there is at least one trial comparing the two treatments. Line width: number of trials. Plot width: number of participants.

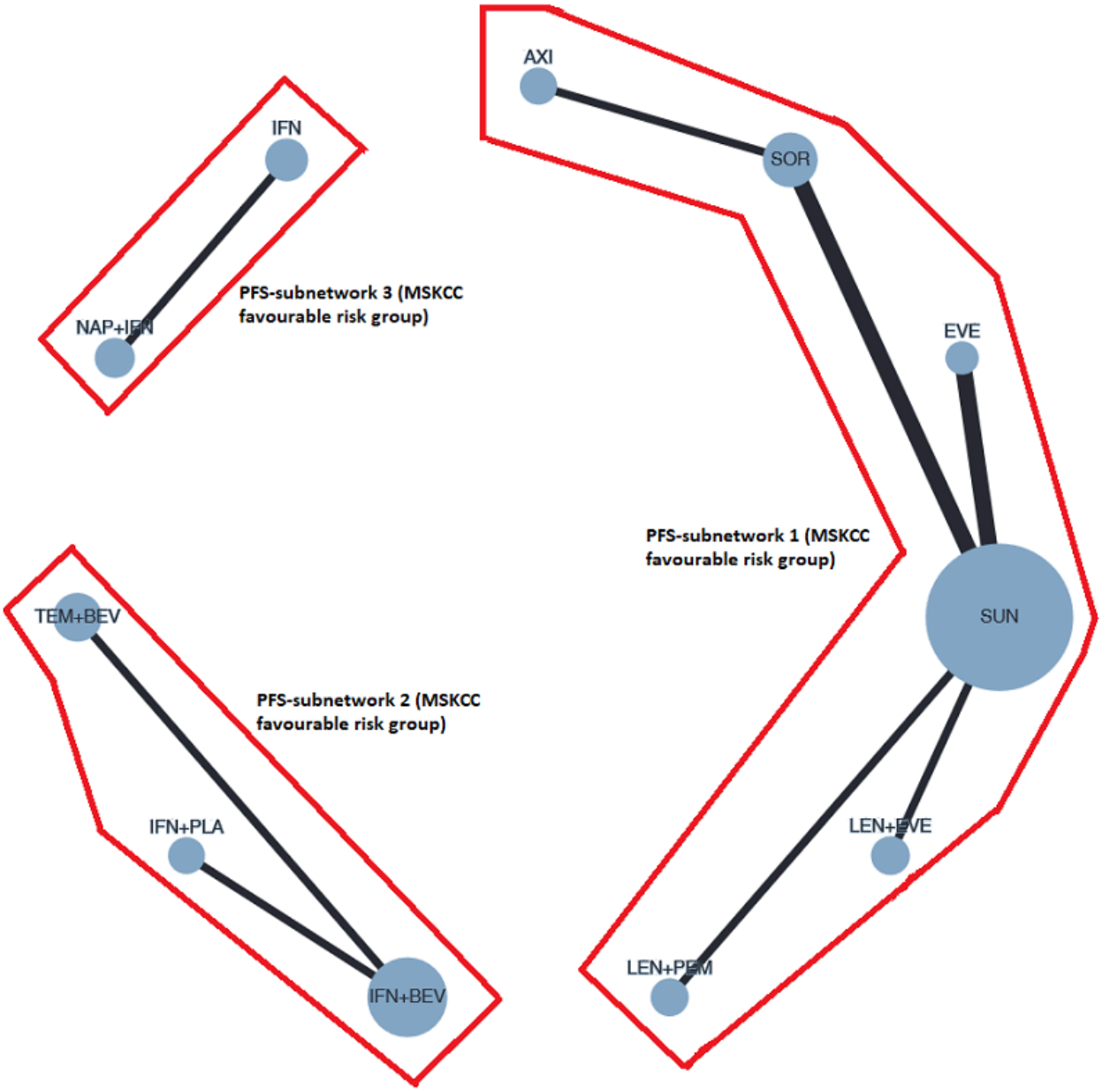
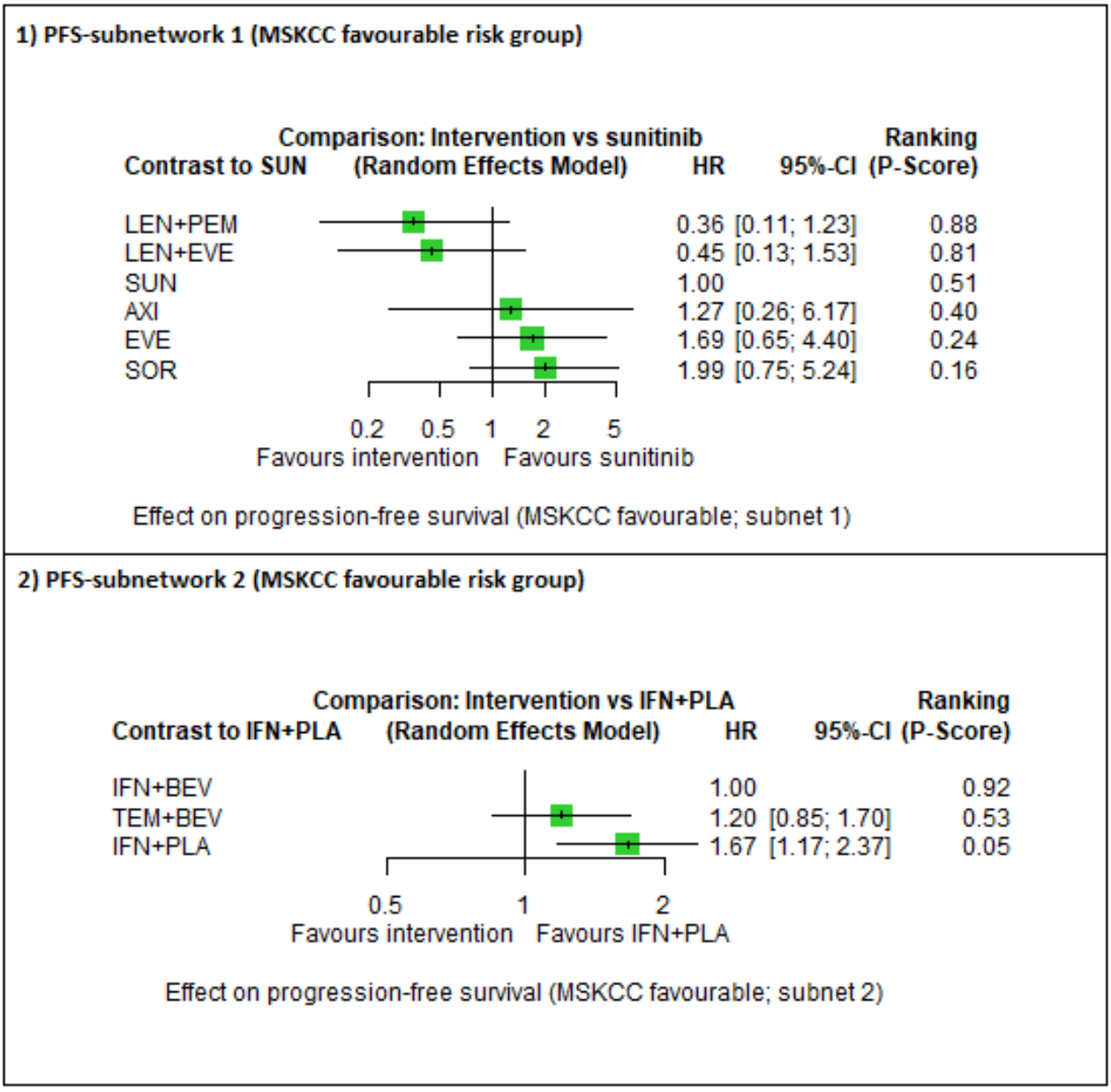


Figure 26. Forest plot for PFS (MSKCC favourable risk group). 1) PFS-subnetwork 1. Reference treatment: sunitinib (SUN); 2) PFS-subnetwork 2. Reference treatment: interferon-alpha + bevacizumab (IFN+BEV). Treatments are ordered by P-score (descending).



In sub-network 1, we observed substantial heterogeneity (Q = 6.16, df = 2, P = 0.046; I^2 = 67.6%, Tau^2 = 0.3473). We are uncertain whether LEN+PEM (HR 0.36, 95% CI 0.11 to 1.23, very low certainty, P-score: 0.88) improves PFS when compared to SUN alone (P-score: 0.51). Comparison data were not available for PEM+AXI, AVE+AXI, NIV+CAB, NIV+IPI, PAZ alone and CAB alone. In the ranking of treatments, LEN-PEM (P-score: 0.88) was the best treatment option, whereas SOR alone was the worst option (P-score: 0.16) (Figure 26). The fixed-effect model yielded different results (see Sensitivity analysis).

In subnetwork 2, each pairwise comparison was reported by a single trial only, so no heterogeneity statistics could be calculated. Here, IFN+BEV was the comparator, and the ranking of treatments suggested that IFN+BEV was the best treatment option (P-score: 0.92) whereas IFN+PLA was the worst option (P-score: 0.05).

Results for IMDC favourable risk groups

We analysed data on the favourable risk group according to the IMDC criteria from five trials (1007 participants) (NCT00420888; NCT02231749; NCT02684006; NCT02811861; NCT03141177). Thereof, one three-arm trial was included, presenting two pairwise comparisons (we did not have data for the third comparison). Figure 74 in Appendix 15 outlines the available direct evidence (six pairwise comparisons). The network was not fully connected and

consisted of two sub-networks (Figure 27). We conducted network meta-analysis for subnetwork 1. Subnetwork 2 contained only one trial, so no further analyses were conducted. Results for all network comparisons in subnet 1, including the ranking of treatments, are shown in Table 14 and in Figure 28.

Figure 27. Network graph for PFS (IMDC favourable risk group). Any two treatments are connected by a line when there is at least one trial comparing the two treatments. Line width: number of trials. Plot width: number of participants.

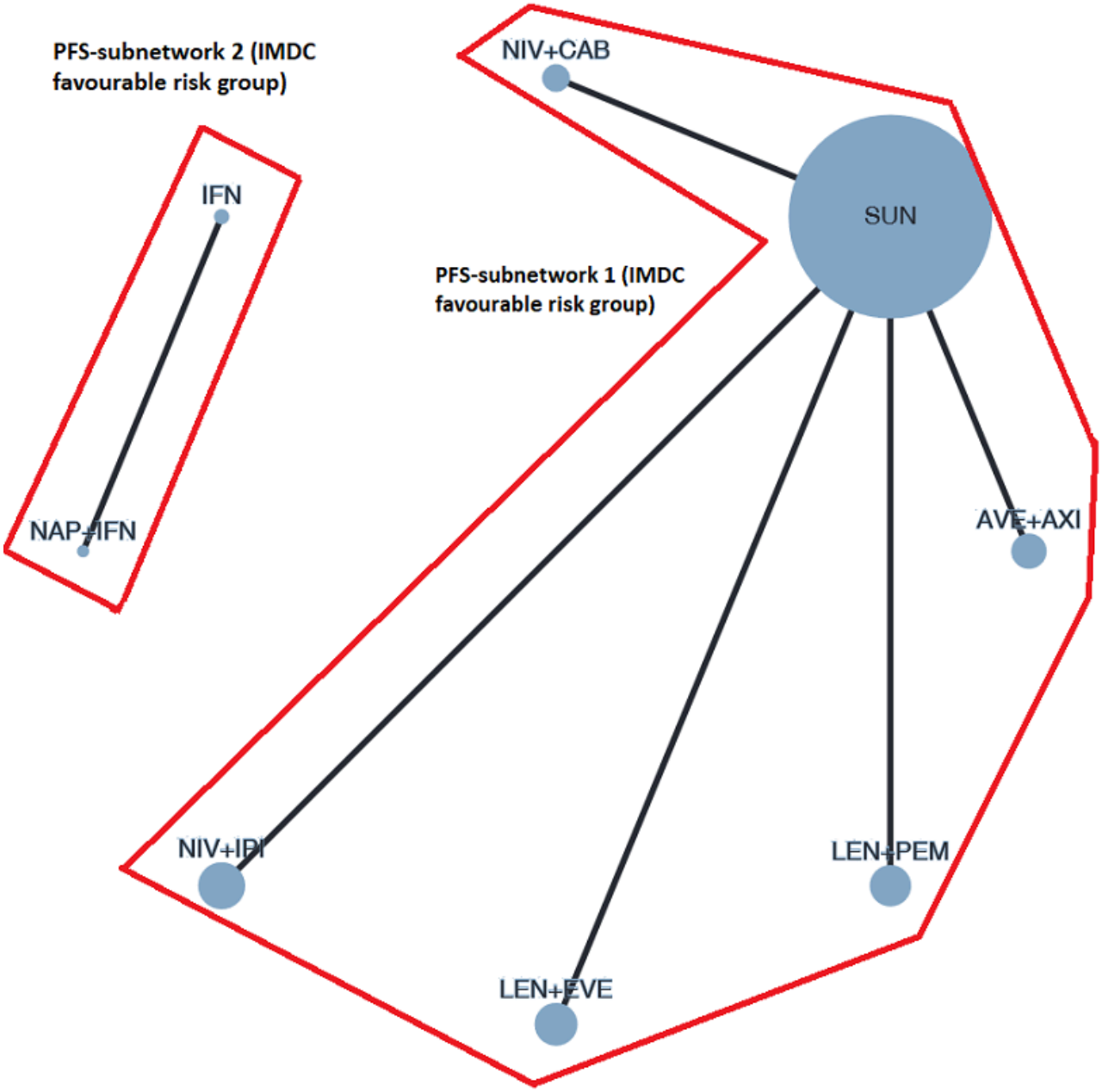
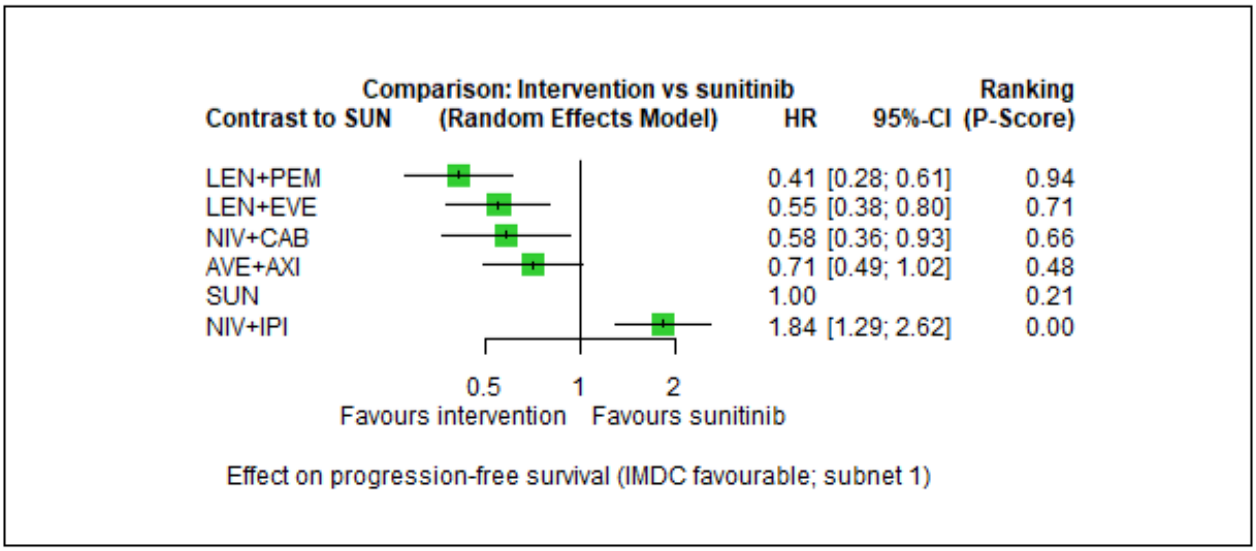


Figure 28. Forest plot for PFS (IMDC favourable risk group). 1) PFS-subnetwork 1. Reference treatment: sunitinib (SUN). Treatments are ordered by P-score (descending).



In sub-network 1, each pairwise comparison was reported by a single trial only, so no heterogeneity statistics could be calculated. The combinations LEN+PEM (HR 0.41, 95% CI 0.28 to 0.61, low certainty, P-score: 0.94), NIV+CAB (HR 0.58, 95% CI 0.36 to 0.93, low certainty, P-score: 0.66) and AVE+AXI (HR 0.71, 95% CI 0.49 to 1.02, low certainty, P-score: 0.48) may improve PFS when compared to SUN alone (P-score: 0.21, respectively. The combination NIV+IPI probably reduces PFS (HR 1.84, 95% CI 1.29 to 2.62, moderate certainty, P-score: 0.00), when compared to SUN. Comparison data were not available for PAZ alone, CAB alone and PEM+AXI. In the ranking of treatments, LEN+PEM was the best treatment option (P-score: 0.94) and NIV+IPI was the worst (P-score: 0.00) (Figure 28).

Results for MSKCC intermediate and poor risk groups

We analysed data on the intermediate and poor risk groups according to the MSKCC criteria from eight trials (2797 participants) (NCT00420888; NCT00631371; NCT00732914; NCT00738530; NCT00903175; NCT00920816; NCT01108445; NCT02811861). Of these, one three-arm trial was included, presenting two pairwise comparisons (we did not have data for the third comparison). Figure 75 in Appendix 15 outlines the available direct evidence (15 pairwise comparisons). The network was not fully connected and consisted of three sub-networks (Figure 29). We conducted network meta-analysis for sub-networks 1 and 2. Subnetwork 3 contained only one trial, so no further analyses were conducted. Results for all network comparisons, including the ranking of treatments, are shown in Table 15 and Figure 30.

Figure 29. Network graph for PFS (MSKCC intermediate and poor risk groups). Any two treatments are connected by a line when there is at least one trial comparing the two treatments. Line width: number of trials. Plot width: number of participants.

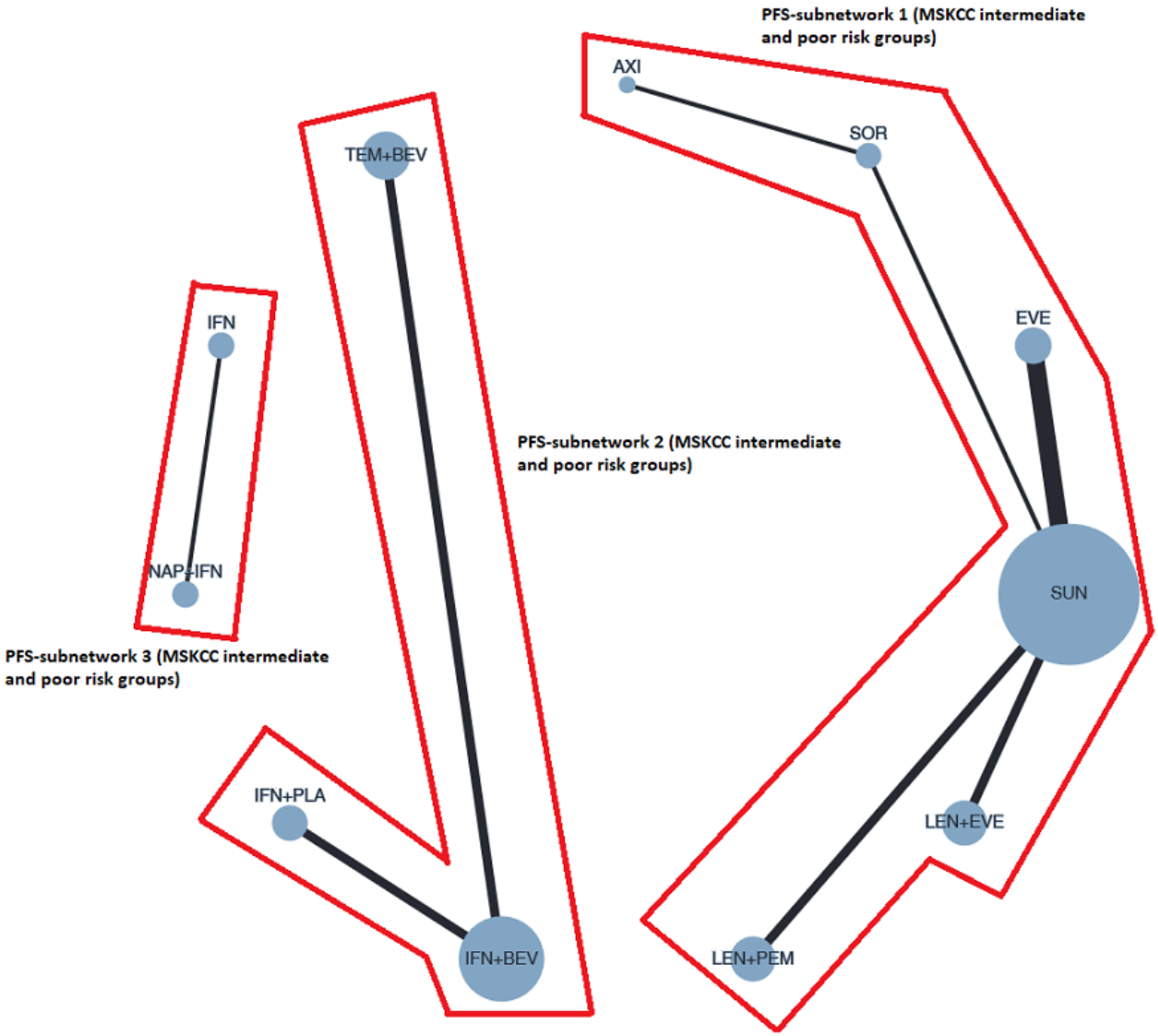
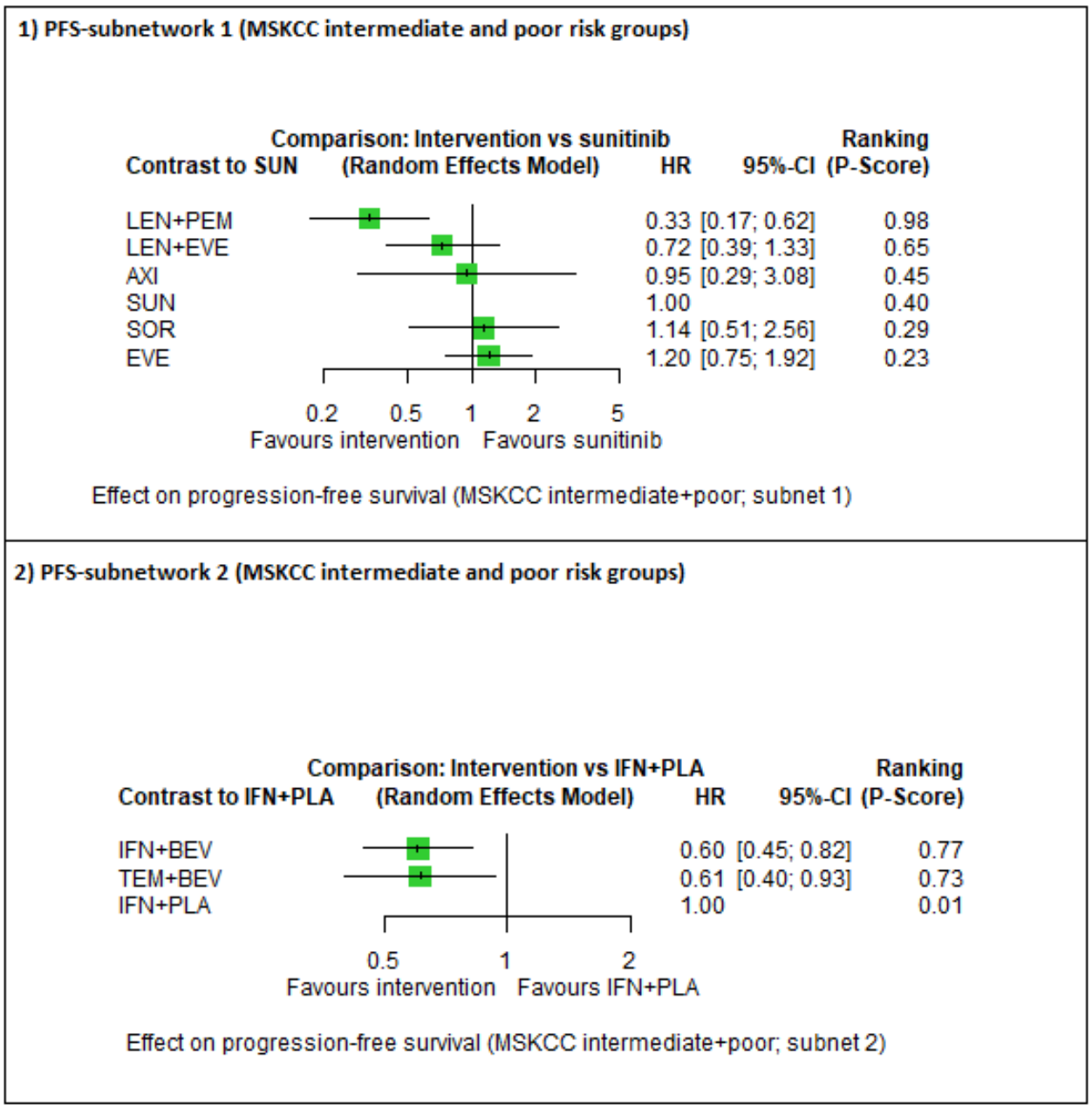


Figure 30. Forest plot for PFS (MSKCC intermediate and poor risk groups). 1) PFS-subnetwork 1. Reference treatment: sunitinib (SUN); 2) PFS-subnetwork 2. Reference treatment: interferon-alpha + placebo (IFN+PLA). Treatments are ordered by P-score (descending).



In sub-network 1, we observed substantial between-study heterogeneity (Q = 14.40, df=5, P = 0.013; I^2 = 65.3%, Tau^2 = 0.1433). The combination LEN+PEM probably improves PFS (HR 0.33, 95% CI 0.17 to 0.62, moderate certainty, P-score: 0.98) when compared to SUN alone (P-score: 0.40). Comparison data were not available for PEM+AXI, AVE+AXI, NIV+IPI, NIV+CAB, PAZ alone and CAB alone. In the ranking of treatments, LEN+PEM was the best treatment option (P-score: 0.98), and EVE alone was the worst option (P-score: 0.23) (Figure 30). The fixed-effect model yielded somewhat different results (Sensitivity analysis).

In sub-network 2, where IFN+PLA was the comparator treatment, we observed moderate between-study heterogeneity (Q=2.79, df=2, P = 0.247; I^2 = 28.3%, Tau^2 = 0.0175). In the ranking of treatments, IFN+BEV was the better treatment option (P-score: 0.77) and IFN+PLA the worst (P-score: 0.01). The fixed-effect model yielded slightly different results (Sensitivity analysis).

Results for IMDC intermediate and poor risk groups

We analysed data on the intermediate and poor risk groups according to the IMDC criteria from eight trials (4042 participants) (NCT00065468; NCT00420888; NCT01392183; NCT01835158;

NCT02231749; NCT02684006; NCT02811861; NCT03141177). Of these, two three-arm trials were included, each presenting two pairwise comparisons (we did not have data for the third comparison). Figure 76 in Appendix 15 outlines the available direct evidence (16 pairwise comparisons). The network was not fully connected and consisted of two sub-networks (Figure 31). We conducted network meta-analysis for both networks. Results for all network comparisons, including the ranking of treatments, are shown in Table 16 and Figure 32, per subnetwork.

Figure 31. Network graph for PFS (IMDC intermediate and poor risk groups). Any two treatments are connected by a line when there is at least one trial comparing the two treatments. Line width: number of trials. Plot width: number of participants.

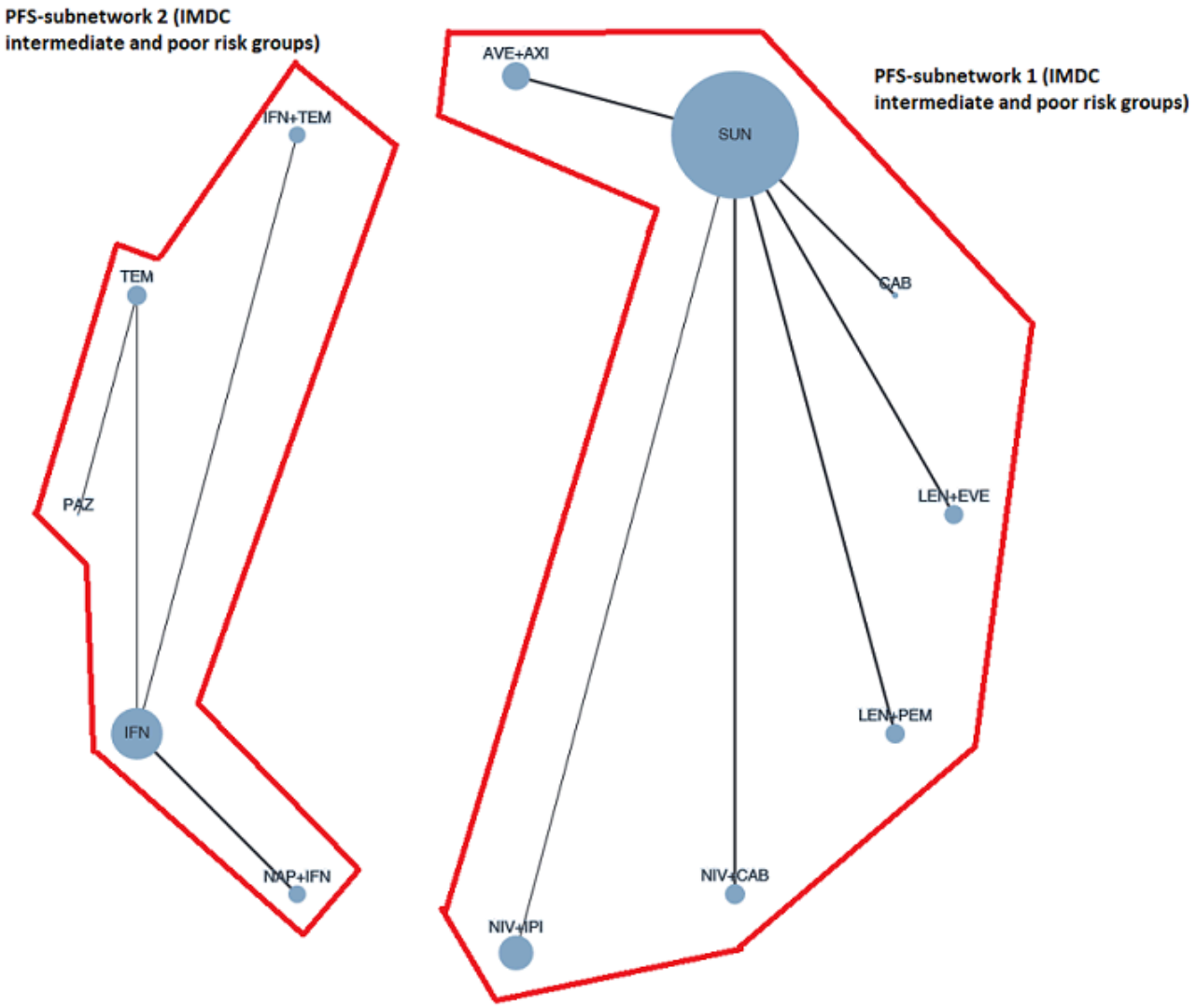
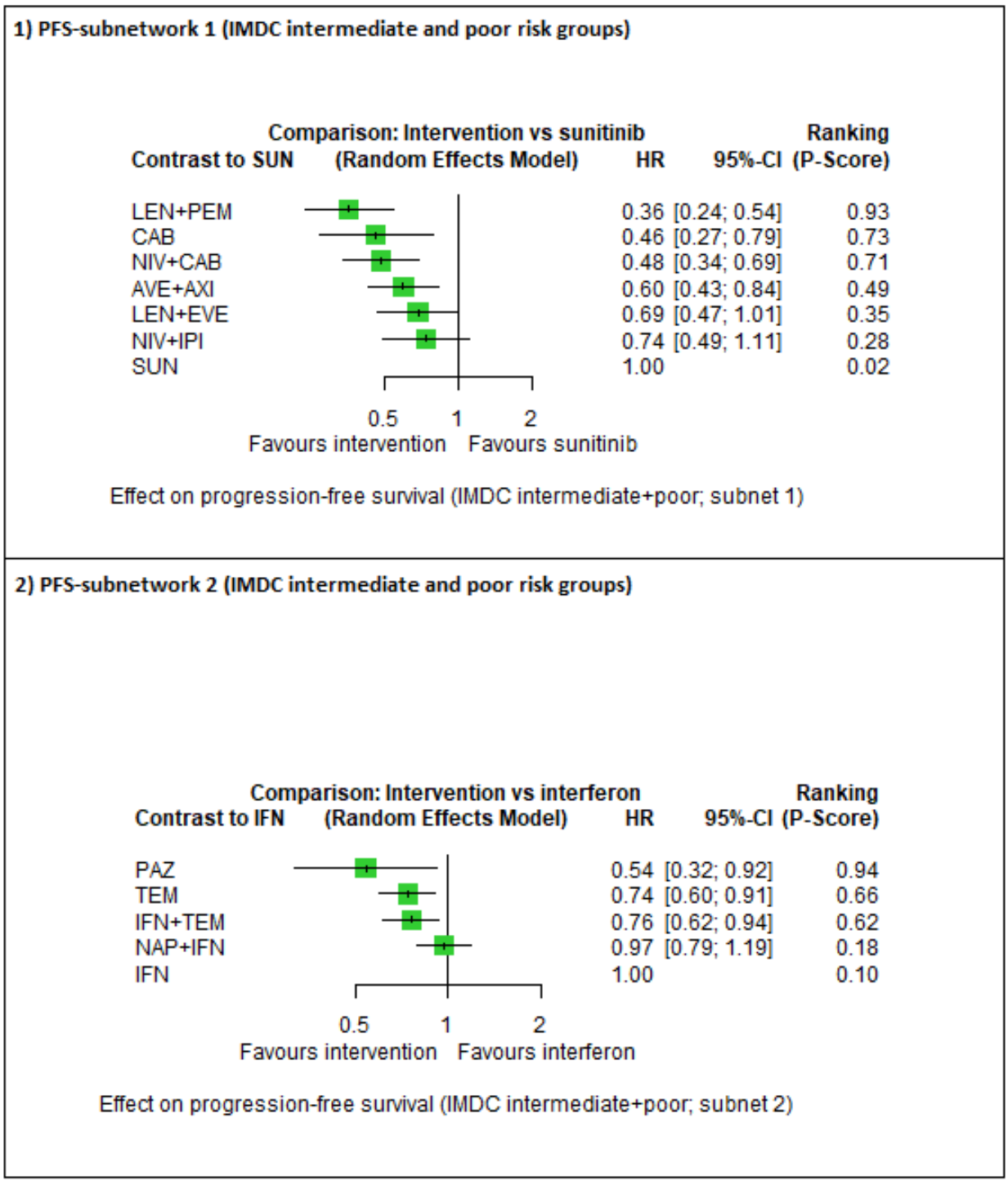


Figure 32. Forest plot for PFS (IMDC intermediate and poor risk groups). 1) PFS-subnetwork 1. Reference treatment: sunitinib (SUN); 2) PFS-subnetwork 2. Reference treatment: interferon-alpha (IFN). Treatments are ordered by P-score (descending).



In sub-network 1, we observed moderate between-study heterogeneity (Q=8.7, df = 5, P = 0.12; I^2 = 42.5%, Tau^2 = 0.0357). Cabozantinib alone (HR 0.46, 95% CI 0.27 to 0.79, moderate certainty, P-score: 0.73), and the combinations LEN+PEM (HR 0.36, 95% CI 0.24 to 0.54, moderate certainty, P-score: 0.93), NIV+CAB (HR 0.48, 95% CI 0.34 to 0.69, moderate certainty, P-score: 0.71) and AVE+AXI (HR 0.60, 95% CI 0.43 to 0.84, moderate certainty, P-score: 0.49) probably improve PFS when compared to SUN alone (P-score: 0.02), respectively. There may be little or no difference in PFS between NIV+IPI (HR 0.74, 95% CI 0.49 to 1.11, low certainty, P-score: 0.28) and SUN alone. Comparison data were not available for PEM+AXI and PAZ alone. In the ranking of treatments, LEN+PEM

was the best treatment option (P-score: 0.93), and SUN alone was the worst option (P-score: 0.02) (Figure 32). The fixed-effect model yielded little differences (Sensitivity analysis).

In subnetwork 2, where IFN alone was the comparator treatment, we did not observe between-study heterogeneity (Q = 0.47, df = 1, P = 0.50; I^2 = 0%, Tau^2 = 0.0). The ranking of treatments suggested that PAZ alone was the best treatment option (P-score: 0.94), whereas IFN alone was the worst option (P-score: 0.10).

Adverse events

Adverse events (AEs) were not consistently reported across trials. To be able to meta-analyse results, we could only consider AEs when the number of participants with at least one all-cause event of grade 3 or 4 was reported. We did not consider cumulated events or treatment-related AEs. All-cause AEs were assessed in a total of 18 trials (NCT00065468; NCT00081614; NCT00719264; NCT00720941; NCT00732914; NCT01024920; NCT01613846; NCT01835158; NCT01984242; NCT02420821; NCT02684006;NCT02811861; NCT03141177; NCT00738530; NCT00920816; NCT01030783; NCT01108445; NCT01274273) (15 two-arm trials, three three-arm trials), for a total of 8423 participants. One of these trials did not report individual AEs, meaning we could only extract data for the total number of participants with at least one grade 3 or 4 AE (NCT01984242). Another five trials did not report the total number of participants with at least one grade 3 or 4 AE, meaning we could only extract data for individual grade 3 or 4 AEs (NCT00738530; NCT00920816; NCT01030783; NCT01108445; NCT01274273). In the three-arm trials, only two comparisons (arm A versus arm C and arm B versus arm C) were reported, so we manually added a third comparison (arm A versus arm B).

Analysis of all-cause AEs (grade 3 or 4)

We conducted a combined analysis of all-cause grade 3 or 4 AEs in all risk groups combined from 13 trials (NCT00065468; NCT00081614; NCT00719264; NCT00720941; NCT00732914; NCT01024920; NCT01613846; NCT01835158; NCT01984242; NCT02420821; NCT02684006; NCT02811861; NCT03141177), for a total of 6909 participants. Thereof, three three-arms trials were included, each presenting three pairwise comparisons (NCT00065468; NCT01984242; NCT02811861). In one trial, AEs occurring in >2% of participants were reported; in another trial the frequency was >3%; in three trials >10% of participants; in four trials >20% of participants; in one trial AEs in >25% of the participants; for three trials, the frequency was not reported. Figure 77 in Appendix 15 outlines the available direct evidence (19 pairwise comparisons). The network was not fully connected and consisted of four sub-networks (Figure 33). We conducted network meta-analysis for subnetwork 1. Sub-networks 2, 3 and 4 contained only one trial, so no further analyses were conducted. Results for all network comparisons, including the ranking of treatments, are shown in Table 17 and Figure 34.

Figure 33. Network graph for all-cause AEs (grades 3-4; all risk groups combined). Any two treatments are connected by a line when there is at least one trial comparing the two treatments. Line width: number of trials. Plot width: number of participants.

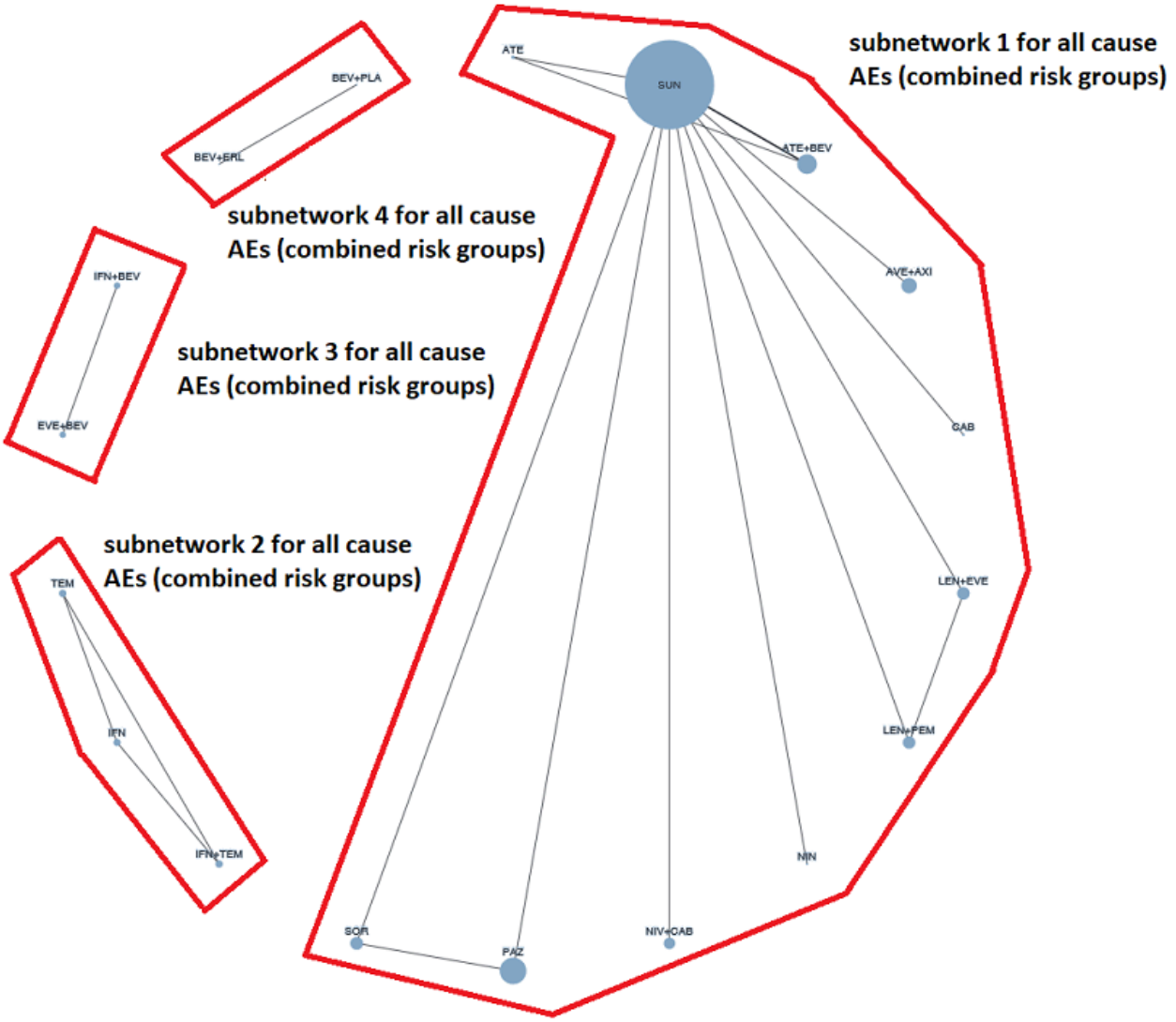
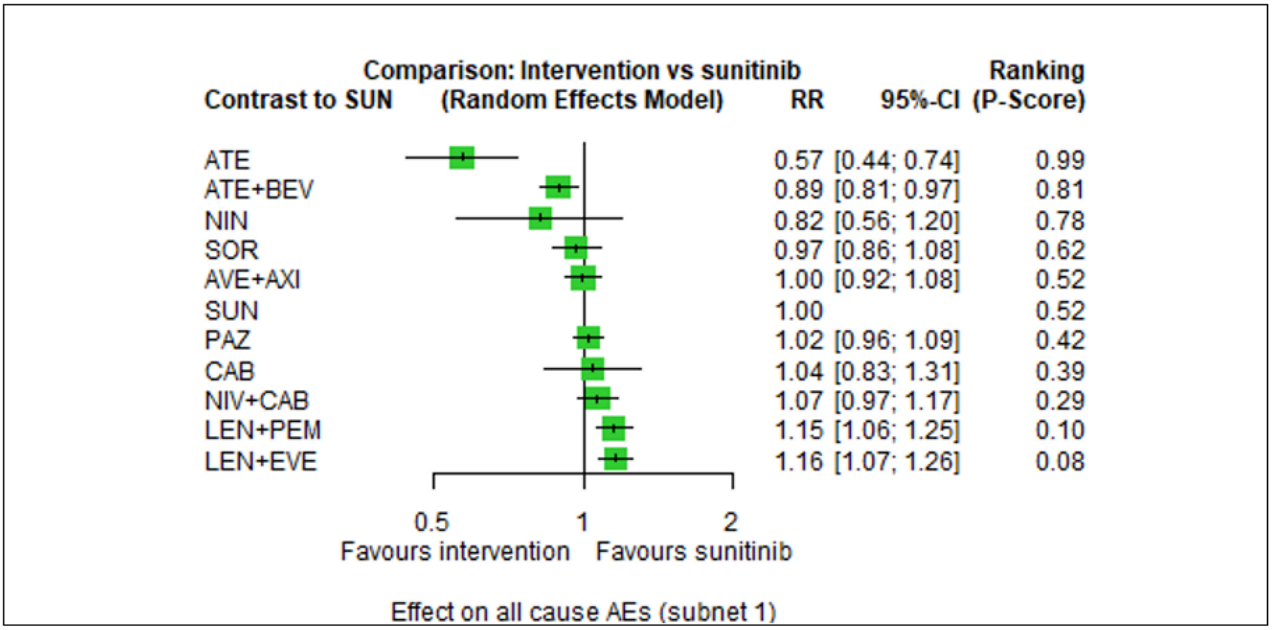


Figure 34. Forest plot for all-cause AEs (grades 3-4; all risk groups combined). Subnet 1. Reference treatment: sunitinib (SUN). Treatments are ordered by P-score (descending).



In sub-network 1, we did not observe between-study heterogeneity (Q_{total} = 0.31, df = 2, P = 0.85; Q_{within} = 0.0, df = 0, P = n.a.; $Q_{between}$ = 0.31, df=2, P = 0.85; I^2 = 0.0%, Tau^2 = 0.0). The combination LEN +PEM probably increases slightly the risk for AEs (RR 1.15, 95% CI 1.06 to 1.25, moderate certainty, P-score: 0.10) when compared to SUN alone (P-score: 0.52). We found that there probably is little or no difference in the risk for AEs between AVE+AXI (RR 1.00, 95% CI 0.92 to 1.08, moderate certainty, P-score: 0.52) and NIV+CAB (RR 1.07, 95% CI 0.97 to 1.17, moderate certainty, P-score: 0.29), when compared to SUN alone, respectively. We are uncertain whether CAB alone reduces or increases the risk for AEs (HR 1.04, 95% CI 0.83 to 1.31, very low certainty, P-score: 0.39), when compared to SUN alone. There is probably little or no difference in the risk for AEs between PAZ alone (RR 1.02, 95% CI 0.96 to 1.09, moderate certainty, P-score: 0.42) and SUN alone. Comparison data were not available for PEM+AXI and NIV+IPI. In the ranking of treatments, ATE alone was the best treatment option (P-score: 0.99), and LEN+EVE was the worst option(P-score: 0.08) (Figure 34).

As shown in Figure 33, there are closed loops in this network. The forest plot of splitting direct and indirect evidence is depicted in Figure 78 in Appendix 15. There was no significant difference between direct and indirect estimates (P = 0.7148) for ATE alone versus SUN alone and ATE alone versus ATE+BEV; P = 0.6702 for PAZ alone versus SOR alone, PAZ alone versus SUN alone and SOR alone versus SUN alone (data not shown)). The net heat plot showed no signs for inconsistency (Figure 79 in Appendix 15).

Analyses of individual AEs

Hand-food syndrome

Hand-food syndrome was assessed in 10 trials (NCT00720941; NCT00920816; NCT01024920; NCT01030783; NCT01108445; NCT01613846; NCT01835158; NCT02684006; NCT02811861; NCT03141177) (nine two-arm trials, one three-arm trial), for a total of 5029 participants. One two-arm trial reported zero events and was therefore excluded from the analyses (NCT01024920). Hence, analyses were conducted with 4933 participants. Figure 80 in Appendix 15 outlines the available direct evidence (11 pairwise comparisons). The network was fully connected (Figure 35). Results for all network comparisons, including the ranking of treatments, is shown in Figure 36. Heterogeneity statistics could not be calculated because each pairwise comparison was reported by a single trial only.

Figure 35. Network graph for the AE hand-foot-syndrome (all risk groups combined). Any two treatments are connected by a line when there is at least one trial comparing the two treatments. Line width: number of trials. Plot width: number of participants.

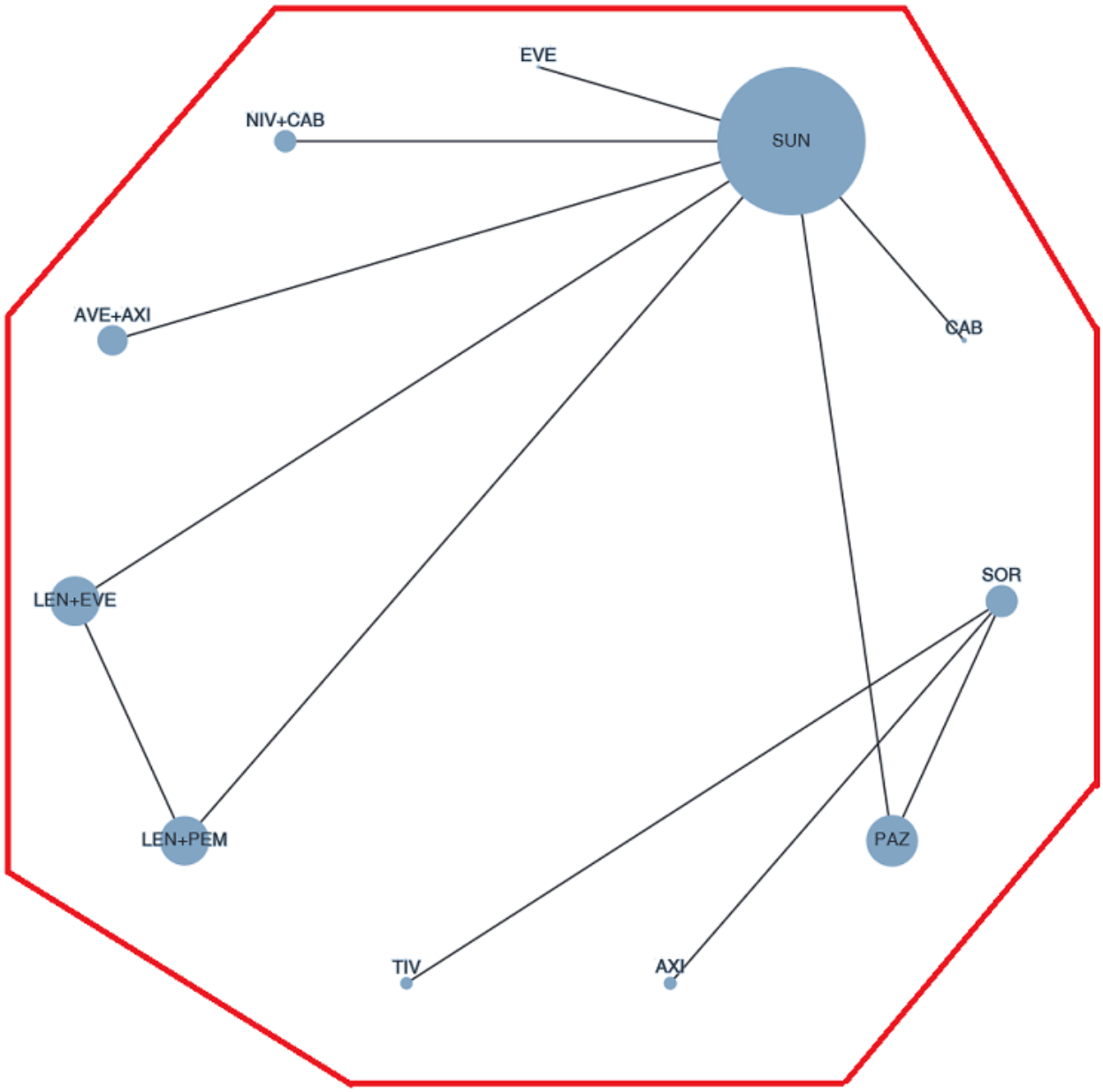
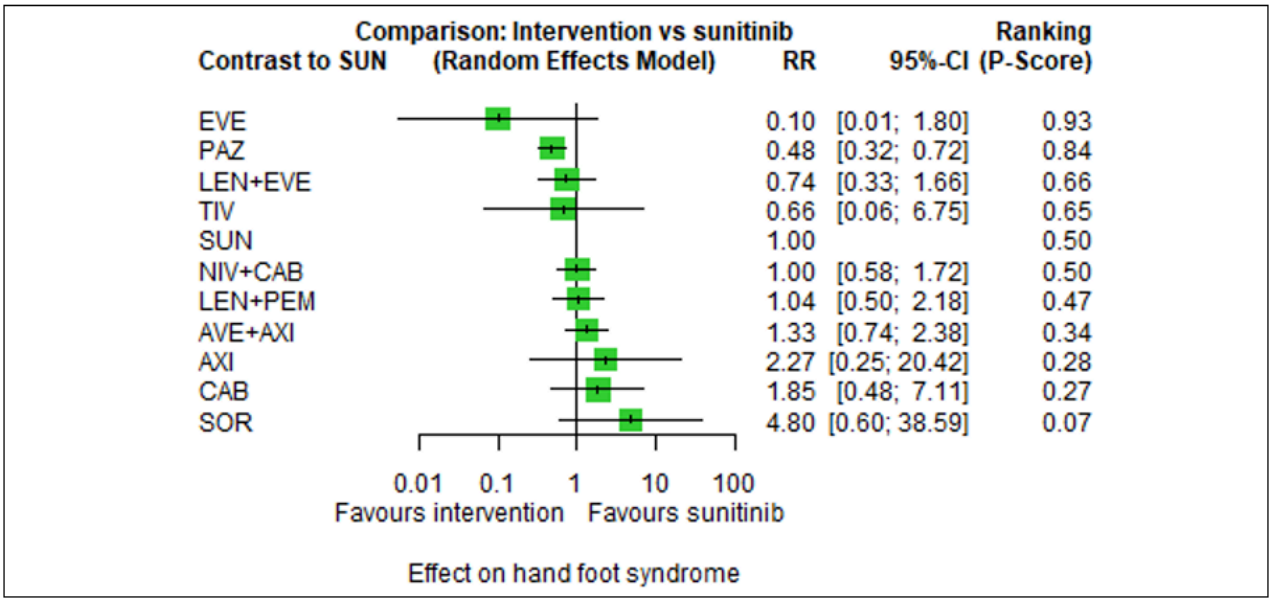


Figure 36. Forest plot for the AE hand-foot-syndrome (all risk groups combined). Reference treatment in the network: sunitinib (SUN). Treatments are ordered by P-score (descending).



The evidence suggests a substantially smaller risk with PAZ alone (RR 0.48, 95% CI 0.32 to 0.72, P-score: 0.84) when compared to SUN alone (P-score: 0.50). The combinations NIV+CAB (RR 1.00, 95% CI 0.58 to 1.72, P-score: 0.50), LEN+PEM (RR 1.04, 95% CI 0.50 to 2.18, P-score: 0.47), AVE+AXI (RR 1.33, 95% CI 0.74 to 2.38, P-score: 0.34) and CAB alone (RR 1.85, 95% CI 0.48 to 7.11, P-score: 0.27) reduce or increase the risk for hand-foot syndrome, when compared to SUN alone, respectively. Comparison data were not available for PEM +AXI and NIV+IPI. In the ranking of treatments, EVE alone was the best treatment option (P-score: 0.93) and SOR alone was the worst treatment option (P-score: 0.07) (Figure 36).

Fatigue

Fatigue was assessed in 14 trials (NCT00719264; NCT00720941; NCT00732914; NCT00738530; NCT00920816; NCT01024920; NCT01030783; NCT01108445; NCT01274273; NCT01613846; NCT01835158; NCT02684006; NCT02811861; NCT03141177) (13 two-arm trials, one three-arm trial), for a total of 6502 participants. Figure 81 in Appendix 15 outlines the available direct evidence (16 pairwise comparisons). The network was not fully connected and consisted of three sub-networks (Figure 37). We conducted analyses for subnets 1 and 2; subnetwork 3 contained only one trial, so no further analyses were conducted. Results for all network comparisons, including the ranking of treatments, is shown in Figure 38.

Figure 37. Network graph for the AE fatigue (all risk groups combined). Any two treatments are connected by a line when there is at least one trial comparing the two treatments. Line width: number of trials. Plot width: number of participants.

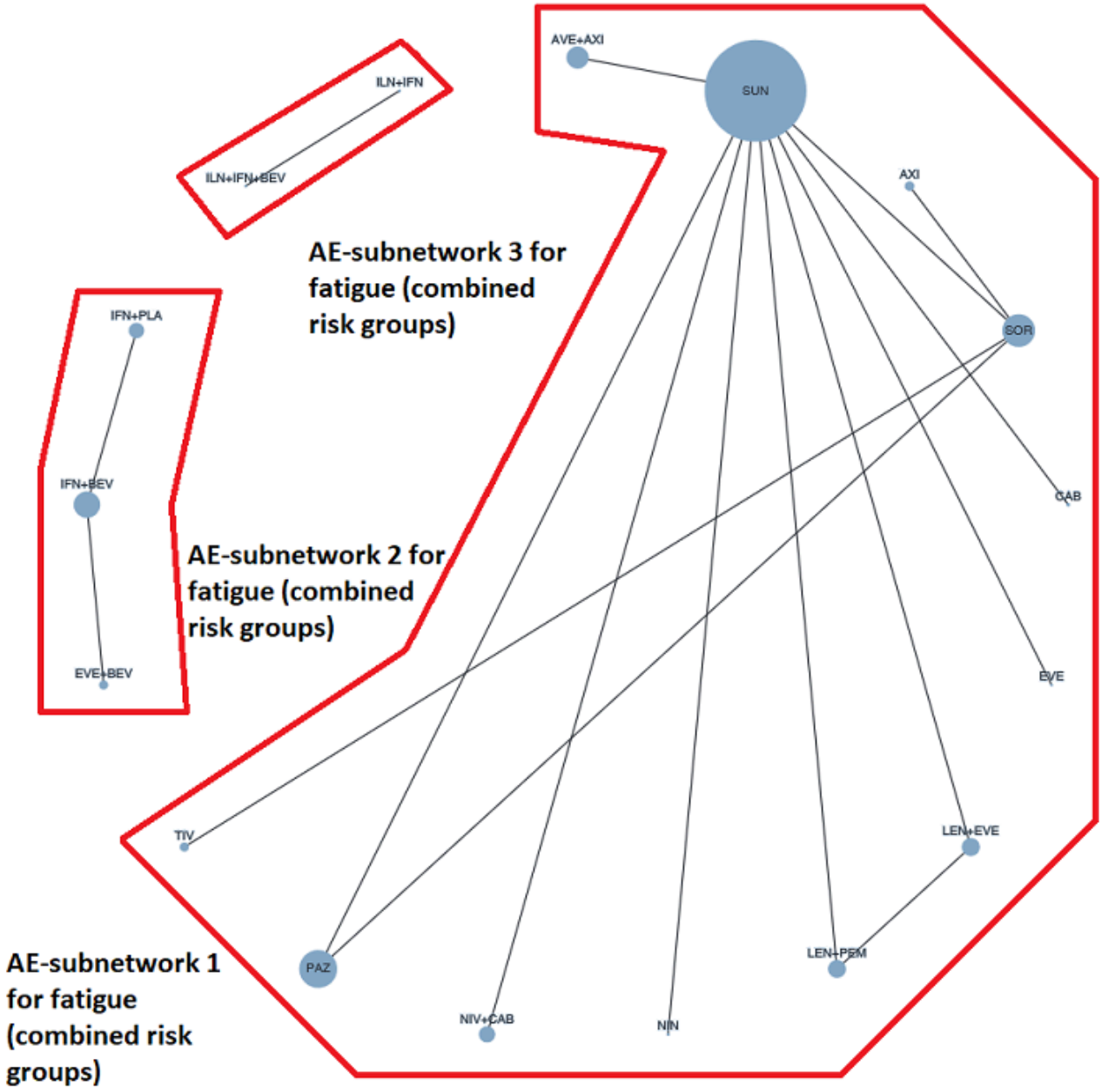
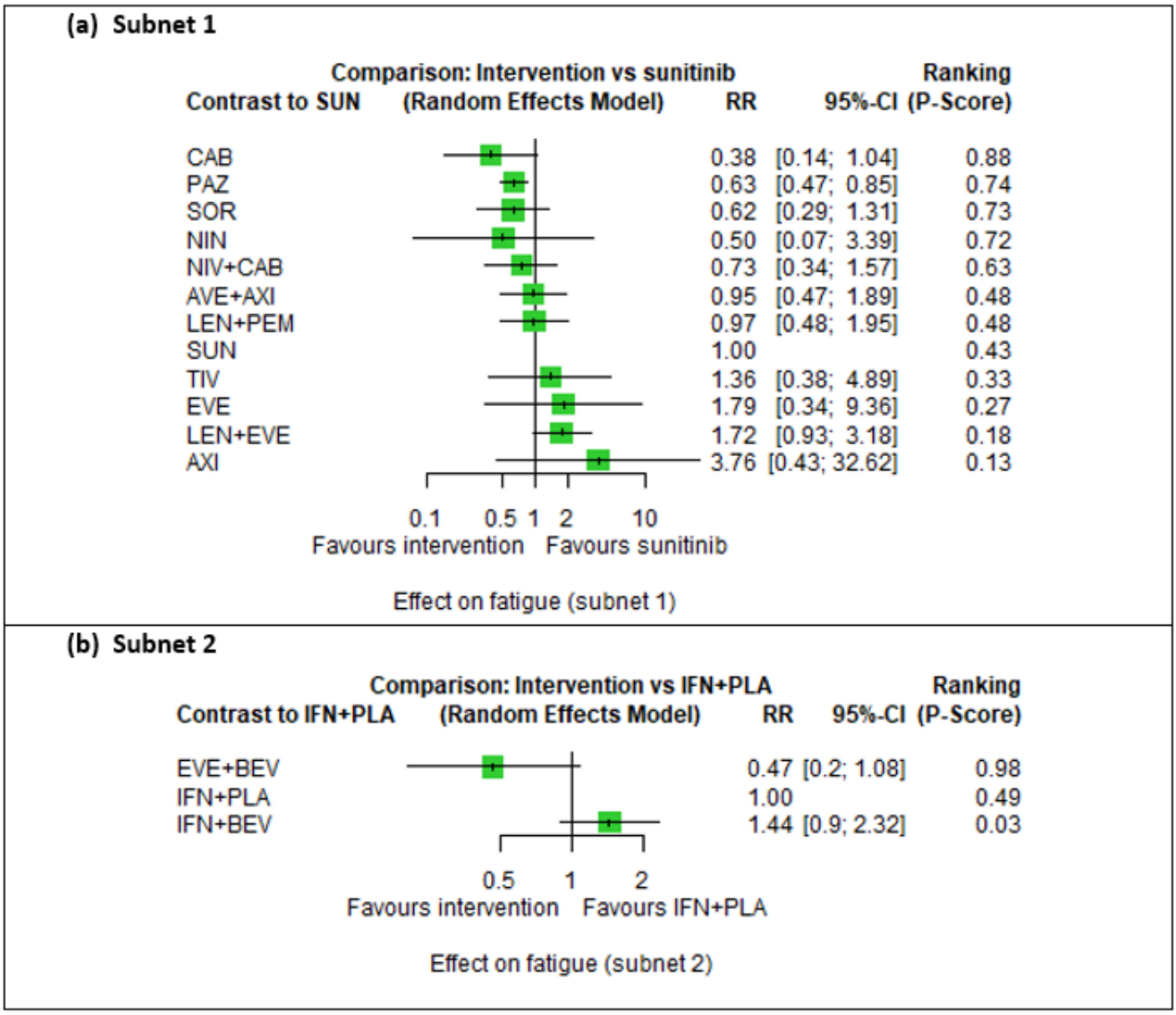


Figure 38. Forest plot for the AE fatigue (all risk groups combined). a) Subnetwork 1. Reference treatment: sunitinib (SUN). b) Subnetwork 2. Reference treatment: interferon-alpha + placebo (IFN+PLA). Treatments are ordered by P-score (descending).



In sub-network 1, we did not observe between-study heterogeneity (Q_{total}=0.0, df = 1, P = 0.97; Q_{within}=0.0, df = 0, P = n.a.; $Q_{between}$=0.0, df=1, P = 0.97; I^2 = 0.0%, Tau^2 = 0.0). Cabozantinib alone reduces or increases the risk for fatigue (RR 0.38, 95% CI 0.14 to 1.04, P-score: 0.88) when compared to SUN alone (P-score: 0.43). The evidence suggests a substantially smaller risk with PAZ alone (RR 0.63, 95% CI 0.47 to 0.85, P-score: 0.74) compared to SUN alone. The combinations NIV+CAB (RR 0.73, 95% CI 0.34 to 1.57, P-score: 0.63), AVE+AXI (RR 0.95, 95% CI 0.47 to 1.89, P-score: 0.48), LEN+PEM (RR 0.97, 95% CI 0.48 to 1.95, P-score: 0.48) decrease or increase the risk for fatigue, when compared to SUN alone, respectively. Comparison data were not available for PEM+AXI and NIV+IPI. In the ranking of treatments, CAB alone was the best treatment option (P-score: 0.88) and AXI alone was the worst treatment option (P-score: 0.13) (Figure 38).

In sub-network 2, heterogeneity statistics could not be calculated because each pairwise comparison was reported by a single trial only. Here, IFN+PLA was the comparator treatment, and the results suggested a lower risk for fatigue with EVE+BEV compared to IFN +PLA (RR 0.47, 95% CI 0.2 to 1.08). In the ranking of treatments, EVE +BEV was the best treatment option (P-score: 0.98) and IFN+BEV was the worst option (P-score: 0.03).

Diarrhoea

Diarrhoea was assessed in 16 trials (NCT00081614; NCT00719264; NCT00720941; NCT00732914; NCT00738530; NCT00920816; NCT01024920; NCT01030783; NCT01108445; NCT01274273; NCT01613846; NCT01835158; NCT02684006; NCT02811861; NCT03141177; NCT00065468) (15 two-arm trials, two three-arm trials), for a total of 7222 participants. Figure 82 in Appendix 15 outlines the available direct evidence (20 pairwise comparisons). The network was not fully connected and consisted of five sub-networks (Figure 39). We conducted analyses for sub-networks 1 and 2; sub-networks 3, 4 and 5 contained only one trial each, so no further analyses were conducted. Results for all

network comparisons, including the ranking of treatments, is shown in Figure 40.

Figure 39. Network graph for the AE diarrhoea (all risk groups combined). Any two treatments are connected by a line when there is at least one trial comparing the two treatments. Line width: number of trials. Plot width: number of participants.

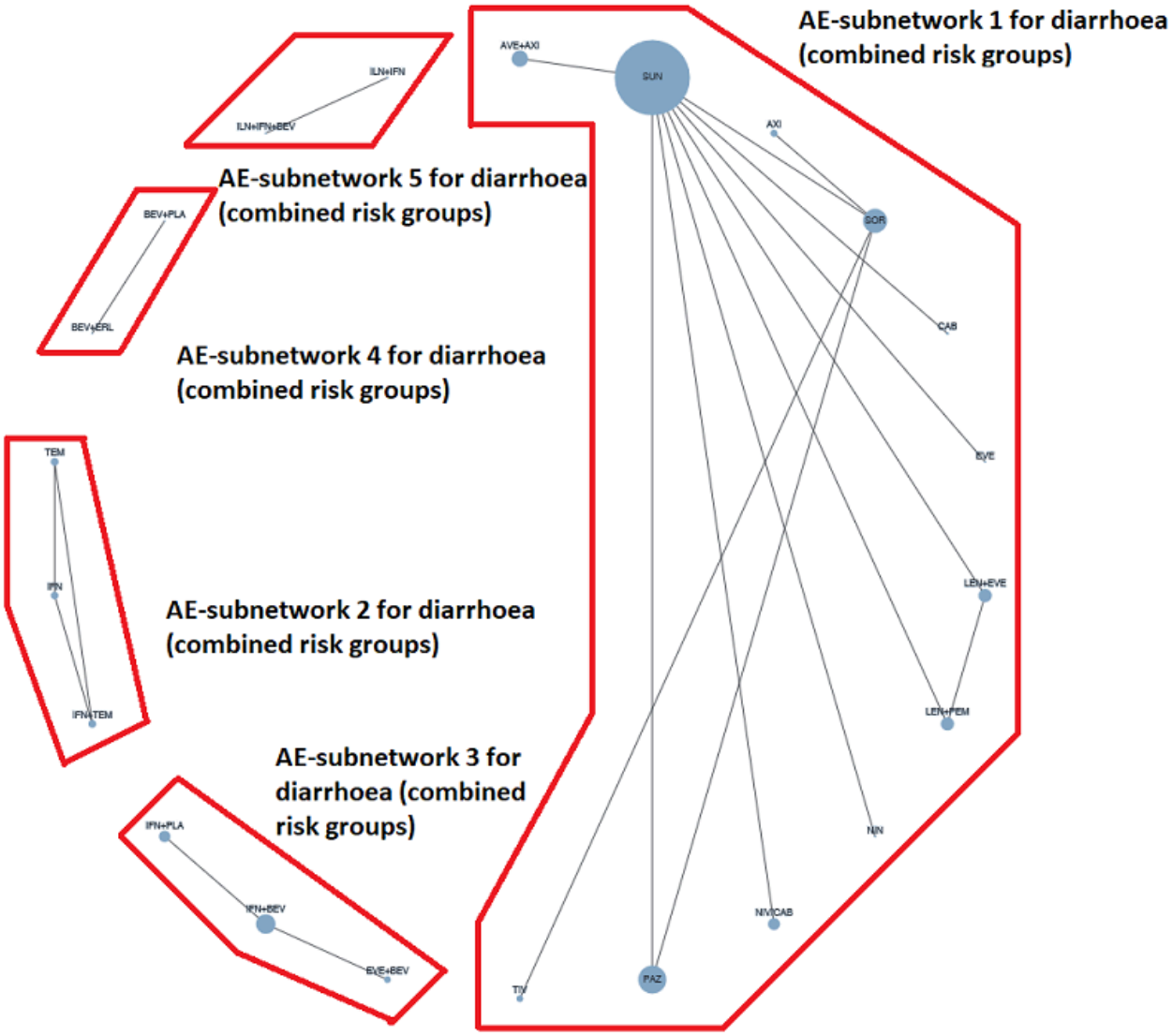
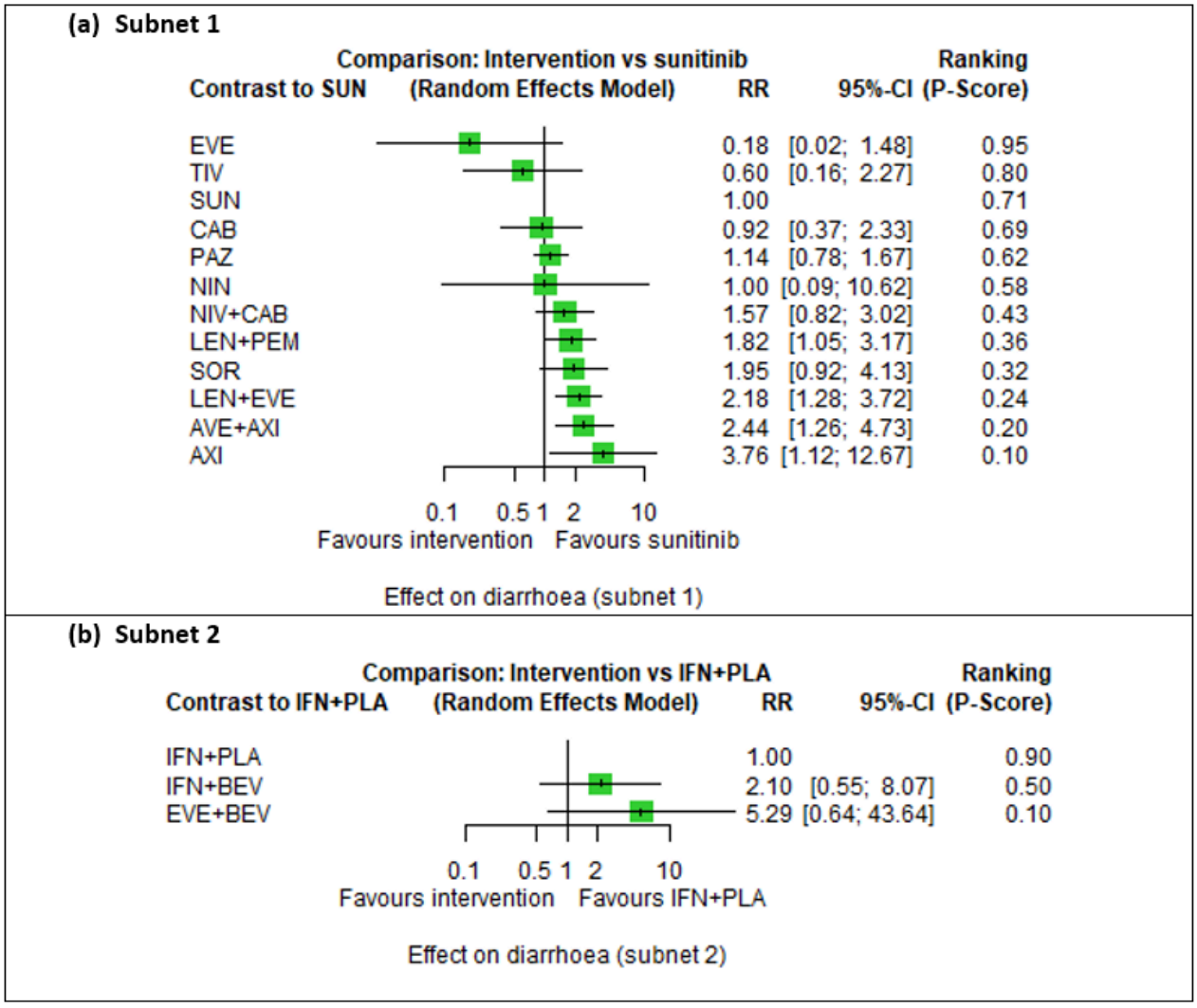


Figure 40. Forest plot for the AE diarrhoea (all risk groups combined). a) Subnetwork 1. Reference treatment: sunitinib (SUN); b) Subnetwork 2. Reference treatment: interferon-alpha + placebo (IFN+PLA). Treatments are ordered by P-score (descending).



In sub-network 1, we did not observe between-study heterogeneity (Q_{total} = 0.05, df=1, P = 0.83; Q_{within} = 0.0, df = 0, P = n.a. ; $Q_{between}$ = 0.05, df = 1, P = 0.83; I^2 = 0.0%, Tau^2 = 0.0). Cabozantinib alone (RR 0.92, 95% CI 0.37 to 2.33, P-score: 0.69), PAZ alone (RR 1.14, 95% CI 0.78 to 1.67, P-score: 0.62) and NIV+CAB (RR 1.57, 95% CI 0.82 to 3.02, P-score: 0.43) reduce or increase the risk for diarrhoea, when compared to SUN alone (P-score: 0.71), respectively. The evidence suggests that LEN+PEM (RR 1.82, 95% CI 1.05 to 3.17, P-score: 0.36) and AVE+AXI (RR 2.44, 95% CI 1.26 to 4.73, P-score: 0.20) substantially increase the risk for diarrhoea, when compared to SUN, respectively. Comparison data were not available for PEM +AXI and NIV+IPI. In the ranking of treatments, EVE alone was the best treatment option (P-score: 0.95) and AXI alone was the worst treatment option (P-score: 0.10) (Figure 40).

As shown in Figure 39, there are closed loops in the network. The forest plot of splitting direct and indirect evidence is depicted in Figure 83 in Appendix 15. There was no significant difference between direct and indirect estimates (P = 0.8272 (data not shown)). The net heat plot showed no signs for inconsistency (Figure 84 in Appendix 15).

In sub-network 2, heterogeneity statistics could not be calculated because each pairwise comparison was reported by a single trial only. Here, IFN+PLA was the comparator treatment, and the ranking of treatments suggested that IFN+PLA was the best treatment option (P-score: 0.90) and EVE+BEV the worst option (P-score: 0.10).

Vomiting

Vomiting was assessed in 10 trials (NCT00720941; NCT00920816; NCT01024920; NCT01108445; NCT01274273; NCT01835158; NCT02684006; NCT02811861; NCT03141177; NCT00065468) (eight two-arm trials, two three-arm trials), for a total of 5035 participants. Figure 85 in Appendix 15 outlines the available direct evidence (14 pairwise comparisons). The network was not fully connected and consisted of four subnets (Figure 41). We conducted analyses for subnetwork 1; sub-networks 2, 3 and 4 contained only one trial, so no further analyses were conducted.

Results for all network comparisons, including the ranking of treatments, is shown in Figure 42.

Figure 41. Network graph for the AE vomiting (all risk groups combined). Any two treatments are connected by a line when there is at least one trial comparing the two treatments. Line width: number of trials. Plot width: number of participants.

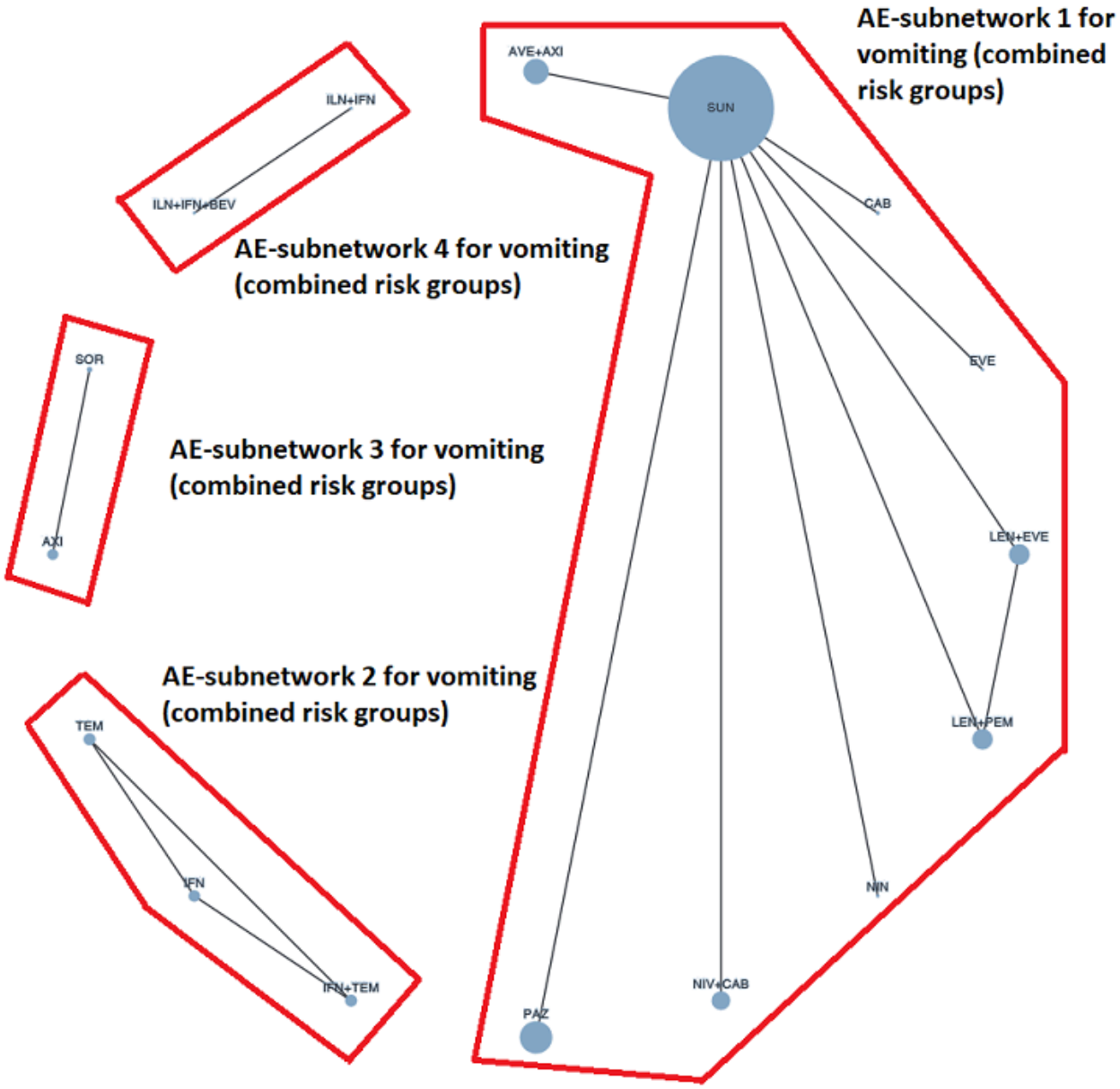
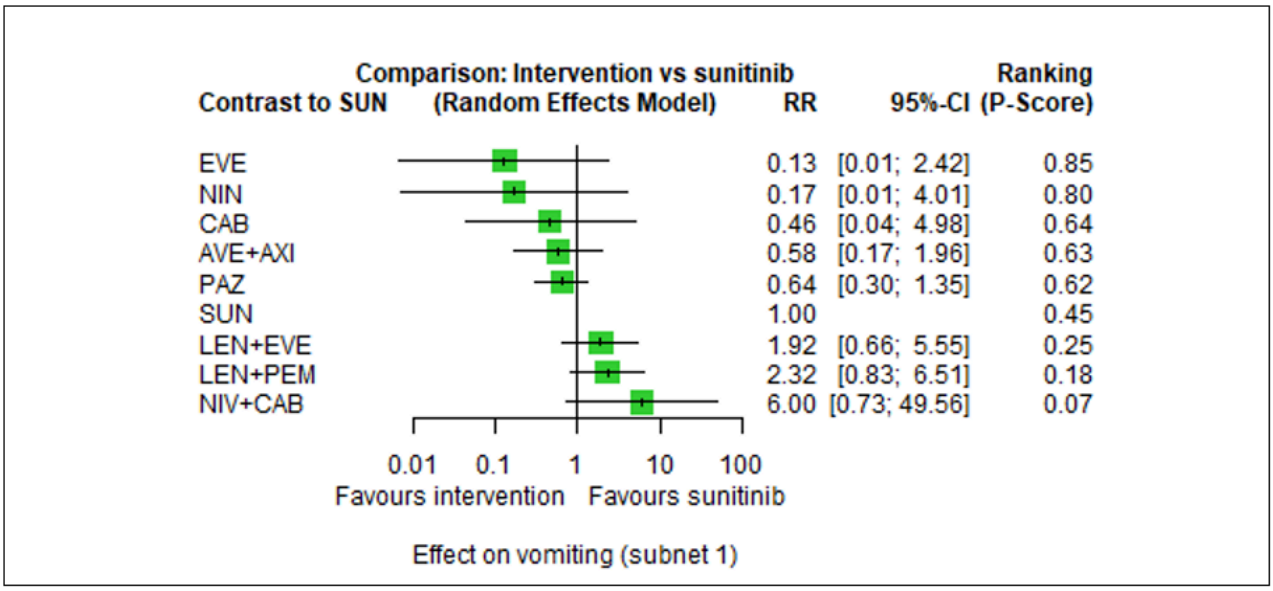


Figure 42. Forest plot for the AE vomiting (all risk groups combined). Subnetwork 1. Reference treatment: sunitinib (SUN). Treatments are ordered by P-score (descending).



In sub-network 1, each comparison was reported by a single trial only, so no heterogeneity statistics could be calculated. We found that CAB alone (RR 0.46, 95% CI 0.04 to 4.98, P-score: 0.64), AVE+AXI (RR 0.58, 95% CI 0.17 to 1.96, P-score: 0.63), PAZ alone (RR 0.64, 95% CI 0.30 to 1.35, P-score. 0.62), LEN+PEM (RR 2.32, 95% CI 0.83 to 6.51, P-score: 0.18) and NIV+CAB (RR 6.00, 95% CI 0.73 to 49.56, P-score: 0.07) decrease or increase the risk for vomiting, when compared to SUN alone (P-score: 0.45), respectively. Comparison data were not available for PEM+AXI and NIV+IPI. In the ranking of treatments, EVE alone was the best treatment option (P-score: 0.85) and NIV+CAB was the worst treatment option (P-score: 0.07) (Figure 42).

Loss of appetite

Loss of appetite was assessed in 11 trials (NCT00719264; NCT00720941; NCT00732914; NCT00920816; NCT01024920; NCT01108445; NCT01613846; NCT01835158; NCT02684006; NCT02811861; NCT03141177) (10 two-arm trials, one three-arm trial), for a total of 5381 participants. Figure 86 in Appendix 15 outlines the available direct evidence (13 pairwise comparisons). However, one trial (NCT01024920) was excluded from analyses because zero events were reported. Hence, data were analysed for 5285 participants. The network was not fully connected and consisted of two subnets (Figure 43). We conducted analyses for subnetwork 1; subnetwork 2 contained only one trial, so no further analyses were conducted. Results for all network comparisons, including the ranking of treatments, is shown in Figure 44.

Figure 43. Network graph for the AE loss of appetite (all risk groups combined). Any two treatments are connected by a line when there is at least one trial comparing the two treatments. Line width: number of trials. Plot width: number of participants.

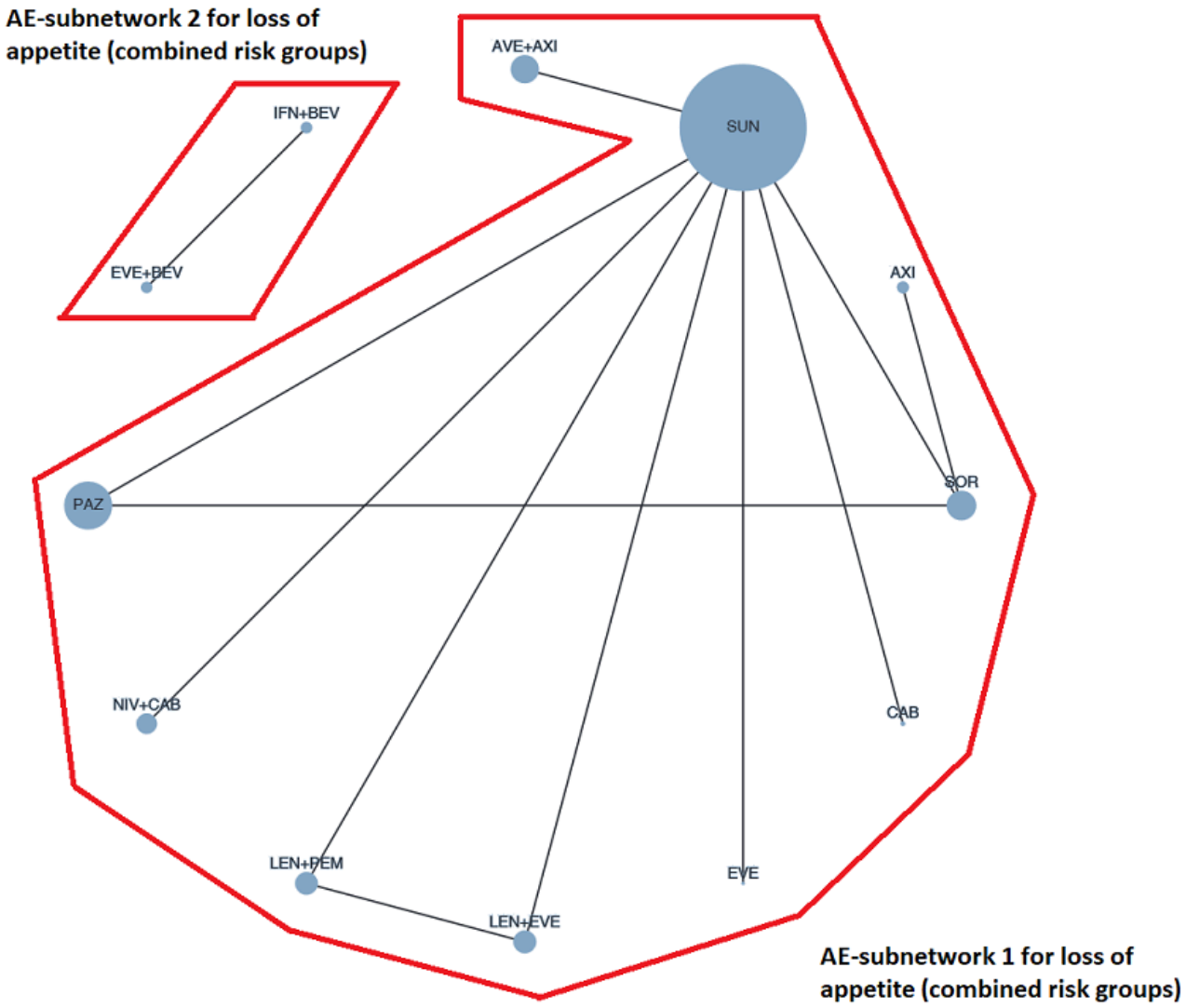
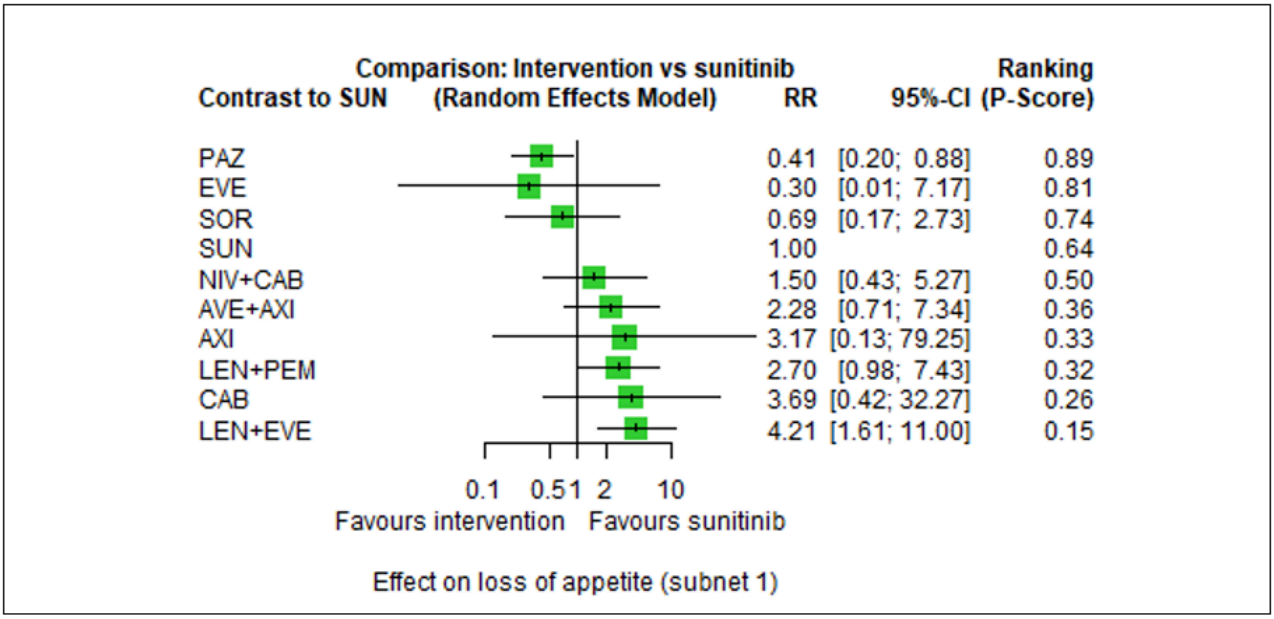


Figure 44. Forest plot for the AE loss of appetite (all risk groups combined). Subnetwork 1. Reference treatment: sunitinib (SUN). Treatments are ordered by P-score (descending).



In sub-network 1, we did not observe between-study heterogeneity (Q_{total} = 0.44, df = 1, P = 0.51; Q_{within} = 0.0, df = 0, P = n.a.; $Q_{between}$ = 0.44, df = 1, P = 0.51; I^2 = 0.0%, Tau^2 = 0.0). The evidence suggests a lower risk for loss of appetite with PAZ alone (RR 0.41, 95% CI 0.20 to 0.88) compared to SUN alone (P-score: 0.64). We found that NIV +CAB (RR 1.50, 95% CI 0.43 to 5.27, P-score: 0.50), AVE+AXI (RR 2.28, 95% CI 0.71 to 7.34, P-score: 0.36), LEN+PEM (RR 2.70, 95% CI 0.98 to 7.43, P-score: 0.32), and CAB alone (RR 3.69, 95% CI 0.42 to 32.27, P-score: 0.26) reduce or increase the risk for loss of appetite, when compared to SUN alone, respectively. WComparison data were not available for PEM+AXI and NIV+IPI. In the ranking of treatments, PAZ alone was the best treatment option (P-score: 0.89), whereas LEN +EVE was the worst treatment option (P-score: 0.15) (Figure 44).

As shown in Figure 43, there are closed loops in the network. Figure 87 in Appendix 15 depicts the forest plot of splitting direct and indirect evidence. There was no significant difference between direct and indirect estimates (P = 0.5071 (data not shown)). The net heat plot showed negligible signs for inconsistency (Figure 88 in Appendix 15).

Weight loss

Weight loss was assessed in 12 trials (NCT00719264; NCT00720941; NCT00920816; NCT01024920; NCT01108445; NCT01274273; NCT01613846; NCT01835158; NCT02684006; NCT02811861; NCT03141177; NCT00065468) (10 two-arm trials, two three-arm trials), for a total of 5762 participants. Figure 89 in Appendix 15 outlines the available direct evidence (16 pairwise comparisons). However, one trial (NCT01108445) was excluded from analyses because zero events were reported. Hence, data were analysed for 5654 participants. The network was not fully connected and consisted of four sub-networks (Figure 45). We conducted analyses for subnetwork 1; sub-networks 2, 3 and 4 contained only one trial, so no further analyses were conducted. Results for all network comparisons, including the ranking of treatments, is shown in Figure 46.

Figure 45. Network graph for the AE weight loss (all risk groups combined). Any two treatments are connected by a line when there is at least one trial comparing the two treatments.

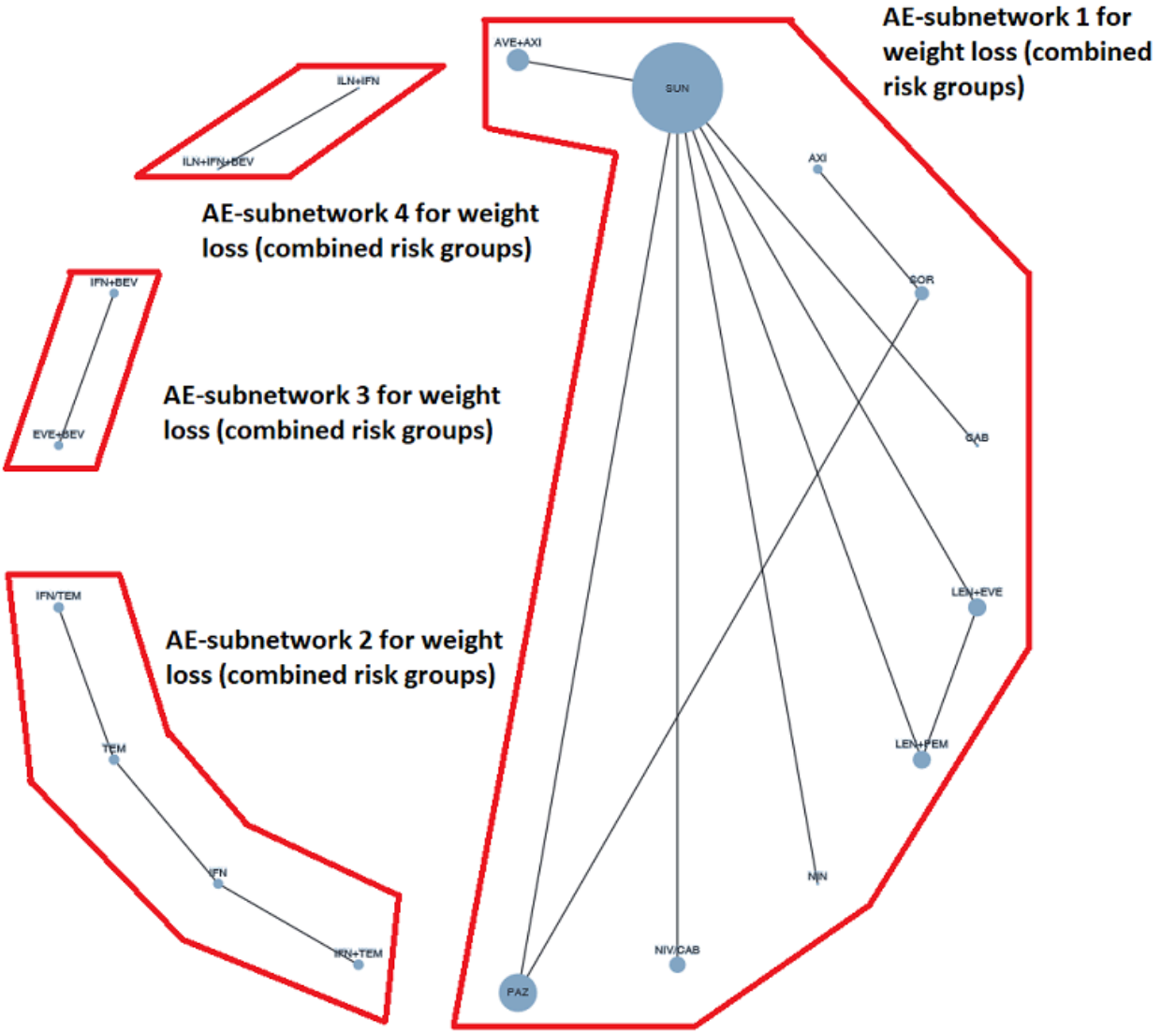
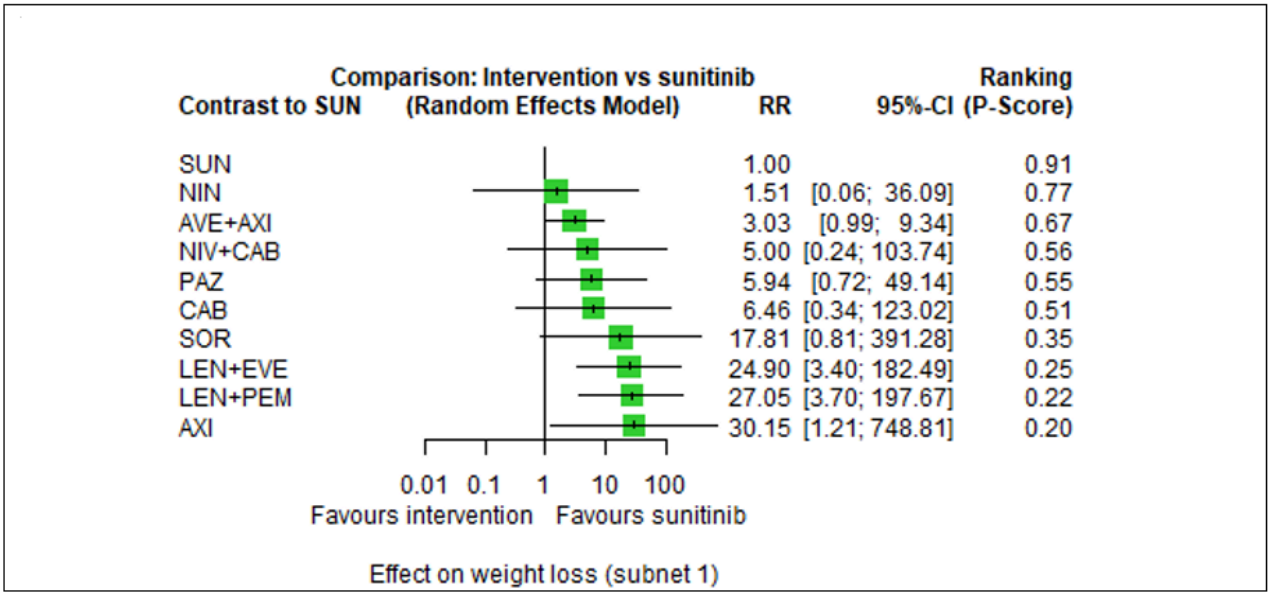


Figure 46. Forest plot for the AE weight loss (all risk groups combined). Subnetwork 1. Reference treatment: sunitinib (SUN). Treatments are ordered by P-score (descending).



In sub-network 1, heterogeneity statistics could not be calculated because each comparison was reported by single trial only. The evidence showed that no treatment had a lower risk for weight loss compared to SUN alone. In the ranking of treatments, SUN alone was the best treatment option (P-score: 0.91) and AXI alone was the worst treatment option (P-score: 0.20) (Figure 46). The risk with SUN alone was substantially lower when compared to LEN+PEM (RR 0.04, 95% CI 0.01 to 0.27, P-score: 0.22). We found that NIV+CAB (RR 5.00, 95% CI 0.24 to 103.74, P-score: 0.56), AVE+AXI (RR 3.03, 95% CI 0.99 to 9.34, P-score: 0.67), and CAB alone (RR 6.46, 95% CI 0.34 to 123.02, P-score: 0.51) reduce or increase the risk for weight loss, when compared to SUN alone, respectively. Comparison data were not available for PEM+AXI and NIV+IPI.

Stomatitis

Stomatitis was assessed in 12 trials (NCT00719264; NCT00720941; NCT00732914; NCT01024920; NCT01108445; NCT01274273; NCT01613846; NCT01835158; NCT02684006; NCT02811861; NCT03141177; NCT00065468) (10 two-arm trials, two three-arm trials), for a total of 5830 participants. Figure 90 in Appendix 15 outlines the available direct evidence (16 pairwise comparisons). However, one trial (NCT01274273) was excluded from analyses because zero events were reported. Hence, data were analysed for 5712 participants. The network was not fully connected and consisted of three sub-networks (Figure 47). We conducted analyses for subnetwork 1; sub-networks 2 and 3 contained only one trial, so no further analyses were conducted. Results for all network comparisons, including the ranking of treatments, is shown in Figure 48.

Figure 47. Network graph for the AE stomatitis (all risk groups combined). Any two treatments are connected by a line when there is at least one trial comparing the two treatments.

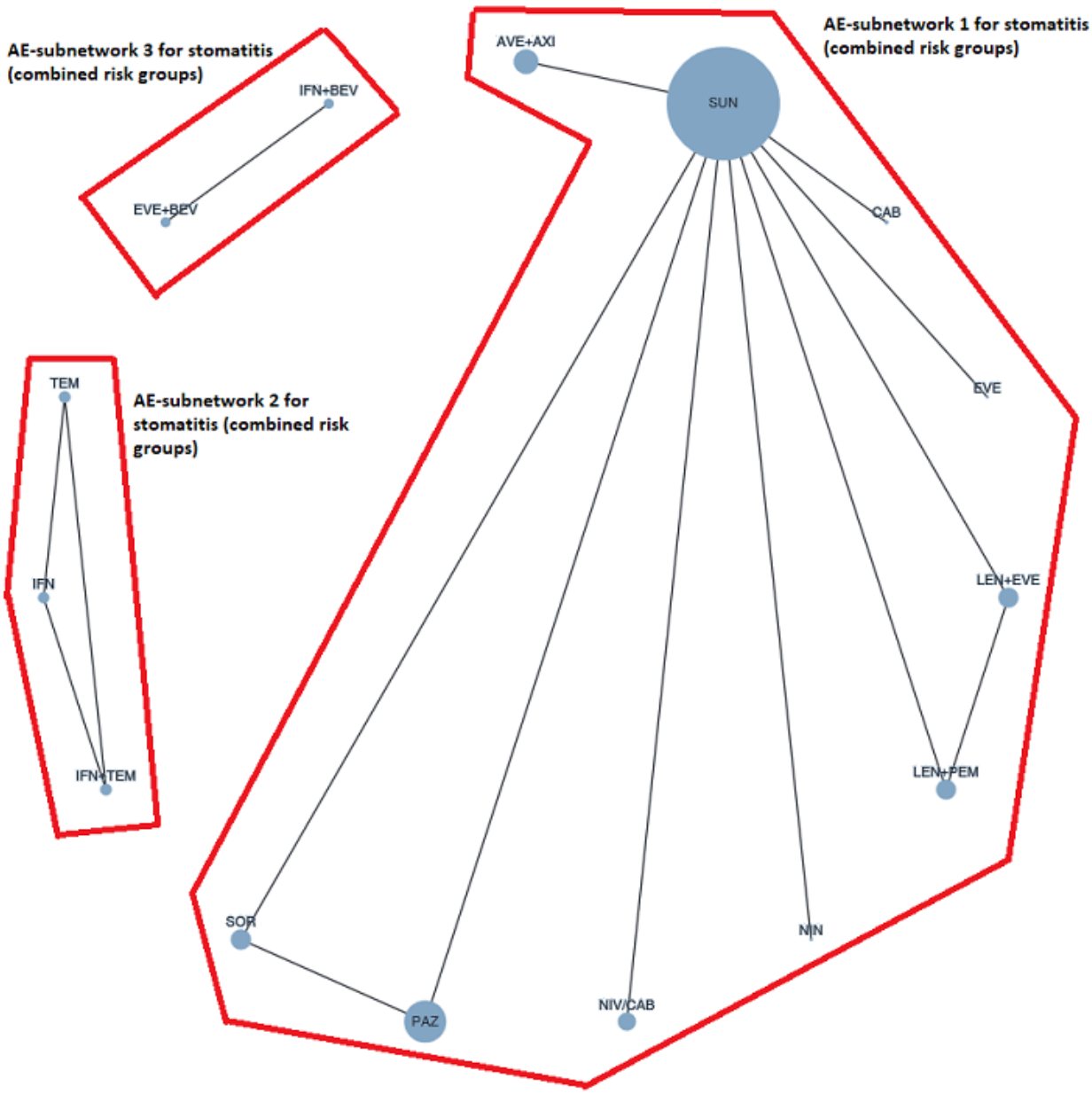
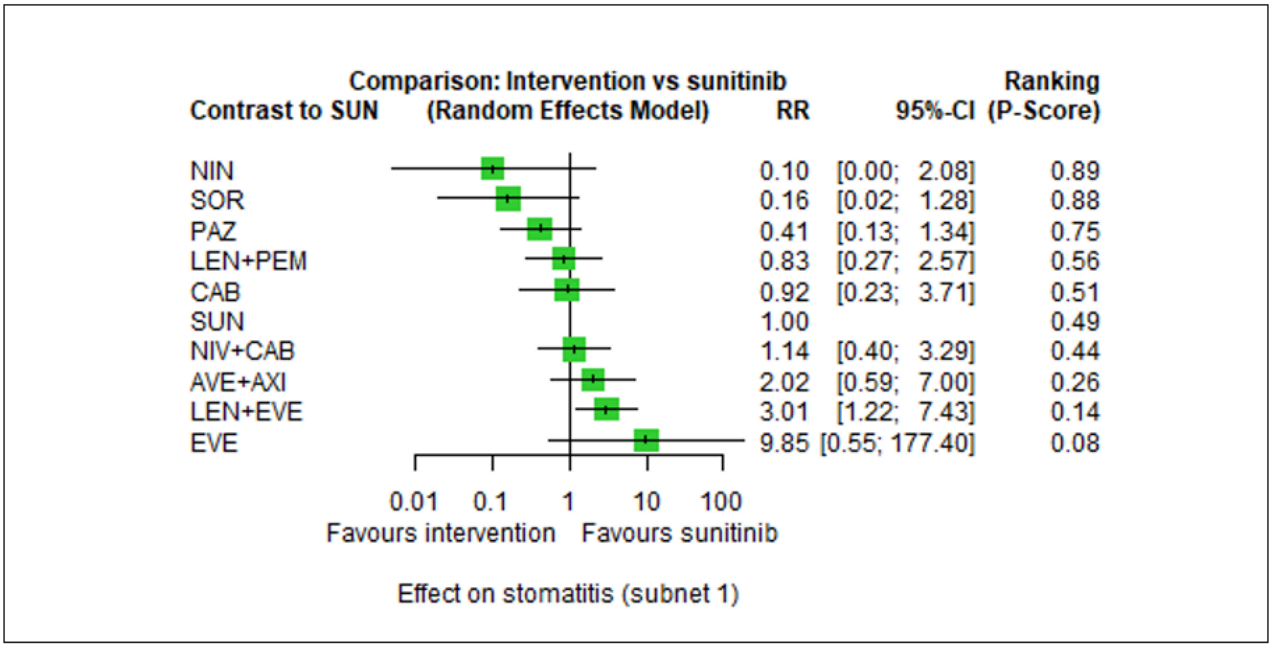


Figure 48. Forest plot for the AE stomatitis (all risk groups combined). Subnetwork 1. Reference treatment: sunitinib (SUN). Treatments are ordered by P-score (descending).



In sub-network 1, we did not observe important between-study heterogeneity (Q_{total} = 1.02, df = 1, P = 0.31; Q_{within} = 0.0, df = 0, P = n.a.; $Q_{between}$ = 1.02, df = 1, P = 0.31; I^2 = 2%, Tau^2 = 0.0302). We found that LEN+PEM (RR 0.83, 95% CI 0.27 to 2.57, P-score: 0.56), NIV+CAB (RR 1.14, 95% CI 0.40 to 3.29, P-score: 0.44), AVE+AXI (RR 2.02, 95% CI 0.59 to 7.00, P-score: 0.26), PAZ alone (RR 0.41, 95% CI 0.13 to 1.34, P-score: 0.75) and CAB alone (RR 0.92, 95% CI 0.23 to 3.71, P-score: 0.51) reduce or increase the risk for stomatitis, when compared to SUN alone (P-score: 0.49), respectively. Comparison data were not available for PEM+AXI and NIV+IPI. In the ranking of treatments, NIN alone (P-score: 0.89) was the best treatment option, and EVE alone was the worst treatment option (P-score: 0.08) (Figure 48).

As shown in Figure 47, there are closed loops in the network. Figure 91 in Appendix 15 depicts the forest plot of splitting direct and indirect evidence. There was no significant difference between direct and indirect estimates (P = 0.3173 (data not shown)). The net heat plot showed negligible signs for inconsistency (Figure 92 in Appendix 15).

Mucosal inflammation

Mucosal inflammation was assessed in four two-arm trials (NCT00720941; NCT01108445; NCT02684006; NCT03141177), for a total of 2723 participants. Figure 93 in Appendix 15 outlines the available direct evidence (four pairwise comparisons). The network was fully connected (Figure 49); results for all network comparisons, including the ranking of treatments, is shown in Figure 50. Because each comparison in this network was reported by a single trial only, heterogeneity statistics could not be calculated.

Figure 49. Network graph for the AE mucosal inflammation (all risk groups combined). Any two treatments are connected by a line when there is at least one trial comparing the two treatments.

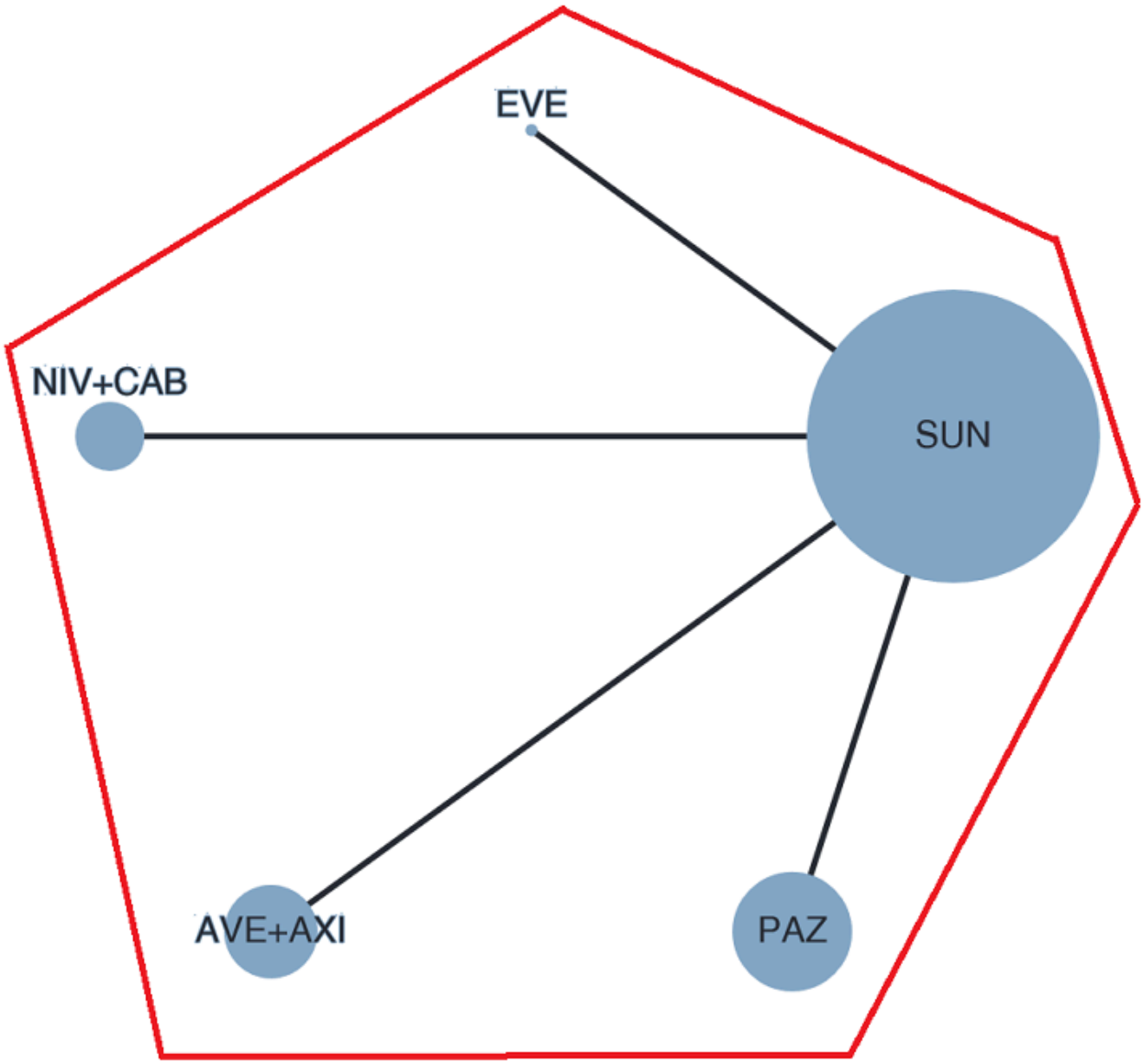
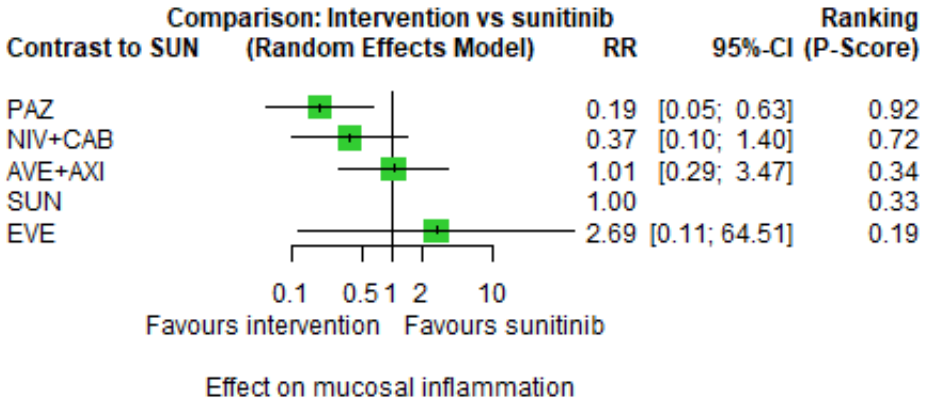


Figure 50. Forest plot for the AE mucosal inflammation (all risk groups combined). Reference treatment: sunitinib (SUN). Treatments are ordered by P-score (descending).



The evidence suggests a substantially lower risk for mucosal inflammation with PAZ alone (RR 0.19, 95% CI 0.05 to 0.63, P-score: 0.92) compared to SUN alone (P-score: 0.33). The combinations NIV+CAB (RR 0.37, 95% CI 0.10 to 1.40, P-score: 0.72) and AVE+AXI (RR 1.01, 95% CI 0.29 to 3.47, P-score: 0.34) reduce or increase the risk for mucosal inflammation, when compared to SUN alone, respectively. Comparison data were not available for PEM+AXI, NIV +IPI, LEN+PEM, and CAB alone. In the ranking of treatments, PAZ alone was the best treatment option (P-score: 0.92) and EVE alone was the worst option (P-score: 0.19) (Figure 50).

Insomnia

Insomnia was assessed in two two-arm trials (NCT00720941; NCT01108445), for a total of 1210 participants. In NCT00720941, participants in the experimental arm received PAZ alone and in NCT01108445, participants in the experimental arm received EVE alone. In both trials, SUN alone was the comparator treatment. We were not able to analyse results as the trials reported null events (i.e. no participant in any arm had an event).

Depression

Depression was assessed in two two-arm trials (NCT00738530; NCT01274273), for a total of 759 participants. We were not able to analyse results as the two trials were not connected in the network. In NCT00738530, IFN+BEV (experimental arm, N=337) versus IFN +PLA (control arm, N = 304) were compared. Ten participants in the experimental arm and four participants in the control arm had at least one grade 3 or 4 event. In NCT01274273, ILN+IFN+BEV (experimental arm, N = 59) versus ILN+IFN (control arm, N = 59) were compared. It was reported that no participant in the experimental arm had an event, whereas one participant in the control arm had at least one event of grade 3 or 4.

Number of participants who discontinued study treatment due to an adverse effect (AE)

The number of participants who discontinued study treatment due to an AE was assessed for 30 trials (NCT00920816; NCT00979966; NCT01024920; NCT01108445; NCT01613846; Jonasch 2010; NCT00065468; NCT00072046; NCT00081614; NCT00098657/NCT00083889; NCT00117637; NCT00609401; NCT00619268; NCT00631371; NCT00719264; NCT00720941; NCT00732914; NCT00903175; NCT01030783; NCT01274273; NCT01392183; NCT01481870; NCT01835158; NCT01984242; NCT02231749; NCT02420821; NCT02684006; NCT02811861; NCT02853331; NCT03141177) (27 two-arm trials, three three-arm trials), for a total of 13,110 participants. Figure 94 in Appendix 15 outlines the available direct evidence (36 comparisons). The network was not fully connected and consisted of three sub-networks (Figure 51). We conducted network meta-analysis for subnetwork 1. Sub-networks 2 and 3 contained only one trial each, so no further analyses were conducted. Results for all network comparisons, including the ranking of treatments, are shown in Table 18 and Figure 52.

Figure 51. Network graph for the outcome Number of participants who discontinued treatment due to an AE (combined risk groups). Any two treatments are connected by a line when there is at least one trial comparing the two treatments.

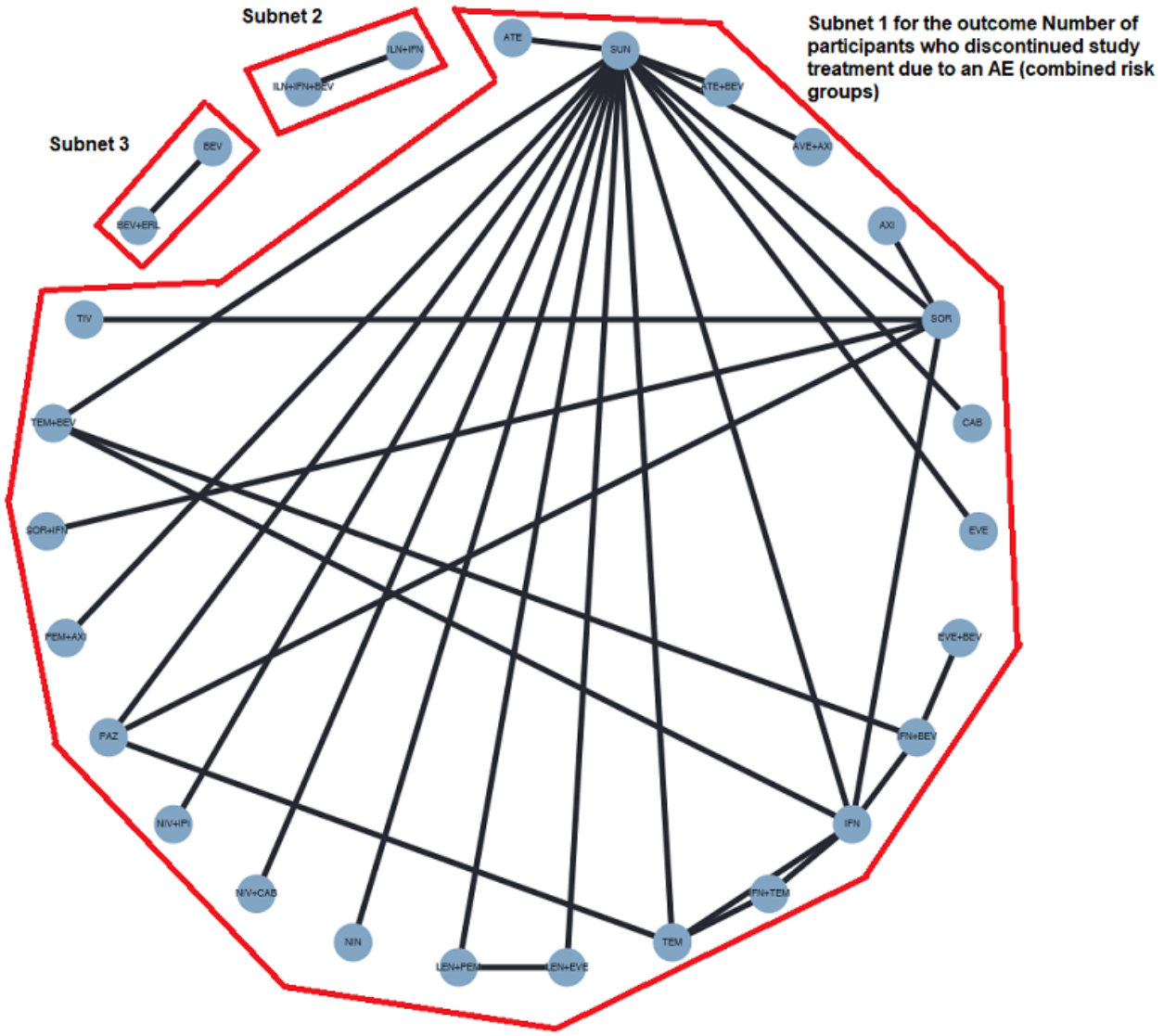
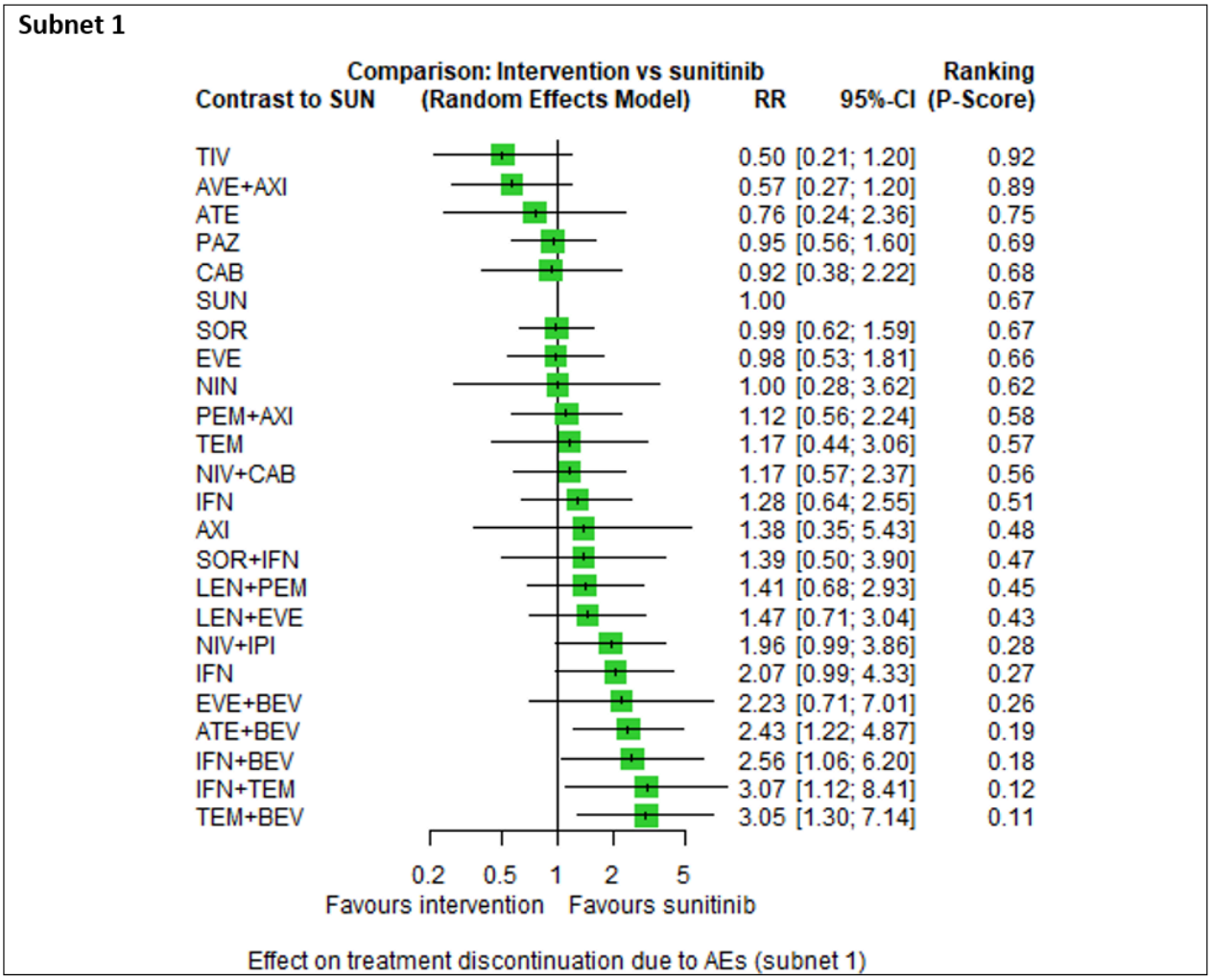


Figure 52. Forest plot for the outcome Number of participants who discontinued treatment due to an AE (combined risk groups). Any two treatments are connected by a line when there is at least one trial comparing the two treatments.



For sub-network 1, we observed moderate heterogeneity (Q_{total} = 18.14, df = 8, P = 0.020; Q_{within} = 3.27, df = 3, P = 0.35; $Q_{between}$ = 14.87, df = 5, P = 0.011; I^2 = 55.9%, Tau^2 = 0.1029). The only closed loops in this subnet contained of multi-arm trials, so inconsistency could not be checked. The evidence suggests that AVE+AXI (P-score: 0.89), PAZ alone (P-score: 0.69), CAB alone (P-score: 0.68), PEM +AXI (P-score: 0.58), NIV+CAB (P-score: 0.56), LEN+PEM (P-score: 0.45), and NIV+IPI (P-score: 0.28) decrease or increase the risk for study discontinuation due to an AE, when compared to SUN alone, respectively. For this outcome, the best treatment option was TIV alone (P-score: 0.92) and the worst option was TEM+BEV (P-score: 0.11).

Time to initiation of the first subsequent anticancer therapy

None of the included trials reported the outcome TFST as a time-to-event outcome. However, 19 trials (NCT00081614; NCT00098657/NCT00083889; NCT00609401; NCT00619268; NCT00719264; NCT00732914; NCT00738530; NCT00920816; NCT01024920; NCT01108445; NCT01274273; NCT01835158; NCT01984242; NCT02231749; NCT02420821; NCT02684006; NCT02811861; NCT02853331; NCT03141177) reported the number of participants who received some subsequent anticancer therapy after discontinuing study treatment. In one trial (NCT00072046) there were discrepancies in the numbers reported in texts and tables, so we refrained from reporting the results. However, reporting of the outcome was very heterogenous: different definitions were provided; different types of therapy reported (subsequent therapy could include only systemic therapy or systemic therapy, radiotherapy or other); participants may have received more than one subsequent therapy. Furthermore, the timing of reporting was not reported in the trials. Hence, we refrained from pooling data and reported results narratively for all available interventions and comparisons in Table 19.

As for the interventions that we chose for the Summary of findings 1, we found that PEM+AXI (RR 0.72, 95% CI 0.64 to 0.81, low certainty), AVE+AXI (RR 0.61, 95% CI 0.52 to 0.72, low certainty) and NIV+IPI (RR 0.86, 95% CI 0.79 to 0.94, low certainty) may reduce the risk for subsequent therapy, when compared to SUN alone, respectively. We are uncertain whether NIV+CAB (RR 0.57, 95% CI 0.44 to 0.75, very low certainty) and LEN+PEM (RR 0.57, 95% CI 0.48

to 0.68, very low certainty) reduce the risk for subsequent therapy, when compared to SUN alone, respectively. Lastly, we are uncertain whether CAB alone reduces or increases the risk for subsequent therapy (RR 0.93, 95% CI 0.74 to 1.16, very low certainty), when compared to SUN. Comparison data were not available for PAZ alone compared to SUN alone.

Subgroup analyses

We were able to conduct only one subgroup analysis for one outcome. Other subgroup analyses were not possible due to the distribution of characteristics across trials or due to a lack of available subgroup data (see Differences between protocol and review).

Follow-up time

We conducted a subgroup analysis for the outcome OS in the combined risk groups (total trial population) for the follow-up times <5 years and ≥ 5 years. For most trials, follow-up time was estimated from the Kaplan-Meier-Curves of the respective effect estimates for OS. Out of 21 trials that were included in the analysis of OS in the total trial population, 14 trials had a follow-up time for this outcome of <5 years (Jonasch 2010; NCT00081614; NCT00098657/NCT00083889; NCT00334282; NCT00631371; NCT00719264; NCT00738530; NCT00920816; NCT01024920; NCT01108445; NCT01984242; NCT02420821; NCT02761057; NCT02811861), while the remaining seven trials had a follow-up time of ≥ 5 years (NCT00072046; NCT00420888; NCT00609401; NCT00720941; NCT00979966; NCT02231749; NCT02853331),

In the analysis of trials with a follow-up of less than five years, the network consisted of three sub-networks. In subnetwork 1, which included SUN as our main comparator, we did not find notable effects between interventions. We observed moderate heterogeneity in this subnetwork (Q=1.81, df = 1, P = 0.18; I^2 = 44.6%, Tau^2 = 0.0284). In the ranking of treatments, LEN+PEM was the best treatment option (P-score: 0.89) and IFN alone the worst (P-score: 0.25). These results are similar to those of the main analyses: LEN +PEM was also rated the best treatment option, whereas IFN alone was the second-worst treatment option. Results for all network comparisons, including the ranking of treatments, are shown in Figure 95 in Appendix 16.

In the analysis of trials with a follow-up time of five or more years, the network consisted of two sub-networks. In subnetwork 1, we found evidence suggesting that OS was higher with NIV+IPI (HR 0.69, 95% CI 0.59 to 0.81) and PEM+AXI (HR 0.73, 95% CI 0.60 to 0.81) when compared to SUN alone, respectively. In the ranking of treatments, NIV+IPI was the best treatment option (P-score: 0.85) and SUN alone was the worst (P-score: 0.15). Results for all network comparisons, including the ranking of treatments, are shown in Figure 96 in Appendix 16.

Sensitivity analyses

Sensitivity analyses were not possible for every outcome that was planned for (see Differences between protocol and review).

Fixed-effects

We conducted fixed-effect NMA for the outcomes OS, SAE and PFS.

For sub-network 1 of the outcome OS in all risk groups combined, the fixed-effect model yielded somewhat different results (Figure 97 in Appendix 16). The results from the fixed-effect model suggested substantially better OS with LEN+PEM, NIV+IPI and PEM+AXI when compared to SUN alone, respectively, and the confidence intervals were more narrow. However, there were no changes in the direction of the effect and there were no changes in the ranking of these three treatments according to their P-score. Furthermore, in the fixed-effect model, ATE+BEV was favoured over SUN alone (HR 0.95, 95% CI 0.80 to 1.12), as opposed to the random-effects model, where SUN alone was favoured over ATE+BEV; hence, the ranking of these two treatments changed according to their P-score.

For the outcome SAEs (all risk groups combined), the fixed-effect model yielded somewhat different results (Figure 98 in Appendix 16). Firstly, the ranking and order of treatments slightly changed. Secondly, the direction of effect for the comparison ATE alone versus SUN alone changed, as ATE alone was favoured over SUN alone (HR 0.99, 95% CI 0.71 to 1.36) in the fixed-effect model. Furthermore, the results suggested a substantially lower risk for SAE with SUN alone when compared to PEM+AXI, LEN+EVE and NIV +IPI, respectively (the direction of effects remained unchanged).

For sub-network 1 of the outcome PFS in the MSKCC favourable risk group, the fixed-effect model yielded different results (Figure 99 Appendix 16). Here, AXI alone was favoured over SUN alone (HR 0.95, 95% CI 0.52 to 1.74) as opposed to the result of the random-effects model, where SUN alone was favoured over AXI alone; hence, the ranking of these two treatments changed. Furthermore, results suggested substantially better PFS with LEN +PEM and LEN+EVE when compared to SUN alone, respectively, and the confidence intervals were more narrow. However, there were no changes in the direction of effects. For subnetwork 1 of the outcome PFS in the MSKCC intermediate and poor risk groups, the fixed-effect model yielded somewhat different results (Figure 100 in Appendix 16). Results suggested substantially better PFS with LEN+EVE versus SUN alone. For subnetwork 2, the fixed-effect model yielded slightly different results (Figure 101 in Appendix 16). The fixed-effect model for subnetwork 1 of the outcome PFS in the IMDC intermediate and poor risk groups yielded only little differences as well (Figure 102 in Appendix 16).

Assumption of proportional hazards

We conducted this sensitivity analysis for the outcome OS (all risk groups combined). Five trials reported that they tested the assumption of proportional hazards, but only four reported that the assumption was also validated (Jonasch 2010; NCT00609401; NCT00920816; NCT02420821). In NCT00334282, the assumption was not validated.

For this sensitivity analysis, the network consisted of two sub-networks. An analysis was conducted for sub-network 1; sub-network 2 contained only one trial, so no further analyses were conducted. For sub-network 1, the main comparator was SOR alone, and we did not find notable effects between interventions. In the ranking of treatments, the best treatment option was SOR+ILN (P-score: 0.74), whereas SOR+IFN was the worst option (P-score: 0.06) (Figure 103 in Appendix 16).

Risk of bias

We conducted sensitivity analyses according to the risk of bias ('low risk of bias' or 'some concerns' versus 'high risk of bias') in the outcomes OS and SAEs in the combined risk groups.

For the outcome OS (all risk groups combined), which included 21 trials, five trials had an overall 'low risk of bias' (NCT00334282; NCT00720941; NCT02231749; NCT02420821; NCT02853331), 12 trials had 'some concerns' (Jonasch 2010; NCT00072046; NCT00081614; NCT00609401; NCT00631371; NCT00719264; NCT00738530; NCT00920816; NCT01024920; NCT01108445; NCT01984242; NCT02761057) and the remaining four trials had a 'high risk of bias' (NCT00098657/NCT00083889; NCT00420888; NCT00979966; NCT02811861). The network of trials at 'low risk of bias' or 'some concerns' consisted of four sub-networks; analyses were conducted for sub-networks 1, 2 and 3, while subnetwork 4 contained only one trial. For sub-network 1, moderate heterogeneity (Q=1.81, df=1, P = 0.18; I^2=44.6%, Tau^2=0.0284) was observed. We did not find notable effects between interventions in subnetwork 1. In the ranking of treatments, NIV+IPI was the best treatment option (P-score: 0.81) and EVE alone was the worst option (P-score: 0.33). Results for all network comparisons, including the ranking of treatments, are shown in Figure 104 in Appendix 16. As for the trials that were at 'high risk of bias', the network was fully connected. We found evidence suggesting substantially better OS with LEM+PEM (HR 0.66, 95% CI 0.49 to 0.88) compared to SUN alone (Figure 105 in Appendix 16). In the ranking of treatments, the best treatment option was LEN+PEM (P-score: 0.95), whereas NAP+IFN (P-score: 0.17) was the worst option.

For the outcome SAE (all risk groups combined), which included 22 trials, one trial had an overall 'low risk of bias' (NCT02853331), five trials had overall 'some concerns' (NCT00720941; NCT00920816; NCT00903175; NCT00126594; NCT01984242), and the remaining 16 trials had an overall 'high risk of bias' (NCT00065468; NCT00098657/NCT00083889; NCT00117637; NCT00619268; NCT00631371; NCT00719264; NCT00732914; NCT00738530; NCT00979966; NCT01024920; NCT01108445; NCT01613846; NCT01835158; NCT02231749; NCT02420821; NCT02811861). The network of trials at 'low risk of bias' or 'some concerns' consisted of two sub-networks. For sub-network 1, the evidence suggests no difference in the risk for SAE between PAZ alone and SUN alone (HR 0.99, 95% CI 0.87 to 1.14). Instead, we found evidence that suggested a substantially lower risk of SAE with SUN alone, when compared to ATE+BEV (RR 0.58, 95% CI 0.40 to 0.84) and PEM+AXI (RR 0.78, 95% CI 0.65 to 0.93), respectively. In the ranking of treatments, the best treatment option was PAZ alone (P-score: 0.81), closely followed by SUN alone (P-score: 0.80) and the worst treatment option was ATE+BEV (P-score: 0.03). Results for all network comparisons, including the ranking of treatments, are shown in Figure 106 in Appendix 16. As for the trials at 'high risk of bias', the network was fully connected. We did not find notable effects between interventions. In the ranking of treatments, TEM alone was the best treatment option (P-score: 0.89), whereas LEN+PEM was the worst treatment option (P-score: 0.17) (Figure 107 in Appendix 16).

DISCUSSION

Summary of main results

The primary objective of this systematic review with network meta-analysis was to evaluate and compare the benefits and harms of first-line therapies for adults with advanced renal cell carcinoma (RCC), and thereby produce a clinically relevant ranking of therapies. Secondary objectives were to maintain the currency of the evidence by using a living systematic review approach, as well as to incorporate data from clinical study reports (CSRs).

We identified a total of 55 eligible trials; of these, 36 randomised-controlled trials (RCTs) were included in quantitative analyses and narrative reporting in this review, with a total of 15,177 participants. In these trials, 22 drugs and 17 different combinations were assessed. The substance sunitinib (SUN) was the main comparator in this review, and also the main comparator in 16 included trials. All trials but one (NCT01064310) were included in the network meta-analyses. Overall risk of bias was mostly judged high across trials because most were open-label trials, hindering blinded outcome assessments. Reporting harms especially lacked details about the method of analysis and method of outcome measurement. The certainty in the evidence for all outcomes ranged from moderate to very low. The main outcomes and comparisons are presented in the Summary of findings 1, Summary of findings 2 and Summary of findings 3.

Primary outcomes

Overall survival (OS)

See Summary of findings 1, Summary of findings 2, and Summary of findings 3.

- **PEM+AXI:** We found that this combination probably improves OS, when compared to SUN alone, in the combined groups. However, we were not able to obtain subgroup data per risk group (neither for International Metastatic RCC Database Consortium (IMDC) nor Memorial Sloan Kettering Cancer Center (MSKCC criteria)).
- **AVE+AXI:** Subgroup data according to IMDC risk groups revealed that AVE+AXI may improve OS in the favourable, intermediate and poor risk groups, when compared to SUN alone. We were not able to obtain data for the risk groups according to MSKCC criteria, and neither for all risk groups combined.
- **NIV+CAB**: For the risk groups according to IMDC criteria, we are uncertain whether NIV+CAB improve or decrease OS in the favourable risk groups. However, we found that NIV+CAB probably improve OS in the intermediate and poor risk groups, when compared to SUN alone, respectively. We were not able to obtain data for the risk groups according to MSKCC criteria, and neither for all risk groups combined.
- **LEN+PEM:** We found that in the favourable risk group according to IMDC criteria, there may be little or no difference in OS between LEN+PEM and SUN in the favourable risk groups. As for the MSKCC favourable risk group, we are uncertain whether LEN+PEM improves OS. However, for both the IMDC and MSKCC intermediate and poor risk groups, we found that LEN+PEM probably improves OS, when compared to SUN alone. Looking at all risk groups combined, we found that LEN+PEM may improve OS.
- **NIV+IPI:** For the risk groups according to IMDC criteria, there probably is little or no difference in OS between NIV+IPI and SUN in the favourable risk groups, but NIV+IPI probably improve OS in the intermediate and poor risk groups. We were not able to obtain data for the risk groups according to MSKCC criteria. Looking at all risk groups combined, we found that NIV+IPI probably improve OS.
- **CAB:** We were not able to obtain subgroup data for the favourable risk groups (neither for IMDC nor MSKCC criteria), but

found that for the IMDC intermediate and poor risk groups, CAB alone may improve slightly OS, when compared to SUN alone. We were not able to obtain data for the MSKC intermediate and poor risk groups. Looking at all risk groups combined, we are uncertain whether CAB alone improves OS.

- **PAZ:** We found that there is probably little or no difference in OS between PAZ alone and SUN alone in the combined groups. We were not able to obtain subgroup data per risk group.

Quality of life (QoL)

See Summary of findings 1, Summary of findings 2, and Summary of findings 3.

We were not able to obtain subgroup data per risk groups for this outcome. Looking at the combined risk groups, we were also not able to obtain data on quality of life for PEM+AXI, AVE+AXI, NIV+CAB, LEN+PEM, NIV+IPI, and CAB alone. We obtained data for the comparison PAZ alone and SUN alone, where one RCT reported that the mean post-score of the intervention group (PAZ) was higher than that of the control group (SUN). However, we are uncertain about the evidence we found.

Serious adverse events (SAEs)

See Summary of findings 1, Summary of findings 2, and Summary of findings 3.

We were not able to obtain subgroup data per risk group for this outcome. Looking at the combined risk groups, we were also unable to obtain data on AVE+AXI and NIV+CAB. As for the other substances, we found that PEM+AXI probably increase slightly the risk for SAEs when compared to SUN. Furthermore, we found that LEN+PEM and NIV+IPI probably increase the risk for SAEs. There probably is little or no difference in the risk for SAEs between PAZ alone and SUN alone, and we are uncertain whether CAB alone reduces or increases the risk for SAEs, when compared to SUN alone.

Secondary outcomes

Progression-free survival (PFS)

See Summary of findings 1, Summary of findings 2, and Summary of findings 3.

- **PEM+AXI:** We found that this combination probably improves slightly PFS, when compared to SUN alone, in the combined groups. However, we were not able to obtain subgroup data per risk group.
- **AVE+AXI:** For the IMDC favourable risk group, we found that AVE+AXI may improve PFS, and for the IMDC intermediate and poor risk groups that AVE+AXI probably improve PFS, when compared to SUN, respectively. We were not able to obtain data for the risk groups according to MSKCC criteria, and neither for all risk groups combined.
- **NIV+CAB**: For the IMDC favourable risk group, NIV+CAB may improve PFS, and for the IMDC intermediate and poor risk groups, NIV+CAB probably improve PFS, when compared to SUN alone, respectively. We were not able to obtain data for the risk groups according to MSKCC criteria, and neither for all risk groups combined.
- **LEN+PEM:** We found that in the IMDC favourable risk groups, LEN+PEM may improve PFS, but we are uncertain whether LEN+PEM improve PFS in the MSKCC favourable risk groups, when compared to SUN alone, respectively. For the IMDC and MSKCC intermediate and poor risk groups, we found that LEN+PEM probably improve PFS, when compared to SUN alone. Looking at the combined risk groups, LEN+PEM probably improve PFS.
- **NIV+IPI:** For the IMDC favourable risk groups we found that NIV+IPI probably reduce PFS, when compared to SUN alone, but there may be little or no difference in PFS between NIV+IPI and SUN in the IMDC intermediate and poor risk groups. We were not able to obtain subgroup data according to MSKCC criteria. Looking at the combined groups, there may be little or no difference between NIV+IPI and SUN in improving PFS.
- **CAB:** For the IMDC intermediate and poor risk groups, we found that CAB alone probably improves PFS when compared to SUN alone. We were not able to obtain data for the other risk groups. Looking at the combined risk groups, we found that CAB alone may improve PFS.
- **PAZ:** We were not able to obtain subgroup data. Looking at the combined risk groups, there probably is little or no difference in PFS between PAZ alone and SUN alone.

Adverse events (AEs)

See Summary of findings 1, Summary of findings 2, and Summary of findings 3.

We were not able to obtain subgroup data per risk group for this outcome. Looking at the combined risk groups; we were also not able to obtain data for PEM+AXI and NIV+IPI. However, we found that there probably is little or no difference in the risk for AEs in AVE+AXI, NIV+CAB and PAZ alone, when compared to SUN alone, respectively. The combination LEN+PEM probably increases slightly the risk for AEs, when compared to SUN alone. Lastly, we are uncertain whether CAB alone reduces or increases the risk for AEs, when compared to SUN.

Number of participants who discontinued treatment due to an AE

We were not able to obtain subgroup data per risk group for this outcome. Looking at the combined risk groups, we observed that AVE+AXI, PAZ alone, CAB alone, PEM+AXI, NIV+CAB, LEN+PEM, and NIV+IPI decrease or increase the risk for study discontinuation due to an AE, when compared to SUN alone, respectively.

Time to initiation of first subsequent therapy

See Summary of findings 1, Summary of findings 2, and Summary of findings 3.

We were not able to analyse this outcome as a time-to-event outcome, mainly due to differences in outcome definition and reporting. Therefore, we reported the results narratively. Moreover, we were not able to obtain subgroup data, and also no data for the substance PAZ. For the combined risk groups, we found that PEM+AXI, AVE+AXI and NIV+IPI may reduce the risk for subsequent therapy, when compared to SUN alone, respectively. We are uncertain whether NIV+CAB, NIV+IPI and CAB alone reduce the risk for subsequent therapy, when compared to SUN alone, respectively.

Overall completeness and applicability of evidence

In this systematic review with network meta-analysis, we included 36 RCTs that assessed first-line treatments for adults with advanced RCC. All trials but one (NCT00126594) were published as full-text publications. For 11 trials, we also identified study protocols (some including a statistical analysis plan) that provided detailed information on the study design, participants, methods and outcomes, which informed our risk of bias assessments (NCT00720941; NCT00903175; NCT01030783; NCT01064310; NCT01835158; NCT02231749; NCT02420821; NCT02684006; NCT02811861; NCT02853331; NCT03141177). For two trials (NCT00334282; NCT00720941), we identified clinical study reports and for another trial (NCT01064310), a scientific result summary was found, which we used as our primary sources for data extraction and risk of bias assessment. In addition, we identified 19 ongoing trials and five trials that are still awaiting classification. Regarding heterogeneity between trials, we determined moderate heterogeneity within the networks for the outcome OS in the analyses for all risk groups combined and for IMDC intermediate and poor risk groups, substantial heterogeneity for PFS in the analysis of the MSKCC favourable risk group, and moderate heterogeneity for PFS in both the MSKCC and IMDC intermediate and poor risk groups. Lastly, we observed substantial heterogeneity for the outcome SAE (all risk groups combined). The heterogeneity probably originates from the slight differences in the included trials with regard to some effect modifiers (for example, regarding histology types or differences in the sites of metastases; and in the analyses of all risk groups combined, also differences with regard to the risk groups). Nevertheless, our included trials remain largely comparable. Unfortunately, subgroup analyses for some of these characteristics were not possible (see Differences between protocol and review), so we were not able to further explore heterogeneity statistically. Looking at the concordance between survival outcomes (OS and PFS) for the interventions presented in the summary of findings tables, we found that overall, across all risk groups, OS and PFS were improved (albeit to different extent) with every treatment, when the treatment was compared to SUN alone. The only instance where a survival outcome was not improved but rather reduced was in the IMDC favourable risk groups, where we found that NIV+IPI probably reduces PFS, when compared to SUN. However, it should be noted that the certainty in the evidence varies across OS and PFS even when we observed improvements. The main reason for this is that for the outcome PFS, some evidence was rated as a high risk of bias, namely when the imaging scans were assessed by unblinded study investigators. In comparison, some studies conducted blinded independent central review to avoid bias due to unblinded outcome assessment (see also Implications for research).

There are several limitations to this review that we want to address. Firstly, the results of the review mostly apply to people with clear cell carcinoma, as most trials in this review included participants with the clear cell type, whereas other types were underrepresented. We aimed to conduct subgroup analyses for the different histology types (clear cell type, papillary type, sarcomatoid type), but this was not possible as in most trials, *only* participants with clear cell RCC were included, followed by trials in which *most* participants had clear cell RCC. In addition, we did not have sufficient subgroup data for histology types to perform subgroup analyses. Hence, results should be interpreted with caution. Secondly, regarding the network meta-analyses of our key interventions of interest, meaning those that we chose for our summary of findings tables, we only had direct evidence from one trial per comparison (except for one comparison, namely CAB alone versus SUN alone in the PFS analysis of the combined risk groups). Due to insufficient data, we were not able to combine direct evidence in pairwise analyses, and we were not able to create so-called closed loops of direct and indirect evidence. Most interventions of interest were compared to SUN alone in the included trials, but not to each other. Hence, there is a great lack of head-to-head comparisons of these interventions. Therefore, there was no additional benefit from network meta-analyses. Thirdly, reporting of AEs differed between trials; therefore, some trials could not be included in the analyses for this outcome. We were not able to include data on AEs from 18 trials (50% of included trials) due to major reporting differences between studies. At protocol stage, we decided to include the number of participants who experienced at least one event, instead of cumulated events to avoid double counting. Moreover, we were interested in grade 3 or 4 adverse events, and in all-cause adverse events. However, we found major reporting differences between trials, meaning that half of the included trials reported cumulated grades of severity, cumulated events and/or treatment-related instead of all-cause AEs, which made it impossible to use all data for our analyses. As for the individual AEs of interest, reporting of insomnia and depression was sparse, so analysing these specific events was not possible at all. As for the timing of outcome measurement and reporting, all AEs and also SAEs reported were those that occurred during treatment. However, because most trials provided continuous therapy, while others provided therapy for a fixed period of time, the time points of occurrence of AEs and SAEs most likely varied between trials. In addition, for trials with continuous therapy (where therapy was administered until progression, or even beyond progression if a clinical benefit was observed according to the treating clinician), inevitably, there is an increased risk for the occurrence of SAEs. Therefore, results for these outcomes should be interpreted with caution. The same applies to the outcome 'number of participants who discontinued treatment due to an AE'. Fourthly, a total of 22 trials reported the outcome health-related QoL. However, a total of 25 different scales were used across trials to measure this outcome, so we prioritised scales for assessment in this review. In the end, we prioritised five scales, from four of which data were extractable. However, neither network meta-analyses nor pairwise meta-analyses were possible for this outcome, mainly due to a lack of available comparisons within our pre-specified time points. Fifthly, the outcome Time to First Subsequent Therapy (TFST) was not reported as such (i.e. as time-to-event outcome). Instead, trials reported the number of participants who received subsequent anticancer therapy after discontinuation of study treatment. Hence, analyses were not feasible, and we reported results narratively, which should also be interpreted with caution because the definition of this outcome varied across trials and the time point of reporting was unclear. We reported data on this outcome in the SoF table 1, but particularly downgraded by two levels in the domain 'indirectness' due to an indirect measurement of our outcome of interest. Hence, the certainty in the evidence for this outcome ranges from low to very low. Sixthly, we were not able to perform relevant subgroup analyses by sex (male, female), age (< 65 years, > 65 years), prior nephrectomy (yes, no), prior radiotherapy (yes, no), histology type (clear cell, papillary, sarcomatoid), and sites of metastases (lung, bone, liver). It was not possible to differentiate by study (e.g.,

by analysing and comparing studies with only women against studies with only men, as all studies included both sexes), and there was a great lack of subgroup data. While few studies did report some subgroup data for the outcome OS, potential network or pairwise meta-analyses would have included no more than two or three studies. Such analyses would not have produced meaningful results. However, we want to stress the importance of assessing the benefits and harms of the different treatments in different subgroups. For example, research has found that immune checkpoint inhibitors seem more effective in men than women; whereas for women, immune checkpoint inhibitors combined with chemotherapy seem more effective than for men (Wang 2019). Additional subgroups that were not considered in this review, but could be assessed in an update of the review, are ethnicity (Nassar 2022) and Eastern Cooperative Oncology Group (ECOG) performance status. Lastly, most trials provided a hazard ratio (HR) for the outcomes OS and PFS, and we were able to include all interventions listed in Table 1 (Description of studies) in our analyses. However, networks were not always fully connected, meaning that only treatments within the same sub-network could be compared to each other. Moreover, only five trials reported that they tested the assumption of proportional hazards, thereof four reported that the assumption was also validated. For the remaining studies it remains unclear whether the assumption of proportional hazards was tested. If the assumption is not validated, it is unclear which impact this would have on the meta-analysis.

Risk of bias

We assessed the risk of bias for the total population (i.e. all risk groups combined) for the outcomes OS, PFS, AEs, SAEs and QoL. For OS and PFS, risk of bias was additionally assessed for each risk group (i.e. favourable, intermediate or poor risk group per IMDC or MSKCC). Risk of bias was predominantly judged as 'high risk of bias' or 'some concerns' across most trials and outcomes. The judgement between the total population and the risk groups differed for only one trial and one outcome (NCT00720941). The main reasons for negative judgements were the lack of detailed information about the randomisation process, the blinding of outcome assessors, the method of analysis and the method of outcome measurement. Furthermore, pre-agreed study protocols and statistical analyses plans (SAPs) were missing for most trials. Only 11 pre-agreed study protocols were available (NCT00720941; NCT00903175; NCT01030783; NCT01064310; NCT01835158; NCT02231749; NCT02420821; NCT02684006; NCT02811861; NCT02853331; NCT03141177). Three of these protocols did not include a SAP (NCT00903175; NCT01030783; NCT01835158), and for one protocol the date of finalisation was not reported (NCT01030783). Clinical study reports (CSRs) were available for two trials (NCT00334282; NCT00720941); a scientific result summary was available for one trial (NCT01064310).

Certainty of the evidence

We rated our certainty in the evidence for the outcomes included in the SoF table (OS, QoL, SAEs, PFS, AEs, and TFST).

All risk groups combined

For all outcomes that were analysed in the combined risk groups (OS, QoL, SAEs, PFS, AEs, and TFST (the latter reported narratively)), our certainty in the evidence ranged from moderate to very low. For OS, we downgraded by one level in the domain 'study limitations' due to a high risk of bias in one comparison; by one level for imprecision in two comparisons, because of a wide confidence interval (CI) and the upper CI limit suggested no difference between interventions; by one level for imprecision in one comparison because the upper CI limit suggested no difference between interventions; by one level for imprecision in one comparison because of a wide CI that favoured either of the compared treatments. Lastly, for one comparison, we downgraded by one level for indirectness because in one trial (NCT02761057) seven per cent of the total study population received previous systemic therapy, and by two levels for imprecision because of a very wide CI that included values that favoured either of the compared treatments, and the evidence stemmed from only one trial with 90 participants (NCT02761057). For QoL, the only available evidence was rated as very low because we downgraded by two levels for study limitations due to a high risk of bias, and by two levels for imprecision because of a very wide CI that included values that favoured either of the compared treatments, and because the evidence stemmed from only one trial with four participants analysed (NCT00720941). For SAEs, the certainty in the evidence ranged from moderate to very low. Most evidence was downgraded by one level for study limitations because of a high risk of bias. In one instance, we downgraded by one level for imprecision because of a wide CI that favoured either of the compared treatments; and in another instance, we downgraded by two levels for imprecision because of a very wide CI that included values that favoured either of the compared treatments, and because the evidence stemmed from only one trial with 157 participants (NCT01835158). For PFS, the certainty in the evidence ranged from moderate to low. For most evidence we downgraded by one level for study limitations because of a high risk of bias. For one comparison, we downgraded by one level for indirectness because in one trial (NCT02761057) seven per cent of the total study population received previous systemic therapy. In two comparisons, we downgraded by one level for imprecision because of a wide CI that favoured either of the compared treatments. For AEs, the certainty of the evidence was mostly moderate, except for one comparison that was rated as very low. For all comparisons, we rated down by one level for study limitations due to a high risk of bias. For one comparison, we additionally downgraded by two levels for imprecision because of a wide CI that included values that favoured either of the interventions, and because the evidence stemmed from only one trial with 157 participants (NCT01835158). Lastly, for TFST, the certainty in the evidence ranged from low to very low. For all comparisons, we rated down by two levels for indirectness due to indirect measurement of the outcome of interest. For three comparisons, we rated down by one level for study limitations due to a high risk of bias. For one comparison, we downgraded by two levels for imprecision because of a wide CI that included values that favoured either of the interventions, and because the evidence stemmed from only one trial with 157 participants (NCT01835158).

Favourable risk groups (according to IMDC and MSKCC)

In the IMDC favourable risk groups for OS, our certainty in the evidence ranged from low to very low. We mostly downgraded by one level for study limitations due to a high risk of bias and/or by one level for 'imprecision' when the CI was wide and included values that favoured either of the compared treatments; when the CI was wide and the upper CI limit suggested no difference between interventions; or when the evidence stemmed from only

one trial with < 150 participants. In some instances, we downgraded by two levels for imprecision, because of a very wide CI that included values that favoured either of the compared treatments, and/or when the evidence stemmed from only one trial with < 150 participants. For PFS, the evidence ranged from moderate to low. We mostly rated down by one level for study limitations due to a high risk of bias. In one instance, we additionally downgraded by one level for imprecision because of a wide CI, where the upper CI limit suggested no difference between interventions; and in another instance we downgraded by one level for imprecision because the evidence stemmed from only one trial with < 150 participants.

In the MSKCC favourable risk groups for OS, the certainty of the evidence ranged from low to very low. We mostly downgraded by one level for study limitations. In one instance, we additionally downgraded by one level for imprecision because of a wide CI that included values that favoured either of the compared treatments; and in another instance we downgraded by two levels for imprecision because of a very wide CI that included values that favoured either of the compared treatments. For PFS, there was only one comparison, where we rated our certainty in the evidence as very low. We downgraded by one level for study limitations due to a high risk of bias, and by two levels for imprecision because of a very wide CI that included values that favoured either of the compared treatments.

Intermediate and poor risk groups (according to IMDC and MSKCC)

In the IMDC risk groups for OS, our certainty in the evidence ranged from moderate to very low. Most evidence was downgraded by one level for study limitations because of a high risk of bias, and by one level for imprecision because of a wide CI that favoured either of the compared treatments. In one instance, we downgraded by two levels for imprecision because of a very wide CI that included values that favoured either of the compared treatments, and because the evidence stemmed from only one trial with 157 participants. For PFS, the certainty of the evidence ranged from moderate to low. Most evidence was downgraded by one level for study limitations because of a high risk of bias, and by one level for imprecision because of a wide CI that favoured either of the compared treatments. In one instance, we downgraded by one level for imprecision because the evidence stemmed from only one trial with 157 participants.

In the MSKCC risk groups for OS, the certainty of the evidence ranged from moderate to low. The evidence was downgraded by one level for study limitations because of a high risk of bias, and/or by one level for imprecision because of a wide CI that favoured either of the compared treatments. For PFS, the only available comparison was rated as moderate because we downgraded by one level for study limitations due to a high risk of bias.

Potential biases in the review process

A key strength of our review is that it is a very comprehensive review and includes all available treatment options in the first-line treatment setting for adults with advanced RCC. We explored the effectiveness of all treatment options (where data were available), i.e. the effectiveness of different combinations of substances from different drug categories as well as the effectiveness of individual substances alone.

To prevent potential bias in our review, the important steps in the review development process were conducted by two review authors independently (i.e. study screening and selection, data extraction, and risk of bias assessments). Only the GRADE assessment was conducted by one review author (AAa) first, and then the assessment was independently examined by another review author (VP). Discrepancies were then resolved by discussion. Overall, seven co-authors (AAa, BB, CH, ET, MG, ND, VP) were involved in the different important steps of the review development. Any conflicts that arose during the review process were resolved by discussion until a consensus was reached, and if necessary, by involving a third review author.

Trials and related publications were identified by a sensitive search strategy developed by an experienced information specialist, and we searched all relevant databases (CENTRAL; MEDLINE; Embase), several trial registries (ISRCTN; EU Clinical Trial Register; ClinicalTrials.gov; WHO ICTRP) as well as conference proceedings of relevant conferences (ASCO; ESMO). We also reviewed other published systematic reviews on first-line therapies for adults with advanced RCC to make sure that we did not miss any trials. The fact that this is a living systematic review during its development process, with the last search conducted in February 2022, and due to our extensive search strategy, we are confident that we have identified all relevant trials to address the research question of our review. Besides the 36 included trials, we identified an additional 19 ongoing trials that could be included in an update of this review. Methodologically, we followed all current Cochrane guidelines and recommendations in every stage of our review process and are not aware of any deficiencies in our review process. For transparency, we have documented and justified all changes to our methods from the published protocol (Goldkuhle 2020) in the Differences between protocol and review section.

Agreements and disagreements with other studies or reviews

We believe that, currently, our systematic review is the most comprehensive systematic review with network meta-analyses that explored different treatment options in the first-line therapy for advanced RCC). However, we identified a number of systematic reviews with meta-analyses or network meta-analyses assessing first-line therapy in advanced RCC. Here, we present the results of those systematic reviews that also conducted network meta-analyses, and we only assessed the most recent reviews of 2021/2022 due to the rapidly evolving treatment landscape. It should be noted that methodologically, the reviews differ from our review in that, firstly, none of the reviews included data from clinical study reports. Secondly, except for one review (Riaz 2021), none were living systematic reviews. Thirdly, not all reviews conducted a risk of bias and/or GRADE assessment. Lastly, most reviews included only a few selected trials that assessed only a selected number of treatment options. Hence, the reviews did not assess the full range of available treatment options for advanced RCC in the first-line treatment setting, making our review the most comprehensive. We compared results mainly for the outcomes OS and PFS. As for harms, most reviews reported treatment-related AEs or treatment-related discontinuations due to AEs, which are not comparable to our data (we assessed all-cause AEs). In one review, health-related quality of life (HRQoL) was assessed, the instruments included being EQ-5D and FACT-FKSI Symptom Index.

We primarily compared the results of our review to the results of the most recent reviews published in 2022 (Bosma 2022; Nocera 2022) and to one review from 2021 that is a living systematic review (Riaz 2021). In addition, we present brief summaries of the results of other reviews published in 2021.

Comparison of results of our review to other recent reviews

We identified one living systematic review with network meta-analyses that includes a total of 14 trials, which are also included in our review (https://rcc.network-meta-analysis.com/RCC.html) (Riaz 2021).

Comparison of results for Overall survival (OS)

The analyses for OS and all risk groups combined showed that LEN+PEM (HR 0.66, 95% CI 0.49 to 0.88), CAB+NIV (HR 0.66, 95% CI 0.50 to 0.87), PEM+AXI (HR 0.68, 95% CI 0.55 to 0.85) and NIV +IPI (HR 0.69, 95% CI 0.59 to 0.81) showed a substantial benefit for OS, compared to SUN. Treatment ranking (according to SUCRA analyses) showed that LEN+PEM (83%) had the highest likelihood of being the preferred treatment option for OS, closely followed by NIV+CAB (82%) and PEM+AXI (80%). In our analysis, we also found that LEN+PEM may improve OS, and in according to our ranking of treatments, it was also the best treatment option (P-score 0.85). Furthermore, we also found that PEM+AXI and NIV+IPI probably improve OS, when compared to SUN, respectively. We are uncertain whether CAB improves OS, and we did not have evidence for the comparison NIV+CAB versus SUN.

For the favourable risk groups (unclear whether IMDC or MSKCC risk groups were reported in the review), AVE+AXI (HR 0.81, 95% CI 0.34 to 1.94), NIV+CAB (HR 0.84, 95% CI 0.36 to 1.99) and PAZ alone (HR 0.88, 95% CI 0.63 to 1.22) may or may not improve OS when compared to SUN. However, AVE+AXI (SUCRA 63%) had the highest likelihood of being the preferred treatment option, closely followed by PAZ (62%) and NIV+CAB (60%). In our analysis of the IMDC favourable risk group, we found that AVE+AXI may improve OS, and it was also the best treatment option according to the ranking of treatments (P-score: 0.83). The reviews are also in agreement that there may be little or no difference between LEN+PEM and SUN in improving OS (our result: HR 1.15, 95% CI 0.55 to 2.40; the other reviews' result: HR 1.14, 95% CI 0.55 to 2.38). We are also uncertain about the effect of NIV+CAB. As for the MSKCC favourable risk groups, we are also uncertain about the effect of LEN+PEM, and we found that there may be little or no difference between PAZ and SUN.

For the intermediate and poor risk groups (unclear whether IMDC or MSKCC risk groups were reported in the review), NIV+CAB (HR 0.52, 95% CI 0.28 to 0.98), LEN+PEM (HR 0.61, 95% CI 0.44 to 0.85), PEM+AXI (HR 0.63, 95% CI 0.49 to 0.80) and NIV+IPI (HR 0.65, 95% CI 0.54 to 0.78) showed a substantial benefit in OS, compared to SUN. Treatment ranking showed that NIV+CAB (SUCRA: 82%) had the highest likelihood of being the preferred treatment option, followed by LEN+PEM (SUCRA: 73%). This is similar to our results, where for the IMDC risk groups, we found that NIV+CAB and LEN +PEM probably improve OS, and CAB alone may improves slightly OS, when compared to SUN, respectively. However, there is a difference in the ranking of these two treatments, because we found that LEN+PEM (P-score: 0.81) was the best treatment option, followed by NIV+CAB (P-score: 0.74). Furthermore, we also found that NIV+IPI probably improves slightly OS, compared to SUN. For the MSKCC risk groups, we also found that LEN+PEM probably improves OS, and it was the best treatment option according to the ranking of treatments (P-score: 0.81).

Comparison of results for Progression-free survival (*PFS)*

For PFS and for all risk groups combined, LEN+PEM (HR 0.39, 95% CI 0.31 to 0.48), CAB alone (HR 0.48, 95% CI 0.31 to 0.74), NIV+CAB (HR 0.52, 95% CI 0.43 to 0.63), AVE+AXI (HR 0.69, 95% CI 0.58 to 0.83) and PEM+AXI (HR 0.71, 95% CI 0.60 to 0.84) showed a substantial benefit in PFS, compared to SUN. Treatment ranking showed that LEN+PEM (SUCRA 98% CI) had the highest likelihood of being the preferred treatment option. We also found that LEN+PEM probably improve PFS, and PEM+AXI probably improve slightly PFS, when compared to SUN, respectively. We also found that CAB alone may improves PFS. In our ranking of treatments, LEN+PEM (P-score: 0.98) was also the best treatment option, closely followed by CAB alone (Pscore: 0.92), LEN+EVE (P-score: 0.87) and PEM+AXI (P-score: 0.86),

As for the favourable risk groups (unclear whether IMDC or MSKCC risk groups were reported in the review), LEN+PEM (HR 0.40, 95% CI 0.27 to 0.60) and AVE+AXI (HR 0.63, 95% CI 0.40 to 0.99) showed a substantial benefit in PFS, compared to SUN. The combination LEN+PEM had the highest likelihood (96%) of being the preferred treatment option. For the IMDC risk group, we also found that LEN+PEM, NIV+CAB and AVE+AXI probably improve PFS, when compared to SUN, respectively. LEN+PEM was also the best treatment option (P-score: 0.94). As the MKSCC risk group, we are uncertain whether LEN+PEM improves PFS, when compared to SUN.

For the intermediate and poor risk groups (unclear whether IMDC or MSKCC risk groups were reported in the review), LEN+PEM (HR 0.37, 95% CI 0.28 to 0.49), NIV+CAB (HR 0.47, 95% CI 0.33 to 0.67), CAB alone (HR 0.48, 95% CI 0.31 to 0.74), AVE+AXI (HR 0.65, 95% CI 0.44 to 0.95), PEM+AXI (HR 0.69, 95% CI 0.56 to 0.84) and NIV+IPI (HR 0.74, 95% CI 0.62 to 0.88) showed a substantial benefit in PFS, compared to SUN. Treatment ranking showed that LEN+PEM (SUCRA 95%) has the highest likelihood of being the preferred treatment option. For the IMDC risk groups, we also found that CAB alone, LEN+PEM, AVE+AXI and NIV+CAB probably improve PFS, when compared to SUN, respectively. However, we found that there may be little or no difference in PFS between NIV+IPI and SUN. LEN+PEM was also the best treatment option in our ranking (P-score: 0.94). For the MSKCC groups, we also found that LEN+PEM probably improve PFS, compared to SUN, and it was also the best treatment option (P-score: 0.98).

Bosma 2022 included six trials, which we also included (NCT02231749; NCT02420821; NCT02684006; NCT02811861; NCT02853331; NCT03141177) and assessed OS, PFS, and HRQoL. Harms were also assessed, however, in this review, treatment-related grade 3–4 AEs and treatment-related drug discontinuation were assessed, so we did compare these results to ours. Across all risk groups and with regard to OS, results suggested better OS with NIV+CAB (HR 0.60, 95% CrI 0.40 to 0.90), NIV+IPI (HR 0.69, 95% CrI 0.59 to 0.81), PEM+AXI (HR 0.68, 95% CrI 0.55 to 0.84), and LEN+PEM (HR 0.66, 95% CrI 0.49 to 0.88) when compared to SUN, respectively. This is similar to the results of our review, where we also found improvements in OS and PFS with PEM+AXI and LEN +PEM in comparison to SUN. However, we did not have evidence for AVE+AXI and NIV+CAB in the combined risk groups. Looking at the different subgroups, treatment ranking showed that AVE

+AXI (65%), PEM+AXI (78%) and LEN+PEM (89%) for the favourable, intermediate and poor risk groups, respectively, had the highest likelihood of being the preferred treatment options in terms of OS. In our review, we found that for the IMDC risk groups, AVE+AXI was the best treatment option for the favourable risk group and LEN +PEM was the best option for the intermediate or poor risk groups. As for the MSKCC risk groups, LEN+EVE was the best option for the favourable risk group and LEN+PEM the best option for the intermediate and poor risk groups.

With regard to PFS, results of the other review suggested better PFS with AVE+AXI (HR 0.69, 95% CrI 0.57 to 0.83), NIV+CAB (HR 0.51, 95% CrI 0.41 to 0.64), PEM+AXI (HR 0.71, 95% CrI 0.60 to 0.84), and LEN+PEM (HR 0.39, 95% CrI 0.32 to 0.48) when compared to SUN, respectively. Treatment ranking (based on SUCRA) revealed that LEN+PEM (99%) had the highest likelihood of being the preferred treatment option for the entire population in terms of PFS. As for the different risk groups, LEN+PEM also had the highest likelihood of being the preferred treatment option for the favourable, intermediate and poor risk groups with a 96%, 98% and 89% likelihood, respectively. We obtained the same result in our review: according to our ranking of treatments, LEN+PEM was the best treatment option across all groups, and both amongst IMDC and MSKCC.

As for HRQoL, analysis of treatment ranking for EQ-5D showed that LEN+PEM (SUCRA 85%) followed by NIV+CAB (SUCRA 75%) were associated with the highest likelihood of being the preferred treatment. For the Functional Assessment of Cancer Therapy Kidney Cancer Symptom Index (FKSI) questionnaire, NIV+IPI (SUCRA 93%) followed by NIV+CAB (SUCRA 66%) was associated with the highest likelihood of being the preferred treatment option. Unfortunately, an analysis of these treatments on QoL was not feasible in our review.

Nocera 2022 only assessed four RCTs with proven OS benefit relative to SUN, which we also included (NCT02231749; NCT02853331; NCT02811861; NCT03141177). Outcomes included OS, PFS and treatment-related grade 3+4 AEs; the main comparator being SUN. Results showed that the combination NIV+CAB (P-score: 0.77), followed by LEN+PEM (P-score: 0.63), PEM+AXI (P-score: 0.57) and NIV+IPI (P-score: 0.53) had the highest likelihood of OS benefit for all risk groups combined. As we did not have results for NIV +CAB in the combined risk groups, the ranking in our review was different: LEN+PEM came first with a P-score of 0.85, followed by NIV +IPI with a P-score: 0.83 and then PEM+AXI with a P-score of 0.78. The results for PFS in the other review suggest that the treatments in the following order showed the highest likelihood of benefit (for all risk groups combined): LEN+PEM (P-score: 0.99), NIV+CAB (P-score: 0.76), PEM+AXI (P-score:0.50), NIV+IPI (P-score: 0.24). Again, we did not have data for NIV+CAB, so our ranking was different: LEN +PEM was the best option (P-score: 0.98), PEM+AXI was the fourth-best option (P-score: 0.86) and NIV+IPI sixth-best option (P-score: 0.71).

Brief summary of results from reviews published in 2021

Cattrini 2021 only assessed immunotherapy and included six trials that we also included (NCT02231749; NCT02420821; NCT02684006; NCT02811861; NCT02853331; NCT03141177). Outcomes assessed were OS in the total population, OS per IMDC subgroup and grade #3 AEs. The main comparator was SUN. In terms of OS benefit, results showed the highest likelihood (based on SUCRA analyses) for the combinations of NIV+CAB (82%), LEN+PEM (72%), PEM +AXI (68%) and NIV+IPI (56%) being the preferred treatments for all risk groups combined. With regard to the IMDC risk groups, PEM+AXI (78%) had the highest likelihood of being the preferred treatment for the intermediate risk group, and PEM+LEN (74%) the highest for the poor risk group. Contradicting results were shown for the favourable risk group. With regard to toxicity: NIV+IPI (96%), followed by ATE+BEV (87%), SUN (55%) and AVE+AXI (54%) were the preferred options with the highest tolerability.

Liu 2021 included five trials that we also included (NCT01984242; NCT02231749; NCT02420821; NCT02684006; NCT02853331) and assessed immunotherapy treatment options only. For PFS, and compared to SUN, results suggested better PFS with AVE+AXI (HR 0.69, 95% CrI 0.56 to 0.85) and PEM+AXI (HR 0.69, 95 CrI 0.57 to 0.83), followed by NIV+IPI (HR 0.82, 95% CrI 0.68 to 0.99) when compared to SUN, respectively. For OS, and compared to SUN, results suggested better OS with PEM+AXI (HR 0.53, 95% CrI 0.38 to 0.74) and NIV+IPI (HR 0.63, 95% CrI 0.48 to 0.83). However, no data were available for AVE+AXI and ATE alone. As for AEs, ATE alone had a lower risk for AEs (odds ratio (OR) 0.26, 95% CI 0.15 – 0.43), followed by NIV+IPI (OR 0.50, 95% CI 0.39 to 0.64) and ATE+BEV (OR 0.60, 95% CI 0.47 to 0.76) when compared to SUN, respectively.

Mori 2021 included five trials that we also included (NCT00720941; NCT02231749; NCT02420821; NCT02684006; NCT02853331) and assessed OS, PFS and treatment-related AEs. For OS and for all risk groups combined, results suggested better OS with PEM+AXI (HR 0.85, 95% CrI 0.73 to 0.98) and NIV+IPI (HR 0.86, 95% CrI 0.75 to 0.99), but the upper CI limits suggested no difference to SUN, respectively. PEM+AXI (P-score 0.80) was the best option based on treatment ranking. For PFS and for all risk groups combined, PEM+AXI (HR 0.86, 95% CrI 0.76 to 0.97) and AVE+AXI (HR 0.85, 95% CrI 0.74 to 0.98) may improve PFS but the upper limit of the CIs suggested no difference. AVE+AXI (P-score: 0.82) and PEM+AXI (P-score: 0.80) were the best options based on treatment ranking.

Quhal 2021 assessed immunotherapies only and included six trials that we also included (NCT02231749; NCT02420821; NCT02684006; NCT02811861; NCT02853331; NCT03141177). For OS and for all risk groups combined, PEM+AXI (HR 0.85, 95% CrI 0.73 to 0.98) and NIV+IPI (HR 0.85, 95% CrI 0.76 to 0.95) may lead to better OS, but the upper CI limits suggest no difference to SUN, respectively. According to SUCRA treatment ranking, NIV+CAB had the highest likelihood of providing the maximal OS (P-score 0.75739). Results of effectiveness were unclear for the favourable risk groups. As for the intermediate and poor risk groups, LEN+PEM (HR 0.71, 95% CI 0.54 to 0.95) and NIV+CAB (HR 0.73, 95% CI 0.55 to 0.97) showed improvements in OS, but the upper CI limits suggested no difference to SUN, respectively. For PFS and for all risk groups combined, results suggested that LEN+PEM (HR 0.66, 95% CrI 0.61 to 0.72), NIV+CAB (HR 0.75, 95% CrI 0.67 to 0.84), AVE+AXI (HR 0.85, 95% CrI 0.75 to 0.96) and PEM+AXI (HR 0.86, 95% CrI 0.76 to 0.97) showed improvements in PFS when compared to SUN, respectively. Treatment ranking according to SUCRA showed that LEN+PEM (P-score: 0.99) had the highest likelihood of providing the maximal PFS, followed by NIV+CAB (P-score: 0.82). As for the favourable risk group, LEN+PEM (HR 0.68, 95% CI 0.57 to 0.80) and PEM+AXI (HR 0.82, 95% CI 0.70 to 0.96) suggest improvement in PFS when compared to SUN, respectively. As for the intermediate and poor risk groups, results suggested that LEN+PEM (HR 0.64, 95% CI 0.55 to 0.74), NIV+CAB (HR 0.71, 95% CI 0.61 to 0.82) and AVE+AXI (HR

0.83, 95% CI 0.70 to 0.97) improve PFS when compared to SUN, respectively.

AUTHORS' CONCLUSIONS

Implications for practice

The findings of our living systematic review with network meta-analyses may be an aid to clinicians and people with advanced renal cell carcinoma in decision-making about treatment options for first-line therapy. However, most evidence for currently recommended treatment options for the different risk groups stems from direct evidence from one trial only; hence, the results of this review should be interpreted with caution. Furthermore, before a decision is met about a treatment option, the results of all outcomes should be taken into consideration, meaning benefits and harms should be contrasted with one another. Because most networks in the network meta-analyses in this review are not fully connected, not all treatment combinations could be compared to each other. Furthermore, the main results of our review on the effectiveness of different combination therapies stem from comparisons to SUN alone. Considering the interventions that are currently most recommended (PEM+AXI, AVE+AXI, NIV+CAB, LEN +PEM, NIV+IPI, CAB alone, and PAZ alone) across clinical practice guidelines (NCCN; ESMO; EAU; German S3 guideline) and across the different risk groups, we found the following evidence on survival, harms and quality of life, when each intervention is compared to SUN alone.

Implications for survival, harms and quality of life in the combined risk groups

For both OS and PFS, we are most certain in the evidence for PEM +AXI (probably improve OS, probably improve slightly PFS), LEN +PEM (may improve OS, probably improve PFS), and CAB alone (may improve PFS, but we are uncertain on the evidence for OS). We found that NIV+IPI probably improve OS, but for PFS, there may be little or no difference between NIV+IPI and SUN. For both OS and PFS, we found that there is probably little or no difference between PAZ alone and SUN alone. We did not have evidence for AVE+AXI and NIV+CAB for neither OS nor PFS. It should be highlighted that some of these treatments come with a higher incidence in SAEs and AEs (severity grades 3 and 4). We found that PEM+AXI probably increase slightly, and NIV+IPI and LEN+PEM probably increase the risk for SAEs, when compared to SUN, respectively. We are uncertain whether CAB alone reduces or increases the risk, and there is probably little or no difference in the risk for SAEs between PAZ alone and SUN alone. We did not have evidence for AVE+AXI and NIV+CAB on the comparative risk for SAEs. As for AEs, we found that LEN+PEM probably increase slightly the risk for AEs, and there is probably little or no difference in the risk for AEs between AVE+AXI, NIV+CAB, and PAZ alone, when compared to SUN alone, respectively. We are uncertain in the evidence for CAB alone, and we did not have evidence on PEM+AXI and NIV+IPI. All in all, when making treatment decisions, it should be individually evaluated whether the benefits of some of these interventions in improving OS or PFS outweigh their increased risk for harms. Unfortunately, we had a great lack of evidence on the impact of these interventions on the quality of life of people with advanced RCC.

Implications for survival, harms and quality of life for favourable versus intermediate + poor risk groups (IMDC and MSKCC)

For the favourable risk groups, we were most certain in the evidence for AVE+AXI, which may improve OS and PFS (in the IMDC group). We were not certain in the remaining evidence for OS. As for PFS, we were most certain in the evidence on NIV+CAB and LEN +PEM, namely that each may improve PFS (in the IMDC group), and that NIV+IPI probably reduce PFS (in IMDC group). We are missing evidence on CAB and PAZ for the favourable risk groups. As for the intermediate and poor risk groups, we were most certain in the evidence for LEN+PEM, namely that this combination probably improves OS and PFS (in the IMDC and MSKCC groups). We were also most certain in that NIV+CAB probably improves OS and PFS (in the IMDC groups), and that AVE+AXI probably improve PFS and may improve OS (in the IMDC groups). We are also certain that NIV +IPI probably improves OS (in the IMDC group). There may be little or no difference in PFS (for IMDC groups). There was a great lack of evidence for the MSKCC risk groups. We did not have subgroup data by risk group for harms and quality of life.

Implications for research

The research field on first-line therapies for adults with advanced renal cell carcinoma is a very fast evolving field due to the continuously changing treatment landscape that includes newer combinations of targeted therapies and immunotherapies. Hence, we identified 19 currently ongoing trials. However, for those interventions that are currently recommended across the different risk groups, thus far we only found direct evidence from one trial only, respectively. Furthermore, in all trials, these interventions were all compared to SUN. Thus, more trials are needed where these interventions and combinations are compared head-to-head, and not only to SUN. In our review, the additional benefit from the network meta-analytic approach is limited because for most interventions, we could not create so-called closed loops (involving at least three interventions) of direct and indirect evidence, where each direct comparison of interventions can be supplemented by an indirect comparison.

Regarding the outcome measurement of PFS, more studies are needed that perform blinded independent central reviews (BICR) of imaging scans when assessing PFS. Most studies in this review were not blinded and PFS was assessed presumably by unblinded study investigators, which we assessed as a high risk of bias. In comparison, few studies in this review that were non-masked conducted BICR to control bias in the outcome measurement, which we assessed as a low risk of bias. In some instances, PFS was assessed both by the unblinded investigators and by BICR, and results were compared by the studies.

In this review, we initially aimed to conduct important subgroup analyses (e.g., by histology type, sex or age), but none were possible due to a great lack of reporting of subgroup data by the primary studies. Assessing the impact of immunotherapies and targeted therapies on different subgroups is essential, for example to understand differences by sex or ethnicity in responding to different therapies. Specifically for RCC, more studies are needed that assess histology types such as the papillary type or the sarcomatoid type. Most participants in this review had clear cell RCC, thus the results of this review are mostly applicable to the clear cell type. A sufficient and thorough analysis of different subgroups could be achieved by analysing individual participant data (IPD) provided by study authors.

Lastly, making study protocols (SPs), statistical analyses plans (SAPs) and clinical study reports (CSRs) publicly available would allow for a more detailed and accurate assessment of studies and data. All SPs, SAPs and CSRs that were found in this review informed data extraction and risk of bias assessments.

ACKNOWLEDGEMENTS

We thank Joanne Platt (Information Specialist, Cochrane Gynaecological, Neuro-oncology and Orphan Cancer Group) for peer-reviewing the search strategy (at protocol stage).

The following people conducted the editorial process for this article:

- Sign-off Editor (final editorial decision): Michael Brown, Cochrane's Editorial Board, Michigan State University College of Human Medicine, USA;
- Managing Editor (selected peer reviewers, collated peer-reviewer comments, provided editorial guidance to authors, edited the article): Sam Hinsley, Cochrane Central Editorial Service;
- Editorial Assistant (conducted editorial policy checks and supported editorial team): Lisa Wydrzynski, Cochrane Central Editorial Service;
- Copy Editor (copy editing and production): Heather Maxwell;
- Peer-reviewers (provided comments and recommended an editorial decision): Stella O'Brien (consumer review), Nuala Livingstone, Cochrane Evidence Production Methods (methods review), Anne Littlewood, Information Specialist at Cochrane Oral Health (search review). Two additional peer reviewers provided clinical peer review, but chose not to be publicly acknowledged.

REFERENCES

References to studies included in this review

Jonasch 2010 *{published data only}*

Jonasch E, Corn P, Asche RG. Randomized phase II study of sorafenib with or without low-dose IFN in patients with metastatic renal cell carcinoma. *Journal of Clinical Oncology. Conference: ASCO Annual Meeting* 2007;**25**(18 Suppl):Abstract 5104.

* Jonasch E, Corn P, Pagliaro LC, Warneke CL, Johnson MM, Tamboli P, et al. Upfront, randomized, phase 2 trial of sorafenib versus sorafenib and low-dose interferon alfa in patients with advanced renal cell carcinoma: clinical and biomarker analysis. *Cancer* 2010;**116**(1):57-65.

NCT00065468 *{published data only}*

Atkins MB, Hidalgo M, Stadler W, Logan T, Dutcher JP, Hudes G, et al. A randomized double-blind phase 2 study of intravenous CCI-779 administered weekly to patients with advanced renal cell carcinoma. In: Proceedings of the American Society of Clinical Oncology. Vol. 21 (Pt 1). 2002:10a, Abstract 36. [ClinicalTrials.gov: NCT00065468]

Bellmunt J, Szczylik C, Feingold J, Strahs A, Berkenblit A. Temsirolimus safety profile and management of toxic effects in patients with advanced renal cell carcinoma and poor prognostic features. *Annals of Oncology* 2008;**19**(8):1387-92. [ClinicalTrials.gov: NCT00065468]

Hudes G, Carducci M, Tomczak P, Dutcher J, Figlin R, Kapoor A, et al. A phase 3, randomized, 3-arm study of temsirolimus (TEMSR) or interferon-alpha (IFN) or the combination of TEMSR + IFN in the treatment of first-line, poor-risk patients with advanced renal cell carcinoma (adv RCC). *Journal of Clinical Oncology: ASCO Annual Meeting Proceedings* 2006;**24**(18S):LBA4. [ClinicalTrials.gov: NCT00065468]

* Hudes G, Carducci M, Tomczak P, Dutcher J, Figlin R, Kapoor A, et al. Temsirolimus, interferon alfa, or both for advanced renal-cell carcinoma. *New England Journal of Medicine* 2007;**356**(22):2271-81. [ClinicalTrials.gov: NCT00065468]

Rajagopalan S, Pullenayegum E, Alemao E, Strahs A, Purvis J. Evaluation of adverse event-related hospitalizations in patients with advanced renal cell carcinoma on treatment with temsirolimus or interferon-alfa: results from a phase 3 randomized trial. *European Journal of Cancer, Supplement* 2009;**7**(2-3):441. [ClinicalTrials.gov: NCT00065468]

NCT00072046 *{published data only}*

Halabi S, Rini BI, Stadler WM, Small EJ. Use of progression-free survival (PFS) to predict overall survival (OS) in patients with metastatic renal cell carcinoma (mRCC). *Journal of Clinical Oncology. Conference: ASCO Annual Meeting* 2010;**28**(15 Suppl 1):4525-4525. [ClinicalTrials.gov: NCT00072046]

Melichar B, Bracarda S, Bellmunt J, Ravaud A, Negrier S, Jethwa S, et al. First-line bevacizumab + reduced-dose interferon-alpha2a in patients (pts) with metastatic renal cell carcinoma (mRCC): An update on overall survival. *European Journal of Cancer, Supplement* 2009;**7**(2-3):430. [ClinicalTrials.gov: NCT00072046]

Nixon AB, Halabi S, Liu Y, Starr MD, Brady JC, Shterev I, et al. Predictive biomarkers of overall survival in patients with metastatic renal cell carcinoma treated with IFNalpha +/- bevacizumab: results from CALGB 90206 (alliance). *Clinical Cancer Research* 2021;**OF1 - OF8**:29. [ClinicalTrials.gov: NCT00072046]

Rini BI, Halabi S, Rosenberg J, Stadler WM, Vaena DA, Atkins JN, et al. Bevacizumab plus interferon-alpha versus interferon-alpha monotherapy in patients with metastatic renal cell carcinoma: results of overall survival for CALGB 90206. *Journal of Clinical Oncology* 2009;**27**(15S Pt 1):239. [ClinicalTrials.gov: NCT00072046]

* Rini BI, Halabi S, Rosenberg JE, Stadler WM, Vaena DA, Archer L, et al. Phase III trial of bevacizumab plus interferon alfa versus interferon alfa monotherapy in patients with metastatic renal cell carcinoma: final results of CALGB 90206. *Journal of Clinical Oncology* 2010;**28**(13):2137-43. [ClinicalTrials.gov: NCT00072046]

Rini BI, Halabi S, Rosenberg JE, Stadler WM, Vaena DA, Ou SS, et al. Bevacizumab plus interferon alfa compared with interferon alfa monotherapy in patients with metastatic renal cell carcinoma: CALGB 90206. *Journal of Clinical Oncology* 2008;**26**(33):5422-8. [ClinicalTrials.gov: NCT00072046]

NCT00081614 *{published data only}*

Bukowski RM, Kabbinavar F, Figlin RA, Flaherty K, Srinivas S, Vaishampayan U, et al. Bevacizumab with or without erlotinib in metastatic renal cell carcinoma (RCC). *Journal of Clinical Oncology: ASCO Annual Meeting Proceedings* 2006;**24**(18S):4523. [ClinicalTrials.gov: NCT00081614]

* Bukowski RM, Kabbinavar FF, Figlin RA, Flaherty K, Srinivas S, Vaishampayan U, et al. Randomized phase II study of erlotinib combined with bevacizumab compared with bevacizumab alone in metastatic renal cell cancer. *Journal of Clinical Oncology* 2007;**25**(29):4536-41. [ClinicalTrials.gov: NCT00081614]

NCT00098657/NCT00083889 *{published data only}*

Blagoev KB, Wilkerson J, Stein WD, Motzer RJ, Bates SE, Fojo AT. Effect of sunitinib (Su) administration on post-treatment survival in patients with metastatic renal cell carcinoma (mRCC) treated on the upfront randomized phase III trial of sunitinib or interferon alfa (IFN). *Journal of Clinical Oncology. Conference: ASCO Annual Meeting* 2011;**29**(15 Suppl 1):4634-4634. [ClinicalTrials.gov: NCT00098657/NCT00083889]

Bushmakin A, Cappelleri JC, Korytowsky B, Sandin R, Matczak E, Cella D. Sunitinib (Su) dosing schedule and data collection timepoints: Impact on quality of life (QoL) outcomes in metastatic renal cell carcinoma (mRCC). *Annals of Oncology* 2012;**9**:ix269. [ClinicalTrials.gov: NCT00098657/NCT00083889]

Cappelleri JC, Bushmakin AG, Charbonneau C, Kim ST, Li JZ, Motzer RJ. Quality of life (QoL) with sunitinib versus interferon-

alfa (IFN) as first-line therapy in patients with metastatic renal cell carcinoma (mRCC): final results. *Journal of Clinical Oncology* 2009;**27**(15S Pt I):330. [ClinicalTrials.gov: NCT00098657/NCT00083889]

Castellano D, del Muro XG, Perez-Gracia JL, Gonzalez-Larriba JL, Abrio MV, Ruiz MA, et al. Patient-reported outcomes in a phase III, randomized study of sunitinib versus interferon-{alpha} as first-line systemic therapy for patients with metastatic renal cell carcinoma in a European population. *Annals of Oncology* 2009;**20**(11):1803-12. [ClinicalTrials.gov: NCT00098657/NCT00083889]

Cella D, Bushmakin AG, Cappelleri JC, Charbonneau C, Li JZ, Kim ST, et al. Health-related quality of life (HrQoL) in metastatic renal cell carcinoma (mRCC) patients receiving sunitinib (Su) or interferon (IFN)-alfa in a randomized phase III trial: updated geographic analysis. *Annals of Oncology* 2009;**19**(Suppl 8):189. [ClinicalTrials.gov: NCT00098657/NCT00083889]

Cella D, Bushmakin AG, Cappelleri JC, Charbonneau C, Sandin R, Negrier S, et al. Comparison of health-related quality of life (HrQoL) in patients reporting the same adverse event (AE) fatigue on different treatments. *European Journal of Cancer* 2011;**47**:233. [ClinicalTrials.gov: NCT00098657/NCT00083889]

Cella D, Li JZ, Cappelleri JC, Bushmakin A, Charbonneau C, Kim ST, et al. Quality of life in patients with metastatic renal cell carcinoma treated with sunitinib or interferon alfa: results from a phase III randomized trial. *Journal of Clinical Oncology* 2008;**26**(22):3763-9. [ClinicalTrials.gov: NCT00098657/NCT00083889]

Cella D, Michaelson MD, Bushmakin AG, Cappelleri JC, Charbonneau C, Kim ST, et al. Final quality of life (QoL) results with geographical analysis for sunitinib versus interferon-alfa, as first-line therapy in patients with metastatic renal cell carcinoma (mRCC). *Value in Health* 2009;**12**(7):A283. [ClinicalTrials.gov: NCT00098657/NCT00083889]

Cella D, Michaelson MD, Cappelleri JC, Bushmakin AG, Charbonneau C, Kim ST, et al. Quality of life (QoL) with sunitinib versus interferon-alfa (IFN-alpha) as first-line therapy in patients with metastatic renal cell carcinoma (mRCC): Final results. *Journal of Clinical Oncology* 2009;**1**:6529. [ClinicalTrials.gov: NCT00098657/NCT00083889]

Doehn C. Sunitinib vs interferon-alpha for the treatment of metastatic renal cell carcinoma.. *Onkologe* 2007;**13**(9):862-4. [ClinicalTrials.gov: NCT00098657/NCT00083889]

Figlin RA, Hutson TE, Tomczak P, Michaelson MD, Bjarnason GA, Lin X, et al. Characterizing prior nephrectomy patients with metastatic renal cell carcinoma (mRCC) treated with first-line sunitinib versus interferon-alfa. *European Journal of Cancer* 2013;**2**:S644-5. [ClinicalTrials.gov: NCT00098657/NCT00083889]

Figlin RA, Hutson TE, Tomczak P, Michelson MD, Bukowski RM, Negrier S, et al. Overall survival with sunitinib versus interferon (IFN)-alfa as first-line treatment of metastatic renal cell carcinoma (mRCC). *Journal of clinical oncology: ASCO annual meeting proceedings* 2008;**26**(15S Pt 1):256. [ClinicalTrials.gov: NCT00098657/NCT00083889]

* Motzer RJ, Hutson TE, Tomczak P, Michaelson MD, Bukowski RM, Oudard S, et al. Overall survival and updated results for sunitinib compared with interferon alfa in patients with metastatic renal cell carcinoma. *Journal of Clinical Oncology* 2009;**27**(22):3584-90. [ClinicalTrials.gov: NCT00098657/NCT00083889]

Motzer RJ, Hutson TE, Tomczak P, Michaelson MD, Bukowski RM, Rixe O, et al. Phase III randomized trial of sunitinib malate (SU11248) versus interferon-alfa (IFN-{alpha}) as first-line systemic therapy for patients with metastatic renal cell carcinoma (mRCC). *Journal of Clinical Oncology: ASCO annual meeting proceedings* 2006;**24**(18S):LBA3. [ClinicalTrials.gov: NCT00098657/NCT00083889]

Motzer RJ, Hutson TE, Tomczak P, Michaelson MD, Bukowski RM, Rixe O, et al. Sunitinib versus interferon alfa in metastatic renal-cell carcinoma. *New England Journal of Medicine* 2007;**356**(2):115-24. [ClinicalTrials.gov: NCT00098657/NCT00083889]

Negrier S, Figlin RA, Hutson TE, Tomczak P, Michaelson MD, Bukowski RM, et al. Overall survival with sunitinib versus interferon (IFN)-alfa as first-line treatment of metastatic renal cell carcinoma (mRCC). *Annals of Oncology* 2008;**19 (S8)**:viii190. [ClinicalTrials.gov: NCT00098657/NCT00083889]

Patel PH, Schwartz LH, Baum MS, Motzer RJ. Superior efficacy of sunitinib compared with interferon as first-line therapy in patients with metastatic renal cell carcinoma. *American Journal of Hematology/Oncology* 2007;**6**(5):260-4. [ClinicalTrials.gov: NCT00098657/NCT00083889]

Patil S, Figlin RA, Hutson TE, Michaelson D, Negrier S, Kim ST, et al. TWiST analysis to estimate overall benefit for metastatic renal cell carcinoma (mRCC) patients (pts) treated in a phase III trial of sunitinib versus interferon-alfa (IFN-alpha). *Journal of Clinical Oncology* 2010;**28**(15 Suppl 1):4594-4594. [ClinicalTrials.gov: NCT00098657/NCT00083889]

Patil S, Figlin RA, Hutson TE, Michaelson MD, Negrier S, Kim ST, et al. Q-TWiST analysis to estimate overall benefit for patients with metastatic renal cell carcinoma treated in a phase III trial of sunitinib vs interferon-alpha. *British Journal of Cancer* 2012;**106**(10):1587-90. [ClinicalTrials.gov: NCT00098657/NCT00083889]

Pickering LM, Pyle L, Larkin JM. Sunitinib is superior to interferon alpha with respect to quality of life for patients with renal cell carcinoma. *Nature Clinical Practice Oncology* 2009;**6**(1):6-7. [ClinicalTrials.gov: NCT00098657/NCT00083889]

Reddy K, Bukowski RM. Phase III study of sunitinib malate (SU11248) versus interferon-α as first-line treatment in patients with metastatic renal cell carcinoma. *Clinical Genitourinary Cancer* 2006;**5**(1):23-5. [ClinicalTrials.gov: NCT00098657/NCT00083889]

Smith AD, Zhang H, Souza F, Roda M, Penman A. Predicting survival in metastatic RCC on sunitinib using MASS criteria: evaluation of a large multicentered prospective phase III trial. *Journal of Clinical Oncology. Conference* 2013;**31**(6 Suppl 1):407. [ClinicalTrials.gov: NCT00098657/NCT00083889]

Szczylik C, Negrier S, Oudard S, Figlin RA, Hutson TE, Tomczak P, et al. Overall survival with sunitinib versus interferon-alfa (IFN) as first-line treatment for metastatic renal cell carcinoma (mRCC). *European Urology, Supplements* 2009;**8**(4):182. [ClinicalTrials.gov: NCT00098657/NCT00083889]

Yang S, Hudes G, DeSouza P, Alemao E, Strahs A, Purvis J. Evaluation of quality of life in patients with advanced renal cell carcinoma treated with temsirolimus vs interferon-alfa: results from a phase III randomized trial. *European Journal of Cancer, Supplement* 2009;**7**(2-3):433. [ClinicalTrials.gov: NCT00098657/NCT00083889]

NCT00117637 *{published data only}*

Escudier B, Szczylik C, Demkow T, Staehler M, Rolland F, Negrier S, et al. Randomized phase II trial of the multi-kinase inhibitor sorafenib versus interferon (IFN) in treatment-naive patients with metastatic renal cell carcinoma (mRCC). *Journal of Clinical Oncology: ASCO annual meeting proceedings* 2006;**24**(18S):4501. [ClinicalTrials.gov: NCT00117637]

* Escudier B, Szczylik C, Hutson TE, Demkow T, Staehler M, Rolland F, et al. Randomized phase II trial of first-line treatment with sorafenib versus interferon alfa-2a in patients with metastatic renal cell carcinoma. *Journal of Clinical Oncology* 2009;**27**(8):1280-9. [ClinicalTrials.gov: NCT00117637]

Escudier. Randomized phase II trial of first-line treatment with sorafenib versus interferon alfa-2a in patients with metastatic renal cell carcinoma. *Journal of Clinical Oncology* 2009;**27**(13):2305. [ClinicalTrials.gov: NCT00117637]

Felsch M, Zaim S, Dicken V, Lehmacher W, Scheuring UJ. Comparison of central and local serial CT assessments of metastatic renal cell carcinoma patients in a clinical phase IIB study. *Acta Radiologica* 2017;**58**(2):249-55. [ClinicalTrials.gov: NCT00117637]

Szczylik C, Demkow T, Staehler M, Rolland F, Negrier S, Hutson TE, et al. Randomized phase II trial of first-line treatment with sorafenib versus interferon in patients with advanced renal cell carcinoma: final results. *Journal of Clinical Oncology: ASCO annual meeting proceedings* 2007;**25**(18 Suppl):Abstract 5025. [ClinicalTrials.gov: NCT00117637]

NCT00126594 *{published data only}*

Sorafenib tosylate with or without recombinant interferon alpha-2b in treating patients with metastatic kidney cancer. https://clinicaltrials.gov/ct2/show/NCT00126594. [ClinicalTrials.gov: NCT00126594]

NCT00334282 *{published and unpublished data}*

Cella D, Pickard AS, Duh MS, Guerin A, Mishagina N, Antras L, et al. Health-related quality of life in patients with advanced renal cell carcinoma receiving pazopanib or placebo in a randomised phase III trial. *European Journal of Cancer* 2012;**48**(3):311-23. [ClinicalTrials.gov: NCT00334282]

Pickard AS, Cella D, Duh MS, Guerin A, Mishagina N, Antras L, et al. Health-related quality of life in patients with advanced renal cell carcinoma receiving pazopanib or placebo in a randomized phase III trial. In: Journal of Clinical Oncology. Vol. 29. 2011. [ClinicalTrials.gov: NCT00334282]

* Sternberg CN, Davis ID, Mardiak J, Szczylik C, Lee E, Wagstaff J, et al. Pazopanib in locally advanced or metastatic renal cell carcinoma: results of a randomized phase III trial. *Journal of Clinical Oncology* 2010;**28**(6):1061-8. [ClinicalTrials.gov: NCT00334282]

Sternberg CN, Hawkins RE, Szczylik C, Davis ID, Wagstaff J, McCann L, et al. A randomized, double-blind phase III study (VEG105192) of pazopanib (paz) versus placebo (pbo) in patients with advanced/metastatic renal cell carcinoma (mRCC): updated safety results. *Journal of Clinical Oncology* 2011;**29**(7 Suppl 1):313. [ClinicalTrials.gov: NCT00334282]

Sternberg CN, Hawkins RE, Szczylik C, Davis ID, Wagstaff J, McCann L, et al. Randomized, double-blind phase III study of pazopanib in patients with advanced/metastatic renal cell carcinoma (mRCC): final overall survival (OS) results. *Annals of Oncology* 2010;**8**:viii10. [ClinicalTrials.gov: NCT00334282]

Sternberg CN, Hawkins RE, Wagstaff J, Salman P, Mardiak J, Barrios CH, et al. A randomised, double-blind phase III study of pazopanib in patients with advanced and/or metastatic renal cell carcinoma: final overall survival results and safety update. *European Journal of Cancer* 2013;**49**(6):1287-96. [ClinicalTrials.gov: NCT00334282]

NCT00420888 *{published data only}*

* Eisen T, Hedlund G, Forsberg G, Nordle O, Hawkins R. Baseline biomarker trend analysis of a randomized phase 2/3 study of naptumomab estafenatox plus IFN-a vs IFN-a in advanced renal cell carcinoma. *European Journal of Cancer* 2013;**2**:S645. [ClinicalTrials.gov: NCT00420888]

Elkord E, Burt DJ, Sundstedt A, Nordle Ö, Hedlund G, Hawkins RE. Immunological response and overall survival in a subset of advanced renal cell carcinoma patients from a randomized phase 2/3 study of naptumomab estafenatox plus IFN-α versus IFN-α. *Oncotarget* 2015;**6**(6):4428-39. [ClinicalTrials.gov: NCT00420888]

Hawkins RE, Gore M, Shparyk Y, Bondar V, Gladkov O, Ganev T, et al. A randomized phase II/III study of naptumomab estafenatox + IFNα versus IFNα in renal cell carcinoma: final analysis with baseline biomarker subgroup and trend analysis. *Clinical Cancer Research* 2016;**22**(13):3172-81. [ClinicalTrials.gov: NCT00420888]

NCT00420888. ABR-217620/Naptumomab Estafenatox With Interferon-alpha (IFN-alpha) Compared to IFN-alpha Alone in Patients With Advanced Renal Cell Carcinoma. clinicaltrials.gov/show/NCT00420888 first received 11 January 2007. [ClinicalTrials.gov: NCT00420888]

NCT00609401 *{published data only}*

Procopio G, Verzoni E, Bracarda S, Conti G, Guadalupi V, Ortega C, et al. Subgroup analysis and updated results of a randomized study comparing sorafenib plus interleukin-2 versus sorafenib alone as first-line treatment in metastatic renal cell carcinoma. *Anticancer Research* 2010;**30**(4):1461-2. [ClinicalTrials.gov: NCT00609401]

Procopio G, Verzoni E, Bracarda S, Ricci S, Miceli R, Bertolini A, et al. A randomized, prospective, phase 2 study, with sorafenib

(So) and interleukin-2 (IL-2) versus So alone as first line treatment in advanced renal cell cancer (RCC): ROSORC trial. *European Journal of Cancer* 2009;**7**(2-3):425. [ClinicalTrials.gov: NCT00609401]

Procopio G, Verzoni E, Bracarda S, Ricci S, Ridolfi L, Sacco C, et al. Overall survival (OS) of sorafenib (So) plus interleukin-2 (IL-2) versus So alone in patients with treatment-naive metastatic renal cell carcinoma (mRCC): final update of the ROSORC trial. *Journal of Clinical Oncology* 2013;**31**(6 Suppl 1):356. [ClinicalTrials.gov: NCT00609401]

Procopio G, Verzoni E, Bracarda S, Ricci S, Sacco C, Ridolfi L, et al. A randomized, open label, prospective study comparing the association between sorafenib (So) and interleukin-2 (IL-2) versus So alone in advanced untreated renal cell cancer (RCC): ROSORC trial. *Journal of Clinical Oncology* 2009;**27**(15S Pt 1):258. [ClinicalTrials.gov: NCT00609401]

Procopio G, Verzoni E, Bracarda S, Ricci S, Sacco C, Ridolfi L, et al. Overall survival for sorafenib plus interleukin-2 compared with sorafenib alone in metastatic renal cell carcinoma (mRCC): final results of the ROSORC trial. *Annals of Oncology* 2013;**24**(12):2967-71. [ClinicalTrials.gov: NCT00609401]

Procopio G, Verzoni E, Bracarda S, Ricci S, Sacco C, Ridolfi L, et al. Sorafenib with interleukin-2 vs sorafenib alone in metastatic renal cell carcinoma: the ROSORC trial. *British Journal of Cancer* 2011;**104**(8):1256-61. [ClinicalTrials.gov: NCT00609401]

Procopio G, Verzoni E, Bracarda S, Ricci S, Sacco C, Ridolfi L, et al. Subgroup analysis and updated results of the randomized study comparing sorafenib (So) plus interleukin-2 versus So alone as first-line treatment in metastatic renal cell carcinoma. *Journal of Clinical Oncology: ASCO annual meeting proceedings* 2010;**28**(15 Suppl):Abstract 4589. [ClinicalTrials.gov: NCT00609401]

* Verzoni E, Bracarda S, Ricci S, Sacco C, Ridolfi L, Porta C, et al. Proof of concept of correlation between progression-free survival (PFS) and overall survival (OS) in treatment-naive patients with advanced renal-cell carcinoma (mRCC): final results of a randomized phase II study comparing sorafenib plus interleukin-2 (IL-2) versus sorafenib alone: the ROSORC trial. *Journal of Clinical Oncology* 2013;**31**(15 Suppl 1):e15589. [ClinicalTrials.gov: NCT00609401]

NCT00619268 *{published data only}*

Bay JO, Negrier S, Perol D, Gravis G, Chevreau C, Delva R, et al. Updated results on long-term overall survival (OS) of the French randomized phase II trial TORAVA in metastatic renal cell carcinoma (mRCC) patients. *Journal of Clinical Oncology* 2012;**30**(15 Suppl 1):4625. [ClinicalTrials.gov: NCT00619268]

Escudier BJ, Negrier S, Gravis G, Chevreau C, Delva R, Bay J, et al. Can the combination of temsirolimus and bevacizumab improve the treatment of metastatic renal cell carcinoma? Results of the randomized TORAVA phase II trial. *Journal of Clinical Oncology: ASCO annual meeting proceedings* 2010;**28**(15 Suppl):Abstract 4516. [ClinicalTrials.gov: NCT00619268]

* Negrier S, Gravis G, Perol D, Chevreau C, Delva R, Bay JO, et al. Temsirolimus and bevacizumab, or sunitinib, or interferon alfa and bevacizumab for patients with advanced renal cell carcinoma (TORAVA): a randomised phase 2 trial. *Lancet Oncology* 2011;**12**(7):673-80. [ClinicalTrials.gov: NCT00619268]

NCT00631371 *{published data only}*

Rini BI, Bellmunt J, Clancy J, Wang K, Niethammer A, Escudier B. Randomized phase IIIb trial of temsirolimus and bevacizumab versus interferon and bevacizumab in metastatic renal cell carcinoma: results from intoract. *Annals of Oncology* 2012;**9**:ix14. [ClinicalTrials.gov: NCT00631371]

* Rini BI, Bellmunt J, Clancy J, Wang K, Niethammer AG, Hariharan S, et al. Randomized phase III trial of temsirolimus and bevacizumab versus interferon alfa and bevacizumab in metastatic renal cell carcinoma: INTORACT trial. *Journal of Clinical Oncology* 2014;**32**(8):752-9. [ClinicalTrials.gov: NCT00631371]

NCT00719264 *{published data only}*

Ravaud A, Bajetta E, Kay AC, Pelov D, Hollaender N, Rouyrre N, et al. Everolimus with bevacizumab versus interferon alfa-2a plus bevacizumab as first-line therapy in patients with metastatic clear cell renal cell carcinoma. In: Journal of Clinical Oncology: ASCO annual meeting proceedings. Vol. 28. 2010. [ClinicalTrials.gov: NCT00719264]

* Ravaud A, Barrios CH, Alekseev B, Tay MH, Agarwala SS, Yalcin S, et al. RECORD-2: phase II randomized study of everolimus and bevacizumab versus interferon alpha-2a and bevacizumab as first-line therapy in patients with metastatic renal cell carcinoma. *Annals of Oncology* 2015;**26**(7):1378-84. [ClinicalTrials.gov: NCT00719264]

Ravaud A, Barrios CH, Alekseev BY, Tay MH, Agarwala SS, Yalcin S, et al. Randomized phase II study of first-line everolimus plus bevacizumab (E+B) versus interferon alpha-2a plus bevacizumab (I+B) in patients (pts) with metastatic renal cell carcinoma (mRCC): Record-2 final overall survival (OS) and safety results. *Journal of Clinical Oncology: ASCO annual meeting proceedings* 2013;**31**(15 Suppl 1):4576. [ClinicalTrials.gov: NCT00719264]

NCT00720941 *{published and unpublished data}*

Beaumont JL, Diaz J, Deen KC, McCann L, Powles T, Hackshaw MD, et al. Q-TWiST analysis of patients with metastatic renal cell carcinoma (mRCC) randomized to pazopanib (PAZ) or sunitinib (SUN). *Journal of Clinical Oncology: ASCO annual meeting proceedings* 2014;**32**(15 Suppl 1):4581. [ClinicalTrials.gov: NCT00720941]

Bjarnason GA, Kollmannsberger C, Ahmad Q, Dezzani L, Elmeliegy M, Han J, et al. Effects of pazopanib (PAZ) and sunitinib (SUN) dose modification on safety and efficacy in patients with metastatic renal cell carcinoma (mRCC) from COMPARZ. *Kidney cancer* 2018;**2**:S18. [ClinicalTrials.gov: NCT00720941]

Bjarnason GA, Kollmannsberger CK, Ahmad Q, Dezzani L, Elmeliegy M, Han J, et al. Effects of pazopanib (PAZ) and sunitinib (SUN) dose modification on safety and efficacy in patients with metastatic renal cell carcinoma (mRCC) from COMPARZ. *Journal of Clinical Oncology* 2017;**35**(15 Suppl 1):4574. [ClinicalTrials.gov: NCT00720941]

Cella D, Hackshaw MD, Diaz J, Huang C, Deen KC, Crescenzo R, et al. Quality of life (QoL) among patients with renal cell carcinoma (RCC) treated with pazopanib versus sunitinib in the COMPARZ study. *Journal of Clinical Oncology. Conference* 2013;**31**(6 Suppl 1):346. [ClinicalTrials.gov: NCT00720941]

Guo J, Jin J, Huang Y, Wang JW, Lim HY, Uemura H, et al. Comparison of PFS and safety for Asian compared to North American and European populations in the phase III trial of pazopanib versus sunitinib in patients with treatment-naive RCC (COMPARZ). *Journal of Clinical Oncology* 2013;**31**(6 Suppl 1):366. [ClinicalTrials.gov: NCT00720941]

Guo J, Jin J, Oya M, Uemura H, Takahashi S, Tatsugami K, et al. Safety of pazopanib and sunitinib in treatment-naive patients with metastatic renal cell carcinoma: Asian versus non-Asian subgroup analysis of the COMPARZ trial. *Journal of Hematology and Oncology* 2018;**11**(1):69. [ClinicalTrials.gov: NCT00720941]

Johnson T, Xu CF, Choueiri TK, Deen KC, Xue Z, Bartlett-Pandite AN, et al. Association of hyperbilirubinemia in pazopanib- or sunitinib-treated patients in COMPARZ with UGT1A1 polymorphisms. *Journal of Clinical Oncology: ASCO annual meeting proceedings* 2013;**31**(15 Suppl):4569. [ClinicalTrials.gov: NCT00720941]

Kattan MW, Sternberg CN, Mehmud F, Bhatt K, Chakrabarti D, McCann L, et al. Development and validation of a prognostic nomogram for progression-free and overall survival in patients with advanced renal cell carcinoma (aRCC) treated with pazopanib. *Journal of Clinical Oncology. Conference* 2014;**32**(4 Suppl 1):405. [ClinicalTrials.gov: NCT00720941]

Motzer R, Hutson TE, Reeves J, Hawkins R, Guo J, Nathan P, et al. Randomized, open-label, phase III trial of pazopanib versus sunitinib in first-line treatment of patients with metastatic renal cell carcinoma (mRCC): results of the COMPARZ trial. *Annals of Oncology* 2012;**9**:ix13. [ClinicalTrials.gov: NCT00720941]

* Motzer RJ, Hutson TE, Cella D, Reeves J, Hawkins R, Guo J, et al. Pazopanib versus sunitinib in metastatic renal-cell carcinoma. *New England Journal of Medicine* 2013;**369**(8):722-31. [ClinicalTrials.gov: NCT00720941]

Pond GR, Dietrich M, Grunwald V. Prognostic ability of early tumor shrinkage on overall survival (OS) in metastatic renal cell carcinoma (mRCC)-a validation study. *Annals of Oncology* 2016;**27 Suppl 6**:VI283. [ClinicalTrials.gov: NCT00720941]

Sheng X, Jin J, He Z, Huang Y, Zhou A, Wang J, et al. Pazopanib versus sunitinib in Chinese patients with locally advanced or metastatic renal cell carcinoma: pooled subgroup analysis from the randomized, COMPARZ studies. *BMC Cancer* 2020;**20**(1):219. [ClinicalTrials.gov: NCT00720941]

Sheng X, Jin J, He Z, Huang Y, Zhou AP, Wang J, et al. Efficacy and safety of pazopanib (PAZ) versus sunitinib (SUN) in patients (pts) with locally advanced or metastatic renal cell carcinoma (RCC): a pooled China subgroup analysis from COMPARZ studies. *Journal of Clinical Oncology* 2018;**36**(15):e16588. [ClinicalTrials.gov: NCT00720941]

Tannir NM, Porta C, Grunwald V, Choueiri TK, Ahmad QI, Carrasco-Alfonso MJ, et al. Long-term response and time to response to pazopanib (PAZ) and sunitinib (SUN) in metastatic renal cell carcinoma (mRCC): COMPARZ subanalysis. *Kidney Cancer* 2018;**2 Suppl 1**:25-6. [ClinicalTrials.gov: NCT00720941]

Xu CF, Johnson T, Choueiri TK, Deen KC, Xue Z, Spraggs CF, et al. Association of IL8 polymorphisms with overall survival in patients with renal cell carcinoma in COMPARZ (pazopanib versus sunitinib phase III study). *Journal of Clinical Oncology: ASCO annual meeting proceedings* 2013;**31**(15 Suppl 1):4519. [ClinicalTrials.gov: NCT00720941]

NCT00732914 *{published data only}*

* Eichelberg C, Vervenne WL, De Santis M, Fischer von Weikersthal L, Goebell PJ, Lerchenmüller C, et al. SWITCH: a randomised, sequential, open-label study to evaluate the efficacy and safety of sorafenib-sunitinib versus sunitinib-sorafenib in the treatment of metastatic renal cell cancer. *European Urology* 2015;**68**(5):837-47. [ClinicalTrials.gov: NCT00732914]

Fischer Von Weikersthal L, Goebell PJ, De Santis M, Lerchenmuller C, Zimmermann U, Bos MMEM, et al. SWITCH: A randomized sequential open-label study to evaluate efficacy and safety of sorafenib (So) / sunitinib (Su) versus Su/So in the treatment of metastatic renal cell cancer (mRCC). *Oncology Research and Treatment* 2014;**5**:5. [ClinicalTrials.gov: NCT00732914]

Fischer Von Weikersthal L, Vervenne WL, Goebell PJ, Eichelberg C, Freier W, De Santis M, et al. Phase III randomized sequential open-label study to evaluate efficacy and safety of sorafenib (So) followed by sunitinib (Su) versus sunitinib followed by sorafenib in patients with advanced / metastatic renal cell carcinoma without prior systemic therapy (SWITCH Study)-Safety interim analysis results. *Onkologie* 2012;**6**:237. [ClinicalTrials.gov: NCT00732914]

Goebell PJ, Vervenne W, De Santis M, Von Weikersthal LF, Lerchenmuller CA, Zimmermann U, et al. Subgroup analyses of a randomized sequential open-label study (SWITCH) to evaluate efficacy and safety of sorafenib (So)/sunitinib (Su) versus Su/So in the treatment of metastatic renal cell cancer (mRCC). *Journal of Clinical Oncology: ASCO annual meeting proceedings* 2014;**32**(15 Suppl 1):4567. [ClinicalTrials.gov: NCT00732914]

Michel MS, Vervenne W, De Santis M, Von Weikersthal LF, Goebell PJ, Lerchenmueller J, et al. SWITCH: a randomized sequential open-label study to evaluate efficacy and safety of sorafenib (So)/sunitinib (Su) versus Su/So in the treatment of metastatic renal cell cancer (mRCC). *Journal of Clinical Oncology* 2014;**32**(4 Suppl 1):393. [ClinicalTrials.gov: NCT00732914]

Michel MS, Vervenne W, Goebell PJ, Von Weikersthal LF, Freier W, De Santis M, et al. Phase III randomized sequential open-label study to evaluate efficacy and safety of sorafenib (So) followed by sunitinib (Su) versus sunitinib followed by sorafenib in patients with advanced/metastatic renal cell carcinoma without prior systemic therapy (SWITCH Study): safety interim analysis results. *Journal of Clinical Oncology* 2012;**30**(15 Suppl 1):4539. [ClinicalTrials.gov: NCT00732914]

Retz M, Bedke J, Bögemann M, Grimm MO, Zimmermann U, Müller L, et al. SWITCH II: phase III randomized, sequential,

open-label study to evaluate the efficacy and safety of sorafenib-pazopanib versus pazopanib-sorafenib in the treatment of advanced or metastatic renal cell carcinoma (AUO AN 33/11). *European Journal of Cancer* 2019;**107**:37-45. [ClinicalTrials.gov: NCT00732914]

NCT00738530 *{published data only}*

Bajetta E, Ravaud A, Bracarda S, Negrier S, Szczylik C, Bellmunt Molins J, et al. Efficacy and safety of first-line bevacizumab (BEV) plus interferon-a2a (IFN) in patients (pts) >65 years with metastatic renal cell carcinoma (mRCC). *Journal of Clinical Oncology: ASCO annual meeting proceedings* 2008;**26**(15S pt I):273. [ClinicalTrials.gov: NCT00738530]

* Escudier B, Bellmunt J, Negrier S, Bajetta E, Melichar B, Bracarda S, et al. Phase III trial of bevacizumab plus interferon alfa-2a in patients with metastatic renal cell carcinoma (AVOREN): final analysis of overall survival. *Journal of Clinical Oncology* 2010;**28**(13):2144-50. [ClinicalTrials.gov: NCT00738530]

Escudier B, Bellmunt J, Negrier S, Melichar B, Bracarda S, Ravaud A, et al. Final results of the phase III, randomized, double-blind AVOREN trial of first-line bevacizumab (BEV) + interferon alfa-2a (IFN) in metastatic renal cell carcinoma (mRCC). *Journal of Clinical Oncology* 2009;**27**(15S Pt 1):239. [ClinicalTrials.gov: NCT00738530]

Escudier B, Koralewski P, Pluzanska A, Ravaud A, Bracarda S, Szczylik C, et al. A randomized, controlled, double-blind phase III study (AVOREN) of bevacizumab/interferon-a2a vs placebo/interferon-α2a as first-line therapy in metastatic renal cell carcinoma. *Journal of Clinical Oncology: ASCO annual meeting proceedings* 2007;**25**(18 Suppl):Abstract 3. [ClinicalTrials.gov: NCT00738530]

Escudier B, Pluzanska A, Koralewski P, Ravaud A, Bracarda S, Szczylik C, et al. Bevacizumab plus interferon alfa-2a for treatment of metastatic renal cell carcinoma: a randomised, double-blind phase III trial. *Lancet* 2007;**370**(9605):2103-11. [ClinicalTrials.gov: NCT00738530]

Sommerauer M, Doehn C. Bevacizumab plus interferon alfa-2a in the treatment of metastatic renal cell carcinoma: a randomized, double-blind phase III study. *Onkologe* 2008;**14**(5):535-537. [ClinicalTrials.gov: NCT00738530]

NCT00903175 *{published data only}*

Hsieh J, Chen D, Wang P, Chen Y, Marker M, Patel P, et al. Differential overall survival (OS) results in RECORD-3 study based on three distinct mRCC molecular subgroups classified by BAP1 and/or PBRM1 mutations. *Journal of Clinical Oncology* 2016;**34**(2 Suppl 1):561. [ClinicalTrials.gov: NCT00738530]

Hsieh J, Chen D, Wang P, Chen YB, Mahtab M, Pate P, et al. Differential overall survival results in RECORD-3 study based on three distinct clear cell metastatic renal cell carcinoma molecular subgroups classified by BAP1 and/ or PBRM1 mutations. *BJU International* 2016;**118 Suppl 5**:10. [ClinicalTrials.gov: NCT00738530]

Hsieh JJ, Chen D, Wang PI, Chen YB, Redzematovic A, Marker M, et al. Identification of efficacy biomarkers in a large metastatic renal cell carcinoma (mRCC) cohort through next-generation sequencing (NGS): results from RECORD-3. *BJU International* 2015;**5**:13. [ClinicalTrials.gov: NCT00738530]

Hsieh JJ, Chen D, Wang PI, Marker M, Redzematovic A, Chen YB, et al. Genomic biomarkers of a randomized trial comparing first-line everolimus and sunitinib in patients with metastatic renal cell carcinoma. *European Urology* 2017;**71**(3):405-14. [ClinicalTrials.gov: NCT00738530]

Knox JJ, Barrios CH, Kim TM, Cosgriff T, Srimuninnimit V, Pittman KB, et al. Final overall survival analysis for the RECORD-3 study of first-line everolimus followed by sunitinib versus first-line sunitinib followed by everolimus in metastatic RCC (mRCC). *Journal of Clinical Oncology* 2015;**33**(15 Suppl 1):4554. [ClinicalTrials.gov: NCT00738530]

Knox JJ, Kay AC, Schiff E, Hollaender N, Rouyrre N, Ravaud A, et al. First-line everolimus followed by second-line sunitinib versus the opposite treatment sequence in patients with metastatic renal cell carcinoma (mRCC). In: Journal of Clinical Oncology: ASCO annual meeting proceedings. Vol. 28. 2010. [ClinicalTrials.gov: NCT00738530]

* Motzer RJ, Barrios CH, Kim TM, Falcon S, Cosgriff T, Harker WG, et al. Phase II randomized trial comparing sequential first-line everolimus and second-line sunitinib versus first-line sunitinib and second-line everolimus in patients with metastatic renal cell carcinoma. *Journal of Clinical Oncology* 2014;**32**(25):2765-72. [ClinicalTrials.gov: NCT00738530]

Motzer RJ, Barrios CH, Kim TM, Falcon S, Cosgriff T, Harker WG, et al. Record-3: phase II randomized trial comparing sequential first-line everolimus (EVE) and second-line sunitinib (SUN) versus first-line SUN and second-line EVE in patients with metastatic renal cell carcinoma (mRCC). *Journal of Clinical Oncology* 2013;**31**(15 Suppl 1):4504. [ClinicalTrials.gov: NCT00738530]

NCT00920816 *{published data only}*

* Hutson TE, Al-Shukri S, Stus VP, Lipatov ON, Shparyk Y, Bair AH, et al. Axitinib versus sorafenib in first-line metastatic renal cell carcinoma: overall survival from a randomized phase III trial. *Clinical Genitourinary Cancer* 2017;**15**(1):72-6. [ClinicalTrials.gov: NCT00920816]

Hutson TE, Al-Shukri S, Stus VP, Lipatov ON, Shparyk Y, Bair AH, et al. Overall survival analysis from a randomised phase III trial of axitinib vs sorafenib as first-line therapy in patients with metastatic renal cell carcinoma. *European Journal of Cancer* 2015;**51**:476-77. [ClinicalTrials.gov: NCT00920816]

Hutson TE, Gallardo J, Lesovoy V, Al-Shukri S, Stus V, Bair AH, et al. Axitinib versus sorafenib as first-line therapy in patients with metastatic renal cell carcinoma (mRCC). *Journal of Clinical Oncology* 2013;**31**(6 Suppl 1):1287-94. [ClinicalTrials.gov: NCT00920816]

Sheng X, Bi F, Ren X, Cheng Y, Wang J, Rosbrook B, et al. First-line axitinib versus sorafenib in Asian patients with metastatic renal cell carcinoma (mRCC): subgroup analysis of data from a phase III trial. *Annals of Oncology* 2017;**28**:x82. [ClinicalTrials.gov: NCT00920816]

Sheng X. First-line axitinib versus sorafenib in Asian patients with metastatic renal cell carcinoma: exploratory subgroup analyses of phase III data. *Future Oncology* 2019;**15**(1):3267-77. [ClinicalTrials.gov: NCT00920816]

NCT00979966 *{published data only}*

Bergmann L, Grunwald V, Maute L, Grimm MO, Weikert S, Schleicher J, et al. A prospective randomized phase-II trial with temsirolimus vs. sunitinib in non-clear renal cell carcinoma. A study of the CESAR central european society for anticancer drug research-EWIV and interdisciplinary renal cell carcinoma group of the German cancer society (IAGN). *European Journal of Cancer* 2015;**3**:517. [ClinicalTrials.gov: NCT00979966]

* Bergmann L, Grunwald V, Maute L, Grimm MO, Weikert S, Schleicher J, et al. A randomized phase IIa trial with temsirolimus versus sunitinib in advanced non-clear cell renal cell carcinoma: an intergroup study of the CESAR Central European Society for Anticancer Drug Research-EWIV and the Interdisciplinary Working Group on Renal Cell Cancer (IAGN) of the German Cancer Society. *Oncology Research and Treatment* 2020;**43**(7-8):333-9. [ClinicalTrials.gov: NCT00979966]

NCT01024920 *{published data only}*

* Eisen T, Loembe AB, Shparyk Y, MacLeod N, Jones RJ, Mazurkiewicz M, et al. A randomised, phase II study of nintedanib or sunitinib in previously untreated patients with advanced renal cell cancer: 3-year results. *British Journal of Cancer* 2015;**113**(8):1140-7. [ClinicalTrials.gov: NCT01024920]

Eisen T, Shparyk Y, Jones R, MacLeod NJ, Temple G, Finnigan H, et al. Phase II efficacy and safety study of nintedanib versus sunitinib in previously untreated renal cell carcinoma (RCC) patients. *Journal of Clinical Oncology* 2013;**31**(15 Suppl 1):4506. [ClinicalTrials.gov: NCT01024920]

Eisen T, Shparyk Y, MacLeod NJ, Wallenstein G, Temple G, Khder Y, et al. Effects of nintedanib (BIBF 1120) on QTc interval in previously untreated patients with renal cell cancer (RCC): results from an open-label, phase II study. In: Journal of Clinical Oncology (Conference). Vol. 30. 2012. [ClinicalTrials.gov: NCT01024920]

Eisen T, Shparyk Y, Macleod N, Jones R, Wallenstein G, Temple G, et al. Effect of small angiokinase inhibitor nintedanib (BIBF 1120) on QT interval in patients with previously untreated, advanced renal cell cancer in an open-label, phase II study. *Investigational New Drugs* 2013;**31**(5):1283-93. [ClinicalTrials.gov: NCT01024920]

NCT01030783 *{published data only}*

Cella D, Ivanescu C, Skaltsa K, Casamayor M, Strahs AL, Esteves B, et al. Treatment benefit of tivozanib hydrochloride versus sorafenib on health-related quality of life (HRQoL) among patients (pts) with advanced/metastatic renal cell carcinoma (mRCC): TIVO-1 study results. *Journal of Clinical Oncology. Conference* 2013;**31**(6 Suppl 1):355. [ClinicalTrials.gov: NCT01030783]

Eisen TQG, Sternberg CN, Tomczak P, Harza M, Jinga V, Esteves B, et al. Detailed comparison of the safety of tivozanib versus sorafenib in patients with advanced/metastatic renal cell carcinoma (mRCC) from a phase 3 trial. *Annals of Oncology* 2012;**9**:ix262-3. [ClinicalTrials.gov: NCT01030783]

Hutson T, Nosov D, Tomczak P, Bondarenko I, Lipatov ON, Sternberg CN, et al. Tivozanib vs sorafenib targeted therapy for advanced renal cell carcinoma: final results of a phase III trial (901) and efficacy results of a 2nd line tivozanib extension study (902). *Journal of Clinical Oncology: ASCO annual meeting proceedings* 2015;**33**(15 Suppl 1):4557. [ClinicalTrials.gov: NCT01030783]

Hutson TE, Nosov D, Eisen T, Lipatov ON, Tomczak P, Alyasova A, et al. Subgroup analyses of a phase III trial comparing tivozanib hydrochloride versus sorafenib as initial targeted therapy for patients (pts) with metastatic renal cell carcinoma (mRCC). *Journal of Clinical Oncology* 2013;**31**(6 Suppl 1):354. [ClinicalTrials.gov: NCT01030783]

Hutson TE, Nosov D, Harza M, Esteves B, Strahs AL, Berkenblit A, et al. Fewer dose adjustments with tivozanib vs. sorafenib in the phase III TIVO-1 study in advanced renal cell carcinoma (RCC). *Urologe* 2013;**52**(1 Suppl 1):89-90. [ClinicalTrials.gov: NCT01030783]

Hutson TE, Nosov D, Harza M, Esteves B, Strahs AL, Berkenblit A, et al. Rates of dose adjustment in patients treated with tivozanib versus sorafenib in the phase III TIVO-1 study. *Journal of Clinical Oncology: ASCO annual meeting proceedings* 2013;**31**(15 Suppl 1):4564. [ClinicalTrials.gov: NCT01030783]

Mehta A, Sonpavde G, Escudier B. Tivozanib for the treatment of renal cell carcinoma: results and implications of the TIVO-1 trial. *Future Oncology* 2014;**10**(11):1819-26. [ClinicalTrials.gov: NCT01030783]

Motzer R, Nosov D, Eisen T, Bondarenko I, Lesovoy V, Lipatov O, et al. Tivozanib versus sorafenib as initial targeted therapy for patients with advanced renal cell carcinoma: results from a phase III randomized, open-label, multicenter trial. *BJU International* 2012;**2**:14. [ClinicalTrials.gov: NCT01030783]

Motzer R, Nosov D, Eisen T, Lipatov O, Tomczak P, Lyulko O, et al. Tivozanib hydrochloride versus sorafenib as initial targeted therapy for patients with advanced renal cell carcinoma: results from the phase III randomized, open-label, multicenter TIVO-1 trial. *European Urology, Supplements* 2012;**11**(5):185. [ClinicalTrials.gov: NCT01030783]

Motzer RJ, Eisen T, Hutson TE, Szczylik C, Krygowski M, Strahs AL, et al. Overall survival results from a phase III study of tivozanib hydrochloride versus sorafenib in patients with renal cell carcinoma. *Journal of Clinical Oncology* 2013;**31**(6 Suppl 1):350. [ClinicalTrials.gov: NCT01030783]

* Motzer RJ, Nosov D, Eisen T, Bondarenko I, Lesovoy V, Lipatov O, et al. Tivozanib versus sorafenib as initial targeted therapy for patients with metastatic renal cell carcinoma: results from a phase III trial. *Journal of Clinical Oncology* 2013;**31**(30):3791-9. [ClinicalTrials.gov: NCT01030783]

Motzer RJ, Nosov D, Eisen T, Bondarenko I, Lesovoy V. Tivozanib versus sorafenib as initial targeted therapy for patients with advanced renal cell carcinoma: results from a phase III randomized, open-label, multicenter trial. In: Clinical Advances

in Hematology & Oncology. Vol. 10. 2012:4-6. [ClinicalTrials.gov: NCT01030783]

Motzer RJ, Nosov D, Tomczak P, Berkenblit A, Esteves B, Strahs AL, et al. Efficacy and safety data from patients with advanced renal cell cancer treated with tivozanib hydrochloride after progression on sorafenib. *Journal of Clinical Oncology. Conference* 2013;**31**(6 Suppl 1):364. [ClinicalTrials.gov: NCT01030783]

Sternberg CN, Eisen T, Tomczak P, Strahs AL, Esteves B, Berkenblit A, et al. Tivozanib in patients treatment-naive for mRCC in the TIVO-1 study. *Der Urologe* 2013;**52**(1 Suppl 1):89. [ClinicalTrials.gov: NCT01030783]

Sternberg CN, Eisen T, Tomczak P, Strahs AL, Esteves B, Berkenblit A, et al. Tivozanib in patients treatment-naive for metastatic renal cell carcinoma: a subset analysis of the phase III TIVO-1 study. *Journal of Clinical Oncology* 2013;**31**(15 Suppl 1):4513. [ClinicalTrials.gov: NCT01030783]

NCT01064310 *{published and unpublished data}*

Cella D, Kaiser K, Beaumont J, Diaz J, McCann L, Mehmud F, et al. Quality of life (QoL) among renal cell carcinoma (RCC) patients in a randomized double blind cross-over patient preference study of pazopanib (P) versus sunitinib (S). *Annals of Oncology* 2012;**9**:ix261-2. [ClinicalTrials.gov: NCT01064310]

* Escudier B, Porta C, Bono P, Powles T, Eisen T, Sternberg CN, et al. Randomized, controlled, double-blind, cross-over trial assessing treatment preference for pazopanib versus sunitinib in patients with metastatic renal cell carcinoma: PISCES Study. *Journal of Clinical Oncology* 2014;**32**(14):1412-8.

Lai JS, Beaumont JL, Diaz J, Khan S, Cella D. Validation of a short questionnaire to measure symptoms and functional limitations associated with hand-foot syndrome and mucositis in patients with metastatic renal cell carcinoma. *Cancer* 2016;**122**(2):287-95.

NCT01108445 *{published data only}*

Armstrong AJ, Broderick S, Eisen T, Stadler WM, Jones RJ, Garcia JA, et al. Final clinical results of a randomized phase II international trial of everolimus vs. Sunitinib in patients with metastatic non-clear cell renal cell carcinoma (ASPEN). *Journal of Clinical Oncology* 2015;**33**(15 Suppl 1):4507. [ClinicalTrials.gov: NCT01064310]

* Armstrong AJ, Halabi S, Eisen T, Broderick S, Stadler WM, Jones RJ, et al. Everolimus versus sunitinib for patients with metastatic non-clear cell renal cell carcinoma (ASPEN): a multicentre, open-label, randomised phase 2 trial. *Lancet Oncology* 2016;**17**(3):378-88. [ClinicalTrials.gov: NCT01064310]

Armstrong AJ, Halabi S, Eisen T, Stadler WM, Jones RR, Vaishampayan UN, et al. ASPEN: a randomized phase II trial of everolimus versus sunitinib in patients with metastatic non-clear cell renal cell carcinoma. In: Journal of Clinical Oncology: ASCO annual meeting proceedings. Vol. 31. 2013. [ClinicalTrials.gov: NCT01064310]

Armstrong AJ, Nixon AB, Carmack A, Eisen T, Stadler WM, Jones RJ, et al. Angiokines associated with outcomes after sunitinib or everolimus treatment in patients with non-clear cell renal cell carcinoma. *Clinical Cancer Research* 2021;**16**:16. [ClinicalTrials.gov: NCT01064310]

NCT01274273 *{published data only}*

Donskov F, Jensen NV, Smidt-Hansen T, Brondum L, Geertsen PF. A randomized phase II trial of interleukin-2/interferon-alpha plus bevacizumab versus interleukin-2/interferon-alpha in metastatic renal cell carcinoma (mRCC): results from the Danish renal cancer group (DARENCA) study 1. *Journal of Clinical Oncology* 2016;**34**(15 Suppl):4563. [ClinicalTrials.gov: NCT01274273]

* Donskov F, Jensen NV, Smidt-Hansen T, Brøndum L, Geertsen P. A randomized phase II trial of interleukin-2 and interferon-α plus bevacizumab versus interleukin-2 and interferon-α in metastatic renal-cell carcinoma (mRCC): results from the Danish renal cancer group (DaRenCa) study-1. *Acta Oncologica* 2018;**57**(5):589-94. [ClinicalTrials.gov: NCT01274273]

NCT01392183 *{published data only}*

Rini BI, Gruenwald V, Fishman MN, Melichar B, Ueda T, Bair AH, et al. Axitinib with or without dose titration for first-line metastatic renal cell carcinoma (mRCC): unblinded results from a randomized phase II study. In: Journal of Clinical Oncology: ASCO annual meeting proceedings. Vol. 31. 2013. [ClinicalTrials.gov: NCT01392183]

Rini BI, Melichar B, Ueda T, Fishman MN, Pithavala Y, Chen Y, et al. Randomised phase II trial of axitinib with or without dose titration in first-line metastatic renal cell carcinoma (mRCC): efficacy and pharmacokinetic (PK) analyses. *European Journal of Cancer* 2013;**2**:S657. [ClinicalTrials.gov: NCT01392183]

* Tannir NM, Msaouel P, Ross JA, Devine CE, Chandramohan A, Gonzalez GM, et al. Temsirolimus versus pazopanib (TemPa) in patients with advanced clear-cell renal cell carcinoma and poor-risk features: a randomized phase II trial. *European Urology Oncology* 2019;**3**(5):687-694. [ClinicalTrials.gov: NCT01392183]

Tannir NM, Msaouel P, Ross JA, Devine CE, Chandramohan A, Gonzalez GM, et al. Temsirolimus versus pazopanib (TemPa) in patients with advanced clear-cell renal cell carcinoma and poor-risk features: a randomized phase II trial. *European Urology Oncology* 2020;**3**(5):687-94. [ClinicalTrials.gov: NCT01392183]

Tannir NM, Msaouel P, Ross JA, Devine CE, Chandramohan A, Gonzalez GM, et al. Temsirolimus versus pazopanib (TemPa) in patients with advanced clear-cell renal cell carcinoma and poor-risk features: a randomized phase II trial. *European Urology Oncology* 2020;**3**(5):687-94. [ClinicalTrials.gov: NCT01392183]

Tannir NM, Ross JA, Devine CE, Chandramohan A, Wang X, Lim ZD, et al. A randomized phase II trial of pazopanib (PAZ) versus temsirolimus (TEM) in patients (PTS) with advanced clear-cell renal cell carcinoma (aCCRCC) of intermediate and poor-risk (the TemPa trial). *Journal of Clinical Oncology* 2018;**36**(6 Suppl 1):583. [ClinicalTrials.gov: NCT01392183]

NCT01481870 *{published data only}*

Tomita Y, Naito S, Sassa N, Takahashi A, Kondo T, Koie T, et al. Sunitinib versus sorafenib as first-line therapy followed by sorefenib and sunitinib for patients with metastatic renal

cell carcinoma (RCC) with clear cell histology: a multicenter randomized trial, CROSS-J-RCC. *Journal of Clinical Oncology* 2017;**35**(6):469. [ClinicalTrials.gov: NCT01481870]

Tomita Y, Naito S, Sassa N, Takahashi A, Kondo T, Koie T, et al. Sunitinib versus sorafenib as first-line therapy for patients with metastatic renal cell carcinoma with favorable or intermediate MSKCC risk factors: a multicenter randomized trial, CROSS-J-RCC. *Journal of Clinical Oncology* 2014;**32**(4 Suppl 1):502. [ClinicalTrials.gov: NCT01481870]

* Tomita Y, Naito S, Sassa N, Takahashi A, Kondo T, Koie T, et al. Sunitinib versus sorafenib as initial targeted therapy for mCC-RCC with favorable/intermediate risk: multicenter randomized trial CROSS-J-RCC. *Clinical Genitourinary Cancer* 2020;**18**(4):e374-85. [ClinicalTrials.gov: NCT01481870]

NCT01613846 *{published data only}*

Gschwend J, Beeker A, De Santis M, Brehmer B, Zaiss MR, Schirrmacher-Memmel S, et al. Phase III randomized sequential open-label study to evaluate the efficacy and safety of sorafenib followed by pazopanib versus pazopanib followed by sorafenib in the treatment of advanced/metastatic renal cell carcinoma (SWITCH-2 study). In: Journal of Clinical Oncology: ASCO annual meeting proceedings. Vol. 31. 2013. [ClinicalTrials.gov: NCT01613846]

* Retz M, Bedke J, Bogemann M, Grimm MO, Zimmermann U, Muller L, et al. SWITCH II: phase III randomized, sequential, open-label study to evaluate the efficacy and safety of sorafenib-pazopanib versus pazopanib-sorafenib in the treatment of advanced or metastatic renal cell carcinoma (AUO AN 33/11). *European Journal of Cancer* 2019;**107**:37-45. [ClinicalTrials.gov: NCT01613846]

Retz M, Bedke J, Herrmann E, Grimm MO, Zimmermann U, Muller L, et al. Phase III randomized, sequential, open-label study to evaluate the efficacy and safety of sorafenib-pazopanib versus pazopanib-sorafenib in the treatment of metastatic renal cell carcinoma (SWITCH-II). *Annals of Oncology* 2017;**28 Suppl 5**:v295. [ClinicalTrials.gov: NCT01613846]

Rexer H. First line therapy of advanced or metastasized renal cell carcinoma: open, randomized phase III sequence study to examine the effectiveness and tolerance of sorafenib followed by pazopanib versus pazopanib followed by sorafenib in the first-line treatment of patients with advanced or metastasized renal cell carcinoma (SWITCH-2-AN 33/11).. *Urologe - Ausgabe A* 2012;**51**(5):724-6. [ClinicalTrials.gov: NCT01613846]

Rexer H. First-line therapy of advanced or metastasized renal cell carcinoma: phase III, open, randomized sequence study to examine efficacy and tolerance of sorafenib followed by pazopanib versus pazopanib followed by sorafenib in the first-line treatment of patients with advanced or metastasized renal cell carcinoma (SWITCH-2 - AN 33/11). *Der Urologe* 2014;**53**(5):735-8. [ClinicalTrials.gov: NCT01613846]

Schirrmacher-Memmel S, Retz M, De Santis M, Brehmer B, Zaiss M, Beeker A, et al. Phase III randomized sequential open-label study to evaluate the efficacy and safety of sorafenib followed by pazopanib vs pazopanib followed by sorafenib in the treatment of advanced / metastatic renal cell carcinoma-design of the SWITCH-2 Study. *Onkologie* 2013;**7**:75. [ClinicalTrials.gov: NCT01613846]

NCT01835158 *{published data only}*

Alba SO, Jose MB, Silvia FC. Cabozantinib for the treatment of advanced renal cell carcinoma in treatment-naive adults. *European Journal of Clinical Pharmacy* 2019;**21**(3):160-3. [ClinicalTrials.gov: NCT01835158]

Chen RC, Choueiri TK, Feuilly M, Meng J, Lister J, Marteau F, et al. Quality-adjusted survival with first-line cabozantinib or sunitinib for advanced renal cell carcinoma in the CABOSUN randomized clinical trial (Alliance). *Cancer* 2020;**126**(24):5311-18. [ClinicalTrials.gov: NCT01835158]

Chen RC, Feuilly M, Meng J, Lister J, Marteau F, Morris MJ, et al. Quality-adjusted time without symptoms or toxicity (Q-TWiST): analysis of cabozantinib (Cabo) vs sunitinib (Sun) in patients with advanced renal cell carcinoma (aRCC) of intermediate or poor risk (Alliance A031203). *Journal of Clinical Oncology* 2018;**36**(15 Suppl 1):4556. [ClinicalTrials.gov: NCT01835158]

Choueiri TK, Halabi S, Sanford B, Hahn O, Michaelson MD, Walsh M, et al. CABOzantinib versus SUNitinib (CABOSUN) as initial targeted therapy for patients with metastatic renal cell carcinoma (mRCC) of poor and intermediate risk groups: results from ALLIANCE A031203 trial. *Annals of Oncology* 2016;**27 Suppl 6**:vi566. [ClinicalTrials.gov: NCT01835158]

Choueiri TK, Halabi S, Sanford BL, Hahn O, Michaelson MD, Walsh MK, et al. Cabozantinib versus sunitinib as initial targeted therapy for patients with metastatic renal cell carcinoma of poor or intermediate risk: the Alliance A031203 CABOSUN trial. *Journal of Clinical Oncology* 2017;**35**(6):591-7. [ClinicalTrials.gov: NCT01835158]

Choueiri TK, Hessel C, Halabi S, Sanford B, Hahn O, Michaelson MD, et al. Progression-free survival (PFS) by independent review and updated overall survival (OS) results from Alliance A031203 trial (CABOSUN): cabozantinib versus sunitinib as initial targeted therapy for patients (pts) with metastatic renal cell carcinoma (mRCC). *Annals of Oncology* 2017;**28 Suppl 5**:v623. [ClinicalTrials.gov: NCT01835158]

* Choueiri TK, Hessel C, Halabi S, Sanford B, Michaelson MD, Hahn O, et al. Cabozantinib versus sunitinib as initial therapy for metastatic renal cell carcinoma of intermediate or poor risk (Alliance A031203 CABOSUN randomised trial): progression-free survival by independent review and overall survival update. *European Journal of Cancer* 2018;**94**:115-25. [ClinicalTrials.gov: NCT01835158]

Feldman DR, Feuilly M, Meng J, Lister J, Marteau F, Morris MJ, et al. ECOG score analysis as a proxy for healthrelated quality of life assessment in patients with poor or intermediate risk metastatic renal cell carcinoma from the CABOSUN trial (Alliance A031203). In: Journal of Clinical Oncology (Conference). Vol. 36. 2018. [ClinicalTrials.gov: NCT01835158]

George DJ, Hessel C, Halabi S, Michaelson MD, Hahn O, Walsh M, et al. Cabozantinib versus sunitinib for untreated patients with advanced renal cell carcinoma of intermediate or poor risk: subgroup analysis of the Alliance A031203 CABOSUN

trial. *Oncologist* 2019;**24**(11):1497-1501. [ClinicalTrials.gov: NCT01835158]

George DJ, Hessel C, Halabi S, Sanford BL, Dror Michaelson M, Hahn OM, et al. Cabozantinib versus sunitinib for previously untreated patients with advanced renal cell carcinoma (RCC) of intermediate or poor risk: subgroup analysis of progression-free survival (PFS) and objective response rate (ORR) in the Alliance A031203 CABOSUN trial. *Journal of Clinical Oncology* 2018;**36**(6 Suppl 1):582. [ClinicalTrials.gov: NCT01835158]

NCT01984242 *{published data only}*

Atkins MB, McDermott DF, Powles T, Motzer RJ, Rini BI, Fong L, et al. IMmotion150: a phase II trial in untreated metastatic renal cell carcinoma (mRCC) patients (pts) of atezolizumab (atezo) and bevacizumab (bev) vs and following atezo or sunitinib (sun). *Journal of Clinical Oncology* 2017;**35**(15):4505. [ClinicalTrials.gov: NCT01984242]

Grunwald V, McDermott DF, Atkins M, Motzer R, Rini B, Escudier B, et al. A phase II study of atezolizumab (atezo) with or without bevacizumab (bev) vs sunitinib (sun) in untreated metastatic renal cell carcinoma (mRCC) patients (pts). *Oncology Research and Treatment* 2017;**40 Suppl 3**:113-4. [ClinicalTrials.gov: NCT01984242]

McDermott DF, Atkins MB, Motzer RJ, Rini BI, Escudier BJ, Fong L, et al. A phase II study of atezolizumab (atezo) with or without bevacizumab (bev) versus sunitinib (sun) in untreated metastatic renal cell carcinoma (mRCC) patients (pts). *Journal of Clinical Oncology* 2017;**35**(6):431. [ClinicalTrials.gov: NCT01984242]

* McDermott DF, Huseni MA, Atkins MB, Motzer RJ, Rini BI, Escudier B, et al. Clinical activity and molecular correlates of response to atezolizumab alone or in combination with bevacizumab versus sunitinib in renal cell carcinoma. *Nature Medicine* 2018;**24**(6):749-57. [ClinicalTrials.gov: NCT01984242]

Pal S, Powles T, McDermott D, Rini B, Motzer R, Atkins M, et al. IMmotion150: novel radiological endpoints and updated data from a randomized phase II trial investigating atezolizumab with or without bevacizumab vs sunitinib in untreated metastatic renal cell carcinoma. *Kidney Cancer* 2018;**2 Suppl 1**:S22-3. [ClinicalTrials.gov: NCT01984242]

Pal SK, McDermott DF, Atkins MB, Escudier B, Rini BI, Motzer RJ, et al. Patient-reported outcomes (PROs) in IMmotion150: atezolizumab (atezo) alone or with bevacizumab (bev) versus sunitinib (sun) in first-line metastatic renal cell carcinoma (mRCC). *Journal of Clinical Oncology* 2019;**37**(15 Suppl):4515. [ClinicalTrials.gov: NCT01984242]

Pal SK, McDermott DF, Atkins MB, Escudier B, Rini BI, Motzer RJ, et al. Patient-reported outcomes in a phase 2 study comparing atezolizumab alone or with bevacizumab vs sunitinib in previously untreated metastatic renal cell carcinoma. *BJU International* 2020;**126**(1):73-82. [ClinicalTrials.gov: NCT01984242]

Powles T, Atkins MB, Escudier B, Motzer RJ, Rini BI, Fong L, et al. Efficacy and safety of atezolizumab plus bevacizumab following disease progression on atezolizumab or sunitinib monotherapy in patients with metastatic renal cell carcinoma in IMmotion150: a randomized phase 2 clinical trial. *European Urology* 2021;**79**(5):665-73. [ClinicalTrials.gov: NCT01984242]

Powles T, McDermott DF, Rini B, Motzer RJ, Atkins MB, Fong L, et al. IMmotion150: novel radiological endpoints and updated data from a randomized phase II trial investigating atezolizumab (atezo) with or without bevacizumab (bev) vs sunitinib (sun) in untreated metastatic renal cell carcinoma (mRCC). *Annals of Oncology* 2017;**28**(Suppl 5):v624. [ClinicalTrials.gov: NCT01984242]

NCT02231749 *{published data only}*

Albiges L, Tannir N, Burotto M, McDermott DF, Plimack ER, Barthelemy P, et al. Nivolumab + ipilimumab (N+I) vs sunitinib (S) for first-line treatment of advanced renal cell carcinoma (aRCC) in CheckMate 214: 4-year follow-up and subgroup analysis of patients (pts) without nephrectomy. *Annals of Oncology* 2020;**31 Suppl 4**:S559-60. [ClinicalTrials.gov: NCT02231749]

Albiges L, Tannir NM, Burotto M, McDermott D, Plimack ER, Barthelemy P, et al. First-line nivolumab plus ipilimumab versus sunitinib in patients without nephrectomy and with an evaluable primary renal tumor in the CheckMate 214 trial. *European Urology* 2021;**81**(3):266-271. [ClinicalTrials.gov: NCT02231749]

* Albiges L, Tannir NM, Burotto M, McDermott D, Plimack ER, Barthelemy P, et al. Nivolumab plus ipilimumab versus sunitinib for first-line treatment of advanced renal cell carcinoma: extended 4-year follow-up of the phase III CheckMate 214 trial. *ESMO Open* 2020;**5**(6):e001079. [ClinicalTrials.gov: NCT02231749]

Ali S, Camarero J, Hennik PV, Bolstad B, Sommerfelt Gronvold M, Syvertsen C, et al. European Medicines Agency extension of indication to include the combination immunotherapy cancer drug treatment with nivolumab (Opdivo) and ipilimumab (Yervoy) for adults with intermediate/poor-risk advanced renal cell carcinoma. *ESMO Open* 2020;**5**(6):e000798. [ClinicalTrials.gov: NCT02231749]

Branchoux S, Négrier S, de Peretti C, Malcolm B, May J, Marié L, et al. PCN214 cost-effectiveness analysis of nivolumab in combination with ipilimumab versus sunitinib for the first-line treatment of intermediate- to poor-risk advanced renal cell carcinoma in France. In: Value in Health. Vol. 22. 2019:S477. [ClinicalTrials.gov: NCT02231749]

Cakar E, May J, Malcolm B, Gooden KM, Klijn S. PCN421 stability of lifetime overall survival estimates of nivolumab+ipilimumab IN first-line advanced/metastatic intermediate- or poor-risk renal cell carcinoma. In: Value in Health. Vol. 22. 2019:S517-8. [ClinicalTrials.gov: NCT02231749]

Cella D, Escudier B, Ivanescu C, Mauer M, Lord-Bessen J, Gooden K. Quality of life in previously untreated patients with advanced renal cell carcinoma (aRCC) in CheckMate 214: updated results. *Annals of Oncology* 2019;**30 Suppl 5**:v383-4. [ClinicalTrials.gov: NCT02231749]

Cella D, Escudier B, Ivanescu C, Maurer M, Lord-Bessen J, Gooden K. Quality of life in previously untreated patients with advanced renal cell carcinoma (aRCC) in CheckMate 214: updated results. *Kidney Cancer* 2020;**4 Suppl 1**:S49-50. [ClinicalTrials.gov: NCT02231749]

Cella D, Escudier B, Saggi SS, Blum S, Ejzykowicz F, Ivanescu C. Time to deterioration in quality of life in previously untreated patients with advanced renal cell carcinoma (aRCC) in CheckMate 214. *Annals of Oncology* 2020;**31 Suppl 4**:S562. [ClinicalTrials.gov: NCT02231749]

Cella D, Grunwald V, Escudier B, Hammers HJ, George S, Nathan P, et al. Patient-reported outcomes of patients with advanced renal cell carcinoma treated with nivolumab plus ipilimumab versus sunitinib (CheckMate 214): a randomised, phase 3 trial. *Lancet Oncology* 2019;**20**(2):297-310. [ClinicalTrials.gov: NCT02231749]

Cella D, Grunwald V, Escudier B, Hammers HJ, George S, Nathan P, et al. Quality of life in patients with advanced renal cell carcinoma in the randomized, open-label CheckMate 214 trial. *Journal of Clinical Oncology* 2018;**36**(15):3073. [ClinicalTrials.gov: NCT02231749]

Characterization of response to nivolumab plus ipilimumab (N1I) or sunitinib (S) in patients (Pts) with previously untreated advanced renal cell carcinoma (aRCC): CheckMate 214. Annals of Oncology 2018;**29**:viii309-10. [ClinicalTrials.gov: NCT02231749]

Choueiri TK, Stwalley B, Huo S, May JR, Malcolm B, Nickel K, et al. Comparison of long-term survival and cost-effectiveness (CE) of first-line (1L) treatment options in advanced renal cell carcinoma (aRCC) with intermediate or poor (I/P) prognostic risk. *Annals of Oncology* 2020;**31 Suppl 4**:S563-4. [ClinicalTrials.gov: NCT02231749]

Correction to Lancet Oncol 2019; 20: 1370–85 (The Lancet Oncology (2019) 20(10) (1370–1385), (S1470204519304139), (10.1016/S1470-2045(19)30413-9)). Lancet Oncology 2020;**21**(6):e304. [ClinicalTrials.gov: NCT02231749]

Correction to Lancet Oncology 2019; 20: 1370-85 (The Lancet Oncology (2019) 20(10) (1370-1385), (S1470204519304139), (10.1016/S1470-2045(19)30413-9)). Lancet Oncology 2020;**21**(6):e304. [ClinicalTrials.gov: NCT02231749]

Correction to nivolumab plus ipilimumab versus sunitinib in first-line treatment for advanced renal cell carcinoma: extended follow-up of efficacy and safety results from a randomised, controlled, phase 3 trial (The Lancet Oncology (2019) 20(10) (1370-1385), (S1470204519304139), (10.1016/S1470-2045(19)30413-9)). Lancet Oncology 2019;**20**(10):e559. [ClinicalTrials.gov: NCT02231749]

Correction to nivolumab plus ipilimumab versus sunitinib in first-line treatment for advanced renal cell carcinoma: extended follow-up of efficacy and safety results from a randomised, controlled, phase 3 trial (The Lancet Oncology (2019) 20(10) (1370-1385), (S1470204519304139), (10.1016/S1470-2045(19)30413-9)). Lancet Oncology 2019;**20**(10):e559. [ClinicalTrials.gov: NCT02231749]

Correction to nivolumab plus ipilimumab versus sunitinib in first-line treatment for advanced renal cell carcinoma: extended follow-up of efficacy and safety results from a randomised, controlled, phase 3 trial (The Lancet Oncology (2019) 20(10) (1370â€“1385), (S1470204519304139), (10.1016/S1470-2045(19)30413-9)). Lancet Oncology 2019;**20**(10):e559. [ClinicalTrials.gov: NCT02231749]

Correction: survival outcomes and independent response assessment with nivolumab plus ipilimumab versus sunitinib in patients with advanced renal cell carcinoma: 42-month follow-up of a randomized phase 3 clinical trial. Journal for Immunotherapy of Cancer 2021;**9**(5). [ClinicalTrials.gov: NCT02231749]

Di Rienzo P, Zarrelli L, Mennini FS, Marcellusi A, Bini C, Malcolm B et al. PCN170 cost-effectiveness of nivolumab in combination with ipilimumab compared to sunitinib in the first-line treatment of advanced or metastatic intermediate- or poor-risk renal cell carcinoma in Italy. In: Value in Health. Vol. 22. 2019:S469. [ClinicalTrials.gov: NCT02231749]

Escudier B, Motzer RJ, Tannir NM, Porta C, Tomita Y, Maurer MA, et al. Efficacy of nivolumab plus ipilimumab according to number of IMDC risk factors in CheckMate 214. *European Urology* 2019;**13**:13. [ClinicalTrials.gov: NCT02231749]

Escudier B, Motzer RJ, Tannir NM, Porta C, Tomita Y, Maurer MA, et al. Efficacy of nivolumab plus ipilimumab according to number of IMDC risk factors in CheckMate 214. *European Urology* 2020;**77**(4):449-53. [ClinicalTrials.gov: NCT02231749]

Escudier B, Tannir NM, McDermott DF, Frontera OA, Melichar B, Plimack ER, et al. CheckMate 214: efficacy and safety of nivolumab 1 ipilimumab (N1I) v sunitinib (S) for treatment-naïve advanced or metastatic renal cell carcinoma (mRCC), including IMDC risk and PD-L1 expression subgroups. *Annals of Oncology* 2017;**28**:v621-2. [ClinicalTrials.gov: NCT02231749]

First-line nivolumab plus ipilimumab is associated with lower costs per responder versus sunitinib among patients with advanced renal cell carcinoma. Journal for Immunotherapy of Cancer 2018;**6**. [ClinicalTrials.gov: NCT02231749]

George S, Betts K, Yang S, Du E, Johansen J, Rao S, et al. Nivolumab plus ipilimumab is associated with lower number needed to treat compared with sunitinib for preventing death in advanced renal cell carcinoma. In: Journal for ImmunoTherapy of Cancer. Conference: 33rd Annual Meeting and Pre Conference Programs of the Society for Immunotherapy of Cancer, SITC. Vol. 6 Suppl 1. 2018. [ClinicalTrials.gov: NCT02231749]

Grimm MO, McDermott DF, Choueiri TK, Motzer RJ, Frontera OA, George S, et al. CheckMate 214 post hoc analyses of nivolumab plus ipilimumab or sunitinib in IMDC intermediate/poor-risk patients with previously untreated advanced renal cell carcinoma (aRCC) with sarcomatoid features (sRCC). *Oncology Research and Treatment* 2020;**43**:70-1. [ClinicalTrials.gov: NCT02231749]

Gruenwald V, Cella D, Escudier BJ, Hammers HJ, George S, Nathan P, et al. Quality of life in patients with advanced renal cell carcinoma in the randomized, open-label CheckMate 214

trial. *Oncology Research and Treatment* 2018;**41 Suppl 4**:42. [ClinicalTrials.gov: NCT02231749]

Gruenwald V, Choueiri T, Rini B, Powles T, George S, Grimm MO, et al. Association between depth of response and overall survival: exploratory analysis in patients with previously untreated advanced renal cell carcinoma (aRCC) in CheckMate 214. *Kidney Cancer* 2020;**4**:S5-6. [ClinicalTrials.gov: NCT02231749]

Gruenwald V, Choueiri TK, Rini BI, Powles T, George S, Grimm MO, et al. Association between depth of response (DepOR) and overall survival (OS): exploratory analysis of nivolumab + ipilimumab (N+I) vs sunitinib (S) in patients with previously untreated advanced renal cell carcinoma (aRCC) in CheckMate 214. *Oncology Research and Treatment* 2020;**43**:74-5. [ClinicalTrials.gov: NCT02231749]

Gruenwald V, Choueiri TK, Rini BI, Powles T, George S, Grimm MO, et al. Association between depth of response and overall survival: exploratory analysis in patients with previously untreated advanced renal cell carcinoma (aRCC) in CheckMate 214. *Annals of Oncology* 2019;**30 Suppl 5**:v382-3. [ClinicalTrials.gov: NCT02231749]

Hammers H, Sternberg C, McDermott DF, Larkin J, Ravaud A, Rini B, et al. A phase 3, randomized, openlabel study of nivolumab combined with ipilimumab versus sunitinib monotherapy in subjects with previously untreated metastatic renal cell carcinoma. *BJU International* 2014;**114**:9. [ClinicalTrials.gov: NCT02231749]

Hammers HJ, Plimack ER, Sternberg C, McDermott DF, Larkin JM, Ravaud A, et al. CheckMate 214: a phase III, randomized, open-label study of nivolumab combined with ipilimumab versus sunitinib monotherapy in patients with previously untreated metastatic renal cell carcinoma. In: Journal of Clinical Oncology. Meeting Abstract | 2015 ASCO Annual Meeting I. Vol. 33. 2015. [ClinicalTrials.gov: NCT02231749]

Jena R. CheckMate 214 trial: immune checkpoint regulators for advanced renal cell carcinoma. *Indian Journal of Urology* 2020;**36**(3):231-3. [ClinicalTrials.gov: NCT02231749]

Lakhdari K, Young KC, Berling M, Gooden KM, Holdgate O. PCN240 evaluating partitioned survival model and response-based modeling approaches for use in cost-effectiveness analysis: estimating and validating survival outcomes. In: Value in Health. Vol. 22. 2019:S101-2. [ClinicalTrials.gov: NCT02231749]

McDermott DF, Choueiri TK, Motzer RJ, Aren OR, George S, Powles T, et al. CheckMate 214 post-hoc analyses of nivolumab plus ipilimumab or sunitinib in IMDC intermediate/poor-risk patients with previously untreated advanced renal cell carcinoma with sarcomatoid features. *Journal of Clinical Oncology* 2019;**37**(15):4513. [ClinicalTrials.gov: NCT02231749]

McDermott DF, Rini BI, Motzer RJ, Tannir NM, Escudier B, Kollmannsberger CK, et al. Treatment-free interval (TFI) following discontinuation of first-line nivolumab plus ipilimumab (N1I) or sunitinib (S) in patients (pts) with advanced renal cell carcinoma (aRCC): CheckMate 214 analysis. *Annals of Oncology* 2018;**29 Suppl 8**:viii309. [ClinicalTrials.gov: NCT02231749]

McDermott DF, Rini BI, Motzer RJ, Tannir NM, Escudier B, Kollmannsberger CK, et al. Treatment-free survival (TFS) after discontinuation of first-line nivolumab (NIVO) plus ipilimumab (IPI) or sunitinib (SUN) in intention-to-treat (ITT) and IMDC favorable-risk patients (pts) with advanced renal cell carcinoma (aRCC) from CheckMate 214. *Journal of Clinical Oncology* 2019;**37**(7):564. [ClinicalTrials.gov: NCT02231749]

Motzer RJ, Choueiri TK, May J, Kwon Y, Rusibamayila N, Botteman M, et al. Long-term trend of qualityadjusted time without symptoms or toxicities (Q-TWiST) of nivolumab +ipilimumab (N+I) versus sunitinib (SUN) for the firstline treatment of advanced renal cell carcinoma (aRCC). *Journal of Clinical Oncology* 2021;**39**(15 Suppl):6568. [ClinicalTrials.gov: NCT02231749]

Motzer RJ, Choueiri TK, McDermott DF, Powles T, Yao J, Ammar R, et al. Biomarker analyses from the phase III CheckMate 214 trial of nivolumab plus ipilimumab (N+I) or sunitinib (S) in advanced renal cell carcinoma (aRCC). *Journal of Clinical Oncology: ASCO annual meeting proceedings* 2020;**38**(15 Suppl):5009. [ClinicalTrials.gov: NCT02231749]

Motzer RJ, Escudier B, McDermott DF, Aren Frontera O, Melichar B, Powles T, et al. Survival outcomes and independent response assessment with nivolumab plus ipilimumab versus sunitinib in patients with advanced renal cell carcinoma: 42-month follow-up of a randomized phase 3 clinical trial. *Journal for Immunotherapy of Cancer* 2020;**8**(2):e000891. [ClinicalTrials.gov: NCT02231749]

Motzer RJ, Rini BI, McDermott DF, Aren Frontera O, Hammers HJ, Carducci MA, et al. Nivolumab plus ipilimumab versus sunitinib in first-line treatment for advanced renal cell carcinoma: extended follow-up of efficacy and safety results from a randomised, controlled, phase 3 trial. *Lancet Oncology* 2019;**20**(10):1370-85. [ClinicalTrials.gov: NCT02231749]

Motzer RJ, Tannir NM, McDermott DF, Arén Frontera O, Melichar B, Choueiri TK, et al. Nivolumab plus ipilimumab versus sunitinib in advanced renal-cell carcinoma. *New England Journal of Medicine* 2018;**378**(14):1277-90. [ClinicalTrials.gov: NCT02231749]

Motzer RJ, Tannir NM, McDermott DF, Frontera OA, Melichar B, Plimack ER, et al. Nivolumab + ipilimumab (N+I) vs sunitinib (S) for treatment-naive advanced or metastatic renal cell carcinoma (aRCC): results from CheckMate 214, including overall survival by subgroups. In: Journal for ImmunoTherapy of Cancer. Vol. 5. 2017. [ClinicalTrials.gov: NCT02231749]

NCT02231749. Metastatic renal cell carcinoma (mRCC) patients treated with sunitinib as first-line therapy. who.int/trialsearch/Trial2.aspx?TrialID=ChiCTR-TRC-13004016 first received 24 July 2013. [ClinicalTrials.gov: NCT02231749]

Oniangue-Ndza C, Gooden KM, May J, Malcolm B, Du EX, Yin L, et al. PCN111 comparison of adverse event costs of nivolumab+ipilimumab versus sunitinib for previously untreated intermediate-/poor-risk advanced renal cell

carcinoma in Switzerland. *Value in Health* 2019;**22 Suppl 3**:S457. [ClinicalTrials.gov: NCT02231749]

Paly V, George S, Youn JH, Borrill J, Ejzykowicz F, May JR, et al. POSC321 estimating long-term survivorship rates for previously untreated intermediate or poor (I/P) risk advanced renal cell carcinoma (aRCC) patients treated with nivolumab plus ipilimumab (NIVO+IPI): analyses from the CheckMate 214 trial. *Value in Health* 2022;**25**(1 Suppl):S210.

Polanco Sanchez C, Gooden KM, May J, Malcolm B, Du EX, Yin L, et al. PCN115 comparison of adverse event costs of nivolumab +ipilimumab versus sunitinib for previously untreated intermediate-/poor-risk advanced renal cell carcinoma in Spain. *Value in Health* 2019;**22 Suppl 3**:S458. [ClinicalTrials.gov: NCT02231749]

Regan MM, Atkins MB, Powles T, Werner L, Mantia C, Yang S, et al. Treatment-free survival, with and without toxicity, as a novel outcome applied to immuno-oncology agents in advanced renal cell carcinoma. *Annals of Oncology* 2019;**30 Suppl 5**:v393-4. [ClinicalTrials.gov: NCT02231749]

Regan MM, Atkins MB, Powles T, Werner L, Mantia C, Yang S, et al. Treatment-free survival, with and without toxicity, as a novel outcome applied to immuno-oncology agents in advanced renal cell carcinoma. *Annals of Oncology* 2019;**30**:v393-4. [ClinicalTrials.gov: NCT02231749]

Regan MM, Jegede OA, Mantia C, Powles T, Werner L, Huo S, et al. Treatment-free survival, with and without toxicity, after immuno-oncology vs targeted therapy for advanced renal cell carcinoma (aRCC): 42-month results of CheckMate 214. *Annals of Oncology* 2020;**31 Suppl 4**:S561. [ClinicalTrials.gov: NCT02231749]

Regan MM, Jegede OA, Mantia CM, Powles T, Werner L, Motzer RJ, et al. Treatment-free survival after immune checkpoint inhibitor therapy versus targeted therapy for advanced renal cell carcinoma: 42-month results of the CheckMate 214 trial. *Clinical Cancer Research* 2021;**27**(24):6687-6695. [ClinicalTrials.gov: NCT02231749]

Rexer H. Therapy of untreated local advanced or metastatic renal cell carcinoma: phase III, randomized, open-label study of nivolumab combined with ipilimumab versus sunitinib monotherapy in subjects with previously untreated, local advanced or metastatic renal cell carcinoma (CheckMate 214 - AN 36/15 of the AUO).. *Urologe* 2015;**54**(10):1443-5. [ClinicalTrials.gov: NCT02231749]

Rini BI, Tannir NM, Escudier B, McDermott DF, Grimm MO, Porta C, et al. Characterization of response to nivolumab plus ipilimumab (N1I) or sunitinib (S) in patients (pts) with previously untreated advanced renal cell carcinoma (aRCC): CheckMate 214. *Annals of Oncology* 2018;**29 Suppl 8**:viii309-10. [ClinicalTrials.gov: NCT02231749]

Rizzo A, Mollica V, Massari F. Re: Nizar M. Tannir, Sabina Signoretti, Toni K. Choueiri, et al. Efficacy and safety of nivolumab plus ipilimumab versus sunitinib in first-line treatment of patients with advanced sarcomatoid renal cell carcinoma. Clin Cancer Res. In press. https://doi.org/10.1158/1078-0432.ccr-20-2063. *European Urology Oncology* 2020;**3**(6):804-5. [ClinicalTrials.gov: NCT02231749]

Schmidinger M, Shariat SF, Fajkovic H. Dual immune checkpoint inhibition in metastatic renal cell carcinoma: editorial re.: nivolumab plus ipilimumab versus sunitinib for first-line treatment of advanced RCC: extended 4-year follow-up of the phase III CheckMate 214 trial. *Esmo Open* 2021;**6**(1):100035. [ClinicalTrials.gov: NCT02231749]

Tannir NM, Frontera OA, Hammers HJ, Carducci MA, McDermott DF, Salman P, et al. Thirty-month follow-up of the phase III CheckMate 214 trial of first-line nivolumab + ipilimumab (N+I) or sunitinib (S) in patients (pts) with advanced renal cell carcinoma (aRCC). *Journal of Clinical Oncology* 2019;**37**(7):547. [ClinicalTrials.gov: NCT02231749]

Tannir NM, Hammers HJ, Amin A, Grimm MO, Rini BI, Mekan S, et al. Characterization of the benefit-risk profile of nivolumab + ipilimumab (N+I) v sunitinib (S) for treatment-naive advanced renal cell carcinoma (aRCC; CheckMate 214). *Journal of Clinical Oncology. Conference* 2018;**36**(6 Suppl 1):686. [ClinicalTrials.gov: NCT02231749]

Tannir NM, McDermott DF, Escudier B, Hammers HJ, Aren OR, Plimack ER, et al. Overall survival and independent review of response in CheckMate 214 with 42-month follow-up: first-line nivolumab + ipilimumab (N+I) versus sunitinib (S) in patients (pts) with advanced renal cell carcinoma (aRCC). *Journal of Clinical Oncology* 2020;**38**(6):609. [ClinicalTrials.gov: NCT02231749]

Tannir NM, Motzer RJ, Albiges L, Plimack ER, George S, Powles T, et al. Patterns of progression in patients treated with nivolumab plus ipilimumab (NIVO+IPI) versus sunitinib (SUN) for first-line treatment of advanced renal cell carcinoma (aRCC) in CheckMate 214. *Journal of Clinical Oncology. Conference* 2021;**39**(6 Suppl):313. [ClinicalTrials.gov: NCT02231749]

Tannir NM, Motzer RJ, Plimack ER, McDermott DF, Barthelemy P, Porta C, et al. Outcomes in patients (pts) with advanced renal cell carcinoma (aRCC) who discontinued (DC) first-line nivolumab + ipilimumab (N+I) or sunitinib (S) due to treatment-related adverse events (TRAEs) in CheckMate 214. *Journal of Clinical Oncology* 2019;**37**(7):581. [ClinicalTrials.gov: NCT02231749]

Tannir NM, Signoretti S, Choueiri TK, McDermott DF, Motzer RJ, Flaifel A, et al. Efficacy and safety of nivolumab plus ipilimumab versus sunitinib in first-line treatment of patients with advanced sarcomatoid renal cell carcinoma. *Clinical Cancer Research* 2020;**27**(1):78–86. [ClinicalTrials.gov: NCT02231749]

Tannir NM, Signoretti S, Choueiri TK, McDermott DF, Motzer RJ, Flaifel A, et al. Efficacy and safety of nivolumab plus ipilimumab versus sunitinib in first-line treatment of patients with advanced sarcomatoid renal cell carcinoma. *Clinical Cancer Research* 2021;**27**(1):78-86. [ClinicalTrials.gov: NCT02231749]

Tomita Y, Kondo T, Kimura G, Inoue T, Wakumoto Y, Yao M, et al. Nivolumab plus ipilimumab versus sunitinib in previously untreated advanced renal-cell carcinoma: analysis of Japanese patients in CheckMate 214 with extended follow-up. *Japanese*

Journal of Clinical Oncology 2019;**50**(1):12-9. [ClinicalTrials.gov: NCT02231749]

Tomita Y, Kondo T, Kimura G, Inoue T, Wakumoto Y, Yao M, et al. Nivolumab plus ipilimumab versus sunitinib in previously untreated advanced renal-cell carcinoma: analysis of Japanese patients in CheckMate 214 with extended follow-up. *Japanese Journal of Clinical Oncology* 2020;**50**(1):12-9. [ClinicalTrials.gov: NCT02231749]

Vyas C, Motzer RJ, Tannir NM, McDermott DF, Frontera OA, Melichar B, et al. Nivolumab + ipilimumab (N+I) vs sunitinib (S) for treatment-naive advanced or metastatic renal cell carcinoma (aRCC): results from CheckMate 214, including overall survival by subgroups. *Journal of Oncology Pharmacy Practice* 2018;**24**(5 Suppl 1):17-8. [ClinicalTrials.gov: NCT02231749]

Wan X, Zhang Y, Tan C, Zeng X, Peng L. First-line nivolumab plus ipilimumab vs sunitinib for metastatic renal cell carcinoma: a cost-effectiveness analysis. *JAMA Oncology* 2019;**5**(4):491-6. [ClinicalTrials.gov: NCT02231749]

NCT02420821 *{published data only}*

Atkins MB, Rini BI, Motzer RJ, Powles T, McDermott DF, Suarez C, et al. Patient-reported outcomes from the phase III randomized IMmotion151 trial: atezolizumab bevacizumab versus sunitinib in treatment-naive metastatic renal cell carcinoma. *Clinical Cancer Research* 2020;**26**(11):2506-14. [ClinicalTrials.gov: NCT02420821]

Escudier B, Motzer RJ, Rini BI, Powles T, McDermott DF, Suarez C, et al. Patient-reported outcomes (PROs) in IMmotion151: atezolizumab (atezo) + bevacizumab (bev) vs sunitinib (sun) in treatment (tx) naive metastatic renal cell carcinoma (mRCC). *Journal of Clinical Oncology* 2018;**36**(15):4511. [ClinicalTrials.gov: NCT02420821]

Grullich C, Motzer RJ, Powles T, Atkins M, Escudier BJ, McDermott D, et al. IMmotion151: a randomized phase III study of atezolizumab plus bevacizumab vs sunitinib in untreated metastatic renal cell carcinoma (mRCC). *Oncology Research and Treatment* 2018;**41 Suppl 4**:41-2. [ClinicalTrials.gov: NCT02420821]

Motzer RJ, Powles T, Atkins MB, Escudier B, McDermott DF, Alekseev BY, et al. Final overall survival and molecular analysis in IMmotion151, apPhase 3 trial comparing atezolizumab plus bevacizumab vs sunitinib in patients with previously untreated metastatic renal cell carcinoma. In: JAMA Oncology. Vol. 23. 2021:23. [ClinicalTrials.gov: NCT02420821]

Motzer RJ, Powles T, Atkins MB, Escudier B, McDermott DF, Suarez C, et al. IMmotion151: a randomized phase III study of atezolizumab plus bevacizumab vs sunitinib in untreated metastatic renal cell carcinoma (mRCC). *Journal of Clinical Oncology* 2018;**36**(6):578. [ClinicalTrials.gov: NCT02420821]

Rexer H, Doehn C. First-line treatment for advanced renal cell carcinoma: a phase 3, open-label, randomized study of atezolizumab (Anti-PD-L1-Antibody) in combination with bevacizumab versus sunitinib in patients with untreated advanced renal cell carcinoma ("IMmotion") - AN 37/15 der AUO. *Urologe* 2016;**55**(9):1242-3. [ClinicalTrials.gov: NCT02420821]

* Rini BI, Atkins MB, Escudier B, Powles T, McDermott DF, Alekseev BY, et al. IMmotion151: updated overall survival (OS) and exploratory analysis of the association of gene expression and clinical outcomes with atezolizumab plus bevacizumab vs sunitinib in patients with locally advanced or metastatic renal cell carcinoma (mRCC). In: Cancer Research. Conference: AACR Annual Meeting. Vol. 81. 2021. [ClinicalTrials.gov: NCT02420821]

Rini BI, Motzer RJ, Powles T, McDermott DF, Escudier B, Donskov F, et al. Atezolizumab (atezo) + bevacizumab (bev) versus sunitinib (sun) in pts with untreated metastatic renal cell carcinoma (mRCC) and sarcomatoid (sarc) histology: iMmotion151 subgroup analysis. *Journal of Clinical Oncology* 2019;**37**(15):4512. [ClinicalTrials.gov: NCT02420821]

Rini BI, Motzer RJ, Powles T, McDermott DF, Escudier B, Donskov F, et al. Atezolizumab plus bevacizumab versus sunitinib for patients with untreated metastatic renal cell carcinoma and sarcomatoid features: a prespecified subgroup analysis of the IMmotion151 clinical trial. *European Urology* 2020;**79**(5):659-662. [ClinicalTrials.gov: NCT02420821]

Rini BI, Powles T, Atkins MB, Escudier B, McDermott DF, Suarez C, et al. Atezolizumab plus bevacizumab versus sunitinib in patients with previously untreated metastatic renal cell carcinoma (IMmotion151): a multicentre, open-label, phase 3, randomised controlled trial. *Lancet* 2019;**393**(10189):2404-15. [ClinicalTrials.gov: NCT02420821]

NCT02684006 *{published data only}*

Albiges L, Rini BI, Haanen JB, Motzer RJ, Kollmannsberger CK, Negrier S, et al. Primary renal tumour shrinkage in patients (pts) who did not undergo upfront cytoreductive nephrectomy (uCN): subgroup analysis from the phase III JAVELIN renal 101 trial of first-line avelumab 1 axitinib (A 1 Ax) vs sunitinib (S) for advanced renal cell carcinoma (aRCC). *Annals of Oncology* 2019;**30 Suppl 5**:v359-60. [ClinicalTrials.gov: NCT02684006]

Bilen MA, Rini BI, Voss MH, Larkin J, Haanen JB, Albiges L, et al. Association of neutrophil-to-lymphocyte ratio with efficacy of first-line avelumab plus axitinib vs. sunitinib in patients with advanced renal cell carcinoma enrolled in the phase 3 JAVELIN renal 101 trial. *Clinical Cancer Research* 2021;**17**:17. [ClinicalTrials.gov: NCT02684006]

Choueiri TK, Larkin JM, Pal SK, Motzer RJ, Venugopal B, Alekseev BY, et al. Efficacy and biomarker analysis of patients (pts) with advanced renal cell carcinoma (aRCC) with sarcomatoid histology (sRCC): subgroup analysis from the phase III JAVELIN renal 101 trial of first-line avelumab plus axitinib (A + Ax) vs sunitinib (S). *Annals of Oncology* 2019;**30 Suppl 5**:v361. [ClinicalTrials.gov: NCT02684006]

Choueiri TK, Motzer RJ, Campbell MT, Alekseev BY, Uemura M, Kollmannsberger CK, et al. Subgroup analysis from JAVELIN renal 101: outcomes for avelumab plus axitinib (A + Ax) versus sunitinib (S) in advanced renal cell carcinoma (aRCC). *Journal of Clinical Oncology* 2019;**37**(7):544. [ClinicalTrials.gov: NCT02684006]

Choueiri TK, Motzer RJ, Rini BI, Haanen J, Campbell MT, Venugopal B, et al. Updated efficacy results from the JAVELIN renal 101 trial: first-line avelumab plus axitinib versus sunitinib in patients with advanced renal cell carcinoma. *Annals of Oncology* 2020;**31**(8):1030-1039. [ClinicalTrials.gov: NCT02684006]

Choueiri TK, Rini BI, Cosgriff T, Miyake H, Vogelzang NJ, Kannourakis G, et al. JAVELIN renal 101: a phase 3 study of avelumab in combination with axitinib vs sunitinib as first-line treatment for patients with advanced renal cell carcinoma (aRCC). *BJU International* 2016;**118 Suppl 5**:29-30. [ClinicalTrials.gov: NCT02684006]

Choueiri TK, Rini BI, Larkin JM, Bjarnason GA, Gravis G, Gurney H, et al. Avelumab plus axitinib vs sunitinib as first-line treatment of advanced renal cell carcinoma: phase 3 study (JAVELIN renal 101). Journal of Clinical Oncology 2017;**35**(15). [ClinicalTrials.gov: NCT02684006]

Haanen J. Efficacy of avelumab + axitinib vs sunitinib by IMDC risk group in advanced renal cell carcinoma (aRCC): extended follow-up results from JAVELIN renal 101. *Asia-Pacific Journal of Clinical Oncology* 2021;**17Suppl 7**:51. [ClinicalTrials.gov: NCT02684006]

* Haanen JB, Larkin J, Choueiri TK, Albiges L, Rini BI, Atkins MB, et al. Efficacy of avelumab + axitinib (A + Ax) versus sunitinib (S) by IMDC risk group in advanced renal cell carcinoma (aRCC): extended follow-up results from JAVELIN renal 101. *Journal of Clinical Oncology* 2021;**39**(15 Suppl):4574. [ClinicalTrials.gov: NCT02684006]

Jonasch E, Hasanov E, Motzer RJ, Hariharan S, Choueiri TK, Huang B, et al. Evaluation of brain metastasis in JAVELIN renal 101: efficacy of avelumab + axitinib (A+Ax) versus sunitinib (S). *Journal of Clinical Oncology* 2020;**38**(6):687. [ClinicalTrials.gov: NCT02684006]

Motzer RJ, Choueiri T, Larkin J, Albiges L, Haanen JB, Schmidinger M, et al. Phase 3 study of avelumab in combination with axitinib versus sunitinib as first-line treatment for patients with advanced renal cell carcinoma (aRCC). *Annals of Oncology* 2016;**27**(6):vi293. [ClinicalTrials.gov: NCT02684006]

Motzer RJ, Penkov K, Haanen J, Rini B, Albiges L, Campbell MT, et al. Avelumab plus axitinib versus sunitinib for advanced renal-cell carcinoma. *New England Journal of Medicine* 2019;**380**(12):1103-15. [ClinicalTrials.gov: NCT02684006]

Motzer RJ, Penkov K, Haanen JB, Rini BI, Albiges L, Campbell MT, et al. JAVELIN renal 101: a randomized, phase III study of avelumab + axitinib vs sunitinib as first-line treatment of advanced renal cell carcinoma (aRCC). *Annals of Oncology* 2018;**29 Suppl 8**:viii724. [ClinicalTrials.gov: NCT02684006]

Motzer RJ, Robbins PB, Powles T, Albiges L, Haanen JB, Larkin J, et al. Avelumab plus axitinib versus sunitinib in advanced renal cell carcinoma: biomarker analysis of the phase 3 JAVELIN renal 101 trial. *Nature Medicine* 2020;**26**(11):1733-41. [ClinicalTrials.gov: NCT02684006]

Tomita Y, Larkin J, Venugopal B, Haanen JB, Kanayama HO, Eto M, et al. Association of c-reactive protein (CRP) with efficacy of avelumab + axitinib (A + Ax) in advanced renal cell carcinoma (aRCC): long-term follow-up results from JAVELIN renal 101. *Journal of Clinical Oncology* 2021;**39**(15 Suppl):4548. [ClinicalTrials.gov: NCT02684006]

Tomita Y, Motzer RJ, Choueiri TK, Rini BI, Miyake H, Oya M, et al. Efficacy of avelumab plus axitinib (A + Ax) versus sunitinib (S) by number of IMDC risk factors and tumor sites at baseline in advanced renal cell carcinoma (aRCC): extended follow-up results from JAVELIN renal 101. *Journal of Clinical Oncology. Conference* 2021;**39**(6 Suppl):302. [ClinicalTrials.gov: NCT02684006]

Tomita Y, Motzer RJ, Choueiri TK, Rini BI, Miyake H, Uemura H, et al. Efficacy and safety of avelumab plus axitinib (A + Ax) versus sunitinib (S) in elderly patients with advanced renal cell carcinoma (aRCC): extended follow-up results from JAVELIN renal 101. *Journal of Clinical Oncology. Conference* 2021;**39**(6 Suppl):301. [ClinicalTrials.gov: NCT02684006]

Uemura M, Tomita Y, Miyake H, Hatakeyama S, Kanayama HO, Numakura K, et al. Avelumab plus axitinib vs sunitinib for advanced renal cell carcinoma: Japanese subgroup analysis from JAVELIN renal 101. *Cancer Science* 2019;**28**:28. [ClinicalTrials.gov: NCT02684006]

Uemura M, Tomita Y, Miyake H, Hatakeyama S, Kanayama HO, Numakura K, et al. Avelumab plus axitinib vs sunitinib for advanced renal cell carcinoma: Japanese subgroup analysis from JAVELIN renal 101. *Cancer Science* 2020;**111**(3):907-23. [ClinicalTrials.gov: NCT02684006]

Uemura M, Tomita Y, Miyake H, Hatakeyama S, Kanayama HO, Numakura K, et al. Randomized phase III trial of avelumab + axitinib vs sunitinib as firstline treatment for advanced renal cell carcinoma: JAVELIN renal 101 Japanese subgroup analysis. *Annals of Oncology* 2019;**30 Suppl 5**:v386-7. [ClinicalTrials.gov: NCT02684006]

Voss MH, Motzer RJ, Albiges L, Rini BI, Schmidinger M, Larkin J, et al. Depth of response (DEPOR) analysis and correlation with clinical outcomes from JAVELIN renal 101. *Oncology Research and Treatment* 2020;**43**:225. [ClinicalTrials.gov: NCT02684006]

NCT02761057 *{published data only}*

NCT02761057. Cabozantinib s-malate, crizotinib, volitinib, or sunitinib malate in treating patients with locally advanced or metastatic kidney cancer. clinicaltrials.gov/show/NCT02761057 first received 4 May 2016. [ClinicalTrials.gov: NCT02761057]

Pal SK, Tangen C, Thompson IM, Balzer-Haas N, George DJ, Heng DY, et al. A comparison of sunitinib with cabozantinib, crizotinib, and savolitinib for treatment of advanced papillary renal cell carcinoma: a randomised, open-label, phase 2 trial. *Lancet* 2021;**397**(10275):695-703. [ClinicalTrials.gov: NCT02761057]

* Pal SK, Tangen C, Thompson IM, Haas NB, George DJ, Heng DY, et al. Sunitinib versus cabozantinib, crizotinib or savolitinib in metastatic papillary renal cell carcinoma (pRCC): results from the randomized phase II SWOG 1500 study. *Journal of Clinical Oncology* 2021;**39**(6 Suppl):270. [ClinicalTrials.gov: NCT02761057]

Pal SK, Tangen C, Thompson IM, Haas NB, George DJ, Heng DY, et al. Sunitinib versus cabozantinib, crizotinib or savolitinib in metastatic papillary renal cell carcinoma (pRCC): results from the randomized phase II SWOG 1500 study. *Journal of Clinical Oncology* 2021;**39**(6 Suppl):270. [ClinicalTrials.gov: NCT02761057]

Pal SK, Tangen CM, Thompson IM, Shuch BM, Haas NB, George DJ, et al. A randomized, phase II efficacy assessment of multiple MET kinase inhibitors in metastatic papillary renal carcinoma (PRCC): SWOG S1500. In: Journal of Clinical Oncology (Conference). Vol. 35. 2017. [ClinicalTrials.gov: NCT02761057]

NCT02811861 *{published data only}*

Choueiri TK, Eto M, Kopyltsov E, Rha SY, Porta CG, Motzer R, et al. Phase III CLEAR trial in advanced renal cell carcinoma (aRCC): outcomes in subgroups and toxicity update. *Annals of Oncology* 2021;**32 Suppl 5**:S683-5. [ClinicalTrials.gov: NCT02811861]

Gruenwald V, Powles T, Choueiri TK, Hutson TE, Porta C, Eto M, et al. Lenvatinib plus everolimus or pembrolizumab versus sunitinib in advanced renal cell carcinoma: study design and rationale. *Future Oncology* 2019;**15**(9):929-41. [ClinicalTrials.gov: NCT02811861]

Gruenwald V, Powles T, Kopyltsov E, Kozlov V, Alonso Gordoa T, Eto M, et al. Analysis of the CLEAR study in patients (pts) with advanced renal cell carcinoma (RCC): depth of response and efficacy for selected subgroups in the lenvatinib (L) + pembrolizumab (P) and sunitinib (S) treatment arms. *Oncology Research and Treatment* 2021;**44 Suppl 2**:25. [ClinicalTrials.gov: NCT02811861]

Gruenwald V, Powles T, Kopyltsov E, Kozlov V, Gordoa TA, Eto M, et al. Analysis of the CLEAR study in patients (pts) with advanced renal cell carcinoma (RCC): depth of response and efficacy for selected subgroups in the lenvatinib (LEN) + pembrolizumab (PEMBRO) and sunitinib (SUN) treatment arms. *Journal of Clinical Oncology* 2021;**39**(15 Suppl):4560. [ClinicalTrials.gov: NCT02811861]

Hutson TE, Choueiri TK, Motzer RJ, Rha SY, Alyasova A, Merchan JR, et al. Post hoc analysis of the CLEAR study in advanced renal cell carcinoma (RCC): effect of subsequent therapy on survival outcomes in the lenvatinib (LEN) + everolimus (EVE) versus sunitinib (SUN) treatment arms. *Journal of Clinical Oncology* 2021;**39**(15 Suppl):4562. [ClinicalTrials.gov: NCT02811861]

* Motzer RJ, Alekseev B, Rha SY, Porta C, Eto M, Powles T, et al. Lenvatinib plus pembrolizumab or everolimus for advanced renal cell carcinoma. *New England Journal of Medicine* 2021;**13**:13. [ClinicalTrials.gov: NCT02811861]

Motzer RJ, Gruenwald V, Hutson TE, Porta C, Powles T, Eto M, et al. A phase 3 trial to compare efficacy and safety of lenvatinib in combination with everolimus or pembrolizumab versus sunitinib alone in first-line treatment of patients with metastatic renal cell carcinoma. In: Journal of Clinical Oncology (Conference). Vol. 36. 2018. [ClinicalTrials.gov: NCT02811861]

Motzer RJ, Gruenwald V, Hutson TE, Porta C, Powles T, Eto M, et al. A phase 3 trial to compare efficacy and safety of lenvatinib in combination with everolimus or pembrolizumab vs sunitinib alone in first-line treatment of patients with metastatic renal cell carcinoma. *Kidney Cancer* 2018;**2 Suppl 1**:S2-3. [ClinicalTrials.gov: NCT02811861]

Motzer RJ, Gruenwald V, Hutson TE, Porta C, Powles T, Eto M, et al. A phase III trial to compare efficacy and safety of lenvatinib in combination with everolimus or pembrolizumab vs sunitinib alone in first-line treatment of patients (pts) with metastatic renal cell carcinoma (RCC). In: Journal of Clinical Oncology (Conference). Vol. 35. 2017. [ClinicalTrials.gov: NCT02811861]

Motzer RJ, Gruenwald V, Hutson TE, Porta C, Powles T, Eto M, et al. A phase III trial to compare efficacy and safety of lenvatinib in combination with everolimus or pembrolizumab vs sunitinib alone in first-line treatment of patients with metastatic renal cell carcinoma (RCC). *Asia-Pacific Journal of Clinical Oncology* 2017;**13**:169. [ClinicalTrials.gov: NCT02811861]

Motzer RJ, Porta C, Alekseev B, Rha SY, Choueiri TK, Mendez-Vidal MJ, et al. Health-related quality-of-life (HRQoL) analysis from the phase 3 CLEAR trial of lenvatinib (LEN) plus pembrolizumab (PEMBRO) or everolimus (EVE) versus sunitinib (SUN) for patients (pts) with advanced renal cell carcinoma (aRCC). *Journal of Clinical Oncology* 2021;**39**(15 Suppl):4502. [ClinicalTrials.gov: NCT02811861]

Motzer RJ, Porta C, Eto M, Powles T, Gruenwald V, Hutson TE, et al. Phase 3 trial of lenvatinib (LEN) plus pembrolizumab (PEMBRO) or everolimus (EVE) versus sunitinib (SUN) monotherapy as a first-line treatment for patients (pts) with advanced renal cell carcinoma (RCC) (CLEAR study). *Journal of Clinical Oncology* 2021;**39**(6 Suppl):269. [ClinicalTrials.gov: NCT02811861]

NCT02853331 *{published data only}*

Correction to Lancet Oncol 2020; 21: 1563–73 (The Lancet Oncology (2020) 21(12) (1563–1573), (S1470204520304368), (10.1016/S1470-2045(20)30436-8)). Lancet Oncology 2020;**21**(12):e553. [ClinicalTrials.gov: NCT02853331]

Gafanov R, Powles TB, Bedke J, Stus V, Waddell TS, Nosov D, et al. Subsequent therapy following pembrolizumab + axitinib or sunitinib treatment for advanced renal cell carcinoma (RCC) in the phase III KEYNOTE-426 study. *Annals of Oncology* 2021;**32 Suppl 5**:S694. [ClinicalTrials.gov: NCT02853331]

Kondoh CN, Bae WK, Tamada S, Matsubara N, Lee HJ, Mizuno R, et al. Pembrolizumab plus axitinib (pembro + axi) vs sunitinib in metastatic renal cell carcinoma (mRCC) outcomes of the KEYNOTE-426 study in patients from eastern Asia. *Annals of Oncology* 2020;**31**(Suppl 6):1319. [ClinicalTrials.gov: NCT02853331]

Plimack ER, Powles T, Bedke J, Pouliot F, Stus V, Waddell T, et al. Outcomes for patients in the pembrolizumab+axitinib arm with advanced renal cell carcinoma (RCC) who completed two years of treatment in the phase III KEYNOTE-426 study. *Journal of Clinical Oncology* 2021;**39**(6 Suppl):327. [ClinicalTrials.gov: NCT02853331]

Plimack ER, Rini BI, Stus V, Gafanov R, Hawkins R, Nosov D, et al. Pembrolizumab plus axitinib versus sunitinib as first-line therapy formetastatic renal cell carcinoma: outcomes in the combined IMDC intermediate-/poor-risk and sarcomatoid subgroups of the phase 3 Keynote-426 study. *Asia-Pacific Journal of Clinical Oncology* 2019;**15**:33-4. [ClinicalTrials.gov: NCT02853331]

Plimack ER, Rini BI, Stus V, Gafanov R, Waddell T, Nosov D, et al. Pembrolizumab plus axitinib versus sunitinib as first-line therapy for advanced renal cell carcinoma (RCC): updated analysis of KEYNOTE-426. *Journal of Clinical Oncology* 2020;**38**(15):5001. [ClinicalTrials.gov: NCT02853331]

Powles T, Plimack ER, Soulieres D, Waddell T, Stus V, Gafanov R, et al. Pembrolizumab plus axitinib versus sunitinib monotherapy as first-line treatment of advanced renal cell carcinoma (KEYNOTE-426): extended follow-up from a randomised, open-label, phase 3 trial. *Lancet Oncology* 2020;**21**(12):1563-73. [ClinicalTrials.gov: NCT02853331]

Powles T, Plimack ER, Stus V, Gafanov RA, Hawkins RE, Nosov D, et al. Pembrolizumab (pembro) plus axitinib (axi) versus sunitinib as first-line therapy for locally advanced or metastatic renal cell carcinoma (mRCC): phase III KEYNOTE-426 study. *Journal of Clinical Oncology* 2019;**37**(7):543. [ClinicalTrials.gov: NCT02853331]

Rexer H, Bedke J. First-line therapy in advanced renal cell carcinoma: a randomized, open-label phase III study evaluating the efficacy and safety of pembrolizumab (MK-3475) in combination with axitinib compared to sunitinib monotherapy as first-line treatment for locally advanced or metastatic renal cell carcinoma (mRCC) (Keynote-426) - AN 39/16 of the AUO. *Urologe* 2017;**56**(3):385-6. [ClinicalTrials.gov: NCT02853331]

Rini BI, Atkins MB, Plimack ER, Soulieres D, McDermott RS, Bedke J, et al. Characterization and management of treatment-emergent hepatic toxicity in patients with advanced renal cell carcinoma receiving first-line pembrolizumab plus axitinib. Results from the KEYNOTE-426 trial. *European Urology Oncology* 2021;**5**(2):225-234. [ClinicalTrials.gov: NCT02853331]

Rini BI, Plimack ER, Stus V, Gafanov R, Hawkins R, Nosov D, et al. Pembrolizumab (pembro) plus axitinib (axi) versus sunitinib as first-line therapy for metastatic renal cell carcinoma (mRCC): outcomes in the combined IMDC intermediate/poor risk and sarcomatoid subgroups of the phase 3 KEYNOTE-426 study. *Journal of Clinical Oncology* 2019;**37**(15 Suppl):4500. [ClinicalTrials.gov: NCT02853331]

Rini BI, Plimack ER, Stus V, Gafanov R, Hawkins R, Nosov D, et al. Pembrolizumab plus axitinib versus sunitinib for advanced renal-cell carcinoma. *New England Journal of Medicine* 2019;**380**(12):1116-27. [ClinicalTrials.gov: NCT02853331]

* Rini BI, Plimack ER, Stus V, Waddell T, Gafanov R, Pouliot F, et al. Pembrolizumab (pembro) plus axitinib (axi) versus sunitinib as first-line therapy for advanced clear cell renal cell carcinoma (ccRCC): results from 42-month follow-up of KEYNOTE-426. *Journal of Clinical Oncology* 2021;**39**(15 Suppl):4500. [ClinicalTrials.gov: NCT02853331]

Rini BI, Powles T, Chen M, Puhlmann M, Atkins MB. Phase 3 KEYNOTE-426 trial: pembrolizumab (pembro) plus axitinib versus sunitinib alone in treatment-naive advanced/metastatic renal cell carcinoma (mRCC). In: Journal of Clinical Oncology (Conference). Vol. 35. 2017. [ClinicalTrials.gov: NCT02853331]

Rini BI, Powles T, Chen M, Song Y, Puhlmann M, Atkins MB. KEYNOTE-426: randomized phase III study of pembrolizumab in combination with axitinib versus sunitinib monotherapy in treatment-naive advanced/metastatic renal cell carcinoma (mRCC). In: Journal for ImmunoTherapy of Cancer. Conference: 31st Annual Meeting and Associated Programs of the Society for Immunotherapy of Cancer, SITC. Vol. 4 Suppl 1. 2016. [ClinicalTrials.gov: NCT02853331]

Soulieres D, Plimack ER, Rini BI, Stus V, Gafanov R, Hawkins R, et al. Pembrolizumab plus axitinib vs sunitinib as first-line therapy for metastatic renal cell carcinoma: outcomes by IMDC risk and sarcomatoid subgroups of the phase 3 KEYNOTE-426 study. *Kidney Cancer* 2020;**4 Suppl 1**:40. [ClinicalTrials.gov: NCT02853331]

Tamada S, Kondoh C, Matsubara N, Mizuno R, Kimura G, Anai S, et al. Pembrolizumab plus axitinib versus sunitinib in metastatic renal cell carcinoma: outcomes of Japanese patients enrolled in the randomized, phase III, open-label KEYNOTE-426 study. *International Journal of Clinical Oncology* 2021;**27**(1):154-64. [ClinicalTrials.gov: NCT02853331]

Tamada S, Kondoh C, Matsubara N, Mizuno R, Kimura G, Anai S, et al. Pembrolizumab plus axitinib versus sunitinib in metastatic renal cell carcinoma: outcomes of Japanese patients enrolled in the randomized, phase III, open-label KEYNOTE-426 study. *International Journal of Clinical Oncology* 2022;**27**(1):154-64. [ClinicalTrials.gov: NCT02853331]

Zhu J, Zhang T, Wan N, Liang Z, Li J, Chen X, et al. Cost-effectiveness of pembrolizumab plus axitinib as first-line therapy for advanced renal cell carcinoma. *Immunotherapy* 2020;**12**(17):1237-46. [ClinicalTrials.gov: NCT02853331]

NCT03141177 *{published data only}*

Apolo AB, Powles T, Burotto M, Bourlon MT, Hsieh JJ, Basso U, et al. Nivolumab plus cabozantinib (N+C) versus sunitinib (S) for advanced renal cell carcinoma (aRCC): outcomes by baseline disease characteristics in the phase 3 CheckMate 9ER trial. *Journal of Clinical Oncology* 2021;**39**(15 Suppl):4553. [ClinicalTrials.gov: NCT03141177]

Basso U, Powles T, Burotto M, Bourlon MT, Hsieh JJ, Shah AY, et al. Nivolumab plus cabozantinib (N+C) vs sunitinib (S) for advanced renal cell carcinoma (aRCC): outcomes by baseline disease characteristics in the phase 3 CheckMate 9ER trial. *Tumori Journal* 2021;**107**(2 Suppl):42-3. [ClinicalTrials.gov: NCT03141177]

Bedke J, Powles T, Burotto M, Bourlon MT, Hsieh JJ, Basso U, et al. Nivolumab plus cabozantinib (N+C) vs sunitinib (S) for advanced renal cell carcinoma (aRCC): outcomes by baseline disease characteristics in the phase 3 CheckMate 9ER trial. *Oncology Research and Treatment* 2021;**44 Suppl 2**:24-5. [ClinicalTrials.gov: NCT03141177]

Cella D, Choueiri TK, Blum SI, Ejzykowicz F, Hamilton M, Zhang J, et al. Patient-reported outcomes of patients with advanced renal cell carcinoma (aRCC) treated with first-line nivolumab plus cabozantinib versus sunitinib: the CheckMate 9ER trial. *Journal of Clinical Oncology* 2021;**39**(6 Suppl):285. [ClinicalTrials.gov: NCT03141177]

Cella D, Motzer RJ, Suarez C, Blum SI, Ejzykowicz F, Hamilton M, et al. Patient-reported outcomes with first-line nivolumab plus cabozantinib versus sunitinib in patients with advanced renal cell carcinoma treated in CheckMate 9ER: an open-label, randomised, phase 3 trial. *Lancet Oncology* 2022;**23**(2):292-303. [ClinicalTrials.gov: NCT03141177]

Choueiri TK, Apolo AB, Powles T, Escudier B, Aren OR, Shah A, et al. A phase 3, randomized, open-label study of nivolumab combined with cabozantinib vs sunitinib in patients with previously untreated advanced or metastatic renal cell carcinoma (RCC; CheckMate 9ER). In: Journal of Clinical Oncology. Meeting Abstract | 2018 ASCO Annual Meeting I. Vol. 36. 2018. [ClinicalTrials.gov: NCT03141177]

Choueiri TK, Powles T, Burotto M, Bourlon MT, Zurawski B, Oyervides Juarez VM, et al. Nivolumab + cabozantinib vs sunitinib in first-line treatment for advanced renal cell carcinoma: first results from the randomized phase III CheckMate 9ER trial. *Annals of Oncology* 2020;**31 Suppl 4**:1159. [ClinicalTrials.gov: NCT03141177]

* Choueiri TK, Powles T, Burotto M, Escudier B, Bourlon MT, Zurawski B, et al. Nivolumab plus cabozantinib versus sunitinib for advanced renal-cell carcinoma. *New England Journal of Medicine* 2021;**384**:829-41. [ClinicalTrials.gov: NCT03141177]

Kfoury M, Oing C. ESMO20 YO for YO: highlights on metastatic renal cell carcinoma-the CheckMate-9ER trial. *ESMO Open* 2021;**6**(1):100025. [ClinicalTrials.gov: NCT03141177]

Mauro G. Nivolumab/cabozantinib combo impresses as frontline RCC treatment: combination reduces risk of disease progression or death by 49% versus sunitinib. *Urology Times* 2020;**48**(11):32. [ClinicalTrials.gov: NCT03141177]

Motzer RJ, Choueiri TK, Powles T, Burotto M, Bourlon MT, Hsieh JJ, et al. Nivolumab + cabozantinib (NIVO+CABO) versus sunitinib (SUN) for advanced renal cell carcinoma (aRCC): outcomes by sarcomatoid histology and updated trial results with extended follow-up of CheckMate 9ER. *Journal of Clinical Oncology* 2021;**39**(6 Suppl):308. [ClinicalTrials.gov: NCT03141177]

Oing C, Bokemeyer C. Increasing combination possibilities for first-line treatment of metastatic clear cell renal cell carcinoma with nivolumab plus cabozantinib: Results of the CheckMate-9ER trial.. *Onkologe* 2021;**27**(2):168-71. [ClinicalTrials.gov: NCT03141177]

Pook D, Powles T, Burotto M, Bourlon MT, Hsieh JJ, Basso U, et al. Nivolumab plus cabozantinib (N+C) vs sunitinib for advanced renal cell carcinoma (aRCC): outcomes by baseline disease characteristics in the phase 3 CheckMate 9ER trial. *Asia-Pacific Journal of Clinical Oncology* 2021;**17 Suppl 9**:131. [ClinicalTrials.gov: NCT03141177]

Shah AY, Motzer RJ, Apolo AB, Powles T, Escudier B, Zhang J, et al. Cabozantinib (C) exposureresponse (ER) analysis for the phase 3 CheckMate 9ER (CM 9ER) trial of nivolumab plus cabozantinib (N+C) versus sunitinib (S) in first-line advanced renal cell carcinoma (1L aRCC). *Journal of Clinical Oncology* 2021;**39**(15 Suppl):4561. [ClinicalTrials.gov: NCT03141177]

References to studies excluded from this review

Aass 2005 *{published data only}*

Aass N, De Mulder PH, Mickisch GH, Mulders P, van Oosterom AT, van Poppel H, et al. Randomized phase II/III trial of interferon alfa-2a with and without 13-cis-retinoic acid in patients with progressive metastatic renal cell carcinoma: the European Organisation for Research and Treatment of Cancer Genito-Urinary Tract Cancer Group (EORTC 30951). *Journal of Clinical Oncology* 2005;**23**(18):4172-8.

Abdel 2018 *{published data only}*

Abdel Hakim AM, Nouira Y, Taran R, Amokrane D, Ghosn M, Larbaoui B, et al. Everolimus as first-line or after cytokine therapy in patients with metastatic recurrent and/or unresectable renal cell carcinoma (RCC) (EVERMORE). *Annals of Oncology* 2018;**29 Suppl 8**:viii316.

Adler 1987 *{published data only}*

Adler A, Gillon G, Lurie H, Shaham J, Loven D, Shachter Y, et al. Active specific immunotherapy of renal cell carcinoma patients: a prospective randomized study of hormono-immuno- versus hormonotherapy. Preliminary report of immunological and clinical aspects. *Journal of Biological Response Modifiers* 1987;**6**(6):610-24.

Amin 2018 *{published data only}*

Amin A, Plimack ER, Ernstoff MS, Lewis LD, Bauer TM, McDermott DF, et al. Safety and efficacy of nivolumab in combination with sunitinib or pazopanib in advanced or metastatic renal cell carcinoma: the CheckMate 016 study. *Journal for ImmunoTherapy of Cancer* 2018;**6**(1):109.

Amin 2018a *{published data only}*

Amin A, Plimack ER, Lewis LD, Ernstoff MS, Bauer TM, McDermott DF, et al. Updated results from a phase I study of nivolumab in combination with sunitinib or pazopanib in metastatic renal cell carcinoma: the CheckMate 016 study. *Kidney Cancer* 2018;**2 Suppl 1**:S45-6.

Atkins 1991 *{published data only}*

Atkins MB, Sparano J, Fisher RI, Weiss GR, Margolin KA, Aronson FR, et al. Randomized phase II trial of high dose IL-2 either alone or in combination with interferon alpha 2B in advanced renal cell carcinoma. In: Proceedings of The American Society of Clinical Oncology. 1991.

Atkins 1993 *{published data only}*

Atkins MB, Sparano J, Fisher RI, Weiss GR, Margolin KA, Fink KI, et al. Randomized phase II trial of high-dose interleukin-2 either alone or in combination with interferon alfa-2b in advanced renal cell carcinoma. *Journal of Clinical Oncology* 1993;**11**(4):661-70.

Atzpodien 1997 *{published data only}*

Atzpodien J, Kirchner H, Franzke A, Wandert T, Probst M, Buer J, et al. Results of a randomized clinical trial comparing SC interleukin-2, SC alpha-2a-interferon, and IV bolus 5-fluorouracil against oral tamoxifen in progressive metastatic renal cell carcinoma patients. In: Proceedings of The American Society of Clinical Oncology. Vol. 16. 1997:326a, Abstract 1164.

Atzpodien 1997a *{published data only}*

Atzpodien J, Kirchner H, Franzke A. A randomized clinical trial comparing sc interleukin-2, sc alpha-2a-interferon, and iv bolus 5-flurouracil against oral tamoxifen in progressive metastatic renal cell carcinoma patients. *European Journal of Cancer* 1997;**33**:S8, A151.

Atzpodien 1999 *{published data only}*

Atzpodien J, Kirchner H, Bergmann L, Oberneder R, Jonas U, Ganser A. 13cis-retinoic acid, IFN-alpha, IL-2 and chemotherapy in advanced renal cell carcinoma: results of a prospectively randomized trial of the German Cooperative Renal Carcinoma Immunotherapy Group (DGCIN). In: Proceedings of The American Society of Clinical Oncology. Vol. 18. 1999:448a, Abstract 1727.

Atzpodien 2001 *{published data only}*

Atzpodien J, Kirchner H, Illiger HJ, Metzner B, Ukena D, Schott H, et al. IL-2 in combination with IFN- alpha and 5-FU versus tamoxifen in metastatic renal cell carcinoma: long-term results of a controlled randomized clinical trial. *British Journal Of Cancer* 2001;**85**(8):1130-6.

Atzpodien 2004 *{published data only}*

Atzpodien J, Kirchner H, Jonas U, Bergmann L, Schott H, Heynemann H, et al. Interleukin-2- and interferon alfa-2a-based immunochemotherapy in advanced renal cell carcinoma: a prospectively randomized trial of the German Cooperative Renal Carcinoma Chemoimmunotherapy Group (DGCIN). *Journal of Clinical Oncology* 2004;**22**(7):1188-94.

Atzpodien 2006 *{published data only}*

Atzpodien J, Kirchner H, Rebmann U, Soder M, Gertenbach U, Siebels M, et al. Interleukin-2/interferon-alpha2a/13-retinoic acid-based chemoimmunotherapy in advanced renal cell carcinoma: results of a prospectively randomised trial of the German Cooperative Renal Carcinoma Chemoimmunotherapy Group (DGCIN). *British Journal Of Cancer* 2006;**95**(4):463-9.

Barrios 2009 *{published data only}*

Barrios CH, Hernandez-Barajas D, Brown MP, Lee SH, Fein L, Liu JH, et al. Phase II trial of continuous once-daily dosing of sunitinib as first-line treatment in patients with metastatic renal cell carcinoma (mRCC): preliminary results. *European Journal of Cancer, Supplement* 2009;**7**(2-3):429-30.

Beaumont 2009 *{published data only}*

Beaumont J, Cella D, Hutson T, Bracarda S, Grunwald V, Thompson J, et al. Patient-reported outcomes in a randomized trial of everolimus with metastatic renal cell carcinoma patients. *Journal of Clinical Oncology* 2009;**1**:e17516.

Beaumont 2011 *{published data only}*

Beaumont JL, Butt Z, Baladi J, Motzer RJ, Haas T, Hollaender N, et al. Patient-reported outcomes in a phase III study of everolimus versus placebo in patients with metastatic carcinoma of the kidney that has progressed on vascular endothelial growth factor receptor tyrosine kinase inhibitor therapy. *Oncologist* 2011;**16**(5):632-40.

Berg 1998 *{published data only}*

Berg WJ, Bukowski R, Thompson JA, Nemunaitis J, Murphy B, Ellerhorst J, et al. A randomized phase II trial of recombinant human interleukin-12 (IL-12) versus interferon alpha-2A (IFN) in advanced renal cell carcinoma. In: Proceedings of The American Society of Clinical Oncology. Vol. 17. 1998:318a, Abstract 1226.

Bex 2017 *{published data only}*

Bex A, Mulders P, Jewett MA, Wagstaff J, Van Velthoven R, Laguna Pes PM, et al. Immediate versus deferred cytoreductive nephrectomy (CN) in patients with synchronous metastatic renal cell carcinoma (mRCC) receiving sunitinib (EORTC 30073 SURTIME). *Annals of Oncology* 2017;**28 Suppl 5**:v622.

Boccardo 1998 *{published data only}*

Boccardo F, Rubagotti A, Canobbio L, Galligioni E, Sorio R, Lucenti A, et al. Interleukin-2, interferon-alpha and interleukin-2 plus interferon-alpha in renal cell carcinoma. A randomized phase II trial. *Tumori Journal* 1998;**84**(5):534-9.

Bracarda 2007 *{published data only}*

Bracarda S, Porta C, Boni C. Randomized prospective phase II trial of two schedules of sorafenib daily and interferon-a2a in metastatic renal cell carcinoma (RAPSODY): GOIRC Study 0681. *Journal Of Clinical Oncology* 2007;**25 Suppl**:Abstract 5100.

Buckley 2019 *{published data only}*

Buckley HL, Collinson FJ, Ainsworth G, Poad H, Flanagan L, Katona E et al. PRISM protocol: a randomised phase II trial of nivolumab in combination with alternatively scheduled ipilimumab in first-line treatment of patients with advanced or metastatic renal cell carcinoma. *BMC Cancer* 2019;**19**(1):1102.

Cella 2016 *{published data only}*

Cella D, Gruenwald V, Nathan PD, Doan J, Dastani H, Taylor F, et al. Quality of life (QoL) and overall survival (OS) in patients (pts) with advanced clear-cell renal cell carcinoma (aRCC) treated with nivolumab (NIVI) vs. everolimus (EVE) in the phase III CheckMate 025 study. *Journal of Clinical Oncology: ASCO annual meeting proceedings* 2016;**34**(15 Suppl):4549.

Choueiri 2017 *{published data only}*

Choueiri TK, Jakacki R, Ghiorghiu D, Haddad V, Kohlmann A, Frigault MM, et al. Savolitinib versus sunitinib in patients with MET-driven, unresectable and locally advanced or metastatic papillary renal cell carcinoma: SAVOIR, a randomised, phase III trial. *Annals of Oncology* 2017;**28**:v328.

Choueiri 2020 *{published data only}*

Choueiri TK, Heng DY, Lee JL, Cancel M, Verheijen RB, Mellemgaard A, et al. Efficacy of savolitinib vs sunitinib in patients with MET -driven papillary renal cell carcinoma: the

SAVOIR phase 3 randomized clinical trial. *JAMA Oncology* 2020;**6**:1247-55.

Choueiri 2020a *{published data only}*

Choueiri TK, Heng DY, Lee JL, Cancel M, Verheijen RB, Mellemgaard A, et al. SAVOIR: a phase III study of savolitinib versus sunitinib in pts with MET-driven papillary renal cell carcinoma (PRCC). *Journal of Clinical Oncology: ASCO annual meeting proceedings* 2020;**38**(15 Suppl):5002.

Cirkel 2016 *{published data only}*

Cirkel GA, Hamberg P, Sleijfer S, Loosveld O, Dercksen W, Los M, et al. A randomized phase II study to compare the efficacy of upfront bi-monthly rotations between pazopanib (PAZ) and everolimus (EVE) versus sequential treatment of first-line PAZ and second-line EVE until progression in patients with metastatic clear cell renal cell cancer (ccRCC) (ROPETAR trial). *Journal of Clinical Oncology* 2016;**34**(15):4550.

Cirkel 2017 *{published data only}*

Cirkel GA, Hamberg P, Sleijfer S, Loosveld OJ, Dercksen MW, Los M, et al. Alternating treatment with pazopanib and everolimus vs continuous pazopanib to delay disease progression in patients with metastatic clear cell renal cell cancer: the ROPETAR randomized clinical trial. *JAMA Oncology* 2017;**3**(4):501-8.

Climent 2020 *{published data only}*

Climent Duran MA, Basterretxea L, Sanchez Escribano R, Llabres E, Heras Lopez L, Zambrana F, et al. Pilot study of cabozantinib efficacy, safety and tolerability in metastatic renal carcinoma in aged fragile patients. *Annals of Oncology* 2020;**31 Suppl 4**:S608.

Cole 2003 *{published data only}*

Cole BF, McDermott D, Parker R, Youmans A, Connolly C, Ernstoff M, et al. The impact of treatmen with high-dose interleukin-2 (HD IL-2) or subcutaneous (SC) IL-2/interferon alfa-2b (IFN) on quality of life (QoL) in patients with metastatic renal cell carcinoma (mRCC). Proceedings of The American Society of Clinical Oncology 2003.

Collinson 2012 *{published data only}*

Collinson FJ, Gregory WM, McCabe C, Howard H, Lowe C, Potrata D, et al. The STAR trial protocol: a randomised multi-stage phase II/III study of sunitinib comparing temporary cessation with allowing continuation, at the time of maximal radiological response, in the first-line treatment of locally advanced/metastatic renal cancer. *BMC Cancer* 2012;**12**:598.

Collinson 2018 *{published data only}*

Collinson F, Brown S, Buckley H, Ainsworth G, Howard H, Poad H, et al. PRISM: a randomised phase II trial of nivolumab in combination with alternatively scheduled ipilimumab in first-line treatment of patients with advanced or metastatic renal cell carcinoma. *Annals of Oncology* 2018;**29 Suppl 8**:viii331.

Colomba 2021 *{published data only}*

Colomba E, Flippot R, Dalban C, Negrier S, Chevreau C, Gravis G, et al. Association of statins and nivolumab activity in patients with metastatic renal cell carcinoma (mRCC): results from the phase II nivoren-GETUG AFU 26 trial. *Journal of Clinical Oncology. Conference* 2021;**39**(6 Suppl):359.

Conter 2013 *{published data only}*

Conter HJ, Wood CG, Matin SF, Tamboli P, Millikan RE, Jonasch E, et al. Ten-year follow-up of patients (pts) with metastatic renal cell carcinoma (mRCC) treated with interferon alfa-2b (IFN) as first-line therapy: results from a randomized trial. *Journal of Clinical Oncology. Conference* 2013;**31**(6 Suppl 1):365.

Demirci 1999 *{published data only}*

Demirci D, Tatlisen A, Ekmekcioglu O, Orskiran G, Gulmez I. Comparison of vinblastine and interferon alpha with 5-flourouracil treatment in metastatic renal cell carcinoma: a preliminary report. *Erciyes Tip Dergisi* 1999;**21**(2):111-6.

de Mulder 1991 *{published data only}*

de Mulder PH, Debruyne FM, van Oosterom A, Bouffioux C, Vermeylen K, Sylvester R. EORTC randomized phase II study of recombinant interferon alpha and recombinant interferon alpha and gamma in advanced renal cell carcinoma. In: Proceedings of The American Society of Clinical Oncology. Vol. 10. 1991:166, Abstract 528.

Dexeus 1988 *{published data only}*

Dexeus F, Logothetis C, Chong C, Sella A, Finn L. Phase III study in metastatic renal cell carcinoma comparing combination chemotherapy versus interferon alternating with combination chemotherapy. In: Proceedings of The American Society of Clinical Oncology. Vol. 7. 1988:131, Abstract 507.

Dexeus 1989 *{published data only}*

Dexeus FH, Logothetis CJ, Sella A, Finn L. Interferon alternating with chemotherapy for patients with metastatic renal cell carcinoma. *American Journal of Clinical Oncology* 1989;**12**(4):350-4.

DRKS00010309 2016 *{published data only}*

DRKS00010309. A phase 2 randomized, double-blind, crossover, controlled, multi-center subject preference study of tivozanib hydrochloride versus sunitinib in the treatment of subjects with metastatic renal cell carcinoma. drks.de/drks_web/navigate.do?navigationId=trial.HTML&TRIAL_ID=DRKS00010309 2016.

Dubois 1997 *{published data only}*

Dubois JS, Trehu EG, Mier JW, Shapiro L, Epstein M, Klempner M, et al. Randomized placebo-controlled clinical trial of high-dose interleukin-2 in combination with a soluble p75 tumor necrosis factor receptor immunoglobulin G chimera in patients with advanced melanoma and renal cell carcinoma. *Journal of Clinical Oncology* 1997;**15**(3):1052-62.

Eisen 2019 *{published data only}*

Eisen TQ, Frangou E, Smith B, Ritchie A, Kaplan RS, Oza B, et al. Primary efficacy analysis results from the SORCE trial (RE05): adjuvant sorafenib for renal cell carcinoma at intermediate or high risk of relapse: an international, randomised double-blind phase III trial led by the MRC CTU at UCL. *Annals of Oncology* 2019;**30**:v891-2.

Elhilali 2000 *{published data only}*

Elhilali MM, Gleave M, Fradet Y, Davis I, Venner P, Saad F, et al. Placebo-associated remissions in a multicentre, randomized, double-blind trial of interferon gamma-1b for the treatment of metastatic renal cell carcinoma. *BJU International* 2000;**86**(6):613-8.

Epaillard 2020 *{published data only}*

Epaillard N, Simonaggio A, Elaidi R, Azzouz F, Brachychenko E, Thibault C, et al. BIONIKK: a phase 2 biomarker driven trial with nivolumab and ipilimumab or VEGFR tyrosine kinase inhibitor (TKI) in naïve metastatic kidney cancer. *Bulletin du Cancer* 2020;**107**(5):eS22-7.

Escudier 2005 *{published data only}*

Escudier B, Szczylik C, Eisen T, Stadler WM, Schwartz B, Shan M, et al. Randomized phase III trial of the raf kinase and VEGFR inhibitor sorafenib (BAY 43-9006) in patients with advanced renal cell carcinoma (RCC). In: Annual meeting proceedings of The American Society of Clinical Oncology. Vol. 23. 2005:380.

Euctr2006-002851-33-AT *{published data only}*

Schmidinger M. Sorafenib and bevacizumab as first- line treatment in patients with advanced renal cell carcinoma. trialsearch.who.int/Trial2.aspx?TrialID=EUCTR2006-002851-33-AT 2006.

Euctr 2006-003429-95-ES *{published data only}*

EUCTR2006-003429-95-ES. Treatment for renal cancer. who.int/trialsearch/Trial2.aspx?TrialID=EUCTR2006-003429-95-ES 2012.

Euctr2006-005751-16-NL *{published data only}*

EUCTR2006-005751-16-NL. An open label, dose escalation safety and tolerability trial of the combination of s.c. recombinant human IL-21 (rIL-21) and sunitinib (phase 1) followed by an open label stratified randomized 2-arm trial of rIL-21 plus sunitinib versus sunitinib alone (phase 2a) in subjects with stage IV renal cell carcinoma. Trial phase: 1/2a. who.int/trialsearch/Trial2.aspx?TrialID=EUCTR2006-005751-16-NL 2007.

Euctr2007-002556-41-AT *{published data only}*

EUCTR2007-002556-41-AT. TWIST- randomized prospective phase II study of temsirolimus with or without low-dose interferon alpha in metastatic non-clear renal cell carcinoma-GOIRC STUDY 02/2008 - TWIST. trialsearch.who.int/Trial2.aspx?TrialID=EUCTR2007%E2%80%90002556%E2%80%9041%E2%80%90AT 2008.

EUCTR2008-002667-13-DE 2008 *{published data only}*

EUCTR2008-002667-13-DE. A single-center controlled pilot study to investigate pre- and perioperative therapy with Sorafenib (Nexavar®) in patients who are candidates for a curative surgery of renal cell cancer (PREST = preoperative Sorafenib Therapy in RCC) - PREST. https://www.clinicaltrialsregister.eu/ctr-search/trial/2008-002667-13/2008.

Euctr2012-001730-33-ES *{published data only}*

EUCTR2012-001730-33-ES. A phase 2 randomized, double-blind, crossover, controlled, multi-center, subject preference study of tivozanib hydrochloride versus sunitinib in the treatment of subjects with metastatic renal cell carcinoma. trialsearch.who.int/Trial2.aspx?TrialID=EUCTR2012-001730-33-ES 2012.

Euctr2015-002133-22-FR *{published data only}*

EUCTR2015-002133-22-FR. Study of the drug substances MLN0128 and MLN0128+MLN1117 compared with everolimus in the treatment of patients with metastatic clear cell renal carcinoma. who.int/trialsearch/Trial2.aspx?TrialID=EUCTR2015-002133-22-FR 2017.

Euctr2018-001495-38-FR *{published data only}*

EUCTR2018-001495-38-FR. A randomized, phase 3, double-blind, placebo-controlled study of Pazopanib with or without abexinostat in patients with locally advanced or metastatic renal cell carcinoma. who.int/trialsearch/Trial2.aspx?TrialID=EUCTR2018-001495-38-FR 2018.

Feldman 2020 *{published data only}*

Feldman DR, Ged Y, Lee CH, Knezevic A, Molina AM, Chen YB, et al. Everolimus plus bevacizumab is an effective first-line treatment for patients with advanced papillary variant renal cell carcinoma: final results from a phase II trial. *Cancer* 2020;**126**(24):5247-55.

Feldman 2020a *{published data only}*

Feldman DR, Ged Y, Lee CH, Knezevic A, Molina AM, Chen YB, et al. Everolimus plus bevacizumab is an effective first-line treatment for patients with advanced papillary variant renal cell carcinoma: final results from a phase II trial. *Cancer* 2020;**126**(24):5247-55.

Figlin 1998 *{published data only}*

Figlin RA, Thompson JA, Roudet C, Lange P, Belldegrun A. Multicenter randomized placebo controlled phase II/III trial of CD8(+) tumor infiltrating lymphocyte therapy (CD8(+) TIL)/recombinant interleukin-2 (IL-2) in metastatic renal cell carcinoma. In: Proceedings of The American Society of Clinical Oncology. Vol. 17. 1998:318a, Abstract 1225.

Figlin 1999 *{published data only}*

Figlin RA, Thompson JA, Bukowski RM, Vogelzang NJ, Novick AC, Lange P, et al. Multicenter, randomized, phase III trial of CD8(+) tumor-infiltrating lymphocytes in combination with recombinant interleukin-2 in metastatic renal cell carcinoma. *Journal of Clinical Oncology* 1999;**17**(8):2521-9.

Figlin 2014 *{published data only}*

Figlin RA, Wood CG. Patient identification and eligibility insights in the synchronous mRCC population: an update from the ongoing ADAPT* phase 3 study experience. In: Journal of Clinical Oncology (Conference). Vol. 33. 2014.

Figlin 2014a *{published data only}*

Figlin RA, Wood CG. Enrollment insights in the synchronous mRCC population: an update from the ongoing ADAPT* phase 3 study experience. In: Journal of Clinical Oncology: ASCO annual meeting proceedings. Vol. 32. 2014.

Figlin 2017 *{published data only}*

Figlin R, Nicolette C, Tannir N, Tykodi SS, Chen D, Master V, et al. Interim analysis of the phase 3 ADAPT trial evaluating rocapuldencel-T (AGS-003), an individualized immunotherapy for the treatment of newly-diagnosed patients with metastatic renal cell carcinoma (mRCC). *Annals of Oncology* 2017;**28 Suppl 5**:v404.

Figlin 2018 *{published data only}*

Figlin RA, Wood CG, Gamble AH, Plachco A, Norris MS, Horvatinovich JM, et al. Interim data analysis of the ADAPT trial using the modified intent to treat (mITT) population re-evaluates Rocapuldencel-T for clinical benefit over standard of care. *Journal of Clinical Oncology: ASCO annual meeting proceedings* 2018;**36**(15 Suppl 1):4557.

Figlin 2020 *{published data only}*

Figlin RA, Tannir NM, Uzzo RG, Tykodi SS, Chen DY, Master V, et al. Results of the ADAPT phase 3 study of rocapuldencel-t in combination with sunitinib as first-line therapy in patients with metastatic renal cell carcinoma. *Clinical Cancer Research* 2020;**26**(10):2327-35.

Flaherty 2015 *{published data only}*

Flaherty KT, Manola JB, Pins M, McDermott DF, Atkins MB, Dutcher JJ, et al. BEST: a randomized phase II study of vascular endothelial growth factor, RAF kinase, and mammalian target of rapamycin combination targeted therapy with bevacizumab, sorafenib, and temsirolimus in advanced renal cell carcinoma - a trial of the ECOG-ACRIN cancer research group (E2804). *Journal of Clinical Oncology* 2015;**33**(21):2384-91.

Foon 1988 *{published data only}*

Foon K, Doroshow J, Bonnem E, Fefer A, Graham S, Grosh B, et al. A prospective randomized trial of alpha(2B)-interferon/gamma- interferon or the combination of advanced metastatic renal cell carcinoma. *Journal of Biological Response Modifiers* 1988;**7**(6):540-5.

Fossa 1989 *{published data only}*

Fossa SD, Raabe N, Moe B. Recombinant interferon-alpha with or without vinblastine in metastatic renal carcinoma. Results of a randomised phase II study. *British Journal of Urology* 1989;**64**(5):468-71.

Fossa 1992 *{published data only}*

Fossa SD, Martinelli G, Otto U, Schneider G, Wander H, Oberling F, et al. Recombinant interferon alfa-2a with or without vinblastine in metastatic renal cell carcinoma: results of a European multi-center phase III study. *Annals of Oncology* 1992;**3**(4):301-5.

Gao 2017 *{published data only}*

Gao J, Karam JA, Wood CG, Matin S, Ahrar K, Jonasch E, et al. Clinical activity, immune and molecular correlates of nivolumab vs. nivolumab plus bevacizumab vs nivolumab plus ipilimumab in metastatic renal cell carcinoma. *Cancer Research* 2017;**77**(13 Suppl):CT083.

Gao 2019 *{published data only}*

Gao J, Karam JA, Tannir NM, Campbell MT, Tidwell RS, Ahrar K, et al. A pilot randomized study evaluating nivolumab (nivo) or nivo + bevacizumab (bev) or nivo + ipilimumab (ipi) in patients with metastatic renal cell carcinoma (MRCC) eligible for cytoreductive nephrectomy, metastasectomy or post-treatment biopsy (Bx). *Journal of Clinical Oncology* 2019;**37**(15):4501.

Gedye 2021 *{published data only}*

Gedye C, Joshi AJ, Zhang AY, Martin AJ, Joshua AM, Harris CA, et al. Denosumab and pembrolizumab in clear cell renal carcinoma (KEYPAD): a phase II trial (ANZUP1601). In: Journal of Clinical Oncology (Conference). Vol. 39. 2021.

Ghiorghiu 2018 *{published data only}*

Ghiorghiu D, Jakacki R, Haddad V, Kohlmann A, Frigault MM, Ottesen L, et al. Savolitinib versus sunitinib in patients with MET-driven, unresectable and locally advanced or metastatic papillary renal cell carcinoma: SAVOIR, a randomised, phase III trial. *Kidney Cancer* 2018;**2 Suppl 1**:S33.

Gleave 1997 *{published data only}*

Gleave M, Elhilai M, Fradet Y, Davis I, Venner P, Saad F, et al. A multicenter randomized, double blind trial of actimmune(r) interferon gamma-1b injection versus placebo for the treatment of metastatic renal cell carcinoma. In: Proceedings of the Annual Meeting of the American Society for Clinical Oncology. 1997.

Gleave 1997a *{published data only}*

Gleave M, Elhilali M, Fradet Y. A phase III randomized double-blind placebo-controlled multicenter trial comparing interferon gamma to placebo in patients with metastatic renal cell carcinoma. *Journal of Urology* 1997;**157**:A1274.

Gleave 1998 *{published data only}*

Gleave ME, Elhilali M, Fradet Y, Davis I, Venner P, Saad F, et al. Interferon gamma-1b compared with placebo in metastatic renal-cell carcinoma. Canadian urologic oncology group. *New England Journal of Medicine* 1998;**338**(18):1265-71.

Gore 2008 *{published data only}*

Gore ME. Interferon-a (IFN), interleukin-2 (IL2) and 5-fluorouracil (5FU) vs IFN alone in patients with metastatic renal cell carcinoma (mRCC): results of the randomised MRC/EORTC RE04 trial. *Journal of Clinical Oncology* 2008;**26**(15 Pt 1):259.

Gore 2010 *{published data only}*

Gore ME, Griffin CL, Hancock B, Patel PM, Pyle L, Aitchison M, et al. Interferon alfa-2a versus combination therapy with interferon alfa-2a, interleukin-2, and fluorouracil in patients with untreated metastatic renal cell carcinoma (MRC RE04/EORTC GU 30012): an open-label randomised trial. *Lancet* 2010;**375**(9715):641-8.

Gruenwald 2020 *{published data only}*

Gruenwald V, Ivanyi P, Jacobasch L, Zahn MO, Hubner A, Von Weikersthal LF, et al. A phase III study testing the role of proactive coaching on patient reported outcome in advanced or metastatic renal cell carcinoma treated with sunitinib or a combination of axitinib + checkpoint inhibitor (CPI) in first

line therapy (PREPARE). *Oncology Research and Treatment* 2020;**43**:224.

Haas 2016 *{published data only}*

Haas NB, Manola J, Uzzo RG, Flaherty KT, Wood CG, Kane C, et al. Adjuvant sunitinib or sorafenib for high-risk, non-metastatic renal-cell carcinoma (ECOG-ACRIN E2805): a double-blind, placebo-controlled, randomised, phase 3 trial. *Lancet* 2016;**387**(10032):2008-16.

Hainsworth 2015 *{published data only}*

Hainsworth JD, Mace JR, Reeves JA, Crane EJ, Hamid O, Stille JR, et al. Randomized phase II study of sunitinib + CXCR4 inhibitor LY2510924 versus sunitinib alone in first-line treatment of patients with metastatic renal cell carcinoma. *Journal of Clinical Oncology: ASCO annual meeting proceedings* 2015;**33**(15 Suppl 1):4547.

Hainsworth 2016 *{published data only}*

Hainsworth JD, Reeves JA, Mace JR, Crane EJ, Hamid O, Stille JR, et al. A randomized, open-label phase 2 study of the CXCR4 inhibitor LY2510924 in combination with sunitinib versus sunitinib alone in patients with metastatic renal cell carcinoma (RCC). *Targeted Oncology* 2016;**11**(5):643-53.

Han 2002 *{published data only}*

Han KR, Pantuck AJ, Belldegrun AS. A randomized phase III trial of high dose interleukin-2 versus subcutaneous interleukin-2/interferon. *Current Urology Reports* 2002;**3**(1):11-2.

Harima 1990 *{published data only}*

Harima M, Yasumoto R, Asakawa M, Horii A, Nishio S, Sakamoto W, et al. Comparison of efficacy between interferon-alpha (IFN) plus fluoropyrimidine (FP) and IFN alone in advanced renal cell cancer (RCC): preliminary results. *Journal of Cancer Research and Clinical Oncology* 1990;**116**:533.

Henriksson 1998 *{published data only}*

Henriksson R, Nilsson S, Colleen S, Wersall P, Helsing M, Zimmerman R, et al. Survival in renal cell carcinoma - a randomized evaluation of tamoxifen vs interleukin 2, alpha-interferon (leucocyte) and tamoxifen. *British Journal of Cancer* 1998;**77**(8):1311-7.

Hutson 2006 *{published data only}*

Hutson TE, Bukowski RM. A phase II study of GW786034 using a randomized discontinuation design in subjects with locally recurrent or metastatic clear-cell renal cell carcinoma. *Clinical Genitourinary Cancer* 2006;**4**(4):296-8.

Hutson 2021 *{published data only}*

Hutson TE, Carthon BC, Yorio J, Babu S, McKean HA, Percent IJ, et al. Safety and efficacy outcomes with nivolumab plus ipilimumab in patients with advanced renal cell carcinoma and low Karnofsky performance status: results from the CheckMate 920 trial. In: Journal of Clinical Oncology. Conference. Vol. 39. 2021.

ISRCTN95351638 *{published data only}*

Collinson F, Brown S, Buckley H, Ainsworth G, Howard H, Poad H, et al. PRISM: a randomised phase II trial of nivolumab in combination with alternatively scheduled ipilimumab in first-line treatment of patients with advanced or metastatic renal cell carcinoma. *Annals of Oncology* 2018;**29 Suppl 8**:viii331.

EUCTR2017-001476-33-GB. A randomised phase II trial comparing two different ways of combining nivolumab and ipilimumab in first-line treatment of patients with advanced or metastatic renal cell carcinoma. who.int/trialsearch/Trial2.aspx?TrialID=EUCTR2017-001476-33-GB 2018.

Jager 2005 *{published data only}*

Jager E, Heinzer H, Grimm MO, Krause S, Scheuring U, Siebels M. Randomized phase III trial of the multiple kinase inhibitor sorafenib (BAY 43-9006) in patients with advanced renal cell carcinoma (RCC). In: Onkologie. Vol. 28 Suppl 3. 2005:1–275.

Jayson 1998 *{published data only}*

Jayson GC, Middleton M, Lee SM, Ashcroft L, Thatcher N. A randomized phase II trial of interleukin 2 and interleukin 2-interferon alpha in advanced renal cancer. *British Journal of Cancer* 1998;**78**(3):366-9.

Jeon 1999 *{published data only}*

Jeon SH, Chang SG. The effect of immunotherapy based on interferon - alpha in advanced renal cell carcinoma. *Journal of the Korean Cancer Association* 1999;**31**(5):986-94.

JPRN-JapicCTI-122014 *{published data only}*

JPRN-JapicCTI-122014. A Phase 3 Study of ONO-4538/BMS-936558. who.int/trialsearch/Trial2.aspx?TrialID=JPRN-JapicCTI-122014 2012.

JPRN-jRCTs031180024 *{published data only}*

JPRN-jRCTs031180024. NIVOIGERCC. who.int/trialsearch/Trial2.aspx?TrialID=JPRN-jRCTs031180024 2018.

JPRN-UMIN000001995 *{published data only}*

JPRN-UMIN000001995. Randomized parallel group phase II clinical study of sorafenib/interferon combination therapy and sunitinib monotherapy for advanced renal cell carcinoma. who.int/trialsearch/Trial2.aspx?TrialID=JPRN-UMIN000001995 2009.

Kinouchi 2004 *{published data only}*

Kinouchi T, Sakamoto J, Tsukamoto T, Akaza H, Kubota J, Ozono S, et al. Prospective randomized trial of natural interferon-alpha (IFN) versus IFN + cimetidine in advanced renal cell carcinoma with pulmonary metastases. *Journal of Clinical Oncology* 2004;**22**(14 Suppl):4676.

Kinouchi 2006 *{published data only}*

Kinouchi T, Sakamoto J, Tsukamoto T, Akaza H, Kubota Y, Ozono S, et al. Prospective randomized trial of natural interferon-alpha versus natural interferon-alpha plus cimetidine in advanced renal cell carcinoma with pulmonary metastasis. *Journal of Cancer Research and Clinical Oncology* 2006;**132**(8):499-504.

Larkin 2019 *{published data only}*

Larkin JM, Tykodi SS, Donskov F, Lee JL, Szczylik C, Malik J, et al. First-line pembrolizumab (pembro) monotherapy in advanced clear cell renal cell carcinoma (ccRCC): updated follow-up for KEYNOTE-427 cohort A. *Annals of Oncology* 2019;**30 Suppl 5**:v381-2.

Law 1995 *{published data only}*

Law TM, Motzer RJ, Mazumdar M, Sell KW, Walther PJ, O'Connell M, et al. Phase III randomized trial of interleukin-2 with or without lymphokine- activated killer cells in the treatment of patients with advanced renal cell carcinoma. *Cancer* 1995;**76**(5):824-32.

Lee 2020 *{published data only}*

Lee JL, Ziobro M, Suarez C, Langiewitz P, Matveev VB, Wiechno P, et al. First-line pembrolizumab (pembro) monotherapy in advanced non-clear cell renal cell carcinoma (nccRCC): updated follow-up for KEYNOTE-427 cohort B. *Journal of Clinical Oncology: ASCO annual meeting proceedings* 2020;**38**(15):5034.

Lee 2021 *{published data only}*

Lee CH, Voss MH, Carlo MI, Chen YB, Reznik E, Knezevic A, et al. Nivolumab plus cabozantinib in patients with non-clear cell renal cell carcinoma: results of a phase 2 trial. *Journal of Clinical Oncology: ASCO annual meeting proceedings* 2021;**39**(15 Suppl):4509.

Lindskog 2020 *{published data only}*

Lindskog M, Laurell A, Kjellman A, Melichar B, Niezabitowski J, Maroto P, et al. A randomized phase II study with ilixadencel, a cell-based immune primer, plus sunitinib versus sunitinib alone in synchronous metastatic renal cell carcinoma. *Journal of Clinical Oncology* 2020;**38**(5):11.

Lissoni 1993 *{published data only}*

Lissoni P, Barni S, Ardizzoia A, Andres M, Scardino E, Cardellini P, et al. A randomized study of low-dose interleukin-2 subcutaneous immunotherapy versus interleukin-2 plus interferon-alpha as first line therapy for metastatic renal cell carcinoma. *Tumori* 1993;**79**(6):397-400.

Liu 2012 *{published data only}*

Liu L, Zhang W, Qi X, Li H, Yu J, Wei S, et al. Randomized study of autologous cytokine-induced killer cell immunotherapy in metastatic renal carcinoma. *Clinical Cancer Research* 2012;**18**(6):1751-9.

Lummen 1996 *{published data only}*

Lummen G, Goepel M, Mollhoff S, Hinke A, Otto T, Rubben H. Phase II study of interferon-gamma versus interleukin-2 and interferon-alpha2b in metastatic renal cell carcinoma. *Journal of Urology* 1996;**155**(2):455-8.

Madhusudan 2004 *{published data only}*

Madhusudan S, Protheroe A, Vasey P, Patel P, Selby P, Altman D, et al. A randomized phase II study of interferon alpha alone or in combination with thalidomide in metastatic renal cancer. *Journal of Clinical Oncology* 2004;**22**(14 Suppl):4742.

McDermott 2001 *{published data only}*

McDermott D, Flaherty L, Clark J, Weiss G, Logan T, Gordon M, et al. A randomized phase III trial of high-dose interleukin-2 (HD IL2) versus subcutaneous (SC) Il2/interferon (IFN) in patients with metastatic renal cell carcinoma (RCC). In: Proceedings of The American Society of Clinical Oncology. Vol. 20 Pt 1. 2001:172a, Abstract 685.

McDermott 2005 *{published data only}*

McDermott DF, Regan MM, Clark JI, Flaherty LE, Weiss GR, Logan TF, et al. Randomized phase III trial of high-dose interleukin-2 versus subcutaneous interleukin-2 and interferon in patients with metastatic renal cell carcinoma. *Journal of Clinical Oncology* 2005;**23**(1):133-41.

McDermott 2013 *{published data only}*

McDermott DF, Manola J, Pins M, Flaherty KT, Atkins MB, Dutcher JP, et al. The BEST trial (E2804): a randomized phase II study of VEGF, RAF kinase, and mTOR combination targeted therapy (CTT) with bevacizumab (bev), sorafenib (sor), and temsirolimus (tem) in advanced renal cell carcinoma (RCC). *Journal of Clinical Oncology. Conference* 2013;**31**(6 Suppl 1):345.

McDermott 2020 *{published data only}*

McDermott DF, Lee JL, Bjarnason GA, Larkin JM, Gafanov R, Kochenderfer MD, et al. First-line pembrolizumab (pembro) monotherapy in advanced clear cell renal cell carcinoma (ccRCC): updated follow-up for KEYNOTE-427 cohort A. *Journal of Clinical Oncology: ASCO annual meeting proceedings* 2020;**38**(15):5069.

Mickisch 2001 *{published data only}*

Mickisch GH, Garin A, van Poppel H, de Prijck L, Sylvester R. Radical nephrectomy plus interferon-alfa-based immunotherapy compared with interferon alfa alone in metastatic renal-cell carcinoma: a randomised trial. *Lancet* 2001;**358**(9286):966-70.

Minasian 1993 *{published data only}*

Minasian LM, Motzer RJ, Gluck L, Mazumdar M, Vlamis V, Krown SE. Interferon alfa-2a in advanced renal cell carcinoma: treatment results and survival in 159 patients with long-term follow-up. *Journal of Clinical Oncology* 1993;**11**(7):1368-75.

Molina 2009 *{published data only}*

Molina AM, Ginsberg M, Sweeney S, Myrie J, Motzer R J. Efficacy and safety of everolimus treatment in metastatic renal cell carcinoma. *American Journal of Hematology/Oncology* 2009;**8**(4):166-169.

Motzer 2001 *{published data only}*

Motzer RJ, Rakhit A, Thompson JA, Nemunaitis J, Murphy BA, Ellerhorst J, et al. Randomized multicenter phase II trial of subcutaneous recombinant human interleukin-12 versus interferon-alpha2a for patients with advanced renal cell carcinoma. *Journal of Interferon and Cytokine Research* 2001;**21**(4):257-63.

Mulders 2012 *{published data only}*

Mulders P, Hawkins R, Nathan P, de Jong I, Osanto S, Porfiri E, et al. Cediranib monotherapy in patients with advanced renal

cell carcinoma: results of a randomised phase II study. *European Journal of Cancer* 2012;**48**(4):527-37.

Naglieri 1998 *{published data only}*

Naglieri E, Gebbia V, Durini E, Lelli G, Abbate I, Selvaggi FP, et al. Standard interleukin-2 (IL-2) and interferon-alpha immunotherapy versus an IL-2 and 4-epirubicin immuno-chemotherapeutic association in metastatic renal cell carcinoma. *Anticancer Research* 1998;**18**(3B):2021-6.

NCT00002737 *{published data only}*

NCT00002737. Interferon alfa with or without isotretinoin in treating patients with metastatic kidney cancer. clinicaltrials.gov/show/NCT00002737 2004.

NCT00005966 *{published data only}*

NCT00005966. Interferon alfa-2b with or without thalidomide in treating patients with metastatic or unresectable kidney cancer. clinicaltrials.gov/show/NCT00005966 2003.

NCT00019539 *{published data only}*

NCT00019539. Monoclonal antibody therapy in treating patients with advanced kidney cancer. clinicaltrials.gov/show/NCT00019539 2009.

NCT00027664 *{published data only}*

NCT00027664. Interferon alfa with or without thalidomide in treating patients with metastatic kidney cancer. clinicaltrials.gov/show/NCT00027664 2003.

NCT00053820 *{published data only}*

NCT00053820. Interferon alfa with or without interleukin-2 and fluorouracil in treating patients with advanced metastatic kidney cancer. clinicaltrials.gov/show/NCT00053820 2003.

NCT00073307 *{published data only}*

Eisen T, Bukowski RM, Staehler M, Szczylik C, Oudard S, Stadler WM, et al. Randomized phase III trial of sorafenib in advanced renal cell carcinoma (RCC): impact of crossover on survival. *Journal of Clinical Oncology: ASCO annual meeting proceedings* 2006;**24**(18S):4524.

Negrier S, Jager E, Porta C, McDermott D, Moore M, Bellmunt J, et al. Efficacy and safety of sorafenib in patients with advanced renal cell carcinoma with and without prior cytokine therapy, a subanalysis of TARGET. *Medical Oncology* 2010;**27**(3):899-906.

NCT00100906 *{published data only}*

NCT00100906. Sequential ATRA then IL-2 for modulation of dendritic cells and treatment of metastatic renal cell cancer. clinicaltrials.gov/show/NCT00100906 2005.

NCT00378703 *{published data only}*

NCT00378703. Bevacizumab, sorafenib tosylate, and temsirolimus in treating patients with metastatic kidney cancer. clinicaltrials.gov/show/NCT00378703 2006.

NCT00416871 *{published data only}*

NCT00416871. Interleukin-2 and interferon in treating patients with metastatic kidney cancer. clinicaltrials.gov/show/NCT00416871 2006.

NCT00467025 *{published data only}*

Inc Amgen. A randomized, double blinded, multi-center study 2 study to estimate the efficacy and evaluate the safety and tolerability of sorafenib in combination with AMG 386 or placebo in subjects with metas. https://clinicaltrials.gov/ct2/show/NCT00467025.

NCT00491738 *{published data only}*

NCT00491738. A study evaluating the efficacy and safety of sunitinib with or without bevacizumab in first-line patients with metastatic renal cell cancer (SABRE-R). clinicaltrials.gov/show/NCT00491738 2007.

NCT00709995 *{published data only}*

NCT00709995. A study for participants with metastatic renal cell carcinoma. clinicaltrials.gov/show/NCT00709995 2008.

NCT00835978 *{published data only}*

Rini BI, Gruenwald V, Fishman MN, Melichar B, Ueda T, Karlov P, et al. Axitinib for first-line metastatic renal cell carcinoma (mRCC): overall efficacy and pharmacokinetic (PK) analyses from a randomized phase II study. *Clinical Advances in Hematology and Oncology* 2012;**10**(9):6-8.

Rini BI, Melichar B, Fishman MN, Oya M, Pithavala YK, Chen Y et al. Axitinib dose titration: analyses of exposure, blood pressure and clinical response from a randomized phase II study in metastatic renal cell carcinoma. *Annals of Oncology* 2015;**26**(7):1372-7.

Rini BI, Melichar B, Ueda T, Grünwald V, Fishman MN, Arranz JA, et al. Axitinib with or without dose titration for first-line metastatic renal-cell carcinoma: a randomised double-blind phase 2 trial. *Lancet Oncology* 2013;**14**(12):1233-42.

Rini BI, Tomita Y, Melichar B, Ueda T, Gruenwald V, Fishman MN, et al. Overall survival analysis from a randomized phase II study of axitinib with or without dose titration in first-line metastatic renal cell carcinoma. *Clinical Genitourinary Cancer* 2016;**14**(6):499-503.

Tomita Y, Uemura H, Oya M, Shinohara N, Habuchi T, Fujii Y, et al. Patients with metastatic renal cell carcinoma who benefit from axitinib dose titration: analysis from a randomised, double-blind phase II study. *BMC Cancer* 2019;**19**(1):17.

NCT00873236 *{published data only}*

EUCTR2008-006414-19-GB. Dynamic contrast enhanced MRI (DCE-MRI) assessment of the vascular changes induced with bevacizumab alone and in combination with interferon-a in patients with advanced renal cell carcinoma. - DCE-MRI changes in tumour vaculature with Bevacizumab and Interferon in advanced RCC. who.int/trialsearch/Trial2.aspx?TrialID=EUCTR2008-006414-19-GB 2009.

NCT01164228 *{published data only}*

NCT01164228. Sunitinib malate with or without gemcitabine hydrochloride in treating patients with advanced kidney cancer that cannot be removed by surgery. clinicaltrials.gov/show/NCT01164228 2010.

NCT01223027 *{published data only}*

NCT01223027. Study of dovitinib versus sorafenib in patients with metastatic renal cell carcinoma. clinicaltrials.gov/show/NCT01223027 2010.

NCT01408004 *{published data only}*

NCT01408004. Rotating pazopanib and everolimus to avoid resistance. clinicaltrials.gov/show/NCT01408004 2011.

NCT01444807 *{published data only}*

NCT01444807. Evaluate the efficacy of sorafenib in renal cell carcinoma patients after a radical resection of the metastases. clinicaltrials.gov/show/NCT01444807 2011.

NCT01616186 *{published data only}*

NCT01616186. Everolimus/sorafenib or sunitinib in patients with metastatic renal cell carcinoma (RCC). clinicaltrials.gov/show/NCT01616186 2012.

NCT01664182 *{published data only}*

NCT01664182. Trebananib with or without bevacizumab, pazopanib hydrochloride, sorafenib tosylate, or sunitinib malate in treating patients with advanced kidney cancer. clinicaltrials.gov/show/NCT01664182 2012.

NCT01673386 *{published data only}*

NCT01673386. A subject treatment preference study of tivozanib hydrochloride versus sunitinib in subjects with metastatic renal cell carcinoma. clinicaltrials.gov/show/NCT01673386 2012.

NCT01727089 *{published data only}*

NCT01727089. Bevacizumab with or without TRC105 in treating patients with metastatic kidney cancer. clinicaltrials.gov/show/NCT01727089 2012.

NCT01727336 *{published data only}*

NCT01727336. Study of dalantercept and axitinib in patients with advanced renal cell carcinoma. clinicaltrials.gov/show/NCT01727336 2012.

NCT01793636 *{published data only}*

NCT01793636. A study comparing AZD2014 vs everolimus in patients with metastatic renal cancer. clinicaltrials.gov/show/NCT01793636 2013.

NCT02014636 *{published data only}*

NCT02014636. Safety and efficacy study of pazopanib and MK 3475 in advanced renal cell carcinoma (RCC; KEYNOTE-018). clinicaltrials.gov/show/NCT02014636 2013.

NCT02127710 *{published data only}*

EUCTR2014-001858-41-ES. This is a study designed to evaluate the effectiveness and safety of AZD6094 in patients with papillary renal cell cancer. who.int/trialsearch/Trial2.aspx?TrialID=EUCTR2014-001858-41-ES 2015.

NCT02667886 *{published data only}*

NCT02667886. Trial of X4P-001 in patients with advanced renal cell carcinoma. clinicaltrials.gov/show/NCT02667886 2016.

NCT02724020 *{published data only}*

NCT02724020. MLN0128 and MLN0128 + MLN1117 compared with everolimus in the treatment of adults with advanced or metastatic clear-cell renal cell carcinoma. clinicaltrials.gov/show/NCT02724020 2016.

NCT02960906 *{published data only}*

Epaillard N, Simonaggio A, Elaidi R, Azzouz F, Braychenko E, Thibault C et al. A BIOmarker driven trial with nivolumab and ipilimumab or VEGFR tKi in naive metastatic kidney cancer (BIONIKK). clinicaltrials.gov/ct2/show/NCT02960906.

NCT03035630 *{published data only}*

NCT03035630. Sunitinib followed by avelumab or the reverse for metastatic renal cell carcinoma. clinicaltrials.gov/show/NCT03035630 2017.

NCT03092856 *{published data only}*

NCT03092856. Axitinib with or without anti-OX40 antibody PF-04518600 in treating patients with metastatic kidney cancer. clinicaltrials.gov/show/NCT03092856 2017.

NCT03095040 *{published data only}*

NCT03095040. CM082 combined with everolimus in chinese patients with metastatic renal cell carcinoma. clinicaltrials.gov/show/NCT03095040 2017.

NCT03173560 *{published data only}*

NCT03173560. Trial to assess safety and efficacy of lenvatinib (18 mg vs. 14 mg) in combination with everolimus in participants with renal cell carcinoma. clinicaltrials.gov/ct2/show/NCT03173560 2017.

NCT03501381 *{published data only}*

NCT03501381. High dose IL 2 and entinostat in RCC. clinicaltrials.gov/show/NCT03501381 2018.

NCT03595124 *{published data only}*

NCT03595124. Axitinib and nivolumab in treating participants with unresectable or metastatic TFE/translocation renal cell carcinoma. clinicaltrials.gov/show/NCT03595124 2018.

NCT03829111 *{published data only}*

NCT03829111. CBM588, nivolumab, and ipilimumab in treating patients with stage IV or advanced kidney cancer. clinicaltrials.gov/show/NCT03829111 2019.

NCT04195750 *{published data only}*

NCT04195750. A study of MK-6482 versus everolimus in participants with advanced renal cell carcinoma (MK-6482-005). clinicaltrials.gov/show/NCT04195750 2019.

NCT04300140 *{published data only}*

NCT04300140. Efficacy and safety study of AVB-S6-500 in patients with clear cell renal cell carcinoma. clinicaltrials.gov/show/NCT04300140 2020.

Negrier 1996 *{published data only}*

Negrier S, Escudier B, Lasset C, Savary J, Douillard JY, Chevreau C, et al. The FNCLCC Crecy trial: interleukin 2 (IL2) + interferon (IFN) is the optimal treatment to induce responses

in metastatic renal cell carcinoma (mRCC). In: Proceedings of The American Society of Clinical Oncology. Vol. 15. 1996:248, Abstract 629.

Negrier 1997 *{published data only}*

Negrier S, Escudier B, Douillard JY, Lesimple T, Rossi JF, Viens P, et al. Randomized study of interleukin-2 and interferon with or without 5-FU (FUCY study) in metastatic renal cell carcinoma. In: Proceedings of The American Society of Clinical Oncology. Vol. 16. 1997:326a, Abstract 1161.

Negrier 1998 *{published data only}*

Negrier S, Escudier B, Lasset C, Douillard JY, Savary J, Chevreau C, et al. Recombinant human interleukin-2, recombinant human interferon alfa-2a, or both in metastatic renal-cell carcinoma. *New England Journal of Medicine* 1998;**338**(18):1272-8.

Negrier 2000 *{published data only}*

Negrier S, Caty A, Lesimple T, Douillard JY, Escudier B, Rossi JF, et al. Treatment of patients with metastatic renal carcinoma with a combination of subcutaneous interleukin-2 and interferon alfa with or without fluorouracil. Groupe Francais d'Immunotherapie, Federation Nationale des Centres de Lutte Contre le Cancer. *Journal of Clinical Oncology* 2000;**18**(24):4009-15.

Negrier 2006 *{published data only}*

Negrier S, Perol D, Ravaud A, Bay JO, Oudard S, Fargeot P, et al. Is intravenous (iv) IL2 superior to subcutaneous (sc) IL2 in good prognosis patients (pts) with metastatic renal cell carcinoma (mRCC) receiving a combination of IL2 and alpha interferon (IFN)? Results of the prospective randomized PERCY duo trial. *Journal of Clinical Oncology: ASCO annual meeting proceedings* 2006;**24**(18 Suppl):4536.

Negrier 2007 *{published data only}*

Negrier S, Perol D, Ravaud A, Chevreau C, Bay JO, Delva R, et al. Medroxyprogesterone, interferon alfa-2a, interleukin 2, or combination of both cytokines in patients with metastatic renal carcinoma of intermediate prognosis: results of a randomized controlled trial. *Cancer* 2007;**110**(11):2468-77.

Negrier 2008 *{published data only}*

Negrier S, Perol D, Ravaud A, Bay JO, Oudard S, Chabaud S, et al. Randomized study of intravenous versus subcutaneous interleukin-2, and IFNalpha in patients with good prognosis metastatic renal cancer. *Clinical Cancer Research* 2008;**14**(18):5907-12.

Nosov 2010 *{published data only}*

Nosov DA, Bhargava P, Esteves B, Al-Adhami M, Lipatov O, Lyulko AA, et al. Phase 2 randomized discontinuation trial (RDT) of tivozanib in patients with renal cell carcinoma (RCC): results in patients randomized to tivozanib vs. placebo. *Annals of Oncology* 2010;**8**:viii271.

Nosov 2012 *{published data only}*

Nosov DA, Bhargava P, Esteves B, Strahs AL, Lipatov ON, Lyulko AA, et al. Final analysis of the phase 2 randomized discontinuation trial of tivozanib (AV-951) versus placebo in patients with renal cell carcinoma. *Journal of Clinical Oncology* 2012;**29**(15 Suppl):4550.

Pal 2015 *{published data only}*

Pal S, Azad A, Bhatia S, Drabkin H, Costello B, Sarantopoulos J, et al. P0125 a phase 1/2 trial of BNC105P with everolimus in metastatic renal cell carcinoma. *Clinical Cancer Research* 2015;**21**(15):3420-7.

Pal 2021a *{published data only}*

Pal SK, Puente J, Heng DY, Glen H, Koralewski P, Stroyakovskiy D, et al. Phase II trial of lenvatinib (LEN) at two starting doses + everolimus (EVE) in patients (pts) with renal cell carcinoma (RCC): results by independent imaging review (IIR) and prior immune checkpoint inhibition (ICI). *Journal of Clinical Oncology* 2021;**39**(6 Suppl):307.

Passalacqua 2010 *{published data only}*

Passalacqua R, Buzio C, Buti S, Porta C, Labianca R, Pezzuolo D, et al. Phase III, randomised, multicentre trial of maintenance immunotherapy with low-dose interleukin-2 and interferon-alpha for metastatic renal cell cancer. *Cancer Immunology, Immunotherapy* 2010;**59**(4):553-61.

Plimack 2015 *{published data only}*

Plimack ER, Hammers HJ, Rini BI, McDermott DF, Redman BG, Kuzel TM, et al. Updated survival results from a randomized, doseranging phase II study of nivolumab (NIVO) in metastatic renal cell carcinoma (MRCC). *Journal of Clinical Oncology* 2015;**33**(15 Suppl):4553.

Pyrhonen 1995 *{published data only}*

Pyrhonen S, Salminen E, Lehtonen T, Nurmi M, Tammela T, Juusela H, et al. Recombinant interferon alfa-2a (IFN) with vinblastine vs vinblastine alone in advanced renal cell carcinoma (RCC). A phase III study. In: Proceedings of The American Society of Clinical Oncology. Vol. 14. 1995:253, Abstract 686.

Pyrhonen 1996 *{published data only}*

Pyrhonen S, Salminen E, Lehtonen T, Nurmi M, Tammela T, Juusela H, et al. Recombinant interferon alfa-2a with vinblastine vs vinblastine alone in advanced renal cell carcinoma: a phase III study. In: Proceedings of The American Society of Clinical Oncology. Vol. 15. 1996:244, Abstract 614.

Pyrhonen 1999 *{published data only}*

Pyrhonen S, Salminen E, Ruutu M, Lehtonen T, Nurmi M, Tammela T, et al. Prospective randomized trial of interferon alfa-2a plus vinblastine versus vinblastine alone in patients with advanced renal cell cancer. *Journal of Clinical Oncology* 1999;**17**(9):2859-67.

Ravaud 2006 *{published data only}*

Ravaud A, Gardner J, Hawkins R, der Maase H, Zantl N, Harper P, et al. Efficacy of lapatinib in patients with high tumor EGFR expression: results of a phase III trial in advanced renal cell carcinoma (RCC). *Journal of Clinical Oncology: ASCO annual meeting proceedings 2006* 2006;**24**(18 Suppl):4502.

Ravaud 2016 *{published data only}*

Ravaud A, Motzer RJ, Pandha HS, Staehler M, George D, Pantuck AJ, et al. PR phase III trial of sunitinib (SU) vs placebo (PBO) as adjuvant treatment for high-risk renal cell carcinoma (RCC) after nephrectomy (S-TRAC). *Annals of Oncology* 2016;**27**(6):1-36.

Rexer 2017 *{published data only}*

Rexer H, Steiner T, Gruenwald V. First-line therapy in advanced renal cell carcinoma: a randomized phase II study to examine early switch of tyrosine kinase inhibitors to nivolumab compared to continued tyrosine kinase inhibitor therapy in patients with advanced or metastatic renal cell carcinoma and stable disease after three months of treatment (NIVOSWITCH)-AN 38/15 of the AUO.. *Urologe* 2017;**56**(4):509-11.

Richards 1977 *{published data only}*

Richards F, Muss HB, White DR, Cooper MR, Spurr CL. CCNU, bleomycin, and methylprednisolone with or without adriamycin in renal cel carcinoma: a randomized trial. *Cancer Treatment Reports* 1977;**61**(8):1591-3.

Rini 2011 *{published data only}*

Rini BI, Szczylik C, Tannir NM, Koralewski P, Tomczak P, Deptala A, et al. AMG 386 in combination with sorafenib in patients (pts) with metastatic renal cell cancer (mRCC): a randomized, double-blind, placebo-controlled, phase II study. *Journal of Clinical Oncology. Conference* 2011;**29**(7 Suppl 1):309.

Rini 2012 *{published data only}*

Rini B, Szczylik C, Tannir NM, Koralewski P, Tomczak P, Deptala A, et al. AMG 386 in combination with sorafenib in patients with metastatic clear cell carcinoma of the kidney: a randomized, double-blind, placebo-controlled, phase 2 study. *Cancer* 2012;**118**(24):6152-61.

Rodriguez-Vida 2020 *{published data only}*

Rodriguez-Vida A, Bamias A, Esteban E, Saez MI, Lopez-Brea M, Castellano D, et al. Randomised phase II study comparing alternating cycles of sunitinib and everolimus vs standard sequential administration in first-line metastatic renal carcinoma (SUNRISES study). *BJU international* 2020;**126**(5):559-67.

Rpcec 2017 *{published data only}*

RPCEC00000229. HEBERFERON in renal cell carcinoma. who.int/trialsearch/Trial2.aspx?TrialID=RPCEC00000229 2017.

Sternberg 2013 *{published data only}*

Sternberg C, Bracarda S, Carteni G, Lo Re G, Ruggeri EM, Masini C. Sunitinib expanded-access trial in metastatic renal cell carcinoma (mRCC) - final Italian results. *European Journal of Cancer* 2013;**49**:S644.

Szarek 2021 *{published data only}*

Szarek M, Needle MN, Rini BI, Pal SK, McDermott DF, Atkins MB, et al. Q-TWiST analysis of tivozanib versus sorafenib in patients with advanced renal cell carcinoma in the TIVO-3 study. *Clinical Genitourinary Cancer* 2021;**19**(5):468.e1-468.e5.

Tannir 2016 *{published data only}*

Tannir NM, Jonasch E, Albiges L, Altinmakas E, Ng CS, Matin SF, et al. Everolimus versus sunitinib prospective evaluation in metastatic non-clear cell renal cell carcinoma (ESPN): a randomized multicenter phase 2 trial. *European Urology* 2016;**69**(5):866-74.

Taylor 2020 *{published data only}*

Taylor MH, Lee CH, Makker V, Rasco D, Dutcus CE, Wu J, et al. Phase Ib/II trial of lenvatinib plus pembrolizumab in patients with advanced renal cell carcinoma, endometrial cancer, and other selected advanced solid tumors. *Journal of Clinical Oncology* 2020;**38**(11):1154-63.

Taylor 2020a *{published data only}*

Taylor MH, Lee CH, Makker V, Rasco D, Ductus CE, Wu J etal. Erratum: phase IB/II trial of lenvatinib plus pembrolizumab in patients with advanced renal cell carcinoma, endometrial cancer, and other selected advanced solid tumors. *Journal of Clinical Oncology* 2020;**38**(23):2702.

Thiam 2010 *{published data only}*

Thiam R, Cuenod CA, Fournier LS, Lamuraglia M, Medioni J, Barascout B, et al. Determination of a new recist threshold using everolimus treatment in metastatic renal cell carcinoma: evaluation from the RECORD-1 study. *Annals of Oncology* 2010;**8**:viii74-5.

Trump 2004 *{published data only}*

Atzpodien J, Kirchner H, Jonas U, Bergmann L, Schott H, Heynemann H, et al. Interleukin-2- and interferon alfa-2a-based immunochemotherapy in advanced renal cell carcinoma: a prospectively randomized trial of the german cooperative renal carcinoma chemoimmunotherapy group (DGCIN). *Journal of Clinical Oncology* 2004;**22**(7):1188-94.

Twardowski 2015 *{published data only}*

Twardowski P, Plets M, Plimack ER, Agarwal N, Tangen CM, Vogelzang NJ, et al. SWOG 1107: parallel (randomized) phase II evaluation of tivantinib (ARQ-197) and tivantinib in combination with erlotinib in patients (pts) with papillary renal cell carcinoma (pRCC). *Journal of Clinical Oncology: ASCO annual meeting proceedings* 2015;**33**(15 Suppl 1):4523.

Twardowski 2017 *{published data only}*

Twardowski PW, Tangen CM, Wu X, Plets MR, Plimack ER, Agarwal N, et al. Parallel (randomized) phase II evaluation of tivantinib (ARQ197) and tivantinib in combination with erlotinib in papillary renal cell carcinoma: SWOG S1107. *Kidney Cancer* 2017;**1**(2):123-32.

Verzoni 2018 *{published data only}*

Verzoni E, Ratta R, Grassi P, Salvioni R, Stagni S, Montone R, et al. TARIBO trial: targeted therapy with or without nephrectomy in metastatic renal cell carcinoma: liquid biopsy for biomarkers discovery. *Tumori* 2018;**104**(5):401-5.

Voss 2015 *{published data only}*

Voss MH, Chen Y, Chaim J, Coskey DT, Woo K, Redzematovic A, et al. A phase II trial of everolimus (E) and bevacizumab (B) in advanced non-clear cell renal cell cancer (ncRCC) to show

efficacy in patients (pts) with papillary features. *Journal of Clinical Oncology: ASCO annual meeting proceedings* 2015;**33**(15 Suppl 1):4522.

Voss 2019 *{published data only}*

Voss MH, Azad AA, Hansen AR, Gray JE, Welsh SJ, Achour I, et al. Results from a randomised phase I/II trial evaluating the safety and antitumour activity of anti-PD-1 (MEDI0680)/anti-PD-L1 (durvalumab) vs anti-PD-1 (nivolumab) alone in metastatic clear cell renal cell carcinoma (ccRCC). *Annals of Oncology* 2019;**30**:v516.

Witte 1995 *{published data only}*

Witte RS, Leong T, Emstoff MS, Krigel RL, Oken MM, Harris J, et al. A phase II study of interleukin-2 with and without beta-interferon in the treatment of advanced renal cell carcinoma. *Investigational New Drugs* 1995;**13**(3):241-7.

Wood 2013 *{published data only}*

Wood GC, Figlin RA. ADAPT: an ongoing international phase 3 randomized trial of autologous dendritic cell immunotherapy (AGS 003) plus standard treatment in advanced renal cell carcinoma (RCC). *Journal of Clinical Oncology* 2013;**32**(4 Suppl):449.

Wright 2020 *{published data only}*

Wright KM. Final data analysis supports tivozanib as superior treatment for patients with RCC. *Oncology* 2020;**34**(7):257.

Yang 2002 *{published data only}*

Yang JC, Haworth L, Steinberg SM, Rosenberg SA, Novotny W. A randomized double-blind placebo-controlled trial of bevacizumab (anti-VEGF antibody) demonstrating a prolongation in time to progression in patients with metastatic renal cancer. Proceedings of The American Society of Clinical Oncology 2002;**21 Pt 1**:5a, Abstract 15.

Yang 2003 *{published data only}*

Yang JC, Haworth L, Sherry RM, Hwu P, Schwartzentruber DJ, Topalian SL, et al. A randomized trial of bevacizumab, an anti-vascular endothelial growth factor antibody, for metastatic renal cancer. *New England Journal of Medicine* 2003;**349**(5):427-34.

Zhou 2016 *{published data only}*

Zhou AP, Ma J, Bai Y, Song Y, Li H, Xie X, et al. Anlotinib versus sunitinib as first line treatment for metastatic renal cell carcinoma (mRCC): preliminary results from a randomized phase II clinical trial. *Journal of Clinical Oncology. Conference* 2016;**34**(15 Suppl):4565.

Zhou 2019 *{published data only}*

Zhou A, Bai Y, Song Y, Luo H, Ren X, Wang X, et al. Anlotinib versus sunitinib as first-line treatment for metastatic renal cell carcinoma: a randomized phase II clinical trial. *Oncologist* 2019;**24**(8):e702.

References to studies awaiting assessment

Liu 2017 *{published data only}*

Liu C, Wang X, Mo, J, Yang H, Liu H, Liao H. Efficacy of sunitinib in the treatment of advanced renal cellcarcinoma and its effect on survival time [Chinese]. *Anti-Tumor Pharmacy* 2017;**7**(2):195-9.

NCT01217931 *{published data only}*

Sequential Two-agent Assessment in Renal Cell Carcinoma Therapy: The START Trial. https://www.clinicaltrials.gov/ct2/show/NCT01217931.

NCT01688973 *{published data only}*

Tivantinib with or without erlotinib hydrochloride in treating patients eith metastatic or locally advanced kidney cancer that cannot be removed by surgery. https://www.clinicaltrials.gov/ct2/show/NCT01688973.

NCT01829841 *{published data only}*

A Study of famitinib in patients with advanced metastatic renal cell cancer. https://www.clinicaltrials.gov/ct2/show/NCT01829841.

NCT03541902 *{published data only}*

Cabozantinib or sunitinib malate in treating participants with metastatic variant histology renal cell carcinoma. https://www.clinicaltrials.gov/ct2/show/NCT03541902.

References to ongoing studies

EUCTR2008-000928-71-IT *{published data only}*

EUCTR2007-002556-41-AT. A randomized, open-label, 2-arm, multicentre, phase II study to evaluate the safety and efficacy of trivax, a dendritic cell-based interleukin-12 secreting autologous cancer vaccine, in combination with sunitinib compared with sunitinib alone as first line treatment for patients with metastatic renal-cell carcinoma (mRCC). - Decavarec study TRX 1.0. who.int/trialsearch/Trial2.aspx?TrialID=EUCTR2007-002556-41-AT 2008.

EUCTR2008-000928-71-IT. TWIST. Randomized prospective phase II study of temsirolimus with or without low-dose interferon alpha in metastatic non-clear renal cell carcinoma: GOIRC study 02/2008 - TWIST. trialsearch.who.int/Trial2.aspx?TrialID=EUCTR2008-000928-71-IT 2008.

EUCTR2008-000928-71-IT. TWIST. Randomized prospective phase II study of temsirolimus with or without low-dose interferon alpha in metastatic non-clear renal cell carcinoma: GOIRC study 02/2008 - TWIST. who.int/trialsearch/Trial2.aspx?TrialID=EUCTR2008-000928-71-IT 2008.

NCT02210117 *{published data only}*

Nivolumab with or without bevacizumab or ipilimumab before surgery in treating patients with metastatic kidney cancer that can be removed by surgery. clinicaltrials.gov/ct2/show/NCT02210117 2014. [ClinicalTrials.gov: NCT02210117]

NCT02996110 *{published data only}*

A study to test combination treatments in people with advanced renal cell carcinoma (FRACTION-RCC). clinicaltrials.gov/ct2/show/NCT02996110 2016. [ClinicalTrials.gov: NCT02996110]

NCT03075423 *{published data only}*

Ahrens M, Escudier B, Boleti E, Grimm MO, Goupil MG, Barthelemy P, et al. A randomized phase II study of nivolumab plus ipilimumab versus standard of care in previously untreated and advanced non-clear cell renal cell carcinoma (suniforecast). *Oncology Research and Treatment* 2020;**43 Suppl 1**:226. [ClinicalTrials.gov: NCT03075423]

Ahrens M, Escudier B, Haanen JB, Boleti E, Goupil MG, Grimm MO, et al. A randomized phase II study of nivolumab plus ipilimumab versus standard of care in previously untreated and advanced non-clear cell renal cell carcinoma (SUNIFORECAST). In: Journal of Clinical Oncology. Meeting Abstract | 2021 ASCO Annual Meeting I. Vol. 39. 2021. [ClinicalTrials.gov: NCT03075423]

Ahrens M, Escudier B, Haanen JB, Boleti E, Gross Goupil M, Grimm MO, et al. An ongoing, randomized phase II study of Nivolumab plus Ipilimumab versus standard of care in previously untreated and advanced non-clear cell renal cell carcinoma (SUNIFORECAST). *Oncology Research and Treatment* 2021;**44 Suppl 2**:74-5. [ClinicalTrials.gov: NCT03075423]

Ahrens M, Scheich S, Gokbuget N, Boleti E, Grunwald V, Escudier B, et al. A randomised phase II study of nivolumab plus ipilimumab versus standard of care in previously untreated and advanced non-clear cell renal cell carcinoma (SUNIFORECAST). *Oncology Research and Treatment* 2018;**41**:73. [ClinicalTrials.gov: NCT03075423]

Rexer H, Bergmann L, Steiner T. A phase 2, randomized, open-label study of nivolumab combined with ipilimumab versus sunitinib monotherapy in subjects with previously untreated and advanced (unresectable or metastatic) non-clear cell renal cell carcinoma - SUNNIFORECAST. *Aktuelle Urologie* 2020;**51**(3):236-8. [ClinicalTrials.gov: NCT03075423]

Rexer H, Steiner T, Bergmann L. Nivolumab combined with ipilimumab versus sunitinib monotherapy-SUNNIFORECAST AN 41/16 of the AUO: a phase 2, randomized, open-label study in subjects with previously untreated and advanced (unresectable or metastatic) non-clear cell renal cell carcinoma.. *Urologe* 2017;**56**(6):802-3. [ClinicalTrials.gov: NCT03075423]

NCT03260894 *{published data only}*

NCT03260894. Pembrolizumab (MK-3475) plus epacadostat vs standard of care in mRCC (KEYNOTE-679/ECHO-302). clinicaltrials.gov/ct2/show/NCT03260894 2017. [ClinicalTrials.gov: NCT03260894]

NCT03592472 *{published data only}*

Aggarwal RR, Thomas S, Hauke RJ, Nordquist LT, Munster PN. RENAVIV: a randomized phase III, double-blind, placebo-controlled study of pazopanib with or without abexinostat in patients with locally advanced or metastatic renal cell carcinoma. In: Journal of Clinical Oncology (Conference). Vol. 37. 2019. [ClinicalTrials.gov: NCT03592472]

NCT03729245 *{published data only}*

NCT03729245. A study of bempegaldesleukin (NKTR-214: BEMPEG) in combination with nivolumab compared with the investigator's choice of a tyrosine kinase inhibitor (TKI) therapy (either sunitinib or cabozantinib monotherapy) for advanced metastatic renal cell carcinoma (RCC). clinicaltrials.gov/ct2/show/NCT03729245 2018. [ClinicalTrials.gov: NCT03729245]

Tannir NM, Agarwal N, Pal SK, Cho DC, Formiga M, Guo J, et al. PIVOT-09: a phase III randomized open-label study of bempegaldesleukin (NKTR-214) plus nivolumab versus sunitinib or cabozantinib (investigator's choice) in patients with previously untreated advanced renal cell carcinoma (RCC). In: Journal of Clinical Oncology. Meeting Abstract | 2020 Genitourinary Cancers Symposium. Vol. 38. 2020. [ClinicalTrials.gov: NCT03729245]

NCT03793166 *{published data only}*

Zhang T, Ballman K, Choudhury A, Chen R, Watt C, Wen Y, et al. PDIGREE: an adaptive phase 3 trial of PD-inhibitor nivolumab and ipilimumab (IPI-NIVO) with VEGF TKI cabozantinib (CABO) in metastatic untreated renal cell cancer (Alliance A031704). *Kidney Cancer* 2020;**4**:S39. [ClinicalTrials.gov: NCT03793166]

Zhang T, Ballman KV, Choudhury AD, Chen RC, Watt C, Wen Y, et al. PDIGREE: an adaptive phase 3 trial of PD-inhibitor nivolumab and ipilimumab (IPI-NIVO) with VEGF TKI cabozantinib (CABO) in metastatic untreated renal cell cancer (Alliance A031704). In: Journal of Clinical Oncology. Meeting Abstract | 2021 Genitourinary Cancers Symposium. Vol. 37. 2019. [ClinicalTrials.gov: NCT03793166]

* Zhang T, Ballman KV, Choudhury AD, Chen RC, Watt C, Wen Y, et al. PDIGREE: an adaptive phase III trial of PD-inhibitor nivolumab and ipilimumab (IPI-NIVO) with VEGF TKI cabozantinib (CABO) in metastatic untreated renal cell cancer (Alliance A031704). In: Journal of Clinical Oncology (Conference). Vol. 39. 2021. [ClinicalTrials.gov: NCT03793166]

NCT03873402 *{published data only}*

An immunotherapy study of nivolumab plus ipilimumab versus nivolumab alone in participants with advanced kidney cancer. clinicaltrials.gov/show/NCT03873402 2019. [ClinicalTrials.gov: NCT03873402]

NCT03873402. A study of nivolumab combined with ipilimumab versus nivolumab alone in participants with advanced kidney cancer. clinicaltrials.gov/ct2/show/NCT03873402 2019. [ClinicalTrials.gov: NCT03873402]

NCT03937219 *{published data only}*

Choueiri T, Albiges L, Powles T, Geng A, Mohamed N, Wang F, et al. A phase 3 study (COSMIC-313) of cabozantinib in combination with nivolumab and ipilimumab in patients with previously untreated advanced renal cell carcinoma of intermediate or poor risk. *European Urology Open Science* 2020;**21Suppl 3**:S188. [ClinicalTrials.gov: NCT03937219]

Choueiri T, Albiges L, Powles T, Mohamed N, Wang F, Motzer R. A phase 3 study (COSMIC-313) of cabozantinib in combination with nivolumab and ipilimumab in patients with previously untreated advanced renal cell carcinoma of intermediate

or poor risk. *Journal for Immunotherapy of Cancer* 2020;**8**(3 Suppl):A209. [ClinicalTrials.gov: NCT03937219]

Choueiri TK, Scheffold C, Wang F, Powles T, Motzer RJ. A phase 3 study (COSMIC-313) of cabozantinib in combination with nivolumab and ipilimumab in patients with previously untreated advanced or metastatic renal cell carcinoma (RCC) of intermediate or poor risk. *Kidney Cancer* 2020;**4 Suppl 1**:S2-3. [ClinicalTrials.gov: NCT03937219]

De Giorgi U, Choueiri TK, Albiges L, Powles T, Scheffold C, Wang F, et al. A phase 3 study (Cosmic-313) of cabozantinib in combination with nivolumab and ipilimumab in patients with previously untreated advanced renal cell carcinoma of intermediate or poor risk. *Tumori* 2020;**106**(2 Suppl):152. [ClinicalTrials.gov: NCT03937219]

NCT04090710 *{published data only}*

SBRT with combination ipilimumab/nivolumab for metastatic kidney cancer (CYTOSHRINK). clinicaltrials.gov/ct2/show/NCT04090710 2019. [ClinicalTrials.gov: NCT04090710]

NCT04203901 *{published data only}*

Dendritic cell immunotherapy plus standard treatment of advanced renal cell carcinoma. clinicaltrials.gov/ct2/show/NCT04203901 2019. [ClinicalTrials.gov: NCT04203901]

NCT04394975 *{published data only}*

Study to evaluate the efficacy and safety of toripalimab in combination with axitinib versus sunitinib monotherapy in advanced renal cell cancer. clinicaltrials.gov/ct2/show/NCT04394975 2020. [ClinicalTrials.gov: NCT04394975]

NCT04523272 *{published data only}*

A study of TQB2450 injection combined with anlotinib hydrochloride capsule versus sunitinib in subjects with advanced renal cancer. clinicaltrials.gov/ct2/show/NCT04523272 2020. [ClinicalTrials.gov: NCT04523272]

NCT04540705 *{published data only}*

A study to compare bempegaldesleukin (BEMPEG: NKTR-214) combined with nivolumab and tyrosine kinase inhibitor (TKI) to nivolumab and TKI alone in participants with previously untreated kidney cancer that is advanced or has spread (PIVOT IO 011). clinicaltrials.gov/ct2/show/NCT04540705 2020. [ClinicalTrials.gov: NCT04540705]

NCT04736706 *{published data only}*

Choueiri T, Plimack E, Powles T, Voss M, Gurney H, Silverman R, et al. Phase 3 study of pembrolizumab + belzutifan + lenvatinib or pembrolizumab/quavonlimab + lenvatinib versus pembrolizumab + lenvatinib as first-line treatment for advanced renal cell carcinoma. *Journal for ImmunoTherapy of Cancer* 2021;**9**(2 Suppl):A447. [ClinicalTrials.gov: NCT04736706]

EUCTR2020-002216-52-CZ. An open-label, randomized phase 3 study of immune and targeted combination therapies, as first-line treatment in participants with advanced clear cell renal cell carcinoma (ccRCC). trialsearch.who.int/Trial2.aspx?TrialID=EUCTR2020-002216-52-CZ 2021. [ClinicalTrials.gov: NCT04736706]

NCT04736706. A study of pembrolizumab (MK-3475) in combination with belzutifan (MK-6482) and lenvatinib (MK-7902), or pembrolizumab/quavonlimab (MK-1308A) in combination with lenvatinib, versus pembrolizumab and lenvatinib, for treatment of advanced clear cell renal cell carcinoma (MK-6482-012). clinicaltrials.gov/ct2/show/NCT04736706 2021. [ClinicalTrials.gov: NCT04736706]

Rini BI, Plimack ER, Powles TB, Voss MH, Gurney HP, Silverman R, et al. 717TiP randomized, open-label, 3-arm phase III study comparing MK-1308A + lenvatinib and pembrolizumab (pembro) + belzutifan + lenvatinib versus pembro + lenvatinib as first-line (1L) treatment for advanced clear cell renal cell carcinoma (ccRCC). *Annals of Oncology* 2021;**32 Suppl 5**:S722-3. [ClinicalTrials.gov: NCT04736706]

NCT05043090 *{published data only}*

EUCTR2021-000336-55-ES. A clinical trial to compare the effectiveness of savolitinib plus durvalumab relative to sunitinib for the treatment of renal cell cancer. trialsearch.who.int/Trial2.aspx?TrialID=EUCTR2021-000336-55-ES 2021.

NCT05043090. Savolitinib plus Durvalumab versus sunitinib and Durvalumab monotherapy in MET-Driven, unresectable and locally advanced or metastatic PRCC. clinicaltrials.gov/ct2/show/NCT05043090 2021.

NCT05096390 *{published data only}*

Axitinib +/- Pembrolizumab in first line treatment of mPRCC. clinicaltrials.gov/ct2/show/NCT05096390 2021. [ClinicalTrials.gov: NCT05096390]

UMIN 000012522 *{published data only}*

Kadono Y, Konaka H, Izumi K, Anai S, Fujimoto K, Ishibashi K, et al. Efficacy and safety of cytokines versus first-line sunitinib and second-line axitinib for patients with metastatic renal cell carcinoma (ESCAPE study): a study protocol for phase III randomized sequential open-label study. *Contemporary Clinical Trials Communications* 2019;**15**:1-6.

UMIN000012522. Efficacy and safety of cytokines versus sunitinib as first-line followed by second-line axitinib in the treatment of patients with metastatic renal cell carcinoma: a phase III randomized sequential open-label study. upload.umin.ac.jp/cgi-open-bin/ctr_e/ctr_view.cgi?recptno=R000014557 2013.

Additional references

Aguiar 2018

Aguiar P, Costa de Padua T, Noia Barreto CM, Del Giglio A. Treatment of metastatic renal cell carcinoma: latest evidence and ongoing challenges. *Clinical Medicine Insights: Urology* 2018;**11**:1-7.

Aldin 2023

Aldin A. First-line therapy for adults with advanced renal cell carcinoma - supplementary file. Open Science Framework 2023. [DOI: 10.17605/OSF.IO/HXBG5]

American Cancer Society 2022

Cancer Facts & Figures 2022. Atlanta: American Cancer Society 2022.

ASCO 2021

American Society of Clinical Oncology. Kidney cancer: introduction. https://www.cancer.net/cancer-types/kidney-cancer/introduction#:~:text=Renal%20cell%20carcinoma%20is%20the,filtration%20units%20in%20each%20kidney. (accessed 21 December 2022).

ASCO 2022

American Society of Clinical Oncology. Kidney Cancer: Statistics. https://www.cancer.net/cancer-types/kidney-cancer/statistics (accessed 21 December 2022).

Bosma 2022

Bosma NA, Warkentin MT, Gan CL, Karim S, Heng DY, Brenner DR, et al. Efficacy and safety of first-line systemic therapy for metastatic renal cell aarcinoma: a systematic review and network meta-analysis. *European Urology Open Science* 2022;**37**:14-26.

Brierley 2016

Brierley JD, Gospodarowicz MK, Wittekind C. TNM Classification of Malignant Tumours. 8th edition. Oxford: John Wiley & Sons, 2016.

Capitanio 2019

Capitanio U, Bensalah K, Bex A, Boorjian SA, Bray F, Coleman J, et al. Epidemiology of renal cell carcinoma. *European Urology* 2019;**75**(1):74-84.

Cattrini 2021

Cattrini C, Messina C, Airoldi C, Buti S, Roviello G, Mennitto A, et al. Is there a preferred first-line therapy for metastatic renal cell carcinoma? A network meta-analysis. *Therapeutic Advances in Urology* 2021;**13**:17562872211053189.

Chaimani 2012

Chaimani A, Salanti G. Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions. *Research Synthesis Methods* 2012;**3**(2):161-76.

Chaimani 2013

Chaimani A, Higgins JP, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. *PLOS One* 2013;**8**(10):e76654.

Chaimani 2019

Chaimani A, Caldwell DM, Li T, Higgins JP, Salanti G. Chapter 11: Undertaking network meta-analyses. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editor(s). Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). Cochrane, 2019. training.cochrane.org/handbook 2019.

Choueiri 2017b

Choueiri TK, Motzer RJ. Systemic therapy for metastatic renal-cell carcinoma. *New England Journal of Medicine* 2017;**376**(4):354-66.

CINeMA [Computer program]

CINeMA [Confidence in Network Meta-Analysis]. Nikolakopoulou A, Higgins JP, Papakonstantinou T, Chaimani A, Del Giovane C, Egger M, et al, Version (accessed prior to 22 October 2020). The University of Bern, The Campbell Collaboration, The Cochrane Collaboration, 2020. cinema.ispm.unibe.ch/.

Coppin 2004

Coppin C, Porzsolt F, Autenrieth M, Kumpf J, Coldman A, Wilt T. Immunotherapy for advanced renal cell cancer. *Cochrane Database of Systematic Reviews* 2004, Issue 3. Art. No: CD001425. [DOI: 10.1002/14651858.CD001425]

Dabestani 2016

Dabestani S, Thorstenson A, Lindblad P, Harmenberg U, Ljungberg B, Lundstam S. Renal cell carcinoma recurrences and metastases in primary non-metastatic patients: a population-based study. *World Journal of Urology* 2016;**34**(8):1081-6.

de Groot 2018

de Groot S, Redekop WK, Versteegh MM, Sleijfer S, Oosterwijk E, Kiemeney LA, et al. Health-related quality of life and its determinants in patients with metastatic renal cell carcinoma. *Quality of Life Research* 2018;**27**(1):115-24.

Deeks 2022

Deeks JJ, Higgins JP, Altman DG (editors). Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.3 (updated February 2022). Cochrane 2022. Available from www.training.cochrane.org/handbook 2022.

Delahunt 2019

Delahunt B, Eble JN, Egevad L, Samaratunga H. Grading of renal cell carcinoma. *Histopathology* 2019;**74**(1):4-17.

Dias 2010

Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Statistics in Medicine* 2010;**29**(7-8):932-44.

Edwards 2018

Edwards SJ, Wakefield V, Cain P, Karner C, Kew K, Bacelar M, et al. Axitinib, cabozantinib, everolimus, nivolumab, sunitinib and best supportive care in previously treated renal cell carcinoma: a systematic review and economic evaluation. *Health Technology Assessment* 2018;**22**(6):1-278.

Egger 1997

Egger M, Smith DG, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;**315**(7109):629.

Elliott 2014

Elliott JH, Turner T, Clavisi O, Thomas J, Higgins JP, Mavergames C, et al. Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. *PLOS Medicine* 2014;**11**(2):e1001603.

Escudier 2019

Escudier B, Porta C, Schmidinger M, Rioux-Leclerq N, Bex A, Khoo V, et al. Renal cell carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology* 2019;**30**(5):706-20.

Furukawa 2006

Furukawa TA, Barbui C, Cipriani A, Brambilla P, Watanabe N. Imputing missing standard deviations in meta-analyses can provide accurate results. *Journal of Clinical Epidemiology* 2006;**59**(1):7-10.

Garner 2016

Garner P, Hopewell S, Chandler J, MacLehose H, Schunemann HJ, Akl EA, et al. When and how to update systematic reviews: consensus and checklist. *BMJ* 2016;**354**:i3507.

GRADEpro GDT [Computer program]

GRADEpro GDT. Version accessed 9 22 October 2020. Hamilton (ON): McMaster University (developed by Evidence Prime). gradepro.org.

Guyatt 2017

Guyatt GH, Ebrahim S, Alonso-Coello P, Johnston BC, Mathioudakis AG, Briel M, et al. GRADE guidelines 17: Assessing the risk of bias associated with missing participant outcome data in a body of evidence. *Journal of Clinical Epidemiology* 2017;**87**:14-22.

Harshman 2014

Harshman LC, Drake CG, Choueiri TK. PD-1 blockade in renal cell carcinoma: to equilibrium and beyond. *Cancer Immunology Research* 2014;**2**(12):1132.

Higgins 2019

Higgins JP, Savović J, Page MJ, Elbers RG, Sterne JA. Chapter 8: assessing risk of bias in a randomized trial. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editor(s). Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). Cochrane, 2019. training.cochrane.org/handbook 2019.

Higgins 2019a

Higgins JP, Eldridge S, Li T. Chapter 23: including variants on randomized trials. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editor(s). Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). Cochrane, 2019. training.cochrane.org/handbook 2019.

Higgins 2019b

Higgins JP, Deeks JJ. Chapter 6: choosing effect measures and computing estimates of effect. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editor(s). Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). Cochrane, 2019. training.cochrane.org/handbook 2019.

Higgins 2019c

Higgins JP, Deeks JJ, Altman DG, on behalf of the Cochrane Statistical Methods Group. Chapter 16: special topics in statistics. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editor(s). Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). Cochrane, 2019. training.cochrane.org/handbook 2019.

Keir 2007

Keir ME, Francisco LM, Sharpe AH. PD-1 and its ligands in T-cell immunity. *Annual Review of Immunology* 2007;**19**(3):309-14.

Krahn 2013

Krahn U, Binder H, König J. A graphical tool for locating inconsistency in network meta-analyses. *BMC Medical Research Methodology* 2013;**13**(1):35.

Lefebvre 2019

Lefebvre C, Glanville J, Briscoe S, Littlewood A, Marshall C, Metzendorf MI, et al. Chapter 4: searching for and selecting studies. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editor(s). Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). Cochrane, 2019. training.cochrane.org/handbook 2019.

Leitlinienprogramm Onkologie

Leitlinienprogramm Onkologie (Deutsche Krebsgesellschaft, Deutsche Krebshilfe, AWMF). Diagnostics, therapy and aftercare of renal cell carcinoma - Longversion 3.0 [Diagnostik, Therapie und Nachsorge des Nierenzellkarzinoms - Langversion 3.0]. https://www.leitlinienprogramm-onkologie.de/fileadmin/user_upload/Downloads/Leitlinien/Nierenzellkarzinom/Version_3/LL_Nierenzellkarzinom_Langversion_3.0.pdf.

Li 2019

Li T, Higgins JP, Deeks JJ. Chapter 5: collecting data. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editor(s). Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). Cochrane, 2019. training.cochrane.org/handbook 2019.

Liu 2021

Liu Z, Chen Y, Wei Z, He Y, Wang J, Mu X, et al. Comparative efficacy and safety of immunotherapy in the first-line treatment of metastatic renal cell carcinoma: a systematic review and network meta-analysis. *Annals of Palliative Medicine* 2021;**10**(3):2805-14.

Ljungberg 2022

Ljungberg B, Albiges L, Abu-Ghanem Y, Bedke J, Capitanio U, Dabestani S, et al. European Association of Urology guidelines on renal cell carcinoma: the 2022 update. *European Urology* 2022;**82**(4):399–410.

Lopez-Beltran 2009

Lopez-Beltran A, Carrasco JC, Cheng L, Scarpelli M, Kirkali Z, Montironi R. 2009 update on the classification of renal epithelial tumors in adults. *International Journal of Urology* 2009;**16**(5):432-43.

Moher 2009

Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Journal of Clinical Epidemiology* 2009;**62**(10):1006-12.

Mori 2021

Mori K, Mostafaei H, Miura N, Karakiewicz PI, Luzzago S, Schmidinger M, et al. Systemic therapy for metastatic renal cell carcinoma in the first-line setting: a systematic review and network meta-analysis. *Cancer Immunology, Immunotherapy* 2021;**70**(2):265-73.

Motzer 2022

Motzer RJ, Jonasch E, Agarwal N, Alva A, Baine M, Beckermann K, et al. Kidney cancer, version 3.2022 - NCCN clinical practice guidelines in oncology. *Journal of the National Comprehensive Cancer Network* 2022;**20**(1):71-90.

Nassar 2022

Nassar AH, Adib E, Abou Alaiwi S, El Zarif T, Groha S, Akl, EW, et al. Ancestry-driven recalibration of tumor mutational burden and disparate clinical outcomes in response to immune checkpoint inhibitors. *Cancer Cell* 2022;**40**(10):1161-72.

Nocera 2022

Nocera L, Karakiewicz P I, Wenzel M, Tian Z, Shariat S F, Saad F, et al. Clinical outcomes and adverse events after first-line.treatment in metastatic renal cell carcinoma: a systematic review and network meta-analysis. *Journal of Urology* 2022;**207(1)**:16-24.

Parmar 1998

Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Statistics in Medicine* 1998;**17**(24):2815-34.

Powles 2021

Powles T, Albiges L, Bex A, Grünwald V, Porta C, Procopio G, et al. ESMO Clinical Practice Guideline update on the use of immunotherapy in early stage and advanced renal cell carcinoma. *Annals of Oncology* 2021;**32**(12):1511-9.

Puhan 2014

Puhan MA, Schunemann HJ, Murad MH, Li T, Brignardello-Petersen R, Singh JA, et al. A GRADE working group approach for rating the quality of treatment effect estimates from network meta-analysis. *BMJ* 2014;**349**:g5630. [DOI: 10.1136/bmj.g5630]

Qin 2018

Qin F, Yu H, Xu CR, Chen HH, Bai JL. Safety of axitinib and sorafenib monotherapy for patients with renal cell carcinoma: a meta-analysis. *Journal of Biomedical Research* 2018;**32**(1):30-8.

Quhal 2021

Quhal F, Mori K, Bruchbacher A, Resch I, Mostafaei H, Pradere B, et al. First-line immunotherapy-based combinations for metastatic renal cell carcinoma: a systematic review and network meta-analysis. *European Urology Oncology* 2021;**4**(5):755-65.

R Core Team 2019 [Computer program]

R: A language and environment for statistical computing. R Core Team 2019. Vienna, Austria: R Foundation for Statistical Computing, 2019. Available at www.R-project.org.

Riaz 2021

Riaz IB, He H, Ryu AJ, Siddiqi R, Naqvi SA, Yao Y, et al. A living, interactive systematic review and network meta-analysis of first-line treatment of metastatic renal cell carcinoma. *European Urology* 2021;**80**(6):712-23.

Rini 2009

Rini BI, Campbell SC, Escudier B. Renal cell carcinoma. *Lancet* 2009;**373**(9669):1119-32.

Robert Koch Institute 2021

Robert Koch Institute. Cancer in Germany in 2017/2018 [Krebs in Deutschland für 2017/2018]. https://www.krebsdaten.de/Krebs/EN/Content/Publications/Cancer_in_Germany/cancer_chapters_2017_2018/cancer_germany_2017_2018.pdf?__blob=publicationFile 2021.

Rücker 2012

Rücker G. Network meta-analysis, electrical networks and graph theory. *Research Synthesis Methods* 2012;**3**(4):312-24.

Rücker 2014

Rücker G, Schwarzer G. Reduce dimension or reduce weights? Comparing two approaches to multi-arm studies in network meta-analysis. *Statistics in Medicine* 2014;**33**(25):4353-69.

Rücker 2015

Rücker G, Schwarzer G. Ranking treatments in frequentist network meta-analysis works without resampling methods. *BMC Medical Research Methodology* 2015;**15**(1):58.

Rücker 2019

Rücker G, Krahn U, König J, Efthimiou O, Schwarzer G. netmeta: Network meta-analysis using frequentist methods. R package version 1.10. CRAN.R-project.org/package=netmeta 2019.

Sabatini 2006

Sabatini DM. mTOR and cancer: insights into a complex relationship. *Nature Reviews Cancer* 2006;**6**(9):729-34.

Salanti 2012

Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Research Synthesis Methods* 2012;**3**(2):80-97.

Salanti 2014

Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JP. Evaluating the quality of evidence from a network meta-analysis. *PLOS One* 2014;**9**(7):e99682.

Sanchez-Gastaldo 2017

Sanchez-Gastaldo A, Kempf E, Gonzalez del Alba A, Duran I. Systemic treatment of renal cell cancer: a comprehensive review. *Cancer Treatment Reviews* 2017;**60**:77-89.

Scelo 2018

Scelo G, Larose TL. Epidemiology and risk factors for kidney cancer. *Journal of Clinical Oncology* 2018;**36**(36):3574-81.

Schwarzer 2007

Schwarzer G. meta: an R package for meta-analysis. *R News* 2007;**7**(3):40-5.

Schwarzer 2015

Schwarzer G, Carpenter JR, Rücker G. Chapter 8: network meta-analysis. In: Meta-Analysis With R. Springer International Publishing Switzerland, 2015.

Schünemann 2019

Schünemann HJ, Vist GE, Higgins JP, Santesso N, Deeks JJ, Glasziou P, et al. Chapter 15: interpreting results and drawing conclusions. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editor(s). Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). Cochrane, 2019. training.cochrane.org/handbook 2019.

Siegel 2022

Siegel RL, Miller KD, Fuchs H, Jemal A. Cancer statistics 2022. *CA: A Cancer Journal for Clinicians* 2022;**72**(1):7-33.

Skoetz 2019

Skoetz N, Goldkuhle M, Weigl A, Dwan K, Labonte V, Dahm P, et al. Methodological review showed correct absolute effect size estimates for time-to-event outcomes in less than one-third of cancer-related systematic reviews. *Journal of Clinical Epidemiology* 2019;**108**:1-9.

Skoetz 2020

Skoetz N, Goldkuhle M, van Dalen EC, Akl EA, Trivella M, Mustafa R, et al. GRADE guidelines 27: how to calculate absolute effects for time-to-event outcomes in summary of findings tables and evidence profiles. *Journal of Clinical Epidemiology* 2020;**118**:124-31.

Sterne 2019

Sterne JA, Savovic J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;**366**:l4898.

Sun 2011

Sun M, Thuret R, Abdollah F, Lughezzani G, Schmitges J, Tian Z, et al. Age-adjusted incidence, mortality, and survival rates of stage-specific renal cell carcinoma in North America: a trend analysis. *European Urology* 2011;**59**(1):135-41.

Swallow 2018

Swallow E, Messali A, Ghate S, McDonald E, Duchesneau E, Perez JR. The additional costs per month of progression-free survival and overall survival: an economic model comparing everolimus with cabozantinib, nivolumab, and axitinib for second-line treatment of metastatic renal cell carcinoma. *Journal of Managed Care & Specialty Pharmacy* 2018;**24**(4):335-43.

Thekdi 2015

Thekdi SM, Milbury K, Spelman A, Wei Q, Wood C, Matin SF, et al. Posttraumatic stress and depressive symptoms in renal cell carcinoma: association with quality of life and utility of single-item distress screening. *Psycho-Oncology* 2015;**24**(11):1477-84.

Tierney 2007

Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials* 2007;**8**:16.

Tudur 2001

Tudur C, Williamson PR, Khan S, Best LY. The value of the aggregate data approach in meta-analysis with time-to-event outcomes. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2001;**164**(2):357-70.

UICC 2017

Union for International Cancer Control (UICC). TNM Classification of Malignant Tumours. 8th edition. Wiley-Blackwell, 2017.

United States Congress 2007

United States Congress 2007. Food and Drug Administration Amendments Act (FDAAA) of 2007: public law no 110-85. govinfo.gov/content/pkg/PLAW-110publ85/pdf/PLAW-110publ85.pdf.

Unverzagt 2017

Unverzagt S, Moldenhauer I, Nothacker M, Rossmeissl D, Hadjinicolaou AV, Peinemann F, et al. Immunotherapy for metastatic renal cell carcinoma. *Cochrane Database of Systematic Reviews* 2017, Issue 5. Art. No: CD011673. [DOI: 10.1002/14651858.CD011673]

Wang 2019

Wang S, Cowley LA, Liu X-S. Sex differences in cancer immunotherapy efficacy, biomarkers, and therapeutic strategy. *Molecules* 2019;**24**(18):3214.

Warren 2018

Warren AY, Harrison D. WHO/ISUP classification, grading and pathological staging of renal cell carcinoma: standards and controversies. *World Journal of Urology* 2018;**36**(12):1913-26.

White 2012

White IR, Barrett JK, Jackson D, Higgins JP. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Research Synthesis Methods* 2012;**3**(2):111-25.

World Medical Association

World Medical Association. World Medical Association Declaration of Helsinki - ethical principles for medical research involving human subjects. *Journal of the American Medical Association* 2013;**310**(20):2191-4.

References to other published versions of this review

Goldkuhle 2020

Goldkuhle M, Aldin A, Jakob T, Adams A, Monsef I, Heidenreich A, et al. First-line therapy for adults with advanced renal cell carcinoma: a systematic review and network meta-analysis. *Cochrane Database of Systematic Reviews* 2020, Issue 12. Art. No: CD013798. [DOI: 10.1002/14651858.CD013798]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Jonasch 2010

Study characteristics	
Methods	**Study name:** - (NCT also not available) **Study design:** randomised, phase II trial **Blinding**: no information **Study dates:** June 24, 2005 - June 18, 2007 (date of enrolment) **Date of data cut-off:** not reported **Location:**USA.; type of centre: cancer centre (1 study location) **Cross-over study or cross over permitted:** not per design; not reported whether cross over was permitted at some point (e.g. upon progression)

Participants	**Inclusion criteria:**
	
	• pathologically confirmed metastatic clear cell RCC
	• no prior systemic therapy
	• ECOG PS of 0 or 1
	• no brain metastases
	• measurable disease according to RECIST
	
	Sample size: N = 80
	
	Age, median in years (range): experimental arm: 60.7 (43-81), control arm: 62.4 (45-83)
	
	Sex (m/f): experimental arm: 29/11, control arm: 32/8
	
	Prognostic factors:
	
	• **ECOG status, n (%)**
	  ○ **0**
	  ■ experimental arm: 25 (62.5), control arm: 25 (62.5)
	  ○ **1**
	  ■ experimental arm: 15 (37.5), control arm: 15 (37.5)
	• **MSKCC prognostic risk, n (%)**
	  ○ **Low**
	  ■ experimental arm: 20 (50), control arm: 21 (52.5),
	  ○ **Intermediate**
	  ■ experimental arm: 18 (45), control arm: 19 (47.5),
	  ○ **Poor**
	  ■ experimental arm: 2 (5), control arm: 0
	• **Previous nephrectomy (N,%)**
	  ○ **Yes**
	  ■ experimental arm: 40 (100), control arm: 39 (98)

Jonasch 2010 *(Continued)*

Interventions	**Experimental arm (n = 40):** Sorafenib 400 mg (oral, twice/day) + Interferon alfa (0.5 MIU, subcutaneous injection, twice/day) **Control arm (n = 40):** Sorafenib, 400mg (oral, twice/day)
Outcomes	**Primary outcome(s)** • Safety (report "toxicities") **Secondary outcome(s)** • Progression-free survival (PFS) • Overall survival (OS) **Outcomes relevant to this review but not reported:** QoL; TFST; number of participants who discontinued treatment due to an AE **Other outcomes (not relevant to this review):** ORR
Notes	**Funding sources:** National Cancer Institute's Cancer Therapy Evaluation Program **Conflict of interest disclosures:** "Supported by the National Cancer Institute's Cancer Therapy Evaluation Program." **Clinical study report available:** no **Study protocol available:** no **Statistical analysis plan available:** no

NCT00065468

Study characteristics	
Methods	**Study name:** - **Study design:** randomised, phase III (three-arm trial) **Blinding**: none, open-label **Study dates:** July 2003 – April 2005 (date of enrolment) **Date of data cut-off:** exact date not reported. The results presented and used in this review are of the second interim analysis. **Location:** 25 countries (Argentina, Australia, Canada, Czech Republic, Former Serbia and Montenegro Germany, Greece, Hungary, Italy, Latvia, Lithuania, Mexico, the Netherlands, Poland, Russian Federation, Serbia, Slovakia, South Africa, Spain, Sweden, Taiwan, Turkey, Ukraine, UK, USA; types of centres: cancer centres, hospitals, university hospitals (153 study locations) **Cross-over study or cross-over permitted:** no
Participants	**Inclusion criteria:** • all sexes • 18 years and older • histologically confirmed, advanced RCC • no prior systemic therapy **Exclusion criteria:**

NCT00065468 *(Continued)*

- participants with central nervous system (CNS) metastases
- prior anticancer therapy for RCC
- prior investigational therapy/agents within 4 weeks of randomisation

Sample size: N = 626

Age, median in years (range): experimental arm I: 60 (23-86), experimental arm II arm: 58 (32-81), control arm: 59 (32-82)

Sex (m/f): experimental arm I: 148/59, experimental arm II: 139/70, control arm: 145/65

Prognostic factors:

- **MSKCC risk classification, n (%)**
  - **Poor risk (≥3 of 5 factors)**
    - experimental arm I: 157 (76), experimental arm II: 145 (69), control arm: 160 (76)
  - **Intermediate risk(1 or 2 or 5 factors)**
    - experimental arm I: 50 (24), experimental arm II: 64 (31), control arm: 50 (24)
- **Previous nephrectomy (N,%)**
  - **Yes**
    - experimental armI: 193 (67), experimental arm II: 168 (80), control arm: 141 (67)

Interventions

Experimental arm I (n = 207): Interferon alfa, 3 MIU (1st week), 9 MIU (2nd week), 18 MIU (thereafter) (subcutaneous injection, three times/week)

Experimental arm II (n = 209): Temsirolimus (25 mg, intravenous, once/week)

Control arm (n = 210): Interferon alfa 3 MIU (1st week), 6 MIU (thereafter) (subcutaneous injection, three times/week) + Temsirolimus (15 mg, intravenous, once/week)

Outcomes

Primary outcome(s)

- OS
  - Time frame: baseline up to month 80

Secondary outcome(s)

- PFS
  - Time frame: at baseline, monthly until tumour progression or death (up to month 80)
- Quality of life, measured with the European Quality of Life Health Questionnaire (EQ-5D) - Index Score
  - time frame: measured at baseline
- Safety (AEs, SAEs)

Relevant to this review but not reported: TFST; number of participants who discontinued treatment due to an AE

Other outcomes (not relevant to this review): objective response (OR); participants with clinical benefit; duration of response (DR); time to treatment failure (TTF); quality-adjusted time without symptoms or toxicity (Q-TWIST)

Notes

Funding sources: Pfizer

Declarations of Interests: Quote: "Dr. Hudes reports receiving consulting and lecture fees from Pfizer Pharmaceuticals and consulting fees from Wyeth Pharmaceuticals; Drs. Carducci and Motzer, consulting fees from Wyeth Pharmaceuticals; Dr. Dutcher, consulting and lecture fees from Novartis, Chiron, Bayer, and Onyx Pharmaceuticals, consulting fees from Wyeth Pharmaceuticals, lecture fees from Pfizer Pharmaceuticals, and research grants from Bayer, Chiron, Genentech, Pfizer, and Wyeth Pharmaceuticals; Dr. Figlin, consulting and lecture fees and research grants from Wyeth Pharmaceuticals; Dr. Kapoor, consulting fees and research grants from Wyeth Pharmaceuticals and research grants from Bayer Pharmaceuticals; Dr. McDermott, consulting fees from Bayer, Onyx, and Genentech Pharmaceuticals and lecture fees from Novartis Pharmaceuticals; and Dr. Schmidt-Wolf, symposium support fees from Wyeth Pharmaceuticals. Mr. O'Toole, Ms. Lustgarten, and Dr. Moore report being full-time employ-

NCT00065468 *(Continued)*

ees of Wyeth Pharmaceuticals. No other potential conflict of interest relevant to this article was reported."

Clinical study report available: no

Study protocol available: no

Statistical analysis plan available: no

NCT00072046

Study characteristics

Methods

Study name: CALGB 90206

Study design: randomised, phase III

Blinding: none, open-label

Study dates: October 2003 – November 2012 (date of enrolment)

Date of data cut-off: March 24, 2009 (for OS), not reported for PFS

Location: 2 countries (Canada, USA.), types of centres: cancer centres, medical centres, hospitals (493 study locations)

Cross-over study or cross over permitted: no

Participants

Inclusion criteria:

- 18 years to 120 years
- all sexes
- histologically or cytologically confirmed renal cell carcinoma (RCC)
  - conventional clear cell carcinoma
  - metastatic or unresectable disease
- Karnofsky 70% to 100%
- not pregnant/nursing
- no pre-existing thyroid abnormality in which normal thyroid function cannot be maintained by medication
- no delayed wound healing, ulcers, or bone fractures
- no uncontrolled psychiatric disorder

Exclusion criteria:

- true papillary cellular type
- sarcomatoid features without a clear cell component
- chromophobe
- oncocytoma
- collecting duct tumour
- transitional cell carcinoma

Sample size: N = 732

Age (years, median with range): experimental arm: 61 (56 to 70), control arm: 62 (55 to 70)

Sex (m/f): experimental arm: 269/100, control arm: 239/124

Prognostic factors:

NCT00072046 *(Continued)*

- **ECOG Performance**
  - **0**
    - experimental arm: 230 (62), control arm: 227 (62)
  - **1**
    - experimental arm: 132 (36), control arm: 133 (37)
  - **2**
    - experimental arm: 7 (2), control arm: 3 (1)
- **Previous nephrectomy (n,%)**
  - **Yes**
    - experimental arm: 312 (85), control arm: 308 (85)

Interventions	**Experimental arm (n = 369):** Bevacizumab (10 mg/kg, intravenous), Interferon alfa (9 MIU, subcutaneous injection) **Control arm (n = 363):** Interferon alfa (9 MIU, subcutaneous injection)

Outcomes	**Primary outcome(s)**
	
	• OS
	  ◦ Time frame: 5 years
	
	Secondary outcome(s)
	
	• Time to progression (unclear whether definition of PFS will be used)
	  ◦ Time frame: 3 cycles
	• Toxicity (AEs)
	  ◦ Time frame: unclear
	
	Relevant to this review but not reported: QoL, SAE, TFST, number of participants who discontinued treatment due to an AE
	
	Other outcomes (not relevant to this review): objective response rate (ORR)

Notes	**Funding sources:** Walter M. Stadler, Genentech; Daniel A. Vaena, Genentech; Janice Dutcher, Novartis, Genentech, Pfizer, sponsor: Alliance for Clinical Trials in Oncology, collaborators: National Cancer Institute (NCI) & NCIC Clinical Trials Group **Declaration of Interest:** Quote: "Although all authors completed the disclosure declaration, the following author(s) indicated a financial or other interest that is relevant to the subject matter under consideration in this article. (...) For a detailed description of the disclosure categories, or for more information about ASCO's conflict of interest policy, please refer to the Author Disclosure Declaration and the Disclosures of Potential Conflicts of Interest section in Information for Contributors." **Clinical study report available:** no **Study protocol available:** no **Statistical analysis plan available:** no

NCT00081614

Study characteristics	
Methods	**Study name:** - **Study design:** randomised, parallel, placebo-controlled phase II trial **Blinding**: double-blind (investigator and participants) **Study dates:** March 2004 - July 2005

NCT00081614 *(Continued)*

Date of data cut-off: not reported

Location: 1 country (USA), types of centres: cancer centres, medical centres, hospitals/clinics, university hospitals (20 study locations)

Cross-over study or cross over permitted: not a cross-over study; no information whether cross over was permitted

Participants	**Inclusion criteria:**
	• 18 years or older
	• all sexes
	• mRCC with predominant (>50%) clear-cell histology
	• prior nephrectomy
	• to have measurable disease
	• ECOG PS 0 or 1
	• previous radiotherapy (exception single-fraction radiotherapy for pain control)
	Exclusion criteria:
	• prior systemic therapy either in the adjuvant setting or for metastatic disease
	• previous use of angiogenesis
	• previous use of EGFR inhibitors
	• currently receiving dialysis
	• undergoing a major surgical procedure within 28 days of initiating study treatment
	Sample size: N = 104
	Age (median (years, range)): experimental arm: 66 (38-86), control arm: 61 (35-78)
	Sex (m/f): experimental arm: male: 33/18, control arm: 40/13
	Prognostic factors:
	• **ECOG PS (n (%))**
	  ◦ **0**
	  ▪ experimental arm: 31 (61); control arm: 34 (64)
	  ◦ **1**
	  ▪ experimental arm: 20 (39); control arm: 19 (36)
	• **MSKCC risk category (n, %)**
	  ◦ **Low**
	  ▪ experimental arm: 16 (31); control arm: 19 (36)
	  ◦ **Intermediate**
	  ▪ experimental arm: 35 (69); control arm: 34 (64)
	• **Previous nephrectomy (n,%)**
	  ◦ **Yes**(all participants)

Interventions	**Experimental arm (n = 51):** bevacizumab (10 mg/kg, intravenous) + erlotinib (150 mg, oral, daily) **Control arm (n = 53):** bevacizumab (10 mg/kg, intravenous) + Placebo (150 mg, oral, daily)
Outcomes	**Primary outcome(s)** • Progression-free survival (PFS), 9 months after enrolment of the last participant • Objective response rate (ORR), 9 months after enrolment of the last participant **Secondary outcome(s)** • Overall survival (OS) • Safety (AE)

NCT00081614 *(Continued)*

	Relevant to this review but not reported: QoL, SAE, TFST, number of participants who discontinued treatment **Other outcomes (not relevant to this review):** ORR
Notes	**Funding sources:** Ronald M. Bukowski, Bayer/Onyx, Genentech, Wyeth, Pfizer, Amgen; Robert A. Figlin, Genentech; Janice P. Dutcher, Bayer, Pfizer, Wyeth, Chiron/Novartis, Idera, Genentech; David F. McDermott, Genentech, Pfizer, Bayer/Onyx **Declarations of Interests:** Quote;"Although all authors completed the disclosure declaration, the following authors or their immediate family members indicated a financial interest. No conflict exists for drugs or devices used in a study if they are not being evaluated as part of the investigation. For a detailed description of the disclosure categories, or for more information about ASCO's conflict of interest policy, please refer to the Author Disclosure Declaration and the Disclosures of Potential Conflicts of Interest section in Information for Contributors." **Clinical study report available**: no **Study protocol available:** no **Statistical analysis plan available:** no

NCT00098657/NCT00083889

Study characteristics	
Methods	**Study name:** - **Study design:** randomised, phase III trial **Blinding:** none, open-label **Study dates:** August 2004 – October 2005 (date of randomisation) **Date of data cut-off:** November 2005 **Location:** 11 countries (Australia, Brazil, Canada, France, Germany, Italy, Poland, Russian Federation, Spain, U.K., USA), types of centres: cancer centres, medical centres, university hospitals (124 study locations) **Cross-over study or cross over permitted:** not a cross-over study per design, but cross over to sunitinib was permitted in case of disease progression
Participants	**Inclusion criteria:** • histologically confirmed renal cell carcinoma of clear cell histology with metastases • evidence of measurable disease by radiographic technique • eastern Cooperative Oncology Group (ECOG) performance status of 0 or 1 **Exclusion criteria:** • prior systemic (including adjuvant or neoadjuvant) therapy of any kind for RCC • history of or known brain metastases • serious acute or chronic illness or recent history of significant cardiac abnormality **Sample Size:** N = 750 **Age (years, median with range):** experimental arm: 62 (27-87), control arm: 59 (34-85) **Sex (M/F):** experimental arm: 267/108, control arm: 269/106

NCT00098657/NCT00083889 *(Continued)*

Prognostic factors:

- **ECOG PS no. (%)**
  - **0**
    - experimental arm: 231 (62), control arm: 229 (61)
  - **1**
    - experimental arm: 144 (38), control arm: 146 (39)
- **MSKCC risk factors no. (%)**
  - **0(favourable)**
    - experimental arm: 143 (38), control arm: 121 (34)
  - **1-2 (intermediate)**
    - experimental arm: 209 (56), control arm: 212 (59)
  - **≥3 (poor)**
    - experimental arm: 23 (6), control arm: 25 (7)
- **Previous nephrectomy (n,%)**
  - **Yes**
    - experimental arm: 390 (91), control arm: 335 (89)

Interventions	**Experimental arm (n = 375):** Sunitinib (50mg, oral, once/day) **Control arm (n = 375):** Interferon alfa (3 MIU (1st week), 6 MIU (2nd week), and 9 MIU (thereafter), subcutaneous injection, thrice/week)

Outcomes	**Primary outcome(s)**
	
	• PFS, core radiology assessment
	  ○ Time frame: day 28 of each 6-week cycle: duration of treatment phase
	• PFS, investigator's assessment
	  ○ Time frame: day 28 of each 6-week cycle: duration of treatment phase
	
	Secondary outcome(s)
	
	• OS
	  ○ Time frame: clinic visit or telephone contact every 2 months until death
	• number of participants who discontinued treatment due to an AE
	• AE & SAE
	• QoL
	
	Relevant to this review but not reported: TFST
	
	Other outcomes (not relevant to this review): objective response (OR), time to tumour progression (TTP), duration of response (DR), laboratory results, pharmacokinetics

Notes	**Funding sources:** Pfizer **Declarations of Interests: Quote:** "Dr. Motzer reports receiving research grants from Pfizer and Genentech, consulting fees from Wyeth, and lecture fees from Bayer Pharmaceuticals; Dr. Hutson, consulting and lecture fees from Pfizer, Bayer Pharmaceuticals, and Onyx Pharmaceuticals; Dr. Michaelson, consulting fees from Pfizer and Wyeth Pharma- ceuticals and lecture fees from Pfizer; Dr. Bukowski, research grants from Pfizer, Bayer Pharmaceuticals, Genentech, Genzyme, and Bristol-Myers Squibb and consulting and lecture fees from Pfizer, Bayer Pharmaceuticals, Onyx Pharmaceuticals, and Genentech; Dr. Rixe, consulting and lecture fees from Pfizer; Dr. Oudard, consulting and lecture fees from Pfizer; Dr. Negrier, consulting fees from Pfizer and Bayer Pharmaceuticals; and Dr. Figlin, research grants from Pfizer, consulting fees from Pfizer and Onyx Pharmaceuticals, and lecture fees from Pfizer and Bayer Pharmaceuticals. Ms. Kim and Drs. Chen, Bycott, and Baum report being full-time employees of Pfizer and having equity ownership in the company. No other potential conflict of interest relevant to this article was reported." **Clinical study report available:** no **Study protocol available:** no

NCT00098657/NCT00083889 *(Continued)*

Statistical analysis plan available: no

NCT00117637

Study characteristics

Methods

Study name: -

Study design: randomised, phase II trial

Blinding: none, open-label

Study dates: June 28, 2005 - September 30, 2005 (date of randomisation)

Date of data cut- off: not reported

Location: 7 countries* (France, Germany, Poland, Russian Federation, Ukraine, UK, USA), types of centres: not reported

Cross-over study or cross over permitted: yes, cross over trial**

*discrepancies between information provided in the publication and information provided on ct.gov; we included information from ct.gov.

**For cross-over trials, we only extracted data from the first period.

Participants

Inclusion criteria:

- ECOG PS ≤ 1
- age ≥ 18 years
- life expectancy ≥ 12 weeks
- complete surgical excision of primary RCC at initial diagnosis
- adequate bone marrow, liver, and renal function assessed 7 days before screening

Exclusion criteria:

- previous malignancy
- distinct in primary site/histology from that evaluated in this study
- complete renal failure that required dialysis
- symptomatic metastatic brain or meningeal tumours

Sample size: N = 189

Age (median in years (range)): experimental arm: 62 (34-78), control arm: 62.5 (18-80)

Sex (m/f): experimental arm: 65/32, control arm: 52/40

Prognostic factors:

- **ECOG PS, n (%)**
  - **0**
    - experimental arm: 56 (57.7), control arm: 49 (53.3)
  - **1**
    - experimental arm: 41 (42.3), control arm: 43 (46.7)
- **MSKCC score, n (%)**
  - **Low**
    - experimental arm: 52 (53.6), control arm: 47 (51.1)
  - **Intermediate**
    - experimental arm: 44 (45.4), control arm: 44 (47.8)

NCT00117637 *(Continued)*

- **High**
  - experimental arm: 1 (1.0), control arm: 0 (0.0)
- **Missing**
  - experimental arm: 0 (0.0), control arm: 1 (1.1)

- **Previous nephrectomy (n,%)**
  - **Yes**
    - experimental arm: 95 (97.9), control arm: 83 (90.2)

Interventions	**Experimental arm (n = 97):** Sorafenib (400mg, oral, twice/day) **Control arm (n = 92):** Interferona alfa (9 MIU, subcutaneous injection, thrice/week)
Outcomes	**Primary outcome(s)** • PFS **Secondary outcome(s)** • Safety (AEs, SAEs) • QoL • number of participants who discontinued treatment due to an AE **Relevant to this review but not reported:** OS, TFST **Other outcomes (not relevant to this review):** response duration, OR, DCR, CR
Notes	**Funding sources:** Thomas E. Hutson, Bayer/Onyx, Pfizer Inc, Wyeth; MichaelStaehler, Bayer Healthcare, Pfizer Inc, Roche, Novartis, Wyeth; DavidCella, Bayer Healthcare; Ronald Bukowski, Bayer Healthcare, Wyeth,Novartis **Declarations of interests:** quote: "Although all authors completed the disclosure declaration, the following author(s) indicated a financial or other interest that is relevant to the subject matter under consideration in this article. Certain relationships marked with a “U” are those for which no compensation was received; those relationships marked with a “C” were compensated. For a detailed description of the disclosure categories, or for more information about ASCO’s conflict of interest policy, please refer to the Author Disclosure Declaration and the Disclosures of Potential Conflicts of Interest section in Information for Contributors." **Clinical study report available:** no **Study protocol available:** no **Statistical analysis plan available:** no

NCT00126594

Study characteristics	
Methods	**Study name:** - **Study design:** randomised, parallel assignment, phase II (three-arm trial) **Blinding:** none, open-label **Study dates:** June, 2005 - August, 2013 **Date of data cut-off:** not reported **Location:** 1 country (USA), type of centre: cancer centre (1 study location)

NCT00126594 *(Continued)*

Cross-over study or cross over permitted: not per design; not reported whether cross over was permitted at some point (e.g. upon progression)

Participants	**Inclusion criteria:**
	• participants with histologically or cytologically confirmed metastatic clear cell RCC
	• 18 years and older
	• participants must have measurable disease, defined as a lesion that can be accurately measured in at least one dimension (longest diameter to be recorded) and measures >= 20 mm with conventional techniques or >= 10 mm with spiral CT scan
	• ECOG performance status =< 1
	Exclusion criteria:
	• no prior malignancy is allowed, except for non-melanoma skin cancer, in situ carcinoma of any site, or other cancers for which the patient has been adequately treated and disease free for 5 years
	• participants must not have received any systemic anticancer therapy for renal cell carcinoma; participants must not have received any radiotherapy for renal cell carcinoma within 4 weeks prior to entering the study or those who have not recovered from adverse events due to agents administered more than 4 weeks earlier
	• participants must not be scheduled to receive another experimental drug while on this study; participants are permitted to be on concomitant bisphosphonates
	• participants must not have a primary brain tumour, any brain metastases, leptomeningeal disease, seizure disorders not controlled with standard medical therapy, or history of stroke
	More inclusion and exclusion criteria on CT.gov.
	Sample size: N = 80
	Age (median in years (range)): experimental arm: 60.7 (43-81), control arm: 62.4 (45-83)
	Sex (m/f): experimental arm: 29/11, control arm: 32/8
	Prognostic factors:
	• **ECOG PS, n**
	  ◦ **0**
	  ▪ experimental arm: 25, control arm: 25
	  ◦ **1**
	  ▪ experimental arm: 15, control arm: 15
	• **MSKCC score, n**
	  ◦ **Low**
	  ▪ experimental arm: 20, control arm: 21
	  ◦ **Intermediate**
	  ▪ experimental arm: 18, control arm: 19
	  ◦ **Poor**
	  ▪ experimental arm: 2, control arm: 0
	• **Previous nephrectomy, n**
	  ◦ **Yes**
	  ▪ experimental arm: 39, control arm: 40

Interventions	**Experimental arm:** sorafenib (400 mg, oral, twice/day) **Control arm:** sorafenib (400 mg, oral, twice/day) and recombinant interferon alfa-2b (0.5 MIU, subcutaneous injection, twice/day)
Outcomes	**Primary outcome(s)** - **Secondary outcome(s)**

NCT00126594 *(Continued)*

- Selected grade 3-4 AEs
  - Time frame: up to 12 months of treatment
- PFS
  - Time frame: from date of randomisation until the date of first documented progression or date of death from any cause, whichever came first, assessed up to 36 months
- Median OS
  - Time frame: from the start of protocol therapy to death or date of last follow-up, up to 36 months

Relevant to this review but not reported: QoL, TFST, number of participants who discontinued treatment due to an AE

Other outcomes (not relevant to this review): ORR, DoR

Notes

Funding sources: National Cancer Institute (NCI), M.D. Anderson Cancer Center

Declarations of Interests: not found

Clinical study report available: no

Study protocol available: no

Statistical analysis plan available: no

NCT00334282

Study characteristics

Methods

Study name: -

Study design: randomised, placebo-controlled trial, phase III

Blinding: quadruple (participant, care provider, investigator, outcomes assessor)

Study dates: April 2006 – April 2007 (date of enrolment)

Date of data cut-off: March 15, 2010 (final analysis of OS and updated safety data), May 23, 2008 (final PFS analysis)

Location: 25 countries (Argentina, Australia, Austria, Brazil, Chile, China, Czech Republic, Estonia, Greece, Hong Kong, India, Ireland, Italy, Republic of Korea, Latvia, Lithuania, Mexico, New Zealand, Pakistan, Poland, Russian Federation, Slovakia, Tunisia, Ukraine, UK.), types of centres: (100 study locations)

Cross-over study or cross over permitted: not per design, but cross over was permitted from placebo to pazopanib

Participants

Inclusion criteria:

- all sexes
- ≥ 18 years of age
- diagnosis of clear cell RCC
- locally advanced RCC
- participants with only one prior systemic treatment for locally advanced or metastatic RCC*
- first-line systemic treatment* must be cytokine based

Or,

- no prior systemic therapy for advanced/metastatic RCC
- ECOG PS 0 or 1

NCT00334282 *(Continued)*

Exclusion criteria:

- history of another malignancy
- current or prior use of an investigational anti-cancer drug within 4 weeks of start of study
- prior use of an investigational or licensed drug that targets VEGF or VEGF receptors (e.g. bevacizumab, sunitinib, sorafenib, etc)

Sample size: N=233 treatment-naive participants

Age (years, median with range): experimental arm: 65 (25-80), control arm: 60 (25-81)

Sex (m/f): experimental arm: 61/19, control arm: 109/36

Prognostic factors:

- **ECOG PS, n(%)**
  - **0**
    - experimental arm: 27(34), control arm: 60 (41)
  - **1**
    - experimental arm: 43 (54), control arm: 85 (59)
  - **2**
    - experimental arm: 10 (13), control arm: 0 (0)
- **MSKCC risk category, n(%)**
  - **Favourable**
    - experimental arm: 31(39), control arm: 57(39)
  - **Intermediate**
    - experimental arm: 38(48), control arm: 77(53)
  - **Poor**
    - experimental arm: 1 (1), control arm: 5 (3)
  - **Unkown**
    - experimental arm: 10 (13), control arm: 6(4)
- **Prior nephrectomy n(%)**
  - **Yes**
    - experimental arm: 74 (93), control arm: 127 (88)

Interventions	**Experimental arm (n = 155):** Pazopanib (800 mg, oral, once/day) **Control arm (n = 78):** Placebo (800mg, oral, once/day)

Outcomes	**Primary outcome(s)**
	
	• PFS
	  ◦ Time frame: up to 2 years
	
	Secondary outcome(s)
	
	• OS
	  ◦ Time frame: up to 2 years
	• Health-related QoL
	  ◦ Time frame: baseline and weeks 6, 12, 18, 24, and 48
	• Safety (SAE)
	
	Relevant to this review but not reported: AE in first-line participants, TFST, number of (first-line) participants who discontinued treatment due to an AE
	
	Other outcomes (not relevant to this review): DoR, CR, ORR

Notes	**Funding sources:** GlaxoSmithKline **Declarations of Interests:** Quote: "No potential conflict of interest." stated by EL.

NCT00334282 *(Continued)*

Clinical study report available: yes

Study protocol available: yes

Statistical analysis plan available: yes

*Trial included both participants who have received prior treatment and participants who are treatment-naive. Results are reported separately for the treatment-naive participants in the publication. Hence, all data reported in this review refers to the treatment-naive group of participants only.

NCT00420888

Study characteristics

Methods

Study name: -

Study design: randomised, parallel-group trial, phase II/ III

Blinding: none, open-label

Study dates: May 2007 - October 2010 (date of randomisation)

Date of data cut-off: not reported

Location: 5 countries (Bulgaria, Romania, Russian Federation, Ukraine, UK), types of centres: hospitals, urology clinics, cancer centres, research centres (51 study locations)

Cross-over study or cross over permitted: not per design; not reported whether cross over was permitted at some point (e.g. upon progression)

Participants

Inclusion criteria:

- participants with confirmed metastatic or inoperable locally advanced RCC eligible for standard therapy with IFN
- histologically or cytologically confirmed clear cell or papillary type RCC
- KPS ≥ 70
- favourable or moderate risk group MSKCC
- life expectancy > 3 months
- acceptable levels of specific haematology and serum chemistry parameters

Sample size: N = 513

Age, mean in years (standard deviation (SD)): experimental arm: 58 (25-79), control arm: 57 (19-83)

Sex (m/f): experimental arm: 183/77), control arm: 183/70

prognostic factors:

- **ECOG PS, n (%)**
  - **0**
    - experimental arm: 164 (65); control arm: 159 (61)
  - **1**
    - experimental arm: 89 (35); control arm: 100 (39)
- **MSKCC risk subgroup, n (%)**
  - **Favourable**
    - experimental arm: 152 (60); control arm: 152 (59)
  - **Intermediate**
    - experimental arm: 101 (40); control arm: 108 (42)
- **Prior nephrectomy n(%**

NCT00420888 *(Continued)*

- **Yes**
  - experimental arm: 206 (81.4), control arm: 209 (80.4%)

Interventions	**Experimental arm (n=253):** naptumomab (15 mg/kg, intravenous, once/day) + IFN-alfa (9 MIU, subcutaneous injection, thrice/week) **Control arm (n = 260):** IFN-alfa (9 MIU, subcutaneous injection, thrice/week)
Outcomes	**Primary outcome(s)** • OS (time frame: every 12 weeks, including after a maximum of 18 months of study treatment) **Secondary outcome(s)** • PFS (time frame: every 12 weeks for the 18-month treatment period and also every 12 weeks after the treatment period) • AE (time frame: every visit through week 73) **Relevant to this review but not reported:** quality of life (QoL), serious adverse events (SAEs), time to first subsequent therapy (TFST), number of participants who discontinued treatment due to an AE **Other outcomes (not relevant to this review):** response rate (RR); immunological response to treatment in participants receiving naptumomab; pharmacokinetics
Notes	**Funding sources:** GlaxoSmithKline, Novartis, Pfizer and Bayer **Declarations of interests: "**R.E. Hawkins reports receiving commercial research grants from GlaxoSmithKline, Novartis, and Pfizer; speakers bureau honoraria from Bristol- Meyers Squibb, GlaxoSmithKline, Novartis, and Pfizer; and is a consultant/ advisory board member for Pfizer. G. Hedlund, G. Forsberg, and O. Nordle have ownership interest (including patents) in Active Biotech. T. Eisen is an employee of AstraZeneca; reports receiving commercial research grants from Bayer, GlaxoSmithKline, and Pfizer and other research grants from AstraZeneca; has ownership interest (including patents) in AstraZeneca; and is a consultant/ advisory board member for Aveo, Bayer, Bristol-Meyers Squibb, GlaxoSmithKline, Immatics, Novartis, and Pfizer. No potential conflicts of interest were disclosed by the other authors." **Clinical study report available:** no **Study protocol available:** no **Statistical analysis plan available:** no

NCT00609401

Study characteristics	
Methods	**Study name:** ROSORC **Study design:** randomised, phase II **Blinding:** none, open-label **Study dates:** October 2006 - February 2008 (date of enrolment) **Date of data cut-off:** September 30, 2012 (for OS) **Location:** 1 country (Italy), types of centres: not reported, but multicentre study **Cross-over study or cross over permitted:** not a cross-over study; not reported whether cross over was permitted at some point (e.g. upon progression)

NCT00609401 *(Continued)*

Participants	**Inclusion criteria:**
	• age >= 18 years
	• all sexes
	• Karnofsky PS >= 60%
	• cytohistological diagnosis of RCC
	• written informed consent
	• measurable disease according to RECIST criteria v. 1.0
	• life expectancy of greater than 3 months and an Eastern Cooperative Oncology Group performance status ≤2
	• histologically based diagnosis of mRCC
	• participants had not been previously treated with systemic therapy for metastatic disease, but they could have undergone nephrectomy
	
	Exclusion criteria:
	• history of brain metastases
	• presence of concomitant illnesses
	• medical conditions like unstable angina, uncontrolled hypertension, unstable diabetes mellitus, or potentially life-threatening autoimmune disorders
	
	Sample size: N = 128
	
	Age (years, median with range): experimental arm: 64 (57-69), control arm: 62 (52-69)
	
	Sex (m/f): experimental arm: 52/14, control arm: 43/19
	
	Prognostic factors:
	• **MSKCC risk group, n(%)**
	  ◦ **Low**
	  ▪ experimental arm: 36 (55), control arm: 34 (55)
	  ◦ **Intermediate**
	  ▪ experimental arm: 27 (41), control arm: 24 (39)
	  ◦ **High**
	  ▪ experimental arm: 3 (5), control arm: 4 (6)
	• **Prior nephrectomy n(%)**
	  ◦ **Yes**
	  ▪ experimental arm: 48 (73), control arm: 46 (74)

Interventions	**Experimental arm (n = 66):** sorafenib (400 mg, oral, twice/day) + IL-2 (3 MIU, subcutaneous injection, 5 days/week)**Control arm (n = 62):** Sorafenib (400mg, oral, twice/day)
Outcomes	**Primary outcome(s)** • PFS ◦ Time frame: 2 years **Secondary outcome(s)** • OS • Safety **Relevant to this review but not reported:** QoL,TFST, number of participants who discontinued treatment **Other outcomes (not relevant to this review):** ORR
Notes	**Funding sources:** editorial assistance for this manuscript was provided by Dragonfly Editorial, funded by Bayer HealthCare. This study was supported in part by Bayer HealthCare.

NCT00609401 *(Continued)*

Declarations of Interests: Quote: "The authors have declared no conflicts of interest."

Clinical study report available: no

Study protocol available: no

Statistical analysis plan available: no

NCT00619268

Study characteristics	
Methods	**Study name:** TORAVA **Study design:** randomised, phase II (three-arm trial) **Blinding:** none, open-label **Study dates:** March 3, 2008 - May 6, 2009 (date of randomisation) **Date of data cut-off:** not reported **Location:** 1 country (France), types of centres: cancer centres/institutes, hospitals, university hospitals (29 study locations) **Cross-over study or cross over permitted:** not a cross-over study per design; not reported whether cross over was permitted at some point (e.g. upon progression)
Participants	**Inclusion criteria:** • male or female participants >= 18 years of age • participants with histological or cytological evidence of metastatic renal cell carcinoma mostly of all type, except for papillary • no prior systemic treatment (chemotherapy, immunotherapy, anti-angiogenic drugs, or treatment under evaluation) for metastatic renal cancer • no brain metastases revealed by MRI or CT-scan within 28 days prior to randomisation • E.C.O.G performance status =<2 • at least one measurable lesion using the RECIST criteria • signed written informed consent • liver, renal, and haematological functions in the range of 1·5 to two times above or below normal values • normal lipid and glycaemic concentrations; normal cardiac function within 6 weeks before randomisation; and no hyper tension • no systemic treatment for the disease and no history of arterial or venous thrombosis in the past 6 months **Exclusion criteria:** • participant with pure papillary renal cell carcinoma • prior systemic treatment for metastatic renal cancer • history of other malignancies • evidence of brain metastasis by computerized tomographic scan or MRI in the 28 days prior to randomisation **Sample size:** N = 171 **Age (years, median with range):** experimental arm: 62.0 (33-83), control arm I: 61.2 (33-83), control arm II: 61.9 (40-79)

NCT00619268 *(Continued)*

Sex (m,f) : experimental arm: 65/23, control arm I: 32/10, Group 3: 27/14

Prognostic factors:

- **ECOG PS, n(%)**
  - **0 or 1**
    - experimental arm: 77(88), control arm I: 37(88), control arm II: 36(88)
  - **2**
    - experimental arm: 11(13), control arm I: 5(12), control arm II: 5(12)
- **MSKCC risk group, n (%)**
  - **Good risk**
    - experimental arm: 25(32), control arm I: 12(31), control arm II: 14(39)
  - **Intermediate risk**
    - experimental arm: 41(53), control arm I: 23(59), control arm II: 16(44)
  - **Poor risk**
    - experimental arm: 11(14), control arm I: 4(10), control arm II: 6(17)
- **Prior nephrectomy n(%)**
  - **Yes**
    - experimental arm I: 73 (83), experimental arm II: 41 (98), control arm: 35 (85)

Interventions	**Experimental arm (n = 88):** Temsirolimus (25mg, intravenous, once/week) + Bevacizumab (10mg/kg, intravenous, every 2 weeks) **Control arm I (n = 42):** Sunitinib (50mg, oral, daily) **Control arm II (n=41):** Bevacizumab (10mg/kg, intravenous, every 2 weeks) + IFN-alpha-2a (9 MIU, subcutaneous injection, thrice/week)

Outcomes	**Primary outcome(s)**
	
	• Progression-free rate
	  ○ Time frame: at 48 weeks post-treatment
	
	Secondary outcome(s)
	
	• Toxicity
	  ○ Time frame: at week 2, week 5-6 and after every 5-6 weeks during 48 weeks
	• Quality of Life (QoL)
	  ○ Time frame: at inclusion, month 6 and at 1 year
	• PFS
	  ○ Time frame: NI
	• OS
	  ○ Time frame: NI
	
	Relevant to this review but not reported: TFST
	
	Other outcomes (not relevant to this review): ORR, DoR, PFR

Notes	**Funding sources:** French Ministry of Health and Wyeth Pharmaceuticals, Centre Leon Berard **Declarations of Interests:** Quote "SN has received honoraria from Novartis, Wyeth, Pfizer, GlaxoSmithKline, and Roche; and has received research funding from Wyeth, Roche, and Novartis. DP has received honoraria from Bayer, Eli Lilly, and Roche. J-OB has received honoraria from Amgen and is a consultant with Novartis. LG and BL have received honoraria from Novartis. BE has received honoraria from Bayer, Roche, Pfizer, Genentech, Novartis, GlaxoSmithKline, and Aveo; and is a consultant with Bayer, Pfizer, and Roche. All other authors declared no conflicts of interest." **Clinical study report available**: no **Study protocol available**: no

NCT00619268 *(Continued)*

Statistical analysis plan available: no

NCT00631371

Study characteristics

Methods

Study name: INTORACT

Study design: randomised, phase III
Blinding: none, open-label

Study dates: April 10, 2008 - October 19, 2010 (date of randomisation)

Date of data cut-off: April 19, 2012

Location: 30 countries (Argentina, Australia, Belgium, Brazil, Canada, Chile, Colombia, Czech Republic, France, Germany, Hong Kong, Hungary, India, Italy, Republic of Korea, Malaysia, Mexico, Netherlands, Poland, Portugal, Russian Federation, Serbia, Singapore, Slovakia, South Africa, Spain, Taiwan, Ukraine, UK., USA), types of centres: hospitals, university hospitals, cancer centres, research centres (172 study locations)
Cross- over study or cross over permitted: not per design; not reported whether cross over was permitted at some point (e.g. upon progression)

Participants

Inclusion criteria:

- all sexes
- 18 years and older
- histologically and/or cytologically confirmed to have advanced (stage IV or recurrent) renal cell carcinoma (RCC)
- majority component of conventional clear-cell type is mandatory
- at least 1 measurable lesion (per RECIST)
- Karnofsky performance status >70%, life expectancy > 12 weeks
- adequate organ function
- written consent

Exclusion criteria:

- prior systemic treatment for RCC
- evidence of current or prior central nervous system (CNS) metastases
- cardiovascular disease, history of major thrombotic or bleeding episode within 6 months, inadequately controlled hypertension (systolic blood pressure>150 mmHg and/or diastolic blood pressure >100 mmHg on medication)
- pregnant or nursing women
- additional criteria applies
- major surgery or radiation therapy within 4 weeks, or chronic use of antiplatelet agents or corticosteroids

Sample size: N = 791

Age (years, median with range): experimental arm: 59 (22-87), control arm: 58 (23-81)

Sex (m/f, %): experimental arm: 286/114, control arm: 270/121

Prognostic factors:

- **MSKCC prognostic group, n(%)**
  - **Favourable**
    - experimental arm: 123 (31), control arm: 114 (29)

NCT00631371 *(Continued)*

- **Intermediate**
  - experimental arm: 230 (58), control arm: 237 (61)
- **Poor**
  - experimental arm: 47 (12), control arm: 40 (10)

- **Prior nephrectomy n(%)**
  - **Yes**
    - experimental arm: 338 (85), control arm: 335 (86)

Interventions	**Experimental arm (n = 400):** Bevacizumab (10 mg/kg, intravenous, every two weeks) + Temsirolimus (25mg, intravenous, once/week) **Control arm (n = 391):** Bevacizumab (10 mg/kg, intravenous, every two weeks) + IFN-alfa (9 MIU, subcutaneous, three times/ week)

Outcomes	**Primary outcome(s)**
	
	• PFS: independent-assessment
	  ◦ Time frame: baseline until disease progression, initiation of new anticancer treatment, or death, assessed every 8 weeks
	
	Secondary outcome(s)
	
	• PFS: investigator-assessment
	  ◦ Time frame: baseline until disease progression, initiation of new anticancer treatment, or death, assessed every 8 weeks
	• OS
	  ◦ Time frame: baseline until death due to any cause, assessed every 8 weeks
	• Safety (AE, SAE)
	• QoL
	• number of participants who discontinued treatment due to an AE
	
	Relevant to this review but not reported: TFST
	
	Other outcomes (not relevant to this review): ORR

Notes	**Funding Sources:** Brian I. Rini, Pfizer **Declarations of Interests:** Quote: "Although all authors completed the disclosure declaration, the following author(s) and/or an author's immediate family member(s) indicated a financial or other interest that is relevant to the subject matter under consideration in this article (...)For a detailed description of the disclosure categories, or for more information about ASCO's conflict of interest policy, please refer to the Author Disclosure Declaration and the Disclosures of Potential Conflicts of Interest section in Information for Contributors." **Clinical study report available**: no **Study protocol available**: no **Statistical analysis plan available**: no

NCT00719264

Study characteristics	
Methods	**Study name:** RECORD-2 **Study design:** a two-arm, RCT, phase II **Blinding:** none, open-label

NCT00719264 *(Continued)*

Date of enrolment/randomisation: not reported

Date of data cut-off: December 31, 2011 (for PFS); August 30, 2012 (for OS and safety)

Location: 21 countries (Belgium, Brazil, Czech Republic, Egypt, France, Germany, Hong Kong, Hungary, Italy, Republic of Korea, the Netherlands, Russian Federation, Singapore, South Africa, Spain, Switzerland, Taiwan, Thailand, Turkey, UK, USA.) types of centres: (108 study locations)
Cross-over study or cross over permitted: not per design; not reported whether cross over was permitted at some point (e.g. upon progression)

Participants

Inclusion criteria:

- all sexes
- 18 years and older
- participants with mRCC
- participants with progressive mRCC
- participants who had a prior partial or complete nephrectomy
- participants with a KPS ≥70%

Exclusion criteria:

- 4 weeks post-major surgery
- participants who had radiation therapy within 28 days prior to start of study
- participants in need for major surgical procedure during the course of the study
- participants who have received prior systemic treatment for their metastatic RCC
- participants who received prior therapy with VEGF pathway inhibitor

Sample size: N=365

Age, Mean (years, SD): experimental arm: 60.71 (10.6) , control arm: 59.9 (10.3)

Sex, m/f): experimental arm: 138/44, control arm: 131/52

Prognostic factors:

- **MSKCC risk, n(%)**
  - **Favourable**
    - experimental arm: 65 (35.7), control arm: 66 (36.1)
  - **Intermediate**
    - experimental arm: 104 (57.1), control arm: 104 (56.8)
  - **Poor**
    - experimental arm: 13 (7.1), control arm: 13 (7.1)
- **Prior nephrectomy n(%)**
  - **Yes**
    - all participants

Interventions

Experimental arm (n = 182): everolimus (10 mg, daily) + bevacizumab (10 mg/kg, every two weeks)

Control arm (n= 183): IFN, dose escalated from 3 MIU during week 1, 6 MIU during week 2, and 9 MIU during week 3 of treatment and subsequently (if tolerated), 3 times per week plus intravenous bevacizumab 10 mg/kg every 2 weeks

Outcomes

Primary outcome(s)

- PFS
  - Time frame: Time from randomisation to the date of radiological progressive disease as per independent central review, death from any cause, or last tumour assessment, reported between date of first participant randomised until 31Dec2011, cut-off date

Secondary outcome(s)

NCT00719264 *(Continued)*

- OS
  - Time frame: time from randomisation to the date of death from any cause, reported between date of first participant randomised and up to 2 years after the last participant randomised (data cutoff: 30 Aug2012)
- Number of participants who experienced AEs, SAEs and deaths
  - Time frame: from the first participant randomised until the last patient discontinued the study treatment + 28 days
- Time to Definitive Deterioration of the Global Health Status and the Physical Functioning (PF) Subscale Scores of the European Organization for the Research and Treatment of Cancer (EORTC)-Core Quality of Life Questionnaire (QLQ-C30) by at least 10%

Relevant to this review but not reported: PFS, TFST

Other outcomes (not relevant to this review): DoE, disease related symptoms, RDD, best OR

Notes

Funding sources: Novartis Pharmaceuticals. No grant numbers applied.

Declarations of Interests: Quote: "All remaining authors have declared no conflicts of interest."

Clinical study report available: no

Study protocol available: no

Statistical analysis plan available: no

NCT00720941

Study characteristics

Methods

Study name: COMPARZ

Study design: a two-arm, randomised, phase III

Blinding: none, open-label

Date of enrolment/randomisation: August 2008 - September 2011

Date of data cut-off: 21st May 2012 (PFS), 30th Sep. 2013 (for OS, AE and SAE)

Location: 14 countries (Australia, Canada, China, Germany, Ireland, Italy, Japan, Republic of Korea, the Netherlands, Spain, Sweden, Taiwan, UK, USA), types of centres: not reported (227 study locations)

Cross-over study or cross over permitted: not per design; not reported whether cross over was permitted at some point (e.g. upon progression)

Participants

Inclusion criteria:

- all sexes
- 18 years and older
- diagnosis of renal cell carcinoma with clear-cell component histology
- measurable disease according to the Response Evaluation Criteria in Solid Tumors (RECIST) guidelines
- received no prior systemic therapy (interleukin-2, interferon-alpha, chemotherapy, bevacizumab, mTOR inhibitor, sunitinib, sorafenib or other VEGF TKI) for advanced or metastatic RCC
- locally advanced or metastatic renal cell carcinoma
- KPS status of >=70

Exclusion criteria:

NCT00720941 *(Continued)*

- pregnant or lactating female (unless agrees to refrain from nursing throughout the treatment period and for 14 days following the last dose of study)
- history of another malignancy (unless have been disease-free for 3 years)
- History or clinical evidence of CNS metastases (unless have previously-treated CNS metastases and meet all 3 of the following criteria are: are asymptomatic, have had no evidence of active CNS metastases for >= 6 months prior to enrolment, and have no requirement for steroids or enzyme-inducing anticonvulsants)
- prior use of an investigational or licensed drug that targets VEGF or VEGF receptors (e.g. bevacizumab, sunitinib, sorafenib, etc), or are mTOR inhibitors (e.g. temsirolimus, everolimus, etc)
- is now undergoing and/or has undergone in the 14 days immediately prior to first dose of study drug, any cancer therapy (surgery, tumour embolisation, chemotherapy, radiation therapy, immunotherapy, biological therapy, or hormonal therapy)

Sample size: N=1110

Age, continuous mean (years, SD): Group 1: 60.9 (10.89), Group 2: 61.2 (10.98)

Sex (m/f): Group 1: 398/159, Group 2: 415/138

Prognostic factors:

- **KPS, n(%)**
  - **70 or 80**
    - experimental arm: 141 (25), control arm: 130 (24)
  - **90 or 100**
    - experimental arm: 416 (75), control arm: 423 (76)
- **Prior nephrectomy, n(%)**
  - **Yes**
    - experimental arm: 459 (82), control arm: 465 (84)
- **MSKCC risk category, n(%)**
  - **Favourable risk**
    - experimental arm: 151 (27), control arm: 152 (27)
  - **Intermediate risk**
    - experimental arm: 322 (40), control arm: 328 (59)
  - **Poor risk**
    - experimental arm: 67 (12), control arm: 52 (9)
  - **Unknown**
    - experimental arm: 17 (3), control arm: 21 (4)
- **Heng risk category, n(%)**
  - **Favourable risk**
    - experimental arm: 142 (23), control arm: 137 (28)
  - **Intermediate risk**
    - experimental arm: 299 (54), control arm: 308 (56)
  - **Poor risk**
    - experimental arm: 106 (19), control arm: 94 (17)
  - **Unknown**
    - experimental arm: 10 (2), control arm: 14 (3)

Interventions	**Experimental arm (n = 557):** Pazopanib (800 mg, oral, once/day) **Control arm (n = 553):** Sunitinib (50 mg, oral, once/day)
Outcomes	**Primary outcome(s)** • PFS ○ Time frame: from randomisation until the earliest date of disease progression or death (up to study week 191) **Secondary outcome(s)**

NCT00720941 *(Continued)*

- OS
  - Time frame: from randomisation until death (up to study week 268)
- Number of participants with SAEs/Non-SAEs (any untoward medical occurrence in a participants administered a pharmaceutical product and which does not necessarily have a causal relationship with this treatment)
  - Time frame: from the time of the first dose of study drug to approximately one month after the discontinuation of study drug (up to study week 268)
- QoL
- Number of participants who discontinued treatment due to an AE

Relevant to this review but not reported: TFST, Safety (AE, SAE)

Other outcomes (not relevant to this review): laboratory results, MRU, CTSQ, SQLQ, FKSI- 19 scale, FACIT-F scale, DoR

Notes

Funding sources: supported by GlaxoSmithKline Pharmaceuticals and Novartis Pharmaceuticals

Declarations of Interests: not reported

Clinical study report available: yes

Study protocol available: yes

Statistical analysis plan available: yes

NCT00732914

Study characteristics

Methods

Study name: SWITCH

Study design: a two-arm, randomised, phase III

Blinding: none, open-label

Study dates: February 2009 - December 2011 (date of randomisation)

Date of data cut-off: August 15, 2013 (primary analysis) *(data used in this review)*; January 14, 2014 (post-hoc analysis of OS)

Location: 1 country (Germany), types of centres: urology clinic (1 study location)

Cross-over study or cross over permitted: yes, cross-over study*

*For cross over trials we only extracted data from the first period.

Participants

Inclusion criteria:

- participants with metastatic / advanced RCC (all histologies), who are not suitable for cytokine therapy and for whom study medication constitutes first-line therapy
- age >= 18 and <= 85years
- ECOG PS of 0 or 1
- MSKCC prognostic score, low or intermediate
- life expectancy of at least 12 weeks
- participants with at least one uni-dimensional (for RECIST) measurable lesion. Lesions must be measured by CT/MRI-scan.

Exclusion criteria:

- history of cardiac disease: congestive heart failure

NCT00732914 *(Continued)*

- history of HIV infection or chronic hepatitis B or C
- active clinically serious infections (> grade 2 NCI-CTC version 3.0)
- symptomatic metastatic brain or meningeal tumours (unless the patient is > 6 months from definitive therapy, has a negative imaging study within 4 weeks of study entry and is clinically stable with respect to the tumour at the time of study entry)
- known allergy to sunitinib or sorafenib or one of its constituents

Excluded therapies and medications, previous and concomitant:

- anticancer chemotherapy or immunotherapy during the study or within 4 weeks of study entry
- radiotherapy during study or within 3 weeks of start of study drug. (Palliative radiotherapy will be allowed). Major surgery within 4 weeks of start of study
- investigational drug therapy outside of this trial during or within 4 weeks of study entry
- prior exposure to the study drug

Sample size: N =365

Median age (years, range): experimental arm: 63 (39-84), control arm: 65 (40-83)

Sex (m/f): experimental arm: 139/43, control arm: 135/48

Prognostic factors:

- **Nephrectomy, n(%)**
  - **Yes**
    - experimental arm: 167 (92), control arm: 168 (92)
- **MSKCC risk score, n(%)**
  - **High**
    - experimental arm: 1 (0.5), control arm: 1 (0.5)
  - **Intermediate**
    - experimental arm: 108 (59), control arm: 94 (51)
  - **Favourable**
    - experimental arm: 71 (39), control arm: 82 (45)
  - **Unknown**
    - experimental arm: 2 (1.1), control arm: 4 (2.2)
  - **Missing**
    - experimental arm: 0 (0), control arm: 2 (1.1)
- **ECOG performance scale, n(%)**
  - **0**
    - experimental arm: 116 (66), control arm: 106 (60)
  - **1**
    - experimental arm: 55 (31), control arm: 66 (38)
  - **2**
    - experimental arm: 0 (0), control arm: 1 (0.6)
  - **Missing**
    - experimental arm: 6 (3.4), control arm: 3 (1.7)

Interventions	**Experimental arm (n = 182):** Sorafenib (400 mg, oral, twice/day) **Control arm (n =183):** Sunitinib (50 mg, oral, once/day)
Outcomes	**Primary outcome(s)** • First-line* PFS **Secondary outcome(s)** • Safety (AE, SAE) in the first-line treatment • Number of participants who discontinued treatment due to an AE during first-line treatment

NCT00732914 *(Continued)*

	Relevant to this review but not reported: first-line OS, QoL, TFST **Other outcomes (not relevant to this review):** OS (after second-line treatment), time to progression (TTP), DCR (disease control rate), cardiotoxicity *First-line refers to first-line treatment (i.e. period 1 in this cross-over study).
Notes	**Funding sources:** Bayer, Pfizer, and Novartis quote: :"The SWITCH trial was sponsored by the German Cancer Society (DKG) with a financial grant from Bayer HealthCare. The Main Association of Austrian Social Security Institutions also supported the study. The specific role of the sponsors was in the design and conduct of the study. Bayer HealthCare also funded medical writing support for the preparation of this article." **Declarations of Interests:** Quote: "Christian Eichelberg certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (...)." **Clinical study report available:** no **Study protocol available:** no **Statistical analysis plan available:** no

NCT00738530

Study characteristics	
Methods	**Study name:** AVOREN **Study design:** randomised, parallel, placebo-controlled phase III **Blinding:** the study was planned as a double-blind trial, but was unblinded after a protocol amendment: quote: "An interim analysis of overall survival was prespecified after 250 deaths. On the basis of new second-line therapies that became available while the trial was in progress, which could have confounded analyses of overall survival data, we agreed with regulatory agencies that the preplanned final analysis of progression-free survival would be acceptable for regulatory submission." The protocol was amended to allow the study to be unblinded at this point. **Study dates:** between June 2004 and October 2005 (date of enrolment) **Date of data cut-off:** September 8, 2006 for final analysis of PFS *(data used in this review)* and interim analysis of OS; cutoff for final analysis of OS *(data used in this review)* was September 2008 **Location:** 18 countries (Australia, Belgium, Czech Republic, Finland, France, Germany, Hungary, Israel, Italy, the Netherlands, Norway, Poland, Russian Federation, Singapore, Spain, Switzerland, Taiwan, UK., types of centres: hospitals, cancer centres (104 study locations) **Cross-over study or cross over permitted:** not per design, but cross over from the control group to receive bevacizumab was recommended for participants who had not progressed
Participants	**Inclusion criteria:** • 18 years or older • all sexes • participants with measurable or non-measurable tumour (according to RECIST criteria) • participants with (>50%) clear-cell renal cell carcinoma • participants that have undergone nephrectomy or partial nephrectomy • KPS ≥ 70% **Exclusion criteria:**

NCT00738530 *(Continued)*

- prior systemic treatment for metastatic RCC
- recent major surgical procedures
- evidence of brain metastases
- ongoing need for full dose anticoagulants
- uncontrolled hypertension
- clinically significant cardiovascular disease.

Sample size: N = 649

Age (years, median with range): experimental arm: 61 (30-82), control arm: 60 (18-81)

Sex (m/f): experimental arm: 222/10, control arm: 234/88

Prognostic factors:

- **MSKCC risk score**
  - **Favourable**
    - experimental arm: 87 (27), control arm: 93 (29)
  - **Intermediate**
    - experimental arm: 183 (56), control arm: 180 (56)
  - **Poor**
    - experimental arm: 29 (9), control arm: 5 (8)
  - **Not available**
    - experimental arm: 28 (9), control arm: 24 (7)
- **Previous nephrectomy (n,%)**
  - **Yes**
    - (all participants)

Interventions	**Experimental arm (n = 327):** Bevacizumab (10mg/kg, intravenous, every two weeks + IFN-alfa (9 MIU, subcutaneous injection, thrice/week) **Control arm (n = 322):** Placebo + IFN-a (9 MIU, subcutaneous injection, thrice/week)
Outcomes	**Primary outcome(s)** • OS **Secondary outcome(s)** • PFS • Safety (AE/SAE) • Number of participants who discontinued treatment due to an AE **Relevant to this review but not reported:** QoL, TFST **Other outcomes (not relevant to this review):** ORR, OR, CR, TTF, TTP
Notes	**Funding sources:** Alain Ravaud, F. Hoffmann-La Roche, GlaxoSmithKline **Declarations of Interests:** Quote: "Although all authors completed the disclosure declaration, the following author(s) indicated a financial or other interest that is relevant to the subject matter under consideration in this article(...), please refer to the Author Disclosure Declaration and the Disclosures of Potential Conflicts of Interest section in Information for Contributors." **Clinical study report available:** no **Study protocol available:** no **Statistical analysis plan available:** no

NCT00903175

Study characteristics	
Methods	**Study name:** RECORD-3 **Study design:** RCT, phase II **Blinding:** none, open-label **Study Dates:** from October 2009 to June 2011 date of enrolment) **Date of data cut-off:** September 3, 2012 (primary analysis) **Location:** 19 countries (Argentina, Australia, Brazil, Canada, Denmark, France, Germany, Hong Kong, Italy, Republic of Korea, Mexico, the Netherlands, Peru, Spain, Taiwan, Thailand, Turkey, UK, USA), types of centres: cancer centres/institutes, hospitals, (84 study locations) **Cross-over study:** yes* *For cross-over trials, we extracted data on the first period only.

Participants	**Inclusion criteria:**
	
	• all sexes
	• 18 years and older
	• participants with advanced renal cell carcinoma
	• participants with at least one measurable lesion
	• participants with a Karnofsky Performance Status ≥70%
	• adequate bone marrow function
	• adequate liver function
	• adequate renal function
	• women of childbearing potential must have had a negative serum pregnancy test within 14 days prior to the administration of the study medication. Adequate contraception must be used while on study
	
	Exclusion criteria:
	
	• less than 4 weeks post-major surgery
	• participants who had radiation therapy within 4 weeks prior to start of study treatment (palliative radiotherapy to bone lesions allowed within 2 weeks prior to study treatment start)
	• participants in need for major surgical procedure during the course of the study
	• participants who have received prior systemic treatment for their metastatic RCC
	
	Sample size: N=471
	
	Age (median in years, (range): experimental arm: 62 (20-89), control arm: 62 (29-84)
	
	Sex (m/f): experimental arm: 166/72, control arm: 176/57
	
	Prognostic factors:
	
	• **MSKCC risk group, n (%)**
	  ◦ **Favorable**
	  ▪ experimental arm: 70 (29), control arm: 69 (30)
	  ◦ **Intermediate**
	  ▪ experimental arm: 132 (56), control arm: 131 (56)
	  ◦ **Poor**
	  ▪ experimental arm: 35 (15), control arm: 32 (14)
	• **Previous nephrectomy, n(%)**
	  ◦ **Yes**

NCT00903175 *(Continued)*

- experimental arm: 159 (67), control arm: 156 (67)

Interventions	**Experimental arm (n = 238):** everolimus (10 mg, oral, daily) **Control arm (n = 233): s**unitinib (50 mg, oral, daily)

Outcomes	**Primary outcome(s)**
	
	• PFS First-Line (PFS 1-L)
	  ◦ Time frame: based on radiological assessments every 3 months until disease progression, start of another antineoplastic therapy or for any other reason up to 35 months
	
	Secondary Outcome(s)
	
	• PFSl Combined (PFS-C)
	  ◦ Time frame: based on radiological assessments every 3 months until disease progression, start of another antineoplastic therapy or for any other reason up to about 56 months
	• AEs/SAEs
	• OS
	  ◦ Time frame: every 2 months from randomisation up to 3 years after last patient randomised
	• Time to Definitive Deterioration of the Global Health Status/QoL Scores of the EORTC QLQ-C30 by First and Second-Line Drugs Combined
	  ◦ Time frame: <=14 days prior to the first dose of study medication, on day 1, day 28 of every cycle, at the end of treatment visit, at the 28 day FUP visit and monthly thereafter for up to 3 months or until initiation of another anticancer therapy up to 35 months
	
	Relevant to this review but not reported: TFST
	
	Other outcomes (not relevant to this review): EORTC, FKSI-DRS Risk Score, DoR, ORR

Notes	**Funding sources:** Robert J. Motzer, Novartis, Pfizer, GlaxoSmithKline; Carlos H. Barrios, Novartis; Thomas Cosgriff, Novartis; Thomas W. Flaig, Amgen, Bayer AG/Onyx Pharmaceuticals, Genentech, GlaxoSmithKline, Novartis, Pfizer, ZymoGenetics; Ray Page, Pfizer; J. Thaddeus Beck, Novartis; Jennifer Knox, Pfizer, Novartis **Declarations of Interests:** Quote: "Although all authors completed the disclosure declaration, the following author(s) and/or an author's immediate family member(s) indicated a financial or other interest that is relevant to the subject matter under consideration in this article (...), (...)please refer to the Author Disclosure Declaration and the Disclosures of Potential Conflicts ofInterest section in Information for Contributors." **Clinical study report available:** no **Study protocol available:** yes **Statistical analysis plan available:** yes

NCT00920816

Study characteristics

Methods	**Study name: -** **Study design:** a two-arm RCT, phase III **Blinding:** none, open-label **Study dates:** August 25 2009 – July 27 2012 **Date of data cut-off:** July 27, 2012 (for PFS) and December 18, 2014 (for OS)

NCT00920816 *(Continued)*

Location: 14 countries (Bosnia and Herzegovina, Bulgaria, Chile, China, India, Malaysia, Mexico, Philippines, Romania, Russian Federation, South Africa, Taiwan, Ukraine, U.S.A.), types of centres: cancer centres, medical centres, hospitals, university hospitals (125 study locations)

Cross-over study or cross over permitted: no

Participants

Inclusion criteria:

- all sexes
- 18 years and older
- histologically documented metastatic renal cell cancer with a component of a clear cell histology
- evidence of measurable disease
- participants with mRCC must have received no prior systemic first-line therapy or must have progressive disease per RECIST (version 1.0) after one prior systemic first line regimen* for metastatic disease containing sunitinib, cytokine(s), or both

Exclusion criteria:

- prior treatment for metastatic renal cell cancer with more than one systemic first line therapy
- major surgery less that 4 weeks or radiation less than 2 weeks of starting study drug

Sample size: N = 288

Age, mean (range): experimental arm: 58 (23-83), control arm: 58 (20-77)

Sex (m/f): experimental arm: 134/58, Group 0: 74/22

Prognostic factors:

- **ECOG Performance Scale, n(%)**
  - **0**
    - experimental arm: 21 (44), control arm: 9 (38)
  - **1**
    - experimental arm: 27 (56), control arm: 15 (62)
- **MSKCC risk group**
  - **Favourable**
    - experimental arm: 22 (46), control arm: 10 (42)
  - **Intermediate**
    - experimental arm: 22 (46), control arm: 13 (54)
  - **Poor**
    - experimental arm: 3 (6), control arm: 1 (4)
  - **Missing**
    - experimental arm: 1 (2), control arm: 0 (0)
  - **Previous nephrectomy, n(%)**
    - **Yes**
    - experimental arm: 164 (85), control arm: 86 (90)

Interventions

Experimental arm (n = 192): axitinib (5 mg, oral, twice/day)

Control arm (n = 96): sorafenib (400 mg, oral, twice/day)

Outcomes

Primary outcome(s)

- PFSl: first-line participants
  - Time frame: baseline until disease progression or death (assessed on Week 6, Week 12 and thereafter every 8 weeks up to Week 107/ 103)

Secondary outcome(s)

- AEs/SAEs
- discontinued treatment

NCT00920816 *(Continued)*

- QoL
- OS: first-line participants
  - Time frame: baseline until death (assessed on Week 6, Week 12 and thereafter every 8 weeks up to Week 103)

Relevant to this review but not reported: TFST

Other outcomes (not relevant to this review): OS and PFS in second-line participants, DoR, OR

Notes

Funding sources: AVEO, Bayer, GlaxoSmithKline, Novartis, and Pfizer

Declarations of Interests: Quote: "Angel H. Bair, Brad Rosbrook, and Glen I. Andrews are employees of and own stock in Pfizer. Nicholas J. Vogelzang has served on a speakers bureau for Pfizer. The remaining authors have stated that they have no conflicts of interest."

Clinical study report available: no

Study protocol available: no

Statistical analysis plan available: no

***Data of treatment-naive participants was extracted for this review.**

NCT00979966

Study characteristics

Methods

Study name: -

Study design: RCT, parallel assignment, phase II

Blinding: none, open-label

Study dates: July 2009 - July 2012

Date of data cut-off: not reported

Location: 1 country (Germany), types of centres: clinics, university hospitals (14 study locations)

Cross-over study or cross over permitted: not per design; not reported whether cross over was permitted at some point (e.g. upon progression)

Participants

Inclusion criteria:

- adult males and females: ≥18 years of age.
- locally advanced or metastatic, histological confirmed, non-clear cell RCC of all subtypes. participants must have advanced non-clear cell of one of the following subtypes: papillary, chromophobe, collecting duct carcinoma (CDC), renal medullary carcinoma (RMC), or unclassified.
- participants with measurable disease (at least one uni-dimensionally measurable target lesion by CT-scan or MRI) according to Response Evaluation Criteria in Solid Tumors (RECIST 1.1) If prior palliative radiotherapy to metastatic lesions: ≥ 1 measurable lesion that has not been irradiated.
- PS 0-2 ECOG

Exclusion criteria:

- predominant clear-cell RCC
- resectability or other curative options
- any investigational drug within the 30 days before inclusion.
- prior systemic treatment for their RCC.
- known or suspected allergy or hypersensitivity reaction to any of the components of study treatments.

NCT00979966 *(Continued)*

- radiotherapy within the last 4 weeks.
- pregnancy (absence to be confirmed by beta-hCG test) or lactation period.
- men or women of child-bearing potential who are sexually active and unwilling to use a medically acceptable method of contraception during the trial.
- clinically symptomatic brain or meningeal metastasis (known or suspected)

More inclusion and exclusion criteria on CT.gov.

Sample size: N = 22

Age, mean (range): experimental arm: 57.4 (29-85), control arm: 64.8 (46-80)

Sex (m/f): experimental arm: 8/4, control arm: 8/2

Prognostic factors:

- **ECOG Performance Scale, n**
  - **0**
    - experimental arm: 7, control arm: 6
  - **1**
    - experimental arm: 5, control arm: 4
- **Previous nephrectomy, n(%)**
  - **Yes**
    - total population: 106 (85.5)

Interventions	**Experimental arm (n = 12):** temsirolimus (25 mg, intravenous, once/week) **Control arm (n = 10):** sunitinib (50 mg, oral, once/day)
Outcomes	**Primary outcome(s)** - **Secondary outcome(s)** • safety assessed using CTCAE v3.0 and safety assessed according to reported SAEs (time frame: 8-12 months (treatment duration + 1 months)) • one year PFS rate (time frame: 1 year) • overall survival (OS) (time frame: will be evaluated in 2013) • Number of participants who discontinued treatment due to an AE **Relevant to this review but not reported:** TFST **Other outcomes (not relevant to this review):** OR, TTP, CR/PR, RB
Notes	**Funding sources:** This study was supported by a grant from Pfizer Germany. **Declarations of Interests:** Full details provided in Bergmann (2020). **Clinical study report available**: no **Study protocol available:** no **Statistical analysis plan available:** no

NCT01024920

Study characteristics	
Methods	**Study name:** -

NCT01024920 *(Continued)*

Study design: two-arm RCT, phase II

Blinding: none, open-label

Study dates: 11 March 2010 - 14 December 2010 (date of enrolment)

Date of data cut-off: 21 February, 2014

Location: 5 countries (Hungary, Poland, Romania, Ukraine, UK), types of centres: cancer centres, hospitals, university hospitals, research centres (15 study locations)

Cross-over study or cross over permitted: not per design; not reported whether cross over was permitted at some point (e.g. upon progression)

Participants

Inclusion criteria:

- all sexes
- 18 years and older
- participants with unresectable or metastatic Renal Cell Cancer, who have received no previous systemic anti-cancer treatment
- histological-confirmed diagnosis of renal cell cancer with clear cell component
- acceptable renal, liver, cardiovascular, bone marrow and other functions to allow sunitinib/nintedanib treatment

Exclusion criteria:

- participants unable to tolerate sunitinib/nintedanib treatment
- treatment with other investigational drugs or participation in another clinical study within the past 4 weeks before start of therapy or concomitantly with this study
- participants unable to comply with the 1199.26 protocol
- pregnancy or breast feeding
- active alcohol or drug abuse
- women of child bearing potential, or men who are able to father a child, unwilling to use a medically acceptable form of contraception during the study period

Sample size: N=96

Age median (yrs, range): experimental arm: 62 (42-86), control arm: 58 (29-79)

Sex (m/f): experimental arm: 44/2, control arm: 22/10

Prognostic factors:

- **ECOG performance status, n(%)**
  - **0**
    - experimental arm: 13 (20.3), control arm: 10 (31.3)
  - **1**
    - experimental arm: 51 (79.7), control arm: 22 (68.8)
- **MSKCC risk category, n(%)**
  - **Favourable/Intermediate**
    - experimental arm: 61 (95.3), control arm: 30 (93.8)
  - **Poor**
    - experimental arm: 3 (4.7), control arm: 2 (6.3)
- **Previous nephrectomy (n,%)**
  - **Yes**
    - experimental arm: 56 (87.5), control arm: 28 (87.5)

Interventions

Experimental arm (n = 64): nintedanib (200 mg, oral, twice/day)

Control arm (n = 32): sunitinib (50 mg, oral, once/day)

NCT01024920 *(Continued)*

	Treatment was four weeks on treatment, two weeks off treatment (one cycle = six weeks).

Outcomes	**Primary outcome(s)**
	• Probability rates of PFSl at 9 Months
	  ○ Time frame: PFS after 9 months
	Secondary outcome(s)
	• PFS
	  ○ Time frame: from the start of study until the cut-off date for 3 year efficacy analysis
	• OSl
	  ○ Time frame: from the start of study until the cut-off date for 3 year efficacy analysis
	• Safety (AEs/SAEs)
	• Number of participants who discontinued treatment due to an AE
	Relevant to this review but not reported: QoL, TFST
	Other outcomes (not relevant to this review): TTP, TTF, OR

Notes	**Funding sources:** Quote: "Pfizer, Bayer and AstraZeneca, and travel funding for a conference from Novartis. MM and YS have received research funding from Boehringer Ingelheim. RJJ has received research funding from Bristol-Myers Squibb, Boehringer Ingelheim, GlaxoSmithKline, Novartis, Pfizer and Roche. Funding for this study was provided by Boehringer Ingelheim." **Declaration of Interest:** Quote: "TE is a part-time employee (from 1 September 2014) of AstraZeneca and owns stock or other ownership interest in the company, and has had a consulting or advisory role and received honoraria from Bayer, Pfizer, AVEO Oncology and GlaxoSmithKline. RJJ has received honoraria from Bristol-Myers Squibb, GlaxoSmithKline, Novartis, Pfizer and Roche. A-BL, GT and HD are all employees of Boehringer Ingelheim. IB and NM report no conflicts of interest." **Clinical study report available**: no **Study protocol available:** no **Statistical analysis plan available:** no

NCT01030783

Study characteristics	
Methods	**Study name:** TIVO-I **Study design:** randomised, parallel, phase III trial **Blinding:** none, open-label **Study dates:** February - August 2010 (date of randomisation) **Date of data cut-off:** December 15, 2011 **Location:** 16 countries (Argentina, Bulgaria, Canada, Chile, Czech Republic, France, Hungary, India, Italy, Poland, Romania, Russian Federation, Serbia, Ukraine, U.K., U.S.A.), types of centres: not reported (86 study locations) **Cross-over study:** not per design, but cross over was permitted upon progression (from sorafenib to pazopanib)
Participants	**Inclusion criteria:** • ≥ 18-years of age

NCT01030783 *(Continued)*

- participants with recurrent or metastatic RCC
- participants must have undergone prior nephrectomy (complete or partial) for excision of the primary tumour
- histologically or cytologically confirmed RCC with a clear cell component
- measurable disease per the RECIST criteria Version 1.0
- treatment-naïve participants or participants who have received no more than one prior systemic treatment for metastatic RCC*
- ECOG performance status of 0 or 1, and life expectancy ≥ 3 months

Exclusion criteria:

- any prior VEGF-directed therapy including VEGF antibody
- any prior therapy with an agent targeting the mTOR pathway
- pregnant or lactating females

Sample size: N = 362 treatment-naive participants

Age (years, median with range): experimental arm: 59 (23-83), control arm: 59 (23-83)

Sex (m/f): experimental arm: 185/75, control arm: 189/68

Prognostic factors:

- **ECOG performance status, n(%)**
  - **0**
    - experimental arm: 116 (45), control arm: 139 (54)
  - **1**
    - experimental arm: 144 (55), control arm: 118 (46)
- **MSKCC risk category, n(%)**
  - **Favourable**
    - experimental arm: 70 (27), control arm: 87 (34)
  - **Intermediate**
    - experimental arm: 173 (67), control arm: 160 (62)
  - **Poor**
    - experimental arm: 17 (7), control arm: 10 (4)
- **Previous nephrectomy (n,%)**
  - **Yes**
    - All participants

Interventions

Experimental arm (n = 181 treatment-naive participants): tivozanib (1.5 mg, oral, once/day) - treatment cycle (four weeks): three weeks on treatment, one week off treatment

Control arm (n = 181 treatment-naive participants): sorafenib (400 mg, oral, twice/day) - treatment cycle was four weeks on treatment

Outcomes

Primary Outcome(s)

- PFS
  - Time frame: from date of randomisation until the date of first documented progression or date of death from any cause, whichever came first. Disease progression was assessed every 8 weeks

Secondary Outcome(s)

- OS
  - Time frame: date of randomisation to date of death
- Safety and Tolerability (AEs/SAEs)
  - Time frame: from start of treatment therapy to completion of treatment therapy, an average of 11 months
- QoL
- Number of participants who discontinued treatment due to an AE

NCT01030783 *(Continued)*

	Relevant to this review but not reported: TFST **Other outcomes (not relevant to this review):** ORR, DoR, pharmacokinetics, health outcome measurements
Notes	**Funding sources:** Robert J. Motzer, AVEO Pharmaceuticals, Pfizer, GlaxoSmithKline; Dmitry Nosov, Bayer Pharmaceuticals, Pfizer, GlaxoSmithKline; Timothy Eisen, Bayer Pharmaceuticals, Pfizer, GlaxoSmithKline; Vladimir Lesovoy, Pfizer; Anna Alyasova, AVEO Pharmaceuticals, Astellas Pharma; David Cella, AVEO Pharmaceuticals; Thomas E. Hutson, Pfizer, AVEO Pharmaceuticals, Novartis, GlaxoSmithKline **Declaration of Interest:** Quote: "Although all authors completed the disclosure declaration, the following author(s) and/or an author's immediate family member(s) indicated a financial or other interest that is relevant to the subject matter under consideration in this article. (...)For a detailed description of the disclosure categories, or for more information about ASCO's conflict of interest policy, please refer to the Author Disclosure Declaration and the Disclosures of Potential Conflicts ofInterest section in Information for Contributors." **Clinical study report available**: no **Study protocol available:** yes **Statistical analysis plan available:** yes ***Seperate data were reported for treatment-naive participants; hence, we extracted the data for analyses only on the treatment-naive.**

NCT01064310

Study characteristics	
Methods	**Study name:** PISCES **Study design:** randomised, phase III b **Blinding:** quadruple (participant, care provider, investigator, outcomes assessor) **Study dates:** May 2010 – October 2011 **Date of data cut-off:** not reported **Location:** 5 countries (Finland, France, Germany, Italy, UK), types of centres: not reported (51 study locations) **Cross-over study or cross over permitted:** yes, cross-over study* *For cross-over studies, we only extracted data on period 1.
Participants	**Inclusion criteria:** • participants ≥ 18 years of age • all sexes • no prior systemic therapy • locally advanced or metastatic renal cell carcinoma of any histology • ECOG PS 0 or 1 **Exclusion criteria:** • poor MSKCC risk group • history of another malignancy

NCT01064310 *(Continued)*

Sample size: N=169*

(*) One patient was randomly assigned in error, and no data were entered into the study for this patient; data were available for 168 participants.

Age (years, median): experimental arm: 64, control arm: 62

Sex (male/f): experimental arm: 61/3, control arm: 52/10

Prognostic factors:

- **ECOG PS, n (%)**
  - **0**
    - experimental arm: 60 (70), control arm: 61 (74)
  - **1**
    - experimental arm: 26 (30), control arm: 21 (26)
- **Previous nephrectomy (n,%)**
  - **Yes**
    - experimental arm 70 (85), control arm 79 (92)

Interventions	**Experimental arm (n = 86):** pazopanib (800 mg, oral, once/day) - 10 weeks on treatment, followed by 2 weeks wash-out **Control arm (n = 82):** sunitinib (50 mg, oral, once/day) - 4 weeks on treatment, 2 weeks placebo, followed by another 4 weeks on treatment

Outcomes	**Primary outcome(s)**
	
	-
	
	Secondary outcome(s)
	
	• Quality of Life as assessed by the EuroQoL-5 Dimensions (EQ-5D) Thermometer and Utility Scores and FACIT-Fatigue
	  ◦ Time frame: Day 1 (Period 1 Pre-dose); during 2-week Wash-out Period (Study Weeks 11 and 12); and End of Study (Week 10 of Period 2 [Study Week 22])
	• Number of participants with grade 1 to grade 5 (AEs)
	  ◦ Time frame: baseline to end of study (maximum of 22 weeks)
	• Number of participants with the indicated AEs leading to permanent discontinuation of study treatment
	  ◦ Time frame: baseline to end of study (maximum of 22 weeks)
	• Number of participants who discontinued treatment due to an AE
	
	Relevant to this review but not reported: PFS, OS, TFST
	
	Other outcomes (not relevant to this review): number of participants with preference for pazopanib versus sunitinib as assessed by the Patient Preference Questionnaire (PPQ), (BL), dose reduction, dose modification

Notes	**Funding sources:** Novartis Pharmaceuticals, Camillo Porta, Bayer Schering Pharma, Novartis, Pfizer; Petri Bono, GlaxoSmithKline, Pfizer; Thomas Powles, GlaxoSmithKline, Pfizer; Tim Eisen, Bayer AG, Pfizer, GlaxoSmithKline; Robert Hawkins, Pfizer, GlaxoSmithKline; David Cella, GlaxoSmithKline, Pfizer, AVEO Pharmaceuticals **Declaration of Interest:** Quote: "Although all authors completed the disclosure declaration, the following author(s) and/or an author's immediate family member(s) indicated a financial or other interest that is relevant to the subject matter under consideration in this article. (...)For a detailed description of the disclosure categories, or for more information about ASCO's conflict of interest policy, please refer to the Author Disclosure Declaration and the Disclosures of Potential Conflicts of Interest section in Information for Contributors." **Clinical study report available**: only a 'scientific result summary' available

NCT01064310 *(Continued)*

Study protocol available: yes

Statistical analysis plan available: yes

NCT01108445

Study characteristics

Methods

Study name: ASPEN

Study design: a two-arm RCT, phase II

Blinding: none, open-label

Study dates: Between Sept 23, 2010, and Oct 28, 2013 (date of randomisation)

Date of data cut-off: December, 2014 (database lock)

Location: 3 countries (Canada, UK, USA), types of centres: medical centres/agencies, clinics, hospitals, (18 study locations)

Cross-over study or cross over permitted: not a cross-over study per design, but cross over was permitted upon progression (from sunitinib to everolimus)

Participants

Inclusion criteria:

- histologically confirmed advanced RCC, with non-clear cell pathology
- at the time of screening, at least 4 weeks since prior palliative radiation therapy and/or major surgery, and resolution of all toxic effects of prior therapy to NCI Common Terminology Criteria for Adverse Events Grade 1
- participant must have radiographic evidence of metastatic disease with at least 1 measurable per RECIST 1.1 criteria
- age > 18 years
- KPS ≥ 60
- life expectancy of at least 3 months

Exclusion criteria:

- participants with a history of or active CNS metastases
- prior systemic therapy for RCC, including mTOR and anti-angiogenic therapy, chemotherapy, biologic or experimental therapy
- participants receiving known strong CYP3A4 isoenzyme inhibitors and/or inducers
- participants receiving immunosuppressive agents and those with chronic viral/bacterial/fungal illnesses such as HIV
- history of other prior malignancy in past 5 years
- known hypersensitivity to any of the components in everolimus or sunitinib product

Sample size: 131 participants signed consent. 22 were screen failures. 1 withdrew consent prior to being randomised. 108 participants were randomised.

Age (years, median with range): experimental arm: 64 (29-90), control arm: 59 (24-100)

Sex (m/f): experimental arm: 44/13, control arm: 37/14

Prognostic factors:

- **MSKCC risk group (0, 1-2, ≥3) (%)**
  - experimental arm:
    - **0** 14 (25),

NCT01108445 *(Continued)*

- **1-2** 32 (56),
- **≥3** 11 (19)

- control arm:
  - **0** 15 (29),
  - **1-2** 32 (63),
  - **≥3** 4 (8)

- **Prior nephrectomy (n, %)**
  - **Yes**
    - experimental arm: 45 (79), control arm: 41 (80)

Interventions	**Experimental arm (n= 57):** everolimus (10 mg, oral, once/day) **Control arm (n= 51): s**unitinib (50 mg, oral, once/day) Treatment was four weeks on treatment, two weeks off treatment (one cycle = six weeks).

Outcomes	**Primary outcome(s)**
	
	• Anti-tumor activity as measured by median PFS time
	  ○ Time frame: 24 months
	
	Secondary outcome(s)
	
	• PFS rates
	  ○ Time frame: 6, 12, and 24 months
	• PFS expressed in months
	  ○ Time frame: 24 Months
	• OSR
	  ○ Time frame: 6, 12, 24, 36 months
	• Median OS
	  ○ Time frame: up to 40 months
	• Percentage of participants with AEs
	  ○ Time frame: 24 months
	• QoL
	  ○ Time frame: baseline up to 40 months
	• Number of participants who discontinued treatment due to an AE
	
	Relevant to this review but not reported: TFST
	
	Other outcomes (not relevant to this review): Time-to-new metastatic disease, DoR, tumour shrinkage, clinical benefit rate, SD, ORR

Notes	**Funding:** Novartis and Pfizer **Declaration of Interest:** Quote: "AJA and DJG reports grants from Novartis and Pfizer during the conduct of the study; grants and personal fees from Dendreon, Sanofi -Aventis, Bayer, Medivation/Astellas, and Janssen, outside the submitted work. DJG also reports grants from Innocrin and Exelixis and personal fees from BMS and Janssen. TE is an employee of AstraZeneca and reports grants from AstraZeneca, personal fees from Novartis, Roche, BMS, and AVEO, grants from Bayer, grants and personal fees from Pfizer, GSK, personal fees and grant to institution from Astellas, outside the submitted work. JAG reports grants and personal fees from Pfizer and Novartis, during the conduct of the study; grants and personal fees from Bayer and Medivation/Astellas, and personal fees from Sanofi-Aventis, outside the submitted work. TFL reports grants from Novartis and Pfizer, during the conduct of the study; grants from Abbott, Abraxis, Acceleron, Amgen, AstraZeneca, Biovex, and Cerulean, Eisai, Eli Lilly grants and personal fees from Argos and Aveo, Bristol-Myers Squibb, Celgene, GlaxoSmithKline, Hoffman-La Roche, Immatics, Merck, Roche, Synta, Threshold, Tracon, EMD Serono, Millennium, and Schering-Plough, personal fees from Genentech, and grants and personal fees from Novartis, Pfizer, Prometheus, and Wyeth, outside the submitted work. CKK reports personal fees from Pfizer, Novartis, BMS, and Sanofi-Aventis, outside the submitted work. UNV reports grants and personal fees from Novartis and Pfizer, outside the submitted work. CWR reports personal fees from Pfizer and Genentech,

NCT01108445 *(Continued)*

research grant to institution from Onyx, outside the submitted work. RJJ reports grants from Pfizer and Novartis, during the conduct of the study; grants and personal fees from Pfizer, and grants, personal fees, and non-financial support from Novartis and GSK, outside the submitted work. WMS reports grants and personal fees from Pfizer, outside the submitted work. LMP reports personal fees from Pfizer and Novartis. SH, SB, JP, REH, JDH, IP, AP, CML, and SO declare no competing interests."

Clinical study report available: no

Study protocol available: no

Statistical analysis plan available: no

NCT01274273

Study characteristics

Methods

Study name: DARENCA

Study design: a two-arm, RCT, phase II

Blinding: none, open-label

Study dates: 26 October 2009 - 21 November 2014 (date of randomisation)

Date of data cut-off: 31 May 2017 (final analysis)

Location: 1 country (Denmark), type of centre: university hospitals (2 study locations)

Cross-over study or cross over permitted: no

Participants

Inclusion criteria:

- age ≥ 18 years
- all sexes
- histologic or cytologic biopsy proven locally advanced or metastatic renal cell carcinoma, considered non-candidates for curative surgery nephrectomy is not mandatory
- MSKCC favourable- and intermediate prognostic group
- measurable or non-measurable disease (as per RECIST 1.1 criteria)
- KPS of 70% or higher

Exclusion criteria:

- prior systemic treatment for metastatic RCC disease
- major surgical procedure, open surgical biopsy, or significant traumatic injury within 28 days prior to randomisation
- evidence of current central nervous system (CNS) metastases or spinal cord compression. Patient must undergo an MRI or CT scan of the brain (with contrast, if possible) within 28 days prior to randomisation
- history or presence of other disease, metabolic dysfunction, physical examination finding, or clinical laboratory finding giving reasonable suspicion of a disease or condition that contraindicates use of an investigational drug or patient at high risk from treatment complications
- known hypersensitivity to interleukin-2, Interferon, alfa or bevacizumab

Sample size: N = 118

Age, Mean (years, range): experimental arm: 58 (28-70), control arm: 55 (37-69)

Sex (m/f): experimental arm: 46/13, control arm:: 47/12

Prognostic factors:

NCT01274273 *(Continued)*

- **KPS, n(%)**
  - **100**
    - experimental arm: 31 (53), control arm: 37 (63)
  - **90**
    - experimental arm: 19 (32), control arm: 16 (27)
  - **80**
    - experimental arm: 6 (10), control arm: 4 (7)
  - **70**
    - experimental arm: 3 (5), control arm: 2 (3)
- **IMDC risk, n(%)**
  - **Favourable**
    - experimental arm: 14 (24), control arm: 12 (20)
  - **Intermediate**
    - experimental arm: 32 (54), control arm: 36 (61)
  - **Poor**
    - experimental arm: 13 (22), control arm: 11 (19)
- **MSKCC risk, n(%)**
  - **Favourable**
    - experimental arm: 30 (51), control arm: 31 (52)
  - **Intermediate**
    - experimental arm: 29 (49), control arm: 28 (48)
- **Nephrectomy, n(%)**
  - **Yes**
    - experimental arm: 50 (85), control arm: 51 (86)

Interventions	**Experimental arm (n = 59):** interleukin-2 (2.4 MIU/m^2, subcutaneous, twice/day, 5 days per week) + interferon (3 MIU, subcutaneous, once/day, 5 days/week) + bevacizumab (10 mg/kg, intravenous, every 2 weeks) **Control arm (n = 59):** interleukin-2 (2.4 MIU/m^2, subcutaneous, twice/day, 5 days per week) + interferon (3 MIU, subcutaneous, once/day, 5 days/week) All cytokines were administered over 4-week cycles for up to a maximum of 9 cycles (i.e., 9 months):

Outcomes	**Primary outcome(s)**
	
	• PFS
	• exact time frame of assessment not reported
	
	Secondary outcome(s)
	
	• OS
	  ○ exact time frame of assessment not reported
	• Toxicity (AEs)
	  ○ exact time frame of assessment not reported
	• Number of participants who discontinued treatment due to an AE
	
	Relevant to this review but not reported: TFST, SAEs
	
	Other outcomes (not relevant to this review): NED, biomarkers in imaging and biopsies, surgical resection, TTF, TTP, DoR, ORR (CR/PR), tolerability, TTF

Notes	**Funding sources:** Quote: "This study was supported financially by Roche and Novartis and BEV was provided by Roche." **Declarations of Interests:** "No potential conflict of interest was reported by the authors. Roche provided BEV. Roche and Novartis did not have access to data." **Clinical study report available**: no

NCT01274273 *(Continued)*

Study protocol available: no

Statistical analysis plan available: no

NCT01392183

Study characteristics

Methods

Study name: TemPa

Study design: a two-arm, RCT, phase II

Blinding: none, open-label

Study dates: November 2012 - June 2017 (date of enrolment)

Date of data cut-off: not reported

Location: 1 country (USA), type of centre: cancer centre (1 study location)

Cross-over study or cross over permitted: yes, cross-over study*

*For cross-over studies, only data on period 1 were extracted for analyses.

Participants

Inclusion criteria:

- all sexes
- 18 years and older
- pathologic confirmation of metastatic or locally advanced RCC with a major clear cell component
- measurable disease by RECIST criteria
- age >/= 18 years
- ECOG PS 0-2 or KPS >/= 60%
- meets criteria for poor-risk defined as 3 or more of the following: ECOG performance status 2, anaemia (haemoglobin lower than reference range), elevated serum LDH > 1.5x upper limit of normal (ULN), hypercalcaemia (corrected serum calcium level > upper limit of normal), time from initial RCC diagnosis to registration on this trial < 1 year, and > 1 metastatic organ sites

Exclusion criteria:

- prior malignancy, except for non-melanoma skin cancer, in situ carcinoma of any site, or other cancers for which the patient has been adequately treated and disease free for 2 years
- prior targeted therapy (anti-VEGF agents or mTOR inhibitors) including adjuvant therapy, and prior chemotherapy for mRCC. However, prior immunotherapy (cytokines or vaccines) is allowed
- any experimental drug while on this study; however, concomitant bone targeted therapy (bisphosphonates or the anti-RANK ligand denosumab) is allowed
- uncontrolled brain metastases and infections. participants with brain metastases treated with Gamma Knife (GK) or whole brain radiation within 24 hours of registration
- major surgery within 28 days prior to registration

Sample size: N=69

Age (median, range (years)): experimental arm: 61 (42-80), control arm: 61 (37-74)

Sex (m/f): experimental arm: 24/11, control arm: 28/6

Prognostic factors:

- **ECOG Performance Scale, n(%)**
  - **0**

NCT01392183 *(Continued)*

- experimental arm: 1 (2.9), control arm: 1 (2.9)
- **1**
  - experimental arm: 14 (40), control arm: 12 (35.3)
- **2**
  - experimental arm: 20 (57.1), control arm: 21 (61.8)

- **Previous nephrectomy**
  - **Yes**
    - experimental arm: 15 (42.9), control arm: 15 (44.1)
- **IMDC risk**
  - **Intermediate**
    - experimental arm: 11 (31.4), control arm: 8 (23.5)
  - **Poor**
    - experimental arm: 24 (68.6), control arm: 26 (76.5)
  - **Previous IL-2**
    - experimental arm: 2 (5.7), control arm: 1(2.9)

Interventions	**Experimental arm (n = 35):** temsirolimus (25 mg, intravenous, once/week) **Control arm(= 34):** pazopanib (800 mg, oral, once/day)

Outcomes	**Primary outcome(s)**
	
	• PFS
	  ◦ Time frame: assessments every 8 weeks from baseline to 1 year
	
	Secondary outcome(s)
	
	• OS
	  ◦ Time frame: assessments every 8 weeks from baseline to 1 year
	• Number of participants who discontinued treatment due to an AE
	• Safety (AEs)
	
	Relevant to this review but not reported: SAEs, QoL, TFST
	
	Other outcomes (not relevant to this review): -

Notes	**Funding sources:** Novartis Pharmaceuticals and in part by the Cancer Center Support Grant (NCI Grant P30 CA016672) **Declarations of Interests:** Quote;"Nizar M. Tannir certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (e.g. employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: Tannir has received grant funding and personal fees for consultancy from Pfizer and Novartis. Karam has received personal fees for consultancy from Pfizer. Wood has received grant funding from Pfizer. Zurita has received grant funding and personal fees for consultancy from Pfizer, and grant funding from Novartis." **Clinical study report available**: no **Study protocol available:** no **Statistical analysis plan available:** no

NCT01481870

Study characteristics

NCT01481870 *(Continued)*

Methods	**Study name:** CROSS-J-RCC **Study design:** randomised, phase III **Blinding:** none, open- label **Study dates:** February 2010 - July 2012 **Date of data cut-off:** June 30, 2015 **Location:** 1 country (Japan), types of centres: unclear, according to the text: 39 institutions: according to CT.gov: 1 location **Crossing-over study or cross over permitted:** yes, cross-over study* *For cross-over studies, only data on period 1 were extracted for analyses.

Participants	**Inclusion criteria:**
	
	• age: 20-80 years old, both inclusive
	• all sexes
	• ECOG performance status of 0, 1, or 2
	• MSKCC risk of favourable or intermediate
	• histologically confirmed renal cell carcinoma
	• no ischaemic heart disease
	
	Exclusion criteria:
	
	• history of any other malignancy
	• central nervous system metastases. However, participants who remain asymptomatic, have no new or enlarging lesion in the CNS within 6 months of enrolment in this study, and require no corticosteroids may be enrolled
	• prior treatment with sunitinib or sorafenib
	• pregnancy or possible pregnancy at any time during the study
	
	Sample size: N = 124, 120 participants were evaluated
	
	Age (years, median with range): experimental arm: 67 (41-79), control arm: 66 (44-79)
	
	Sex (m/f): experimental arm: 46/11, control arm: 53/10
	
	Prognostic factors:
	
	• **MSKCC risk group**
	• **Favourable**
	  ◦ experimental arm: 12; control arm: 14
	• **Intermediate**
	  ◦ experimental arm: 45; control arm: 49

Interventions	**Experimental arm (n = 60):** sunitinib (50 mg, oral, once/day) - treatment was 4 weeks on treatment, 2 weeks off treatment (1 cycle = 6 weeks) **Control arm (n = 64):** sorafenib (400 mg, oral, twice/week) - no breaks
Outcomes	**Primary outcome(s)** • PFS in first-line treatment ◦ Time frame: time of progression in first line treatment **Secondary outcome(s)** • Total PFS in first-line and second-line treatments

NCT01481870 *(Continued)*

- Time frame: time of progression in second-line treatment
- OS
- AEs
- Number of participant who discontinued treatment due to an AE

Relevant to this review but not reported: QoL, SAEs TFST

Other outcomes (not relevant to this review): -

Notes

Funding sources: Quote: "The present study was supported in part by the Japanese Urological Research Network and Pfizer."

Declaration of Interest: "Y.T. has received grants and lecture and advisory fees from Novartis, Japan; Ono, and Astellas, grants and lecture fees from Astellas and Pfizer, Japan lecture fees from Bristol-Myers Squibb, Japan and grants from Takeda, Japan. T. Kondo has received lecture fees from Intuitive Surgical, Novartis, Ono, and Pfizer. W.O. has received grants form Nipuro and Takeda and lecture fees from Astellas, Bristol-Myers Squibb, Ono, and Pfizer. J.T. has received lecture fees from Pfizer. M.T. has received lecture fees from Pfizer.
H.M. has received grants and lecture fees from Janssen, grants from Kyowa-Hakko-Kirin, Japan;Merck biopharma, Japan; Sharp & Dohme, Takeda Terumo, and Toyo Steel. R.S. has received lecture fees from Intuitive Surgical and Novartis and a grant from Astellas. T.S. has received grants and lecture fees from Astellas and Takeda and lecture fees from Novartis. N. Shinohara has received grants and lecture fees from Astellas and Ono, lecture fees from Bayer, and lecture fees from Bayer, GlaxoSmithKline, Pfizer, and Takeda. M.O. has received grants and lecture fees from Pfizer and lecture fees from Bayer, Bristol-Myers Squibb, Novartis, and Ono. H.K. has received grant and lecture fees from Pfizer and grants from Astellas, Ono, Sanofi, Japan; Taiho, and Takeda. The remaining authors declare that they have no competing interests."

Clinical study report available: no

Study protocol available: no

Statistical analysis plan available: no

NCT01613846

Study characteristics

Methods

Study name: SWITCH II

Study design: a two-arm, RCT, phase III

Blinding: none, open-label

Study dates: June 14th, 2012 - November 14th, 2016

Date of data cut-off: not reported

Location: 3 countries (Austria, Germany, the Netherlands), types of centres: cancer centres/clinics, hospitals, university hospitals (72 study locations)

Cross-over study or cross over permitted: yes, cross-over study*

*For cross-over trials, we extracted data on period 1 only.

Participants

Inclusion criteria:

- all sexes
- age ≥ 18 and ≤ 85 years

NCT01613846 *(Continued)*

- participants with metastatic / advanced RCC (all histologies), who are not suitable for cytokine therapy and for whom study medication constitutes first-line treatment
- non-clear cell histology RCC
- intermediate risk according to MSKCC score
- ECOG ≥ 1 and> 1 organ metastasis + < 24 months between diagnosis and establishing indication for interleukin-2-therapy
- ECOG ≥ 1 and "unable to carry on normal activity or do active work" (Karnofsky Index 70%)
- Karnofsky Index ≥ 70%
- MSKCC prognostic score (2004), low or intermediate

Exclusion criteria:

- major surgery within 4 weeks of start of study
- prior exposure to study drugs
- investigational drug therapy within 4 weeks of study entry
- radiotherapy within 3 weeks of start of study drug and planned radiotherapy during the study
- concomitant medication: Any condition at the discretion of the investigator that precludes compliance with concomitant therapy restrictions described below

Sample size: N = 377

Age, median (years (range)): experimental arm: 68 (31-84), control arm: 68 (26-86)

Sex (m/f, %): experimental arm: 136/53 (72/28), control arm: 137/51 (73/27)

Prognostic factors:

- **KPS, n(%)**
  - **100**
    - experimental arm: 96 (51), control arm: 85 (45)
  - **90**
    - experimental arm: 32 (17), control arm: 46 (25)
  - **80**
    - experimental arm: 52 (27), control arm: 44 (23)
  - **70**
    - experimental arm: 9 (5), control arm: 12 (6)
  - **Missing**
    - experimental arm: 0 (0), control arm: 1 (1)
- **MSKCC risk score, n(%)**
  - **Low**
    - experimental arm: 95 (50), control arm: 91 (48)
  - **Intermediate**
    - experimental arm: 90 (48), control arm: 89 (47)
  - **High**
    - experimental arm: 4 (2), control arm: 5 (3)
  - **Missing/Unknown**
    - experimental arm: 0 (0), control arm: 3 (2)
- **Nephrectomy n(%)**
  - **Total**
    - experimental arm: 167 (88), control arm: 161 (86)
  - **Partial**
    - experimental arm: 19 (10), control arm: 24 (13)

Interventions	**Experimental arm (n = 189):** sorafenib (400 mg, oral, twice/day) **Control arm (n = 188):** pazopanib (800 mg, oral, once/day)
Outcomes	**Primary outcome(s)**

NCT01613846 *(Continued)*

-

Secondary outcome(s)

- PFS in first-line, descriptively
  - Time frame: 4 years
- OS, descriptively (data cut-off same as for primary endpoint)
  - Time frame: 4 years
- Health-related QoL (FACIT-F, FKSI-10)
  - Time frame: 4 years
- Safety and tolerability (AEs, SAEs)
  - Time frame: 4 years
- Number of participants who discontinued treatment due to an AE

Relevant to this review but not reported: TFST

Other outcomes (not relevant to this review): PFS in second-line treatment, DCR (CR,PR,SD, RECIST), biomarker programme, time to treatment failure, TTP

Notes

Funding sources: Quote: "The SWITCH II trial was sponsored by the Technical University of Munich, Germany with financial grants from Bayer HealthCare GmbH and Novartis GmbH. Bayer HealthCare GmbH and Novartis GmbH were not involved in the trial concept and design. The Association for Urologic Oncology (AUO) of the German Cancer Society supported this study (AN 33/11) as well as the main Association of Austrian Social Security Institutions. The specific role of the sponsors was in the design and conduct of the study."

Declaration of Interest: Quote:"It is certified that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (e.g. employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or
patents filed, received, or pending), are the following:" Margitta Retz, Janssen-Cilag, Martin Bögemann, Marc-Oliver Grimm, Maria De Santis, Christian Bolenz, Carsten Bokemeyer, Jürgen E. Geschwend. No conflicts of Interest declare: Uwe Zimmermann, Lothar Müller, Christian Leiber, Dogu Teber, Manfred Wirth, Aart Becker, Jan Lehmann, Robbert van Alphen, Martin Indorf, Melanie Frank.

Clinical study report available: no

Study protocol available: no

Statistical analysis plan available: no

NCT01835158

Study characteristics

Methods

Study name: CABOSUN

Study design: a two-arm, RCT, phase II

Blinding: none, open-label

Study dates: July 9, 2013 to April 6, 2015 (date of randomisation)

Date of data cut-off: July 01, 2017 (for OS), September 15, 2016 (for PFS per IRC)

Location: 1 country (USA), types of centres: cancer centres/clinics, medical centres, hospitals (488 study locations)

Cross-over study or cross over permitted: no

NCT01835158 *(Continued)*

Participants	**Inclusion criteria:**
	• 18 years and older
	• all sexes
	• renal cell carcinoma with some component of clear cell histology; histologic documentation of metastatic disease is not required
	• locally advanced (defined as disease not amenable to curative surgery or radiation therapy) or metastatic RCC (equivalent to stage IV RCC, according to AJCC staging)
	• eligible participants must be intermediate/poor risk, per the International mRCC Database Consortium (Heng) criteria; participants must therefore have as one or more of the following six factors:
	1. Time from diagnosis of RCC to systematic treatment <1 year
	2. Haemoglobin < the lower limit of normal (ULN)
	3. Corrected calcium > the upper limit of normal (ULN)
	4. Karnosfky performance status < 80%
	5. Neutrophil count > ULN
	6. Latelet count > ULN
	• no prior systemic treatment for RCC; supportive therapies such as bisphosphonates (zoledronic acid) or denosumab are permitted
	• participants must have measurable disease by RECIST criteria; lesions that can be accurately measured in at least one dimension (longest diameter to be recorded) as >= 2 cm with conventional techniques or as >= 1 cm with spiral CT scan
	• performance status: ECOG 0-2
	
	Sample size: N = 157
	
	Age (years, median with range): experimental arm: 63 (56-69) , control arm: 64 (57-71)
	
	Sex (m/f): experimental arm: 66/13, control arm: 57/21
	
	Prognostic factors:
	• **ECOG Performance Status, n(%)**
	  ◦ **0**
	  ▪ experimental arm: 36 (46), control arm: 36 (46)
	  ◦ **1**
	  ▪ experimental arm: 33 (42), control arm: 32 (41)
	  ◦ **2**
	  ▪ experimental arm: 10 (13), control arm: 10 (13)
	• **IMDC Risk Group**
	  ◦ **Intermediate**
	  ▪ experimental arm: 64 (81), control arm: 63 (81)
	  ◦ **Poor**
	  ▪ experimental arm: 15 (19), control arm: 15 (19)
	  ◦ **Prior nephrectomy**
	  ▪ **Yes**
	  ▪ experimental arm: 57 (72), control arm: 60 (77)

Interventions	**Experimental arm (n = 79):** cabozantinib (60 mg, oral, once/day) **Control arm (n = 78):** sunitinib (50 mg, oral, once/day) A treatment cycle was defined as 6 weeks in both study groups (4 weeks on treatment, 2 weeks off).
Outcomes	**Primary outcome(s)** • PFS ◦ Time frame: up to 5 years • OS

NCT01835158 *(Continued)*

- Time frame: up to 5 years

Secondary outcome(s)

- Safety (AEs/SAEs)
- Number of participants who discontinued treatment due to an AE

Relevant to this review but not reported: QoL, TFST

Other outcomes (not relevant to this review): ORR

Notes	**Funding sources:**Quote: "The study was designed by the Alliance for Clinical Trials in Oncology, endorsed by the ECOGeAmerican College of Radiology Imaging Network Group and approved by the Cancer Therapy Evaluation Program of the National Cancer Institute part of the National Institutes of Health (the funder)." **Declaration of Interest:** "TKC reports personal fees for an advisory/consulting role from Pfizer, GlaxoSmithKline, Novartis, Merck, Bayer, Eisai, Roche, Prometheus Labs Inc., Foundation Medicine Inc., Bristol-Myers Squibb, and research funding from Pfizer, GlaxoSmithKline, Novartis, Bristol-Myers Squibb, Merck, Exelixis Inc., Roche, AstraZeneca, Tracon and Peloton. MDM reports attendance at advisory boards for Pfizer and Exelixis, Inc., outside the submitted work. OH reports relevant financial activities outside the submitted work, and participation at an advisory board for Pfizer. MJM reports attendance at advisory boards for Bayer, Astellas and Progenics, personal fees and research support from Progenics, and research support from Endocyte, outside the submitted work. DRF reports research support from Seattle Genetics and Novartis, outside the submitted work. DG reports personal fees from Dendreon, Novartis, Sanofi, Bayer, Medivation, Biopharm, Axess Oncology, Exelixis, Inc., Pfizer, GlaxoSmithKline, Astellas Pharma, Innocrin Pharma, Bristol-Myers Squibb, Genentech, Janssen, Acceleron Pharma, Celgene,Merk Sharp & Dohme, and Myovant Sciences, Inc, and research funding from Dendreon, Novartis, Bayer, Exelixis, Inc., Pfizer, Astellas Pharma, Innocrin Pharma, Bristol-Myers Squibb, Genentech, Janssen, Millennium, Acerta Pharma, outside the submitted work. CH, MM and CS are the employees of Exelixis, Inc. All other authors declare no competing interests." **Clinical study report available**: no **Study protocol available:** yes **Statistical analysis plan available:** no

NCT01984242

Study characteristics	
Methods	**Study name:** IMmotion150 **Study design:** a two-arm, RCT, phase II (three-arm trial) **Blinding:** none, open-label **Study dates:** 8 January 2014 to 16 March 2015 (date of enrolment) **Date of data cut-off:** 17 October 2016 (clinical cutoff date) **Location:** 9 countries (Czech Republic, France, Germany, Italy, Poland, Romania, Spain, UK, USA), types of centres: cancer centres, medical centres, hospitals, research institutions (45 study locations) **Cross-over study or cross over permitted:** not per design, but participants enrolled in atezolizumab (except EU participants) or sunitinib group could crossover to receive atezolizumab and bevacizumab combination therapy in case of disease progression
Participants	**Inclusion criteria:**

NCT01984242 *(Continued)*

- all sexes
- 18 years and older
- unresectable advanced or metastatic renal cell carcinoma with component of clear cell histology and/or component of sarcomatoid histology that has not been previously treated with any systemic agents, including treatment in the adjuvant setting
- measurable disease, as defined by RECIST v1.1
- KPS (>/=) 70

Exclusion criteria:

Disease-Specific Exclusions:

- radiotherapy for renal cell carcinoma within 14 days prior to Cycle 1, Day 1 with the exception of single-fraction radiotherapy given for the indication of pain control
- known active malignancies or metastasis of the brain or spinal cord or leptomeningeal disease, as determined by computed tomography (CT) or magnetic resonance imaging (MRI) evaluation during screening and prior radiographic assessments
- malignancies other than renal cell carcinoma within 5 years prior to Cycle 1, Day 1, with the exception of those with a negligible risk of metastasis or death, treated with expected curative outcome

Sample size: N=305

Age, median (years, range): experimental arm I: 62 (32-88), experimental arm II: 61 (27-81), control arm: 61 (25-85)

Sex (m/f): experimental arm I: 74/27, experimental arm II: 77/26, control arm: 79/22

Prognostic factors:

- **MSKCC risk category, n(%)**
  - **Favourable (0)**
    - experimental arm I: 30 (30), experimental arm II: 26 (25), control arm: 21 (21)
  - **Intermediate (1 or 2)**
    - experimental arm I: 62 (61), experimental arm II: 69 (67), control arm: 70 (69)
  - **Poor (≥3)**
    - experimental arm I: 9 (9), experimental arm II: 8 (8), control arm: 10 (10)
- **Prior nephrectomy**
  - **Yes**
    - experimental arm I: 88 (87), experimental arm II: 88 (87), control arm: 89 (86)

Interventions	**Experimental arm I (n = 101):** atezolizumab (1200 mg, intravenous, every three weeks) + Bevacizumab (15 mg/kg, intravenous, every three weeks) **Experimental arm II (n = 103):** atezolizumab (1200 mg, intravenous, every three weeks) **Control arm (n = 101):** sunitinib (50 mg, oral, once/day) - treatment was 4 weeks on treatment, 2 weeks off treatment (1 cycle - 6 weeks)

Outcomes	**Primary outcome(s)**
	
	• PFS per RECIST v1.1 via IRC assessment in ITT population
	  ○ Time frame: from randomisation until disease progression or death due to any cause (until data cut-off date 17 October 2016, up to approximately 2.75 years)
	
	Secondary outcome(s)
	
	• OS in ITT population
	  ○ Time frame: randomisation until death due to any cause (until data cut-off date 17 October 2016, up to approximately 2.75 years)
	• AEs, SAEs
	  ○ Time frame: baseline up to approximately 60 months

NCT01984242 *(Continued)*

	• Number of participants who discontinued treatment due to an AE **Relevant to this review but not reported:** TFST, QoL **Other outcomes (not relevant to this review):** OS and PFS in different subgroups, pharmacokinetics of study drugs, laboratory parameters, disease progression, DoR,OR (CR, PR), number of deaths, DP
Notes	**Funding sources:** Quote; Prometheus Laboratories. M.B.A., Roche/Genentech, Novartis, Pfizer, Eisai, and Exelixis, F. Hoffmann-La Roche, AG (...) **Declaration of Interest:** "D.F.M. reports a consulting/advisory role for Bristol-Myers Squibb, Merck, Roche/ Genentech, Pfizer, Exelixis, Novartis, Eisai, X4 Pharmaceuticals, and Array BioPharma (...). J.A.R., J.H., T.H., C. Suárez, and R.D. have nothing to disclose." **Clinical study report available**: no **Study protocol available:** no **Statistical analysis plan available:** no

NCT02231749

Study characteristics	
Methods	**Study name:** Checkmate 214 **Study design:** a two-arm RCT, phase III **Blinding:** none, open-label **Study dates:** October 16, 2014 - February 23, 2016 (date of randomisation) **Date of data cut-off:** For OS and PFS: 25 February 2020 (Albiges 2020); for QoL: August 7, 2017 (Cella 2019) **Location:** 28 countries (Argentina, Australia, Austria, Belgium, Brazil, Canada, Chile, Colombia, Czech Republic, Denmark, Finland, France, Germany, Hungary, Ireland, Israel, Italy, Japan, Republic of Korea, Mexico, the Netherlands, Poland, Spain, Sweden, Taiwan, Turkey, UK USA), types of centres: not reported (190 study locations) **Cross-over study or cross over permitted:** not per design, but cross over was permitted from the control arm to the experimental arm for intermediate and poor risk participants
Participants	**Inclusion criteria:** • all sexes • 18 years and older • histological confirmation of RCC with a clear-cell component • advanced (not amenable to curative surgery or radiation therapy) or metastatic (AJCC Stage IV) RCC • no prior systematic therapy for RCC with the following exception: ◦ One prior adjuvant or neoadjuvant therapy for completely resectable RCC if such therapy did not include an agent that targets VEGF or VEGF receptors and if recurrence occurred at least 6 months after the last dose of adjuvant or neoadjuvant therapy • KPS of at least 70% • measurable disease as per RECIST 1.1 **Exclusion criteria:** • prior systemic treatment with VEGF or VEGF receptor targeted therapy (including, but not limited to, Sunitinib, Pazopanib, Axitinib, Tivozanib, and Bevacizumab)

NCT02231749 *(Continued)*

- prior treatment with an anti-programmed death (PD)-1, anti-PD-L1, anti-PD-L2, anti-CD137, or anti-cytotoxic T-lymphocyte antigen 4 (CTLA-4) antibody, or any other antibody or drug specifically targeting T-cell co-stimulation or checkpoint pathways

Sample size: N=1096

Age, median (years, range): experimental arm: 62 (26-85), control arm: 62 (21-85)

Sex (m/f): experimental arm: 413/137, control arm: 395/151

Prognostic factors:

- **IMDC risk group, n(%)**
  - **Favourable**
    - experimental arm: 125 (23), control arm: 124 (23)
  - **Intermediate**
    - experimental arm: 334 (61), control arm: 333 (61)
  - **Poor**
    - experimental arm: 91 (17), control arm: 89 (16)
- **Previous nephrectomy, n(%)**
  - **Yes**
    - experimental arm: 453 (82), control arm: 437 (80)

Interventions

Experimental arm (n = 550): nivolumab (3 mg/kg, intravenous) + ipilimumab (1 mg/kg, intravenous), every 3 weeks for four doses (induction phase) followed by nivolumab montherapy (maintenance therapy)

Control arm (n = 546): sunitinib (50 mg, oral, once/day) - treatment was 4 weeks on treatment, 2 weeks off treatment (1 cycle = 6 weeks)

Outcomes

Primary outcome(s)

- OS in intermediate/poor-risk participants with previously untreated mRCC
  - Time frame: From the date of randomisation to the date of death (Approximately 31 months)
- PFS in intermediate/poor-risk participants with previously untreated mRCC
  - Time frame: approximately 31 months (from date of first dose to date of documented disease progression or death due to any cause, whichever occurs first)

Secondary outcome(s)

- OS in any risk participants with previously untreated mRCC
  - Time frame: approximately 31 months (from the date of randomisation to the date of death)
- PFS in any risk participants with previously untreated mRCC
  - Time frame: approx. 31 months (from date of first dose to date of documented disease progression or death due to any cause, whichever occurs first)
- AEs
  - Time frame: approx. 31 months (from first dose to 30 days after last dose of study therapy)
- SAE
  - Time frame: approx. 31 months (from first dose to 30 days after last dose of study therapy)
- Other (not including serious) AEs
  - Time frame: approx. 31 months (from first dose to 30 days after last dose of study therapy)
- Number of participants who discontinued treatment
- QoL, measured with FACT-G and EQ-5D VAS

Relevant to this review but not reported: QoL

Other outcomes (not relevant to this review): ORR

Notes

Funding sources: Bristol-Myers Squibb and ONO Pharmaceutical and grants and personal fees from Pfizer, Novartis, Eisai, Exelixis, and Genentech/Roche (...).

NCT02231749 *(Continued)*

Declarations of Interests: "PS, VN, and BR declare no competing interests." For a more detailed description please refer to the publication.

Clinical study report available: no

Study protocol available: yes

Statistical analysis plan available: yes

NCT02420821

Study characteristics

Methods

Study name: IMmotion 151

Study design: a two-arm, RCT, phase III

Blinding: none, open-label

Study dates: May 20, 2015 - Oct 12, 2016 (date of enrolment)

Date of data cut-off: September 29, 2017 (PFS; first interim analysis); August 13, 2018 (second interim analysis); safety and OS data were updated at the cutoff date February 14, 2020 (final analysis)

Location: 21 countries (Australia, Bosnia and Herzegovina, Brazil, Canada, Czech Republic, Denmark, France, Germany, Italy, Japan, Republic of Korea, Mexico, Poland, Russian Federation, Singapore, Spain, Taiwan, Thailand, Turkey, UK USA), types of centres: cancer centres/institutes, medical centres, hospitals, university hospitals (154 study locations)

Cross-over study or cross over permitted: no

Participants

Inclusion criteria:

- all sexes
- 18 years and older
- definitive diagnosis of unresectable locally advanced or metastatic RCC with clear-cell histology and/or a component of sarcomatoid carcinoma, with no prior treatment in the metastatic setting
- evaluable MSKCC risk score
- measurable disease, as defined by RECIST v1.1
- KPS greater than or equal to 70%

Exclusion criteria:

Disease-specific exclusions:

- radiotherapy for RCC within 14 days prior to treatment
- active central nervous system disease

General medical exclusions:

- life expectancy less than 12 weeks
- participation in another experimental drug study within 4 weeks prior to treatment

Sample size: N=915

Age, median (years, range): experimental arm: 62 (56-69), control arm: 60 (54-66)

Sex (m/f): experimental arm: 317/137, control arm: 352/109

Prognostic factors:

NCT02420821 *(Continued)*

- **MSKCC risk score, n (%)**
  - **Favourable (0)**
    - experimental arm: 89 (20), control arm: 90 (20)
  - **Intermediate (1 or 2)**
    - experimental arm: 311 (69), control arm: 318 (69)
  - **Poor (≥3)**
    - experimental arm: 54 (12%) control arm: 53 (12%)
- **Previous nephrectomy, n(%)**
  - **Yes**
    - experimental arm: 334 (74), control arm: 330 (72)

Interventions	**Experimental arm (n = 454):** atezolizumab (1200 mg, intravenous) + bevacizumab (15 mg/kg, intravenous), once every 3 weeks **Control arm (n= 461):** sunitinib (50 mg, oral, once/day) - treatment was 4 weeks on treatment, 2 weeks off treatment (1 cycle = 6 weeks)

Outcomes	**Primary outcome(s)**
	
	• OS in ITT Population
	  ○ Time frame: baseline until death from any cause (until data cut-off date 29 September 2017, up to approximately 27 months)
	
	Secondary outcome(s)
	
	• PFS as determined by an IRC according to RECIST v1.1 in ITT population
	  ○ Time frame: baseline until documented PD or death, whichever occurred first (until data cut-off date 29 September 2017, up to approximately 24 months)
	• SAEs
	  ○ Time frame: baseline up to data cut-off date 29 September 2017(overall approximately 27 months)
	• Other (not including serious) AEs
	  ○ Time frame: baseline up to data cut-off date 29 September 2017(overall approximately 27 months)
	• Number of participants who discontinued due to an AE
	
	Relevant to this review but not reported: QoL, TFST
	
	Other outcomes (not relevant to this review): OS and PFS in different subgroups, pharmacokinetics, laboratory parameters (ATA), PD, DoR (OR, CR, PR), PD

Notes	**Funding sources:** Hoffmann–La Roche Ltd and Genentech Inc. **Declaration of Interest:** quote: "CSu, FP, and BMell have nothing to disclose." For a detailed description please refer to publication. **Clinical study report available**: no **Study protocol available:** yes **Statistical analysis plan available:** yes

NCT02684006

Study characteristics	
Methods	**Study name:** JAVELIN **Study design:** a two-arm, RCT, phase III **Blinding:** none, open-label

NCT02684006 *(Continued)*

Study dates: March 29, 2016 - December 19, 2017 (date of randomisation)

Date of data cut-off: June 20, 2018 (for safety); exact dates for OS and PFS not reported, but data for PFS was reported and extracted from the second interim analysis for OS, and OS data were reported/extracted from the third interim analysis (longest follow-up available for both outcomes)

Location: 21 countries (Australia, Austria, Belgium, Canada, Denmark, France, Germany, Hungary, Israel, Italy, Japan, Republic of Korea, Mexico, the Netherlands, New Zealand, Romania, Russian Federation, Spain, Sweden, UK, USA), types of centres: hospitals, university hospitals, medical centres, centres/institutes (280 study locations)

Cross-over study or cross over permitted: no

Participants

Inclusion criteria:

- all sexes
- 18 years and older
- histologically or cytologically confirmed advanced or metastatic RCC with clear cell component
- at least one measurable lesion as defined by RECIST version 1.1 that has not been previously irradiated
- ECOG performance status 0 or 1

Exclusion criteria:

- prior systemic therapy directed at advanced or metastatic RCC
- prior adjuvant or neoadjuvant therapy for RCC if disease progression or relapse has occurred during or within 12 months after the last dose of treatment
- prior therapy with axitinib and/or sunitinib as well as any prior therapies with other VEGF pathway inhibitors
- newly diagnosed or active brain metastasis

Sample size: N = 886

Age, median (years, range): experimental arm: 62 (29-83),control arm: 61 (27-88)

Sex (m/f): experimental arm: 316/126 (71.5/28.5), control arm: 344/100 (77.5/22.5)

Prognostic factors:

- **MSKCC prognostic risk group, n(%)**
  - **Favourable**
    - experimental arm: 96 (21.7), control arm: 100 (22.5)
  - **Intermediate**
    - experimental arm: 283 (64.0), control arm: 293 (66.0)
  - **Poor**
    - experimental arm: 51 (11.5), control arm: 45 (10.1)
  - **Not reported**
    - experimental arm: 12 (2.7), control arm: 6 (1.4)
- **IMDC prognostic risk group, n(%)**
  - **Favourable**
    - experimental arm: 94 (21.3), control arm: 96 (21.6)
  - **Intermediate**
    - experimental arm: 271 (61.3), control arm: 276 (62.2)
  - **Poor**
    - experimental arm: 72 (16.3), control arm: 71 (16.0)
  - **Not reported**
    - experimental arm: 5 (1.1), control arm: 1 (0.2)
- **Previous nephrectomy, n(%)**
  - **Yes**
    - experimental arm: 352 (79.6), control arm: 355 (80.0)

NCT02684006 *(Continued)*

Interventions	**Experimental arm (n = 442):** avelumab (10 mg/kg, intravenous, every 2 weeks) + axitinib (mg, oral, twice/day) **Control arm (n = 444):** sunitinib (50 mg, oral, once/day) - treatment was 4 weeks on, 2 weeks off (1 cycle = 6 weeks)

Outcomes	**Primary outcome(s)**
	
	-
	
	Secondary outcome(s)
	
	• OS
	  ○ Time frame: every 3 months up to 8 years
	• PFS by investigator assessment
	  ○ Time frame: every 6 weeks up to 18 months from patient enrolment in the study, then every 12 weeks up to 40 months from randomisation
	• PFS
	  ○ Time frame: from randomisation up to 40 months
	• Treatment discontinuation/failure due to toxicity
	  ○ Time frame: from Cycle 1 Day 1, every 6 weeks up to the end of treatment
	
	Relevant to this review but not reported: QoL
	
	Other outcomes (not relevant to this review): OS and PFS in different subgroups, TTF, VAS, DR, TTR, DC, OR, biomarker status, laboratory parameters (ADA), pharmacokinetics

Notes	**Funding sources:** Pfizer and Merck **Declaration of Interest**: Disclosure forms provided by the authors are available at NEJM.org. **Clinical study report available**: no **Study protocol available:** yes **Statistical analysis plan available:** yes

NCT02761057

Study characteristics	
Methods	**Study name:** SWOG 1500 **Study design:** RCT, phase II **Blinding:** none, open-label **Study dates:** April 5, 2016 - Dec 15, 2019 (date of randomisation) **Date of data cut-off:** October 16, 2020 (for first analysis); exact date for updated analyses not reported **Location:** 2 countries (Canada, U.S.A.), types of centres: cancer centres, medical centres, hospitals, (597 study locations) **Cross-over study or cross over permitted:** not per design; not reported whether cross over was allowed at some point (e.g. at progression)
Participants	**Inclusion / exclusion criteria:**

NCT02761057 *(Continued)*

- Patients must have histologically or cytologically confirmed papillary renal cell carcinoma which is metastatic or locally advanced disease not amenable to surgical resection
- Patients must also have measurable disease
- Patients with a history of treated brain metastases who are asymptomatic and have not received steroid therapy in the 14 days prior to registration are eligible; anti-seizure medications are allowed provided they are non-enzyme
- Patients may have received prior surgery; at least 28 days must have elapsed since surgery and patient must have recovered from any adverse effects of surgery
- Patients may have received up to one prior systemic therapy* for advanced or metastatic renal cell carcinoma with the exception of another VEGF inhibitor Food and Drug Administration (FDA)-approved for advanced RCC (i.e., pazopanib, bevacizumab, sorafenib or axitinib); if a patient develops metastatic disease within six months of discontinuation of adjuvant therapy, this will constitute one prior systemic therapy for advanced or metastatic renal cell carcinoma (RCC); if a patient develops metastatic disease and more than six months has elapsed since discontinuation of adjuvant therapy, this will not constitute prior systemic therapy for advanced or metastatic RCC; patients may have also received prior immunotherapy; patients must not have received a MET/hepatocyte growth factor (HGF) inhibitor or sunitinib as prior therapy; at least 14 days must have elapsed since completion of prior systemic therapy; patients must have recovered from all associated toxicities at the time of registration
- Patients may have received prior radiation therapy, but must have measurable disease outside the radiation port; at least 14 days must have elapsed since completion of prior radiation therapy; patients must have recovered from all associated toxicities at the time of registration

More inclusion & exclusion criteria on CT.gov.

Sample size: N=147

Age, median (years, range): 66 (58-75) (across all participants)

Sex (m/f): 112 females; 35 males (across all participants)

Prognostic factors:

- **IMDC prognostic risk group, n(%)**
  - **Favourable**
    - experimental arm I: 10 (23), experimental arm II: 8 (29), experimental arm III: 6 (21) control arm: 14 (30)
  - **Intermediate**
    - experimental arm I: 28 (64), experimental arm II: 16 (57), experimental arm III: 19 (66), control arm: 26 (75)
  - **Poor**
    - experimental arm I: 6 (14), experimental arm II: 4 (14), experimental arm III: 4 (14), control arm: 6 (13)
- **Previous nephrectomy, n(%)**
  - **Yes**
    - experimental arm: 32 (73), experimental arm II: 26 (93), experimental arm III: 21 (72), control arm: 34 (77)

Interventions	**Experimentarl arm I (N = 46):** sunitinib (50 mg, oral, once/day) - treatment was 4 weeks on treatment, 2 weeks off treatment (1 cycle = 6 weeks) **Experimental arm II (N = 44):** cabozantinib (60 mg, oral, once/day) **Experimental arm III (N = 28):** crizotinib (250 mg, oral, twice/day) **Experimental arm IIII (N = 29):** savolitinib (600 mg, oral, once/day)
Outcomes	**Primary outcome(s)** • PFS **Secondary outcome(s)**

NCT02761057 *(Continued)*

- Toxicity
- OS
- Number of participants who discontinued study drug due to an AE

Relevant to this review but not reported: QoL, TFST

Other outcomes (not relevant to this review): ORR

Notes

Funding sources: National Institutes of Health and National Cancer Institute.

Declaration of Interest: yes

Clinical study report available: no

Study protocol available: yes

Statistical analysis plan available:

**Other: ***7% of the study population received one prior line of systemic therapy (excluding VEGF-directed or MET-directed drugs).

NCT02811861

Study characteristics

Methods

Study name: CLEAR

Study design: RCT, phase III (three-arm trial)

Blinding: none, open-label

Study dates: October 13, 2007 - July 24, 2019 (date of randomisation)

Date of data cut-off: August 28, 2020 (for final analysis of PFS and interim analysis of OS)

Location: 20 countries (Australia, Austria, Belgium, Canada, Czech Republic, France, Germany, Greece, Ireland, Israel, Italy, Japan, Republic of Korea, the Netherlands, Poland, Russian Federation, Spain, Switzerland, UK USA), types of centres: hospitals, university hospitals, medial centres, cancer centres (183 study locations)

Cross-over study or cross over permitted: not a cross-over study; not reported whether cross over was permitted at some point (e.g. after progression)

Participants

Inclusion criteria:

- histological or cytological confirmation of renal cell carcinoma (RCC) with a clear-cell component
- at least 1 measurable target lesion according to Response Evaluation in Solid Tumors (RECIST) 1.1
- Karnofsky Performance Status (KPS) of ≥70
- adequately controlled blood pressure (BP) with or without antihypertensive medications, defined as BP ≤150/90 mmHg at Screening and no change in antihypertensive medications within 1 week prior to Cycle 1/Day 1 (C1/D1)
- adequate organ function per blood work

Exclusion criteria:

- participants who have received any systemic anticancer therapy for RCC, including anti-vascular endothelial growth factor (VEGF) therapy, or any systemic investigational anticancer agent
- participants with central nervous system (CNS) metastases are not eligible, unless they have completed local therapy (e.g. whole brain radiation therapy (WBRT), surgery or radiosurgery) and have discontinued the use of corticosteroids for this indication for at least 4 weeks before starting treatment

NCT02811861 *(Continued)*

in this study. Any signs (e.g. radiologic) or symptoms of CNS metastases must be stable for at least 4 weeks before starting study treatment

- active malignancy (except for RCC, definitively treated basal or squamous cell carcinoma of the skin, and carcinoma in-situ of the cervix or bladder) within the past 24 months. Participants with history of localised and low risk prostate cancer are allowed in the study if they were treated with curative intent and there is no prostate specific antigen (PSA) recurrence within the past 5 years
- prior radiation therapy within 21 days prior to start of study treatment with the exception of palliative radiotherapy to bone lesions, which is allowed if completed 2 weeks prior to study treatment start

More exclusion criteria on CT.gov.

Sample size: N = 1069

Age, median (years, range): experimental arm I: 62 (32-86), experimental arm II: 64 (34-88), control arm: 61 (29-82)

Sex (m/f): experimental arm I: 266/91, experimental arm II: 255/100, control arm: 275/82

Prognostic factors:

- **MSKCC prognostic risk group, n(%)**
  - **Favourable**
    - experimental arm I: 98 (27.5), experimental arm II: 96 (27.0) , control arm: 97 (27.2)
  - **Intermediate**
    - experimental arm I: 227 (63.9), experimental arm II: 227 (63.9), control arm: 228 (63.9)
  - **Poor**
    - experimental arm I: 42 (11.8), experimental arm II: 33 (9.3), control arm: 37 (10.4)
  - **Could not be evaluated**
    - experimental arm I: 6 (1.7), experimental arm II: 2 (0.6), control arm: 4 (1.1)
- **IMDC prognostic risk group, n(%)**
  - **Favourable**
    - experimental arm I: 114 (31.9), experimental arm II: 110 (31.0), control arm: 124 (34.7)
  - **Intermediate**
    - experimental arm I: 195 (54.6), experimental arm II: 210 (59.2) , control arm: 192 (53.8)
  - **Poor**
    - experimental arm I: 42 (11.8), experimental arm II: 33 (9.3), control arm: 37 (10.4)
  - **Could not be evaluated**
    - experimental arm I: 6 (1.7), experimental arm II: 2 (0.6), control arm: 4 (1.1)
- **Previous nephrectomy, n(%)**
  - **Yes**
    - experimental arm: 260 (72.8), experimental arm II: 262 (73.8), control arm: 275 (77.0)

Interventions	**Experimental arm I (n = 355):** lenvatinib (18 mg, oral, once/day) + Everolimus (5mg, oral, once/day) **Experimental arm II (n = 352):** lenvatinib (20 mg, oral, once/day), Pembrolizumab (200mg, intravenous, every 3 weeks) **Control arm (n=340):** Sunitinib (50 mg, oral, once/day) - treatment was 4 weeks on treatment, 2 weeks off treatment (1 cycle = 6 weeks)
Outcomes	**Primary outcome(s)** • PFS by independent review (time frame: up to 47 months approximately) **Secondary outcome(s)** • OS (time frame: up to 67 months approximately) • Number of TEAEs and SAEs (time frame: up to 67 months approximately) • Number of participants who discontinued treatment due to toxicity (time frame: up to 67 months approximately)

NCT02811861 *(Continued)*

- Health-related QoL scores (time frame: up to 47 months)
- PFS by investigator assessment (time frame: up to 47 months approximately)

Relevant to this review but not reported: TFST

Other outcomes (not relevant to this review): PFS on next-line therapy, ORR, TTF, AUC, time of clearance

Notes	**Funding sources:** Eisai and Merck Sharp and Dohme **Declaration of Interest**: Disclosure forms provided by the authors are available at NEJM.org. **Clinical study report available**: no **Study protocol available:** yes **Statistical analysis plan available:** yes

NCT02853331

Study characteristics

Methods	**Study name:** KEYNOTE- 426 **Study design:** a two-arm, RCT, phase III **Blinding:** none, open-label **Study dates:** Oct 24, 2016 - Jan 24, 2018 (date of randomisation) **Date of data cut-off:** exact dates for OS and PFS not reported, but data for both outcomes was extracted at the longest follow-up available (median 42.8 months). Data for OS subgroups were available for a shorter follow-up (median 30.6 months) **Location:** 16 countries (not reported), types of centres: hospitals, cancer centres (129 study locations) **Cross-over study or cross over permitted:** no
Participants	**Inclusion criteria:** • all sexes • 18 years and older • has histologically confirmed diagnosis of RCC with clear cell component with or without sarcomatoid features • has locally advanced/metastatic disease (i.e., newly diagnosed Stage IV RCC per American Joint Committee on Cancer) or has recurrent disease • has measurable disease per RECIST 1.1 as assessed by the investigator/site radiologist • has received no prior systemic therapy for advanced RCC • has KPS ≥ 70% as assessed within 10 days prior to randomisation **Exclusion criteria:** • is currently participating in or has participated in a study of an investigational agent or has used an investigational device within 4 weeks prior to randomisation • has had major surgery within 4 weeks, received radiation therapy within 2 weeks prior to randomisation, or has not recovered (i.e., ≤ Grade 1 or at baseline) from Aes due to prior treatment • has had prior treatment with any anti-programmed cell death (anti-PD-1), or programmed cell death ligand 1 (PD-L1), or PD-L2 agent or an antibody targeting any other immune-regulatory receptors or mechanisms

NCT02853331 *(Continued)*

- has received prior systemic anti-cancer therapy for RCC with VEGF/VEGFR or mTOR targeting agents

Sample size: N=861

Age, median (years, range): experimental arm: 62 (55-68), control arm: 61 (53-68)

Sex (m/f): experimental arm: 308/124, control arm: 320/109

Prognostic factors:

- **IMDC prognostic risk, n(%)**
  - **Favourable**
    - experimental arm: 138 (32), control arm: 131 (31)
  - **Intermediate**
    - experimental arm: 238 (55), control arm: 246 (57)
  - **Poor**
    - experimental arm: 56 (13), control arm: 52 (12)
- **Previous nephrectomy, n(%)**
  - **Yes**
    - experimental arm: 359 (83), control arm: 359 (84)

Interventions

Experimental arm (n = 432): pembrolizumab (200 mg, intravenous, every 3 weeks for up to 35 cycles) + sxitinib (5 mg, oral, twice/day)

Control arm (n = 429): sunitinib (50 mg, oral, once/day) - treatment was 4 weeks on treatment, 2 weeks off treatment (1 cycle = 6 weeks)

Outcomes

Primary outcome(s)

- PFS per RECIST 1.1 as assessed by blinded independent central imaging review
  - Time frame: through database cut-off date of 24-Aug-2018 (up to approximately 22 months)
- OS
  - Time frame: through database cut-off date of 24-Aug-2018 (up to approximately 22 months)

Secondary outcome(s)

- Number of participants who experienced an AE
  - Time frame: through database cut-off date of 24-Aug-2018 (up to approximately 22 months)
- Number of participants who discontinued study drug due to an AE
  - Time frame: through database cut-off date of 24-Aug-2018 (up to approximately 22 months)
- PFS rate at month 12, 18 and 24 in all participants
- OS rate at month 12, 18 and 24 in all participants
- SAEs
  - Time frame: through database cut-off date of 24-Aug-2018 (up to approximately 22 months)
- Other (not including serious) AEs
  - Time frame: through database cut-off date of 24-Aug-2018 (up to approximately 22 months)

Relevant to this review but not reported: TFST, QoL scale (reporting is planned)

Other outcomes (not relevant to this review): TTD, DoR, ORR, DCR

Notes

Funding sources: Merck Sharp & Dohme Corp, a subsidiary of Merck & Co, Inc.

Declaration of Interest: For a very detailed description please refer to the publication.

Clinical study report available: no

Study protocol available: yes

Statistical analysis plan available: yes

NCT03141177

Study characteristics	
Methods	**Study name:** CheckMate 9ER **Study design:** RCT, phase III **Blinding:** none, open-label **Study Dates:** September 2017 - May 2019 (date of randomisation) **Date of data cut-off:** March 30, 2020 (for safety); exact dates for OS and PFS not reported, but we extracted data for the longest follow-up time available (median 23.5 months) **Location:** 18 countries, (Argentina, Australia, Brazil, Chile, Czech Republic, Germany, Greece, Israel, Italy, Japan, Mexico, Poland, Romania, Russia, Spain, Turkey, UK, USA), types of centres: not reported (only "local institutions") **Cross-over study or cross over permitted:** no
Participants	**Inclusion criteria:** • all sexes • 18 years and older • histological confirmation of RCC with a clear-cell component, including participants who may also have sarcomatoid features • advanced (not amenable to curative surgery or radiation therapy) or metastatic (AJCC Stage IV) RCC • no prior systemic therapy for RCC with the following exception: ◦ one prior adjuvant or neoadjuvant therapy for completely resectable RCC if such therapy did not include an agent that targets VEGF or VEGF receptors and if recurrence occurred at least 6 months after the last dose of adjuvant or neoadjuvant therapy **Exclusion criteria:** • any active CNS metastases • any active, known or suspected autoimmune disease • any condition requiring systemic treatment with either corticosteroids (>10 mg daily prednisone equivalent) or other immunosuppressive medications within 14 days of randomisation Statement on CT.gov: "Other protocol defined inclusion/exclusion criteria could apply." **Sample size:** N=651 **Age, median (years, range):** experimental arm: 62 (29-90), control arm: 61 (28-86) **Sex (m/f, (%)):** experimental arm: 249/74 (77.1/22.9), control arm: 232/96 (70.7/29.3) **Prognostic factors:** • **IMDC risk score, n(%)** • **Favourable** • experimental arm: 74 (22.9), control arm: 72 (22.0) • **Intermediate** • experimental arm: 188 (58.2), control arm: 188 (57.3) • **Poor** • experimental arm: 61 (18.9), control arm: 68 (20.7) • **Previous nephrectomy, n(%)** • **Yes** • experimental arm: 222 (68.7), control arm: 233 (71.0)

NCT03141177 *(Continued)*

- **Karnofsky performance-status score, n(%)**
- **90 or 100**
- experimental arm: 257 (79.6), control arm: 241 (73.5)
- **70 or 80**
- experimental arm: 66 (20.4), control arm: 85 (25.9)
- **Not reported**
- experimental arm: 0 (0), control arm: 2 (0.6)

Interventions	**Experimental group (n = 323):** nivolumab (240 mg, intravenous, every 2 weeks) + cabozantinib (40 mg, oral, once/day) **Control group (n =3 28):** sunitinib (50 mg, oral, once/day) - treatment was 4 weeks on treatment, 2 weeks off treatment (1 cycle = 6 weeks)

Outcomes	**Primary outcome(s)**
	
	• PFS per BICR
	  ◦ Time frame: up to 29 months
	
	Secondary outcome(s)
	
	• OS
	  ◦ Time frame: up to 40 months
	• Incidence of AEs
	  ◦ Time frame: up to 40 months
	• Incidence of SAEs
	  ◦ Time frame: up to 40 months
	• Incidence of AEs leading to discontinuation
	  ◦ Time frame: up to 40 months
	• Incidence of deaths
	  ◦ Time frame: up to 40 months
	• QoL (not stated on CT.gov).
	
	Relevant to this review but not reported: TFST
	
	Other outcomes (not relevant to this review): laboratory values, ORR

Notes	**Funding sources:** Bristol-Myers Squibb and others **Declarations of Interests:**Quote: "Disclosure forms provided by the authors are available with the full text of this article at NEJM.org." **Clinical study report available**: not yet **Study protocol available:** yes **Statistical analysis plan available:** yes

AEs: adverse events; ;**v; CNS**: central nervous system; **CT:** computed tomography; **DCR:** disease control rate; **DR:** duration of response;**ECOG:** Eastern Cooperative Oncology Group; **ITT:** intention-to-treat; **KPS:** Karnofsky Performance Status**LDH**: lactate dehydrogenase;**MRI:** magnetic resonance imaging;**ORR:** objective response rate;**mTOR:** mammalian target of rapamycin; **OS: overall survival ; PFS**: progression-free survival**; QoL:** quality of life; **RCC:** renal cell carcinoma; **RR:** response rate; **SAEs:** serious adverse events; **SD:** standard deviation; **TTP:** time to progression; **VEGF:** Vascular endothelial growth factor.
NCT not available for this study.

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Aass 2005	Irrelevant interventions (interferon alfa-2a with/without 13-cis-retinoic acid).
Abdel 2018	Wrong study design (single-arm) and wrong patient population (recurrent/refractory).
Adler 1987	Irrelevant interventions (hormono-immuno- versus hormonotherapy).
Amin 2018	CheckMate 016 study. Wrong study design (non-randomised).
Amin 2018a	CheckMate 016 study. Wrong study design (non-randomised).
Atkins 1991	Irrelevant comparison (interferon vs. interleukin).
Atkins 1993	Irrelevant comparison (interferon vs. interleukin).
Atzpodien 1997	Irrelevant comparison (interleukin vs. interferon and 5-fluorouracil).
Atzpodien 1997a	Irrelevant comparison (interleukin vs. interferon and 5-fluorouracil).
Atzpodien 1999	Irrelevant comparison (13-cis-retinoic acid, IFN-alpha, IL-2 and chemotherapy).
Atzpodien 2001	Irrelevant comparison (interferon+interleuking and 5-FU versus tamoxifen).
Atzpodien 2004	Irrelevant comparison (interleukin vs. interferon).
Atzpodien 2006	Irrelevant comparison (interleukin-2/interferon-alpha2a/13-retinoic acid-based chemoimmunotherapy).
Barrios 2009	Wrong study design (single-group assignment).
Beaumont 2009	RECORD-1 study. Second-line treatment.
Beaumont 2011	RECORD-1 study. Second-line treatment.
Berg 1998	Irrelevant comparison (interleukin versus interferon alpha-2A).
Bex 2017	Irrelevant intervention (nephrectomy).
Boccardo 1998	Irrelevant comparison (interleukin versus interferon alpha-2A).
Bracarda 2007	Wrong study design (same intervention, different schedules).
Buckley 2019	PRISM study. Dose-finding study.
Cella 2016	Checkmate025 study. Second-line treatment.
Choueiri 2017	Prior therapy allowed (more than 10% of patients).
Choueiri 2020	Prior therapy allowed (more than 10% of patients).
Choueiri 2020a	Prior therapy allowed (more than 10% of patients).
Cirkel 2016	Wrong study design (rotating treatments).
Cirkel 2017	ROPETAR study. Wrong study design (rotating treatments).

Study	Reason for exclusion
Climent 2020	ROPETAR study. Wrong study design (rotating treatments).
Cole 2003	Irrelevant comparison (interleukin versus interferon alpha-2A).
Collinson 2012	STAR Trial. Irrelevant comparison (sunitinib; temporary cessation versus continuation).
Collinson 2018	PRISM. Dose-finding study.
Colomba 2021	Wrong study design (single-group assignment).
Conter 2013	Wrong study design (dose-finding study).
de Mulder 1991	Irrelevant comparison (interferon versus interleukin).
Demirci 1999	Irrelevant interventions (vinblastine and interferon alpha with 5-flourouracil).
Dexeus 1988	Irrelevant comparison (chemotherapy versus interferon).
Dexeus 1989	Irrelevant comparison (chemotherapy versus interferon).
DRKS00010309 2016	TAURUS. Terminated study.
Dubois 1997	Irrelevant comparator (p75 tumour necrosis factor receptor immunoglobulin G chimera).
Eisen 2019	Irrelevant intervention (adjuvant therapy).
Elhilali 2000	Irrelevant comparison (interferon vs. placebo).
Epaillard 2020	Wrong study design (biomarker-driven trial).
Escudier 2005	Irrelevant comparator.
Euctr 2006-003429-95-ES	Wrong study design (non-randomised).
Euctr2006-002851-33-AT	Ended prematurely.
Euctr2006-005751-16-NL	Study ended prematurely.
Euctr2007-002556-41-AT	Irrelevant comparison (trivax (cancer vaccine) with sunitinib versus sunitinib alone).
EUCTR2008-002667-13-DE 2008	Adjuvant setting.
Euctr2012-001730-33-ES	Study ended prematurely.
Euctr2015-002133-22-FR	Irrelevant intervention(s).
Euctr2018-001495-38-FR	Study ended prematurely.
Feldman 2020	Wrong study design (cohort study).
Feldman 2020a	Wrong study design (cohort study).
Figlin 1998	Irrelevant comparison (CD8(+) tumour-infiltrating lymphocytes in combination with recombinant interleukin-2).

Study	Reason for exclusion
Figlin 1999	Irrelevant comparison (CD8(+) tumor-infiltrating lymphocytes in combination with recombinant interleukin-2).
Figlin 2014	ADAPT trial. Terminated (no results).
Figlin 2014a	ADAPT trial. Terminated (no results).
Figlin 2017	ADAPT trial. Terminated (no results).
Figlin 2018	ADAPT trial. Terminated (no results).
Figlin 2020	ADAPT trial. Terminated (no results).
Flaherty 2015	Prior therapy allowed.
Foon 1988	Irrelevant comparison (interferon alpha(2B)-interferon/gamma- interferon or the combination).
Fossa 1989	Irrelevant comparison (recombinant interferon-alpha with or without vinblastine).
Fossa 1992	Irrelevant comparison (recombinant interferon-alpha with or without vinblastine).
Gao 2017	Prior therapy allowed.
Gao 2019	Prior therapy allowed.
Gedye 2021	Wrong study design (single-group assignment).
Ghiorghiu 2018	More than 10% received prior therapy.
Gleave 1997	Irrelevant comparison (interferon gamma-1b injection versus placebo).
Gleave 1997a	Irrelevant comparison (interferon gamma-1b injection versus placebo).
Gleave 1998	Irrelevant comparison (interferon gamma-1b injection versus placebo).
Gore 2008	Irrelevant comparison (interferon-a (IFN), interleukin-2 (IL2) and 5-fluorouracil (5FU) vs IFN alone).
Gore 2010	Irrelevant comparison (interferon-a (IFN), interleukin-2 (IL2) and 5-fluorouracil (5FU) vs IFN alone).
Gruenwald 2020	Irrelevant intervention (behavioral intervention, concomitant coaching).
Haas 2016	Wrong intervention (adjuvant therapy).
Hainsworth 2015	Irrelevant comparator (CXCR4 inhibitor LY2510924).
Hainsworth 2016	Irrelevant comparator (CXCR4 inhibitor LY2510924).
Han 2002	Irrelevant comparison (interleukin-2 versus subcutaneous interleukin-2/interferon).
Harima 1990	Irrelevant comparison (interferon-alpha (IFN) plus fluoropyrimidine (FP) and IFN alone).
Henriksson 1998	Irrelevant comparison (tamoxifen vs interleukin 2, alpha-interferon (leucocyte) and tamoxifen).
Hutson 2006	Wrong study design (discontinuation design) and wrong patient population (recurrent).

Study	Reason for exclusion
Hutson 2021	CheckMate 920 trial. Wrong study design (non-randomised).
ISRCTN95351638	PRISM. Dose-finding study.
Jager 2005	Second-line therapy.
Jayson 1998	Irrelevant comparison (interleukin 2 and interleukin 2-interferon alpha).
Jeon 1999	Wrong study design (cohort study).
JPRN-JapicCTI-122014	Prior therapy allowed.
JPRN-jRCTs031180024	Irrelevant comparison (nivolumab combined with image-guided three dimensional beam-convergent and extremely hypofractionated radiotherapy).
JPRN-UMIN000001995	Prior therapy allowed.
Kinouchi 2004	Irrelevant comparison (interferon-alpha (IFN) versus IFN + cimetidine).
Kinouchi 2006	Irrelevant comparison (interferon-alpha (IFN) versus IFN + cimetidine).
Larkin 2019	KEYNOTE-427. Wrong study design (non-randomised).
Law 1995	Irrelevant comparison (interleukin-2 with or without lymphokine- activated killer cells).
Lee 2020	KEYNOTE-427. Wrong study design (non-randomised).
Lee 2021	Wrong study design (cohort study).
Lindskog 2020	Irrelevant comparison (ilixadencel plus sunitinib versus sunitinib alone).
Lissoni 1993	Irrelevant comparison (interleukin-2 subcutaneous immunotherapy versus interleukin-2 plus interferon-alpha).
Liu 2012	Irrelevant comparison (autologous CIK cell immunotherapy versus interleukin-2 treatment combination with IFN-α-2a).
Lummen 1996	Irrelevant comparison (interferon-gamma versus interleukin-2 and interferon-alpha2b).
Madhusudan 2004	Irrelevant comparison (interferon alpha alone or in combination with thalidomide).
McDermott 2001	Irrelevant comparison (interleukin-2 versus interleuking + interferon).
McDermott 2005	Irrelevant comparison (interleukin-2 versus interleukin + interferon).
McDermott 2013	BEST trial. Prior therapy allowed.
McDermott 2020	KEYNOTE-427. Wrong study design (non-randomised).
Mickisch 2001	Irrelevant comparison (radical nephrectomy plus interferon-alfa-based immunotherapy compared with interferon alfa alone).
Minasian 1993	Wrong study design (cohort study).
Molina 2009	RECORD-1 study. Prior therapy allowed.

Study	Reason for exclusion
Motzer 2001	Irrelevant comparison (interleukin-12 versus interferon-alpha2a).
Mulders 2012	Prior therapy (50% of patients).
Naglieri 1998	Irrelevant comparison (interleukin-2 (IL-2) and interferon-alpha immunotherapy versus an IL-2 and 4-epirubicin immuno-chemotherapy).
NCT00002737	Irrelevant comparison (interferon alfa plus isotretinoin versus interferon alfa alone).
NCT00005966	Irrelevant comparison (interferon-alfa2b alone versus interferon-alfa2b plus thalidomide).
NCT00019539	Irrelvant comparison (bevacizumab versus thalidomide) and prior therapy allowed.
NCT00027664	Irrelevant comparison (interferon alfa with thalidomide versus interferon alfa alone).
NCT00053820	Irrelevant comparison (interleukin-2 and fluorouracil versus interferon alfa alone).
NCT00073307	Second-line therapy.
NCT00100906	Irrelevant interventions and dose-finding study.
NCT00378703	One line of prior therapy was allowed.
NCT00416871	Irrelevant comparison (interleukin infusion versus interleukin by injection).
NCT00467025	Irrelevant intervention (AMG386).
NCT00491738	SABRE-R study. Terminated (no results).
NCT00709995	Phase II not conducted.
NCT00835978	Dose-finding study.
NCT00873236	Study ended prematurely.
NCT01164228	Irrelevant comparison (sunitinib with or without gemcitabine hydrochloride).
NCT01223027	Second-line treatment.
NCT01408004	Wrong study design (efficacy of rotating regimen).
NCT01444807	Wrong study design (single-group assignment).
NCT01616186	Study withdrawn.
NCT01664182	Second-line therapy.
NCT01673386	Study terminated (no results).
NCT01727089	Second-line therapy.
NCT01727336	Second-line therapy.
NCT01793636	Second-line therapy.

Study	Reason for exclusion
NCT02014636	Phase II (randomised part) not conducted.
NCT02127710	Wrong study design (single-group assignment).
NCT02667886	Second-line therapy.
NCT02724020	Second-line therapy.
NCT02960906	Wrong-study design (BIOmarker-driven trial).
NCT03035630	Study terminated (no results).
NCT03092856	Second-line therapy.
NCT03095040	Second-line therapy.
NCT03173560	Dose-finding study.
NCT03501381	Second-line therapy.
NCT03595124	Second-line therapy.
NCT03829111	Irrelevant intervention (probiotics in addition to nivolumab and ipilimumab).
NCT04195750	Second-line therapy.
NCT04300140	Second-line therapy.
Negrier 1996	Irrelevant comparison (interleukin-2 versus interferon alfa).
Negrier 1997	Irrelevant comparison (interleukin-2 and interferon with or without fluorouracil).
Negrier 1998	Irrelevant comparison (interleukin-2 versus interferon alfa).
Negrier 2000	Irrelevant comparison (interleukin-2 and interferon with or without fluorouracil).
Negrier 2006	Irrelevant comparison (intravenous interleukin versus subcutaneous interleukin).
Negrier 2007	Irrelevant comparison (medroxyprogesterone with interferon alfa-2a or interleukin 2 versus a combination of both).
Negrier 2008	Irrelevant comparison (intravenous interleukin versus subcutaneous interleukin).
Nosov 2010	Wrong study design (randomised discontinuation trial).
Nosov 2012	Wrong study design (randomised discontinuation trial).
Pal 2015	Second-line therapy.
Pal 2021a	Second-line therapy.
Passalacqua 2010	Irrelevant comparison (interleukin-2 versus interferon-alpha).
Plimack 2015	Wrong study design (dose-finding study).

Study	Reason for exclusion
Pyrhonen 1995	Irrelevant comparison (interferon alfa-2a with vinblastine versus vinblastine alone).
Pyrhonen 1996	Irrelevant comparison (interferon alfa-2a with vinblastine versus vinblastine alone).
Pyrhonen 1999	Irrelevant comparison (interferon alfa-2a with vinblastine versus vinblastine alone).
Ravaud 2006	Full-text not available; inclusion criteria unclear.
Ravaud 2016	Adjuvant therapy.
Rexer 2017	Terminated study.
Richards 1977	Irrelevant interventions (chemotherapy).
Rini 2011	Irrelevant comparator (AMG 386).
Rini 2012	Irrelevant comparator (AMG 386).
Rodriguez-Vida 2020	Terminated study (no results).
Rpcec 2017	Irrelevant comparison (HeberFERON intravenous versus HeberFERON subcutaneous).
Sternberg 2013	Wrong study design (cohort study) and wrong study population (refractory).
Szarek 2021	Second-line treatment.
Tannir 2016	Terminated study.
Taylor 2020	Wrong study design (single-group assignment) and wrong patient population (selected solid Tumours).
Taylor 2020a	Wrong study design (single-group assignment) and wrong patient population (selected solid Tumours).
Thiam 2010	RECORD-1 study. Second-line treatment.
Trump 2004	Irrelevant comparison (subcutaneous interferon alfa-2a, subcutaneous interleukin-2 and intravenous fluorouracil versus oral 13-cis-retinoic acid versus IFN-alpha-2a and vinblastine.).
Twardowski 2015	Prior therapy allowed (more than 10% with prior therapy).
Twardowski 2017	Prior therapy allowed (more than 10% with prior therapy).
Verzoni 2018	Irrelevant comparator (cytoreductive nephrectomy).
Voss 2015	Wrong study design (single-group assignment).
Voss 2019	Second-line therapy (1–3 prior therapy lines).
Witte 1995	Irrelevant comparison (interleukin versus interferon).
Wood 2013	ADAPT study. Terminated (no results).
Wright 2020	Second-line therapy.

Study	Reason for exclusion
Yang 2002	Prior therapy (majority of patients).
Yang 2003	Prior therapy (majority of patients).
Zhou 2016	Prior chemotherapy (13% of patients).
Zhou 2019	Prior chemotherapy (13% of patients).

Characteristics of studies awaiting classification *[ordered by study ID]*

Liu 2017

Methods	**Study type:** randomised, parallel trial **Blinding:** no information* **Study dates:** no information* **Cross-over study:** no **Status:** no information*
Participants	**Estimated enrolment:** N = 56 was the final sample size **Risk groups:** no information* **Inclusion criteria:** no information*. Included were patients with diagnosed advanced renal cell carcinoma. **Exclusion criteria:** no information*
Interventions	**Experimental arm:** sunitinib **Control arm:** Interleukin- 2 combined with interferon-alpha treatment
Outcomes	**Primary outcome(s)** • OS • PFS **Secondary outcome(s)** - **Outcomes relevant to this review but not to be assessed:** AEs/SAEs, QoL, TFST, number of patients who discontinued treatment due to an AE **Other outcomes to be assessed (not relevant to this review):** -
Notes	*Only abstract available, no further information present so far

NCT01217931

Methods	**Study type:** randomised, phase II study **Blinding:** no, open-label

NCT01217931 *(Continued)*

	Study dates: January 2011 - January 2022 (final data collection date for primary outcome measures) **Cross-over study:** sequential two-agent assessment **Status:** active, not recruiting (as of May 4, 2022)
Participants	**Estimated enrolment:** 240 **Risk groups:** no information **Inclusion criteria:** • confirmed metastatic RCC with a clear cell component • prior radical or partial nephrectomy required. Participants whose primary tumour was treated with cryoablation or radiofrequency ablation would also be eligible • measurable disease • Age >/= 18 years • ECOG performance status 0 or 1 • adequate organ and marrow function within 14 days (see CT.gov for specifics) • non-pregnant female participants • participants of child fathering or childbearing potential must be on birth control while on study • participants must give written informed consent prior to initiation of study-related procedures. participants with a history of major psychiatric illness must be judged able to fully understand the investigational nature of the study and the risks associated with the therapy **Exclusion criteria:** • no patient with any concurrent active malignancy, i.e. a patient requiring or receiving systemic therapy for another malignancy at the same time of treatment for RCC • participants must not have received any prior targeted therapy (anti-VEGF agents or mTOR inhibitors), including adjuvant therapy, and must not have received any prior chemotherapy for mRCC. However, participants who had received prior immunotherapy, such as cytokines or vaccines, are permitted to enrol. • participants must not be scheduled to receive another experimental drug while on this study. Participants are permitted to receive concomitant bisphosphonates. • participants must not have multiple brain metastases or leptomeningeal disease. Participants with controlled solitary brain metastasis are eligible. More exclusion criteria on CT.gov.
Interventions	Group 1: Pazopanib + possible Bevacizumab Group 2: Pazopanib + possible Everolimus Group 3: Everolimus + possible Bevacizumab Group 4: Everolimus + possible Pazopanib Group 5: Bevacizumab + possible Pazopanib Group 6: Bevacizumab + possible Everolimus
Outcomes	**Primary outcome(s)** - **Secondary outcome(s)** -

NCT01217931 *(Continued)*

	Outcomes relevant to this review but not to be assessed: PS, PFS, TFST, AE, SAE, QoL, number of participants who discontinued treatment due to an AE **Other outcomes to be assessed (not relevant to this review):** time to overall treatment failure
Notes	Prior systemic therapy:quote: "participants must not have received any prior targeted therapy (anti-VEGF agents or mTOR inhibitors), including adjuvant therapy, and must not have received any prior chemotherapy for mRCC. However, participants who had received prior immunotherapy, such as cytokines or vaccines, are permitted to enroll." --> Awaiting results to check number of participants with prior immunotherapy (if any), and whether results are reported separately for the treatment-naive participants. **Funding sources:** M.D. Anderson Cancer Center, Novartis

NCT01688973

Methods	**Study type:** interventional, randomised, parallel assignment, phase II **Blinding:** no, open-label **Accrual period:** August 20, 2012 - April 30, 2017 **Cross-over study:** no **Status:** completed (as of January 3, 2019)
Participants	**Estimated enrolment:** N =55 **Risk groups:** no information **Inclusion criteria:** • participants must have histologically or cytologically confirmed papillary histology renal cell carcinoma which is metastatic, or locally advanced and unresectable; mixed histologies will be allowed provided that they contain >= 50% of the papillary component • participants must have measurable disease, defined as at least one lesion that can be accurately measured in at least one dimension • participants with metastatic disease who have a resectable primary tumour and are deemed a surgical candidate may have undergone resection • participants with a history of brain metastases who are asymptomatic and have not received steroid therapy in the 14 days prior to registration are eligible; anti-seizure medications are allowed provided they are non-enzyme inducing • participants may have received up to one prior systemic therapy for advanced or metastatic renal cell carcinoma; participants must not have received a MET inhibitor or erlotinib as prior therapy; at least 21 days must have elapsed since completion of prior systemic therapy, 42 days for nitrosourea or mitomycin C; participants must have recovered from all associated toxicities at the time of registration • participants may have received prior radiation therapy, but must have measurable disease outside the radiation port; at least 21 days must have elapsed since completion of prior radiation therapy; participants must have recovered from all associated toxicities at the time of registration • participants must not be receiving or planning to receive any other investigational agents More inclusion criteria on CT.gov. **Exclusion criteria:** not reported **Sample size:** N = 55, 50 participants were analysed **Age, median (years, range):** experimental arm: 63.6 (22.8-81.9), control arm: 62.1 (20.3 - 76.1)

NCT01688973 *(Continued)*

	Sex (m/f, (%)): experimental arm: 15/10 (60/40), control arm: 19/6 (76/24) **Prognostic factors:** • **Previous nephrectomy, n(%)** • **Yes** • experimental arm: 18 (72), control arm: 21 (84)
Interventions	**Experimental arm (n = 25):** tivantinib orally twice daily and erlotinib hydrochloride orally once daily on days 1-28 **Control arm (n = 25):** tivantinib orally twice daily on days 1-28
Outcomes	**Primary outcome(s)** - **Secondary outcome(s)** • Frequency and severity of toxicities, graded by the National Cancer Institute Common Terminology Criteria for Adverse Events Version 4.0 (time frame: up to 3 years) • PFS (time frame: 30 months) **Relevant to this review but not listed on CT.gov:** OS, TFST, QoL, number of participants who discontinued treatment **Other outcomes (not relevant to this review):** response rate
Notes	**Previous therapy:** participants may have received prior therapy (see inclusion criteria). Awaiting results to check whether results for treatment-naive participants are reported separately.

NCT01829841

Methods	**Study type:** RCT, parallel assignment, phase II **Blinding:** unclear, "double-blind" stated in the title, but "none (open-label)" in the description on CT.gov **Study dates:** May 2011- May 2016 (actual study completion period) **Countries:** multicentre **Cross-over study:** no **Status:** completed (as of May 3, 2018)
Participants	**Estimated enrolment:** N = 150 **Risk groups:** no information **Inclusion criteria:** • participants with histologically confirmed advanced renal cell carcinoma including clear cell component and not available for surgery • first-line therapy or second-line treatment (second-line treatment e.g. chemotherapy or cytokine therapy as first-line treatment failure or resistant participants) • with measurable disease (using RECIST1.0 standard conventional CT scan ≥ 20 mm, spiral CT scan ≥ 10 mm, target lesion did not receive radiation therapy, cryotherapy) • male or female, age ≥18 and ≤75 • ECOG 0-1

NCT01829841 *(Continued)*

- life expectancy ≥ 3 months
- participants received surgery, chemotherapy, radiation therapy, cytokines treatment caused the damage has been restored, the time interval ≥ 4 weeks, and the wound has completely healed
- normal major organ function
- signed and dated informed consent

Exclusion criteria:

- previously received targeted therapy of the metastatic renal cell carcinoma (such as sunitinib, sorafenib)
- past or suffering from other cancer, but other than cure basal cell carcinoma and cervical carcinoma in situ
- participated in other clinical trials within four weeks
- a variety of factors that affect the oral medication (such as inability to swallow, gastrointestinal resection, chronic diarrhoea and intestinal obstruction)
- known brain metastases, spinal cord compression, cancer, meningitis, or screening CT or MRI examination revealed brain or leptomeningeal disease

Age, median (years, range): 18-75

Sex (m/f, (%)): all sexes are eligible

More inclusion criteria on CT.gov.

Interventions	**Experimental arm:** Famitinib (Famitinib 25 mg once daily orally) **Control arm:** Sunitinib (Sunitinib 50 mg orally once daily)
Outcomes	**Primary outcome(s)** - **Secondary outcome(s)** • PFS (time frame: 3 years) • OS (time frame: 3 years) • QoL (time frame: 42-day cycle visit until disease progression) • number of participants with AEs (time frame: 3 years) **Relevant to this review but not to be assessed:** TFST, SAE, number of participants who discontinued treatment due to an AE **Other outcomes (not relevant to this review):** ORR, DCR, body vitals, laboratory parameters
Notes	**Funding sources:** Jiangsu HengRui Medicine Co., Ltd.Cancer Institute and Hospital, Chinese Academy of Medical Sciences **Previous therapy:** may include participants in second-line therapy (see inclusion criteria). Awaiting results to check whether results for treatment-naive participants are reported separately.

NCT03541902

Methods	**Study type:** interventional, randomised, parallel assignment, phase II **Blinding:** no, open-label **Accrual period:** May 15, 2018 - July 31, 2022 (estimated study completion date) **Countries:** multicenter (N = 3 in the USA)

NCT03541902 *(Continued)*

	Cross-over study: no **Status:** active, not recruiting (as of June 1st, 2022)
Participants	**Estimated enrolment:** N = 84 **Risk groups:** no information **Inclusion criteria:** • the participant has a histologic or cytologic diagnosis of a variant histology renal cell carcinoma including papillary, chromophobe, Xp.11 translocation, undifferentiated, or unclassified which is treatment-naïve or has previously been treated with one systemic treatment line not containing any vascular endothelial growth factor antibody or vascular endothelial growth factor receptor tyrosine kinase inhibitors. The patient may have received treatment with immune checkpoint therapy including nivolumab as a single agent or nivolumab plus ipilimumab in combination. Previous treatment with mammalian target of rapamycin agents such as temsirolimus or everolimus is acceptable • measurable disease per RECIST v1.1 as determined by the investigator • the participant has had an assessment of all known disease sites e.g. by computerized tomography (CT) scan, magnetic resonance imaging (MRI), bone scan as appropriate, within 28 days before the first dose of cabozantinib or sunitinib • the participant is >/=18 years old on the day of consent • the participant has an Eastern Cooperative Oncology Group (ECOG) performance status of </=2 • recovery to baseline or </= Grade 1 CTCAE v.4.0 from toxicities related to any prior treatments, unless AE(s) are clinically non significant and/or stable on supportive therapy **Exclusion criteria:** • the participant has a variant histology that includes renal medullary carcinoma or collecting duct renal cell carcinoma. Any clear cell component in the tumour will lead to exclusion • the participant has received any previous anti-angiogenic agent. Prior treatment with cabozantinib • radiation therapy for bone metastasis within 2 weeks, any other external radiation therapy within 4 weeks before the first dose of study treatment. Systemic treatment with radionuclides within 6 weeks before the first dose of study treatment. Participants with clinically relevant ongoing complications from prior radiation therapy are not eligible • the participant has received any other type of investigational agent within 28 days before the first dose of study treatment • known brain metastases or cranial epidural disease unless adequately treated with radiotherapy and/or surgery (including radiosurgery) and stable for at least 4 weeks before the first dose of study treatment. Eligible participants must be neurologically asymptomatic and without corticosteroid treatment at the time of the start of study treatment More inclusion and exclusion criteria on CT.gov.
Interventions	**Experimental arm:** cabozantinib orally once daily on days 1-42) **Control arm:** sunitinib malate (orally once daily on days 1-28
Outcomes	**Primary outcome(s)** • PFS evaluated using RECIST 1.1 Criteria (time frame: from randomisation up to the time of disease progression or death up to two years) **Secondary outcome(s)** • OS (time frame: from randomisation to death or last contact if still alive up to two years) • AE rates (time frame: start of study drug up to 30 days after last dose of study drug)

NCT03541902 *(Continued)*

	Relevant to this review but not reported: SAE, TFST, participants who discontinued treatment due to an AE, QoL **Other outcomes (not relevant to this review):** ORR
Notes	**Funding sources:** M.D. Anderson Cancer CenterExelixisNational Cancer Institute (NCI) **Previous therapy:** participants may have received prior therapy (see inclusion criteria). Awaiting results to check whether results for treatment-naive participants are reported separately.

AEs: adverse events;**CT**: computed tomography;**ECOG:** Eastern Cooperative Oncology Group (ECOG**);MRI:** magnetic resonance imaging;**mTOR**: mechanistic target of rapamycin ;**OS: overall survival ; PFS: progression-free survival ; QoL:** quality of life; RCC: renal cell carcinoma; **SAEs:** serious adverse events.

Characteristics of ongoing studies *[ordered by study ID]*

EUCTR2008-000928-71-IT

Study name	-
Methods	**Study type:** RCT, phase II **Blinding:** no, open-label **Accrual period:** no information **Country:** Italy **Cross-over study:** no **Status:** ongoing
Participants	**Estimated enrolment:** no information **Inclusion criteria:** • Signed informed consent • Histologically or cytologically documented non-clear Cell Renal Carcinoma (with centralised review of histological specimens). In the case of a mixed histology the presence of a documented component of clear cell histology <50% is mandatory. Urothelial upper urinary tract tumours are excluded. In cases with initial diagnosis of non-clear RCC of more than 2 years (RFS >2 years) a histological/cytological confirmation of renal cell carcinoma origin of actual metastases is mandatory • Metastatic measurable disease (at least one uni-dimensional measurable lesion by CT-scan or MRI) according to RECIST criteria (reported in Appendix Karnofsky performance status (KPS) • Patients must be accessible for treatment and follow-up **Exclusion criteria:** • CNS metastases • Previous malignancy except for basal cell skin cancer and cervical carcinoma in situ adequately treated, or any other cancer from which the patient has been disease-free for > = 5 years • Any of the concomitant illness or medical condition indicated below: Serious respiratory or cardiovascular disease such as: congestive heart failure (^{3} NYHA Class II -refer to Appendix-); previous history (within 6 months) of myocardial infarction, angina pectoris or cardiac arrhythmias requiring anti-arrhythmics (excluding beta blockers or digoxin). Active coronary artery disease, uncontrolled hypertension Unstable diabetes mellitus, significant neurological or psychiatric disorders or seizure disorder requiring medication (such as anti-epileptics). Uncontrolled hypertension (systolic pressure ^{3} 160 mm Hg and/or diastolic ^{3} 90mm Hg) while receiving chronic medication. Active clinically serious bacterial or fungal infections (> grade 2 NCI-CTC, Version 3) or active human immunodeficiency virus (HIV) infection or chronic hepatitis B or C

EUCTR2008-000928-71-IT *(Continued)*

	• Previous or concomitant treatment with antiangiogenic agents (e.g.: bevacizumab, sorafenib, sunitinib) or m-TOR inhibitors • Previous treatment with chemotherapy, immunotherapy (IFN and/or Interleukin-2) for advanced disease is allowed prior isotope treatment (e.g. strontium or samarium) • Participation in clinical trials with other experimental agents within 30 days of study entry or concomitant treatment with other experimental drug use of immunosuppressive agents including systemic steroids) • History of organ allograft or autologous bone marrow transplant or stem cell rescue within four months of start of study drug • Pregnant or breast-feeding patients • Women of childbearing potential must have a negative pregnancy test performed within seven days prior to the start of study drug • Both men and women enrolled in this trial must use adequate barrier birth control measures during the course of the trial • Known or suspected allergy to the investigational agent or any agent given in association with this trial
Interventions	**Experimental arm:** Temsirolimus+Interferon-alpha **Control arm:** Temsirolimus monotherapy
Outcomes	**Primary outcome(s)** • PFS **Secondary outcome(s)** • OS • Safety **Relevant to this review but not reported:** QoL, TFST, participants who discontinued treatment due to an AE **Other outcomes (not relevant to this review):** ORR, TTP
Starting date	No information
Contact information	-
Notes	**Funding source:** Gruppo Oncologico Italiano Di Ricerca

NCT02210117

Study name	-
Methods	**Study type:** RCT, early phase I, parallel assignment **Blinding:** no, open-label **Accrual period:** November 25, 2014 - May 21, 2020 (final data collection date for primary outcome measure)) **Countries:** no information **Cross-over study:** no **Status:** active, not recruiting (as of March 23, 2020)

NCT02210117 *(Continued)*

Participants	**Estimated enrolment:** N=105 **Inclusion criteria:** • participants must give written informed consent prior to initiation of therapy, in keeping with the policies of the institution; patients with a history of major psychiatric illness must be judged able to fully understand the investigational nature of the study and the risks associated with the therapy • participants with histologically or cytologically confirmed metastatic clear cell RCC who are eligible for cytoreductive nephrectomy, metastasectomy or post-treatment biopsy; diagnosis must be confirmed by pathologist review of screening biopsy; the determination of resectability will ultimately lie in the clinical judgment of the urologist and medical oncologist involved in the care of the patient • participants must have measurable disease and is defined as a lesion that can be accurately measured on the long axis with a minimum size of 10 mm or a lymph node that can be accurately measured along the short axis of a minimum size of 15 mm (computed tomography [CT] scan slice thickness can be no greater than 5 mm) • participants can have had prior treatment for RCC including prior surgery, radiation therapy, immunotherapy with interleukin (IL)-2 or interferon (but not anti-programmed cell death [PD]1 or anti-cytotoxic T-lymphocyte-associated protein 4 [CTLA-4]), target therapy with receptor tyrosine kinase (RTK) inhibitors/mammalian target of rapamycin (mTOR) inhibitors, such as sunitinib, sorafenib, pazopanib, axitinib, everolimus, and temsirolimus (but not bevacizumab) or chemotherapy **Exclusion criteria:** • any other malignancy from which the patient has been disease-free for less than 2 years, except for non-melanoma skin cancer, in situ carcinoma of any site • participants who have organ allografts • participants who have had a major surgical procedure, open biopsy, or significant traumatic injury with poorly healed wound within 6 weeks prior to first dose of study drug; or anticipation of need for major surgical procedure during the course of the study (other than defined by protocol); or fine needle aspirations or core biopsies within 7 days prior to first dose of study drug • known or suspected autoimmune disease; participants with a history of inflammatory bowel disease (including Crohn's disease and ulcerative colitis) are excluded from this study as are participants with a history of autoimmune disease (e.g., rheumatoid arthritis, systemic progressive sclerosis [scleroderma], systemic lupus erythematosus, autoimmune vasculitis [e.g., Wegener's granulomatosis]) are excluded from this study; any condition requiring systemic treatment with corticosteroids (> 10 mg daily prednisone equivalents) or other immunosuppressive medications within 14 days prior to first dose of study drug; inhaled steroids and adrenal replacement steroids doses > 10 mg daily prednisone equivalents are permitted in the absence of active autoimmune disease • known history of testing positive for human immunodeficiency virus (HIV) or known acquired immunodeficiency syndrome (AIDS); positive test for hepatitis B virus (HBV) using HBV surface antigen (HBV sAg) test or positive test for hepatitis C virus (HCV) using HCV ribonucleic acid (RNA) or HCV antibody test indicating acute or chronic infection • any underlying medical condition, which in the opinion of the investigator, will make the administration of study drug hazardous or obscure the interpretation of adverse events, such as a condition associated with frequent diarrhoea More inclusion and exclusion criteria on CT.gov.
Interventions	**Experimental arm I:** Nivolumab + Bevacizumab + surgery **Experimental arm II:** Nivolumab + Ipilimumab + surgery **Control arm:** Nivolumab + surgery
Outcomes	**Primary outcome(s)**

NCT02210117 *(Continued)*

	• incidence of adverse events, defined any grade 3 or higher adverse event that is possibly, probably, or definitely related to any therapy received on this protocol (time frame: 6 weeks) **Secondary outcome(s)** • PFS (time frame: up to 5 years) • OS (Time frame: Up to 5 years) **Relevant to this review but not reported:** TFST, participants who discontinued treatment due to an AE, safety (AEs/SAEs) **Other outcomes (not relevant to this review):** ORR, DoR, immunological changes in tumour tissue and peripheral blood
Starting date	25.11.2014
Contact information	Padmanee Sharma (M.D. Anderson Cancer Center)
Notes	**Funding sources:** M.D. Anderson Cancer Center, National Cancer Institute (NCI)

NCT02996110

Study name	**FRACTION-RCC**
Methods	**Study type:** RCT, phase II, parallel assignment **Blinding:** no, open-label **Accrual period:** January 17, 2017- January 18, 2023 (estimated study completion date) **Countries:** multicentre (35 study centres) **Cross-over study:** no **Status:** Active, not recruiting (as of March 9, 2022)
Participants	**Estimated enrolment:** N=200 **Inclusion criteria:** • advanced Renal Cell Carcinoma • must have at least 1 lesion with measurable disease • life expectancy of at least 3 months • Karnofsky Performance Status (KPS) must be =>70% **Exclusion criteria:** • participants with suspected or known central nervous system metastases unless adequately treated • participants with autoimmune disease • participants who need daily oxygen therapy Other protocol defined inclusion/exclusion criteria could apply
Interventions	**Experimental arm I:** Nivolumab + Relatlimab **Experimental arm II:** Nivolumab + BMS-986205 **Experimental arm III:** Nivolumab + BMS-813160

NCT02996110 *(Continued)*

	Control arm: Nivolumab + Ipilimumab
Outcomes	**Primary outcome(s):** • PFSR (time frame: up to 24 weeks) **Secondary outcome(s)** • Safety (AEs, SAEs) (time frame: up to 2 years) **Relevant to this review but not reported:** QoL, TFST, number of patients who discontinued treatment **Other outcomes (not relevant to this review):** ORR, DoR
Starting date	17.01.2017
Contact information	Bristol-Myers Squibb
Notes	**Funding source:** Bristol-Myers Squibb

NCT03075423

Study name	**SUNIFORECAST**
Methods	**Study type:** RCT, phase II, parallel assignment **Blinding:** no, open-label **Accrual period:** November 1, 2017 - December 31, 2023 (estimated study completion date) **Countries:** international (Belgium, Czechia, France, Germany, the Netherlands, Spain, the UK), multicentre (40 study locations) **Cross-over study:** no **Status:** Recruiting (as of February 23, 2022)
Participants	**Estimated enrolment:** N = 306 **Inclusion criteria:** • Histological confirmation of non-clear cell renal cell carcinoma (nccRCC) with at least 50% non-clear cell component according to actual World Health Organization (WHO) classification • Advanced (not amenable to curative surgery or radiation therapy) or metastatic (AJCC Stage IV) nccRCC • Performance status: Karnofsky (KPS) > 70% (See Appendix 2, 14.2) d) Measurable disease as per RECIST v 1.1 (See Appendix 3, 14.3) documented by an English radiology report • Participants with all risk categories will be eligible for the study. Patients will be stratified for papillary or non-papillary non-clear cell histology and IMDC risk score. Patients will be categorised according to favourable versus intermediate versus poor risk status at registration according to the International Metastatic RCC Database Consortium (IMDC) criteria • Males and females, > 18 years of age **Exclusion criteria:** • any active brain metastases requiring systemic corticosteroids. Baseline imaging of the brain by MRI is required in participants with clinical signs of potential central nervous system (CNS) involvement within 28 days prior to randomisation

NCT03075423 *(Continued)*

	• tumours with a clear-cell component of > 50%Medical History and Concurrent Diseases • prior systemic treatment with vascular endothelial growth factor (VEGF) or VEGF receptor targeted therapy (including, but not limited to, sunitinib, pazopanib, axitinib, tivozanib, and bevacizumab) or prior treatment with an mammalian target of rapamycin (mTOR) inhibitor or cytokines • prior treatment with an immune checkpoint inhibitor as anti-programmed cell death (PD)PD-1, anti-PD-L1, anti-PD-L2, anti cytotoxic T-lymphocyte-associated Protein 4 (CTLA 4) antibody, or any other antibody or drug specifically targeting T-cell co-stimulation or checkpoint pathway • history of deep vein thrombosis (DVT) unless adequately treated with low molecular weight heparin • prior malignancy active within the previous 3 years except for locally curable cancers that have been apparently cured, such as basal or squamous cell skin cancer, superficial bladder cancer, or carcinoma in situ of the prostate, cervix, or breast • known medical condition (e.g., a condition associated with diarrhoea or acute diverticulitis) that, in the investigator's opinion, would increase the risk associated with study participation or study drug administration or interfere with the interpretation of safety results • major surgery (e.g., nephrectomy) < 28 days prior to the first dose of study drug • anti-cancer therapy < 28 days prior to the first dose of study drug or palliative, focal radiation therapy < 14 days prior to the first dose of study drug • receiving concomitant CYP3A4 inducers or strong CYP3A4 inhibitors (See Appendix 4, 14.4). • hypersensitivity to sunitinib or any of the excipients More inclusion & exclusion criteria on CT.gov.
Interventions	**Experimental arm:** Iplilmumab + Nivolumab **Control arm:** Sunitinib
Outcomes	**Primary outcome(s):** • OS (Time frame: 12 months) **Secondary outcome(s):** • OS (time frame: 6 and 18 months) • OS (lime frame: 5 years) • PFS (time frame: 5 years) • Safety (AEs/SAEs) (time frame: 5 years) • QoL (time frame: 5 years) **Relevant to this review but not reported:** TFST, number of patients who discontinued treatment **Other outcomes (not relevant to this review):** ORR
Starting date	01.11.2017
Contact information	Lothar Bergmann, MD; Nicola Goekbuget, MD
Notes	**Funding sources:** Nicola Goekbuget

NCT03260894

Study name	**KEYNOTE-679/ECHO-302**
Methods	**Study type:** RCT, phase III **Blinding:** no, open-label

NCT03260894 *(Continued)*

	Accrual period: December 7, 2017 - February 8, 2022 (estimated study completion date) **Countries:** multicentre (140 study locations) **Cross-over study:** no **Status**: Active, not recruiting (as of February 28, 2022)
Participants	**Estimated enrolment:** N = 129 **Inclusion criteria:** • histologic confirmation of locally advanced or metastatic RCC with a clear-cell component with or without sarcomatoid features • must not have received any prior systemic therapy for their mRCC • measurable disease based on RECIST v1.1 • archival tumour tissue sample or newly obtained core or excisional biopsy of a tumour lesion as required • Karnofsky performance status ≥ 70% • adequate organ function per protocol-defined criteria **Exclusion criteria:** • use of protocol-defined prior/concomitant therapy • currently receiving or has received an investigational treatment as part of a study of an investigational agent or has used an investigational device within 4 weeks before randomisation • history of severe hypersensitivity reaction to study treatments or their excipients • active autoimmune disease that has required systemic treatment in past 2 years • known additional malignancy that has progressed or has required active treatment in the last 3 years • known active central nervous system metastases and/or carcinomatous meningitis • history of (noninfectious) pneumonitis that required steroids or current pneumonitis • history or presence of an abnormal electrocardiogram that, in the investigator's opinion, is clinically meaningful • significant cardiac event within 12 months before Cycle 1 Day 1
Interventions	**Experimental arm:** pembrolizumab + epacadostat **Control arm:** sunitinib or Pazopanib
Outcomes	**Primary outcome(s):** **Secondary outcome(s):** • Safety and tolerability (AEs) (time frame: data reported from start of study to data cutoff 28-Feb-2019, up to 15 months) **Relevant to this review but not reported:** SAEs, OS, PFS, QoL, TFST, number of patients who discontinued treatment **Other outcomes (not relevant to this review):** ORR
Starting date	07.12.2017
Contact information	Mark Jones, MD
Notes	**Funding sources:** Incyte Corporation, Merck Sharp & Dohme Corp.

NCT03592472

Study name	-
Methods	**Study type:** RCT, phase III **Blinding:** Double-blind (participant, investigator) **Accrual period:** July 17, 2018 - June 30, 2022 (estimated study completion date) **Countries:** multicentre study (38 locations in the US, China, Italy, Korea, Poland, Spain) **Cross-over study:** yes **Status**:- Recruiting (as of May 12, 2021)
Participants	**Estimated enrolment:** N = 413 **Inclusion criteria:** • Patients aged ≥ 18 years at time of study entry • Patients have histologically confirmed RCC with clear cell component • Patients have locally advanced and unresectable or metastatic disease • Measurable disease as assessed only by the investigator (not verified by IRC) according to RECIST version 1.1 • Patients must not have had any prior vascular endothelial growth factor (VEGF) tyrosine kinase inhibitor treatment in either (neo)adjuvant or locally advanced/metastatic setting. Up to 1 line of prior cytokine or immune checkpoint inhibitor treatment is allowed in either the (neo)adjuvant or metastatic setting provided screening scans indicate progressive disease (PD) during or following completion of treatment • Patients have Eastern Cooperative Oncology Group (ECOG) performance status of 0 or 1 • Patients have adequate baseline organ function.Patients have adequate baseline haematologic function • Patient must be at least 2 weeks from last systemic treatment or dose of radiation prior to date of randomisation **Exclusion criteria:** • Has persistent clinically significant toxicities (Grade ≥ 2; per NCI CTCAE version 5 from previous anticancer therapy (excluding alopecia which is permitted and excluding grades 2 and 3 laboratory abnormalities if they are not associated with symptoms, are not considered clinically significant by the investigator, and can be managed with available medical therapies) • Has untreated central nervous system (CNS) metastases. Patients with treated CNS metastases are eligible provided imaging demonstrates no new or progressive metastases obtained at least 4 weeks following completion of treatment. CNS imaging during Screening is not required unless clinically indicated • Has an additional malignancy requiring treatment within the past 3 years • Patients with the following concomitant neoplastic diagnoses are eligible: non-melanoma skin cancer, carcinoma in situ, and non-muscle invasive urothelial carcinoma • Poorly controlled hypertension, defined as systolic blood pressure ≥ 160 or diastolic blood pressure ≥ 100 mmHg. Use of anti-hypertensives and re-screening is permitted • A new pulmonary embolism or deep venous thrombosis diagnosed within 3 months prior to randomisation • Has a QTcF interval > 480 msec. New York Heart Association Class III or IV congestive heart failure • Use of prohibited medication within 7 days or 5 half-lives, whichever is shorter, prior to first dose of study drug
Interventions	**Experimental arm:** pazopanib plus abexinostat **Control arm:** pazopanib plus placebo

NCT03592472 *(Continued)*

Outcomes	**Primary outcome(s):** **Secondary outcome(s):** • PFS assessed by blinded Independent Review Committee (IRC) (time frame: from randomisation date to date of first documentation of progression OR death (up to approximately 4 years) • PFS by investigator assessment according to RECIST version 1.1. (time frame: from randomisation date to date of first documentation of progression OR death (up to approximately 4 years) • OS (time frame: from progression or end of study, every 3 months follow up until death, patient withdrawal from study follow-up, or study closure, whichever occurs first (up to approximately 4 years) • Adverse events by NCI CTCAE v. 5 (time frame: from Day 1 until end of treatment visit (up to approximately 4 years) • QoL, assessed by FKSI-19 and FACIT-F **Relevant to this review but not reported:** TFST, number of patients who discontinued treatment **Other outcomes (not relevant to this review):** ORR, DOR
Starting date	July 17, 2018
Contact information	-
Notes	**Funding source**: Xynomic Pharmaceuticals, Inc.

NCT03729245

Study name	BEMPEG
Methods	**Study type:** RCT, phase III, parallel assignment **Blinding:** no, open-label **Accrual period:** December 18, 2018 - June 2024 (estimated study completion date) **Countries:** 116 study locations **Cross-over study:** no **Status:** active, not recruiting (as of April 4, 2022)
Participants	**Estimated enrolment:** N=623 **Inclusion criteria:** • provide written, informed consent to participate in the study and follow the study procedures • Karnofsky Performance Status (KPS) of at least 70% • measurable disease per mRECIST 1.1 criteria • histologically confirmed RCC with a clear-cell component (may have sarcomatoid features); advanced (not amenable to curative surgery or radiation therapy) or metastatic (AJCC Stage IV) RCC • participants with any International Metastatic Renal Cell Carcinoma Database Consortium (IMDC) score (favourable-, intermediate-, or poor-risk) are eligible. At least one IMDC prognostic factor must be present to qualify as either intermediate- or poor-risk renal cell carcinoma • no prior systemic therapy (including neoadjuvant, adjuvant, or vaccine therapy) for RCC • participants with stable brain metastases following local treatment may be enrolled if certain criteria are met • tumour tissue (archival or fresh biopsy) identified and available for PD-L1 testing

NCT03729245 *(Continued)*

	• adequate organ function without growth factor or transfusion support **Exclusion criteria:** • an active, known or suspected autoimmune disease that has required systemic treatment within the past 3 months (exceptions exist) • participants who have a known additional malignancy that is progressing or requires active treatment (exceptions exist) • any tumour invading the wall of a major blood vessels • any tumour invading the gastrointestinal (GI) tract or any evidence of endotracheal or endobronchial tumour within 28 days prior to randomisation • need for >2 medications for management of hypertension (including diuretics) • history of pulmonary embolism, deep vein thrombosis (not including tumour thrombus), or clinically significant thromboembolic event within 3 months of randomisation Additional protocol defined inclusion/exclusion criteria and exceptions apply
Interventions	**Experimental arm:** Bempegaldesleukin + Nivolumab **Control arm:** Sunitinib or Cabozantinib
Outcomes	**Primary outcome(s):** • OS in IMDC participants (Time frame: 32-59 months) **Secondary outcome(s):** • PFS in participants (time frame: 32-59 months) • AEs (time frame: up to 5 years) • OS (time frame: 32-59 months) • QoLb (time frame: 32-59 months) **Relevant to this review but not reported:** SAEs, TFST, number of patients who discontinued treatment **Other outcomes (not relevant to this review):** ORR, changes in cancer- related symptoms
Starting date	18.12.2018
Contact information	
Notes	**Funding sources:** Nektar Therapeutics, Bristol-Myers Squib

NCT03793166

Study name	PDIGREE
Methods	**Study type:** RCT, phase III, parallel assignment **Blinding:** no, open-label **Accrual period:** May 9, 2019 - April 9, 2022 (estimated study completion date) **Countries:** multicentre, 804 study locations **Cross-over study:** no **Status:** Recruiting (as of June 3, 2022)

NCT03793166 *(Continued)*

Participants	**Estimated enrolment:** N=1046 **Inclusion criteria:** • histologically documented renal cell carcinoma with clear cell component, including patients who have sarcomatoid features • any metastatic disease, including visceral, lymph node, other soft tissue and bone, measurable per RECIST 1.1 • measurable disease as defined in the protocol • Must be intermediate or poor risk patient per International Metastatic Renal Cell Carcinoma Database (IMDC) criteria • Central nervous system (CNS) disease permitted, if stable and not otherwise causing symptoms or needing active treatment • Karnofsky performance status >= 70%. • no prior treatment with PD-1, PD-L1, or CTLA-4 targeting agents (including but not limited to nivolumab, pembrolizumab, pidilizumab, durvalumab, atezolizumab, tremelimumab, and ipilimumab), or any other drug or antibody specifically targeting T-cell co-stimulation or checkpoint pathways. The only exception is for prior treatment with nivolumab or other PD-1/PD-L1/CTLA-4 targeting therapy on pre- or post-operative trials, as long as > 1 year since completion of systemic therapy • no prior previous systemic therapy for renal cell carcinoma (prior HD IL-2 [> 28 days] and prior adjuvant sunitinib > 180 days since completion and prior immunotherapy as above are allowed) • no cancer therapy less than 28 days prior to registration; this includes radiation therapy, except for bone lesions less than 14 days prior to registration. There must be a complete recovery and no ongoing complications from radiotherapy • all sexes, age >= 18 years • STEP 2 registration eligibility criteria • successful completion of at least 1 cycle of ipilimumab/nivolumab • resolution of any treatment-related adverse events to grade 1 or less per dose modification section (this criteria does not include any adverse events [AEs] not attributable to treatment which are present due to disease). Exceptions for this criteria include patients receiving replacement hormone treatments (such as levothyroxine for treatment-related hypothyroidism or glucocorticoid replacement for adrenal insufficiency). Please contact study chair if further discussion is needed • no more than 70 days from last dose of ipilimumab/nivolumab **Exclusion criteria:** • active autoimmune disease requiring ongoing therapy • ongoing acute toxicity > grade 2 from previous treatment • major surgery less than 28 days prior to registration • significant cardiac ischemias events (ST elevation myocardial infarction [STEMI] or non-ST elevation myocardial infarction [NSTEMI]) within 6 months or active NY Heart Association class 3-4 heart failure symptom More inclusion criteria on CT.gov.
Interventions	**Experimental arm:** nivolumab + cabozantinib **Control arm:** nivolumab + ipilimumab
Outcomes	**Primary outcome(s):** • OS (time frame: from registration to date of death from any cause for non-randomised patients, from time of randomisation until death from any cause for randomised patients, assessed up to 5 years) **Secondary outcome(s):**

NCT03793166 *(Continued)*

	• PFS (time frame: from date of registration to date of progression or death from any cause, whichever occurs first, assessed up to 5 years) • proportion of participants who discontinue protocol-directed treatment (time frame: up to 5 years) • AEs (time frame: up to 5 years) **Relevant to this review but not reported:** QoL, SAEs, TFST **Other outcomes (not relevant to this review):** CR, OR
Starting date	09.05.2019
Contact information	Tian Zhang
Notes	**Funding sources:** National Cancer Institute (NCI)

NCT03873402

Study name	-
Methods	**Study type:** RCT, parallel assignment, phase IIIB **Blinding:** quadruple blinding (participant, care provider, investigator, outcomes assessor) **Accrual period:** April 29, 2019 - April 19, 2025 (estimated study completion date) **Countries:** multicentre (80 study locations) **Cross-over study:** no **Status:** active, not recruiting (as of February 10, 2022)
Participants	**Estimated enrolment:** N=418 **Inclusion criteria:** • histological confirmation of renal carcinoma with clear cell component including participants who may have sarcomatoid features • advanced (not amenable to curative surgery or radiation therapy) renal cell carcinoma (RCC) or metastatic RCC (mRCC) • measurable disease by computed tomography (CT) or magnetic resonance imaging (MRI) per Response Evaluation Criteria in Solid Tumors version 1.1 (RECIST 1.1) criteria • no prior systemic therapy for RCC • must be intermediate or poor risk as per International Metastatic RCC Database Consortium (IMDC) • all sexes, older than 18 years **Exclusion criteria:** • any active central nervous system (CNS) metastases • active, known, or suspected autoimmune disease • prior treatment with an anti-PD-1, anti-PD-L1, anti-PD-L2, anti-CD137, anti-CTLA-4 antibody, or any other agents specifically targeting T-cell co-stimulation or checkpoint pathways *Other protocol-defined inclusion/exclusion criteria apply
Interventions	**Experimental arm I:** Nivolumab **Experimental arm II:** Ipilimumab

NCT03873402 *(Continued)*

	Control arm: Ipilimumab placebo
Outcomes	**Primary outcome(s)** • PFS by blinded independent central review (BICR) (time frame: up to 34 months) **Secondary outcome(s)** • OS (time frame: up to 4 years) • PFS by investigator (time frame: up to 4 years) • Progression free survival secondary objective (PFS2) by investigator (time frame: up to 4 years9 • Incidence of AEs (time frame: up to 4 years) • Incidence of SAEs (time frame: up to 4 years) **Relevant to this review but not reported:** QoL, TFST, number of patients who discontinued treatment **Other outcomes (not relevant to this review):** ORR, DCR, DoR, TTR, clinical laboratory results
Starting date	April 29, 2019
Contact information	-
Notes	**Funding source:** Bristol-Myers Squibb

NCT03937219

Study name	**COSMIC-313**
Methods	**Study type:** RCT, parallel assignment, phase III **Blinding:** yes, double- blind **Accrual period:** June 25, 2019 - March 2025 (estimated study completion date) **Countries:** multicentre (167 study locations) **Cross-over study:** no **Status:** Active, not recruiting (as of March 10, 2022)
Participants	**Estimated enrolment:** N = 840 **Inclusion criteria:** • histologically confirmed advanced (not amenable to curative surgery or radiation therapy) or metastatic (AJCC Stage IV) renal cell carcinoma with a clear-cell component • intermediate- or poor-risk RCC as defined by International Metastatic RCC Database Consortium (IMDC) criteria • measurable disease per RECIST 1.1 as determined by the Investigator • Karnofsky Performance Status (KPS) ≥ 70%. • adequate organ and marrow function • all sexes, 18 years and older **Exclusion criteria:** • prior systemic anticancer therapy for unresectable locally advanced or metastatic RCC including investigational agents • uncontrolled, significant intercurrent or recent illness

NCT03937219 *(Continued)*

	• other clinically significant disorders such as: Autoimmune disease that has been symptomatic or required treatment within the past two years from the date of randomisation. Any condition requiring systemic treatment with either corticosteroids (> 10 mg daily prednisone equivalent) or other immunosuppressive medications within 14 days of randomisation. Known history of COVID-19 unless the participant has clinically recovered from the disease at least 30 days prior to randomisation • major surgery (e.g., nephrectomy, GI surgery, removal or biopsy of brain metastasis) within 4 weeks prior to randomisation • any other active malignancy at time of randomisation or diagnosis of another malignancy within 3 years prior to randomisation that requires active treatment, except for locally curable cancers that have been apparently cured, such as basal or squamous cell skin cancer, superficial bladder cancer, or carcinoma in situ of the prostate, cervix, or breast *Other protocol-defined inclusion/exclusion criteria apply
Interventions	**Experimental arm:** cabozantinib + nivolumab + ipilimumab (4 doses) followed by cabozantinib + nivolumab **Control arm:** Cabozantinib-matched placebo + nivolumab + ipilimumab (4 doses) followed by cabozantinib-matched placebo + nivolumab
Outcomes	**Primary outcome(s)** • PFS per RECIST 1.1 as determined by blinded independent radiology committee (time frame: up to 23 months after first participant randomised) **Secondary outcome(s)** • OS (time frame: up to 69 months after first participant randomised) **Relevant to this review but not reported:** QoL, AEs/SAEs, TFST, number of patients who discontinued treatment **Other outcomes (not relevant to this review):** -
Starting date	June 25, 2019
Contact information	-
Notes	**Funding source:** Exelixis

NCT04090710

Study name	**CYTOSHRINK**
Methods	**Study type:** RCT, parallel assignment, phase II **Blinding:** no, open-label **Accrual period:** January 29, 2020 - December 31, 2023 **Countries:** international (Australia, Canada), multicentre (7 study locations) **Cross-over study:** no **Status:** Recruiting (as of March 31, 2022)
Participants	**Estimated enrolment:** N=78 **Inclusion criteria:**

NCT04090710 *(Continued)*

- biopsy proven renal cell carcinoma of any histology
- imaging proven metastatic disease based on CT or MRI within 10 weeks of screening
- intermediate/poor risk disease based on IMDC criteria (see Appendix II)
- primary kidney lesion amenable to SBRT
- eligible for standard of care delivery of ipilimumab and nivolumab (I/N) according to approved product monograph
- all sexes, 18 years and older

Exclusion criteria:

- a maximum primary renal lesion size of 20 cm or greater
- candidate for cytoreductive nephrectomy, unless a patient has refused cytoreductive nephrectomy (in this case, a discussion of cytoreductive nephrectomy and patient refusal must be documented)
- treatment with prior systemic therapy in the adjuvant or metastatic setting for renal cell carcinoma
- Kanofsky Performance (KPS) score below 60 (see Appendix III)
- history of auto-immune disorder precluding treatment with ipilimumab or nivolumab
- chronic corticosteroid use or other chronic immune suppressive therapy. (Participants are permitted the use of topical, ocular, intra-articular, intranasal, and inhalational corticosteroids (with minimal systemic absorption). Adrenal replacement steroid doses of prednisone ≤ 10 mg daily are permitted)
- inability to lie flat for at least 30 minutes without moving

*Other protocol-defined inclusion/exclusion criteria apply

Interventions	**Experimental arm:** Radiation: SBRT + ipilimumab/nivolumab **Control arm:** ipilimumab/ nivolumab
Outcomes	**Primary outcome(s)** • PFS (time frame: 2 years) The primary outcome of this study is the hazard ratio for progression-free survival (PFS), defined from the date of randomisation until the date of progression (PFS truncated at subsequent systemic therapy) as determined by RECIST 1.1, or death due to any cause, whichever comes first **Secondary outcome(s)** • Pparticipant safety (AEs/SAEs) (time frame: date of randomisation until 1 year post treatment), using NCI CTCAE v5. and incidence and attribution of deaths • OS (time frame: 2 years) • QoL: EORTC QLQ-C30 questionnaire (time frame: 1 year), which will be evaluated using the EORTC QLQ-C30 questionnaire **Relevant to this review but not reported:** TFST, number of patients who discontinued treatment **Other outcomes (not relevant to this review):** ORR, drug tolerability, stool microbiome, blood immune signature changes
Starting date	January 29, 2020
Contact information	Ontario Clinical Oncology Group (OCOG)
Notes	**Funding source:** Ontario Clinical Oncology Group (OCOG)

NCT04203901

Study name	-
Methods	**Study type:** RCT, phase IIb **Blinding:** no, open-label **Accrual period:** July 22, 2020 - March, 2022 (estimated study completion date) **Countries:** national (USA), single- centre (Texas) **Cross-over study:** no **Status:** Recruiting (as of Februrary 11, 2022)
Participants	**Estimated enrolment:** N=120 **Inclusion criteria:** • age ≥ 18 years, all sexes • advanced disease histologically assessed as RCC, with predominantly clear cell histology • metastatic disease (measurable or non-measurable) that can be monitored throughout the course of study participation per iRECIST • participants who are candidates for standard first-line therapy • time from initial RCC diagnosis to initiation of systemic treatment (Nivolumab+Ipilimumab) of <1 year • Karnofsky Performance Status (KPS) ≥ 70% • resolution of all acute toxic effects of prior radiotherapy or surgical procedures to Grade ≤ 1 according to National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) Version 4.0 **Exclusion criteria:** • prior systemic therapy (including adjuvant or neoadjuvant) of any kind for RCC, including immunotherapy, chemotherapy, hormonal, or investigational therapy • prior history of malignancy within the preceding 3 years, except for adequately treated in situ carcinomas or non-melanoma skin cancer, adequately treated early stage breast cancer, superficial bladder cancer, and non-metastatic prostate cancer with a normal PSA • history of or known brain metastases, spinal cord compression, or carcinomatous meningitis, or evidence of brain or leptomeningeal disease • participants will be excluded if they have <2 of the following risk factors at Screening: Time from diagnosis to systemic treatment < 1 year Hgb < LLN Corrected calcium > 10.0 mg/dLKPS < 80% Neutrophils > ULNPlatelets > ULN • NCI CTCAE Grade 3 haemorrhage < 28 days before Visit 1 (Week 0) • any serious medical condition or illness considered by the investigator to constitute an unwarranted high risk for investigational treatment *Other protocol-defined inclusion/exclusion criteria apply
Interventions	**Experimental arm:** CMN-001 and Nivolumab+Ipilimumab (1st line therapy), Lenvatinib + Everolimus (2nd line therapy after progression) **Control arm:** Nivolumab+Ipilimumab (1st line therapy), Lenvatinib + Everolimus (2nd line therapy after progression)
Outcomes	**Primary outcome(s)** • OS (time frame: through study completion, an average of 2 years), participants will be followed for OS until the completion of the study **Secondary Outcome(s)**

NCT04203901 *(Continued)*

	• treatment emergent adverse events (TEAEs) between both arms (time frame: through study completion, an average of 2 years) • PFS (time frame: through study completion, an average of 2 years, assessed by the investigator per iRECIST) **Relevant to this review but not reported:** QoL, SAEs, TFST, number of patients who discontinued treatment **Other outcomes (not relevant to this review):** tumour response
Starting date	July 22, 2020
Contact information	CoImmune; Mark DeBenedette, PhD
Notes	**Funding source:** CoImmune

NCT04394975

Study name	-
Methods	**Study type:** RCT, sequential assignment, phase III **Blinding:** no, open-label **Accrual period:** August 20, 2020 - June 30, 2023 estimated study completion date) **Countries:** multicentre **Cross-over study:** no **Status:** Recruiting (as of January 21, 2021)
Participants	**Estimated enrolment:** N= 380 **Inclusion criteria:** • male or female with age ≥ 18 years and <80 years • have received no prior systemic therapy after previous metastasis for RCC, histologically confirmed diagnosis of unresectable, recurrent or metastatic RCC with clear cell component with or without sarcomatoid features, prior cytokine therapy was allowed • the IDMC score was medium to high risk • having at least one measurable disease per RECIST 1.1. Lesions situated in a previously irradiated area are considered measurable if re-progression has been demonstrated • provide archival tumour tissues or newly obtained biopsies if patients participate in the exploratory study • ECOG PS 0 or 1 • adequate function of vital organs **Exclusion criteria:** participants with any of the following conditions will not be included in the study: • prior Anti-PD-1, PD-L1 or CTLA-4 agents • prior systemic anti-cancer therapy after metastasis (e.g., VEGF/VEGFR or mTOR targeting agents, including (but not limited to) sunitinib, axitinib, sorafenib, pazopanib, cabozantinib, lenvatinib, bevacizumab or everolimus). • progression or recurrence during neoadjuvant/adjuvant therapy for renal cell cancer or within 12 months after the last dose treatment

NCT04394975 *(Continued)*

	• has participated or is currently participating in a trial of investigational agent within 4 weeks prior to the first dose of study treatment, unless observational (non-interventional) clinical study or follow-up period of interventional study • had major surgery (judged by investigators) within 4 weeks prior to the first dose of study treatment or has not recovered from prior surgery • requiring corticosteroids (Prednisone >10 mg/day or equivalent analogue) or other immunosuppressive agents within 2 weeks prior to the first dose of study treatment. Patients without active autoimmune disease using inhaled prednisone >10 mg/day will not be excluded from the study • has a history of organ transplantation or required long-term treatment with corticosteroids • has an additional malignancy that has progressed or required treatment within 5 years prior to randomisation • has a history of active central nervous system (CNS) metastasis or CNS metastasis had been confirmed by radiological examination (MRI or CT) at baseline within 30 days prior to the first dose of study drug • has current use (within 7 days of randomisation) or anticipated need for treatment drugs what are known strong CYP3A4/5 inhibitor and CYP3A4/5 inducer (including, but not limited to, carbamazepine, phenobarbital, phenytoin, rifabutin, rifampin and St. John's wort) or the drugs that are known with proarrhythmic potential (including, but not limited to, terfenadine, quinidine, procainamide, disopyramide, sotalol, probucol and benazapril, etc.) *Other protocol-defined inclusion/exclusion criteria apply
Interventions	**Experimental arm:** axitinib **Control arm: s**unitinib
Outcomes	**Primary outcome(s):** • PFS assessed by IRC per RECIST 1.1. (time frame: 3 years) **Secondary outcome(s):** • PFS assessed by investigators per RECIST 1.1 (time frame: 3 years) • overall survival rate (OSR) assessed by investigators and IRC per RECIST 1.1, respectively; (time frame: 3 years) • OS assessed by investigators and IRC per RECIST 1.1, respectively; (time frame: 3 years) • incidence and grade of AEs and SAEs per NCI-CTCAE version 5.0, incidence of ≥ grade 3 AE; (time frame: 3 years) **Relevant to this review but not reported:** QoL, TFST, number of patients who discontinued treatment **Other outcomes (not relevant to this review):** ORR, DoR, DCR, biomarkers, incidence and grade of AEs and SAEs related to study drugs
Starting date	August 20, 2020
Contact information	Shanghai Junshi Bioscience Co., Ltd., Fugui Wang
Notes	**Funding source:** Shanghai Junshi Bioscience Co., Ltd.

NCT04523272

Study name	-
Methods	**Study type:** RCT, parallel assignment, phase III

NCT04523272 *(Continued)*

Blinding: no, open-label

Accrual period: August 25, 2020 - June 2023

Countries: multicentre (26 study locations)

Cross-over study: no

Status: Recruiting (as of September 10, 2020)

Participants

Estimated enrolment: N = 418

Inclusion criteria:

- histopathologically confirmed renal clear cell cancer, including advanced renal cell carcinoma with clear cell components
- has not received systemic therapy for local advanced/metastatic disease
- at least has one measurable lesion
- Eastern Cooperative Oncology Group (ECOG) performance status of 0 to 1; Life expectancy ≥ 3 months
- adequate laboratory indicators
- agree to provide at least 5 slices tumour tissue samples for biomarker detection
- serum or urine pregnancy tests are negative within 7 days before randomisation; Men and women should agree to use effective contraception during the study period and after the end of the study period within 6 months
- all sexes, 18 - 80 years old

Exclusion criteria:

- has symptomatic central nervous system (CNS) disease and / or cancerous meningitis, pia mater disease
- has received anti-angiogenesis targeted therapy or targeted PD-1 and PD-L1 immunotherapy
- has active virus, bacteria, fungal infection; cardiovascular and cerebrovascular diseases; gastrointestinal abnormalities; Immunodeficiency; bleeding risk; lung disease; neurological or psychiatric disorders
- has participated in other clinical trials within 30 days before randomisation
- has received attenuated live vaccine within 28 days before randomisation or planned to received attenuated live vaccine during the study period

*Other protocol-defined inclusion/exclusion criteria apply

Interventions

Experimental arm: TQB2450 + anlotinib

Control arm: sunitinib mMalate capsules

Outcomes

Primary outcome(s):

- PFS evaluated by Independent Review Committee(IRC) [Time frame: up to 60 weeks] PFS defined as the time from randomisation until the first documented progressive disease (PD) or death from any cause, based on IRC

Secondary outcome(s):

- Progression-free survival (PFS) evaluated by investigator (time frame: up to 60 weeks)
- OS (time frame: up to 60 weeks)
- PFS at 12 months (time frame: up to 12 months)
- OS at 12 months (time frame: up to 12 months)
- OS at 24 months (time frame: up to 24 months)

Relevant to this review but not reported: QoL, AEs/SAEs, TFST, number of patients who discontinued treatment

NCT04523272 *(Continued)*

	Other outcomes (not relevant to this review): DCR, DoR,
Starting date	August 25, 2020
Contact information	Jun Guo, Doctor
Notes	**Funding source:** Chia Tai Tianqing Pharmaceutical Group Co., Ltd

NCT04540705

Study name	PIVOT IO 011
Methods	**Study type:** RCT, parallel assignment, phase 1/2 study **Blinding:** no, open-label **Accrual period:** September 11, 2020 - January 17, 2026 (estimated study completion date) **Countries:** international (5 countries: Brazil, Argentina, USA, Spain, Canada,), multicentre **Cross-over study:** no **Status:** active, not recruiting (as of May 18, 2022)
Participants	**Estimated enrolment:** N = 250 **Inclusion criteria:** • histological confirmation of renal cell carcinoma (RCC) with clear cell component including participants who may also have sarcomatoid features • advanced (not amenable to curative surgery or radiation therapy) or metastatic (American Joint Committee on Cancer (AJCC) Stage 4) RCC • no prior systemic therapy, including prior PD-L1 therapy, for RCC is allowed with the following exception: i) One prior adjuvant or neoadjuvant therapy for completely resectable RCC is allowed. Therapy must have included an agent that targets vascular endothelial growth factor (VEGF) pathway or VEGF receptors and recurrence must have occurred at least 6 months after the last dose of adjuvant or neoadjuvant therapy • life expectancy ≥ 12 weeks • Karnofsky Performance Status (KPS) of at least 70% • measurable disease by computed tomography (CT) or magnetic resonance imaging (MRI) per RECIST 1.1 criteria • males and females must agree to follow specific methods of contraception, if applicable • all sexes, older than 18 years **Exclusion criteria:** • active CNS brain metastases or leptomeningeal metastases • active, known or suspected autoimmune disease • inadequately treated adrenal insufficiency • history of pulmonary embolism (PE), deep vein thrombosis (DVT), or prior clinically significant venous or non-CVA/TIA arterial thromboembolic event (e.g., internal jugular vein thrombosis) within 3 months prior to treatment assignment (part 1) and randomisation (part 2) *Other protocol-defined inclusion/exclusion criteria apply
Interventions	**Experimental arm I:** part 1A (part 1): nivolumab + bempegaldesleukin + bxitinib **Experimental arm II:** part 1B (part 1): nivolumab + bempegaldesleukin + cabozantinib

NCT04540705 *(Continued)*

	Experimental arm III: arm A (part 2): nivolumab + bempegaldesleukin + cabozantinib **Control arm:** arm B (part 2): nivolumab + cabozantinib
Outcomes	**Primary outcome(s) :** • incidence of AEs by severity (part 1) (time frame: up to 5 years) • incidence of SAEs (part 1) (time frame: up to 5 years) • incidence of AEs leading to discontinuation (part 1) (time frame: up to 5 years) • incidence of immune-mediated adverse events (imAEs) (part 1) (time frame: up to 5 years) **Secondary outcome(s) :** • PFS by RECIST 1.1 by Investigator (part 2) (time frame: up to 32 months from start of part 2) • OS (part 2) (time frame: up to 60 months • incidence of AEs by severity (part 2) (time frame: up to 5 years) • incidence of SAEs (part 2) (time frame: up to 5 years) • incidence of AEs leading to discontinuation (part 2) (time frame: up to 5 years) • incidence of imAEs (part 2) (time frame: up to 5 years) **Relevant to this review but not reported:** QoL, TFST, number of patients who discontinued treatment **Other outcomes (not relevant to this review):** DLTs, laboratory results, ORR
Starting date	September 11, 2020
Contact information	Bristol-Myers Squibb
Notes	**Funding sources:** Bristol-Myers Squibb

NCT04736706

Study name	-
Methods	**Study type:** RCT, parallel assignment, phase III **Blinding:** no, open-label **Accrual period:** April 14, 2021 - October 29, 2026 (estimated study completion date) **Countries:** international (USA, Australia, Chile, Czechia, Denmark, Finland, Guatemala, Hungary, Korea, Norway, Poland, Russia, Spain, Sweden, Taiwan, Turkey, Ukraine), multicentre (92 study locations) **Cross-over study:** no **Status:** Recruiting (as of May 31, 2022)
Participants	**Estimated enrolment:** N = 1431 **Inclusion criteria:** • has histologically confirmed diagnosis of RCC with clear cell component • has received no prior systemic therapy for advanced ccRCC • dose of lenvatinib or belzutifan, whichever occurs last • has adequately controlled blood pressure with or without antihypertensive medications • has adequate organ function

NCT04736706 *(Continued)*

- participants receiving bone resorptive therapy must have therapy initiated at least 2 weeks prior to randomisation/allocation

Exclusion criteria:

- has a known additional malignancy that is progressing or has required active treatment within the past 3 years
- has had major surgery, other than nephrectomy within 4 weeks prior to randomisation
- has known central nervous system (CNS) metastases and/or carcinomatous meningitis
- has received prior radiotherapy within 2 weeks prior to first dose of study intervention
- has hypoxia or requires intermittent supplemental oxygen or requires chronic supplemental oxygen
- has clinically significant cardiac disease within 12 months from first dose of study intervention
- has a history of interstitial lung disease
- has symptomatic pleural effusion; a participant who is clinically stable following treatment of this condition is eligible
- has preexisting gastrointestinal or non-gastrointestinal fistula
- has a diagnosis of immunodeficiency or is receiving chronic systemic steroid therapy or any other form of immunosuppressive therapy within 7 days prior to the first dose of study treatment
- has a known psychiatric or substance abuse disorder that would interfere with requirements of the study
- has received a live or live-attenuated vaccine within 30 days before the first dose of study drug; killed vaccines are allowed
- has an active autoimmune disease that has required systemic treatment in the past 2 years
- has a history of noninfectious pneumonitis that required steroids or has current pneumonitis
- has an active infection requiring systemic therapy
- has radiographic evidence of intratumoural cavitation, encasement or invasion of a major blood vessel
- has clinically significant history of bleeding within 3 months prior to randomisation
- has had an allogenic tissue/solid organ transplant

*Other protocol-defined inclusion/exclusion criteria apply

Interventions	**Experimental arm I:** pembrolizumab + belzutifan + lenvatinib **Experimental arm II:** pembrolizumab/quavonlimab + lenvatinib **Control arm:** pembrolizumab + lenvatinib
Outcomes	**Primary outcome(s) :** • PFS according to RECIST 1.1 as assessed by blinded independent central review (BICR) (time frame: up to approximately 46 months) • OS (time frame: up to approximately 66 months) **Secondary outcome(s):** • number of participants who experienced at least one AE (time frame: up to approximately 66 months) • number of participants who discontinue study treatment due to an AE (time frame: up to approximately 66 months) **Relevant to this review but not reported:** QoL, TFST, SAEs **Other outcomes (not relevant to this review):** ORR, DoR
Starting date	April 14, 2021

NCT04736706 *(Continued)*

Contact information	Merck Sharp & Dohme Corp.
Notes	**Funding sources:** Merck Sharp & Dohme Corp., Eisai Inc.

NCT05043090

Study name	**SAMETA**
Methods	**Study type:** RCT, phase III, parallel assignment **Blinding:** no, open-label **Accrual period:** October 28, 2021 - June 9, 2025 (estimated study completion date) **Countries:** multicentre (172 locations in the USA, Argentina, Australia, Brazil, Canada, Chile, Czech Republic, France, Germany, China, India, Isreal, Italy, Korea, Mexico, the Netherlands, Poland, Romania, Russia, Singapore, Spain, Taiwan, Turkey, Ukraine, the UK) **Cross-over study:** no **Status:** recruiting (as of May 17, 2022)
Participants	**Estimated enrolment:** N=220 **Inclusion criteria:** • histologically confirmed unresectable and locally advanced or metastatic PRCC • PRCC must be centrally confirmed as MET-driven using a sponsor-designated central laboratory validated NGS assay • No prior systemic anti-cancer treatment in the metastatic setting; no prior exposure to MET inhibitors, Durvalumab or Sunitinib in any setting • Karnofsky Score >70 • at least one lesion, not previously irradiated, that can be accurately measured at baseline • adequate organ and bone marrow function • ;life expectancy ≥12weeks at Day 1 **Exclusion criteria:** • history of liver cirrhosis of any origin and clinical stage; or history of other serious liver disease or chronic disease with relevant liver involvement, with or without normal LFTs • spinal cord compression or brain metastases, unless asymptomatic and stable on treatment for at least 14 days prior to study intervention • active or prior cardiac disease (within past 6 months) or clinically significant ECG abnormalities and/or factors/medications that may affect QT and/or QTc intervals • active infection including HIV, TB, HBV and HCV • active or prior documented autoimmune or inflammatory disorders • receipt of live attenuated vaccine within 30 days prior to the first dose of study intervention
Interventions	**Experimental arm I:** savolitinib + durvalumab **Experimental arm II:** durvalumab **Control arm: s**unitinib
Outcomes	**Primary outcome(s) :** • PFS assessed by BICR - savolitinib plus durvalumab relative to sunitinib (time frame: approximately 28 months post first participant randomised)

NCT05043090 *(Continued)*

	• OS - savolitinib plus durvalumab relative to sunitinib (time frame: approximately 28 months and approximately 42 months post first participant randomised) **Secondary outcome(s):** • PFS assessed by BICR - savolitinib plus durvalumab relative to durvalumab monotherapy (time frame: approximately 28 months post first participant randomised) • Assessment of patient-reported symptoms, functioning, and HRQoL **Relevant to this review but not reported:** TFST, AEs, SAEs, number of participants who discontinued treatment due to an AE **Other outcomes (not relevant to this review):** ORR, DoR, DCR
Starting date	October 28, 2021
Contact information	Toni Choueiri, Dana-Farber Cancer Institute
Notes	**Funding source:** AstraZeneca

NCT05096390

Study name	**PAXIPEM**
Methods	**Study type:** RCT, phase II, parallel assignment **Blinding:** no, open-label **Accrual period:** December 2021 - December 2025 (estimated study completion date) **Countries:** multicentre (11 locations in France) **Cross-over study:** no **Status:** Not yet recruiting (as of October 27, 2021)
Participants	**Estimated enrolment:** N = 72 **Inclusion criteria:** • Age ≥ 18 years on the day of signing informed consent • metastatic or locally advanced (inoperable) type 2 or mixed PRCC, histologically confirmed by central review: FFPE blocks (or all HES and IHC slides) with the initial histology report must be sent for central reading before confirmation of inclusion in the study • no prior systemic treatment for renal cancer (chemotherapy, immunotherapy, anti-angiogenic drugs, or treatment under evaluation) even in adjuvant setting • at least one measurable site of disease according to RECIST v1.1 • Eastern Cooperative Oncology Group (ECOG) Performance Status (PS) ≤ 1 evaluated within 7 days prior to the date of inclusion • in case of prior radiation therapy, discontinuation of irradiation for at least 3 weeks before first dose of study treatment, with at least 1 site kept/preserved for evaluation. participants must have recovered from all radiation-related toxicities, not require corticosteroids, and not have had radiation pneumonitis. A 1-week washout is permitted for palliative radiation (≤ 2 weeks - limited field (<10% of the whole body)) to non-CNS disease **Exclusion criteria:** • presence of brain metastases on Magnetic Resonance Imaging (MRI) or Computed Tomography-scan (CT-scan) performed within 28 days prior to inclusion. Patients with a history of brain

NCT05096390 *(Continued)*

	metastases treated by surgery or stereotactic surgery, with normal brain MRI or CT-scan are allowed to participate • metastases with high risk of nervous compression or bone lesion with high risk of fracture • Prior history of other malignancies other than PRCC (except for curatively treated basal cell or squamous cell carcinoma of the skin or in situ uterine cervix carcinoma) unless the participant has been free of the disease for at least 5 years • major surgical procedure, open biopsy, or serious none healing wound within 28 days prior to inclusion • significant cardiovascular disease • Any anti-coagulation therapy except prophylactic low dose More inclusion & exclusion criteria on CT.gov.
Interventions	**Experimental arm:** axitinib + pembrolizumab **Control arm:** axitinib monotherapy
Outcomes	**Primary outcome(s) :** • Efficacy of axitinib + pembrolizumab versus axitinib in patients with locally advanced or metastatic type 2 papillary renal carcinoma in first-line treatment (time Frame: at 6 months for each patient) Secondary outcome(s): • PFS (time frame: up to 24 months for each patient) • OS (time frame: up to 48 months) • Incidence of adverse events (time frame: up to 48 months) **Relevant to this review but not reported:** TFST, QoL, number of participants who discontinued treatment due to an AE **Other outcomes (not relevant to this review):** DoR, BoR
Starting date	December 2021
Contact information	Sylvie Negrier
Notes	**Funding source:** Centre Leon Berard

UMIN 000012522

Study name	**ESCAPE**
Methods	**Study type:** RCT, phase III, parallel assignment **Blinding:** no, open-label **Accrual period:** no information **Countries:** national (Japan) **Cross-over study:** no information **Status:** recruiting
Participants	**Estimated enrolment:** N = 144 **Inclusion criteria:**

UMIN 000012522 *(Continued)*

	• participants who have already performed nephrectomy with metastatic renal cell carcinoma (RCC) • participants with confirmed clear cell RCC • participants who had not received any prior systemic treatment for metastatic RCC • participants with the Memorial Sloan-Kettering Cancer Center (MSKCC) risk criteria of favourable or intermediate • participants who have at least one measurable lesion on CT or MRI at baseline as per the RECIST 1.1 criteria • age: 20-80 years old, both inclusive • participants with ECOG performance status of 0 or 1 • participants who are expected to have more than 3 months of life expectancy **Exclusion criteria:** • participants with history of hypersensitivity against IFN, IL-2, sunitinib, or axitinib • participants with a history of hypersensitivity to biological preparations such as vaccines • participants having Shou-Sai-Kotou (special herbal drug) • participants with autoimmune hepatitis • participants with a history of interstitial pneumonia • participants treated for another primary malignancy within 3 years of enrolment • participants judged ineligible to participate in the study by the investigator *Other protocol-defined inclusion/exclusion criteria apply
Interventions	**Experimental arm:** cytokines (IL-2+ IFN) as 1st-line followed by 2nd-line axitinib **Control arm:** sunitinib as 1st-line followed by 2nd-line axitinib
Outcomes	**Primary outcome(s):** • PFS from randomisation to progression or death during second-line therapy (total PFS) of cytokines followed by axitinib is superior compared to sunitinib followed by axitinib **Secondary outcome(s):** • OS, descriptively in each arm • PFS in 1st-line and 2nd-line treatment, descriptively in each arm • safety 1st-line treatment, descriptively in each arm (AEs/SAEs) • health-related Quality-of-life (HRQOL) in 1st-line and 2nd-line treatment, descriptively in each arm **Relevant to this review but not reported:** TFST, number of patients who discontinued treatment **Other outcomes (not relevant to this review):** ORR, TTF, DCR
Starting date	-
Contact information	Kanazawa University Hospital
Notes	**Funding sources:** Innovative Clinical Research Center, Kanazawa University Hospital

AEs: adverse events **CNS**: central nervous system'; DVT: deep vein thrombosis; **MRI:** magnetic resonance imaging;**mTOR**: mechanistic target of rapamycin ;**OS: overall survival ; PFS**: progression-free survival**; QoL:** quality of life; **RCC:** renal cell carcinoma; **RRCT:** randomised controlled trial;**SAEs:** serious adverse events.

ADDITIONAL TABLES

Table 1. NMA results for OS (combined risk groups)

Results of network meta-analysis for outcome overall survival (combined risk groups). Treatments are ordered by P-Score (descending). Only subnetworks with >1 designs. Upper triangle: direct estimate. Lower triangle: network estimate.

Subnet 1

No. of studies: 19. No. of pairwise comparisons: 19. No. of treatments: 19. No. of designs: 18

Heterogeneity/Inconsistency: Q=1.81, df=1, P = 0.18; I^2 = 44.6%, Tau^2 = 0.0284

Treatment Effects + 95%-CIs (Hazard Ratios, random effects model):

LEN+PEM	.	.	.	.	.	.	.	0.66 [0.42, 1.03]	.	.	.	.	.	.	.	.	.	.	.
0.96 [0.54, 1.70]	NIV+IPI	.	.	.	.	.	.	0.69 [0.48, 1.00]	.	.	.	.	.	.	.	.	.	.	.
0.90 [0.50, 1.62]	0.95 [0.56, 1.60]	PEM +AXI	.	.	.	.	.	0.73 [0.50, 1.07]	.	.	.	.	.	.	.	.	.	.	.
0.79 [0.35, 1.75]	0.82 [0.38, 1.76]	0.87 [0.40, 1.88]	CAB	.	.	.	.	0.84 [0.43, 1.64]	.	.	.	.	.	.	.	.	.	.	.
0.72 [0.41, 1.28]	0.75 [0.45, 1.26]	0.80 [0.47, 1.35]	0.92 [0.43, 1.97]	PAZ	1.01 [0.63, 1.62]	.	.	0.92 [0.64, 1.32]	.	.	.	.	.	.	.	.	.	.	.
0.73 [0.35, 1.53]	0.76 [0.38, 1.53]	0.81 [0.40, 1.64]	0.93 [0.38, 2.28]	1.01 [0.63, 1.62]	PLA	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.72 [0.33, 1.54]	0.75 [0.36, 1.55]	0.79 [0.38, 1.65]	0.91 [0.37, 2.28]	0.99 [0.48, 2.05]	0.98 [0.41, 2.34]	NIN	.	0.92 [0.49, 1.72]	.	.	.	.	.	.	.	.	.	.	.
0.67 [0.19, 2.41]	0.70 [0.20, 2.46]	0.74 [0.21, 2.61]	0.86 [0.22, 3.38]	0.93 [0.27, 3.26]	0.92 [0.24, 3.52]	0.94 [0.24, 3.62]	TEM	0.98 [0.30, 3.24]	.	.	.	.	.	.	.	.	.	.	.

Table 1. NMA results for OS (combined risk groups) *(Continued)*

0.66 [0.42, 1.03]	0.69 [0.48, 1.00]	0.73 [0.50, 1.07]	0.84 [0.43, 1.64]	0.91 [0.64, 1.32]	0.91 [0.50, 1.65]	0.92 [0.49, 1.72]	0.98 [0.30, 3.24]	SUN	0.99 [0.72, 1.36]	.	.	.	0.94 [0.52, 1.70]	0.89 [0.36, 2.21]	.	0.87 [0.57, 1.33]	0.82 [0.56, 1.21]	.
0.65 [0.38, 1.13]	0.68 [0.42, 1.11]	0.72 [0.44, 1.19]	0.83 [0.40, 1.75]	0.91 [0.56, 1.47]	0.90 [0.46, 1.77]	0.91 [0.45, 1.84]	0.97 [0.28, 3.35]	0.99 [0.72, 1.36]	ATE +BEV	.	.	.	.	.	.	.	.	.
0.64 [0.28, 1.44]	0.67 [0.30, 1.45]	0.70 [0.32, 1.55]	0.81 [0.31, 2.12]	0.88 [0.40, 1.92]	0.87 [0.35, 2.17]	0.89 [0.35, 2.25]	0.95 [0.24, 3.76]	0.96 [0.48, 1.92]	0.97 [0.46, 2.07]	EVE +BEV	0.99 [0.64, 1.53]	.	.	.	.	.	.	.
0.63 [0.32, 1.26]	0.66 [0.35, 1.26]	0.70 [0.36, 1.34]	0.80 [0.34, 1.89]	0.87 [0.46, 1.66]	0.86 [0.39, 1.93]	0.88 [0.39, 2.00]	0.94 [0.25, 3.47]	0.95 [0.56, 1.63]	0.96 [0.52, 1.79]	0.99 [0.64, 1.53]	IFN +BEV	1.00 [0.69, 1.46]	.	.	0.91 [0.62, 1.33]	.	0.86 [0.60, 1.24]	.
0.63 [0.29, 1.39]	0.66 [0.31, 1.39]	0.70 [0.33, 1.49]	0.80 [0.31, 2.04]	0.87 [0.41, 1.84]	0.86 [0.36, 2.10]	0.88 [0.36, 2.17]	0.94 [0.24, 3.66]	0.95 [0.50, 1.83]	0.96 [0.47, 1.99]	0.99 [0.56, 1.76]	1.00 [0.69, 1.46]	TEM +BEV	.	.	.	.	.	.
0.62 [0.30, 1.30]	0.65 [0.32, 1.30]	0.69 [0.34, 1.39]	0.79 [0.32, 1.94]	0.86 [0.43, 1.73]	0.85 [0.37, 1.98]	0.87 [0.37, 2.05]	0.92 [0.24, 3.51]	0.94 [0.52, 1.70]	0.95 [0.49, 1.86]	0.98 [0.39, 2.42]	0.99 [0.45, 2.19]	0.99 [0.41, 2.38]	ATE	.	.	.	.	.
0.59 [0.22, 1.61]	0.62 [0.23, 1.64]	0.65 [0.24, 1.74]	0.75 [0.24, 2.31]	0.82 [0.31, 2.17]	0.81 [0.27, 2.39]	0.82 [0.27, 2.47]	0.87 [0.20, 3.92]	0.89 [0.36, 2.21]	0.90 [0.34, 2.35]	0.93 [0.30, 2.89]	0.94 [0.33, 2.67]	0.94 [0.31, 2.86]	0.95 [0.32, 2.79]	EVE	.	.	.	.
0.57 [0.26, 1.26]	0.60 [0.28, 1.27]	0.63 [0.30, 1.35]	0.73 [0.29, 1.86]	0.79 [0.38, 1.68]	0.79 [0.32, 1.91]	0.80 [0.32, 1.97]	0.85 [0.22, 3.33]	0.87 [0.45, 1.67]	0.88 [0.42, 1.81]	0.90 [0.51, 1.61]	0.91 [0.62, 1.33]	0.91 [0.53, 1.55]	0.92 [0.38, 2.22]	0.97 [0.32, 2.97]	IFN +PLA	.	.	.
0.57 [0.31, 1.06]	0.60 [0.34, 1.05]	0.63 [0.36, 1.12]	0.73 [0.33, 1.62]	0.80 [0.46, 1.39]	0.79 [0.38, 1.64]	0.80 [0.38, 1.70]	0.85 [0.24, 3.03]	0.87 [0.57, 1.33]	0.88 [0.52, 1.49]	0.90 [0.40, 2.03]	0.91 [0.46, 1.80]	0.91 [0.42, 1.99]	0.92 [0.45, 1.91]	0.97 [0.36, 2.65]	1.00 [0.46, 2.18]	LEN +EVE	.	.
0.54 [0.30, 0.97]	0.57 [0.33, 0.96]	0.60 [0.35, 1.03]	0.69 [0.32, 1.49]	0.75 [0.44, 1.28]	0.74 [0.37, 1.51]	0.76 [0.36, 1.57]	0.80 [0.23, 2.83]	0.82 [0.56, 1.21]	0.83 [0.50, 1.36]	0.85 [0.48, 1.51]	0.86 [0.60, 1.24]	0.86 [0.51, 1.46]	0.87 [0.43, 1.76]	0.92 [0.34, 2.46]	0.95 [0.56, 1.60]	0.94 [0.53, 1.67]	IFN	0.93 [0.63, 1.37]
0.50 [0.25, 1.01]	0.52 [0.27, 1.01]	0.55 [0.28, 1.08]	0.64 [0.27, 1.52]	0.70 [0.36, 1.34]	0.69 [0.31, 1.55]	0.70 [0.30, 1.61]	0.74 [0.20, 2.78]	0.76 [0.44, 1.31]	0.77 [0.41, 1.44]	0.79 [0.39, 1.57]	0.80 [0.47, 1.36]	0.80 [0.41, 1.53]	0.81 [0.36, 1.80]	0.85 [0.30, 2.45]	0.88 [0.45, 1.69]	0.87 [0.44, 1.75]	0.93 [0.63, 1.37]	NAP +IFN

Table 1. NMA results for OS (combined risk groups) *(Continued)*

Subnet 2

No. of studies: 3. No. of pairwise comparisons: 3. No. of treatments: 4. No. of designs: 3

Heterogeneity/Inconsistency: Q=0, df=0, p=n.a.; I^2=n.a., Tau^2=n.a.

Treatment Effects + 95%-CIs (Hazard Ratios, random effects model):

ILN+SOR	.	0.91 [0.59, 1.41]	.
0.91 [0.54, 1.56]	AXI	1.00 [0.73, 1.36]	.
0.91 [0.59, 1.41]	0.99 [0.73, 1.36]	SOR	0.51 [0.22, 1.19]
0.47 [0.18, 1.21]	0.51 [0.21, 1.26]	0.51 [0.22, 1.19]	SOR+IFN

Table 2. NMA results for OS (MSKCC favourable risk groups)

Results of network meta-analysis for outcome overall survival (MSKCC favourable risk group). Treatments are ordered by P-Score (descending). Only subnetworks with >1 designs. Upper triangle: direct estimate. Lower triangle: network estimate.

Subnet 1

No. of studies: 3. No. of pairwise comparisons: 3. No. of treatments: 4. No. of designs: 3

Heterogeneity/Inconsistency: Q = 0, df = 0, p=n.a.; I^2=n.a., Tau^2=n.a.

Treatment Effects + 95%-CIs (Hazard ratios, random-effects model):

LEN+EVE	.	.	0.54 [0.21, 1.37]
0.62 [0.23, 1.65]	PAZ	.	0.88 [0.63, 1.21]
0.63 [0.18, 2.16]	1.02 [0.43, 2.44]	LEN+PEM	0.86 [0.38, 1.93]
0.54 [0.21, 1.37]	0.88 [0.63, 1.21]	0.86 [0.38, 1.93]	SUN

Subnet 2

No. of studies: 3. No. of pairwise comparisons: 3. No. of treatments: 4. No. of designs: 3

Heterogeneity/Inconsistency: Q = 0, df = 0, P = n.a.; I^2=n.a., Tau^2 = n.a.

Treatment Effects + 95%-CIs (Hazard Ratios, random effects model):

IFN+BEV	.	0.92 [0.62, 1.37]	0.90 [0.64, 1.25]
0.93 [0.58, 1.49]	NAP+IFN	.	0.96 [0.69, 1.34]
0.92 [0.62, 1.37]	0.99 [0.53, 1.83]	IFN+PLA	.
0.89 [0.64, 1.25]	0.96 [0.69, 1.34]	0.97 [0.58, 1.64]	IFN

Table 3. NMA results for OS (IMDC favourable risk group)

Results of network meta-analysis for outcome overall survival (IMDC favourable risk group). Treatments are ordered by P-Score (descending). Only subnetworks with >1 designs. Upper triangle: direct estimate. Lower triangle: network estimate.

Subnet 1

No. of studies: 5. No. of pairwise comparisons: 5. No. of treatments: 6. No. of designs: 5

Heterogeneity/Inconsistency: Q = 0, df = 0, P = n.a.; I^2=n.a., Tau^2 = n.a.

Treatment Effects + 95%-CIs (Hazard ratios, random-effects model):

AVE+AXI	.	.	.	0.66 [0.36, 1.22]	.
0.71 [0.34, 1.48]	NIV+IPI	.	.	0.93 [0.62, 1.40]	.
0.70 [0.27, 1.80]	0.99 [0.43, 2.25]	NIV+CAB	.	0.94 [0.46, 1.92]	.
0.65 [0.24, 1.77]	0.92 [0.38, 2.22]	0.93 [0.32, 2.68]	LEN+EVE	1.01 [0.46, 2.20]	.

Table 3. NMA results for OS (IMDC favourable risk group) *(Continued)*

0.66 [0.36, 1.22]	0.93 [0.62, 1.40]	0.94 [0.46, 1.92]	1.01 [0.46, 2.20]	SUN	0.87 [0.42, 1.82]
0.57 [0.22, 1.50]	0.81 [0.35, 1.88]	0.82 [0.29, 2.28]	0.88 [0.30, 2.57]	0.87 [0.42, 1.82]	LEN+PEM

Table 4. NMA results for OS (MSKCC intermediate and poor risk groups)

Results of network meta-analysis for outcome overall survival (MSKCC intermediate and poor risk groups). Treatments are ordered by P-Score (descending). Only subnetworks with >1 designs. Upper triangle: direct estimate. Lower triangle: network estimate.

No. of studies: 15. No. of pairwise comparisons: 15. No. of treatments: 10. No. of designs: 9

Heterogeneity/Inconsistency: Q = 1.45, df = 6, P = 0.96; I^2 = 0%, Tau^2 = 0.0

Treatment Effects + 95%-CIs (Hazard ratios, random-effects model):

LEN+PEM	.	0.63 [0.46, 0.86]	.	.	.	.	.	.	.
0.71 [0.49, 1.02]	PAZ	0.89 [0.75, 1.06]	.	.	.	.	.	.	.
0.63 [0.46, 0.86]	0.89 [0.75, 1.06]	SUN	.	.	.	0.81 [0.61, 1.07]	0.77 [0.61, 0.96]	.	.
0.62 [0.40, 0.97]	0.88 [0.61, 1.25]	0.98 [0.72, 1.35]	TEM	.	.	.	0.78 [0.63, 0.97]	.	.
0.57 [0.37, 0.88]	0.80 [0.57, 1.13]	0.91 [0.67, 1.21]	0.92 [0.69, 1.22]	IFN+BEV	.	.	0.85 [0.70, 1.02]	0.83 [0.67, 1.04]	.
0.52 [0.33, 0.81]	0.73 [0.51, 1.05]	0.83 [0.61, 1.13]	0.84 [0.62, 1.14]	0.91 [0.69, 1.21]	IFN+TEM	.	0.93 [0.75, 1.15]	.	.
0.51 [0.33, 0.77]	0.72 [0.51, 1.00]	0.81 [0.61, 1.07]	0.82 [0.54, 1.24]	0.89 [0.59, 1.34]	0.98 [0.64, 1.48]	LEN+EVE	.	.	.
0.48 [0.33, 0.71]	0.68 [0.51, 0.91]	0.77 [0.61, 0.96]	0.78 [0.63, 0.97]	0.85 [0.70, 1.02]	0.93 [0.75, 1.15]	0.95 [0.67, 1.37]	IFN	.	0.88 [0.68, 1.15]
0.48 [0.29, 0.77]	0.67 [0.45, 1.01]	0.75 [0.52, 1.09]	0.77 [0.53, 1.10]	0.83 [0.67, 1.04]	0.91 [0.64, 1.31]	0.94 [0.59, 1.49]	0.98 [0.74, 1.31]	IFN+PLA	.
0.43 [0.27, 0.68]	0.60 [0.41, 0.89]	0.68 [0.48, 0.96]	0.69 [0.49, 0.97]	0.75 [0.55, 1.03]	0.82 [0.59, 1.15]	0.84 [0.54, 1.32]	0.88 [0.68, 1.15]	0.90 [0.61, 1.33]	NAP+IFN

Table 5. NMA results for OS (IMDC intermediate and poor risk groups)

Results of network meta-analysis for outcome overall survival (IMDC intermediate and poor risk groups). Treatments are ordered by P-Score (descending). Only subnetworks with >1 designs. Upper triangle: direct estimate. Lower triangle: network estimate.

Table 5. NMA results for OS (IMDC intermediate and poor risk groups) *(Continued)*

Subnet 1

No. of studies: 10. No. of pairwise comparisons: 10. No. of treatments: 7. No. of designs: 6

Heterogeneity/Inconsistency: Q = 9.1, df = 4, P = 0.059; I^2 = 56.1%, Tau^2 = 0.0635

Treatment Effects + 95%-CIs (Hazard rratios, random-effects model):

LEN+PEM	.	.	.	.	0.55 [0.33, 0.91]	.
0.91 [0.46, 1.83]	NIV+CAB	.	.	.	0.60 [0.37, 0.96]	.
0.84 [0.40, 1.75]	0.92 [0.45, 1.86]	NIV+IPI	.	.	0.65 [0.38, 1.10]	.
0.75 [0.39, 1.45]	0.82 [0.44, 1.54]	0.89 [0.46, 1.75]	AVE+AXI	.	0.73 [0.48, 1.11]	.
0.68 [0.30, 1.55]	0.75 [0.34, 1.66]	0.81 [0.35, 1.87]	0.91 [0.42, 1.96]	CAB	0.80 [0.42, 1.52]	.
0.55 [0.33, 0.91]	0.60 [0.37, 0.96]	0.65 [0.38, 1.10]	0.73 [0.48, 1.11]	0.80 [0.42, 1.52]	SUN	0.93 [0.58, 1.48]
0.51 [0.25, 1.01]	0.55 [0.28, 1.07]	0.60 [0.30, 1.22]	0.67 [0.36, 1.26]	0.74 [0.33, 1.64]	0.93 [0.58, 1.48]	LEN+EVE

Subnet 2

No. of studies: 2. No. of pairwise comparisons: 2. No. of treatments: 2. No. of designs: 1

Heterogeneity/Inconsistency: Q=0.12, df=1, p=0.73; I^2=0%, Tau^2=0

Treatment Effects + 95%-CIs (Hazard Ratios, random effects model):

IFN	0.85 [0.68, 1.05]
0.85 [0.68, 1.05]	NAP+IFN

Table 6. Short-term results (1 month after initiation of treatment) for QoL (all risk groups combined)

Trial	time point of measure-ment (in months)	Intervention (N analysed)	Intervention post mean score (SD)	Comparator (N analysed)	Comparator post mean score (SD)
Scale: FKSI-DRS (score range 0-36; higher scores represent better QoL)					
NCT00098657/ NCT00083889	**1**	**SUN (N = 348)**	**27.73 (5.080)**	**IFN (N = 317)**	**26.68 (5.195)**
Scale: EQ-5D-VAS (score range 0-100; higher scores represent better QoL)					
NCT00098657/ NCT00083889	**1**	**SUN (N=347)**	**69.35 (18.992)**	**IFN (N=315)**	**67.66 (20.058)**
Scale: FACT-G (score range 0-108; higher scores represent better QoL)					
NCT00098657/ NCT00083889	**1**	**SUN (N=348)**	**42.71 (8.959)**	**IFN (N=317)**	**40.93 (9.292)**
Scale: FACIT-F (score range 0-52; higher scores represent better QoL)					
NCT00720941	**1**	**PAZ (N=388)**	**35.2 (11.67)**	**SUN (N=393)**	**33.8 (12.56)**

Comparisons including SUN are bold. The scales are listed in no particular order.

Table 7. Mid-term results (6 months after initiation of treatment) for QoL (all risk groups combined)

Trial	time point of measure-ment (in months)	Intervention (N analysed)	Intervention post mean score (SD)	Comparator (N analysed)	Comparator post mean score (SD)
Scale: FKSI-DRS (score range 0-36; higher scores represent better QoL)					
NCT01108445	**6.9**	**EVE (N=13)**	**29.8 (3.76)**	**SUN (N = 19)**	**27.6 (4.37)**
NCT00098657/ NCT00083889	**6.4**	**SUN (N=242)**	**29.43 (4.280)**	**IFN (N = 109)**	**28.37 (4.726)**
NCT00920816	6.4	AXI (N=131)	28.557 (4.308)	SOR (N = 60)	30.296 (3.890)
Scale: EQ-5D-VAS (score range 0-100; higher scores represent better QoL)					
NCT00098657/ NCT00083889	**6.4**	**SUN (N=240)**	**75.13 (16.771)**	**IFN (N=104)**	**72.57 (16.007)**
NCT00920816	6.4	AXI (N=131)	71.031 (19.081)	SOR (N = 60)	73.183 (16.674)
Scale: FACT-G (score range 0-108; higher scores represent better QoL)					
NCT00098657/ NCT00083889	**6.4**	**SUN (N=241)**	**82.23 (15.124)**	**IFN (N = 106)**	**80.60 (15.527)**

Table 7. Mid-term results (6 months after initiation of treatment) for QoL (all risk groups combined) *(Continued)*

Scale: FACIT-F (score range 0-52; higher scores represent better QoL)					
NCT00720941	**6.5**	**PAZ (N=230)**	**38.8 (9.55)**	**SUN (N = 226)**	**36.2 (10.26)**

Comparisons including SUN are bold. The scales are listed in no particular order.

Table 8. Mid-term results (12 months after initiation of treatment) for QoL (all risk groups combined)

Trial	time point of measure-ment (in months)	Intervention (N analysed)	Intervention post mean score (SD)	Comparator (N analysed)	Comparator post mean score (SD)
Scale: FKSI-DRS (score range 0-36;higher scores represent better QoL)					
NCT00098657/ NCT00083889	**12**	**SUN (N = 166)**	**29.72 (4.245)**	**IFN (N=46)**	**29.36 (4.418)**
NCT00920816	12	AXI (N = 95)	29.579 (4.186)	SOR (N=37)	31.027 (3.790)
Scale: EQ-5D-VAS (score range 0-100; higher scores represent better QoL)					
NCT00098657/ NCT00083889	**12**	**SUN (N=168)**	**76.24 (15.740)**	**IFN (N = 46)**	**76.57 (17.924)**
NCT00920816	12	AXI (N = 95)	73.147 (17.546)	SOR (N = 37)	75.108 (18.371)
Scale: FACT-G (score range 0-108; higher scores represent better QoL)					
NCT00098657/ NCT00083889	**12**	**SUN (N = 166)**	**82.71 (15.276)**	**IFN (N = 46)**	**83.14 (17.067)**
Scale: FACIT-F (score range 0-52; higher scores represent better QoL)					
NCT00720941	**12**	**PAZ (N = 138)**	**39.5 (9.36)**	**SUN (N = 144)**	**37.7 (8.32)**

Comparisons including SUN are bold. The scales are listed in no particular order.

Table 9. Long-term results (approximately 24 months after initiation of treatment) for QoL (all risk groups combined)

Trial	time point of measure-ment (in months)	Intervention (N analysed)	Intervention post mean score (SD)	Comparator (N analysed)	Comparator post mean score (SD)
Scale: EQ-5D-VAS (score range 0-100; higher scores represent better QoL)					
NCT02231749	**23.7**	**NIV+IPI (N = 342)**	**(mean (SD) not reported)** **mean change 10.07 (4.35 to 15.80)**	**SUN (N = 351)**	**(mean (SD) not reported)** **mean change 6.40 (-1.36 to 14.16)**

Table 9. Long-term results (approximately 24 months after initiation of treatment) for QoL (all risk groups combined) *(Continued)*

Scale: FACT-G (score range 0-108; higher scores represent better QoL)					
NCT02231749	**23.7**	**NIV+IPI (N = 352)**	**(mean (SD) not reported)** **mean change 4.77 (1.73 to 7.82)**	**SUN (N = 356)**	**(mean (SD) not reported)** **mean change -4.32 (-8.54 to -0.11)**
Scale: FACIT-F (score range 0-52; higher scores represent better QoL)					
NCT00720941	**23.5**	**PAZ (N=39)**	**39.9 (8.33)**	**SUN (N = 36)**	**37.7 (8.46)**

Comparisons including SUN are bold. The scales are listed in no particular order.

Table 10. Long-term results (at the end of treatment) for QoL

Trial	time point of measure-ment (in months)	Intervention (N analysed)	Intervention post mean score (SD)	Comparator (N analysed)	Comparator post mean score (SD)
Scale: FKSI-DRS (score range 0-36; higher scores represent better QoL)					
NCT01108445	**40**	**EVE (N = 33)**	**26.6 (6.85)**	**SUN (N = 47)**	**26.6 (6.13)**
NCT00098657/ NCT00083889	**28.5**	**SUN (N = 53)**	**29.44 (4.210)**	**IFN (N = 351)**	**29.22 (7.694)**
NCT00920816	24.6	AXI (N =72)	26.556 (5.487)	SOR (N=95)	26.786 (5.982)
Scale: EQ-5D-VAS (score range 0-100; higher scores represent better QoL)					
NCT00098657/ NCT00083889	**28.5**	**SUN (N=54)**	**76.85 (16.863)**	**IFN (N = 352)**	**75.44 (25.060)**
NCT00920816	24.6	AXI (N = 71)	67.254 (19.495)	SOR (N=94)	67.048 (22.570)
Scale: FACT-G (score range 0-108; higher scores represent better QoL)					
NCT00098657/ NCT00083889	**28.5**	**SUN (N = 52)**	**84.62 (16.257)**	**IFN (N = 351)**	**79.54 (26.109)**
Scale: FACIT-F (score range 0-52; higher scores represent better QoL)					
NCT00720941	**38**	**PAZ (N = 2)**	**38.5 (13.59)**	**SUN (N = 2)**	**29.5 (0.71)**

Table 11. NMA results for SAEs (all risk groups combined)

Results of network meta-analysis for outcome serious adverse events (all risk groups combined). Treatments are ordered by P-Score (descending). Only subnetworks with >1 designs. Upper triangle: direct estimate. Lower triangle: network estimate.

No. of studies: 22. No. of pairwise comparisons: 30. No. of treatments: 21. No. of designs: 21.

Q_{total}=15.40, df=6, p=0.017; Q_{within}= 3.44, df=1, P = 0.064; $Q_{between}$=11.96, df=5, P = 0.035; I^2 = 61.0%, Tau^2 = 0.0256

Treatment Effects + 95%-CIs (Risk ratios, random-effects model):

TEM	0.80 [0.54, 1.17]	.	0.67 [0.46, 0.98]	.	.	.	.	.	0.97 [0.45, 2.09]	.	.	.	.	.	.	.	.	.	.	.	.
0.89 [0.63, 1.26]	IFN	.	0.84 [0.59, 1.21]	.	.	.	0.83 [0.53, 1.30]	.	0.57 [0.39, 0.83]	.	.	.	.	.	.	.	.	.	.	.	.
0.84 [0.36, 1.94]	0.94 [0.43, 2.07]	IFN +PLA	.	.	.	.	.	.	.	.	.	.	.	0.55 [0.36, 0.85]	.	.	.	.	.	.	.
0.71 [0.50, 1.03]	0.80 [0.56, 1.14]	0.85 [0.36, 2.00]	IFN+TEM	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.68 [0.42, 1.11]	0.77 [0.52, 1.13]	0.81 [0.37, 1.77]	0.96 [0.57, 1.60]	EVE	.	.	.	.	0.91 [0.70, 1.19]	.	.	.	.	.	.	.	.	.	.	.	.
0.68 [0.37, 1.23]	0.76 [0.45, 1.28]	0.81 [0.34, 1.90]	0.95 [0.51, 1.76]	0.99 [0.59, 1.65]	CAB	.	.	.	0.92 [0.60, 1.43]	.	.	.	.	.	.	.	.	.	.	.	.
0.69 [0.31, 1.52]	0.77 [0.37, 1.61]	0.82 [0.30, 2.22]	0.96 [0.43, 2.16]	1.01 [0.49, 2.08]	1.02 [0.45, 2.28]	NIN	.	.	0.91 [0.46, 1.79]	.	.	.	.	.	.	.	.	.	.	.	.
0.64 [0.41, 1.00]	0.72 [0.52, 0.99]	0.76 [0.35, 1.66]	0.89 [0.56, 1.43]	0.93 [0.64, 1.36]	0.94 [0.56, 1.58]	0.93 [0.45, 1.93]	SOR	0.99 [0.68, 1.43]	1.08 [0.74, 1.58]	0.97 [0.64, 1.45]	.	.	0.74 [0.44, 1.23]	.	.	.	.	.	.	.	.
0.63 [0.39, 1.02]	0.71 [0.49, 1.03]	0.75 [0.34, 1.65]	0.88 [0.54, 1.45]	0.92 [0.63, 1.35]	0.93 [0.55, 1.57]	0.92 [0.44, 1.91]	0.99 [0.74, 1.32]	PAZ	0.99 [0.71, 1.40]	.	.	.	.	.	.	.	.	.	.	.	.

Table 11. NMA results for SAEs (all risk groups combined) *(Continued)*

0.62 [0.41, 0.94]	0.70 [0.52, 0.94]	0.75 [0.36, 1.55]	0.87 [0.56, 1.36]	0.91 [0.70, 1.19]	0.92 [0.60, 1.43]	0.91 [0.46, 1.79]	0.98 [0.75, 1.28]	0.99 [0.75, 1.31]	SUN	.	0.78 [0.47, 1.30]	0.78 [0.54, 1.12]	.	0.69 [0.36, 1.31]	0.77 [0.58, 1.03]	.	0.72 [0.50, 1.04]	0.71 [0.51, 1.00]	0.70 [0.38, 1.27]	0.66 [0.46, 0.94]
0.62 [0.34, 1.13]	0.69 [0.41, 1.16]	0.74 [0.31, 1.78]	0.86 [0.46, 1.61]	0.90 [0.52, 1.57]	0.91 [0.47, 1.76]	0.90 [0.39, 2.07]	0.97 [0.64, 1.45]	0.98 [0.59, 1.62]	0.99 [0.61, 1.61]	SOR +IFN	.	.	.	.	.	.	.	.	.	.
0.58 [0.32, 1.07]	0.65 [0.38, 1.12]	0.70 [0.29, 1.64]	0.82 [0.43, 1.53]	0.85 [0.51, 1.44]	0.86 [0.46, 1.62]	0.85 [0.38, 1.92]	0.91 [0.54, 1.54]	0.93 [0.54, 1.57]	0.93 [0.60, 1.47]	0.95 [0.49, 1.84]	ATE	.	.	.	0.74 [0.47, 1.16]	.	.	.	.	.
0.48 [0.28, 0.84]	0.54 [0.34, 0.87]	0.58 [0.26, 1.31]	0.68 [0.38, 1.20]	0.71 [0.45, 1.11]	0.72 [0.41, 1.27]	0.71 [0.33, 1.52]	0.76 [0.48, 1.19]	0.77 [0.49, 1.22]	0.78 [0.54, 1.12]	0.79 [0.43, 1.44]	0.83 [0.47, 1.48]	PEM +AXI	.	.	.	.	.	.	.	.
0.47 [0.24, 0.93]	0.53 [0.29, 0.96]	0.56 [0.22, 1.43]	0.66 [0.33, 1.32]	0.69 [0.37, 1.30]	0.70 [0.34, 1.44]	0.69 [0.28, 1.67]	0.74 [0.44, 1.23]	0.75 [0.42, 1.34]	0.75 [0.42, 1.34]	0.76 [0.40, 1.47]	0.81 [0.39, 1.68]	0.97 [0.49, 1.92]	AXI	.	.	.	.	.	.	.
0.46 [0.22, 0.95]	0.52 [0.27, 1.00]	0.55 [0.36, 0.85]	0.64 [0.31, 1.35]	0.67 [0.35, 1.29]	0.68 [0.33, 1.42]	0.67 [0.27, 1.65]	0.72 [0.38, 1.38]	0.73 [0.38, 1.40]	0.74 [0.41, 1.33]	0.75 [0.35, 1.61]	0.79 [0.38, 1.66]	0.95 [0.47, 1.90]	0.98 [0.43, 2.23]	IFN +BEV	.	0.96 [0.65, 1.42]	.	.	0.91 [0.68, 1.22]	.
0.48 [0.29, 0.80]	0.54 [0.36, 0.81]	0.57 [0.26, 1.26]	0.67 [0.40, 1.14]	0.71 [0.48, 1.04]	0.71 [0.42, 1.20]	0.70 [0.34, 1.46]	0.75 [0.51, 1.12]	0.76 [0.51, 1.14]	0.77 [0.58, 1.03]	0.78 [0.44, 1.38]	0.83 [0.54, 1.27]	0.99 [0.63, 1.58]	1.02 [0.54, 1.94]	1.05 [0.54, 2.02]	ATE +BEV	.	.	.	.	.
0.44 [0.19, 1.00]	0.49 [0.23, 1.06]	0.53 [0.29, 0.94]	0.62 [0.27, 1.42]	0.64 [0.30, 1.37]	0.65 [0.28, 1.50]	0.64 [0.24, 1.71]	0.69 [0.32, 1.47]	0.70 [0.33, 1.50]	0.70 [0.35, 1.43]	0.71 [0.30, 1.69]	0.75 [0.33, 1.75]	0.91 [0.41, 2.02]	0.93 [0.37, 2.33]	0.96 [0.65, 1.42]	0.91 [0.42, 1.96]	EVE +BEV	.	.	.	.
0.45 [0.26, 0.78]	0.50 [0.32, 0.80]	0.54 [0.24, 1.22]	0.63 [0.35, 1.11]	0.66 [0.42, 1.03]	0.66 [0.38, 1.18]	0.65 [0.30, 1.41]	0.70 [0.45, 1.11]	0.71 [0.45, 1.13]	0.72 [0.50, 1.04]	0.73 [0.40, 1.34]	0.77 [0.43, 1.38]	0.93 [0.55, 1.55]	0.95 [0.48, 1.88]	0.98 [0.49, 1.96]	0.93 [0.59, 1.48]	1.02 [0.46, 2.27]	LEN +EVE	.	.	0.91 [0.64, 1.29]
0.45 [0.26, 0.76]	0.50 [0.32, 0.78]	0.53 [0.24, 1.19]	0.62 [0.36, 1.09]	0.65 [0.43, 1.00]	0.66 [0.38, 1.15]	0.65 [0.30, 1.38]	0.70 [0.45, 1.08]	0.71 [0.46, 1.10]	0.71 [0.51, 1.00]	0.72 [0.40, 1.31]	0.76 [0.44, 1.34]	0.92 [0.56, 1.51]	0.95 [0.49, 1.84]	0.97 [0.49, 1.92]	0.93 [0.59, 1.44]	1.01 [0.46, 2.23]	0.99 [0.60, 1.63]	NIV +IPI	.	.
0.42 [0.21, 0.85]	0.47 [0.25, 0.89]	0.50 [0.30, 0.84]	0.59 [0.28, 1.21]	0.61 [0.33, 1.15]	0.62 [0.30, 1.27]	0.61 [0.25, 1.48]	0.65 [0.35, 1.23]	0.66 [0.35, 1.26]	0.67 [0.38, 1.19]	0.68 [0.32, 1.44]	0.72 [0.35, 1.49]	0.86 [0.44, 1.70]	0.89 [0.39, 2.00]	0.91 [0.68, 1.22]	0.87 [0.46, 1.65]	0.95 [0.58, 1.55]	0.93 [0.47, 1.84]	0.94 [0.48, 1.83]	TEM +BEV	.

Table 11. NMA results for SAEs (all risk groups combined) *(Continued)*

0.41 [0.24, 0.71]	0.46 [0.29, 0.73]	0.49 [0.22, 1.11]	0.57 [0.32, 1.02]	0.60 [0.38, 0.94]	0.61 [0.34, 1.07]	0.60 [0.28, 1.29]	0.64 [0.41, 1.01]	0.65 [0.41, 1.03]	0.66 [0.46, 0.94]	0.67 [0.36, 1.22]	0.70 [0.39, 1.26]	0.85 [0.51, 1.41]	0.87 [0.44, 1.72]	0.89 [0.45, 1.79]	0.85 [0.54, 1.35]	0.93 [0.42, 2.07]	0.91 [0.64, 1.29]	0.92 [0.56, 1.51]	0.98 [0.50, 1.94]	LEN +PEM

Table 12. NMA results for PFS (all risk groups combined)

Results of network meta-analysis for outcome progression-free survival (combined risk groups). Treatments are ordered by P-Score (descending). Only subnetworks with >1 designs. Upper triangle: direct estimate. Lower triangle: network estimate.

Subnet 1

No. of studies: 27. No. of pairwise comparisons: 27. No. of treatments: 23. No. of designs: 23.

Q_{total}= 6.93, df=5, P = 0.23; Q_{within}= 2.02, df = 4, p=0.73; $Q_{between}$=4.91, df = 1, P = 0.027; I^2=27.9%, Tau^2 = 0.0155

Treatment Effects + 95%-CIs (Hazard ratios, random-effects model):

LEN +PEM	.	.	.	.	.	0.39 [0.29, 0.53]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.73 [0.46, 1.16]	CAB	.	.	.	.	0.54 [0.37, 0.76]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.60 [0.39, 0.91]	0.82 [0.52, 1.31]	LEN +EVE	.	.	.	0.65 [0.48, 0.87]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.57 [0.38, 0.86]	0.79 [0.50, 1.23]	0.96 [0.64, 1.42]	PEM +AXI	.	.	0.68 [0.52, 0.89]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.45 [0.31, 0.65]	0.61 [0.40, 0.93]	0.74 [0.51, 1.08]	0.78 [0.55, 1.10]	ATE +BEV	.	0.87 [0.70, 1.10]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.44 [0.29, 0.65]	0.60 [0.39, 0.94]	0.73 [0.49, 1.09]	0.76 [0.53, 1.11]	0.98 [0.69, 1.39]	NIV +IPI	0.89 [0.68, 1.16]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Table 12. NMA results for PFS (all risk groups combined) *(Continued)*

0.39 [0.29, 0.53]	0.54 [0.37, 0.76]	0.65 [0.48, 0.87]	0.68 [0.52, 0.89]	0.87 [0.70, 1.10]	0.89 [0.68, 1.16]	SUN	0.95 [0.73, 1.23]	.	.	0.89 [0.53, 1.50]	.	.	0.84 [0.55, 1.28]	.	.	0.71 [0.55, 0.92]	0.79 [0.60, 1.04]	.	0.57 [0.22, 1.47]	0.54 [0.41, 0.71]	.	.	
0.37 [0.25, 0.55]	0.51 [0.33, 0.79]	0.62 [0.42, 0.92]	0.65 [0.45, 0.94]	0.83 [0.59, 1.17]	0.85 [0.58, 1.23]	0.95 [0.73, 1.23]	PAZ	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0.40 [0.25, 0.63]	
0.36 [0.22, 0.60]	0.50 [0.29, 0.86]	0.60 [0.36, 1.00]	0.63 [0.38, 1.03]	0.81 [0.50, 1.30]	0.82 [0.50, 1.35]	0.93 [0.61, 1.40]	0.97 [0.60, 1.59]	TIV	.	.	.	.	.	.	.	.	0.76 [0.54, 1.06]	.	.	.	.	.	
0.35 [0.21, 0.61]	0.49 [0.27, 0.86]	0.59 [0.35, 1.01]	0.62 [0.37, 1.04]	0.79 [0.48, 1.31]	0.81 [0.48, 1.36]	0.91 [0.58, 1.42]	0.95 [0.57, 1.60]	0.98 [0.59, 1.63]	AXI	.	.	.	.	.	.	.	0.77 [0.53, 1.12]	.	.	.	.	.	
0.35 [0.19, 0.63]	0.48 [0.25, 0.90]	0.58 [0.32, 1.05]	0.61 [0.34, 1.09]	0.78 [0.44, 1.37]	0.79 [0.44, 1.42]	0.89 [0.53, 1.50]	0.94 [0.53, 1.67]	0.96 [0.50, 1.87]	0.98 [0.50, 1.95]	NIN	.	.	.	.	.	.	.	.	.	.	.	.	
0.33 [0.21, 0.53]	0.46 [0.28, 0.76]	0.56 [0.35, 0.88]	0.58 [0.37, 0.91]	0.75 [0.49, 1.14]	0.76 [0.49, 1.18]	0.85 [0.60, 1.22]	0.90 [0.58, 1.39]	0.92 [0.55, 1.54]	0.94 [0.55, 1.62]	0.96 [0.51, 1.79]	IFN +BEV	.	.	0.92 [0.65, 1.29]	0.91 [0.69, 1.20]	.	.	.	.	0.71 [0.55, 0.92]	0.63 [0.48, 0.83]	.	
0.32 [0.01, 17.44]	0.44 [0.01, 24.04]	0.54 [0.01, 29.06]	0.56 [0.01, 30.33]	0.72 [0.01, 38.92]	0.73 [0.01, 39.70]	0.83 [0.02, 44.22]	0.87 [0.02, 46.82]	0.89 [0.02, 48.10]	0.91 [0.02, 49.16]	0.92 [0.02, 51.21]	0.97 [0.02, 52.33]	SOR +IFN	.	.	.	.	0.85 [0.02, 45.10]	.	.	.	.	.	
0.33 [0.20, 0.55]	0.45 [0.26, 0.78]	0.55 [0.33, 0.91]	0.57 [0.35, 0.94]	0.73 [0.45, 1.19]	0.75 [0.45, 1.23]	0.84 [0.55, 1.28]	0.88 [0.54, 1.45]	0.91 [0.50, 1.64]	0.92 [0.50, 1.71]	0.94 [0.48, 1.84]	0.98 [0.57, 1.71]	1.02 [0.02, 55.79]	ATE	.	.	.	.	.	.	.	.	.	
0.31 [0.17, 0.55]	0.42 [0.23, 0.77]	0.51 [0.29, 0.91]	0.53 [0.30, 0.93]	0.69 [0.40, 1.18]	0.70 [0.40, 1.22]	0.78 [0.48, 1.29]	0.82 [0.47, 1.44]	0.85 [0.46, 1.57]	0.86 [0.45, 1.64]	0.88 [0.43, 1.80]	0.92 [0.65, 1.29]	0.95 [0.02, 52.25]	0.93 [0.49, 1.79]	EVE +BEV	.	.	.	.	.	.	.	.	
0.30 [0.18, 0.52]	0.42 [0.23, 0.74]	0.51 [0.29, 0.87]	0.53 [0.31, 0.89]	0.68 [0.41, 1.13]	0.69 [0.41, 1.17]	0.78 [0.49, 1.22]	0.82 [0.48, 1.37]	0.84 [0.47, 1.51]	0.85 [0.46, 1.57]	0.87 [0.44, 1.73]	0.91 [0.69, 1.20]	0.94 [0.02, 51.51]	0.92 [0.50, 1.72]	0.99 [0.64, 1.54]	TEM +BEV	.	.	.	.	.	.	.	
0.28 [0.19, 0.41]	0.38 [0.25, 0.59]	0.46 [0.31, 0.68]	0.48 [0.34, 0.70]	0.62 [0.44, 0.88]	0.63 [0.44, 0.92]	0.71 [0.55, 0.92]	0.75 [0.52, 1.08]	0.77 [0.47, 1.25]	0.78 [0.47, 1.32]	0.80 [0.45, 1.42]	0.83 [0.54, 1.29]	0.86 [0.02, 46.67]	0.85 [0.52, 1.39]	0.91 [0.52, 1.59]	0.92 [0.55, 1.54]	EVE	.	.	.	.	.	.	

Table 12. NMA results for PFS (all risk groups combined) *(Continued)*

0.27 [0.19, 0.40]	0.37 [0.24, 0.58]	0.45 [0.31, 0.67]	0.48 [0.33, 0.68]	0.61 [0.44, 0.85]	0.62 [0.43, 0.89]	0.70 [0.55, 0.89]	0.73 [0.52, 1.05]	0.76 [0.54, 1.06]	0.77 [0.53, 1.12]	0.78 [0.44, 1.39]	0.82 [0.56, 1.21]	0.85 [0.02, 45.10]	0.83 [0.51, 1.36]	0.89 [0.53, 1.50]	0.90 [0.56, 1.45]	0.98 [0.69, 1.40]	SOR	.	.	1.14 [0.75, 1.74]	.	.
0.26 [0.16, 0.41]	0.35 [0.21, 0.59]	0.43 [0.27, 0.69]	0.45 [0.28, 0.71]	0.58 [0.37, 0.89]	0.59 [0.37, 0.92]	0.66 [0.46, 0.95]	0.69 [0.44, 1.09]	0.71 [0.42, 1.20]	0.73 [0.42, 1.26]	0.74 [0.39, 1.39]	0.77 [0.53, 1.13]	0.80 [0.01, 43.36]	0.78 [0.45, 1.38]	0.84 [0.50, 1.41]	0.85 [0.53, 1.36]	0.92 [0.59, 1.45]	0.94 [0.63, 1.41]	NAP +IFN	.	0.92 [0.70, 1.22]	.	.
0.22 [0.08, 0.60]	0.30 [0.11, 0.84]	0.37 [0.14, 1.00]	0.39 [0.14, 1.04]	0.50 [0.19, 1.32]	0.51 [0.19, 1.36]	0.57 [0.22, 1.47]	0.60 [0.22, 1.60]	0.61 [0.22, 1.73]	0.63 [0.22, 1.79]	0.64 [0.22, 1.88]	0.66 [0.24, 1.83]	0.69 [0.01, 41.24]	0.68 [0.24, 1.91]	0.72 [0.25, 2.11]	0.73 [0.26, 2.09]	0.80 [0.30, 2.13]	0.81 [0.30, 2.16]	0.86 [0.31, 2.39]	TEM	.	.	.
0.24 [0.16, 0.35]	0.32 [0.21, 0.50]	0.39 [0.27, 0.58]	0.41 [0.29, 0.59]	0.53 [0.38, 0.74]	0.54 [0.38, 0.77]	0.61 [0.48, 0.77]	0.64 [0.45, 0.91]	0.66 [0.42, 1.02]	0.67 [0.42, 1.07]	0.68 [0.38, 1.20]	0.71 [0.55, 0.92]	0.74 [0.01, 39.50]	0.72 [0.44, 1.18]	0.77 [0.50, 1.19]	0.78 [0.53, 1.14]	0.85 [0.60, 1.21]	0.87 [0.65, 1.15]	0.92 [0.70, 1.22]	1.07 [0.40, 2.84]	IFN	.	.
0.21 [0.12, 0.36]	0.29 [0.16, 0.51]	0.35 [0.20, 0.60]	0.37 [0.22, 0.62]	0.47 [0.28, 0.78]	0.48 [0.28, 0.81]	0.54 [0.34, 0.85]	0.57 [0.34, 0.95]	0.58 [0.32, 1.04]	0.59 [0.32, 1.09]	0.60 [0.30, 1.20]	0.63 [0.48, 0.83]	0.65 [0.01, 35.70]	0.64 [0.35, 1.19]	0.69 [0.44, 1.07]	0.69 [0.47, 1.03]	0.76 [0.45, 1.27]	0.77 [0.48, 1.24]	0.82 [0.51, 1.31]	0.95 [0.33, 2.71]	0.89 [0.61, 1.30]	IFN +PLA	.
0.15 [0.08, 0.27]	0.20 [0.11, 0.38]	0.25 [0.14, 0.45]	0.26 [0.14, 0.46]	0.33 [0.19, 0.59]	0.34 [0.19, 0.61]	0.38 [0.23, 0.64]	0.40 [0.25, 0.63]	0.41 [0.21, 0.80]	0.42 [0.21, 0.83]	0.43 [0.20, 0.89]	0.45 [0.24, 0.84]	0.46 [0.01, 25.58]	0.45 [0.23, 0.89]	0.49 [0.24, 1.00]	0.49 [0.25, 0.98]	0.53 [0.30, 0.95]	0.54 [0.31, 0.97]	0.58 [0.31, 1.09]	0.67 [0.23, 1.98]	0.63 [0.35, 1.11]	0.71 [0.36, 1.41]	PLA

Table 13. NMA results for PFS (MSKCC favourable risk group)

Results of network meta-analysis for outcome progression-free survival (MSKCC favourable risk group). Treatments are ordered by P-Score (descending). Only subnetworks with >1 designs. Upper triangle: direct estimate. Lower triangle: network estimate.

Subnet 1

No. of studies: 7. No. of pairwise comparisons: 7. No. of treatments: 6. No. of designs: 5

Heterogeneity/Inconsistency: Q = 6.16, df = 2, P = 0.046; I^2 = 67.6%, Tau^2 = 0.3473

Treatment Effects + 95%-CIs (Hazard ratios, random-effects model):

LEN+PEM	.	0.36 [0.11, 1.23]	.	.	.
0.80 [0.14, 4.54]	LEN+EVE	0.45 [0.13, 1.53]	.	.	.

Table 13. NMA results for PFS (MSKCC favourable risk group) *(Continued)*

0.36 [0.11, 1.23]	0.45 [0.13, 1.53]	SUN	.	0.59 [0.23, 1.55]	0.50 [0.19, 1.33]
0.28 [0.04, 2.10]	0.35 [0.05, 2.61]	0.79 [0.16, 3.81]	AXI	.	0.64 [0.18, 2.23]
0.21 [0.04, 1.02]	0.27 [0.06, 1.26]	0.59 [0.23, 1.55]	0.75 [0.12, 4.78]	EVE	.
0.18 [0.04, 0.87]	0.23 [0.05, 1.08]	0.50 [0.19, 1.33]	0.64 [0.18, 2.23]	0.85 [0.22, 3.32]	SOR

Subnet 2

No. of studies: 2. No. of pairwise comparisons: 2. No. of treatments: 3. No. of designs: 2

Heterogeneity/Inconsistency: Q = 0.0, df=0, P = n.a.; I^2 = n.a., Tau^2 = n.a.

Treatment Effects + 95%-CIs (Hazard ratios, random-effects model):

IFN+BEV	0.83 [0.59, 1.18]	0.60 [0.42, 0.85]
0.83 [0.59, 1.18]	TEM+BEV	.
0.60 [0.42, 0.85]	0.72 [0.44, 1.18]	IFN+PLA

Table 14. NMA results for PFS (IMDC favourable risk group)

Results of network meta-analysis for outcome progression-free survival (IMDC favourable risk group). Treatments are ordered by P-Score (descending). Only subnetworks with >1 designs. Upper triangle: direct estimate. Lower triangle: network estimate.

Subnet 1

No. of studies: 5. No. of pairwise comparisons: 5. No. of treatments: 6. No. of designs: 5

Heterogeneity/Inconsistency: Q = 0, df=0, P = n.a.; I^2=n.a., Tau^2 = n.a.

Treatment Effects + 95%-CIs (Hazard ratios, random-effects model):

LEN+PEM	.	.	.	0.41 [0.28, 0.61]	.
0.75 [0.43, 1.29]	LEN+EVE	.	.	0.55 [0.38, 0.80]	.
0.71 [0.38, 1.31]	0.95 [0.52, 1.74]	NIV+CAB	.	0.58 [0.36, 0.93]	.
0.58 [0.34, 0.99]	0.77 [0.46, 1.31]	0.82 [0.45, 1.49]	AVE+AXI	0.71 [0.49, 1.02]	.
0.41 [0.28, 0.61]	0.55 [0.38, 0.80]	0.58 [0.36, 0.93]	0.71 [0.49, 1.02]	SUN	0.54 [0.38, 0.77]
0.22 [0.13, 0.38]	0.30 [0.18, 0.50]	0.32 [0.17, 0.57]	0.39 [0.23, 0.64]	0.54 [0.38, 0.77]	NIV+IPI

Table 15. NMA results for PFS (MSKCC intermediate and poor risk groups)

Results of network meta-analysis for outcome progression-free survival (MSKCC intermediate and poor risk groups). Treatments are ordered by P-Score (descending). Only subnetworks with >1 designs. Upper triangle: direct estimate. Lower triangle: network estimate.

Subnet 1

No. of studies: 10. No. of pairwise comparisons: 10. No. of treatments: 6. No. of designs: 5

Heterogeneity/Inconsistency: Q = 14.40, df = 5, P = 0.013; I^2 = 65.3%, Tau^2 = 0.1433

Treatment Effects + 95%-CIs (Hazard ratios, random-effects model):

LEN+PEM	.	.	0.33 [0.17, 0.62]	.	.
0.45 [0.19, 1.10]	LEN+EVE	.	0.72 [0.39, 1.33]	.	.
0.34 [0.09, 1.32]	0.76 [0.20, 2.88]	AXI	.	0.83 [0.35, 1.96]	.
0.33 [0.17, 0.62]	0.72 [0.39, 1.33]	0.95 [0.29, 3.08]	SUN	0.88 [0.39, 1.97]	0.84 [0.52, 1.34]
0.29 [0.10, 0.81]	0.63 [0.23, 1.75]	0.83 [0.35, 1.96]	0.88 [0.39, 1.97]	SOR	.
0.27 [0.12, 0.61]	0.60 [0.28, 1.31]	0.79 [0.22, 2.82]	0.84 [0.52, 1.34]	0.95 [0.37, 2.43]	EVE

Subnet 2

No. of studies: 2. No. of pairwise comparisons: 2. No. of treatments: 3. No. of designs: 2

Heterogeneity/Inconsistency: Q = 2.79, df = 2, P = 0.247; I^2=28.3%, Tau^2 = 0.0175

Treatment Effects + 95%-CIs (Hazard ratios, random-effects model):

IFN+BEV	0.99 [0.74, 1.32]	0.60 [0.45, 0.82]
0.99 [0.74, 1.32]	TEM+BEV	.
0.60 [0.45, 0.82]	0.61 [0.40, 0.93]	IFN+PLA

Table 16. NMA results for PFS (IMDC intermediate and poor risk groups)

Results of network meta-analysis for outcome progression-free survival (IMDC intermediate and poor risk groups). Treatments are ordered by P-Score (descending). Only subnetworks with >1 designs. Upper triangle: direct estimate. Lower triangle: network estimate.

Table 16. NMA results for PFS (IMDC intermediate and poor risk groups) *(Continued)*

Subnet 1

No. of studies: 11. No. of pairwise comparisons: 11. No. of treatments: 7. No. of designs: 6

Heterogeneity/Inconsistency: Q = 8.7, df = 5, P = 0.12; I^2 = 42.5%, Tau^2 = 0.0357

Treatment Effects + 95%-CIs (Hazard ratios, random-effects model):

LEN+PEM	.	.	.	.	.	0.36 [0.24, 0.54]
0.78 [0.40, 1.52]	CAB	.	.	.	.	0.46 [0.27, 0.79]
0.75 [0.43, 1.29]	0.96 [0.51, 1.81]	NIV+CAB	.	.	.	0.48 [0.34, 0.69]
0.60 [0.35, 1.02]	0.77 [0.41, 1.45]	0.81 [0.50, 1.32]	AVE+AXI	.	.	0.60 [0.43, 0.84]
0.52 [0.30, 0.92]	0.67 [0.35, 1.29]	0.70 [0.42, 1.18]	0.87 [0.52, 1.44]	LEN+EVE	.	0.69 [0.47, 1.01]
0.49 [0.27, 0.87]	0.63 [0.32, 1.22]	0.65 [0.38, 1.13]	0.81 [0.48, 1.37]	0.93 [0.53, 1.63]	NIV+IPI	0.74 [0.49, 1.11]
0.36 [0.24, 0.54]	0.46 [0.27, 0.79]	0.48 [0.34, 0.69]	0.60 [0.43, 0.84]	0.69 [0.47, 1.01]	0.74 [0.49, 1.11]	SUN

Subnet 2

No. of studies: 5. No. of pairwise comparisons: 5. No. of treatments: 5. No. of designs: 4

Heterogeneity/Inconsistency: Q 0.47, df = 1, P = 0.50; I^2 = 0%, Tau^2 = 0.0

Treatment Effects + 95%-CIs (Hazard ratios, random-effects model):

PAZ	0.73 [0.45, 1.19]	.	.	.
0.73 [0.45, 1.19]	TEM	.	.	0.74 [0.60, 0.91]
0.71 [0.40, 1.25]	0.97 [0.73, 1.31]	IFN+TEM	.	0.76 [0.62, 0.94]
0.56 [0.32, 0.98]	0.76 [0.57, 1.02]	0.78 [0.58, 1.05]	NAP+IFN	0.97 [0.79, 1.19]
0.54 [0.32, 0.92]	0.74 [0.60, 0.91]	0.76 [0.62, 0.94]	0.97 [0.79, 1.19]	IFN

Table 17. NMA results for AEs (all risk groups combined)

Results of network meta-analysis for outcome all-cause AEs (grades 3 and 4). Treatments are ordered by P-Score (descending). Only subnetworks with >1 designs. Upper triangle: direct estimate. Lower triangle: network estimate.

Subnet 1

No. of studies: 10. No. of pairwise comparisons: 14. No. of treatments: 11. No. of designs: 10.

Q_{total}= 0.31, df = 2, P =0.85; Q_{within}= 0.0, df = 0, P = n.a.; $Q_{between}$=0.31, df = 2, P = 0.85; I^2 = 0.0%, Tau^2 = 0.0

Treatment Effects + 95%-CIs (Risk ratios, random-effects model):

ATE	0.63 [0.47, 0.83]	.	.	.	0.58 [0.44, 0.76]	.	.	.	.	.
0.64 [0.49, 0.83]	ATE+BEV	.	.	.	0.89 [0.81, 0.97]	.	.	.	.	.
0.70 [0.44, 1.11]	1.09 [0.74, 1.61]	NIN	.	.	0.82 [0.56, 1.20]	.	.	.	.	.
0.59 [0.44, 0.78]	0.92 [0.80, 1.06]	0.84 [0.57, 1.26]	SOR	.	0.99 [0.85, 1.14]	0.92 [0.78, 1.09]	.	.	.	.
0.57 [0.43, 0.75]	0.89 [0.79, 1.01]	0.82 [0.55, 1.21]	0.97 [0.84, 1.12]	AVE+AXI	1.00 [0.92, 1.08]	.	.	.	.	.
0.57 [0.44, 0.74]	0.89 [0.81, 0.97]	0.82 [0.56, 1.20]	0.97 [0.86, 1.08]	1.00 [0.92, 1.08]	SUN	0.98 [0.92, 1.05]	0.96 [0.77, 1.21]	0.94 [0.85, 1.03]	0.87 [0.80, 0.95]	0.86 [0.80, 0.94]
0.56 [0.43, 0.73]	0.87 [0.78, 0.97]	0.80 [0.54, 1.18]	0.95 [0.84, 1.06]	0.98 [0.88, 1.09]	0.98 [0.92, 1.05]	PAZ	.	.	.	.
0.55 [0.39, 0.77]	0.85 [0.67, 1.09]	0.78 [0.50, 1.22]	0.93 [0.72, 1.20]	0.96 [0.75, 1.22]	0.96 [0.77, 1.21]	0.98 [0.77, 1.24]	CAB	.	.	.
0.53 [0.40, 0.70]	0.83 [0.73, 0.95]	0.77 [0.52, 1.13]	0.91 [0.78, 1.05]	0.93 [0.82, 1.06]	0.94 [0.85, 1.03]	0.96 [0.85, 1.07]	0.98 [0.76, 1.25]	NIV+CAB	.	.
0.50 [0.38, 0.65]	0.77 [0.69, 0.87]	0.71 [0.48, 1.05]	0.84 [0.73, 0.97]	0.87 [0.77, 0.98]	0.87 [0.80, 0.95]	0.89 [0.80, 0.99]	0.91 [0.71, 1.15]	0.93 [0.82, 1.05]	LEN+PEM	0.99 [0.93, 1.06]
0.49 [0.37, 0.64]	0.77 [0.68, 0.87]	0.70 [0.48, 1.04]	0.83 [0.73, 0.96]	0.86 [0.76, 0.97]	0.86 [0.80, 0.94]	0.88 [0.79, 0.98]	0.90 [0.71, 1.14]	0.92 [0.81, 1.04]	0.99 [0.93, 1.06]	LEN+EVE

Table 18. NMA results for Number of participants who discontinued study treatment due to an AE (all risk groups combined)

Results of network meta-analysis for the outcome Number of participants who discontinued treatment due to an AE. Treatments are ordered by P-Score (descending). Only subnetworks with >1 designs. Upper triangle: direct estimate. Lower triangle: network estimate.

Subnet 1

No. of studies: 28. No. of pairwise comparisons: 34. No. of treatments: 24. No. of designs: 25.

Q_{total}= 18.14, df = 8, P = 0.020; Q_{within}= 3.27, df = 3, P = 0.35; $Q_{between}$= 14.87, df = 5, P = 0.011; I^2=55.9%, Tau^2 = 0.1029

Treatment Effects + 95%-CIs (Risk ratios, random-effects model):

TIV	.	.	.	.	.	0.51 [0.24, 1.05]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.89 [0.28, 2.80]	AVE +AXI	.	.	.	0.57 [0.27, 1.20]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.67 [0.16, 2.79]	0.75 [0.19, 2.92]	ATE	.	.	0.76 [0.24, 2.36]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.53 [0.22, 1.32]	0.60 [0.24, 1.49]	0.80 [0.23, 2.80]	PAZ	.	1.24 [0.64, 2.43]	0.72 [0.35, 1.47]	.	.	.	0.69 [0.11, 4.31]	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.55 [0.16, 1.88]	0.61 [0.19, 1.94]	0.82 [0.19, 3.45]	1.02 [0.37, 2.85]	CAB	0.92 [0.38, 2.22]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.50 [0.21, 1.20]	0.57 [0.27, 1.20]	0.76 [0.24, 2.36]	0.95 [0.56, 1.60]	0.92 [0.38, 2.22]	SUN	1.42 [0.80, 2.54]	1.02 [0.55, 1.87]	1.00 [0.28, 3.62]	0.89 [0.45, 1.79]	0.40 [0.02, 9.32]	0.86 [0.42, 1.74]	0.78 [0.39, 1.56]	.	.	0.71 [0.34, 1.47]	0.68 [0.33, 1.41]	0.51 [0.26, 1.01]	0.29 [0.09, 0.98]	.	0.41 [0.21, 0.82]	.	.	0.23 [0.07, 0.72]
0.51 [0.24, 1.05]	0.57 [0.24, 1.38]	0.76 [0.22, 2.61]	0.95 [0.56, 1.63]	0.93 [0.34, 2.53]	1.01 [0.63, 1.62]	SOR	.	.	.	.	.	.	0.72 [0.20, 2.58]	0.71 [0.28, 1.78]	.	.	.	0.75 [0.28, 1.96]	.	.	.	.	.
0.51 [0.18, 1.48]	0.58 [0.22, 1.51]	0.77 [0.21, 2.80]	0.96 [0.43, 2.15]	0.94 [0.32, 2.74]	1.02 [0.55, 1.87]	1.01 [0.47, 2.18]	EVE	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Table 18. NMA results for Number of participants who discontinued study treatment due to an AE (all risk groups combined) *(Continued)*

0.50 [0.11, 2.38]	0.57 [0.13, 2.51]	0.76 [0.14, 4.21]	0.95 [0.24, 3.80]	0.92 [0.19, 4.39]	1.00 [0.28, 3.62]	0.99 [0.25, 3.90]	0.98 [0.24, 4.08]	NIN	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.45 [0.15, 1.37]	0.51 [0.18, 1.40]	0.67 [0.18, 2.56]	0.84 [0.35, 2.02]	0.82 [0.27, 2.53]	0.89 [0.45, 1.79]	0.89 [0.38, 2.05]	0.88 [0.35, 2.21]	0.89 [0.21, 3.85]	PEM +AXI	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.43 [0.13, 1.45]	0.49 [0.14, 1.65]	0.65 [0.15, 2.88]	0.81 [0.30, 2.19]	0.79 [0.21, 2.92]	0.86 [0.33, 2.25]	0.85 [0.32, 2.23]	0.84 [0.27, 2.64]	0.86 [0.17, 4.29]	0.96 [0.29, 3.16]	TEM	.	.	.	.	.	.	.	0.51 [0.22, 1.22]	.	.	.	0.36 [0.15, 0.83]	.	
0.43 [0.14, 1.33]	0.48 [0.17, 1.36]	0.65 [0.17, 2.47]	0.81 [0.34, 1.96]	0.79 [0.26, 2.45]	0.86 [0.42, 1.74]	0.85 [0.36, 1.99]	0.84 [0.33, 2.15]	0.86 [0.20, 3.73]	0.96 [0.36, 2.59]	1.00 [0.30, 3.31]	NIV +CAB	.	.	.	.	.	.	.	.	.	.	.	.	.
0.39 [0.13, 1.19]	0.44 [0.16, 1.22]	0.59 [0.16, 2.23]	0.74 [0.31, 1.76]	0.72 [0.24, 2.20]	0.78 [0.39, 1.56]	0.77 [0.34, 1.78]	0.77 [0.31, 1.93]	0.78 [0.18, 3.36]	0.88 [0.33, 2.33]	0.91 [0.28, 2.98]	0.91 [0.34, 2.45]	IFN	.	.	.	.	.	.	.	.	.	.	.	.
0.36 [0.08, 1.59]	0.41 [0.09, 1.94]	0.55 [0.09, 3.23]	0.68 [0.17, 2.74]	0.67 [0.13, 3.39]	0.72 [0.18, 2.83]	0.72 [0.20, 2.58]	0.71 [0.16, 3.17]	0.72 [0.11, 4.72]	0.81 [0.17, 3.75]	0.84 [0.17, 4.18]	0.84 [0.18, 3.93]	0.92 [0.20, 4.27]	AXI	.	.	.	.	.	.	.	.	.	.	.
0.36 [0.11, 1.17]	0.41 [0.11, 1.45]	0.54 [0.12, 2.52]	0.68 [0.24, 1.96]	0.66 [0.17, 2.57]	0.72 [0.26, 2.01]	0.71 [0.28, 1.78]	0.71 [0.21, 2.34]	0.72 [0.14, 3.74]	0.80 [0.23, 2.79]	0.84 [0.22, 3.16]	0.84 [0.24, 2.93]	0.92 [0.27, 3.18]	0.99 [0.21, 4.81]	SOR +IFN	.	.	.	.	.	.	.	.	.	.
0.36 [0.11, 1.11]	0.40 [0.14, 1.14]	0.53 [0.14, 2.06]	0.67 [0.27, 1.64]	0.65 [0.21, 2.05]	0.71 [0.34, 1.47]	0.70 [0.29, 1.67]	0.69 [0.27, 1.80]	0.71 [0.16, 3.10]	0.79 [0.29, 2.17]	0.82 [0.25, 2.76]	0.83 [0.30, 2.28]	0.91 [0.33, 2.47]	0.98 [0.21, 4.61]	0.98 [0.28, 3.48]	LEN +PEM	0.96 [0.47, 1.95]	.	.	.	.	.	.	.	.
0.34 [0.11, 1.06]	0.38 [0.14, 1.09]	0.51 [0.13, 1.98]	0.64 [0.26, 1.57]	0.63 [0.20, 1.96]	0.68 [0.33, 1.41]	0.67 [0.28, 1.60]	0.67 [0.26, 1.72]	0.68 [0.16, 2.98]	0.76 [0.28, 2.08]	0.79 [0.24, 2.65]	0.79 [0.29, 2.19]	0.87 [0.32, 2.37]	0.94 [0.20, 4.42]	0.95 [0.27, 3.34]	0.96 [0.47, 1.95]	LEN +EVE	.	.	.	.	.	.	.	.
0.26 [0.09, 0.78]	0.29 [0.11, 0.79]	0.39 [0.10, 1.45]	0.48 [0.20, 1.14]	0.47 [0.16, 1.43]	0.51 [0.26, 1.01]	0.51 [0.22, 1.16]	0.50 [0.20, 1.25]	0.51 [0.12, 2.19]	0.57 [0.22, 1.51]	0.60 [0.18, 1.94]	0.60 [0.22, 1.59]	0.65 [0.25, 1.72]	0.71 [0.15, 3.26]	0.71 [0.21, 2.45]	0.72 [0.27, 1.96]	0.75 [0.28, 2.03]	NIV +IPI	.	.	.	.	.	.	.
0.24 [0.09, 0.68]	0.27 [0.10, 0.78]	0.36 [0.09, 1.42]	0.46 [0.20, 1.03]	0.45 [0.14, 1.41]	0.48 [0.23, 1.01]	0.48 [0.23, 0.99]	0.47 [0.18, 1.24]	0.48 [0.11, 2.13]	0.54 [0.20, 1.49]	0.56 [0.26, 1.23]	0.56 [0.20, 1.57]	0.62 [0.23, 1.70]	0.67 [0.15, 2.92]	0.67 [0.21, 2.16]	0.68 [0.24, 1.93]	0.71 [0.25, 2.01]	0.95 [0.35, 2.58]	IFN	.	.	0.81 [0.41, 1.62]	0.70 [0.33, 1.50]	0.77 [0.34, 1.74]	

Table 18. NMA results for Number of participants who discontinued study treatment due to an AE (all risk groups combined) *(Continued)*

0.23 [0.06, 0.88]	0.25 [0.06, 1.00]	0.34 [0.07, 1.70]	0.42 [0.13, 1.40]	0.41 [0.10, 1.75]	0.45 [0.14, 1.41]	0.44 [0.14, 1.39]	0.44 [0.12, 1.61]	0.45 [0.08, 2.51]	0.50 [0.13, 1.91]	0.52 [0.16, 1.74]	0.52 [0.14, 2.01]	0.57 [0.15, 2.18]	0.62 [0.11, 3.46]	0.62 [0.14, 2.70]	0.63 [0.16, 2.46]	0.66 [0.17, 2.56]	0.88 [0.23, 3.32]	0.93 [0.37, 2.34]	EVE +BEV	.	0.87 [0.42, 1.80]	.	.
0.21 [0.07, 0.63]	0.23 [0.08, 0.64]	0.31 [0.08, 1.18]	0.39 [0.16, 0.93]	0.38 [0.12, 1.16]	0.41 [0.21, 0.82]	0.41 [0.18, 0.94]	0.40 [0.16, 1.02]	0.41 [0.10, 1.77]	0.46 [0.17, 1.23]	0.48 [0.15, 1.57]	0.48 [0.18, 1.29]	0.53 [0.20, 1.40]	0.57 [0.12, 2.63]	0.57 [0.16, 1.98]	0.58 [0.21, 1.59]	0.60 [0.22, 1.65]	0.80 [0.30, 2.12]	0.85 [0.31, 2.34]	0.92 [0.24, 3.50]	ATE +BEV	.	.	.
0.20 [0.06, 0.62]	0.22 [0.07, 0.70]	0.30 [0.07, 1.25]	0.37 [0.14, 0.96]	0.36 [0.10, 1.26]	0.39 [0.16, 0.95]	0.39 [0.16, 0.94]	0.38 [0.13, 1.12]	0.39 [0.08, 1.86]	0.44 [0.14, 1.35]	0.46 [0.17, 1.19]	0.46 [0.15, 1.42]	0.50 [0.16, 1.53]	0.54 [0.11, 2.57]	0.54 [0.15, 1.94]	0.55 [0.18, 1.74]	0.58 [0.18, 1.81]	0.76 [0.25, 2.33]	0.81 [0.46, 1.44]	0.87 [0.42, 1.80]	0.95 [0.31, 2.93]	IFN +BEV	.	0.84 [0.42, 1.68]
0.16 [0.05, 0.56]	0.18 [0.05, 0.65]	0.25 [0.05, 1.12]	0.31 [0.11, 0.88]	0.30 [0.08, 1.14]	0.33 [0.12, 0.89]	0.32 [0.12, 0.87]	0.32 [0.10, 1.04]	0.33 [0.06, 1.67]	0.36 [0.11, 1.24]	0.38 [0.17, 0.85]	0.38 [0.11, 1.30]	0.42 [0.12, 1.41]	0.45 [0.09, 2.29]	0.45 [0.12, 1.76]	0.46 [0.13, 1.59]	0.48 [0.14, 1.66]	0.64 [0.19, 2.14]	0.67 [0.32, 1.42]	0.73 [0.22, 2.38]	0.79 [0.23, 2.69]	0.83 [0.33, 2.13]	IFN +TEM	.
0.16 [0.05, 0.51]	0.19 [0.06, 0.58]	0.25 [0.06, 1.02]	0.31 [0.12, 0.78]	0.30 [0.09, 1.03]	0.33 [0.14, 0.77]	0.32 [0.14, 0.77]	0.32 [0.11, 0.92]	0.33 [0.07, 1.53]	0.37 [0.12, 1.10]	0.38 [0.15, 1.00]	0.38 [0.13, 1.16]	0.42 [0.14, 1.25]	0.45 [0.10, 2.12]	0.46 [0.13, 1.60]	0.46 [0.15, 1.42]	0.48 [0.16, 1.48]	0.64 [0.22, 1.90]	0.68 [0.37, 1.24]	0.73 [0.29, 1.85]	0.80 [0.27, 2.39]	0.84 [0.47, 1.49]	1.01 [0.39, 2.61]	TEM +BEV

Table 19. Results for TFST (all risk groups combined, narratively reported)

Trial	Definition of subsequent therapy	Participants	Intervention (N randomised)	n with subsequent anticancer treatment	Comparator (N randomised)	n with subsequent anticancer treatment
Comparisons including SUN (in no particular order)						
NCT00098657/ NCT00083889	"any poststudy cancer treatment for patients who discontinued from the trial"	all risk groups (according to MSKCC)	SUN (N=375)	n=182	IFN (N=375)	n=213
NCT00619268 comparison 1	"second-line therapy after study treatment failure"	all risk groups (according to MSKCC)	TEM+BEV (N=88)	n=61	SUN (N=42)	n=20
NCT00619268 comparison 2	"second-line therapy after study treatment failure"	all risk groups (according to MSKCC)	IFN+BEV (N=41)	n=27	SUN (N=42)	n=20
NCT01108445	"subsequent therapy after progression"	all risk groups (according to MSKCC)	EVE (N=57)	n=33	SUN (N=51)	n=36
NCT01835158	"subsequent anticancer therapy"	intermediate and poor risk groups (according to IMDC)	CAB (N=79)	n=51	SUN (N=72)	n=50
NCT02231749	"any subsequent therapy"	all risk groups (according to IMDC)	NIV+IPI (N=550)	n=330	SUN (N=546)	n=382
NCT01984242 comparison 1**	"subsequent therapy for patients who progressed"	all risk groups (according to MSKCC)	ATE (N=103)	n=62	SUN (N=101)	n=79
NCT02420821	"subsequent systemic treatment"	all risk groups (according to MSKCC)	ATE+BEV (N=454)	n=193	SUN (N=461)	n=238
NCT02684006	"subsequent anticancer therapy"	all risk groups (according to IMDC)	AVE+AXI (N=442)	n=138	SUN (N=444)	n= 227
NCT02853331	"subsequent anticancer therapy"	all risk groups (according to IMDC)	PEM+AXI (N=432)	n=204	SUN (N=429)	n=281
NCT00732914	"subseuqent therapy for patients who discontinued the study after first-line therapy" (cross over trial)	all risk groups (according to MSKCC)	SOR (N=74)***	n=13	SUN (N=100)***	n=24
NCT01024920	"post-study anticancer therapy"	all risk groups (according to MSKCC)	NIN (N=64)	n=25	SUN (N=32)	n=8
NCT03141177	"subsequent anticancer therapy"	all risk groups (according to IMDC)	NIV+CAB (N=323)	n=61	SUN (N=328)	n=108
NCT02811861 comparison 1	"any subsequent therapy"	all risk groups (according to MSKCC and IMDC)	LEN+PEM (N=355)	n=117	SUN (N=357)	n=206

Table 19. Results for TFST (all risk groups combined, narratively reported) *(Continued)*

NCT02811861 comparison 2	"any subsequent therapy"	all risk groups (according to MSKCC and IMDC)	LEN+EVE (N=357)	n=167	SUN (N=357)	n=206
Other comparisons (in no particular order)						
NCT00081614	"second-line therapy"	favourable and intermediate risk groups (according to MSKCC)	BEV+ERL (N=51)	n=7	BEV+PLA (N=53)	n=17
NCT00738530	"post-protocol therapy (not limited to second-line therapy) of patients progressive disease or those in whom trial therapy was discontinued"	all risk groups (according to MSKCC)	IFN+BEV (N=327)	n=180	IFN+PLA (N=322)	n=202
NCT00609401	"subsequent therapy (second-line) at relapse"	all risk groups (according to MSKCC)	SOR+ILN (N=66)	n=49	SOR (N=62)	n=48
NCT00619268 comparison 3*	"second-line therapy after study treatment failure"	all risk groups (according to MSKCC)	TEM+BEV (N=88)	n=61	IFN+BEV (N=41)	n=27
NCT01274273	"subsequent systemic anticancer therapy following study treatment discontinuation"	favourable and intermediate risk groups according to MSKCC; all risk groups according to IMDC	ILN+IFN +BEV (N=59)	n=50	ILN+IFN (N=59)	n=46
NCT00719264	"new cancer therapy after treatment discontinuation"	all risk groups (according to MSKCC)	EVE+BEV (N=182)	n=3	IFN+BEV (N=183)	n=2
NCT00920816	"follow-up systemic therapy after discontinuation of study treatment"	all risk groups (according to MSKCC)	AXI (N=192)	n=29	SOR (N=96)	n=19
NCT02811861 comparison 3*	"any subsequent therapy"	all risk groups (according to MSKCC and IMDC)	LEN+PEM (N=355)	n=117	LEN+EVE (N=357)	n=167

*In the three-arm trials, only two comparisons (arm A versus arm C (comparison 1) and arm B versus arm C (comparsion 2)) were reported, so we manually added a third comparison (arm A versus arm B (comparison 3)).
**Three-arm trial, only data for this comparison (arm A versus arm C) was reported.
***Cross over trial. N represents number of participants who received only first-line (first-period) treatment.

APPENDICES

Appendix 1. Updated search strategy for CENTRAL (for the update search in February 2022)

Cochrane Central Register of Controlled Trials in the Cochrane Library

ID Search

#1 MeSH descriptor: [Carcinoma, Renal Cell] explode all trees

#2 ((collecting duct* or hypernephroid* or nephroid*) NEAR/2 carcinoma*):ti,ab,kw

#3 ((grawitz NEAR/11 tumo?r*) or hypernephroma*):ti,ab,kw

#4 ((renal* or kidney* or nephron*) NEAR/6 (cancer* or neoplasms* or carcinoma* or tumour* or tumor* or adenocarcinoma*)):ti,ab,kw

#5 #1 or #2 or #3 or #4

#6 (advance* or metasta*):ti,ab,kw

#7 #5 and #6

#8 (mRCC or RCC):ti,ab,kw

#9 #7 or #8

#10 MeSH descriptor: [Antineoplastic Combined Chemotherapy Protocols] explode all trees

#11 MeSH descriptor: [Antibodies, Monoclonal, Humanized] explode all trees

#12 MeSH descriptor: [Antibodies, Monoclonal, Humanized] explode all trees

#13 MeSH descriptor: [Antineoplastic Agents] explode all trees

#14 (antibod* near/2 monoclonal*):ti,ab,kw

#15 ((anticancer* or "anti-cancer*" or anticarcinogen* or antineoplastic* or antitumo?r* or anti-tumo?r*) NEAR/2 (agent* or drug*)):ti,ab,kw

#16 antineoplastic*:ti,ab,kw

#17 (antibod* NEAR/2 (neoplasm* or tumo?r* or cancer* or antitumor?r*)):ti,ab,kw

#18 #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17

#19 (car NEAR/1 t cell therap*):ti,ab,kw

#20 (targeted NEAR therap*):ti,ab,kw

#21 MeSH descriptor: [Protein Kinase Inhibitors] explode all trees

#22 tyrosine kinase inhibitor*:ti,ab,kw

#23 (tivozanib* or AV-951 or AV951 or KRN-951 or KRN951):ti,ab,kw

#24 MeSH descriptor: [Sunitinib] explode all trees

#25 (sunitinib* or "SU-011248" or SU11248 or sutent):ti,ab,kw

#26 (pazopanib* or GW-786034 or GW786034 or votrient*):ti,ab,kw

#27 MeSH descriptor: [Sorafenib] explode all trees

#28 (sorafenib* or "bay-43-9006" or "bay439006" or "bay-439006" or "bay-5459085" or "bay5459085" or "bay-673472" or "bay673472" or nexavar):ti,ab,kw

#29 MeSH descriptor: [Lapatinib] explode all trees

#30 (lapatinib* or GW-282974* or GW-2016 or GW-572016 or GW2016 or GW572016 or GW282974* or tykerb or tyverb or AZD6094 or AZD-6094 or HMPL504 or MPL-504):ti,ab,kw

#31 (savolitinib* or volitinib* or AZD6094 or AZD-6094 or HMPL504 or HMPL-504):ti,ab,kw

#32 MeSH descriptor: [Cetuximab] this term only

#33 (cetuximab* or erbitux or c225 or anti-EGFR monoclonal antibod*):ti,ab,kw

#34 (urelumab* or BMS-663513 or BMS663513 or anti-CD137 monoclonal antibod*):ti,ab,kw

#35 (dovitinib* or TKI258 or TKI 258 or chir 258 or chir258):ti,ab,kw

#36 MeSH descriptor: [Axitinib] explode all trees

#37 (axitinib* or AG-013736 or AG013736 or AG-13736 or AG-13736 or inlyta*):ti,ab,kw

#38 (cabozantinib* or XL184 or XL-184 or cabometyx* or cometriq* or BMS 907351 or BMS907351):ti,ab,kw

#39 (SILA-9268A or SILA9268A or WY-090217 or WY090217):ti,ab,kw

#40 cell cycle inhibitor 779:ti,ab,kw

#41 MeSH descriptor: [Erlotinib Hydrochloride] this term only

#42 (Erlotinib* or tarceva* or OSI-774 or OSI774):ti,ab,kw

#43 #19 or #20 or #21 or #22 or #23 or #24 or #25 or #26 or #27 or #28 or #29 or #30 or #31 or #32 or #33 or #34 or #35 or #36 or #37 or #38 #39 or #40 or #41 or #42

#44 mTOR inhibitor*:ti,ab,kw

#45 (apitolisib* or GDC-0980 or GDC0980 or RG7422 or RG 7422):ti,ab,kw

#46 MeSH descriptor: [Sirolimus] explode all trees

#47 (sirolimus* or rapamycin* or AY-22989 or AY22989 or I 2190a or I2190a or cypher or opsiria or perceiva or rapamune):ti,ab,kw

#48 MeSH descriptor: [Everolimus] explode all trees

#49 (everolimus* or rad001 or rad-001 or sdz-rad or afinitor* or certican* or zortress):ti,ab,kw

#50 (temsirolimus* or CCI779 or CCI-779 or rapamune* or torisel*):ti,ab,kw

#51 (tivozanib* or AV-951 or AV951 or KRN-951 or KRN951 or KIL-8951 or KIL8951 or fotivda):ti,ab,kw

#52 #44 or #45 or #46 or #47 or #48 or #49 or #50 or #51

#53 MeSH descriptor: [Angiogenesis Inhibitors] explode all trees

#54 ((angiogenes* or angiogenic* or angiostatic* or angiogenet*) NEAR/2 (antagonist* or inhibitor* or agent*)):ti,ab,kw

#55 ((anti-angiogenetic* or antiangiogenetic* or anti angiogenes*) NEAR/2 (effect* or agent* or drug*)):ti,ab,kw

#56 (neovascularization* NEAR/2 inhibitor*):ti,ab,kw

#57 (lenvatinib* or E7080 or E-7080 or "ER-203492-00" or "ER-20349200" or "ER20349200" or "ER203492-00" or lenvima* or kisplyx):ti,ab,kw

#58 MeSH descriptor: [Angiogenesis Inhibitors] explode all trees

#59 (bevacizumab* or antiVEGF* or rhuMab-VEGF or ABP-215 or ABP215 or ainex* or altuzan* or avastin*):ti,ab,kw

#60 #53 or #54 or #55 or #56 or #57 or #58 or #59

#61 MeSH descriptor: [Immunotherapy] explode all trees

#62 immunotherap*:ti,ab,kw

#63 MeSH descriptor: [Interferons] explode all trees

#64 (interfer?on* or cl 884 or cl884):ti,ab,kw

#65 MeSH descriptor: [Interleukins] explode all trees

#66 interleukin*:ti,ab,kw

#67 (atezolizumab* or mpdl3280a or "mpdl-3280a" or anti-PDL1 or antiPDL1 or tec?ntriq*):ti,ab,kw

#68 #61 or #62 or #63 or #64 or #65 or #66 or #67

#69 checkpoint inhibitor*:ti,ab,kw

#70 MeSH descriptor: [Nivolumab] explode all trees

#71 (nivolumab* or opdivo* or MDX 1106 or MDX1106 or BMS936558 or BMS-936558 or CMAB-819 or CMAB819 or ONO-4538 or ONO4538):ti,ab,kw

#72 (pembrolizumab* or MK-3475 or MK3475 or keytruda or lambrolizumab* or sch900475 or sch 900475):ti,ab,kw

#73 (durvalumab* or MEDI4736 or MEDI-4736 or imfinzi):ti,ab,kw

#74 (tremelimumab* or ticilimumab* or CP-675 206 or CP-675206 or CP675206 or CP-675 or CP675):ti,ab,kw

#75 MeSH descriptor: [Ipilimumab] explode all trees

#76 (ipilimumab* or yervoy* or MDX-010 or MDX010 or mdx-ctla-4 or Anti-CTLA-4 MAb or MDX-101 or MDX101 or BMS-734016 or BMS34016):ti,ab,kw

#77 (LY2510924 or LY 2510924):ti,ab,kw

#78 #69 or #70 or #71 or #72 or #73 or #74 or #75 or #76 or #77

#79 #18 or #43 or #52 or #60 or #68 or #78

#80 #9 and #79

#81 #80 in Trials

Appendix 2. Search strategy for CENTRAL (for all searches up to April 2021)

Cochrane Central Register of Controlled Trials in the Cochrane Library

ID Search

#1 MeSH descriptor: [Carcinoma, Renal Cell] explode all trees

#2 ((collecting duct* or hypernephroid* or nephroid*) NEAR/2 carcinoma*):ti,ab,kw

#3 ((grawitz NEAR/11 tumo?r*) or hypernephroma*).ti,ab,kw

#4 ((renal* or kidney* or nephron*) NEAR/6 (cancer* or neoplasms* or carcinoma* or tumour* or tumor* or adenocarcinoma*)):ti,ab,kw

#5 #1 or #2 or #3 or #4

#6 (advance* or metasta*):ti,ab,kw

#7 #5 and #6

#8 (mRCC or RCC):ti,ab,kw

#9 #7 or #8

#10 MeSH descriptor: [Antineoplastic Combined Chemotherapy Protocols] explode all trees

#11 MeSH descriptor: [Antibodies, Monoclonal, Humanized] explode all trees

#12 MeSH descriptor: [Antibodies, Monoclonal, Humanized] explode all trees

#13 MeSH descriptor: [Antineoplastic Agents] explode all trees

#14 (antibod* near/2 monoclonal*):ti,ab,kw

#15 ((anticancer* or "anti-cancer*" or anticarcinogen* or antineoplastic* or antitumo?r*) NEAR/2 (agent* or drug*)):ti,ab,kw

#16 antineoplastic*:ti,ab,kw

#17 (antibod* NEAR/2 (neoplasm* or tumo?r* or cancer* or antitumor?r*)):ti,ab,kw

#18 #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17

#19 (car NEAR/1 t cell therap*):ti,ab,kw

#20 (targeted NEAR therap*):ti,ab,kw

#21 MeSH descriptor: [Protein Kinase Inhibitors] explode all trees

#22 tyrosine kinase inhibitor*:ti,ab,kw

#23 (tivozanib* or AV-951 or AV951 or KRN 951 or KRN951):ti,ab,kw

#24 MeSH descriptor: [Sunitinib] explode all trees

#25 (sunitinib* or "su 011248" or su11248 or sutent):ti,ab,kw

#26 (pazopanib* or GW 786034 or GW786034 or votrient*):ti,ab,kw

#27 MeSH descriptor: [Sorafenib] explode all trees

#28 (sorafenib* or "bay 43-9006" or "bay 439006" or "bay 5459085" or "bay5459085" or "bay 673472" or "bay673472" or nexavar):ti,ab,kw

#29 MeSH descriptor: [Lapatinib] explode all trees

#30 (lapatinib* or gw 282974x or gw 2016 or gw2016 or gw 572016 or gw572016 or gw282974x or tykerb or tyverb):ti,ab,kw

#31 (savolitinib* or volitinib*):ti,ab,kw

#32 MeSH descriptor: [Lapatinib] explode all trees

#33 (cetuximab* or erbitux or c225 or anti-EGFR monoclonal antibod*):ti,ab,kw

#34 (urelumab* or BMS-663513 or BMS663513 or anti-CD137 monoclonal antibod*):ti,ab,kw

#35 (dovitinib* or TKI258 or TKI 258 or chir 258 or chir258):ti,ab,kw

#36 MeSH descriptor: [Axitinib] explode all trees

#37 (axitinib* or AG-013736 or AG013736 or AG-13736 or AG-13736 or inlyta*):ti,ab,kw

#38 (cabozantinib* or XL184 or XL 184 or cabometyx* or cometriq* or BMS 907351 or BMS907351):ti,ab,kw

#39 (certican* or cci 779 or cci779 or zortress or AY 22989 or AY22989):ti,ab,kw

#40 (SILA 9268A or SILA9268A or WY-090217 or WY090217):ti,ab,kw

#41 cell cycle inhibitor 779:ti,ab,kw

#42 MeSH descriptor: [Axitinib] explode all trees

#43 (Erlotinib* or tarceva* or OSI 774 or OSI774):ti,ab,kw

#44 #19 or #20 or #21 or #22 or #23 or #24 or #25 or #26 or #27 or #28 or #29 or #30 or #31 or #32 or #33 or #34 or #35 or #36 or #37 or #38 or #39 or #40 or #41 or #42 or #43

#45 mTOR inhibitor*:ti,ab,kw

#46 (apitolisib* or GDC-0980 or GDC0980 or RG7422 or RG 7422):ti,ab,kw

#47 MeSH descriptor: [Sirolimus] explode all trees

#48 (rapamycin* or sirolimus* or ay 22989 or ay22989 or i 2190a or i2190a or cypher or opsiria or perceiva or rapamune):ti,ab,kw

#49 MeSH descriptor: [Everolimus] explode all trees

#50 (everolimus* or "rad 001" or sdz rad or afinitor* or certican*):ti,ab,kw

#51 (temsirolimus* or rapamycin* or "RAD 001" or rapamune* or afinitor* or torisel*):ti,ab,kw

#52 (tivozanib* or AV-951 or AV951 or KRN 951 or KRN951 or kil 8951 or kil8951 or fotivda):ti,ab,kw

#53 #45 or #46 or #47 or #48 or #49 or #50 or #51 or #52

#54 MeSH descriptor: [Angiogenesis Inhibitors] explode all trees

#55 ((angiogenes* or angiogenic* or angiostatic* or angiogenet*) NEAR/2 (antagonist* or inhibitor* or agent*)):ti,ab,kw

#56 ((anti-angiogenetic* or antiangiogenetic* or anti angiogenes*) NEAR/2 (effect* or agent* or drug*)):ti,ab,kw

#57 (neovascularization* NEAR/2 inhibitor*):ti,ab,kw

#58 (lenvatinib* or E7080 or E 7080 or "er 203492-00" or "er203492-00" or lenvima* or kisplyx):ti,ab,kw

#59 MeSH descriptor: [Angiogenesis Inhibitors] explode all trees

#60 (bevacizumab* or antiVEGF* or rhuMab-VEGF or abp 215 or abp215 or ainex* or altuzan* or avastin*):ti,ab,kw

#61 #54 or #55 or #56 or #57 or #58 or #59 or #60

#62 MeSH descriptor: [Immunotherapy] explode all trees

#63 immunotherap*:ti,ab,kw

#64 MeSH descriptor: [Interferons] explode all trees

#65 (interfer?on* or cl 884 or cl884):ti,ab,kw

#66 MeSH descriptor: [Interleukins] explode all trees

#67 interleukin*:ti,ab,kw

#68 (atezolizumab* or mpdl3280a or "mpdl 3280a" or anti-PDL1 or antiPDL1 or tec?ntriq*):ti,ab,kw

#69 #62 or #63 or #64 or #65 or #66 or #67 or #68

#70 Checkpoint inhibitor*:ti,ab,kw

#71 MeSH descriptor: [Nivolumab] explode all trees

#72 (nivolumab* or opdivo* or MDX 1106 or MDX1106 or bms936558 or bms 936558 or cmab 819 or cmab819 or ono 4538 or ono4538):ti,ab,kw

#73 (pembrolizumab* or MK-3475 or MK3475 or keytruda or lambrolizumab* or sch900475 or sch 900475):ti,ab,kw

#74 (durvalumab* or MEDI4736 or MEDI 4736 or imfinzi):ti,ab,kw

#75 (tremelimumab* or ticilimumab* or CP-675 206 or CP-675206 or CP675206):ti,ab,kw

#76 MeSH descriptor: [Ipilimumab] explode all trees

#77 (ipilimumab* or yervoy* or MDX-010 or MDX010 or mdx-ctla-4 or MDX-101 or MDX101 or bms 734016 or bms734016):ti,ab,kw

#78 (LY2510924 or LY 2510924):ti,ab,kw

#79 #70 or #71 or #72 or #73 or #74 or #75 or #76 or #77 or #78

#80 #18 or #44 or #53 or #61 or #69 or #79

#81 #9 and #80

Appendix 3. Updated search strategy MEDLINE (for the update search in February 2022)

Medline (OVID)

Searches

1 Carcinoma, Renal Cell/

2 ((collecting duct* or hypernephroid* or nephroid*) adj2 carcinoma*).tw,kf.

3 ((grawitz adj1 tumo?r*) or hypernephroma*).tw,kf.

4 ((renal* or kidney* or nephron*) adj6 (cancer* or neoplasms* or carcinoma* or tumour* or tumor* or adenocarcinoma*)).tw,kf.

5 or/1-4

6 (advance* or metasta*).tw,kf.

7 5 and 6

8 (mRCC or RCC).tw,kf.

9 7 or 8

10 antineoplastic combined chemotherapy protocols/

11 exp Antibodies, Monoclonal, Humanized/

12 Antineoplastic Agents/ad, ae

13 *Antineoplastic Agents/tu

14 antineoplastic.hw.

15 (antibod* adj2 monoclonal*).tw,kf.

16 ((anticancer* or "anti-cancer*" or anticarcinogen* or antineoplastic* or antitumo?r*) adj2 (agent* or drug*)).tw,kf.

17 antineoplastic*.tw,kf,nm.

18 (antibod* adj2 (neoplasm* or tumo?r* or cancer* or antitumor?r*)).tw,kf.

19 or/10-18

20 (car adj1 t cell therap*).tw,kf,nm.

21 (targeted adj therap*).tw,kf.

22 Protein Kinase Inhibitors/

23 tyrosine kinase inhibitor*.tw,kf,nm.

24 (tivozanib* or AV-951 or AV951 or KRN 951 or KRN951).tw,kf,nm.

25 sunitinib/

26 (sunitinib* or "su 011248" or su11248 or sutent).tw,kf,nm.

27 (pazopanib* or GW 786034 or GW786034 or votrient*).tw,kf,nm.

28 sorafenib/

29 (sorafenib* or bay 43-9006 or bay 439006 or bay 5459085 or bay5459085 or bay 673472 or bay673472 or nexavar).tw,kf,nm.

30 Lapatinib/

31 (lapatinib* or GW-282974* or GW-2016 or GW-572016 or GW2016 or GW572016 or GW282974* or tykerb or tyverb or AZD6094 or AZD-6094 or HMPL504 or HMPL-504).tw,kf,nm.

32 (savolitinib* or volitinib* or AZD6094 or AZD-6094 or HMPL504 or HMPL-504).tw,kf,nm.

33 Cetuximab/

34 (cetuximab* or erbitux or c225 auch c-225 or anti-EGFR monoclonal antibod*).tw,kf,nm.

35 (urelumab* or BMS-663513 or BMS663513 or anti-CD137 monoclonal antibod*).tw,kf,nm.

36 (dovitinib* or TKI258 or TKI 258 or chir 258 or chir258).tw,kf,nm.

37 Axitinib/

38 (axitinib* or AG-013736 or AG013736 or AG-13736 or AG-13736 or inlyta*).tw,kf,nm.

39 (cabozantinib* or XL184 or XL-184 or XL184cpd or XL-184cpd or cabometyx* or cometriq* or BMS-907351 or BMS907351).tw,kf,nm.

40 (certican* or cci 779 or cci779 or zortress or AY 22989 or AY22989).tw,kf,nm. Rausnehmen!!!

41 (SILA 9268A or SILA9268A or WY-090217 or WY090217).tw,kf,nm.

42 cell cycle inhibitor 779.tw,kf,nm.

43 Erlotinib Hydrochloride/

44 (Erlotinib* or tarceva* or OSI 774 or OSI774).tw,kf,nm.

45 or/20-44

46 mTOR inhibitor*.tw,kf.

47 (apitolisib* or GDC-0980 or GDC0980 or RG7422 or RG 7422).tw,kf,nm.

48 Sirolimus/

49 (rapamycin* or sirolimus* or ay 22989 or ay22989 or i 2190a or i2190a or cypher or opsiria or perceiva or rapamune).tw,kw,nm.

50 Everolimus/

51 (everolimus* or rad001 or rad-001 or sdz-rad or afinitor* or certican* or zortress).tw,kf,nm.

52 (temsirolimus* or rapamycin* or CCI779 or CCI-779 or rapamune* or torisel*).tw,kf,nm.

53 (tivozanib* or AV-951 or AV951 or KRN 951 or KRN951 or kil 8951 or kil8951 or fotivda).tw,kf,nm.

54 or/46-53

55 exp Angiogenesis Inhibitors/

56 ((angiogenes* or angiogenic* or angiostatic* or angiogenet*) adj2 (antagonist* or inhibitor* or agent*)).tw,kf.

57 ((anti-angiogenetic* or antiangiogenetic* or anti angiogenes*) adj2 (effect* or agent* or drug*)).tw,kf.

58 (neovascularization* adj2 inhibitor*).tw,kf.

59 (lenvatinib* or E7080 or E 7080 or er 203492-00 or er203492-00 or lenvima* or kisplyx).tw,kf,nm.

60 Bevacizumab/

61 (bevacizumab* or antiVEGF* or rhuMab-VEGF or abp 215 or abp215 or ainex* or altuzan* or avastin*).tw,kf,nm.

62 or/55-61

63 Immunotherapy/

64 immunotherap*.tw,kf.

65 exp Interferons/

66 (interfer?on* or cl 884 or cl884).tw,kf.

67 exp Interleukins/

68 interleukin*.tw,kf,nm.

69 (atezolizumab* or mpdl3280a or mpdl 3280a or RG-7446 or RG744 or anti-PDL1 or antiPDL1 or tec?ntriq*).tw,kf.

70 or/63-69

71 Checkpoint inhibitor*.tw,kf.

72 Nivolumab/

73 (nivolumab* or opdivo* or MDX 1106 or MDX1106 or bms936558 or bms 936558 or cmab 819 or cmab819 or ono 4538 or ono4538).tw,kf,nm.

74 (pembrolizumab* or MK-3475 or MK3475 or keytruda or lambrolizumab* or sch900475 or sch 900475).tw,kf,nm.

75 (durvalumab* or MEDI4736 or MEDI 4736 or imfinzi).tw,kf,nm.

76 (tremelimumab* or ticilimumab* or CP-675 206 or CP-675206 or CP675206 or CP-675 or CP675).tw,kf,nm.

77 Ipilimumab/

78 (ipilimumab* or yervoy* or MDX-010 or MDX010 or mdx-ctla-4 or Anti-CTLA-4 MAb or MDX-101 or MDX101 or bms 734016 or bms734016).tw,kf,nm.

79 (LY2510924 or LY 2510924).tw,kf,nm.

80 or/71-79

81 19 or 45 or 54 or 62 or 70 or 80

82 9 and 81

83 randomized controlled trial.pt.

84 controlled clinical trial.pt.

85 randomi?ed.ab.

86 placebo.ab.

87 drug therapy.fs.

88 randomly.ab.

89 trial.ab.

90 groups.ab.

91 or/83-90

92 exp animals/ not humans/

93 91 not 92

94 clinical trial, phase iii/

95 ("Phase 3" or "phase3" or "phase III" or P3 or "PIII").ti,ab,kw.

96 (94 or 95) not 92

97 93 or 96

98 82 and 97

Appendix 4. Search strategy for MEDLINE (for all searches up to April 2021)

MEDLINE (Ovid)

Searches

1 CARCINOMA, RENAL CELL/

2 ((collecting duct* or hypernephroid* or nephroid*) adj2 carcinoma*).tw,kf.

3 ((grawitz adj1 tumo?r*) or hypernephroma*).tw,kf.

4 ((renal* or kidney* or nephron*) adj6 (cancer* or neoplasms* or carcinoma* or tumour* or tumor* or adenocarcinoma*)).tw,kf.

5 or/1-4

6 (advance* or metasta*).tw,kf.

7 5 and 6

8 (mRCC or RCC).tw,kf.

9 7 or 8

10 ANTINEOPLASTIC COMBINED CHEMOTHERAPY PROTOCOLS/

11 exp ANTIBODIES, MONOCLONAL, HUMANIZED/

12 ANTINEOPLASTIC AGENTS/ad, ae

13 *ANTINEOPLASTIC AGENTS/tu

14 antineoplastic.hw.

15 (antibod* adj2 monoclonal*).tw,kf.

16 ((anticancer* or "anti-cancer*" or anticarcinogen* or antineoplastic* or antitumo?r*) adj2 (agent* or drug*)).tw,kf.

17 antineoplastic*.tw,kf.

18 (antibod* adj2 (neoplasm* or tumo?r* or cancer* or antitumor?r*)).tw,kf.

19 or/10-18

20 (car adj1 t cell therap*).tw,kf,nm.

21 (targeted adj therap*).tw,kf.

22 PROTEIN KINASE INHIBITORS/

23 tyrosine kinase inhibitor*.tw,kf.

24 (tivozanib* or AV-951 or AV951 or KRN 951 or KRN951).tw,kf,nm.

25 SUNITINIB/

26 (sunitinib* or "su 011248" or su11248 or sutent).tw,kf,nm.

27 (pazopanib* or GW 786034 or GW786034 or votrient*).tw,kf,nm.

28 SORAFENIB/

29 (sorafenib* or bay 43-9006 or bay 439006 or bay 5459085 or bay5459085 or bay 673472 or bay673472 or nexavar).tw,kf,nm.

30 LAPATINIB/

31 (lapatinib* or gw 282974x or gw 2016 or gw2016 or gw 572016 or gw572016 or gw282974x or tykerb or tyverb).tw,kw.

32 (savolitinib* or volitinib*).tw,kf,nm.

33 CETUXIMAB/

34 (cetuximab* or erbitux or c225 or anti-EGFR monoclonal antibod*).tw,kf,nm.

35 (urelumab* or BMS-663513 or BMS663513 or anti-CD137 monoclonal antibod*).tw,kf,nm.

36 (dovitinib* or TKI258 or TKI 258 or chir 258 or chir258).tw,kf,nm.

37 AXITINIB/

38 (axitinib* or AG-013736 or AG013736 or AG-13736 or AG-13736 or inlyta*).tw,kf,nm.

39 (cabozantinib* or XL184 or XL 184 or cabometyx* or cometriq* or BMS 907351 or BMS907351).tw,kf,nm.

40 (certican* or cci 779 or cci779 or zortress or AY 22989 or AY22989).tw,kf,nm.

41 (SILA 9268A or SILA9268A or WY-090217 or WY090217).tw,kf,nm.

42 cell cycle inhibitor 779.tw,kf,nm.

43 ERLOTINIB HYDROCHLORIDE/

44 (Erlotinib* or tarceva* or OSI 774 or OSI774).tw,kf,nm.

45 or/20-44

46 mTOR inhibitor*.tw,kf.

47 (apitolisib* or GDC-0980 or GDC0980 or RG7422 or RG 7422).tw,kf,nm.

48 SIROLIMUS/

49 (rapamycin* or sirolimus* or ay 22989 or ay22989 or i 2190a or i2190a or cypher or opsiria or perceiva or rapamune).tw,kw,nm.

50 EVEROLIMUS/

51 (everolimus* or "rad 001" or sdz rad or afinitor* or certican*).tw,kf,nm.

52 (temsirolimus* or rapamycin* or "RAD 001" or rapamune* or afinitor* or torisel*).tw,kf,nm.

53 (tivozanib* or AV-951 or AV951 or KRN 951 or KRN951 or kil 8951 or kil8951 or fotivda).tw,kf,nm.

54 or/46-53

55 exp ANGIOGENESIS INHIBITORS/

56 ((angiogenes* or angiogenic* or angiostatic* or angiogenet*) adj2 (antagonist* or inhibitor* or agent*)).tw,kf.

57 ((anti-angiogenetic* or antiangiogenetic* or anti angiogenes*) adj2 (effect* or agent* or drug*)).tw,kf.

58 (neovascularization* adj2 inhibitor*).tw,kf.

59 (lenvatinib* or E7080 or E 7080 or er 203492-00 or er203492-00 or lenvima* or kisplyx).tw,kf,nm.

60 BEVACIZUMAB/

61 (bevacizumab* or antiVEGF* or rhuMab-VEGF or abp 215 or abp215 or ainex* or altuzan* or avastin*).tw,kf,nm.

62 or/55-61

63 IMMUNOTHERAPY/

64 immunotherap*.tw,kf.

65 exp INTERFERONS/

66 (interfer?on* or cl 884 or cl884).tw,kf.

67 exp INTERLEUKINS/

68 interleukin*.tw,kf,nm.

69 (atezolizumab* or mpdl3280a or mpdl 3280a or anti-PDL1 or antiPDL1 or tec?ntriq*).tw,kf.

70 or/63-69

71 Checkpoint inhibitor*.tw,kf.

72 NIVOLUMAB/

73 (nivolumab* or opdivo* or MDX 1106 or MDX1106 or bms936558 or bms 936558 or cmab 819 or cmab819 or ono 4538 or ono4538).tw,kf,nm.

74 (pembrolizumab* or MK-3475 or MK3475 or keytruda or lambrolizumab* or sch900475 or sch 900475).tw,kf,nm.

75 (durvalumab* or MEDI4736 or MEDI 4736 or imfinzi).tw,kf,nm.

76 (tremelimumab* or ticilimumab* or CP-675 206 or CP-675206 or CP675206).tw,kf,nm.

77 IPILIMUMAB/

78 (ipilimumab* or yervoy* or MDX-010 or MDX010 or mdx-ctla-4 or MDX-101 or MDX101 or bms 734016 or bms734016).tw,kf,nm.

79 (LY2510924 or LY 2510924).tw,kf,nm.

80 or/71-79

81 19 or 45 or 54 or 62 or 70 or 80

82 9 and 81

83 randomized controlled trial.pt.

84 controlled clinical trial.pt.

85 randomi?ed.ab.

86 placebo.ab.

87 drug therapy.fs.

88 randomly.ab.

89 trial.ab.

90 groups.ab.

91 or/83-90

92 exp ANIMALS/ not HUMANS/

93 91 not 92

94 CLINICAL TRIAL, PHASE III/

95 ("Phase 3" or "phase3" or "phase III" or P3 or "PIII").ti,ab,kw.

96 (94 or 95) not 92

97 93 or 96

98 82 and 97

99 limit 98 to dt=20200218-20201025

Appendix 5. Updated search strategy for Embase (for the update search in February 2022)

Embase (Ovid)

Searches

1 renal cell carcinoma/

2 ((collecting duct* or hypernephroid* or nephroid*) adj2 carcinoma*).tw,kw.

3 ((grawitz adj1 tumo?r*) or hypernephroma*).tw,kw.

4 ((renal* or kidney* or nephron*) adj6 (cancer* or neoplasms* or carcinoma* or tumor* or tumour* or adenocarcinoma*)).tw,kw.

5 or/1-4

6 (advance* or metasta*).tw,kw.

7 5 and 6

8 (mRCC or RCC).tw,kw.

9 7 or 8

10 exp monoclonal antibody/

11 antineoplastic agent/ae, ad [Adverse Drug Reaction, Drug Administration]

12 *antineoplastic agent/dt [Drug Therapy]

13 antineoplastic.hw.

14 (antibod* adj2 monoclonal*).tw,kw.

15 ((anticancer* or "anti cancer*" or anitcarcinogen* or antitumo?r* or "anti tumo?r*") adj2 (agent* or drug*)).tw,kw.

16 antineoplastic*.tw,kw.

17 (antibod* adj2 (neoplasm* or tumo?r* or cancer* or antitumo?r*)).tw,kw.

18 or/10-17

19 (car adj1 t cell therap*).tw,kw.

20 "CAR T cell therapy"/

21 chimeric antigen receptor T-cell/dt

22 protein kinase inhibitor/

23 tyrosine kinase inhibitor*.tw,kw.

24 (tivozanib* or AV-951 or AV951 or KRN 951 or KRN951).tw,kw.

25 sunitinib/

26 (sunitinib* or "su 011248" or su11248 or sutent).tw,kw.

27 pazopanib/

28 (pazopanib* or GW 786034 or GW786034 or votrient*).tw,kw.

29 sorafenib/

30 (sorafenib* or bay 43-9006 or bay 439006 or bay439006 or bay 5459085 or bay5459085 or bay 673472 or bay673472 or nexavar).tw,kw.

31 lapatinib/ or lapatinib plus pazopanib/

32 (lapatinib* or gw 282974x or gw 2016 or gw2016 or gw 572016 or gw572016 or gw282974x or tykerb or tyverb).tw,kw.

33 savolitinib/

34 (volitinib* or savolitinib* or azd 6094 or azd6094 or hmpl 504 or hmpl504).tw,kw.

35 cetuximab/

36 (cetuximab* or c 225 or c225 or erbitux or imc 225 or imc c225 or imc-c225 or imc225 or imcc 225 or monoclonal antibody C 225).tw,kw.

37 urelumab/

38 (urelumab* or BMS-663513 or BMS663513 or anti-CD137 monoclonal antibod*).tw,kw.

39 dovitinib/

40 (dovitinib* or TKI258 or TKI 258 or chir 258 or chir258).tw,kw.

41 axitinib/

42 (axitinib* or AG-013736 or AG013736 or AG-13736 or AG-13736 or inlyta*).tw,kw.

43 cabozantinib/

44 (cabozantinib* or XL184 or XL 184 or cabometyx* or cometriq* or BMS 907351 or BMS907351).tw,kw.

45 (SILA 9268A or SILA9268A or WY-090217 or WY090217).tw,kw.

46 cell cycle inhibitor 779.tw,kw.

47 erlotinib/

48 (erlotinib* or tarceva* or OSI 774 or OSI774).tw,kw.

49 or/19-48

50 mTOR inhibitor*.tw,kw.

51 (apitolisib* or GDC-0980 or GDC0980 or RG7422 or RG 7422).tw,kw.

52 rapamycin/

53 (rapamycin* or sirolimus* or ay 22989 or ay22989 or i 2190a or i2190a or cypher or opsiria or perceiva or rapamune).tw,kw.

54 everolimus/

55 (everolimus* or "rad 001" or rad001 or sdz rad or afinitor* or certican* or zortress).tw,kw.

56 temsirolimus/

57 (temsirolimus* or rapamycin* or CCI779 or CCI-779 or rapamune* or torisel*).tw,kw.

58 tivozanib/

59 (tivozanib* or AV-951 or AV951 or KRN 951 or KRN951 or kil 8951 or kil8951 or fotivda).tw,kw.

60 or/50-59

61 angiogenesis inhibitor/

62 ((angiogenes* or angiogenic* or angiostatic* or angiogenet*) adj2 (antagonist* or inhibitor* or agent*)).tw,kw.

63 ((anti-angiogenetic* or antiangiogenetic* or anti angiogenes* or antiangiogenic*) adj2 (effect* or agent* or drug*)).tw,kw.

64 (neovascularization* adj2 inhibitor*).tw,kw.

65 ((neovascularization* or vascularization*) adj2 inhibitor*).tw,kw.

66 lenvatinib/

67 (lenvatinib* or E7080 or E 7080 or er 203492-00 or er203492-00 or lenvima* or kisplyx).tw,kw.

68 bevacizumab/

69 (bevacizumab* or antiVEGF* or rhuMab-VEGF or abp 215 or abp215 or ainex* or altuzan* or avastin*).tw,kw.

70 or/61-69

71 chimeric antigen receptor immunotherapy/

72 cancer immunotherapy/

73 immunotherap*.tw,kw.

74 interferon/

75 (interfer?on* or cl 884 or cl884).tw,kw.

76 interleukin derivative/

77 interleukin*.tw,kw.

78 atezolizumab/

79 (atezolizumab* or mpdl3280a or mpdl 3280a or anti-PDL1 or antiPDL1 or tec?ntriq*).tw,kw.

80 checkpoint inhibitor*.tw,kw.

81 nivolumab/

82 (nivolumab* or opdivo* or MDX 1106 or MDX1106 or bms936558 or bms 936558 or cmab 819 or cmab819 or ono 4538 or ono4538).tw,kw.

83 pembrolizumab/

84 (pembrolizumab* or MK-3475 or MK3475 or keytruda or lambrolizumab* or sch900475 or sch 900475).tw,kw.

85 durvalumab/

86 (durvalumab* or MEDI4736 or MEDI 4736 or imfinzi).tw,kw.

87 ticilimumab/

88 (tremelimumab* or ticilimumab* or CP-675 206 or CP-675206 or CP675206).tw,kw.

89 ipilimumab/

90 (ipilimumab* or yervoy* or MDX-010 or MDX010 or mdx-ctla-4 or MDX-101 or MDX101 or mdx010 or bms 734016 or bms734016).tw,kw.

91 (LY2510924 or LY 2510924).tw,kw.

92 or/71-91

93 Randomized controlled trial/

94 Controlled clinical study/

95 random*.ti,ab.

96 randomization/

97 intermethod comparison/

98 placebo.ti,ab.

99 (compare or compared or comparison).ti.

100 (open adj label).ti,ab.

101 ((double or single or doubly or singly) adj (blind or blinded or blindly)).ti,ab.

102 double blind procedure/

103 parallel group$1.ti,ab.

104 (crossover or cross over).ti,ab.

105 ((assign$ or match or matched or allocation) adj5 (alternate or group$1 or intervention$1 or patient$1 or subject$1 or participant $1)).ti,ab.

106 (controlled adj7 (study or design or trial)).ti,ab.

107 (volunteer or volunteers).ti,ab.

108 trial.ti.

109 or/93-108

110 (animal experiment/ or Animal experiment/) not (human experiment/ or human/)

111 109 not 110

112 18 or 49 or 60 or 70 or 92

113 9 and 112 and 111

Appendix 6. Search strategy for Embase (for all searches up to April 2021)

Embase (Ovid)

Searches

1 RENAL CELL CARCINOMA/

2 ((collecting duct* or hypernephroid* or nephroid*) adj2 carcinoma*).tw,kw.

3 ((grawitz adj1 tumo?r*) or hypernephroma*).tw,kw.

4 ((renal* or kidney* or nephron*) adj6 (cancer* or neoplasms* or carcinoma* or tumor* or tumour* or adenocarcinoma*)).tw,kw.

5 or/1-4

6 (advance* or metasta*).tw,kw.

7 5 and 6

8 (mRCC or RCC).tw,kw.

9 7 or 8

10 exp MONOCLONAL ANTIBODY/

11 ANTINEOPLASTIC AGENT/ae, ad

12 *antineoplastic agent/dt

13 antineoplastic.hw.

14 (antibod* adj2 monoclonal*).tw,kw.

15 ((anticancer* or "anti cancer*" or anitcarcinogen* or antitumo?r* or anti tumo?r*) adj2 (agent* or drug*)).tw,kw.

16 antineoplastic*.tw,kw.

17 (antibod* adj2 (neoplasm* or tumo?r* or cancer* or antitumo?r*)).tw,kw.

18 or/10-17

19 (car adj1 t cell therap*).tw,kw.

20 "CAR T cell therapy"/

21 CHIMERIC ANTIGEN RECEPTOR T-CELL/dt

22 PROTEIN KINASE INHIBITOR/

23 tyrosine kinase inhibitor*.tw,kw.

24 (tivozanib* or AV-951 or AV951 or KRN 951 or KRN951).tw,kw.

25 SUNITINIB/

26 (sunitinib* or "su 011248" or su11248 or sutent).tw,kw.

27 PAZOPANIB/

28 (pazopanib* or GW 786034 or GW786034 or votrient*).tw,kw.

29 SORAFENIB/

30 (sorafenib* or bay 43-9006 or bay 439006 or bay 5459085 or bay5459085 or bay 673472 or bay673472 or nexavar).tw,kw.

31 LAPATINIB/ or LAPATINIB PLUS PAZOPANIB/

32 (lapatinib* or gw 282974x or gw 2016 or gw2016 or gw 572016 or gw572016 or gw282974x or tykerb or tyverb).tw,kw.

33 SAVOLITINIB/

34 (volitinib* or savolitinib* or azd 6094 or azd6094 or hmpl 504 or hmpl504).tw,kw.

35 CETUXIMAB/

36 (cetuximab* or c 225 or c225 or erbitux or imc 225 or imc c225 or imc-c225 or imc225 or imcc 225 or monoclonal antibody C 225).tw,kw.

37 URELUMAB/

38 (urelumab* or BMS-663513 or BMS663513 or anti-CD137 monoclonal antibod*).tw,kw.

39 DOVITINIB/

40 (dovitinib* or TKI258 or TKI 258 or chir 258 or chir258).tw,kw.

41 AXITINIB/

42 (axitinib* or AG-013736 or AG013736 or AG-13736 or AG-13736 or inlyta*).tw,kw.

43 CABOZANTINIB/

44 (cabozantinib* or XL184 or XL 184 or cabometyx* or cometriq* or BMS 907351 or BMS907351).tw,kw.

45 (certican* or cci 779 or cci779 or zortress or AY 22989 or AY22989).tw,kw.

46 (SILA 9268A or SILA9268A or WY-090217 or WY090217).tw,kw.

47 cell cycle inhibitor 779.tw,kw.

48 ERLOTINIB/

49 (erlotinib* or tarceva* or OSI 774 or OSI774).tw,kw.

50 or/19-49

51 mTOR inhibitor*.tw,kw.

52 (apitolisib* or GDC-0980 or GDC0980 or RG7422 or RG 7422).tw,kw.

53 RAPAMYCIN/

54 (rapamycin* or sirolimus* or ay 22989 or ay22989 or i 2190a or i2190a or cypher or opsiria or perceiva or rapamune).tw,kw.

55 EVEROLIMUS/

56 (everolimus* or "rad 001" or rad001 or sdz rad or afinitor* or certican*).tw,kw.

57 TEMSIROLIMUS/

58 (temsirolimus* or cci 779 or cci779 or afinitor* or torisel*).tw,kw.

59 TIVOZANIB/

60 (tivozanib* or AV-951 or AV951 or KRN 951 or KRN951 or kil 8951 or kil8951 or fotivda).tw,kw.

61 or/51-60

62 ANGIOGENESIS INHIBITOR/

63 ((angiogenes* or angiogenic* or angiostatic* or angiogenet*) adj2 (antagonist* or inhibitor* or agent*)).tw,kw.

64 ((anti-angiogenetic* or antiangiogenetic* or anti angiogenes* or antiangiogenic*) adj2 (effect* or agent* or drug*)).tw,kw.

65 (neovascularization* adj2 inhibitor*).tw,kw.

66 ((neovascularization* or vascularization*) adj2 inhibitor*).tw,kw.

67 LENVATINIB/

68 (lenvatinib* or E7080 or E 7080 or er 203492-00 or er203492-00 or lenvima* or kisplyx).tw,kw.

69 BEVACIZUMAB/

70 (bevacizumab* or antiVEGF* or rhuMab-VEGF or abp 215 or abp215 or ainex* or altuzan* or avastin*).tw,kw.

71 or/62-70

72 CHIMERIC ANTIGEN RECEPTOR IMMUNOTHERAPY/

73 cancer immunotherapy/

74 immunotherap*.tw,kw.

75 INTERFERON/

76 (interfer?on* or cl 884 or cl884).tw,kw.

77 interleukin derivative/

78 interleukin*.tw,kw.

79 ATEZOLIZUMAB/

80 (atezolizumab* or mpdl3280a or mpdl 3280a or anti-PDL1 or antiPDL1 or tec?ntriq*).tw,kw.

81 checkpoint inhibitor*.tw,kw.

82 NIVOLUMAB/

83 (nivolumab* or opdivo* or MDX 1106 or MDX1106 or bms936558 or bms 936558 or cmab 819 or cmab819 or ono 4538 or ono4538).tw,kw.

84 PEMBROLIZUMAB/

85 (pembrolizumab* or MK-3475 or MK3475 or keytruda or lambrolizumab* or sch900475 or sch 900475).tw,kw.

86 DURVALUMAB/

87 (durvalumab* or MEDI4736 or MEDI 4736 or imfinzi).tw,kw.

88 TICILIMUMAB/

89 (tremelimumab* or ticilimumab* or CP-675 206 or CP-675206 or CP675206).tw,kw.

90 IPILIMUMAB/

91 (ipilimumab* or yervoy* or MDX-010 or MDX010 or mdx-ctla-4 or MDX-101 or MDX101 or mdx010 or bms 734016 or bms734016).tw,kw.

92 (LY2510924 or LY 2510924).tw,kw.

93 or/72-92

94 RANDOMIZED CONTROLLED TRIAL/

95 CONTROLLED CLINICAL STUDY/

96 random*.ti,ab.

97 RANDOMIZATION/

98 INTERMETHOD COMPARISON/

99 placebo.ti,ab.

100 (compare or compared or comparison).ti.

101 (open adj label).ti,ab.

102 ((double or single or doubly or singly) adj (blind or blinded or blindly)).ti,ab.

103 DOUBLE BLIND PROCEDURE/

104 parallel group$1.ti,ab.

105 (crossover or cross over).ti,ab.

106 ((assign$ or match or matched or allocation) adj5 (alternate or group$1 or intervention$1 or patient$1 or subject$1 or participant $1)).ti,ab.

107 (controlled adj7 (study or design or trial)).ti,ab.

108 (volunteer or volunteers).ti,ab.

109 trial.ti.

110 or/94-109

111 (ANIMAL EXPERIMENT/ or ANIMAL EXPERIMENT/) not (HUMAN EXPERIMENT/ or HUMAN/)

112 110 not 111

113 18 or 50 or 61 or 71 or 93

114 9 and 113

115 112 and 114

Appendix 7. Updated search strategy for clinical trial registries (for the update search in February 2022)

ClinicalTrials.gov (https://clinicaltrials.gov/)

expert search

(advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer) AND (axitinib OR cabozantinib OR dovitinib OR erlotinib OR lapatinib OR pazopanib OR savolitinib OR sorafenib OR sunitinib OR bevacizumab OR lenvatinib OR temsirolimus OR everolimus OR nivolumab OR ipilimumab OR pembrolizumab OR lambrolizumab OR atezolizumab OR durvalumab OR tremelimumab OR ticilimumab OR cetuximab OR urelumab OR interferon OR interleukin OR apitolisib OR LY2510924 OR tivozanib OR certican OR "cci 779" OR cci779 OR zortress OR "AY 22989" OR AY22989 OR SILA 9268A OR SILA9268A OR WY-090217 OR WY090217 OR rapamycin or sirolimus OR "ay 22989" OR ay22989 OR "i 2190a" OR i2190a OR cypher OR opsiria OR perceiva OR rapamune)

Interventional Studies

WHO ICTRP (https://trialsearch.who.int/)

Advanced search

1. Condition: "advanced renal cell cancer" OR "advanced renal cell carcinoma" OR "metastatic renal cell carcinoma" OR "metastatic renal cell cancer"
Intervention: axitinib OR cabozantinib OR dovitinib OR erlotinib OR lapatinib OR pazopanib OR savolitinib OR sorafenib OR sunitinib
Recruitment status: ALL

2. Condition: "advanced renal cell cancer" OR "advanced renal cell carcinoma" OR "metastatic renal cell carcinoma" OR "metastatic renal cell cancer"
Intervention: bevacizumab OR lenvatinib OR temsirolimus OR everolimus OR nivolumab OR ipilimumab OR pembrolizumab OR lambrolizumab OR atezolizumab
Recruitment status: ALL

3. Condition: "advanced renal cell cancer" OR "advanced renal cell carcinoma" OR "metastatic renal cell carcinoma" OR "metastatic renal cell cancer"

Intervention: durvalumab OR tremelimumab OR ticilimumab OR cetuximab OR urelumab OR interferon OR interleukin OR apitolisib OR LY2510924 OR tivozanib
Recruitment status: ALL

4. Condition: "advanced renal cell cancer" OR "advanced renal cell carcinoma" OR "metastatic renal cell carcinoma" OR "metastatic renal cell cancer"
Intervention: certican* OR "cci 779" OR cci779 OR zortress OR "AY 22989" OR AY22989 OR "SILA 9268A" OR SILA9268A OR "WY-090217" OR WY090217
Recruitment status: ALL

EU-clincal trials register (www.clinicaltrialsregister.eu)

1. ("advanced renal cell cancer" OR "advanced renal cell carcinoma" OR "metastatic renal cell carcinoma" OR "metastatic renal cell cancer") AND (axitinib OR cabozantinib OR dovitinib OR erlotinib OR lapatinib OR pazopanib OR savolitinib OR sorafenib OR sunitinib)

2. ("advanced renal cell cancer" OR "advanced renal cell carcinoma" OR "metastatic renal cell carcinoma" OR "metastatic renal cell cancer") AND (bevacizumab OR lenvatinib OR temsirolimus OR everolimus)

3. ("advanced renal cell cancer" OR "advanced renal cell carcinoma" OR "metastatic renal cell carcinoma" OR "metastatic renal cell cancer") AND (nivolumab OR ipilimumab OR pembrolizumab OR lambrolizumab OR atezolizumab)

4. ("advanced renal cell cancer" OR "advanced renal cell carcinoma" OR "metastatic renal cell carcinoma" OR "metastatic renal cell cancer") AND (durvalumab OR tremelimumab OR ticilimumab OR cetuximab OR urelumab)

5. ("advanced renal cell cancer" OR "advanced renal cell carcinoma" OR "metastatic renal cell carcinoma" OR "metastatic renal cell cancer") AND (interferon OR interleukin OR apitolisib OR LY2510924 OR tivozanib)

6. ("advanced renal cell cancer" OR "advanced renal cell carcinoma" OR "metastatic renal cell carcinoma" OR "metastatic renal cell cancer") AND (certican* OR "cci 779" OR cci779 OR zortress OR "AY 22989" OR AY22989)

7. ("advanced renal cell cancer" OR "advanced renal cell carcinoma" OR "metastatic renal cell carcinoma" OR "metastatic renal cell cancer") AND ("SILA 9268A" OR SILA9268A OR "WY-090217" OR WY090217)

Appendix 8. Search strategy for clinical trial registers (for all searches up to April 2021)

ClinicalTrials.gov (clinicaltrials.gov)

Basic Search
Advanced Renal Cell Carcinoma AND axitinib
Advanced Renal Cell Carcinoma AND cabozantinib
Advanced Renal Cell Carcinoma AND pazopanib
Advanced Renal Cell Carcinoma AND sorafenib
Advanced Renal Cell Carcinoma AND sunitinib

Advanced search
Conditions: Advanced Renal Cell Carcinoma
Interventions: axitinib OR cabozantinib OR pazopanib OR sorafenib OR sunitinib 133/ nicht verwendbar !!!!!
axitinib OR cabozantinib OR dovitinib OR erlotinib OR lapatinib OR pazopanib OR savolitinib OR sorafenib OR sunitinib
bevacizumab OR levantinib
temsirolimus OR everolimus
nivolumab OR ipilimumab OR pembrolizumab OR Lambrolizumab OR Atezolizumab
Durvalumab OR tremelimumab OR ticilimumab OR Cetuximab OR Urelumab
Interferon
Interleukin
Recruitment: All studies
Study type: Interventional studies
Age: adult, older adult

previous searches in ClincalTrials.gov

Interventional Studies | advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer | axitinib OR cabozantinib OR dovitinib OR erlotinib OR lapatinib OR pazopanib OR savolitinib OR sorafenib OR sunitinib | Adult, Older Adult

Interventional Studies | advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer | bevacizumab OR levantinib | Adult, Older Adult

Interventional Studies | advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer | temsirolimus OR everolimus | Adult, Older Adult

Interventional Studies | advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer | nivolumab OR ipilimumab OR pembrolizumab OR Lambrolizumab OR Atezolizumab | Adult, Older Adult

Interventional Studies | advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer | Durvalumab OR tremelimumab OR ticilimumab OR Cetuximab OR Urelumab | Adult, Older Adult

Interventional Studies | advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer | Interferon | Adult, Older Adult

Interventional Studies | advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer | Interleukin | Adult, Older Adult

WHO ICTRP apps.who.int/trialsearch/AdvSearch.aspx

Advanced search
1. Condition: advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer
Intervention: axitinib OR cabozantinib OR dovitinib OR erlotinib OR lapatinib OR pazopanib OR savolitinib OR sorafenib OR sunitinib
Recruitment status: ALL
2. Condition: advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer
Intervention: bevacizumab OR levantinib
Recruitment status: ALL
3. Condition: advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer
Intervention: temsirolimus OR everolimus
Recruitment status: ALL
4. Condition: advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer
Intervention: nivolumab OR ipilimumab OR pembrolizumab OR Lambrolizumab OR Atezolizumab
Recruitment status: ALL
5. Condition: advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer
Intervention: durvalumab OR tremelimumab OR ticilimumab OR Cetuximab OR urelumab
Recruitment status: ALL
6. Condition: advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer
Intervention: Interferon
Recruitment status: ALL
7. Condition: advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer
Intervention: Interleukin
Recruitment status: ALL

EU-clincal trials register (www.clinicaltrialsregister.eu)

(advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer) and (axitinib OR cabozantinib OR dovitinib OR erlotinib OR lapatinib OR pazopanib OR savolitinib OR sorafenib OR sunitinib)

(advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer) AND (bevacizumab OR levantinib)

(advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer) AND (temsirolimus OR everolimus)

(advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer) and (nivolumab OR ipilimumab OR pembrolizumab OR Lambrolizumab OR Atezolizumab)

(advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer) and (durvalumab OR tremelimumab OR ticilimumab OR Cetuximab OR urelumab)

(advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer) and interferon

(advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer) and interleukin

ISRCTN (www.isrctn.com)

(advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer) and (axitinib OR cabozantinib OR dovitinib OR erlotinib OR lapatinib OR pazopanib OR savolitinib OR sorafenib OR sunitinib)

(advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer) AND (bevacizumab OR levantinib)

(advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer) AND (temsirolimus OR everolimus)

(advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer) AND (nivolumab OR ipilimumab OR pembrolizumab OR Lambrolizumab OR Atezolizumab)

(advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer) and (durvalumab OR tremelimumab OR ticilimumab OR Cetuximab OR urelumab)

(advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer) and interferon

(advanced renal cell cancer OR advanced renal cell carcinoma OR metastatic renal cell carcinoma OR metastatic renal cell cancer) and interleukin

Appendix 9. Risk of bias assessment for the outcome overall survival

Trial	Risk of bias											
	Randomisation process		Deviations from intended interventions		Missing outcome data		Measurement of the outcome		Selection of the reported results		Overall	
	Authors' judgement	Support for judgement	Authors' judgement	Support for judgement	Authors' judgement	Support for judgement	Authors' judgement	Support for judgement	Authors' judgement	Support for judgement	Authors' judgement	Support for judgement
NCT03141177 **IMDC favourable risk group**	Low risk of bias	Interactive Response Technology was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. IMDC risk group was available for all randomised participants.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 8 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	High risk of bias	1.7% did not receive the intended interventions and therefore did not have outcome data. No information whether there was loss to follow-up. 3% of those who received treatment discontinued due to "other" reasons (not explained further).	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement (objective outcome).	High risk of bias	A study protocol with SAP available. All reported subgroup analyses (including analyses per IMDC risk group) were prespecified in the protocol and SAP. However, the time point that produced this numerical result was not pre-	High risk of bias	Overall judged high risk of bias due to lack of information about missing outcome data and inconsistency with preplanned analyses.

(Continued)

										specified.		
NCT03141177 **IMDC intermediate risk group**	Low risk of bias	Interactive Response Technology was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. IMDC risk group was available for all randomised participants.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 8 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	High risk of bias	1.7% did not receive the intended interventions and therefore did not have outcome data. No information whether there was loss to follow-up. 3% of those who received treatment discontinued due to "other" reasons (not explained further).	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement (objective outcome).	High risk of bias	A study protocol with SAP available. All reported subgroup analyses (including analyses per IMDC risk group) were prespecified in the protocol and SAP. However, the time point that produced this numerical result was not prespecified.	High risk of bias	Overall judged high risk of bias due to lack of information about missing outcome data and inconsistency with preplanned analyses.
NCT03141177 **IMDC poor risk group**	Low risk of bias	Interactive Response Technology was used for randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 partic-	High risk of bias	1.7% did not receive the intended interventions and therefore	Low risk of bias	No precise information provided about the out-	High risk of bias	A study protocol with SAP available. All reported sub-	High risk of bias	Overall judged high risk of bias due to lack of infor-

(Continued)

		There were no baseline imbalances that would suggest a problem with randomisation. IMDC risk group was available for all randomised participants.		ipants randomised to the experimental arm and 8 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).		did not have outcome data. No information whether there was loss to follow-up. 3% of those who received treatment discontinued due to "other" reasons (not explained further).		come assessors, but either way knowledge of intervention received could not have affected outcome measurement (objective outcome).		group analyses (including analyses per IMDC risk group) were prespecified in the protocol and SAP. However, the time point that produced this numerical result was not prespecified.		mation about missing outcome data and inconsistency with preplanned analyses.
NCT02811861 Comparison 1 (LEN +PEM vs.SUN) **Total trial population (combined risk groups)**	Low risk of bias	Interactive voice and web response system was used for randomisation. There were no baseline imbalances that would suggest a problem with ran-	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 17 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	High risk of bias	2.8% did not receive the intended interventions and therefore did not have outcome data. No information whether there was loss to follow-up. 2% discontin-	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received could not have	Low risk of bias	A study protocol with SAP available. All reported analyses were prespecified in the protocol and SAP.	High risk of bias	Overall judged high risk of bias due to lack of information about missing outcome data.

(Continued)

		domisation.				ued treatment due to "other" reasons (not explained further).		affected outcome measurement (objective outcome).				
NCT02811861 Comparison 1 (LEN +PEM vs.SUN) **MSKCC favourable risk group**	Low risk of bias	Interactive voice and web response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was available for all randomised participants.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 17 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	High risk of bias	2.8% did not receive the intended interventions and therefore did not have outcome data. No information whether there was loss to follow-up. 2% discontinued treatment due to "other" reasons (not explained further).	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement (objective outcome).	Low risk of bias	A study protocol with SAP available. All reported subgroup analyses (including analyses per MSKCC risk group) were prespecified in the protocol and SAP.	High risk of bias	Overall judged high risk of bias due to lack of information about missing outcome data.
NCT02811861 Comparison 1 (LEN +PEM vs.SUN)	Low risk of bias	Interactive voice and web response system was used for randomisa-	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised	High risk of bias	2.8% did not receive the intended interventions and therefore did not	Low risk of bias	No precise information provided about the outcome as-	Low risk of bias	A study protocol with SAP available. All reported subgroup	High risk of bias	Overall judged high risk of bias due to lack of information

(Continued)

MSKCC intermediate risk group		tion. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was available for all randomised participants.		to the experimental arm and 17 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).		have outcome data. No information whether there was loss to follow-up. 2% discontinued treatment due to "other" reasons (not explained further).		sessors, but either way knowledge of intervention received could not have affected outcome measurement (objective outcome).		analyses (including analyses per MSKCC risk group) were prespecified in the protocol and SAP.		about missing outcome data.
NCT02811861 Comparison 1 (LEN +PEM vs.SUN) **MSKCC poor risk group**	Low risk of bias	Interactive voice and web response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was available for all randomised	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 17 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	High risk of bias	2.8% did not receive the intended interventions and therefore did not have outcome data. No information whether there was loss to follow-up. 2% discontinued treatment due to "other" reasons (not explained further).	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement (objective outcome).	Low risk of bias	A study protocol with SAP available. All reported subgroup analyses (including analyses per MSKCC risk group) were prespecified in the protocol and SAP.	High risk of bias	Overall judged high risk of bias due to lack of information about missing outcome data.

(Continued)

		participants.										
NCT02811861 Comparison 1 (LEN +PEM vs.SUN) **IMDC favourable risk group**	Low risk of bias	Interactive voice and web response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. IMDC risk group was not available for 2 participants randomised to the experimental arm and 4 participants randomised to the control arm.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 17 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT). Participants without IMDC risk group allocation were excluded from subgroup analyses.	High risk of bias	2.8% did not receive the intended interventions and therefore did not have outcome data. No information whether there was loss to follow-up. 2% discontinued treatment due to "other" reasons (not explained further).	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement (objective outcome).	Low risk of bias	A study protocol with SAP available. All reported subgroup analyses (including analyses per IMDC risk group) were prespecified in the protocol and SAP.	High risk of bas	Overall judged high risk of bias due to lack of information about missing outcome data.
NCT02811861 Comparison 1 (LEN +PEM vs.SUN)	Low risk of bias	Interactive voice and web response system was used	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interven-	High risk of bias	2.8% did not receive the intended interventions and	Low risk of bias	No precise information provided about	Low risk of bias	A study protocol with SAP available. All report-	High risk of bias	Overall judged high risk of bias due to lack of

(Continued)

IMDC intermediate risk group		for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. IMDC risk group was not available for 2 participants randomised to the experimental arm and 4 participants randomised to the control arm.		tions. Only 3 participants randomised to the experimental arm and 17 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT). Participants without IMDC risk group allocation were excluded from subgroup analyses.		therefore did not have outcome data. No information whether there was loss to follow-up. 2% discontinued treatment due to "other" reasons (not further explained).		the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement (objective outcome).		ed subgroup analyses (including analyses per IMDC risk group) were prespecified in the protocol and SAP.		information about missing outcome data.
NCT02811861 Comparison 1 (LEN +PEM vs.SUN) **IMDC poor risk group**	Low risk of bias	Interactive voice and web response system was used for randomisation. There were no baseline imbalances that would suggest a prob-	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 17 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT). Partici-	High risk of bias	2.8% did not receive the intended interventions and therefore did not have outcome data. No information whether there was loss to follow-up.	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received	Low risk of bias	A study protocol with SAP available. All reported subgroup analyses (including analyses per IMDC) were pre-	High risk of bias	Overall judged high risk of bias due to lack of information about missing outcome data.

(Continued)

		lem with randomisation. IMDC risk group was not available for 2 participants randomised to the experimental arm and 4 participants randomised to the control arm.		pants without IMDC risk group allocation were excluded from subgroup analyses.		2% discontinued treatment due to "other" reasons (not explained further).		could not have affected outcome measurement (objective outcome).		specified in the protocol and SAP.		
NCT02811861 Comparison 2 (LEN +EVE vs. SUN) **Total trial population** **(combined risk groups)**	Low risk of bias	Interactive voice and web response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 2 participants randomised to the experimental arm and 17 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	High risk of bias	2.7% did not receive the intended interventions and therefore did not have outcome data. No information whether there was loss to follow-up. 2% discontinued treatment due to "other" reasons (not explained further).	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement (objec-	Low risk of bias	A study protocol with SAP available. All reported analyses were prespecified in the protocol and SAP.	High risk of bias	Overall judged high risk of bias due to lack of information about missing outcome data.

(Continued)

								tive outcome).				
NCT02811861 Comparison 2 (LEN +EVE vs. SUN) **MSKCC favourable risk group**	Low risk of bias	Interactive voice and web response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was available for all randomised participants.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 2 participants randomised to the experimental arm and 17 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	High risk of bias	2.7% did not receive the intended interventions and therefore did not have outcome data. No information whether there was loss to follow-up. 2% discontinued treatment due to "other" reasons (not explained further).	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement.	Low risk of bias	A study protocol with SAP available. All reported subgroup analyses (including analyses per MSKCC risk group) were prespecified in the protocol and SAP.	High risk of bias	Overall judged high risk of bias due to lack of information about missing outcome data.
NCT02811861 Comparison 2 (LEN +EVE vs. SUN) **MSKCC intermediate risk group**	Low risk of bias	Interactive voice and web response system was used for randomisation. There were no baseline imbalances that would	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 2 participants randomised to the experimental arm and 17 participants randomised to the control arm did not receive any treatment. The method of	High risk of bais	2.7% did not receive the intended interventions and therefore did not have outcome data. No information whether there was	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of interven-	Low risk of bias	A study protocol with SAP available. All reported subgroup analyses (including analyses per MSKCC	High risk of bias	Overall judged high risk of bias due to lack of information about missing outcome data.

(Continued)

		suggest a problem with randomisation. MSKCC risk group was available for all randomised participants.		analysis was appropriate (ITT).		loss to follow-up. 2% discontinued treatment due to "other" reasons (not further explained).		tion received could not have affected outcome measurement (objective outcome).		risk group) were pre-specified in the protocol and SAP.		
NCT02811861 Comparison 2 (LEN +EVE vs. SUN) **MSKCC poor risk group**	Low risk of bias	Interactive voice and web response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was available for all randomised participants.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 2 participants randomised to the experimental arm and 17 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	High risk of bias	2.7% did not receive the intended interventions and therefore did not have outcome data. No information whether there was loss to follow-up. 2% discontinued treatment due to "other" reasons (not explained further).	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement (objective outcome).	Low risk of bias	A study protocol with SAP available. All reported subgroup analyses (including analyses per MSKCC risk group) were pre-specified in the protocol and SAP.	High risk of bias	Overall judged high risk of bias due to lack of information about missing outcome data.
NCT02811861 Comparison	Low risk of bias	Interactive voice and web	Low risk of bias	The study was open-label: both participants and those de-	High risk of bias	2.7% did not receive the	Low risk of bias	No precise informa-	Low risk of bias	A study protocol with SAP	High risk of bias	Overall judged high risk

(Continued)

2 (LEN +EVE vs. SUN) **IMDC favourable risk group**		response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. IMDC risk group was not available for 6 participants randomised to the experimental arm and 4 participants randomised to the control arm.		livering the intervention were aware of assigned interventions. Only 2 participants randomised to the experimental arm and 17 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT). Participants without IMDC risk group allocation were excluded from subgroup analyses.		intended interventions and therefore did not have outcome data. No information whether there was loss to follow-up. 2% discontinued treatment due to "other" reasons (not explained further).		tion provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement.		available. All reported subgroup analyses (including analyses per IMDC risk group) were prespecified in the protocol and SAP.		of bias due to lack of information on about missing outcome data.
NCT02811861 Comparison 2 (LEN +EVE vs. SUN) **IMDC intermediate risk group**	Low risk of bias	Interactive voice and web response system was used for randomisation. There were no baseline imbalances that	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 2 participants randomised to the experimental arm and 17 participants randomised to the control arm did not receive any treat-	High risk of bias	2.7% did not receive the intended interventions and therefore did not have outcome data. No information whether	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of	Low risk of bias	A study protocol with SAP available. All reported subgroup analyses (including analyses per IMDC	High risk of bias	Overall judged high risk of bias due to lack of information about missing outcome data.

(Continued)

		would suggest a problem with randomisation. IMDC risk group was not available for 6 participants randomised to the experimental arm and 4 participants randomised to the control arm.		ment. The method of analysis was appropriate (ITT). Participants without IMDC risk group allocation were excluded from subgroup analyses.		there was loss to follow-up. 2% discontinued treatment due to "other" reasons (not explained further).		intervention received could not have affected outcome measurement.		risk group) were prespecified in the protocol and SAP.		
NCT02811861 Comparison 2 (LEN +EVE vs. SUN) **IMDC poor risk group**	Low risk of bias	Interactive voice and web response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. IMDC risk group was not available for	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 2 participants randomised to the experimental arm and 17 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT). Participants without IMDC risk group allocation were excluded from subgroup analyses.	High risk of bias	2.7% did not receive the intended interventions and therefore did not have outcome data. No information whether there was loss to follow-up. 2% discontinued treatment due to "other" reasons (not	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement	Low risk of bias	A study protocol with SAP available. All reported subgroup analyses (including analyses per IMDC risk group) were prespecified in the protocol and SAP.	High risk of bias	Overall judged high risk of bias due to lack of information about missing outcome data.

(Continued)

		6 participants randomised to the experimental arm and 4 participants randomised to the control arm.				explained further).		(objective outcome).				
NCT01108445 **Total trial population (combined risk groups)**	Low risk of bias	Participants were randomised in a 1:1 ratio. Randomisation was done under allocation concealment. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned intervention. Only 1 participant withdrew after consent, but before randomisation and before study drug was assigned. The method of analysis was appropriate (ITT).	Low risk of bias	All 108 participants were evaluable.	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement (objective outcome).	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to missing study protocol and SAP.
NCT01392183 **Total trial population**	Some concerns	Participants were randomised in a 1:1 ra-	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of	Low risk of bias	Detailed flow diagram provided, no indication	Low risk of bias	No precise information provided	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to

(Continued)

(only intermediate and poor risk groups included in the trial)		tio, but no information provided about the allocation concealment. There were no baseline imbalances that would suggest a problem with randomisation.		assigned interventions. No statement about the method of analysis.		of loss to follow-up.		about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement.				lack of information about the allocation concealment and method of analysis; missing study protocol and SAP.
NCT00334282 **Total trial population** **(combined risk groups)**	Low risk of bias	Interactive voice response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was blinded: both participants and those delivering the intervention were not aware of assigned interventions. The method of analysis was appropriate (ITT).	Low risk of bias	1.3% of those who received treatment were lost to follow-up. No analysis to correct for bias, but numbers are low and probably did not have an effect on the outcome.	Low risk of bias	Outcome assessors were not aware of the assigned intervention.	Low risk of bias	CSR and study protocol with SAP available. All reported analyses were prespecified in the protocol and SAP.	Low risk of bias	Overall judged low risk of bias.
NCT00065468 Comparison 1	Some concerns	Participants were randomised, but no in-	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of	Low risk of bias	1.9% did not receive the intended interven-	Low risk of bias	No precise information provided	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to

(Continued)

(TEM vs. IFN) **Total trial population** **(only intermediate and poor risk groups included in the trial)**		formation provided about the allocation concealment. There were no baseline imbalances that would suggest a problem with randomisation.		assigned interventions. Only 1 participant randomised to the single-drug arm and 7 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).		tions and therefore did not have outcome data. 2% of those who received treatment were lost to follow-up. However, these numbers are low and probably did not have an effect on the outcome.		about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement.				lack of information about the allocation concealment; missing study protocol and SAP.
NCT00065468 Comparison 2 (IFN +TEM vs. IFN) **Total trial population** **(only intermediate and poor risk groups included in the trial)**	Some concerns	Participants were randomised, but no information provided about the allocation concealment. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 2 participants randomised to the combination arm and 7 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	2.2% did not receive the intended interventions and therefore did not have outcome data. 1.7% of those who received treatment were lost to follow-up. However, these numbers are low and probably did	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement.	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to lack of information about the allocation concealment; missing study protocol and SAP.

(Continued)

						not have an effect on the outcome.						
NCT00738530 **Total trial population (combined risk groups)**	Low risk of bias	Interactive voice recognition system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was double-blind: both participants and those delivering the intervention were not aware of assigned interventions. Only 6 participants randomised to the experimental arm and 2 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	1.2% did not receive the intended interventions and therefore did not have outcome data. 4.4% of those who received treatment withdrew consent or were lost to follow-up before final data cut off for OS analysis. However, these numbers are low and probably did not have an effect on the outcome.	Low risk of bias	Outcome assessors were not aware of the assigned intervention.	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to missing study protocol and SAP.
NCT00738530 **MSKCC favourable risk group**	Low risk of bias	Interactive voice recognition system was used for	Low risk of bias	The study was double-blind: both participants and those delivering the intervention were not aware of assigned	Low risk of bias	1.2% did not receive the intended interventions and	Low risk of bias	Outcome assessors were not aware of the as-	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to missing

(Continued)

		randomisation. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was not available for 28 participants randomised to the experimental arm and 24 participants randomised to the control arm.		interventions. Only 6 participants randomised to the experimental arm and 2 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT). Participants without MSKCC risk group allocation were excluded from subgroup analyses.		therefore did not have outcome data. 4.4% of those who received treatment withdrew consent or were lost to follow-up before final data cut off for OS analysis. Unclear to which risk group they were assigned to. However, these numbers are low and probably did not have an effect on the outcome.		signed intervention.				study protocol and SAP.
NCT00738530 **MSKCC intermediate risk group**	Low risk of bias	Interactive voice recognition system was used for randomisation. There were no baseline imbalances that	Low risk of bias	The study was double-blind: both participants and those delivering the intervention were not aware of assigned interventions. Only 6 participants randomised to the experimental arm and 2 participants randomised to the control arm did not re-	Low risk of bias	1.2% did not receive the intended interventions and therefore did not have outcome data. 4.4% of those who received	Low risk of bias	Outcome assessors were not aware of the assigned intervention.	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to missing study protocol and SAP.

(Continued)

		would suggest a problem with randomisation. MSKCC risk group was not available for 28 participants randomised to the experimental arm and 24 participants randomised to the control arm.		ceive any treatment. The method of analysis was appropriate (ITT). Participants without MSKCC risk group allocation were excluded from subgroup analyses.		treatment withdrew consent or were lost to follow-up before final data cut off for OS analysis. Unclear to which risk group they were assigned to. However, these numbers are low and probably did not have an effect on the outcome.						
NCT00738530 **MSKCC poor risk group**	Low risk of bias	Interactive voice recognition system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC	Low risk of bias	The study was double-blind: both participants and those delivering the intervention were not aware of assigned interventions. Only 6 participants randomised to the experimental arm and 2 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT). Participants without MSKCC risk group allocation	Low risk of bias	1.2% did not receive the intended interventions and therefore did not have outcome data. 4.4% of those who received treatment withdrew consent or were lost to follow-up before final	Low risk of bias	Outcome assessors were not aware of the assigned intervention.	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to missing study protocol and SAP.

(Continued)

		risk group was not available for 28 participants randomised to the experimental arm and 24 participants randomised to the control arm.		were excluded from subgroup analyses.		data cut off for OS analysis. Unclear to which risk group they were assigned to. However, these numbers are low and probably did not have an effect on the outcome.						
NCT00072046 **Total trial population** **(combined risk groups)**	Some concerns	Participants were randomised via a stratified random block design, but no information provided about the allocation concealment. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 13 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	2.2% did not receive the intended interventions and therefore did not have outcome data. Of those who received treatment, less than 1% were lost to follow-up. However, these numbers are low and probably did not have an effect	Low risk of bias	Outcome assessors were aware of the assigned intervention, but knowledge of intervention received could not have affected outcome measurement (objective outcome).	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to missing information about the allocation concealment; missing study protocol and SAP.

(Continued)

						on the outcome.						
NCT00072046 **MSKCC favourable risk group**	Some concerns	Participants were randomised via a stratified random block design, but no information provided about the allocation concealment. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was available for all randomised participants.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 13 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	2.2% did not receive the intended interventions and therefore did not have outcome data. Of those who received treatment, less than 1% were lost to follow-up. Unclear to which risk group they were assigned to. However, these numbers are low and probably did not have an effect on the outcome.	Low risk of bias	Outcome assessors were aware of the assigned intervention, but knowledge of intervention received could not have affected outcome measurement.	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to missing information about the allocation concealment; missing study protocol and SAP.
NCT00072046 **MSKCC intermediate risk group**	Some concerns	Participants were randomised via a stratified random block	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 partic-	Low risk of bias	2.2% did not receive the intended interventions and there-	Low risk of bias	Outcome assessors were aware of the assigned	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to missing infor-

(Continued)

		design, but no information provided about the allocation concealment. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was available for all randomised participants.		ipants randomised to the experimental arm and 13 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).		fore did not have outcome data. Of those who received treatment, less than 1% were lost to follow-up. Unclear to which risk group they were assigned to. However, these numbers are low and probably did not have an effect on the outcome.		intervention, but knowledge of intervention received could not have affected outcome measurement.				mation about the allocation concealment; missing study protocol and SAP.
NCT00072046 **MSKCC poor risk group**	Some concerns	Participants were randomised via a stratified random block design, but no information provided about the allocation concealment. There were no baseline	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 13 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	2.2% did not receive the intended interventions and therefore did not have outcome data. Of those who received treatment, less than 1% were lost to follow-up.	Low risk of bias	Outcome assessors were aware of the assigned intervention, but knowledge of intervention received could not have affected outcome	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to missing information about the allocation concealment; missing study protocol and SAP.

(Continued)

		imbalances that would suggest a problem with randomisation. MSKCC risk group was available for all randomised participants.				Unclear to which risk group they were assigned to. However, these numbers are low and probably did not have an effect on the outcome.		measurement.				
NCT00609401 **Total trial population (combined risk groups)**	Low risk of bias	Randomisation was performed centrally. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. The method of analysis was appropriate (ITT).	Low risk of bias	6.3% were lost to follow-up. However, these numbers are low and probably did not have an effect on the outcome.	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement.	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to missing study protocol and SAP.
NCT00081614 **Total trial**	Low risk of bias	Interactive voice response service	Low risk of bias	The study was double-blind: both participants and those delivering the inter-	Low risk of bias	1 participant randomised to the ex-	Low risk of bias	Outcome assessors were not	Some concerns	No study protocol or SAP	Some concerns	Overall judged some concerns

(Continued)

population **(only favourable and intermediate risk groups included in the trial)**		was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.		vention were not aware of assigned interventions. The method of analysis was appropriate (ITT).		perimental arm was lost to follow-up, which probably did not have an effect on the outcome.		aware of the assigned intervention.		available.		due to missing study protocol and SAP.
Jonasch 2010 **Total trial population** **(combined risk groups)**	Low risk	Randomisation method appropriate and allocation concealed. No imbalances.	Low risk of bias	No information provided about whether the participants or those delivering the intervention were blinded or not. Only 1 participant randomised to the experimental arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk	1 participant did not receive the intended intervention and therefore did not have outcome data. Of those who received treatment, 8.8% came off study before the first 8-week response assessment. However, these numbers are low and probably did not have an effect	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement (objective outcome).	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to missing study protocol and SAP.

(Continued)

				on the outcome.								
NCT00098657; NCT00083889 **Total trial population** **(combined risk groups)**	Some concerns	Participants were randomised in a 1:1 ratio, but it is not mentioned who conducted the randomisation or whether it was conducted centrally so that nobody could foresee assignment. However, there were no baseline imbalances that would suggest a problem with randomisation.	Some concerns	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 15 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	High risk of bias	2% did not receive the intended interventions and therefore did not have outcome data. No information whether there was loss to follow-up.	Low risk of bias	Outcome assessors were aware of the assigned intervention, but knowledge of intervention received could not have affected outcome measurement (objective outcome).	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about the randomisation process and allocation concealment; deviations from intended interventions only in the control group; lack of information about missing data; missing study protocol and SAP.
NCT00098657; NCT00083889 **MSKCC intermedi-**	Some concerns	Participants were randomised in a 1:1	Some concerns	The study was open-label: both participants and those delivering the intervention were aware of	High risk of bias	2% did not receive the intended interven-	Low risk of bias	Outcome assessors were aware of	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to

(Continued)

ate risk group		ratio, but it is not mentioned who conducted the randomisation or whether it was conducted centrally so that nobody could foresee assignment. However, there were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was available for all randomised participants.		assigned interventions. Only 15 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).		tions and therefore did not have outcome data. No information whether there was loss to follow-up.		the assigned intervention, but knowledge of intervention received could not have affected outcome measurement (objective outcome).				lack of information about the randomisation process and allocation concealment; deviations from intended interventions only in the control group; lack of information about missing data; missing study protocol and SAP.
NCT00098655/ NCT00083889 **MSKCC poor risk group**	Some concerns	Participants were randomised in a 1:1 ratio, but it is not mentioned	Some concerns	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 15 participants randomised to the control arm did	High risk of bias	2% did not receive the intended interventions and therefore did not have out-	Low risk of bias	Outcome assessors were aware of the assigned intervention, but	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about

(Continued)

		who conducted the randomisation or whether it was conducted centrally so that nobody could foresee assignment. However, there were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was available for all randomised participants.		not receive any treatment. The method of analysis was appropriate (ITT).		come data. No information whether there was loss to follow-up.		knowledge of intervention received could not have affected outcome measurement (objective outcome).				the randomisation process and allocation concealment; deviations from intended interventions only in the control group; lack of information about missing data; missing study protocol and SAP.
NCT00920816 **Total trial population** **(combined risk groups)**	Low risk of bias	A centralised registration system was used for randomisation. There were no baseline imbalances that	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm did not receive any treatment. The method of analy-	Low risk of bias	1% did not receive the intended interventions and therefore did not have outcome data. 2.1% of those who received	Low risk of bias	Outcome assessors were not aware of the assigned intervention.	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to missing study protocol and SAP.

(Continued)

		would suggest a problem with randomisation.		sis was appropriate (ITT).		treatment were lost to follow-up. However, these numbers are low and probably did not have an effect on the outcome.						
NCT01024920 **Total trial population** **(combined risk groups)**	Low risk of bias	Interactive voice randomisation system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm did not receive any treatment. The method of analysis was appropriate.	Low risk of bias	3% did not receive the intended interventions and therefore did not have outcome data. 1 participant randomised to the control arm was lost to follow-up. However, these numbers are low and probably did not have an effect on the outcome.	Low risk of bias	Outcome assessors were aware of the assigned intervention, but knowledge of intervention received could not have affected outcome measurement.	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to missing study protocol and SAP.

(Continued)

NCT00631371 **Total trial population** **(combined risk groups)**	Low risk of bias	A computerised centrally located randomisation system was used. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 7 participants randomised to the experimental arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	Less than 1% did not receive the intended interventions and therefore did not have outcome data. 5.5% of those who received treatment were lost to follow-up. However, these numbers are low and probably did not have an effect on the outcome.	Low risk of bias	Outcome assessors were aware of the assigned intervention, but knowledge of intervention received could not have affected outcome measurement.	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to missing study protocol and SAP.
NCT01835158 **Total trial population** **(only intermediate and poor risk groups included in the trial)**	Low risk of bias	Randomisation was performed centrally. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 1 participant randomised to the experimental arm and 6 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	High risk of bias	4.5% did not receive the intended interventions and therefore did not have outcome data. There is a statement about loss to fol-	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received	Some concerns	Study protocol available with some statistical considerations briefly described, but no separate SAP available	High risk of bias	Overall judged high risk of bias due to lack of information about missing outcome data; missing SAP.

(Continued)

						low-up, but not how many.		could not have affected outcome measurement.		to fully check the pre-planned analyses.		
NCT02231749 **Total trial population (combined risk groups)**	Low risk of bias	Interactive voice response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 11 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	1.3% did not receive the intended interventions and therefore did not have outcome data. Less than 1% of those who received treatment were lost to follow-up. However, these numbers are low and probably did not have an effect on the outcome.	Low risk of bias	Outcome assessors were aware of the assigned intervention, but knowledge of intervention received could not have affected outcome measurement.	Low risk of bias	Study protocol and SAP available. Final revisions of both done before data cut-off (with extended follow-up). Analyses were pre-planned and reported.	Low risk of bias	Overall judged low risk of bias.
NCT02231749 **IMDC favourable risk group**	Low risk of bias	Interactive voice response system was used for randomisation. There	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised	Low risk of bias	1.3% did not receive the intended interventions and therefore did	Low risk of bias	Outcome assessors were aware of the assigned interven-	Low risk of bias	Study protocol and SAP available. Final revisions of both	Low risk of bias	Overall judged low risk of bias.

(Continued)

		were no baseline imbalances that would suggest a problem with randomisation. IMDC risk group was available for all randomised participants.		to the experimental arm and 11 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).		not have outcome data. Less than 1% of those who received treatment were lost to follow-up. Unclear to which risk group they were assigned to. However, these numbers are low and probably did not have an effect on the outcome.		tion, but knowledge of intervention received could not have affected outcome measurement.		done before data cut-off (with extended follow-up). Analyses were pre-planned and reported.		
NCT02231749 **IMDC intermediate&poor risk groups combined**	Low risk of bias	Interactive voice response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. IMDC risk group was avail-	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned Only 3 participants randomised to the experimental arm and 11 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	1.3% did not receive the intended interventions and therefore did not have outcome data. Less than 1% of those who received treatment were lost to follow-up. Unclear to which	Low risk of bias	Outcome assessors were aware of the assigned intervention, but knowledge of intervention received could not have affected outcome mea-	Low risk of bias	Study protocol and SAP available. Final revisions of both done before data cut-off (with extended follow-up). Analyses were pre-planned	Low risk of bias	Overall judged low risk of bias.

(Continued)

		able for all randomised participants.				risk group they were assigned to. However, these numbers are low and probably did not have an effect on the outcome.		surement.		and reported.		
NCT01984242 Comparison 1 (ATE vs. SUN) **Total trial population** **(combined risk groups)**	Low risk of bias	Interactive voice/web response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 1 participant randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	0.3% did not receive the intended interventions and therefore did not have outcome data. 1.5% of those who received treatment were lost to follow-up. However, these numbers are very low and probably did not have an effect on the outcome.	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement.	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to missing study protocol and SAP.
NCT01984242 Comparison	Low risk of bias	Interactive voice/web response	Low risk of bias	The study was open-label: both participants and those de-	Low risk of bias	0.3% did not receive the	Low risk of bias	No precise information	Some concerns	No study protocol or SAP	Some concerns	Overall judged some

(Continued)

2 (ATE +BEV vs. SUN) **Total trial population** **(combined risk groups)**		system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.		livering the intervention were aware of assigned interventions. Only 1 participant randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).		intended interventions and therefore did not have outcome data. 3.5% of those who received treatment were lost to follow-up. However, these numbers are very low and probably did not have an effect on the outcome.		tion provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement.		available.		concerns due to missing study protocol and SAP.
NCT02420821 **Total trial population** **(combined risk groups)**	Low risk of bias	Interactive voice and web response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk if bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 15 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	2% did not receive the intended interventions and therefore did not have outcome data. Less than 1% were lost to follow-up. However, these numbers are low and probably did	Low risk of bias	Outcome assessors were aware of the assigned intervention, but knowledge of intervention received could not have affected outcome measurement.	Low risk of bias	Study protocol and SAP available. All reported analyses were prespecified in the protocol and SAP.	Low risk of bias	Overall judged low risk of bias.

(Continued)

						not have an effect on the outcome.						
NCT02684006 **IMDC favourable risk group**	Low risk of bias	Interactive voice response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. IMDC risk group was not available for 5 participants randomised to the experimental arm and 1 participant randomised to the control arm.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 8 participants randomised to the experimental arm and 5 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate. Participants without IMDC risk group allocation were excluded from the analysis.	Some concerns	1.5% did not receive the intended interventions and therefore did not have outcome data. No information about study flow for the second interim analysis (which is the result considered in this review).	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement.	High risk of bias	Study protocol and SAP available but discrepancies were found between statements in the publications and the SAP about the prespecification of subgroup analyses. However, the time point (third interim analysis for OS) was prespecified.	High risk of bias	Overall judged high risk of bias due to lack of information about potential losses of follow-up; lack of information about the subgroup analyses in the study protocol and SAP.
NCT02684006 **IMDC intermedi-**	Low risk of bias	Interactive voice response system	Low risk of bias	The study was open-label: both participants and those delivering the interven-	Some concerns	1.5% did not receive the intend-	Low risk of bias	No precise information pro-	High risk of bias	Study protocol and SAP avail-	High risk of bias	Overall judged high risk of bias

(Continued)

ate risk group		was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. IMDC risk group was not available for 5 participants randomised to the experimental arm and 1 participant randomised to the control arm.		tion were aware of assigned interventions. Only 8 participants randomised to the experimental arm and 5 participants randomised to the control arm did not receive any treatment. unclear. The method of analysis was appropriate. Participants without IMDC risk group allocation were excluded from subgroup analyses.		ed interventions and therefore did not have outcome data. No information about study flow for the second interim analysis (which is the result considered in this review).		vided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement.		able but discrepancies were found between statements in the publications and the SAP about the prespecification of subgroup analyses. However, the time point (third interim analysis for OS) was prespecified.		due to lack of information about potential losses of follow-up; lack of information about the subgroup analyses in the study protocol and SAP.
NCT02684006 **IMDC poor risk group**	Low risk of bias	Interactive voice response system was used for randomisation. There were no baseline imbalances that would suggest	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 8 participants randomised to the experimental arm and 5 participants randomised to the control arm did not receive any treatment.	Some concerns	1.5% did not receive the intended interventions and therefore did not have outcome data. No information about study flow	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of interven-	High risk of bias	Study protocol and SAP available but discrepancies were found between statements in the publications	High risk of bias	Overall judged high risk of bias due to lack of information about potential losses of follow-up; lack of

(Continued)

		a problem with randomisation. IMDC risk group was not available for 5 participants randomised to the experimental arm and 1 participant randomised to the control arm.		The method of analysis was appropriate. Participants without IMDC risk group allocation were excluded from subgroup analyses.		for the second interim analysis (which is the result considered in this review).		tion received could not have affected outcome measurement.		and the SAP about the prespecification of subgroup analyses. However, the time point (third interim analysis for OS) was prespecified.		information about the subgroup analyses in the study protocol and SAP.
NCT02853331 **Total trial population** **(combined risk groups)**	Low risk of bias	Interactive voice response system or integrated web response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 4 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	Less than 1% did not receive the intended interventions and therefore did not have outcome data. No indication of loss to follow-up. However, these numbers are low and probably did not have an effect	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement.	Low risk of bias	Study protocol and SAP available. All reported analyses were prespecified in the protocol and SAP.	Low risk of bias	Overall judged low risk of bias.

(Continued)

						on the outcome.						
NCT00719264 **Total trial population (combined risk groups)**	Some concerns	Participants were randomised in a 1:1 ratio, but no information provided about who conducted the randomisation and whether the allocation was concealed. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 1 participant randomised to the experimental arm and 2 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate.	Low risk of bias	Less than 1% did not receive the intended interventions and therefore did not have outcome data. Less than 1% of those who received treatment were lost to follow-up. However, these numbers are low and probably did not have an effect on the outcome.	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement.	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to lack of information about the randomisation and allocation concealment; missing study protocol and SAP.
NCT00720941 **Total trial population (combined risk groups)**	Low risk of bias	Interactive voice response system was used for randomisation. There were no baseline imbalances that	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 5 participants randomised to the control arm did	Low risk of bias	Less than 1% did not receive the intended interventions and therefore did not have outcome data. 2.9% of	Low risk of bias	No precise information provided about the outcome assessors, but either way knowl-	Low risk of bias	CSR and study protocol with SAP available. All reported analyses were pre-specified in the	Low risk of bias	Overall judged low risk of bias.

(Continued)

		would suggest a problem with randomisation.		not receive any treatment. The method of analysis was appropriate (ITT).		those who received treatment were lost to follow-up. However, these numbers are small and probably did not have an effect on the outcome.		edge of intervention received could not have affected outcome measurement.		protocol and SAP.		
NCT00720941 **MSKCC favourable risk group**	Low risk of bias	Interactive voice response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was not available for 17 participants randomised to the experimental arm	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 5 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT). Participants without MSKCC risk group allocation were excluded from subgroup analyses.	Low risk of bias	Less than 1% did not receive the intended interventions and therefore did not have outcome data. 2.9% of those who received treatment were lost to follow-up. Unclear to which risk group they were assigned to. However, these numbers are small and probably did	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement.	High risk of bias	A post-hoc analysis according to risk group was conducted.	High risk of bias	Overall judged high risk of bias due to post-hoc analysis.

(Continued)

		and 21 participants randomised to the control arm.				not have an effect on the outcome.						
NCT00720941 **MSKCC intermediate risk group**	Low risk of bias	Interactive voice response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was not available for 17 participants randomised to the experimental arm and 21 participants randomised to the control arm.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 5 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT). Participants without MSKCC risk group allocation were excluded from subgroup analyses.	Low risk of bias	Less than 1% did not receive the intended interventions and therefore did not have outcome data. 2.9% of those who received treatment were lost to follow-up. Unclear to which risk group they were assigned to. However, these numbers are small and probably did not have an effect on the outcome.	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement.	High risk of bias	A post-hoc analysis according to risk group was conducted.	High risk of bias	Overall judged high risk of bias due to post-hoc analysis.
NCT00720941	Low risk of bias	Interactive voice	Low risk of bias	The study was open-label: both partici-	Low risk of bias	Less than 1% did	Low risk of bias	No precise in-	High risk of bias	A post-hoc	High risk of bias	Overall judged

(Continued)

MSKCC poor risk group		response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was not available for 17 participants randomised to the experimental arm and 21 participants randomised to the control arm.		pants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 5 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT). Participants without MSKCC risk group allocation were excluded from subgroup analyses.		not receive the intended interventions and therefore did not have outcome data. 2.9% of those who received treatment were lost to follow-up. Unclear to which risk group they were assigned to. However, these numbers are small and probably did not have an effect on the outcome.		formation provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement.		analysis according to risk group was conducted.		high risk of bias due to post-hoc analysis.
NCT00420889 **Total trial population**	Some concerns	Participants were randomised in a 1:1 ratio, but no information provided about who conducted the ran-	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 8 participants did not receive any treatment.	High risk of bias	1.5% did not receive the assigned interventions and therefore did not have outcome data. No information	Low risk of bias	No precise information provided about the outcome assessors, but either way knowl-	Some concerns	No study protocol or SAP available	High risk of bias	Overall judged high risk of bias due to lack of information about randomisation

(Continued)

		domisation and whether the allocation was concealed. There were no baseline imbalances that would suggest a problem with randomisation.		The method of analysis was appropriate (ITT).		whether there was loss to follow-up.		edge of intervention received could not have affected outcome measurement.				process and allocation concealment; missing outcome data; missing study protocol and SAP.
NCT00420888 **MSKCC favourable risk group**	Some concerns	Participants were randomised in a 1:1 ratio, but no information provided about who conducted the randomisation and whether the allocation was concealed. There were no baseline imbalances that would suggest a problem with randomi-	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 8 participants did not receive any treatment. The method of analysis was appropriate (ITT).	High risk of bias	1.5% did not receive the assigned interventions and therefore did not have outcome data. No information whether there was loss to follow-up.	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement.	Some concerns	No study protocol or SAP available	High risk of bias	Overall judged high risk of bias due to lack of information about randomisation process and allocation concealment; missing outcome data; missing study protocol and SAP.

(Continued)

		sation. MSKCC risk group was available for all randomised participants.										
NCT00420888 **MSKCC intermediate risk group**	Some concerns	Participants were randomised in a 1:1 ratio, but no information provided about who conducted the randomisation and whether the allocation was concealed. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was available for all randomised	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 8 participants did not receive any treatment. The method of analysis was appropriate (ITT).	High risk of bias	1.5% did not receive the assigned interventions and therefore did not have outcome data. No information whether there was loss to follow-up.	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement.	Some concerns	No study protocol or SAP available	High risk of bias	Overall judged high risk of bias due to lack of information about randomisation process and allocation concealment; missing outcome data; missing study protocol and SAP.

(Continued)

		participants.										
NCT00420888 **IMDC favourable risk group**	Some concerns	Participants were randomised in a 1:1 ratio, but no information provided about who conducted the randomisation and whether the allocation was concealed. There were no baseline imbalances that would suggest a problem with randomisation. HENG risk group was available for all randomised participants.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 8 participants did not receive any treatment. The method of analysis was appropriate (ITT).	High risk of bias	1.5% did not receive the assigned interventions and therefore did not have outcome data. No information whether there was loss to follow-up.	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement.	Some concerns	No study protocol or SAP available	High risk of bias	Overall judged high risk of bias due to lack of information about randomisation process and allocation concealment; missing outcome data; missing study protocol and SAP.
NCT00420888 **IMDC intermedi-**	Some concerns	Participants were randomised in a 1:1 ra-	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of	High risk of bias	1.5% did not receive the assigned interven-	Low risk of bias	No precise information provided	Some concerns	No study protocol or SAP available	High risk of bias	Overall judged high risk of bias due to

(Continued)

ate risk group		tio, but no information provided about who conducted the randomisation and whether the allocation was concealed There were no baseline imbalances that would suggest a problem with randomisation. HENG risk group was available for all randomised participants.		assigned interventions. Only 8 participants did not receive any treatment. The method of analysis was appropriate (ITT).		tions and therefore did not have outcome data. No information whether there was loss to follow-up.		about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement.				lack of information about randomisation process and allocation concealment; missing outcome data; missing study protocol and SAP.
NCT00420888 **IMDC poor risk group**	Some concerns	Participants were randomised in a 1:1 ratio, but no information provided about who conducted the randomisation and	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 8 participants did not receive any treatment. The method of analysis was appropriate (ITT).	High risk of bias	1.5% did not receive the assigned interventions and therefore did not have outcome data. No information whether there was	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of interven-	Some concerns	No study protocol or SAP available	High risk of bias	Overall judged high risk of bias due to lack of information about randomisation process and al-

(Continued)

		whether the allocation was concealed. There were no baseline imbalances that would suggest a problem with randomisation. HENG risk group was available for all randomised participants.				loss to follow-up.		tion received could not have affected outcome measurement.				location concealment; missing outcome data; missing study protocol and SAP.
NCT00979966 **Total trial population**	High risk of bias	Participants were randomised in a 1:1 ratio, but no information provided about who conducted the randomisation and whether the allocation was concealed. There were baseline imbalances that could	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. No information whether there were deviations form intended interventions and no information provided about the method of analysis.	High risk of bias	No information whether there was loss to follow-up.	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about the randomisation, the allocation concealment, the deviations from intended interventions, the

(Continued)

		suggest a problem with randomisation.										method of analysis and missing outcome data; missing study protocol and SAP.
NCT02761057 **SWOG** Comparison 1 (CAB vs. SUN) Total trial population (combined risk groups)	Low risk of bias	Randomisation was done by the Statistical Center. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 2 participants randomised to the experimental arm and 2 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	4.3% did not receive the intended interventions and therefore did not have outcome data. 2.2% had no protocol treatment. Only 1 participant was lost to follow-up. However, these numbers are low and probably did not have an effect on the outcome.	Low risk of bias	No precise information provided about the outcome assessors, but either way knowledge of intervention received could not have affected outcome measurement.	Some concerns	Study protocol available, but no original SAP.	Some concerns	Overall judged some concerns due to missing SAP.

Appendix 10. Risk of bias assessment for the outcome quality of life at the end of treatment

Trial	Risk of bias											
	Randomisation process		Deviations from intended interventions		Missing outcome data		Measurement of the outcome		Selection of the reported results		Overall	
	Authors' judgement	Support for judgement	Authors' judgement	Support for judgement	Authors' judgement	Support for judgement	Authors' judgement	Support for judgement	Authors' judgement	Support for judgement	Authors' judgement	Support for judgement
NCT00720941 **Instrument: FACIT-F** **Total trial population** **(combined risk groups)**	Low risk of bias	Interactive voice response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 5 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	High risk of bias	Less than 1% did not receive the intended interventions and therefore did not have outcome data. 90.7% of those randomised did not have evaluable outcome data at the end of treatment. No analysis to correct for missing outcome data.	High risk of bias	QoL is a participant-reported outcome, therefore outcome assessors were aware of the assigned intervention. Knowledge of intervention received could have affected outcome measurement.	Low risk of bias	CSR and study protocol with SAP available. Scale and time point were prespecified in the protocol and SAP.	High risk of bias	Overall judged high risk of bias due to the high number of participants without outcome data and the outcome assessors' awareness of assigned intervention.
NCT00098657/ NCT00083889	Some concerns	Participants were randomised	Low risk of bias	The study was open-label: both participants	Low risk of bias	2% did not receive the intended interventions	High risk of bias	QoL is a participant-reported outcome,	Some concerns	No study protocol or SAP	High risk of bias.	Overall judged high risk of bias due to lack

(Continued)

Instrument:FKSI-DRS **Total trial population** **(combined risk groups)**		in a 1:1 ratio, but no information about allocation concealment. However, there were no baseline imbalances that would suggest a problem with randomisation.		and those delivering the intervention were aware of assigned interventions. Only 15 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).		and therefore did not have outcome data. 91.6% of those randomised did not have evaluable outcome data the end of treatment. Analysis to correct for missing outcome data were conducted.		therefore outcome assessors were aware of the assigned intervention. Knowledge of intervention received could have affected outcome measurement.		available.		of information about the randomisation process and allocation concealment; the outcome assessors' awareness of assigned intervention; missing study protocol and SAP.
NCT00098652, NCT00083889 **Instrument: EQ-5D (VAS)** **Total trial population** **(combined risk groups)**	Some concerns	Participants were randomised in a 1:1 ratio, but no information about allocation concealment. However, there were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 15 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	2% did not receive the intended interventions and therefore did not have outcome data. 91.4% of those randomised did not have evaluable outcome data the end of treatment. Analysis to correct for missing outcome data were conducted.	High risk of bias	QoL is a participant-reported outcome, therefore outcome assessors were aware of the assigned intervention. Knowledge of intervention received could have affected outcome measurement.	Some concerns	No study protocol or SAP available.	High risk of bias.	Overall judged high risk of bias due to lack of information about the randomisation process and allocation concealment; the outcome assessors' awareness of assigned intervention; missing study protocol and SAP.

(Continued)

NCT00098657; NCT00083889 **Instrument: FACT-G** **Total trial population** **(combined risk groups)**	Some concerns	Participants were randomised in a 1:1 ratio, but no information about allocation concealment. However, there were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 15 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	2% did not receive the intended interventions and therefore did not have outcome data. 91.7% of those randomised did not have evaluable outcome data the end of treatment. Analysis to correct for missing outcome data were conducted.	High risk of bias	QoL is a participant-reported outcome, therefore outcome assessors were aware of the assigned intervention. Knowledge of intervention received could have affected outcome measurement.	Some concerns	No study protocol or SAP available.	High risk of bias.	Overall judged high risk of bias due to lack of information about the randomisation process and allocation concealment; the outcome assessors' awareness of assigned intervention; missing study protocol and SAP.
NCT01108445 **Instrument: FKSI-DRS** **Total trial population** **(combined risk groups)**	Low risk of bias	Participants were randomised in a 1:1 ratio. Randomisation was done under allocation concealment. There were no baseline imbalances that would suggest a problem with ran-	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned intervention. Only 1 participant withdrew after consent, but before randomisation and before study	High risk of bias	43.5% of those randomised did not have evaluable outcome data at the end of treatment. No analysis to correct for bias.	High risk of bias	QoL is a participant-reported outcome, therefore outcome assessors were aware of the assigned intervention. Knowledge of intervention received could have affected outcome measurement.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about method of analysis; high number of participants without outcome data; the outcome assessors' awareness of assigned intervention; missing

(Continued)

		domisation.		drug was assigned. No precise information provided about the method of analysis.								study protocol and SAP.
NCT00920816 **Instrument: FKSI-DRS** **Total trial population (combined risk groups)**	Low risk of bias	A centralised registration system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm did not receive any treatment. The method of analysis was appropriate (ITT).	HIgh risk of bias	1% did not receive the intended interventions and therefore did not have outcome data. 60.4% of those randomised did not have evaluable outcome data. No analysis to correct for bias.	High risk of bias	QoL is a participant-reported outcome, therefore outcome assessors were aware of the assigned intervention. Knowledge of intervention received could have affected outcome measurement.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to high number of participants without outcome data; the outcome assessors' awareness of assigned intervention; missing study protocol and SAP.
NCT00920816 **Instrument: EQ-5D (VAS)** **Total trial population**	Low risk of bias	A centralised registration system was used for randomisation. There were no baseline imbal-	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3	High risk of bias	1% did not receive the intended interventions and therefore did not have outcome data. 60.8% of those randomised did not have	High risk of bias	QoL is a participant-reported outcome, therefore outcome assessors were aware of the assigned intervention. Knowledge	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged high risk of bias due to high number of participants without outcome data; the outcome assessors' awareness of as-

(Continued)

(combined risk groups)	ances that would suggest a problem with randomisation.	participants randomised to the experimental arm did not receive any treatment. The method of analysis was appropriate (ITT).	evaluable outcome data. No analysis to correct for bias.	of intervention received could have affected outcome measurement.		signed intervention; missing study protocol and SAP.

Appendix 11. Risk of bias assessment for the outcome serious adverse events

Trial	Risk of bias											
	Randomisation process		Deviations from intended interventions		Missing outcome data		Measurement of the outcome		Selection of the reported results		Overall	
	Authors' judgement	Support for judgement	Authors' judgement	Support for judgement	Authors' judgement	Support for judgement	Authors' judgement	Support for judgement[B-B1]	Authors' judgement	Support for judgement	Authors' judgement	Support for judgement
NCT02811861 Comparison 1 (LEN +PEM vs. SUN) **Total trial population** **(combined risk groups)**	Low risk of bias	Interactive voice and web response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 17 participants randomised to the control arm did not receive any treatment. The method of analysis was not appropriate (as-treated).	Low risk of bias	2.8% did not receive the intended interventions and therefore did not have outcome data. However, these numbers are low and probably did not have an effect on the outcome.	High risk of bias	Measurement of SAEs could have differed between intervention groups due to longer follow-up of the intervention arm.	Low risk of bias	A study protocol with SAP available. Safety analysis was prespecified in the protocol and SAP.	High risk of bias	Overall judged high risk of bias due to inappropriate method of analysis; probable differences in outcome measurement between intervention arms.
NCT02811861 Comparison 2 (LEN +EVE vs. SUN) **Total trial**	Low risk of bias	Interactive voice and web response system was used for randomisation. There were no baseline	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 2 participants randomised to the experimen-	Low risk of bias	2.7% did not receive the intended interventions and therefore did not have outcome data. Howev-	Low risk of bias	Outcome assessors were not blinded. However, a standardised definition of SAEs was used and in-	Low risk of bias	A study protocol with SAP available. Safety analysis was prespecified in the	High risk of bias	Overall judged high risk of bias due to inappropriate method of analysis.

(Continued)

population (combined risk groups)		imbalances that would suggest a problem with randomisation.		tal arm and 17 participants randomised to the control arm did not receive any treatment. The method of analysis was not appropriate (as-treated).		er, these numbers are low and probably did not have an effect on the outcome.		cluded objective outcome events.		protocol and SAP.		
NCT00065468 Comparison 1 (TEM +IFN) **Total trial population (only intermediate and poor risk groups included in the trial)**	Some concerns	Participants were randomised, but no information provided about the allocation concealment. There were no baseline imbalances that would suggest a problem with randomisation.	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 1 participant randomised to the single-drug arm and 7 participants randomised to the control arm did not receive any treatment. No precise information provided about the method of analysis (as-treated is indicated).	Low risk of bias	1.9% did not receive the intended interventions and therefore did not have outcome data. However, these numbers are low and probably did not have an effect on the outcome.	Some concerns	No information provided about method of measuring SAEs. Outcome assessors were not blinded. However, a standardised definition of SAEs was used and included objective outcome events.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about allocation concealment, method of analysis and method of outcome measurement; missing study protocol and SAP.
NCT00065468 **comparison 2** Comparison 2 (IFN +TEM vs. IFN)	Some concerns	Participants were randomised, but no information provided about the allocation con-	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 2 participants randomised to	Low risk of bias	2.2% did not receive the intended interventions and therefore did not have outcome da-	Some concerns	No information provided about method of measuring SAEs. Outcome assessors were not	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about allocation concealment,

(Continued)

Total trial population (only intermediate and poor risk groups included in the trial)		cealment. There were no baseline imbalances that would suggest a problem with randomisation.		the combination arm and 7 participants randomised to the control arm did not receive any treatment. No precise information provided about the method of analysis (as-treated is indicated).		ta. However, these numbers are low and probably did not have an effect on the outcome.		blinded. However, a standardised definition of SAEs was used and included objective outcome events.				method of analysis and method of outcome measurement; missing study protocol and SAP.
NCT01024920 **Total trial population (combined risk groups)**	Low risk of bias	Interactive voice randomisation system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm did not receive any treatment. No precise information provided about the method of analysis.	Low risk of bias	3% did not receive the intended interventions and therefore did not have outcome data. However, these numbers are low and probably did not have an effect on the outcome.	Some concerns	No information provided about method of measuring SAEs. Outcome assessors were not blinded. However, a standardised definition of SAEs was used and included objective outcome events.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about method of analysis and method of outcome measurement; missing study protocol and SAP.
NCT01835158 **Total trial population (only inter-**	Low risk of bias	Randomisation was performed centrally. There were no baseline imbal-	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. On-	Low risk of bias	4.5% did not receive the intended interventions and therefore did not	Low risk of bias	Outcome assessors were not blinded. However, a standardised definition	Some concerns	Study protocol available with some statistical consider-	High risk of bias	Overall judged high risk of bias due to lack of information about the

(Continued)

mediate and poor risk groups included in the trial)		ances that would suggest a problem with randomisation.		ly 1 participant randomised to the experimental arm and 6 participants randomised to the control arm did not receive any treatment. No precise information provided about the method of analysis.		have outcome data. However, these numbers are low and probably did not have an effect on the outcome.		of SAEs was used and included objective outcome events.		ations briefly described, but no separate SAP available to fully check the preplanned analyses.		method of analysis; missing SAP.
NCT01984242 Comparison 1 (ATE vs. SUN) **Total population** **(combined risk groups)**	Low risk of bias	Interactive voice/web response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 1 participant randomised to the control arm did not receive any treatment. No precise information provided about the method of analysis, but data for crossed-over participants were reported separately. We assume participants were analysed as randomised in period 1.	Low risk of bias	0.3% did not receive the intended interventions and therefore did not have outcome data. However, these numbers are very low and probably did not have an effect on the outcome.	Some concerns	No information provided about method of measuring SAEs. Outcome assessors were not blinded. However, a standardised definition of SAEs was used and included objective outcome events.	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to lack of information about method of outcome measurement; missing study protocol and SAP.
NCT01984242 Comparison	Low risk of bias	Interactive voice/web response	Low risk of bias	The study was open-label: both participants and	Low risk of bias	0.3% did not receive the	Some concerns	No information provid-	Some concerns	No study protocol or SAP	Some concerns	Overall judged some con-

(Continued)

2 (ATE +BEV vs. SUN) **Total trial population** **(combined risk groups)**		system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.		those delivering the intervention were aware of assigned interventions. Only 1 participant randomised to the control arm did not receive any treatment. No precise information provided about the method of analysis, but data for crossed-over participants were reported separately. We assume participants were analysed as randomised in period 1.		intended interventions and therefore did not have outcome data. However, these numbers are very low and probably did not have an effect on the outcome.		ed about method of measuring SAEs. Outcome assessors were not blinded. However, a standardised definition of SAEs was used and included objective outcome events.		available.		cerns due to lack of information about method of outcome measurement; missing study protocol and SAP.
NCT00719264 **Total trial population** **(combined risk groups)**	Some concerns	Participants were randomised in a 1:1 ratio, but no information provided about who conducted the randomisation and whether the allocation was concealed. There were no baseline imbal-	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 1 participant randomised to the experimental arm and 2 participants randomised to the control arm did not receive any treatment; 1 participant randomised to the experimental arm had no post baseline safety assess-	Low risk of bias	1.1% did not receive the intended interventions and therefore did not have outcome data. However, these numbers are low and probably did not have an effect on the outcome.	Some concerns	No information provided about method of measuring SAEs. Outcome assessors were not blinded. However, a standardised definition of SAEs was used and included objective outcome events.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about randomisation process, allocation concealment, method of analysis and method of outcome measurement; missing study pro-

(Continued)

		ances that would suggest a problem with randomisation.		ment. No precise information provided about the method of analysis.								tocol and SAP.
NCT00720941 **Total trial population (combined risk groups)**	Low risk of bias	Interactive voice response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	Some concerns	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 5 participants randomised to the control arm did not receive any treatment. The method of analysis was not appropriate (as-treated), but this probably did not have an effect on the outcome as there is evidence that participants actually received the assigned intervention.	Low risk of bias	Less than 1% did not receive the intended interventions and therefore did not have outcome data. However, these numbers are small and probably did not have an effect on the outcome.	Low risk of bias	Outcome assessors were not blinded. However, a standardised definition of SAEs was used and included objective outcome events.	Low risk of bias	CSR and study protocol with SAP available. Safety analysis was prespecified in the protocol and SAP.	Some concerns	Overall judged some concerns due to inappropriate method of analysis.
NCT00732914 **Total trial population**	Low risk of bias	Randomisation was performed centrally. There were no baseline	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned in-	Low risk of bias	3.3% did not receive the intended interventions and therefore	High risk of bias	No information provided about outcome assessors and	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about

(Continued)

(combined risk groups)		imbalances that would suggest a problem with randomisation.		terventions. Only 5 participants randomised to the experimental arm and 7 participants randomised to the control arm did not receive any treatment. No precise information provided about the method of analysis, but data for first period reported separately. We assume participants in first period received their allocated intervention.		did not have outcome data. However, these numbers are low and probably did not have an effect on the outcome.		method of measuring SAEs.				outcome assessors and method of outcome measurement; missing study protocol and SAP.
NCT01613846 **Total trial population** **(combined risk groups)**	Some concerns	Participants were randomised in a 1:1 ratio, but no information provided about who conducted the randomisation and whether the allocation was concealed. There were no baseline imbal-	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 6 participants randomised to the one experimental arm and 5 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate.	Low risk of bias	2.9% did not receive the intended interventions and therefore did not have outcome data. However, these numbers are low and probably did not have an effect on the outcome.	High risk of bias	No information provided about outcome assessors and method of measuring SAEs.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about randomisation process, allocation concealment, outcome assessors and method of outcome measurement; missing study pro-

(Continued)

		ances that would suggest a problem with randomisation.										tocol and SAP.
NCT00738530 **Total trial population (combined risk groups)**	Low risk of bias	Interactive voice recognition system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	High risk of bias	The study was double-blind: both participants and those delivering the intervention were not aware of assigned interventions. Only 2 participants from the control arm and 6 participants from the intervention arm did not receive any treatment. The method of analysis was not appropriate (as-treated).	Low risk of bias	1.2% did not receive the intended interventions and therefore did not have outcome data. However, these numbers are low and probably did not have an effect on the outcome.	Some concerns	No information provided about method of measuring SAEs and outcome assessors. However, we assume SAEs were assessed by the investigators and this was a double-blind study. Furthermore, a standardised definition of SAEs was used and included objective outcome events.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to inappropriate method of analysis; lack of information about method of outcome measurement; missing study protocol and SAP.
NCT00117637 NCT00117637 **Total trial**	High risk of bias	Participants were randomised in a 1:1 ra-	Low risk of bias	The study was open-label: both participants and those delivering the intervention	Low risk of bias	No indication of loss to follow-up.	Some concerns	No information provided about method of	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack

(Continued)

population **(combined risk groups)**		tio, but no information provided about the allocation concealment. There were some baseline imbalances that could suggest a problem with randomisation.		were aware of assigned interventions. No precise information provided about the method of analysis, but data for crossed-over participants were reported separately. We assume participants were analysed as randomised in period 1.				measuring SAEs. Outcome assessors were not blinded. However, a standardised definition of SAEs was used and included objective outcome events.			of information about the allocation concealment and method of outcome measurement; baseline imbalances; missing study protocol and SAP.
NCT00098655/ NCT00083889 **Total trial population** **(combined risk groups)**	Some concerns	Participants were randomised in a 1:1 ratio, but no information about allocation concealment. However, there were no baseline imbalances that would suggest a problem with randomisation.	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 15 participants randomised to the control arm did not receive any treatment. No precise information provided about the method of analysis.	Low risk of bias	2% did not receive the intended interventions and therefore did not have outcome data.	Some concerns	No information provided about method of measuring SAEs. Outcome assessors were not blinded. However, a standardised definition of SAEs was used and included objective outcome events.	Some concerns	No study protocol or SAP available.	High risk of bias.	Overall judged high risk of bias due to lack of information about the randomisation process, allocation concealment, method of analysis and method of outcome measurement; missing study protocol and SAP.

(Continued)

NCT01108445 **Total trial population** **(combined risk groups)**	Low risk of bias	Participants were randomised in a 1:1 ratio. Randomisation was done under allocation concealment. There were no baseline imbalances that would suggest a problem with randomisation.	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned intervention. Only 1 participant withdrew after consent, but before randomisation and before study drug was assigned. No precise information provided about the method of analysis.	Low risk of bias	All 108 participants were evaluable.	Some concerns	No information provided about method of measuring SAEs. Outcome assessors were not blinded. However, a standardised definition of SAEs was used and included objective outcome events.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about method of analysis and method of outcome measurement; missing study protocol and SAP.
NCT00903175 **Total trial population** **(combined risk groups)**	Low risk of bias	Interactive voice response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 2 participants randomised to the control arm did not receive any treatment. No precise information provided about the method of analysis, but data for crossed-over participants were reported separate-	Low risk of bias	Less than 1% did not receive the intended interventions and therefore did not have outcome data. However, these numbers are low and probably did not have an effect on the outcome.	Low risk of bias	Outcome assessors were not blinded. However, a standardised definition of SAEs was used and included objective outcome events.	Some concerns	Study protocol available with some statistical methods described, but no separate SAP available to fully check the preplanned analyses.	Some concerns	Overall judged some concerns due to missing SAP.

(Continued)

				ly. We assume participants were analysed as randomised in period 1.								
NCT00619268 Comparison 1 (BEV +TEM vs. SUN) **Total trial population** **(combined risk groups)**	Low risk of bias	A computerised centrally located randomisation system was Used. There were no baseline imbalances that would suggest a problem with randomisation.	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned intervention. Only 1 participant randomised to the control arm did not receive any treatment. No precise information provided about the method of analysis.	Low risk of bias	Only 1 participant did not receive the intended intervention and therefore did not have outcome data.	High risk of bias	No information provided about outcome assessors and method of measuring SAEs.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about method of analysis, outcome assessors and method of outcome measurement; missing study protocol and SAP.
NCT00619268 Comparison 2 (BEV +IFN vs. SUN) **Total trial population** **(combined risk groups)**	Low risk of bias	A computerised centrally located randomisation system was Used. There were no baseline imbalances that would suggest a problem with ran-	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned intervention. Only 1 participant randomised to the control arm did not receive any treatment. No precise information provided about the method of analysis.	Low risk of bias	Only 1 participant did not receive the intended intervention and therefore did not have outcome data.	High risk of bias	No information provided about outcome assessors and method of measuring SAEs.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about method of analysis, outcome assessors and method of outcome measurement; missing study pro-

(Continued)

		domisation.										tocol and SAP.
NCT00631371 **Total trial population** **(combined risk groups)**	Low risk of bias	A computerised centrally located randomisation system was used. There were no baseline imbalances that would suggest a problem with randomisation.	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 7 participants randomised to the experimental arm did not receive any treatment. No precise information provided about the method of analysis.	Low risk of bias	Less than 1% did not receive the intended interventions and therefore, did not have outcome data. However, these numbers are low and probably did not have an effect on the outcome.	Some concerns	No information provided about method of measuring SAEs. Outcome assessors were not blinded. However, a standardised definition of SAEs was used and included objective outcome events.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about method of analysis; missing study protocol and SAP.
NCT02231749 **Total trial population** **(combined risk groups)**	Low risk of bias	Interactive voice response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm, and 11 participants randomised to the control arm did not receive any treatment. No precise information provid-	Low risk of bias	1.3% did not receive the intended interventions and therefore did not have outcome data. However, these numbers are very low and probably did not have an ef-	Low risk of bias	Outcome assessors were not blinded. However, a standardised definition of SAEs was used and included objective outcome events.	Low risk of bias	A study protocol with SAP available. Safety analysis was prespecified in the protocol and SAP.	High risk of bias	Overall judged high risk of bias due to lack of information about method of analysis.

(Continued)

				ed about the method of analysis.		fect on the outcome.						
NCT02420821 **Total trial population (combined risk groups)**	Low risk of bias	Interactive voice and web response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 15 participants randomised to the control arm did not receive any treatment. Conflicting information about method of analysis in the protocol.	Low risk of bias	2% did not receive the intended interventions and therefore did not have outcome data. However, these numbers are low and probably did not have an effect on the outcome.	Low risk of bias	Outcome assessors were not blinded. However, a standardised definition of SAEs was used and included objective outcome events.	Low risk of bias	A study protocol with SAP available. Safety analysis was prespecified in the protocol and SAP.	High risk of bias	Overall judged high risk of bias due to conflicting information about method of analysis.
NCT02853331 **Total trial population (combined risk groups)**	Low risk of bias	Interactive voice response system or integrated web response system was used for randomisation. There were no baseline imbalances that would suggest a	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 4 participants randomised to the control arm did not receive any treatment. The method of analy-	Low risk of bias	Less than 1% did not receive the intended interventions and therefore did not have outcome data. However, these numbers are low and probably did not have	Low risk of bias	Outcome assessors were not blinded. However, a standardised definition of SAEs was used and included objective outcome events.	Low risk of bias	A study protocol with SAP available. Safety analysis was prespecified in the protocol and SAP.	Low risk of bias	Overall judged low risk of bias.

(Continued)

		problem with randomisation.		sis was appropriate.		an effect on the outcome.						
NCT00920816 **Total trial population (combined risk groups)**	Low risk of bias	A centralised registration system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm did not receive any treatment. No precise information provided about the method of analysis, but data for first period reported separately. We assume participants in first period received their allocated intervention.	Low risk of bias	1% did not receive the intended interventions and therefore did not have outcome data. However, these numbers are low and probably did not have an effect on the outcome.	Some concerns	No information provided about method of measuring SAEs. Outcome assessors were not blinded. However, a standardised definition of SAEs was used and included objective outcome events.	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to lack of information about method of outcome measurement; missing study protocol and SAP.
NCT00979966 **Total trial population (combined risk groups)**	High risk of bias	Participants were randomised in a 1:1 ratio, but no information provided about who conducted randomisation and	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. No information whether there were deviations form intended interventions and no in-	Low risk of bias	No indication of loss to follow-up.	Some concerns	No information provided about method of measuring SAEs. Outcome assessors were not blinded. However, a stan-	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about randomisation process, allocation concealment,

(Continued)

		whether allocation was concealed. There were baseline imbalances that could suggest a problem with randomisation. Small study population.		formation provided about the method of analysis.				dardised definition of SAEs was used and included objective outcome events.				deviations from intended interventions, method of analysis and method of outcome measurement; baseline imbalances; missing study protocol and SAP.
NCT00126594 **Total trial population** **(combined risk groups)**	Some concerns	No information provided about randomisation process and allocation concealment. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. The method of analysis was appropriate.	Low risk of bias	No indication of loss to follow-up.	Some concerns	No information provided about method of measuring SAEs. Outcome assessors were not blinded. However, a standardised definition of SAEs was used and included objective outcome events.	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to lack of information about randomisation process, allocation concealment, method of outcome measurement; missing study protocol and SAP.

Appendix 12. Risk of bias assessment for the outcome progression-free survival

Trial	Risk of bias											
	Randomisation process	Deviations from intended interventions		Missing outcome data		Measurement of the outcome		Selection of the reported results		Overall	
	Authors' judgement	Support for judgement	Authors' judgement	Support for judgement	Authors' judgement	Support for judgement	Authors' judgement	Support for judgement	Authors' judgement	Support for judgement	Authors' judgement	Support for judgement
NCT03141177 IMDC favourable risk group	Low risk of bias	In- ter-bias ac- tive Re- sponse Tech- nol- o- gy was used for ran- domi- sa- tion. There were no base- line im- bal- ances that would sug- gest a prob- lem	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 8 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	High risk of bias	1.7% did not receive the intended interventions and therefore did not have outcome data. No information whether there was loss to follow-up. 3% of those who received treatment discontinued due to "other" reasons (not explained further).	Low risk of bias	Outcome assessors were not aware of the assigned intervention. PFS was assessed by a blinded independent central review committee.	High risk of bias	A study protocol with SAP available. The subgroup analyses according to IMDC risk group were pre-specified in the protocol and SAP. However, the time point that produced this result was not pre-specified in the protocol. The results of the final PFS analysis were already reported in a previous publication.	High risk of bias	Overall judged high risk of bias due to lack of information about missing outcome data and inconsistency with the protocol regarding the time point of analyses and reporting.

(Continued)

		with randomisation. IMDC risk group was available for all randomised participants.									
NCT03141177 **IMDC intermediate risk group**	Low risk of bias	Inter-active Response Technology was used for randomisation. There were no baseline imbalances that	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 8 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	High risk of bias	1.7% did not receive the intended interventions and therefore did not have outcome data. No information whether there was loss to follow-up. 3% of those who received treatment discontinued due to "other" reasons (not explained further).	Low risk of bias	Outcome assessors were not aware of the assigned intervention. PFS was assessed by a blinded independent central review committee.	High risk of bias	A study protocol with SAP available. The subgroup analyses according to IMDC risk group were pre-specified in the protocol and SAP. However, the time point that produced this result was not pre-specified in the protocol. The results of the final PFS analysis were already reported in a previous publication.	High risk of bias	Overall judged high risk of bias due to lack of information about missing outcome data and inconsistency with the protocol regarding the time point of analyses and reporting.

(Continued)

		would suggest a problem with randomisation. IMDC risk group was available for all randomised participants.									
NCT03141177 **IMDC poor risk group**	Low risk of bias	Inter-active Response Technology was used for randomisation. There were no	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 8 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	High risk of bias	1.7% did not receive the intended interventions and therefore did not have outcome data. No information whether there was loss to follow-up. 3% of those who received treatment discontinued due to "other" reasons (not explained further).	Low risk of bias	Outcome assessors were not aware of the assigned intervention. PFS was assessed by a blinded independent central review committee.	High risk of bias	A study protocol with SAP available. The subgroup analyses according to IMDC risk group were pre-specified in the protocol and SAP. However, the time point that produced this result was not pre-specified in the protocol. The results of the final PFS analysis were already reported in a previous publication.	High risk of bias	Overall judged high risk of bias due to lack of information about missing outcome data and inconsistency with the protocol regarding the time point of analyses and reporting.

(Continued)

		baseline imbalances that would suggest a problem with randomisation. IMDC risk group was available for all randomised participants.									
NCT02811861 Comparison 1 (LEN+PEM vs.SUN) **Total tri-**	Low risk of bias	Inter-bias active voice and web response system was used for	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 17 participants randomised to the	High risk of bias	2.8% did not receive the intended interventions and therefore did not have outcome data. No information whether there was loss to follow-up. 2% discontinued treatment due to	High risk of bias	Independent review was conducted but no statement whether it was blinded. Knowledge of intervention received could have affected outcome measurement.	Low risk of bias	A study protocol with SAP available. All reported analyses were prespecified in the protocol and SAP.	High risk of bias	Overall judged high risk of bias due to lack of information about missing outcome data; the outcome assessors' probable awareness

(Continued)

al pop-u-la-tion (com-bined risk groups)		ran-domi-sa-tion. There were no base-line im-bal-ances that would sug-gest a prob-lem with ran-domi-sa-tions.	control arm did not receive any treatment. The method of analy-sis was appropri-ate (ITT).		“other” rea-sons (not ex-plained fur-ther).						of the as-signed in-terven-tions.
NCT02811861 Com-par-i-son 1 (LEN +PEM vs.SUN) **MSKCC favourable risk group**	Low risk of bias	In-ter-bias ac-tive voice and web re-sponse sys-tem was used for ran-domi-sa-tion. There were no	The study was open-label: both participants and those deliver-ing the interven-tion were aware of assigned in-terventions. On-ly 3 participants randomised to the experimen-tal arm and 17 participants ran-domised to the control arm did not receive any treatment. The method of analy-sis was appropri-ate (ITT).	High risk of bias	2.8% did not receive the intended in-terventions and there-fore did not have outcome data. No in-formation whether there was loss to follow-up. 2% discontinued treatment due to “oth-er” reasons (not further explained).	High risk of bias	Indepen-dent re-view was conduct-ed but no statement whether it was blind-ed. Knowl-edge of in-tervention received could have affected outcome measure-ment.	Low risk of bias	A study protocol with SAP avail-able. The sub-group analy-ses according to IMDC risk group were pre-speci-fied in the proto-col and SAP.	High risk of bias	Overall judged high risk of bias due to lack of informa-tion about missing outcome data; the outcome assessors’ probable awareness of the as-signed in-terven-tions.

(Continued)

		baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was available for all randomised participants.									
NCT02811861 Comparison 1 (LEN+PEM vs.SUN) **MSKCC intermediate/poor risk**	Low risk of bias	Interbias active voice and web response system was used for	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 17 participants randomised to the	High risk of bias	2.8% did not receive the intended interventions and therefore did not have outcome data. No information whether there was loss to follow-up. 2% discontinued treatment due to	High risk of bias	Independent review was conducted but no statement whether it was blinded. Knowledge of intervention received could have affected	Low risk of bias	A study protocol with SAP available. The subgroup analyses according to IMDC risk group were pre-specified in the protocol and SAP.	High risk of bias	Overall judged high risk of bias due to lack of information about missing outcome data; the outcome assessors' probable awareness

(Continued)

medi-ate risk group		ran-domi-sa-tion. There were no base-line im-bal-ances that would sug-gest a prob-lem with ran-domi-sa-tion. MSKCC risk group was avail-able for all ran-domised par-tic-i-pants.	control arm did not receive any treatment. The method of analy-sis was appropri-ate (ITT).		“other” rea-sons (not ex-plained fur-ther).		outcome measure-ment.				of the as-signed in-terven-tions.
NCT02811861	Low risk of bias	In- Low risk of ter-bias ac-tive voice and web	The study was open-label: both participants and those deliver-ing the interven-tion were aware of assigned in-	High risk of bias	2.8% did not receive the intended in-terventions and there-fore did not have outcome	High risk of bias	Indepen-dent re-view was conduct-ed but no statement whether it	Low risk of bias	A study protocol with SAP avail-able. The sub-group analy-ses according to IMDC risk group were pre-speci-	High risk of bias	Overall judged high risk of bias due to lack of informa-tion about
Com-par-i-son 1											

(Continued)

(LEN +PEM vs.SUN)						
MSKCC poor risk group	response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was available for all randomised participants.	terventions. Only 3 participants randomised to the experimental arm and 17 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	data. No information whether there was loss to follow-up. 2% discontinued treatment due to "other" reasons (not further explained).	was blinded. Knowledge of intervention received could have affected outcome measurement.	fied in the protocol and SAP.	missing outcome data; the outcome assessors' probable awareness of the assigned interventions.

(Continued)

NCT02811861 Comparison 1 (LEN+PEM vs.SUN) **IMDC favourable risk group**	Low risk of bias	Low risk of bias	In-active voice and web response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. IMDC risk group was not available	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 17 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT). Participants without IMDC risk group allocation were excluded from subgroup analyses.	High risk of bias	2.8% did not receive the intended interventions and therefore did not have outcome data. No information whether there was loss to follow-up. 2% discontinued treatment due to "other" reasons (not explained further).	High risk of bias	Independent review was conducted but no statement whether it was blinded. Knowledge of intervention received could have affected outcome measurement.	Low risk of bias	A study protocol with SAP available. The subgroup analyses according to IMDC risk group were pre-specified in the protocol and SAP.	High risk of bias	Overall judged high risk of bias due to lack of information about missing outcome data; the outcome assessors' probable awareness of the assigned interventions.

(Continued)

		for 2 participants randomised to the experimental arm and 4 participants randomised to the control arm.									
NCT02811861 Comparison 1 (LEN+PEM vs.SUN) **IMDC intermedi-**	Low risk of bias	Inter-active voice and web response system was used for randomi-	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 17 participants randomised to the control arm did not receive any	High risk of bias	2.8% did not receive the intended interventions and therefore did not have outcome data. No information whether there was loss to follow-up. 2% discontinued treatment due to "other" reasons (not ex-	High risk of bias	Independent review was conducted but no statement whether it was blinded. Knowledge of intervention received could have affected outcome	Low risk of bias	A study protocol with SAP available. The subgroup analyses according to IMDC risk group were pre-specified in the protocol and SAP.	High risk of bias	Overall judged high risk of bias due to lack of information about missing outcome data; the outcome assessors' probable awareness of the assigned in-

(Continued)

ate risk group	sation. There were no baseline imbalances that would suggest a problem with randomisation. IMDC risk group was not available for 2 participants randomised to the experimental arm and	treatment. The method of analysis was appropriate (ITT). Participants without IMDC risk group allocation were excluded from subgroup analyses.	plained further).	measurement.	terventions.

(Continued)

		4 participants randomised to the control arm.									
NCT02811861 Comparison 1 (LEN+PEM vs.SUN) **IMDC poor risk group**	Low risk of bias	Low risk of bias active voice and web response system was used for randomisation. There were no baseline imbalances that would suggest a problem	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 17 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT). Participants without IMDC risk group allocation were excluded from subgroup analyses.	High risk of bias	2.8% did not receive the intended interventions and therefore did not have outcome data. No information whether there was loss to follow-up. 2% discontinued treatment due to "other" reasons (not explained further).	High risk of bias	Independent review was conducted but no statement whether it was blinded. Knowledge of intervention received could have affected outcome measurement.	Low risk of bias	A study protocol with SAP available. The subgroup analyses according to IMDC risk group were pre-specified in the protocol and SAP.	High risk of bias	Overall judged high risk of bias due to lack of information about missing outcome data; the outcome assessors' probable awareness of the assigned interventions.

(Continued)

		with randomisation. IMDC risk group was not available for 2 participants randomised to the experimental arm and 4 participants randomised to the control arm.									
NCT02811861 Comparator	Low risk of bias	Inter-active	Low risk of bias	The study was open-label: both participants and those deliver-	High risk of bias	2.7% did not receive the intended interventions and	High risk of bias	Independent review was conducted	Low risk of bias	A study protocol with SAP available. All reported analyses were	High risk of bias	Overall judged high risk of bias due

(Continued)

i-son 2 (LEN +EVE vs. SUN) **Total trial population (combined risk groups)**		voice and web response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	ing the intervention were aware of assigned interventions. Only 2 participants randomised to the experimental arm and 17 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).		therefore did not have outcome data. No information whether there was loss to follow-up. 2% discontinued treatment due to "other" reasons (not explained further).		ed but no statement whether it was blinded. Knowledge of intervention received could have affected outcome measurement.		pre-specified in the protocol and SAP.		to lack of information about missing outcome data; the outcome assessors' probable awareness of the assigned interventions.
NCT02811861 Comparison 2 (LEN +EVE vs. SUN)	Low risk of bias	In-ter-bias active voice and web response system	Low risk of bias The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 2 participants randomised to the experimen-	High risk of bias	2.7% did not receive the intended interventions and therefore did not have outcome data. No information whether there was loss to follow-up.	High risk of bias	Independent review was conducted but no statement whether it was blinded. Knowledge of intervention	Low risk of bias	A study protocol with SAP available. The subgroup analyses according to IMDC risk group were pre-specified in the protocol and SAP.	High risk of bias	Overall judged high risk of bias due to lack of information about missing outcome data; the outcome

(Continued)

MSKCC favourable risk group	was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was available for all randomised participants.	tal arm and 17 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).		2% discontinued treatment due to "other" reasons (not explained further).		received could have affected outcome measurement.				assessors' probable awareness of the assigned interventions.
NCT02811861 Comparator	Low risk of bias	Inter-active	Low risk of bias	The study was open-label: both participants and those deliver-	High risk of bias	2.7% did not receive the intended interventions and	High risk of bias	Independent review was conduct-	Low risk of bias	A study protocol with SAP available. The subgroup analy-	High risk of bias	Overall judged high risk of bias due

(Continued)

i-son 2 (LEN +EVE vs. SUN)						
MSKCC intermediate risk group	voice and web response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was available for all randomised participants according to IMDC risk group were pre-specified in the protocol and SAP.	ing the intervention were aware of assigned interventions. Only 2 participants randomised to the experimental arm and 17 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	therefore did not have outcome data. No information whether there was loss to follow-up. 2% discontinued treatment due to "other" reasons (not explained further).	ed but no statement whether it was blinded. Knowledge of intervention received could have affected outcome measurement.		to lack of information about missing outcome data; the outcome assessors' probable awareness of the assigned interventions.

(Continued)

		i-pants.									
NCT02811861 Comparison 2 (LEN +EVE vs. SUN) **MSKCC poor risk group**	Low risk of bias	In- ter-bias ac- tive voice and web re- sponse sys- tem was used for ran- domi- sa- tion. There were no base- line im- bal- ances that would sug- gest a prob- lem with ran- domi- sa- tion. MSKCC risk group was avail-	Low risk of bias	The study was open-label: both participants and those deliver- ing the interven- tion were aware of assigned in- terventions. On- ly 2 participants randomised to the experimen- tal arm and 17 participants ran- domised to the control arm did not receive any treatment. The method of analy- sis was appropri- ate (ITT).	High risk of bias	2.7% did not receive the in- tended inter- ventions and therefore did not have out- come data. No informa- tion whether there was loss to follow-up. 2% discon- tinued treat- ment due to "other" rea- sons (not ex- plained fur- ther).	High risk of bias	Indepen- dent re- view was conduct- ed but no statement whether it was blind- ed. Knowl- edge of in- tervention received could have affected outcome measure- ment.	Low risk of bias	A study protocol with SAP avail- able. The sub- group analy- ses according to IMDC risk group were pre-speci- fied in the proto- col and SAP.	High risk of bias	Overall judged high risk of bias due to lack of informa- tion about missing outcome data; the outcome assessors' probable awareness of the as- signed in- terven- tions.

(Continued)

		able for all randomised participants.									
NCT02811861 Comparison 2 (LEN+EVE vs. SUN) **IMDC favourable risk group**	Low risk of bias	Inter-active voice and web response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomi-	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 2 participants randomised to the experimental arm and 17 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT). Participants without IMDC risk group allocation were excluded from subgroup analyses.	High risk of bias	2.7% did not receive the intended interventions and therefore did not have outcome data. No information whether there was loss to follow-up. 2% discontinued treatment due to "other" reasons (not explained further).	High risk of bias	Independent review was conducted but no statement whether it was blinded. Knowledge of intervention received could have affected outcome measurement.	Low risk of bias	A study protocol with SAP available. The subgroup analyses according to IMDC risk group were pre-specified in the protocol and SAP.	High risk of bias	Overall judged high risk of bias due to lack of information about missing outcome data; the outcome assessors' probable awareness of the assigned interventions.

(Continued)

		sation. IMDC risk group was not available for 6 participants randomised to the experimental arm and 4 participants randomised to the control arm.									
NCT02811861 Comparison 2	Low risk of bias	Low risk of bias	In-ter-ac-tive voice and web	The study was open-label: both participants and those delivering the intervention were aware of assigned intervention.	High risk of bias	2.7% did not receive the intended interventions and therefore did not have outcome data.	High risk of bias	Independent review was conducted but no statement whether it was blinded.	Low risk of bias	A study protocol with SAP available. The subgroup analyses according to IMDC risk group were pre-specified.	High risk of bias	Overall judged high risk of bias due to lack of information about

(Continued)

(LEN +EVE vs. SUN) **IMDC intermediate risk group**	response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. IMDC risk group was not available for 6 participants randomised	terventions. Only 2 participants randomised to the experimental arm and 17 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT). Participants without IMDC risk group allocation were excluded from subgroup analyses.	No information whether there was loss to follow-up. 2% discontinued treatment due to "other" reasons (not explained further).	was blinded. Knowledge of intervention received could have affected outcome measurement.	fied in the protocol and SAP.	missing outcome data; the outcome assessors' probable awareness of the assigned interventions.

(Continued)

		to the experimental arm and 4 participants randomised to the control arm.									
NCT02811861 Comparison 2 (LEN+EVE vs. SUN) **IMDC poor risk group**	Low risk of bias	Inter-active voice and web response system was used for randomisation. There were no baseline im-	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 2 participants randomised to the experimental arm and 17 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT). Participants without IMDC risk group allocation were excluded from	High risk of bias	2.7% did not receive the intended interventions and therefore did not have outcome data. No information whether there was loss to follow-up. 2% discontinued treatment due to "other" reasons (not explained further).	High risk of bias	Independent review was conducted but no statement whether it was blinded. Knowledge of intervention received could have affected outcome measurement.	Low risk of bias	A study protocol with SAP available. The subgroup analyses according to IMDC risk group were pre-specified in the protocol and SAP.	High risk of bias	Overall judged high risk of bias due to lack of information about missing outcome data; the outcome assessors' probable awareness of the assigned interventions.

(Continued)

balances that would suggest a problem with randomisation. IMDC risk group was not available for 6 participants randomised to the experimental arm and 4 participants randomised to	subgroup analyses.

(Continued)

		the control arm.									
NCT01108445 **Total trial population (combined risk groups)**	Low risk of bias	Participants were randomised in a 1:1 ratio. Randomisation was done under allocation concealment. There were no baseline imbalances that would suggest a prob-	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned intervention. Only 1 participant withdrew after consent, but before randomisation and before study drug was assigned. The method of analysis was appropriate (ITT).	Low risk of bias	All 108 participants were evaluable.	High risk of bias	Scans were read by a trained radiologist, but no information whether the person was blinded. Knowledge of intervention received could have affected outcome measurement.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to the outcome assessors' probable awareness of the assigned interventions; missing study protocol and SAP.

(Continued)

		lem with randomisation.								
NCT01108445 **MSKCC favourable risk group**	Low risk of bias	Participants were randomised in a 1:1 ratio. Randomisation was done under allocation concealment. There were no baseline imbalances that would suggest	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned intervention. Only 1 participant withdrew after consent, but before randomisation and before study drug was assigned. The method of analysis was appropriate (ITT).	Low risk of bias	All 108 participants were evaluable.	High risk of bias	Scans were read by a trained radiologist, but no information whether the person was blinded. Knowledge of intervention received could have affected outcome measurement.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to the outcome assessors' probable awareness of the assigned interventions; missing study protocol and SAP.

(Continued)

		a problem with randomisation. MSKCC risk group was available for all randomised participants.									
NCT01108445 **MSKCC intermediate risk group**	Low risk of bias	Par Low risk of tic- bias i- pants were randomised in a 1:1 ratio. Randomisation was done under allo- ca-	The study was open-label: both participants and those delivering the intervention were aware of assigned intervention. Only 1 participant withdrew after consent, but before randomisation and before study drug was assigned enrolment and randomisation). The method of analysis was appropriate (ITT).	Low risk of bias	All 108 participants were evaluable	High risk of bias	Scans were read by a trained radiologist, but no information whether the person was blinded. Knowledge of intervention received could have affected outcome measurement.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to the outcome assessors' probable awareness of the assigned interventions; missing study protocol and SAP.

(Continued)

		tion concealment. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was available for all randomised participants.									
NCT01108445 **MSKCC poor risk group**	Low risk of bias	Par-tic-i-pants were ran-domised	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assignment	Low risk of bias	All 108 participants were evaluable	High risk of bias	Scans were read by a trained radiologist, but no information	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to the outcome assessors'

(Continued)

	in a 1:1 ratio. Randomisation was done under allocation concealment. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was available	signed intervention. Only 1 participant withdrew after consent, but before randomisation and before study drug was assigned. The method of analysis was appropriate (ITT).	whether the person was blinded. Knowledge of intervention received could have affected outcome measurement.	probable awareness of the assigned interventions; missing study protocol and SAP.

(Continued)

		for all randomised participants.									
NCT01354002	Some concerns	Participants were randomised in a 1:1 ratio, but no information provided about the allocation concealment. There were no baseline imbal-	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. No statement about the method of analysis.	Low risk of bias	Detailed flow diagram provided, no indication of loss to follow-up.	Low risk of bias	Outcome assessors were not aware of the assigned intervention. The radiographic response was assessed by blinded radiologists.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about the allocation concealment and method of analysis; missing study protocol and SAP.
Total trial population (only intermediate and poor risk groups included in the trial)											

(Continued)

		ances that would suggest a problem with randomisation.									
NCT00134823 **Total trial population (combined risk groups)**	Low risk of bias	Inter-active voice response system was used for randomisation. There were no baseline imbalances that would suggest a problem with	Low risk of bias	The study was blinded: both participants and those delivering the intervention were not aware of assigned interventions. The method of analysis was appropriate (ITT).	Low risk of bias	1.3% of those who received treatment were lost to follow-up. No analysis to correct for bias, but numbers are low and probably did not have an effect on the outcome.	Low risk of bias	Outcome assessors were not aware of the assigned intervention. All imaging scans were evaluated by an independent imaging-review committee (IRC) blinded to study treatment.	Low risk of bias	CSR and study protocol with SAP available. All reported analyses were pre-specified in the protocol and SAP.	Low risk of bias	Overall judged low risk of bias.

(Continued)

		ran-domi-sa-tion.									
NCT01307813 **To-tal tri-al pop-u-la-tion (com-bined risk groups)**	Low risk of bias	In-ter-bias ac-tive voice re-sponse sys-tem was used for ran-domi-sa-tion. There were no base-line im-bal-ances that would sug-gest a prob-lem with ran-domi-sa-tion.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned inter-ventions. Only 1 participant ran-domised to the experimental arm did not receive any treatment. The method of analysis was ap-propriate (ITT).	High risk of bias	No informa-tion provid-ed about loss to follow-up. 2.3% discon-tinued due to "other" rea-sons (not ex-plained fur-ther).	Low risk of bias	Outcome assessors were not aware of the as-signed in-terven-tion. PFS was as-sessed by a blinded indepen-dent radi-ology re-view.	Some con-cerns	No SAP available. Study protocol available, but un-clear whether it was finalized be-fore unblinded outcome data were available.	High risk of bias	Overall judged high risk of bias due to lack of informa-tion about miss-ing out-come da-ta; missing study pro-tocol and SAP.
NCT00326469 Com-par-	Some con-cerns	Par-tic-bias i-	Low risk of	The study was open-label: both participants and	Low risk	1.9% did not receive the intended in-	Low risk of bias	Outcome assessors were not	Some con-cerns	No study protocol or SAP available.	Some con-cerns	Overall judged some con-									

(Continued)

i-son 1 (TEM vs. IFN) **Total trial population (only intermediate and poor risk groups included in the trial)**		pants were randomised, but no information provided about allocation concealment. There were no baseline imbalances that would suggest a problem with randomisation.	those delivering the intervention were aware of assigned interventions. Only 1 participant randomised to the single-drug arm and 7 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	of bias	terventions and therefore did not have outcome data. 2% of those who received treatment were lost to follow-up. However, these numbers are low and probably did not have an effect on the outcome.		aware of the assigned intervention. PFS was assessed by a blinded independent central review.				cerns due to lack of information about allocation concealment; missing study protocol and SAP.
NCT00065468	Some concerns	Par-tic-i-pants	Low risk of bias	The study was open-label: both participants and those deliver-	Low risk of bias	2.2% did not receive the intended in-terventions	Low risk of bias	Outcome assessors were not aware of	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due
Com-par-												

(Continued)

ison 2 (TEM +IFN vs. IFN) **Total trial population (only intermediate and poor risk groups included in the trial)**		were randomised, but no information provided about allocation concealment. There were no baseline imbalances that would suggest a problem with randomisation.	ing the intervention were aware of assigned interventions. Only 2 participants randomised to the combination arm and 7 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).		and therefore did not have outcome data. 1.7% of those who received treatment were lost to follow-up. However, these numbers are low and probably did not have an effect on the outcome.		the assigned intervention. PFS was assessed by a blinded independent central review.				to lack of information about allocation concealment; missing study protocol and SAP.
NCT00138333 **Total tri-**	Low risk of bias	In- Low risk of ter-bias active voice	The study was double-blind: both participants and those delivering the in-	Low risk of bias	1.2% did not receive the intended interventions and therefore	Low risk of bias	Outcome assessors were not aware of the as-	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to missing

(Continued)

al pop-u-la-tion (com-bined risk groups)		recog-ni-tion sys-tem was used for ran-domi-sa-tion. There were no base-line im-bal-ances that would sug-gest a prob-lem with ran-domi-sa-tion.	tervention were not aware of as-signed interven-tions. Only 6 par-ticipants ran-domised to the experimental arm and 2 par-ticipants ran-domised to the control arm did not receive any treatment. The method of analy-sis was appropri-ate (ITT).		did not have outcome da-ta. 3.9% of those who re-ceived treat-ment with-drew consent or were lost to follow-up before inter-im data cut. However, these num-bers are low and probably did not have an effect on the outcome.		signed in-terven-tion.				study pro-tocol and SAP.
NCT00138723 **MSKCC favourable risk group**	Low risk of bias	In-ter-bias ac-tive voice recog-ni-tion sys-tem was used for	Low risk of bias	The study was double-blind: both participants and those de-livering the in-tervention were not aware of as-signed interven-tions. Only 6 par-ticipants ran-domised to the experimental arm and 2 par-	Low risk of bias	1.2% did not receive the intended in-terventions and therefore did not have outcome da-ta. 3.9% of those who re-ceived treat-ment with-drew consent or were lost	Low risk of bias	Outcome assessors were not aware of the as-signed in-terven-tion.	Some con-cerns	No study protocol or SAP available.	Some con-cerns	Overall judged some con-cerns due to missing study pro-tocol and SAP.

(Continued)

randomisation. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was not available for 28 participants randomised to the experimental	ticipants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT). Participants without MSKCC risk group allocation were excluded from subgroup analyses.	to follow-up before interim data cut. Unclear to which risk group they were assigned to. However, these numbers are low and probably did not have an effect on the outcome.

(Continued)

		arm and 24 participants randomised to the control arm.									
NCT00138723 **MSKCC intermediate risk group**	Low risk of bias	Inter- active voice recognition system was used for randomisation. There were no baseline imbalances that would suggest a prob-	Low risk of bias	The study was double-blind: both participants and those delivering the intervention were not aware of assigned interventions. Only 6 participants randomised to the experimental arm and 2 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT). Participants without MSKCC risk group allocation were excluded from subgroup analyses.	Low risk of bias	1.2% did not receive the intended interventions and therefore did not have outcome data. 3.9% of those who received treatment withdrew consent or were lost to follow-up before interim data cut. Unclear to which risk group they were assigned to. However, these numbers are low and probably did not have an effect on the outcome.	Low risk of bias	Outcome assessors were not aware of the assigned intervention.	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to missing study protocol and SAP.

(Continued)

		lem with ran-domi-sa-tion. MSKCC risk group was not avail-able for 28 par-tic-i-pants ran-domised to the ex-per-i-men-tal arm and 24 par-tic-i-pants ran-domised to the con-trol arm.									
NCT00138736	Low risk of bias	In- ter-bias ac-	Low risk of bias	The study was double-blind: both participants	Low risk	1.2% did not receive the intended in-	Low risk of bias	Outcome assessors were not	Some con-cerns	No study protocol or SAP available.	Some con-cerns	Overall judged some con-

(Continued)

MSKCC poor risk group	tive voice recognition system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was not available for 28 participants	and those delivering the intervention were not aware of assigned interventions. Only 6 participants randomised to the experimental arm and 2 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT). Participants without MSKCC risk group allocation were excluded from subgroup analyses.	of bias	terventions and therefore did not have outcome data. 3.9% of those who received treatment withdrew consent or were lost to follow-up before interim data cut. Unclear to which risk group they were assigned to. However, these numbers are low and probably did not have an effect on the outcome.	aware of the assigned intervention.	cerns due to missing study protocol and SAP.

(Continued)

		pants randomised to the experimental arm and 24 participants randomised to the control arm.									
NCT00032396	Some concerns	Par-tic-i-pants were ran-domised via a strat-i-fied ran-dom block de-sign, but no in-for-	Low risk of bias	The study was open-label: both participants and those deliver-ing the interven-tion were aware of assigned in-terventions. On-ly 3 participants randomised to the experimen-tal arm and 13 participants ran-domised to the control arm did not receive any treatment. The method of analy-sis was appropri-ate (ITT).	Low risk of bias	2.2% did not receive the intended in-terventions and therefore did not have outcome da-ta. Less than 1% of those who received the interven-tion were lost to follow-up. These num-bers are low and probably did not have an effect on the outcome.	High risk of bias	Outcome asses-sors were aware of the as-signed in-terven-tion and knowl-edge of in-tervention received could have affected outcome measure-ment.	Some con-cerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of informa-tion about the alloca-tion con-cealment; the as-sessors' aware-ness of as-signed in-terven-tion; miss-ing study protocol and SAP.
To-tal tri-al pop-u-la-tion (com-bined risk groups)											

(Continued)

			mation provided about the allocation concealment. There were no baseline imbalances that would suggest a problem with randomisation.								
NCT00018161 (**only favourable and intermediate**	Low risk of bias	Inter-active voice response service was used	High risk of bias	The study was double-blind: both participants and those delivering the intervention were not aware of assigned interventions. No information provided about the	Low risk of bias	1 participant randomised to the experimental arm was lost to follow-up, which probably did not have an effect on the outcome.	High risk of bias	No precise information provided about the outcome assessors. Knowledge of intervention received	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about the method of analysis; the out-

(Continued)

risk groups included in the trial)		for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	method of analysis.				could have affected outcome measurement.				come assessors' probable awareness of the assigned interventions; missing study protocol and SAP.
NCT00113763	High risk of bias	Par- tic- i- pants were ran- domised in a 1:1 ra- tio, but no in- for- ma- tion pro- vid-	The study was open-label: both participants and those deliver- ing the interven- tion were aware of assigned in- terventions. The method of analy- sis was appropri- ate (ITT).	Low risk of bias	1.1% ran- domised to the control arm were lost to follow-up.	Low risk of bias	Outcome assessors were not aware of the as- signed in- terven- tion. PFS assessed by blinded indepen- dent radi- ological review.	Some con- cerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of informa- tion about the alloca- tion con- cealment; baseline imbal- ances; missing study pro- tocol and SAP.
To- tal tri- al pop- u- la- tion (com- bined risk groups)											

(Continued)

		ed about the allocation concealment. There were some baseline imbalances that could suggest a problem with randomisation.									
Jonasch 2010 **Total trial population (combined**	Low risk of bias	Random domisation method appropriate and allocation con-	Low risk of bias	No information provided about whether the participants or those delivering the intervention were blinded or not. Only 1 participant randomised to the experimental arm did not receive any treatment.	Low risk of bias	1 participant did not receive the intended intervention and therefore did not have outcome data. Of those who received treatment, 8.8% came off study before the first 8-week re-	High risk of bias	No information whether the investigator (outcome assessor) was blinded. We can only assume no (not blinded) because investiga-	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about blinding of outcome assessor; missing study protocol and SAP.

(Continued)

risk groups)		cealed. No imbalances.	The method of analysis was appropriate (ITT).		sponse assessment. However, these numbers are low and probably did not have an effect on the outcome.		tor assessment was compared to blinded review of 20 participants' scans. Knowledge of intervention received could have affected outcome measurement.				
	NCT00326677, NCT00930889 Some concerns	Par Some concerns i- pants were randomised in a 1:1 ratio, but no information about allocation concealment. However,	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 15 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	High risk of bias	2% did not receive the intended interventions and therefore did not have outcome data. No information whether there was loss to follow-up.	Low risk of bias	Outcome assessors were not aware of the assigned intervention. PFS was assessed by blinded independent central review.	Some concerns	No study protocol or SAP available.	High risk of bias.	Overall judged high risk of bias due to lack of information about the randomisation process and allocation concealment; deviations from intended interventions only in the control group; lack of information about missing data; missing
Total trial population (combined risk groups)											

(Continued)

		er, there were no baseline imbalances that would suggest a problem with randomisation.									study protocol or SAP.
NCT00138931	Low risk of bias	Randomisation was performed centrally. There were no baseline imbalances that would suggest a problem	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 5 participants randomised to the experimental arm and 7 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	3.3% did not receive the intended interventions and therefore, did not have outcome data. 4.5% of those who received treatment discontinued due to "other" reasons, including (but not limited to) loss to follow-up. However, these numbers are low and probably did not have an effect on the outcome.	High risk of bias	Outcome assessors were aware of the assigned intervention and knowledge of intervention received could have affected outcome measurement.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to the assessors' awareness of assigned intervention; missing study protocol and SAP.
Total trial population (combined risk groups)											

(Continued)

		lem with randomisation.									
NCT00320911 **MSKCC favourable risk group**	Low risk of bias	Randomisation was performed centrally. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was not available for 8	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 5 participants randomised to the experimental arm and 7 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT). Participants without MSKCC risk group allocation were excluded from subgroup analyses.	Low risk of bias	3.3% did not receive the intended interventions and therefore, did not have outcome data. 4.5% of those who received treatment discontinued due to "other" reasons, including (but not limited to) loss to follow-up. Unclear to which risk group they were assigned to. However, these numbers are low and probably did not have an effect on the outcome.	High risk of bias	Outcome assessors were aware of the assigned intervention and knowledge of intervention received could have affected outcome measurement.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to the assessors' awareness of assigned intervention; missing study protocol and SAP.

(Continued)

		participants.									
NCT00753911 **MSKCC intermediate risk group**	Low risk of bias	Randomisation was performed centrally. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was not available for 8 participants.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 5 participants randomised to the experimental arm and 7 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT). Participants without MSKCC risk group allocation were excluded from subgroup analyses.	Low risk of bias	3.3% did not receive the intended interventions and therefore, did not have outcome data. 4.5% of those who received treatment discontinued due to "other" reasons, including (but not limited to) loss to follow-up. Unclear to which risk group they were assigned to. However, these numbers are low and probably did not have an effect on the outcome.	High risk of bias	Outcome assessors were aware of the assigned intervention and knowledge of intervention received could have affected outcome measurement.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due the assessors' awareness of assigned intervention; missing study protocol and SAP.

(Continued)

		i-pants.									
NCT00910078 **Total trial population (combined risk groups)**	Low risk of bias	A cen-tralized reg-is-tration sys-tem was used for ran-domi-sa-tion. There were no base-line im-bal-ances that would sug-gest a prob-lem with ran-domi-sa-tion.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned inter-ventions. Only 3 participants ran-domised to the experimental arm did not receive any treatment. The method of analysis was ap-propriate (ITT).	Low risk of bias	1% did not receive the intended in-terventions and therefore did not have outcome da-ta. 2.1% of those who re-ceived treat-ment were lost to fol-low-up. How-ever, these numbers are low and prob-ably did not have an effect on the out-come.	Low risk of bias	Outcome assessors were not aware of the as-signed in-terven-tion. PFS was as-sessed by a masked indepen-dent re-view com-mittee.	Some con-cerns	No study protocol or SAP available.	Some con-cerns	Overall judged some con-cerns due to missing study pro-tocol and SAP.
NCT00910078 **MSKCC favourable**	Low risk of bias	A cen-tralized	Low risk of bias	The study was open-label: both participants and those delivering	Low risk of bias	1% did not re-ceive the in-tended inter-ventions and	Low risk of bias	Outcome assessors were not aware of	Some con-cerns	No study protocol or SAP available.	Some con-cerns	Overall judged some con-cerns due

(Continued)

risk group	registration system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was not available for 7 participants	the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm did not receive any treatment. The method of analysis was appropriate (ITT). However, 8 participants for whom MSKCC risk group was not available were still included in the analysis and allocated to the intermediate/poor risk group.	therefore did not have outcome data. 2.1% of those who received treatment were lost to follow-up. Unclear how many of these were assigned to the favourable risk group. However, these numbers are low and probably did not have an effect on the outcome.	the assigned intervention. PFS was assessed by a masked independent review committee.	to missing study protocol and SAP.	

(Continued)

		randomised to the experimental arm and 1 participant randomised to the control arm.									
NCT00920816 **MSKCC intermediate+poor risk groups combined**	Low risk of bias	A centralized registration system was used for randomisation. There were no baseline	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm did not receive any treatment. The method of analysis was appropriate (ITT). However, 8 participants for whom MSKCC risk group was not available	Low risk of bias	1% did not receive the intended interventions and therefore did not have outcome data. 2.1% of those who received treatment were lost to follow-up. Unclear how many of these were assigned to the intermediate/poor risk group. However, these numbers are low and probably	Low risk of bias	Outcome assessors were not aware of the assigned intervention. PFS was assessed by a masked independent review committee.	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to missing study protocol and SAP.

(Continued)

imbalances that would suggest a problem with randomisation. MSKCC risk group was not available for 7 participants randomised to the experimental arm and 1 participant randomised	were still included in the analysis and allocated to the intermediate/poor risk group.	did not have an effect on the outcome.

(Continued)

		to the control arm								
NCT01244971 **Total trial population (combined risk groups)**	Low risk of bias	Inter-active voice randomisation system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm did not receive any treatment. The method of analysis was appropriate.	Low risk of bias	3% did not receive the intended interventions and therefore did not have outcome data. 1 participant randomised to the control arm was lost to follow-up. However, these numbers are low and probably did not have an effect on the outcome.	High risk of bias	Outcome assessors were aware of the assigned intervention. Knowledge of intervention received could have affected outcome measurement.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to the assessors' awareness of assigned intervention; missing study protocol and SAP.

(Continued)

NCT00613171 **Total trial population (combined risk groups)**	Low risk of bias	A computerized centrally located randomisation system was used. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 7 participants randomised to the experimental arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	Less than 1% did not receive the intended interventions and therefore did not have outcome data. 5.5% of those who received treatment were lost to follow-up. However, these numbers are low and probably did not have an effect on the outcome.	Low risk of bias	Outcome assessors were not aware of the assigned intervention. An independent blinded assessment was conducted.	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to missing study protocol and SAP.
NCT00613171 **MSKCC favourable**	Low risk of bias	A computer-	Low risk of bias	The study was open-label: both participants and those delivering	Low risk of bias	Less than 1% did not receive the intended inter-	Low risk of bias	Outcome assessors were not aware of	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due

(Continued)

risk groups	ized centrally located randomisation system was used. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was available for all randomised	the intervention were aware of assigned interventions. Only 7 participants randomised to the experimental arm did not receive any treatment. The method of analysis was appropriate (ITT).	ventions and therefore did not have outcome data. 5.5% of those who received treatment were lost to follow-up. Unclear how many of these were assigned to the favourable risk group. However, these numbers are low and probably did not have an effect on the outcome.	the assigned intervention. An independent blinded assessment was conducted.	to missing study protocol and SAP.

(Continued)

		participants.									
NCT00631371; Low risk of bias **MSKCC intermediate risk groups**	Low risk of bias	A computerized centrally located randomisation system was used. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 7 participants randomised to the experimental arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	Less than 1% did not receive the intended interventions and therefore did not have outcome data. 5.5% of those who received treatment were lost to follow-up. Unclear how many of these were assigned to the intermediate risk group. However, these numbers are low and probably did not have an effect on the outcome.	Low risk of bias	Outcome assessors were not aware of the assigned intervention. An independent blinded assessment was conducted.	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to missing study protocol and SAP.

(Continued)

		risk group was available for all randomised participants.									
NCT00631371 **MSKCC poor risk groups**	Low risk of bias	A computerized centrally located randomisation system was used. There were no baseline imbalances that would suggest	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 7 participants randomised to the experimental arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	Less than 1% did not receive the intended interventions and therefore did not have outcome data. 5.5% of those who received treatment were lost to follow-up. Unclear how many of these were assigned to the poor risk group. However, these numbers are low and probably did not have an effect on the outcome.	Low risk of bias	Outcome assessors were not aware of the assigned intervention. An independent blinded assessment was conducted.	Some concerns	No study protocol or SAP available.	Some concerns	Overall judged some concerns due to missing study protocol and SAP.

(Continued)

		a problem with randomisation. MSKCC risk group was available for all randomised participants.									
NCT01835158 **Total trial population (only intermediate and poor risk groups in-**	Low risk of bias	Random-ization was performed centrally. There were no baseline imbalances that would suggest a	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 1 participant randomised to the experimental arm and 6 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	4.5% did not receive the intended interventions and therefore did not have outcome data. 16% of those who received treatment had missing radiographic images or were unevaluable for tumour response assessments, but there is evidence that the result was not biased by	Low risk of bias	Outcome assessors were not aware of the assigned intervention. PFS was assessed by a blinded independent radiology review committee.	Some concerns	Study protocol available with some statistical considerations briefly described, but no separate SAP available to fully check the pre-planned analyses.	Some concerns	Overall judged some concerns due to missing SAP.

(Continued)

cluded in the trial)		problem with randomisation.			missing outcome data.						
NCT01235715 **IMDC intermediate risk group**	Low risk of bias	Randomisation was performed centrally. There were no baseline imbalances that would suggest a problem with randomisation. IMDC risk group was available for all	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 1 participant randomised to the experimental arm and 6 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	4.5% did not receive the intended interventions and therefore did not have outcome data. 16% of those who received treatment had missing radiographic images or were unevaluable for tumour response assessments, but there is evidence that the result was not biased by missing outcome data.	Low risk of bias	Outcome assessors were not aware of the assigned intervention. PFS was assessed by a blinded independent radiology review committee.	Some concerns	Study protocol available with some statistical considerations briefly described, but no separate SAP available to fully check the pre-planned analyses.	Some concerns	Overall judged some concerns due to missing SAP.

(Continued)

		ran-domised par-tic-i-pants.									
NCT01835158 **IMDC poor risk group**	Low risk of bias	Ran-dom-isa-tion was per-formed cen-tral-ly. There were no base-line im-bal-ances that would sug-gest a prob-lem with ran-domi-sa-tion. IMDC risk group was avail-able for all ran-	Low risk of bias	The study was open-label: both participants and those deliver-ing the interven-tion were aware of assigned in-terventions. On-ly 1 participant randomised to the experimen-tal arm and 6 par-ticipants ran-domised to the control arm did not receive any treatment. The method of analy-sis was appropri-ate (ITT).	Low risk of bias	4.5% did not receive the intended in-terventions and there-fore did not have outcome data. 16% of those who re-ceived treat-ment had missing radi-ographic im-ages or were unevaluable for tumour response as-sessments, but there is evidence that the result was not biased by missing out-come data.	Low risk of bias	Outcome assessors were not aware of the as-signed in-terven-tion. PFS was as-sessed by a blinded indepen-dent radi-ology re-view com-mittee.	Some con-cerns	Study protocol available with some statistical considerations briefly described, but no separate SAP available to fully check the pre-planned analyses.	Some con-cerns	Overall judged some con-cerns due to missing SAP.

(Continued)

		domised participants.									
NCT02231749 **Total trial population (combined risk groups)**	Low risk of bias	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 11 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	1.3% did not receive the intended interventions and therefore did not have outcome data. Less than 1% of those who received treatment were lost to follow-up. However, these numbers are low and probably did not have an effect on the outcome.	High risk of bias	No information whether the independent radiological review committee was blinded. Knowledge of intervention received could have affected outcome measurement.	Low risk of bias	Study protocol and SAP available. Final revisions of both done before data cutoff (with extended follow-up). Analyses were preplanned and reported.	High risk of bias	Overall judged high risk of bias due to lack of information about the outcome assessor and blinding to outcome assessment.
NCT02231749	Low risk of bias	Low risk of bias	The study was open-label: both	Low risk	1.3% did not receive the	High risk of bias	No information	Low risk of bias	Study protocol and SAP avail-	High risk of bias	Overall judged

(Continued)

IMDC favourable risk group	active voice response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. IMDC risk group was available for all randomised participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 11 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	of bias	intended interventions and therefore did not have outcome data. Less than 1% of those who received treatment were lost to follow-up. Unclear to which risk group they were assigned to. However, these numbers are low and probably did not have an effect on the outcome	whether the independent radiological review committee. Knowledge of intervention received could have affected outcome measurement.	able. Final revisions of both done before data cutoff (with extended follow-up). Analyses were preplanned and reported.	high risk of bias due to lack of information about the outcome assessor and blinding to outcome assessment.

(Continued)

		i-pants.									
NCT02231749 **IMDC intermediate and poor risk groups combined**	Low risk of bias	In- Low risk of ter-bias active voice response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. IMDC risk group was available for	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 11 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	1.3% did not receive the intended interventions and therefore did not have outcome data. Less than 1% of those who received treatment were lost to follow-up. Unclear to which risk group they were assigned to. However, these numbers are low and probably did not have an effect on the outcome	High risk of bias	No information whether the independent radiological review committee. Knowledge of intervention received could have affected outcome measurement.	Low risk of bias	Study protocol and SAP available. Final revisions of both done before data cutoff (with extended follow-up). Analyses were preplanned and reported.	High risk of bias	Overall judged high risk of bias due to lack of information about the outcome assessor and blinding to outcome assessment.

(Continued)

		all randomised participants.									
NCT00720941 **Total trial population (combined risk groups)**	Low risk of bias	Inter-active voice response system was used for randomsization. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 5 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	Less than 1% did not receive the intended interventions and therefore did not have outcome data. 2.9% of those who received treatment were lost to follow-up. However, these numbers are small and probably did not have an effect on the outcome.	Low risk of bias	Outcome assessors were not aware of the assigned intervention. PFS was assessed by a blinded independent review committee.	Low risk of bias	CSR and study protocol with SAP available. All reported analyses were pre-specified in the protocol and SAP.	Low risk of bias	Overall judged low risk of bias.

(Continued)

NCT01184742 Comparison 1 (ATE vs. SUN) **Total trial population (combined risk groups)**	Low risk of bias	Interactive voice/web response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 1 participant randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	0.3% did not receive the intended interventions and therefore did not have outcome data. 1.5% of those who received treatment were lost to follow-up. These numbers are very low and probably did not have an effect on the outcome.	High risk of bias	No precise information provided about whether the outcome assessors were blinded. Knowledge of intervention received could have affected outcome measurement.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about blinding of outcome assessor; missing study protocol and SAP.
NCT01184742 Comparison 2	Low risk of bias	Interactive voice/web re-	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned inter-	Low risk of bias	0.3% did not receive the intended interventions and therefore did not have outcome data.	High risk of bias	No precise information provided about whether the out-	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about

(Continued)

(ATE +BEV vs. SUN) **Total trial population (combined all risk groups)**				sponse system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.		ventions. Only 1 participant randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).		3.5% of those who received treatment were lost to follow-up. These numbers are very low and probably did not have an effect on the outcome.		come assessors were blinded. Knowledge of intervention received could have affected outcome measurement.				blinding of outcome assessor; missing study protocol and SAP.
NCT02420821 **Total trial population (combined risk groups)**	Low risk of bias	Interactive voice and web response system was used for randomisation. There were	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 15 par-	Low risk of bias		2% did not receive the intended interventions and therefore	High risk of bias	The investigators were the outcome assessors and they were not blinded to treatment allocation. Knowledge of intervention received could have	Low risk of bias	Study protocol and SAP available. All reported analyses were	High risk of bias		Overall judged high risk of bias due to lack of blinding of the

(Continued)

		no baseline imbalances that would suggest a problem with randomisation.	ticipants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).		did not have outcome data. Less than 1% of those who received treatment were lost to follow-up. However, these numbers are low and probably did not have an effect on the outcome.	affected outcome measurement.		pre-specified in the protocol and SAP.		outcome assessors.
NCT02420821 **MSKCC favourable risk group**	Low risk of bias	In- ter-bias active voice and	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware	Low risk of bias	2% did not receive the intended interventions and therefore did not have	High risk of bias	No precise information provided about the outcome	Low risk of bias	Study protocol and SAP available. All reported subgroup analyses were pre-	High risk of bias	Overall judged high risk of bias due to lack of blinding

(Continued)

	web response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was available for all randomised participants.	of assigned interventions. Only 3 participants randomised to the experimental arm and 15 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	outcome data. Less than 1% of those who received treatment were lost to follow-up. Unclear to which risk groups they were assigned to. However, these numbers are low and probably did not have an effect on the outcome.	assessors. Knowledge of intervention received could have affected outcome measurement.	specified in the protocol and SAP.	of the outcome assessors.

(Continued)

		i-pants.									
NCT02420821 **MSKCC intermediate risk group**	Low risk of bias	In- ter-bias active voice and web response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was avail-	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 15 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	2% did not receive the intended interventions and therefore did not have outcome data. Less than 1% of those who received treatment were lost to follow-up. Unclear to which risk groups they were assigned to. However, these numbers are low and probably did not have an effect on the outcome.	High risk of bias	No precise information provided about the outcome assessors. Knowledge of intervention received could have affected outcome measurement.	Low risk of bias	Study protocol and SAP available. All reported subgroup analyses were prespecified in the protocol and SAP.	High risk of bias	Overall judged high risk of bias due to lack of blinding of the outcome assessors.

(Continued)

		able for all randomised participants.								
NCT02420821 **MSKCC poor risk group**	Low risk of bias	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 15 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	2% did not receive the intended interventions and therefore did not have outcome data. Less than 1% of those who received treatment were lost to follow-up. Unclear to which risk groups they were assigned to. However, these numbers are low and probably did not have an effect on the outcome.	High risk of bias	No precise information provided about the outcome assessors. Knowledge of intervention received could have affected outcome measurement.	Low risk of bias	Study protocol and SAP available. All reported subgroup analyses were prespecified in the protocol and SAP.	High risk of bias	Overall judged high risk of bias due to lack of blinding of the outcome assessors.

(Continued)

		sation. MSKCC risk group was available for all randomised participants.									
NCT02684006 **IMDC favourable risk group**	Low risk of bias	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 8 participants randomised to the experimental arm and 5 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate. Participants without IMDC risk group allocation were excluded from the analysis.	Some concerns	1.5% did not receive the intended interventions and therefore did not have outcome data. No information about study flow for the second interim analysis (which is the result considered in this review).	Low risk of bias	Outcome assessors were not aware of the assigned intervention. A blinded independent central review was conducted.	High risk of bias	Study protocol and SAP available but discrepancies were found between statements in the publications and the SAP about the pre-specification of subgroup analyses. Also the time point that produced this result was not pre-specified in the protocol (final PFS analysis was already reported).	High risk of bias	Overall judged high risk of bias due to lack of information about potential losses of follow-up; lack of information about the subgroup analyses in the study protocol and SAP.

(Continued)

	problem with randomisation. IMDC risk group was not available for 5 participants randomised to the experimental arm and 1 participant randomised to the control arm.

(Continued)

NCT02184003	Low risk of bias	In- ter-bias ac- tive voice re- sponse sys- tem was used for ran- domi- sa- tion. There were no base- line im- bal- ances that would sug- gest a prob- lem with ran- domi- sa- tion. IMDC risk group was not avail- able for 5	The study was open-label: both participants and those deliver- ing the interven- tion were aware of assigned in- terventions. On- ly 8 participants randomised to the experimen- tal arm and 5 par- ticipants ran- domised to the control arm did not receive any treatment. The method of analysis was ap- propriate. Partic- ipants without IMDC risk group allocation were excluded from subgroup analy- ses.	Some con- cerns	1.5% did not receive the intended in- terventions and therefore did not have outcome da- ta. No infor- mation about study flow for the second in- terim analysis (which is the result consid- ered in this re- view).	Low risk of bias	Outcome assessors were not aware of the as- signed in- tervention. A blinded indepen- dent cen- tral review was con- ducted.	High risk of bias	Study protocol and SAP available but discrepan- cies were found between state- ments in the pub- lications and the SAP about the pre-specifica- tion of subgroup analyses. Also the time point that produced this re- sult was not pre- specified in the protocol (final PFS analysis was already report- ed).	High risk of bias	Overall judged high risk of bias due to lack of informa- tion about potential losses of follow-up; lack of in- formation about the subgroup analy- ses in the study pro- tocol and SAP.
IMDC in- ter- me- di- ate risk group											

(Continued)

		participants randomised to the experimental arm and 1 participant randomised to the control arm.									
NCT02181075 **IMDC poor risk group**	Low risk of bias	Low risk of bias	In-ter-active voice re-sponse sys-tem was used for ran-domi-sa-tion. There were	The study was open-label: both participants and those deliver-ing the interven-tion were aware of assigned in-terventions. On-ly 8 participants randomised to the experimen-tal arm and 5 par-ticipants ran-domised to the control arm did not receive any treatment.	Some con-cerns	1.5% did not receive the intended in-terventions and therefore did not have outcome da-ta. No infor-mation about study flow for the second in-terim analysis (which is the result consid-ered in this re-view).	Low risk of bias	Outcome assessors were not aware of the as-signed in-terven-tion. A blinded indepen-dent cen-tral review was con-ducted.	High risk of bias	Study protocol and SAP available but discrepan-cies were found between state-ments in the pub-lications and the SAP about the pre-specifica-tion of subgroup analyses. Also the time point that produced this re-sult was not pre-specified in the protocol (final PFS analysis was	High risk of bias	Overall judged high risk of bias due to lack of informa-tion about potential losses of follow-up; lack of in-formation about the subgroup analy-ses in the study pro-

(Continued)

			already reported).	tocol and SAP.
		no baseline imbalances that would suggest a problem with randomisation. IMDC risk group was not available for 5 participants randomised to the experimental arm and 1 participant	The method of analysis was appropriate. Participants without IMDC risk group allocation were excluded from subgroup analyses.	

(Continued)

		pant randomised to the control arm.									
NCT02815332 **Total trial population (combined risk groups)**	Low risk of bias	Inter-active voice response system or integrated web response system was used for randomisation. There were no baseline imbalances that would sug-	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 4 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	Low risk of bias	Less than 1% did not receive the intended interventions and therefore did not have outcome data. No indication of loss to follow-up. However, these numbers are low and probably did not have an effect on the outcome.	Low risk of bias	Outcome assessors were not aware of the assigned intervention. A blinded independent central review was conducted.	High risk of bias	Final PFS analysis was already reported in a previous publication. The timing of analysis (which is the time point for the final OS analysis) for this numerical result of PFS was not pre-specified in the protocol.	High risk of bias	Overall judged high risk of bias due to the analysis time point not being pre-specified.

(Continued)

		gest a prob-lem with ran-domi-sa-tion.									
NCT00338364 **To-tal tri-al pop-u-la-tion (com-bined risk groups)**	Some con-cerns	Par-tic-i-pants were ran-domised in a 1:1 ra-tio, but no in-for-ma-tion pro-vid-ed about who con-duct-ed the ran-domi-sa-tion and whether the al-lo-	Low risk of bias	The study was open-label: both participants and those deliver-ing the interven-tion were aware of assigned in-terventions. On-ly 1 participant randomised to the experimen-tal arm and 2 par-ticipants ran-domised to the control arm did not receive any treatment. The method of analy-sis was appropri-ate.	Low risk of bias	Less than 1% did not re-ceive the in-tended in-terventions and therefore did not have outcome da-ta. Less than 1% of those who received treatment were lost to follow-up. However, these num-bers are low and probably did not have an effect on the outcome.	High risk of bias	No infor-mation provid-ed about whether the out-come asses-sors were blinded. Knowl-edge of in-tervention received could have affected outcome measure-ment.	Some con-cerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of informa-tion about the ran-domisa-tion and allocation conceal-ment; the outcome assessors' probable awareness of the as-signed in-terven-tions; missing study pro-tocol and SAP.

(Continued)

		ca-tion was con-cealed There were no base-line im-bal-ances that would sug-gest a prob-lem with ran-domi-sa-tion.									
NCT00458068 **To-tal tri-al pop-u-la-tion (on-ly favourable and in-ter-me-di-ate risk**	Some con-cerns	Par Low risk of tic- bias i-pants were ran-domised in a 1:1 ra-tio, but no in-for-ma-tion pro-vid-	The study was open-label: both participants and those delivering the intervention were aware of as-signed interven-tions. Only 8 participants did not receive any treatment. The method of analysis was ap-propriate (ITT).	High risk of bias	1.5% did not receive the assigned in-terventions and there-fore did not have outcome data. No in-formation whether there was loss to follow-up.	High risk of bias	No pre-cise infor-mation provided about the outcome assessors (includ-ing lack of informa-tion about blinding). Knowl-edge of in-tervention received could have affected outcome	Some con-cerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of informa-tion about randomi-sation process and al-location conceal-ment; missing outcome data; the outcome assessors' probable

(Continued)

groups included in the trial)	ed about who conducted ed randomisation and whether allocation was concealed. There were no baseline imbalances that would suggest a problem with randomisation.						measurement.				awareness of the assigned intervention; missing study protocol and SAP.
NCT00943098 **MSKCC favourable risk group**	Some concerns	Par Low risk of tic- bias i- pants were ran-	The study was open-label: both participants and those delivering the intervention were aware of as-	High risk of bias	1.5% did not receive the assigned interventions and therefore did not	High risk of bias	No precise information provided about the outcome	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of informa-

(Continued)

domised in a 1:1 ratio, but no information provided about who conducted randomisation and whether allocation was concealed. There were no baseline imbalances that would suggest a	signed interventions. Only 8 participants did not receive any treatment. The method of analysis was appropriate (ITT).	have outcome data. No information whether there was loss to follow-up.	assessors (including lack of information about blinding). Knowledge of intervention received could have affected outcome measurement.	tion about randomisation process and allocation concealment; missing outcome data; the outcome assessors' probable awareness of the assigned intervention; missing study protocol and SAP.

(Continued)

		problem with randomisation. MSKCC risk group was available for all randomised participants.									
NCT00458289 **MSKCC intermediate risk group**	Some concerns	Par Low risk of tic- bias i- pants were randomised in a 1:1 ratio, but no information provided about who con-	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 8 participants did not receive any treatment. The method of analysis was appropriate (ITT).	High risk of bias	1.5% did not receive the assigned interventions and therefore did not have outcome data. No information whether there was loss to follow-up.	High risk of bias	No precise information provided about the outcome assessors (including lack of information about blinding). Knowledge of intervention received could have affected outcome measurement.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about randomisation process and allocation concealment; missing outcome data; the outcome assessors' probable awareness of the assigned interven-

(Continued)

	tion; missing study protocol and SAP.
	ducted randomisation and whether allocation was concealed. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was available for all randomised par-

(Continued)

		tic-i-pants.									
NCT00430989 **IMDC favourable risk group**	Some concerns	Par-tic-i-pants were ran-domised in a 1:1 ra-tio, but no in-for-ma-tion pro-vid-ed about al-lo-ca-tion con-ceal-ment. There were no base-line im-bal-ances that would sug-gest a	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of as-signed interven-tions. Only 8 partici-pants did not re-ceive any treat-ment. The method of analysis was ap-propriate (ITT).	High risk of bias	1.5% did not receive the assigned in-terventions and there-fore did not have outcome data. No in-formation whether there was loss to follow-up.	High risk of bias	No pre-cise infor-mation provided about the outcome assessors (includ-ing lack of informa-tion about blinding). Knowl-edge of in-tervention received could have affected outcome measure-ment.	Some con-cerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of informa-tion about randomi-sation process and al-location conceal-ment; missing outcome data; the outcome assessors' probable awareness of the as-signed in-terven-tion; miss-ing study protocol and SAP.

(Continued)

		problem with randomisation. HENG risk group was available for all randomised participants.									
NCT00458289 **IMDC intermediate risk group**	Some concerns	Par Low risk of tic- bias i- pants were randomised in a 1:1 ratio, but no information provided about allo-	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 8 participants did not receive any treatment. The method of analysis was appropriate (ITT).	High risk of bias	1.5% did not receive the assigned interventions and therefore did not have outcome data. No information whether there was loss to follow-up.	High risk of bias	No precise information provided about the outcome assessors (including lack of information about blinding). Knowledge of intervention received could have affected outcome measurement.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about randomisation process and allocation concealment; missing outcome data; the outcome assessors' probable awareness of the assigned interven-

(Continued)

		ca-tion con-ceal-ment. There were no base-line im-bal-ances that would sug-gest a prob-lem with ran-domi-sa-tion. HENG risk group was avail-able for all ran-domised par-tic-i-pants.									tion; miss-ing study protocol and SAP.
NCT00458809 **IMDC poor risk group**	Some con-cerns	Par-tic-i-pants were ran-	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of as-	High risk of bias	1.5% did not receive the assigned in-terventions and there-fore did not	High risk of bias	No pre-cise infor-mation provided about the outcome	Some con-cerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of informa-

(Continued)

	domised in a 1:1 ratio, but no information provided about allocation concealment. There were no baseline imbalances that would suggest a problem with randomisation. HENG risk group	signed interventions. Only 8 participants did not receive any treatment. The method of analysis was appropriate (ITT).	have outcome data. No information whether there was loss to follow-up.	assessors (including lack of information about blinding). Knowledge of intervention received could have affected outcome measurement.	tion about randomisation process and allocation concealment; missing outcome data; the outcome assessors' probable awareness of the assigned intervention; missing study protocol and SAP.

(Continued)

		was available for all randomised participants.									
NCT00930066 **Total trial population (combined risk groups)**	High risk of bias	Participants were randomised in a 1:1 ratio, but no information provided about who conducted randomisation and whether allo-	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. No information whether there were deviations form intended interventions and no information provided about the method of analysis.	High risk of bias	No information about loss to follow-up.	High risk of bias	No precise information provided about who assessed the outcome and whether they were blinded. Knowledge of intervention received could have affected outcome measurement.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about the randomisation, the allocation concealment, the deviations from intended interventions, the method of analysis and missing outcome data; the outcome assessors' probable awareness of the assigned interventions; missing study pro-

(Continued)

		cation was concealed. There were baseline imbalances that could suggest a problem with randomisation. Small study population.										tocol and SAP.
NCT00910377 **Total trial population (combined risk groups)**	Low risk of bias	Inter-active voice response system was used for randomisa-	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 2 participants randomised to the control arm did not receive any treatment. The method of analy-	Low risk of bias	Less than 1% did not receive the intended interventions and therefore did not have outcome data. 1 participant randomised to the control arm was lost to follow-up. However, these num-	High risk of bias	Outcome assessors were aware of the assigned intervention. Knowledge of intervention received could have affected outcome	Some concerns	Study protocol available with some statistical methods described, but no separate SAP available to fully check the preplanned analyses.	High risk of bias	Overall judged high risk of bias due to the outcome assessors' awareness of the assigned intervention; missing SAP.

(Continued)

		tion. There were no baseline imbalances that would suggest a problem with randomisation.	sis was appropriate (ITT).		bers are low and probably did not have an effect on the outcome.		measurement.				
NCT00910715 **MSKCC favourable risk group**	Low risk of bias	Inter-active voice response system was used for randomisation. There were no baseline imbalances	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 2 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT). Participants without MSKCC risk group allocation were excluded from subgroup analyses.	Low risk of bias	Less than 1% did not receive the intended interventions and therefore did not have outcome data. 1 participant randomised to the control arm was lost to follow-up. Unclear to which risk group they were assigned to. However, these numbers are low and probably did not have an effect on the outcome.	High risk of bias	Outcome assessors were aware of the assigned intervention. Knowledge of intervention received could have affected outcome measurement.	Some concerns	Study protocol available with some statistical methods described, but no separate SAP available to fully check the preplanned analyses..	High risk of bias	Overall judged high risk of bias due to the outcome assessors' awareness of the assigned intervention; missing SAP.

(Continued)

		that would suggest a problem with randomisation. MSKCC risk group was not available for 2 randomised participants.									
NCT00910377 **MSKCC intermediate risk group**	Low risk of bias	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 2 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT). Participants without MSKCC risk group	Low risk of bias	Less than 1% did not receive the intended interventions and therefore did not have outcome data. 1 participant randomised to the control arm was lost to follow-up. Unclear to which risk groups they were assigned to. However,	High risk of bias	Outcome assessors were aware of the assigned intervention. Knowledge of intervention received could have affected outcome measurement.	Some concerns	Study protocol available with some statistical methods described, but no separate SAP available to fully check the preplanned analyses.	High risk of bias	Overall judged high risk of bias due to the outcome assessors' awareness of the assigned intervention; missing SAP.

(Continued)

		no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was not available for 2 randomised participants.	allocation were excluded from subgroup analyses.		these numbers are low and probably did not have an effect on the outcome.						
NCT00910377 **MSKCC poor risk group**	Low risk of bias	Inter-active voice response system was used for	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 2 participants randomised to the control arm did not receive any	Low risk of bias	Less than 1% did not receive the intended interventions and therefore did not have outcome data. 1 participant randomised to the control arm was lost	High risk of bias	Outcome assessors were aware of the assigned intervention. Knowledge of intervention received	Some concerns	Study protocol available with some statistical methods described, but no separate SAP available to fully check the preplanned analyses..	High risk of bias	Overall judged high risk of bias due to the outcome assessors' awareness of the assigned interven-

(Continued)

		ran-domi-sa-tion. There were no base-line im-bal-ances that would sug-gest a prob-lem with ran-domi-sa-tion. MSKCC risk group was not avail-able for 2 ran-domised par-tic-i-pants.	treatment. The method of analy-sis was appropri-ate (ITT). Partic-ipants without MSKCC risk group allocation were excluded from subgroup analy-ses.		to follow-up. Unclear to which risk groups they were assigned to. However, these num-bers are low and probably did not have an effect on the outcome.		could have affected outcome measure-ment.				tion; miss-ing SAP.
NCT01483573	Low risk of bias	Par-tic-i-pants were ran-	The study was open-label: both participants and those delivering the intervention were aware of	High risk of bias	3.2% did not receive the in-tended inter-ventions and therefore did not have out-	High risk of bias	No infor-mation provided about who assessed the out-	Some con-cerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to high number
To-tal tri-al	Low risk of bias	Low risk of bias		High risk of bias		High risk of bias		Some con-cerns		High risk of bias	

(Continued)

population (only favourable and intermediate risk groups included)	domised in a 1:1 ratio. The assignment was obtained at enrolment by the investigator via the Internet. There were no baseline imbalances that would suggest a problem with	assigned interventions. Only 3 participants randomised to the experimental arm and 1 participant randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	come data. 22% of those who received treatment did not have outcome data.	come and whether they were blinded. Knowledge of intervention received could have affected outcome measurement.	of participants with missing outcome data; the outcome assessors' probable awareness of the assigned intervention; missing study protocol and SAP.	

(Continued)

		randomisation.								
NCT01483781 **MSKCC favourable risk group**	Low risk of bias	Participants were randomised in a 1:1 ratio. The assignment was obtained at enrolment by the investigator via the Internet There were no baseline	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 1 participant randomised to the control arm did not receive any treatment. The method of analysis was appropriate (ITT).	High risk of bias	3.2% did not receive the intended interventions and therefore did not have outcome data. 22% of those who received treatment did not have outcome data.	High risk of bias	No information provided about who assessed the outcome and whether they were blinded. Knowledge of intervention received could have affected outcome measurement.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to high number of participants with missing outcome data; the outcome assessors' probable awareness of the assigned intervention; missing study protocol and SAP.

(Continued)

		imbalances that would suggest a problem with randomisation. MSKCC risk group was available for all randomised participants.									
NCT02210117 Comparison 1 (CAB vs. SUN) **Total trial pop-**	Low risk of bias	The randomisation was done by the Study Center. There were no baseline imbalances that would suggest a problem with randomisation. MSKCC risk group was available for all randomised participants.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 2 participants randomised to the experimental arm and 2 participants randomised to the	Low risk of bias	4.3% did not receive the intended interventions and therefore did not have outcome data. 2.2% had no protocol treatment. Only 1 participant was lost to follow-up. However, these	High risk of bias	No precise information provided about method of measuring PFS. Outcome assessors were aware of the assigned inter-	Some concerns	Study protocol available, but no original SAP.	High risk of bias	Overall judged high risk of bias due to lack of information about method of outcome measurement; the outcome assessors' awareness of the assigned in-

(Continued)

ulation (combined risk groups)	line imbalances that would suggest a problem with randomisation.	control arm did not receive any treatment. The method of analysis was appropriate (ITT).	numbers are low and probably did not have an effect on the outcome.	vention. Knowledge of intervention received could have affected outcome measurement.	tervention; missing SAP.

Appendix 13. Risk of bias assessment for the outcome adverse events

Trial	Risk of bias											
	Randomisation process		Deviations from intended interventions		Missing outcome data		Measurement of the outcome		Selection of the reported results		Overall	
	Authors' judgement	Support for judgement	Authors' judgement	Support for judgement	Authors' judgement	Support for judgement	Authors' judgement	Support for judgement	Authors' judgement	Support for judgement	Authors' judgement	Support for judgement
NCT03141177 **Total trial population** **(combined all risk groups)**	Low risk of bias	Interactive Response Technology was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 8 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate.	Low risk of bias	1.7% did not receive the intended interventions and therefore did not have outcome data. However, these numbers are low and probably did not have an effect on the outcome.	High risk of bias	Adverse events (AEs) were reported by the participants and the investigator was responsible for detecting, documenting and reporting events. Both were aware of the assigned intervention. Knowledge of intervention received could have affected outcome measurement.	Low risk of bias	A study protocol with SAP available. Safety analysis was prespecified in the protocol and SAP.	High risk of bias	Overall judged high risk of bias due to the outcome assessors' awareness of assigned intervention.
NCT02811861 Comparison 1 (LEN	Low risk of bias	Interactive voice and web response	High risk of bias	The study was open-label: both participants and those delivering the intervention	Low risk of bias	2.8% did not receive the intended inter-	High risk of bias	AEs were most likely reported by the participants and	Low risk of bias	A study protocol with SAP available.	High risk of bias	Overall judged high risk of bias due to inap-

(Continued)

+PEM vs. SUN) **Total trial population** **(all combined risk groups)**		system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.		were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 17 participants randomised to the control arm did not receive any treatment. The method of analysis was not appropriate (as-treated).		ventions and therefore did not have outcome data. However, these numbers are low and probably did not have an effect on the outcome.		assessed by the investigator. Both were aware of the assigned intervention. Knowledge of intervention received could have affected outcome measurement.		Safety analysis was prespecified in the protocol and SAP.		propriate method of analysis and the outcome assessors' awareness of assigned intervention.
NCT02811861 Comparison 2 (LEN +EVE vs. SUN) **Total trial population** **(combined risk groups)**	Low risk of bias	Interactive voice and web response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 2 participants randomised to the experimental arm and 17 participants randomised to the control arm did not receive any treatment. The method of analysis was not appropriate (as-treated).	Low risk of bias	2.7% did not receive the intended interventions and therefore did not have outcome data. However, these numbers are low and probably did not have an effect on the outcome.	High risk of bias	AEs were most likely reported by the participants and assessed by the investigator. Both were aware of the assigned intervention. Knowledge of intervention received could have affected outcome measurement.	Low risk of bias	A study protocol with SAP available. Safety analysis was prespecified in the protocol and SAP.	High risk of bias	Overall judged high risk of bias due to inappropriate method of analysis and the outcome assessors' awareness of assigned intervention.

(Continued)

NCT00065468 Comparison 1 (TEM vs. IFN) **Total trial population** **(only intermediate and poor risk groups included in the trial)**	Some concerns	Participants were randomised, but no information provided about the allocation concealment. There were no baseline imbalances that would suggest a problem with randomisation.	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 1 participant randomised to the single-drug arm and 7 participants randomised to the control arm did not receive any treatment. No precise information provided about the method of analysis (as-treated is indicated).	Low risk of bias	1.9% did not receive the intended interventions and therefore did not have outcome data. However, these numbers are low and probably did not have an effect on the outcome.	High risk of bias	No precise information provided about method of measuring AEs. However, AEs were defined according to the National Cancer Institute Common Toxicity Criteria (NCICTC), version 3. Measurement of AEs could have differed between intervention groups due to differences in number of visits to the healthcare provider. AEs were assessed by the participants who were aware of the assigned intervention. Knowledge of intervention received could have affected outcome	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about allocation concealment, method of analysis and method of outcome measurement; probable differences in outcome measurement between intervention arms; the outcome assessors' awareness of assigned intervention; missing study protocol and SAP.

(Continued)

								measurement.				
NCT00065468 **comparison 2** Comparison 2 (IFN +TEM vs. IFN) **Total trial population** **(only intermediate and poor risk groups included in the trial)**	Some concerns	Participants were randomised, but no information provided about the allocation concealment. There were no baseline imbalances that would suggest a problem with randomisation.	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 2 participants randomised to the combination arm and 7 participants randomised to the control arm did not receive any treatment. No precise information provided about the method of analysis (as-treated is indicated).	Low risk of bias	2.2% did not receive the intended interventions and therefore did not have outcome data. However, these numbers are low and probably did not have an effect on the outcome.	High risk of bias	No precise information provided about method of measuring AEs. However, AEs were defined according to NCICTC, version 3. AEs were assessed by the participants who were aware of the assigned intervention. Knowledge of intervention received could have affected outcome measurement.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about allocation concealment, method of analysis and method of outcome measurement; the outcome assessors' awareness of assigned intervention; missing study protocol and SAP.
NCT00081614 **Total trial population** **(only favourable and intermediate risk**	Low risk of bias	Interactive voice response service was used for randomisation. There were no baseline	High risk of bias	The study was double-blind: both participants and those delivering the intervention were not aware of assigned interventions. No information provided about the method of analysis.	Low risk of bias	Only 3 randomised did not have outcome data and 1 from the control group was lost	Some concerns	No precise information provided about method of measuring AEs. However, AEs were defined according to NCICTC, version 3. Outcome asses-	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk due to lack of information about method of analysis and method of outcome measure-

(Continued)

groups included in the trial)		imbalances that would suggest a problem with randomisation.				to follow-up.		sors were not aware of the assigned intervention.				ment; missing study protocol and SAP.
NCT01024920 **Total trial population** **(combined risk groups)**	Low risk of bias	Interactive voice randomisation system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm did not receive any treatment. No precise information provided about the method of analysis.	Low risk of bias	3% did not receive the intended interventions and therefore did not have outcome data. However, these numbers are low and probably did not have an effect on the outcome.	High risk of bias	No precise information provided about method of measuring AEs. However, AEs were defined according to NCICTC, version 3. Outcome assessors were aware of the assigned intervention. Knowledge of intervention received could have affected outcome measurement.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about method of analysis and method of outcome measurement; the outcome assessors' awareness of assigned intervention; missing study protocol and SAP.
NCT01835158 **Total trial population**	Low risk of bias	Randomisation was performed centrally. There were no	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 1	Low risk of bias	4.5% did not receive the intended interventions and therefore	High risk of bias	No precise information provided about method of measuring AEs. However, AEs were	Some concerns	Study protocol available with some statistical consider-	High risk of bias	Overall judged high risk of bias due to lack of information about method

(Continued)

(only intermediate and poor risk groups included in the trial)		baseline imbalances that would suggest a problem with randomisation.		participant randomised to the experimental arm and 6 participants randomised to the control arm did not receive any treatment. No precise information provided about the method of analysis.		fore did not have outcome data. However, these numbers are low and probably did not have an effect on the outcome.		defined according to NCICTC, version 4. AEs were assessed by the participants and the investigator. Both were aware of the assigned intervention. Knowledge of intervention received could have affected outcome measurement.		ations briefly described, but no separate SAP available to fully check the preplanned analyses.		of analysis and method of outcome measurement; the outcome assessors' awareness of assigned intervention; missing SAP.
NCT01984242 **Comparison 1 (ATE vs. SUN)** **Total trial population** **(combined risk groups)**	Low risk of bias	Interactive voice/web response system was used for randomisation. There were no baseline imbalances that would suggest a problem with ran-	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 1 participant randomised to the control arm did not receive any treatment. No precise information provided about the method of analysis, but data for crossed-over participants were reported separately. We assume participants were analyzed as ran-	Low risk of bias	0.3% did not receive the intended interventions and therefore did not have outcome data. However, these numbers are very low and probably did not have an effect on the outcome.	High risk of bias	No information provided about method of measuring AEs. Outcome assessors were aware of the assigned intervention. Knowledge of intervention received could have affected outcome measurement.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about method of outcome measurement; the outcome assessors' awareness of assigned intervention; missing study protocol and SAP.

(Continued)

		domisation.		domised in period 1.								
NCT01984242 Comparison 2 (ATE +BEV vs. SUN) **Total trial population (combined risk groups)**	Low risk of bias	Interactive voice/web response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 1 participant randomised to the control arm did not receive any treatment. No precise information provided about the method of analysis, but data for crossed-over participants were reported separately. We assume participants were analyzed as randomised in period 1.	Low risk of bias	0.3% did not receive the intended interventions and therefore did not have outcome data. However, these numbers are very low and probably did not have an effect on the outcome.	High risk of bias	No information provided about method of measuring AEs. Outcome assessors were aware of the assigned intervention. Knowledge of intervention received could have affected outcome measurement.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about method of outcome measurement; the outcome assessors' awareness of assigned intervention; missing study protocol and SAP.
NCT02684006 **Total trial population** **(combined risk groups)**	Low risk of bias	Interactive voice response system was used for randomisation. There were no baseline imbalances that would	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 8 participants randomised to the experimental arm and 5 participants randomised to the control arm did not receive any treatment. The method	Low risk of bias	1.5% did not receive the intended interventions and therefore did not have outcome data. However, these numbers are low and	High risk of bias	No precise information provided about method of measuring AEs. However, AEs were defined according to NCICTC, version 4.03. Outcome assessors were aware of the assigned in-	Some concerns	Study protocol and SAP available. However, the exact time point of outcome measurement is unclear.	High risk of bias	Overall judged high risk of bias due to lack of information about method and time point of outcome measurement; the outcome assessors' awareness of as-

(Continued)

		sug-gest a problem with ran-domisa-tion.		of analysis was ap-propriate.		probably did not have an effect on the out-come.		tervention. Knowledge of inter-vention re-ceived could have affect-ed outcome measure-ment.				signed in-terven-tion.
NCT00719264 **Total trial popula-tion (com-bined risk groups)**	Some concerns	Partic-ipants were ran-domised in a 1:1 ratio, but no informa-tion pro-vided about who con-ducted the ran-domisa-tion and whether the allo-cation was con-cealed. There were no baseline imbal-ances that would sug-gest a problem with ran-	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned inter-ventions. Only 1 participant ran-domised to the ex-perimental arm and 2 participants randomised to the control arm did not receive any treat-ment; 1 participant randomised to the experimental arm had no post base-line safety assess-ment. No precise information pro-vided about the method of analysis.	Low risk of bias	1.1% did not re-ceive the intend-ed inter-ventions and there-fore did not have outcome data. Howev-er, these num-bers are low and probably did not have an effect on the out-come.	High risk of bias	No precise informa-tion provid-ed about method of measuring AEs. Howev-er, AEs were defined ac-cording to NCICTC, ver-sion 3. Out-come asses-sors were aware of the assigned in-tervention. Knowledge of inter-vention re-ceived could have affect-ed outcome measure-ment.	Some concerns	No study protocol or SAP avail-able.	High risk of bias	Overall judged high risk of bias due to lack of informa-tion about randomiz sation process, alloca-tion con-cealment, method of analy-sis and method of outcome measure-ment; the outcome assessors' aware-ness of as-signed in-tervention; miss-ing study protocol and SAP.

(Continued)

		domisation.										
NCT00720941 **Total trial population (combined risk groups)**	Low risk of bias	Interactive voice response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	Some concerns	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 5 participants randomised to the control arm did not receive any treatment. The method of analysis was not appropriate (as-treated), but this probably did not have an effect on the outcome as there is evidence that participants actually received the assigned intervention.	Low risk of bias	Less than 1% did not receive the intended interventions and therefore did not have outcome data. However, these numbers are small and probably did not have an effect on the outcome.	High risk of bias	AEs were most likely reported by the participants. The investigator and site staff were responsible for detecting, documenting and reporting events. All were aware of the assigned intervention. Knowledge of intervention received could have affected outcome measurement.	Low risk of bias	CSR and study protocol with SAP available. Safety analysis was prespecified in the protocol and SAP.	High risk of bias	Overall judged high risk of bias due to inappropriate method of analysis and the outcome assessors' awareness of assigned intervention.
NCT00732914 **Total trial population (combined risk groups)**	Low risk of bias	Randomisation was performed centrally. There were no baseline imbalances that would sug-	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 5 participants randomised to the experimental arm and 7 participants randomised to the control arm did not	Low risk of bias	3.3% did not receive the intended interventions and therefore did not have outcome data. However, these	High risk of bias	No precise information provided about method of measuring AEs. However, AEs were defined according to NCICTC, version 3. Outcome assessors were	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about the method of outcome measurement; outcome assessors'

(Continued)

		gest a problem with randomisation.		receive any treatment. No precise information provided about the method of analysis, but data for first period reported separately. We assume participants in first period received their allocated intervention.		numbers are low and probably did not have an effect on the outcome.		aware of the assigned intervention. Knowledge of intervention received could have affected outcome measurement.				awareness of assigned intervention; missing study protocol and SAP.
NCT01613846 **Total trial population** **(combined risk groups)**	Some concerns	Participants were randomised in a 1:1 ratio, but no information provided about who conducted the randomisation and whether the allocation was concealed. There were no baseline imbalances that would suggest a problem	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 6 participants randomised to the experimental arm and 5 participants randomised to the control arm did not receive any treatment. The method of analysis was appropriate.	Low risk of bias	2.9% did not receive the intended interventions and therefore did not have outcome data. However, these numbers are low and probably did not have an effect on the outcome.	High risk of bias	No precise information provided about method of measuring AEs. However, AEs were defined according to NCICTC, version 4.03. Outcome assessors were aware of the assigned intervention. Knowledge of intervention received could have affected outcome measurement.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about randomistion process, allocation concealment and method of outcome measurement; the outcome assessors' awareness of assigned intervention; missing study protocol and SAP.

(Continued)

		with randomisation.										
NCT02420821 **Total trial population (combined risk groups)**	Low risk of bias	Interactive voice and web response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm and 15 participants randomised to the control arm did not receive any treatment. Conflicting information about method of analysis in the protocol.	Low risk of bias	2% did not receive the intended interventions and therefore did not have outcome data. However, these numbers are low and probably did not have an effect on the outcome.	High risk of bias	Adverse events were reported by the participants and/or study personnel was responsible for detecting, documenting and reporting events. Both were aware of the assigned intervention. Knowledge of intervention received could have affected outcome measurement.	Low risk of bias	A study protocol with SAP available. Safety analysis was prespecified in the protocol and SAP.	High risk of bias	Overall judged high risk of bias due to conflicting information about method of analysis; the outcome assessors' awareness of assigned intervention.
NCT00920816 **Total trial population (combined risk groups)**	Low risk of bias	A centralized registration system was used for randomisation. There were no baseline imbalances	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 3 participants randomised to the experimental arm did not receive any treatment. The method	Low risk of bias	1% did not receive the intended interventions and therefore did not have outcome data. However, these	High risk of bias	Outcome assessors were aware of the assigned intervention. Knowledge of intervention received could have affected outcome measurement.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to the outcome assessors' awareness of assigned intervention; missing study

(Continued)

		that would suggest a problem with randomisation.		of analysis was appropriate.		numbers are low and probably did not have an effect on the outcome.						protocol and SAP.
NCT01030783 **Total trial population (combined risk groups)**	Low risk of bias	Interactive voice response system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 1 participant randomised to the experimental arm did not receive any treatment. The method of analysis was not appropriate (as-treated).	Low risk of bias	0.2% did not receive the intended interventions and therefore did not have outcome data. However, these numbers are very low and probably did not have an effect on the outcome.	High risk of bias	Measurement of AEs could have differed between intervention groups due to differences in number of visits to the healthcare provider. AEs were assessed by the participants and the investigator. Both were aware of the assigned intervention. Knowledge of intervention received could have affected outcome measurement.	Some concerns	No SAP available. Study protocol available, but unclear whether it was finalized before unblinded outcome data were available.	High risk of bias	Overall judged high risk of bias due to inappropriate method of analysis; probable differences in outcome measurement between intervention arms; the outcome assessors' awareness of assigned intervention; missing study protocol and SAP.
NCT00738530 **Total trial**	Low risk of bias	Interactive voice	High risk of bias	The study was double-blind: both participants and those	Low risk of bias	1.2% did not receive the	Low risk of bias	Outcome assessors were not	Some concerns	No study protocol or SAP	High risk of bias	Overall judged high risk

(Continued)

population **(combined risk groups)**		recognition system was used for randomisation. There were no baseline imbalances that would suggest a problem with randomisation.		delivering the intervention were not aware of assigned interventions. Only 2 participants from the control arm and 6 participants from the intervention arm did not receive any treatment. The method of analysis was not appropriate (as-treated).		intended interventions and therefore did not have outcome data. However, these numbers are low and probably did not have an effect on the outcome.		aware of the assigned intervention.		available.		of bias due to inappropriate method of analysis; missing study protocol and SAP.
NCT01274273 **Total trial population** **(combined risk groups)**	Low risk of bias	Participants were randomised in a 1:1 ratio. Allocation was probably controlled by an external unit. There were no baseline imbalances that would suggest a problem with	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. No information provided about the method of analysis.	Low risk of bias	We assume all participants received the intended interventions.	High risk of bias	No precise information provided about method of measuring AEs. However, AEs were defined according to NCICTC, version 3. Outcome assessors were aware of the assigned intervention. Knowledge of intervention received could have affected outcome measurement.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about method of analysis and method of outcome measurement; the outcome assessors' awareness of assigned intervention; missing study protocol and SAP.

(Continued)

		random-szation.										
NCT01108445 **Total trial population** **(combined risk groups)**	Low risk of bias	Participants were randomised in a 1:1 ratio. Randomisation was done under allocation concealment. There were no baseline imbalances that would suggest a problem with randomisation.	High risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned intervention. Only 1 participant withdrew after consent, but before randomisation and before study drug was assigned. No precise information provided about the method of analysis.	Low risk of bias	All 108 participants were evaluable.	High risk of bias	No precise information provided about method of measuring AEs. However, AEs were defined according to NCICTC, version 4. Outcome assessors were aware of the assigned intervention. Knowledge of intervention received could have affected outcome measurement.	Some concerns	No study protocol or SAP available.	High risk of bias	Overall judged high risk of bias due to lack of information about method of analysis and method of outcome measurement; the outcome assessors' awareness of assigned intervention; missing study protocol and SAP.
NCT00903175 **Total trial population** **(combined risk groups)**	Low risk of bias	Interactive voice response system was used for randomsation. There were no baseline imbal-	Low risk of bias	The study was open-label: both participants and those delivering the intervention were aware of assigned interventions. Only 2 participants randomised to the control arm did not receive any treatment. No precise infor-	Low risk of bias	Less than 1% did not receive the intended interventions and therefore did not have outcome	High risk of bias	AEs were assessed by the participants who were aware of the assigned intervention. Knowledge of intervention received could have affect-	Some concerns	Study protocol available with some statistical methods described, but no separate SAP	High risk of bias	Overall judged high risk of bias due to the outcome assessors' awareness of assigned intervention; missing SAP.

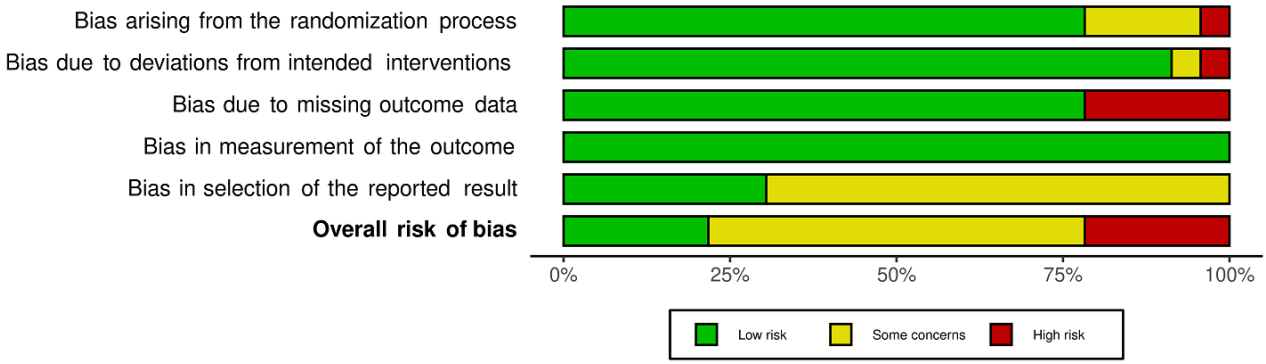
(Continued)

	ances that would suggest a problem with randomisation.	mation provided about the method of analysis, but data for crossed-over participants were reported separately. We assume participants were analyzed as randomised in period 1.	data. However, these numbers are low and probably did not have an effect on the outcome.	ed outcome measurement.	available to fully check the pre-planned analyses.

Appendix 14. Additional figures (risk of bias summary plots)

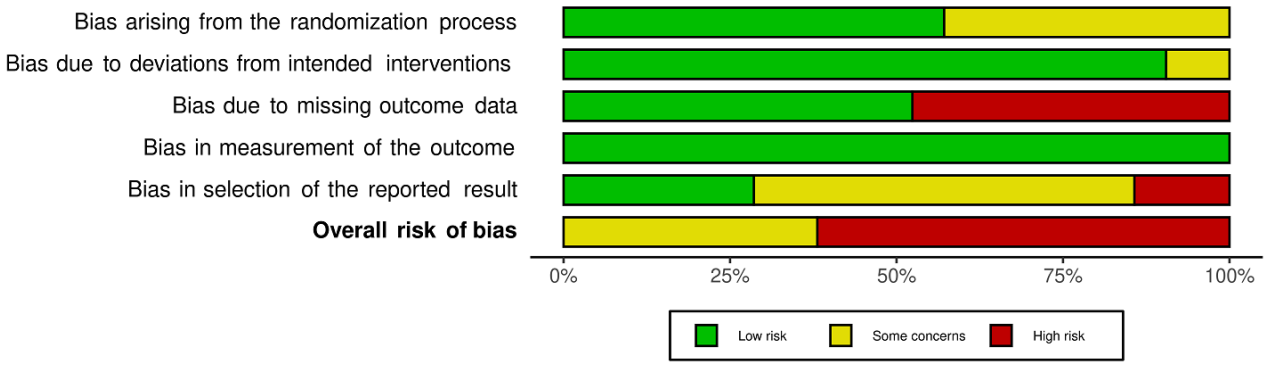
1. Summary plot for OS for all risk groups combined: Figure 53

Figure 53. Summary plot for OS for all risk groups combined



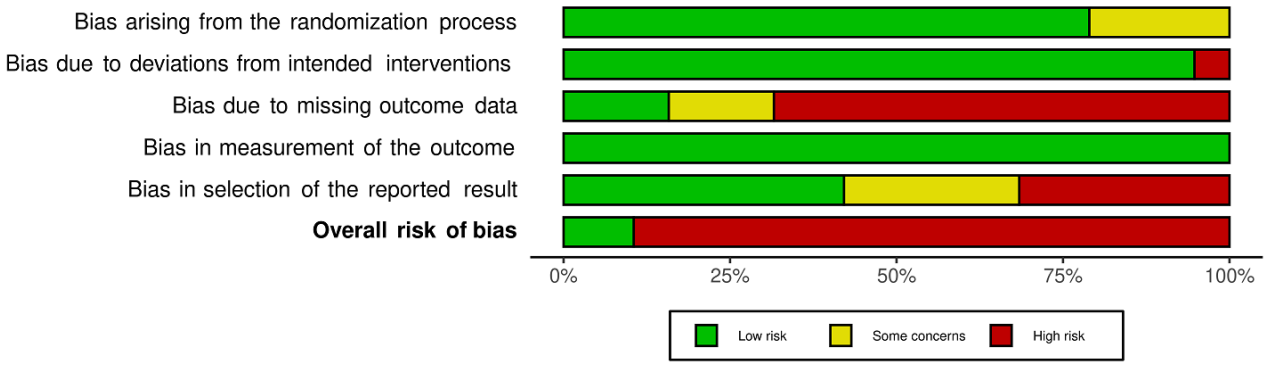
2. Summary plot for OS per MSKCC favourable, intermediate, poor risk, respectively: Figure 54

Figure 54. Summary plot for OS per MSKCC favourable, intermediate, poor risk, respectively



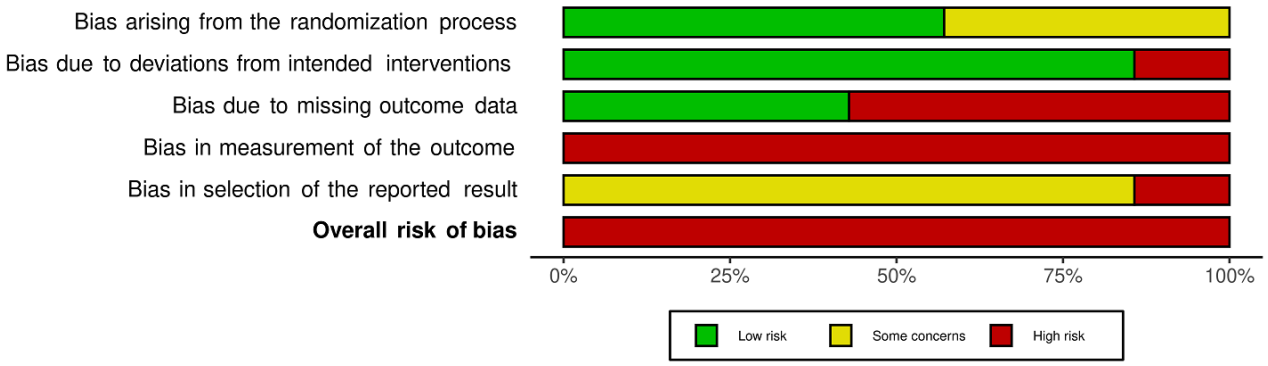
3. Summary plot for OS per IMDC favourable, intermediate, poor risk, respectively: Figure 55

Figure 55. Summary plot for OS per IMDC favourable, intermediate, poor risk, repsectively



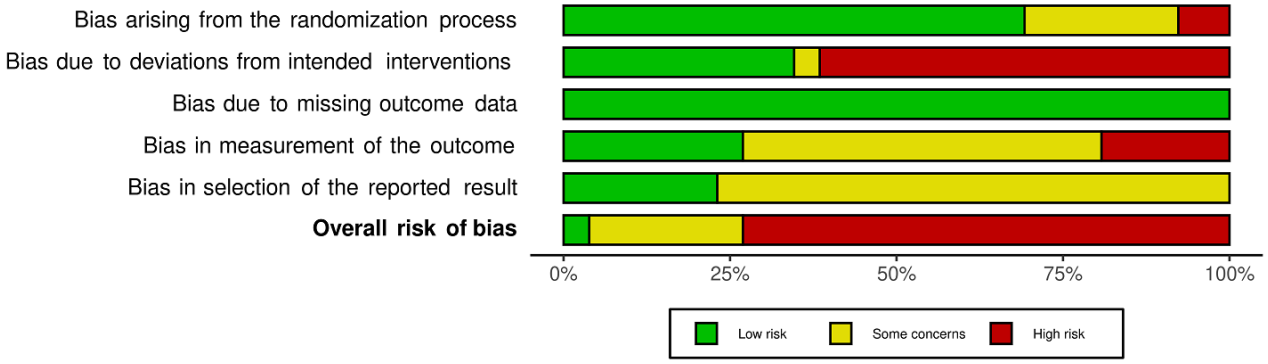
4. Summary plot for QoL for all risk groups combined: Figure 56

Figure 56. Summary plot for QoL for all risk groups combined



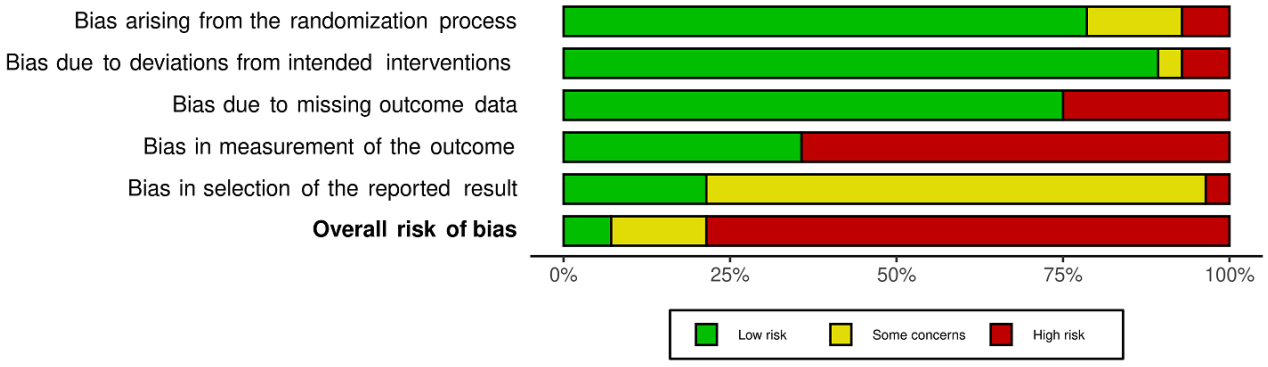
5. Summary plot for SAEs for all risk groups combined: Figure 57

Figure 57. Summary plot for SAEs for all risk groups combined



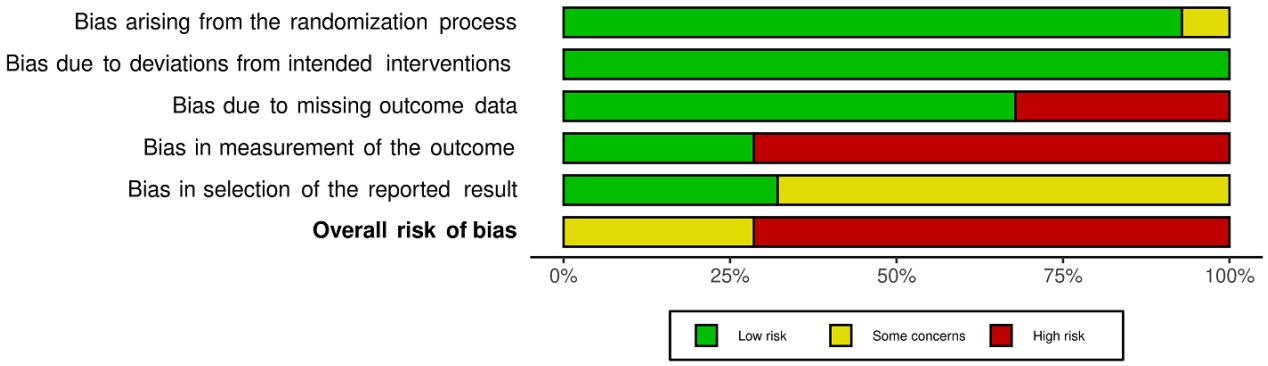
6. Summary plot for PFS (all risk groups combined): Figure 58

Figure 58. Summary plot for PFS (all risk groups combined)



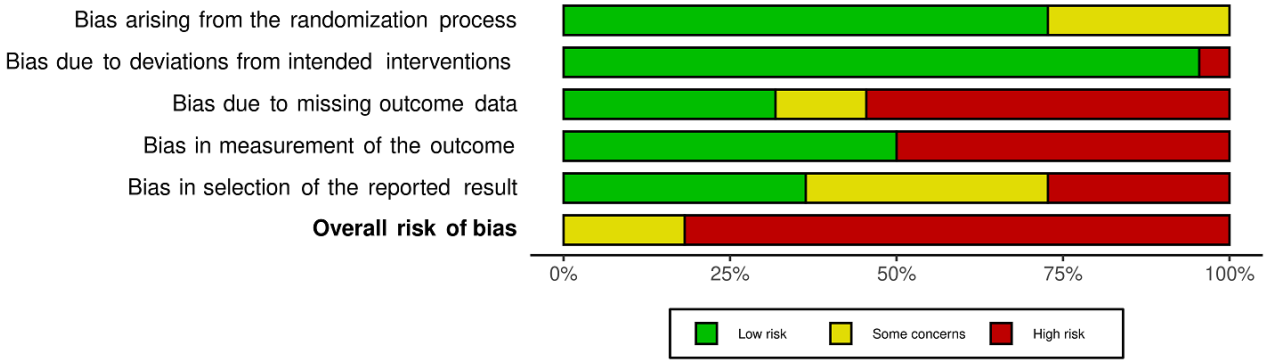
7. Summary plot for PFS per MSKCC favourable, intermediate, poor risk, respectively: Figure 59

Figure 59. Summary plot for PFS per MSKCC favourable, intermediate, poor risk, respectively



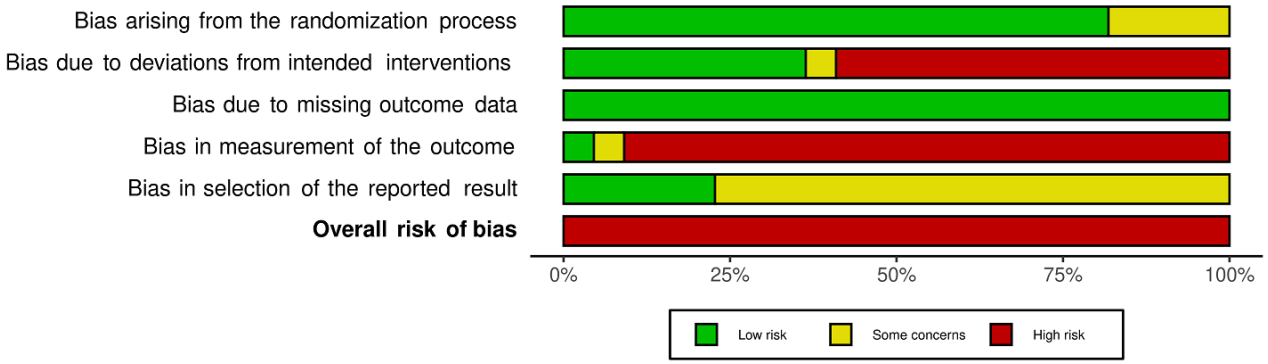
8. Summary plot for PFS per IMDC favourable, intermediate, poor risk, respectively: Figure 60

Figure 60. Summary plot for PFS per IMDC favourable, intermediate, poor risk, respectively



9. Summary plot for all-cause grade 3 or 4 AEs for all risk groups combined: Figure 61

Figure 61. Summary plot for all-cause grade 3 or 4 AEs for all risk groups combined



Appendix 15. Additional figures (main analyses of OS, SAEs, PFS, AEs, and Number of participants who discontinued study treatment due to an AE)

1. Pairwise comparison for OS (all risk groups combined): Figure 62

Figure 62. Pairwise comparison for OS (all risk groups combined)

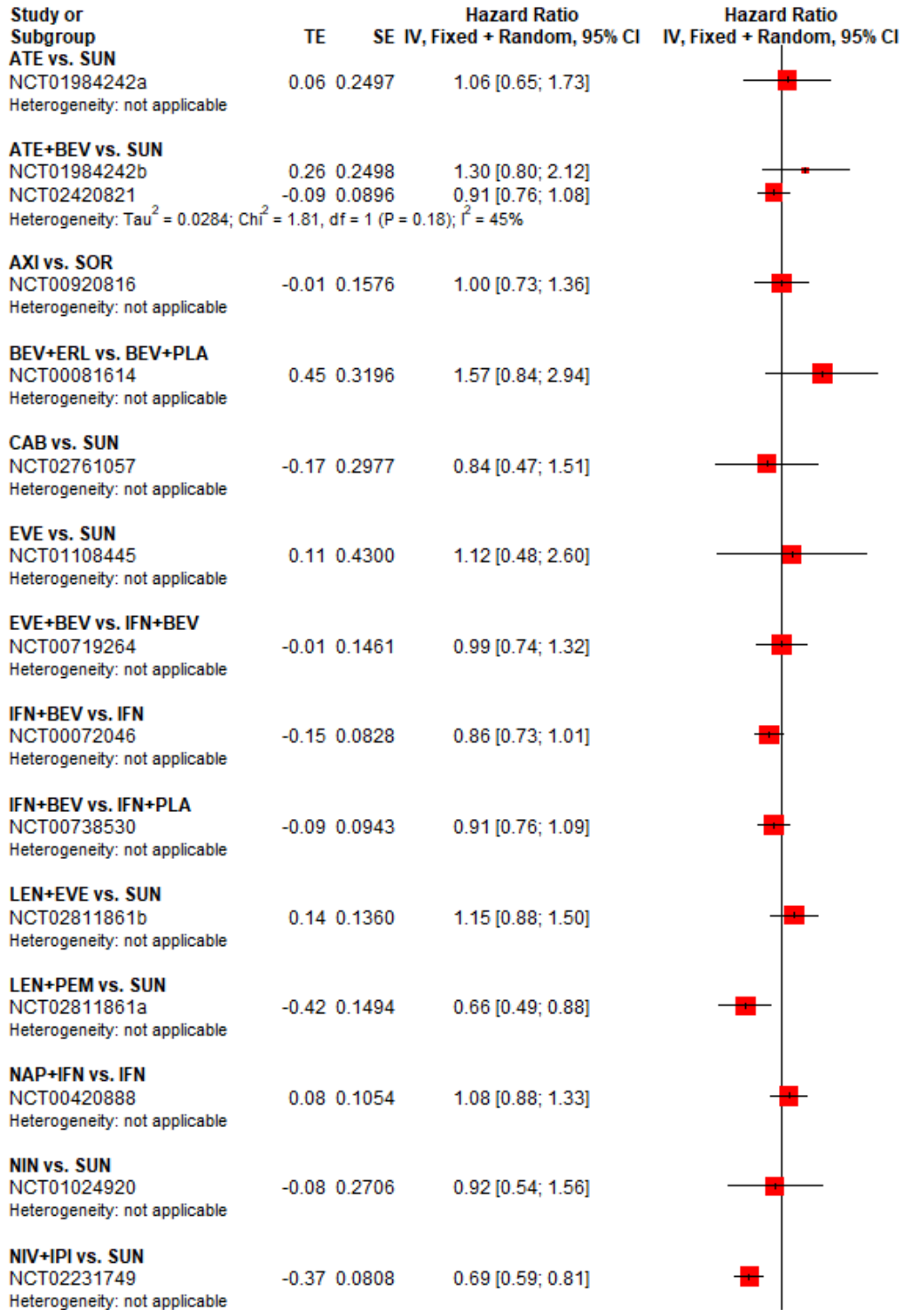
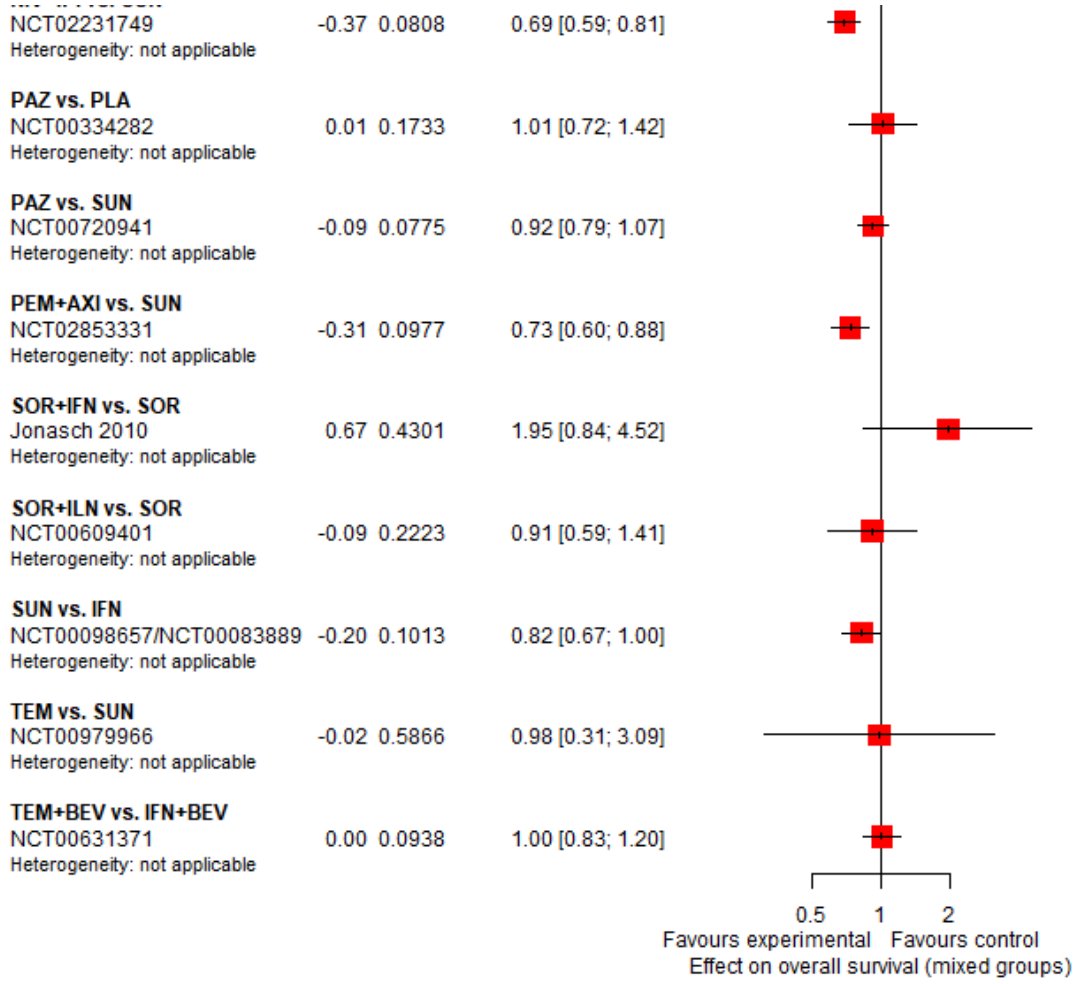
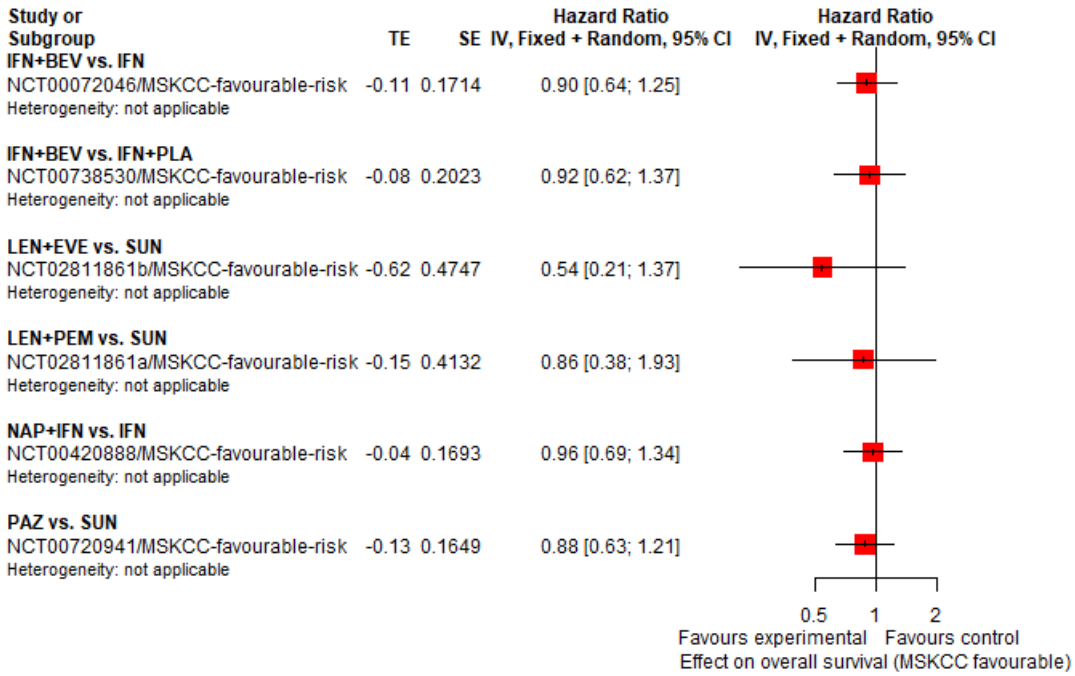


Figure 62. (Continued)



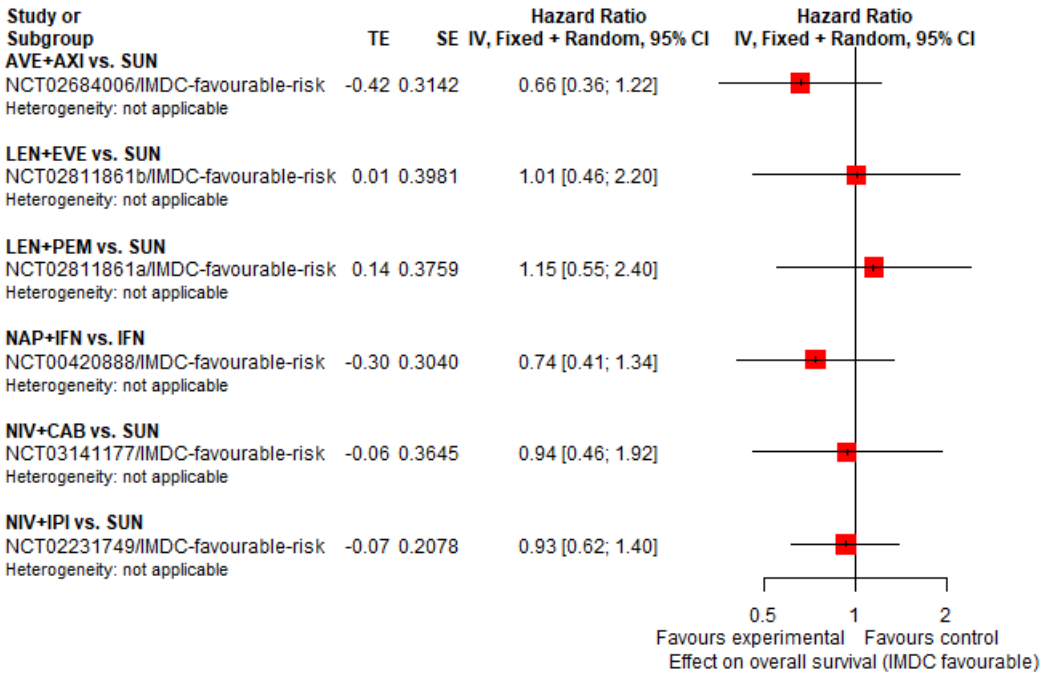
2. Pairwise comparison for OS (MSKCC favourable): Figure 63

Figure 63. Pairwise comparison for OS (MSKCC favourable)



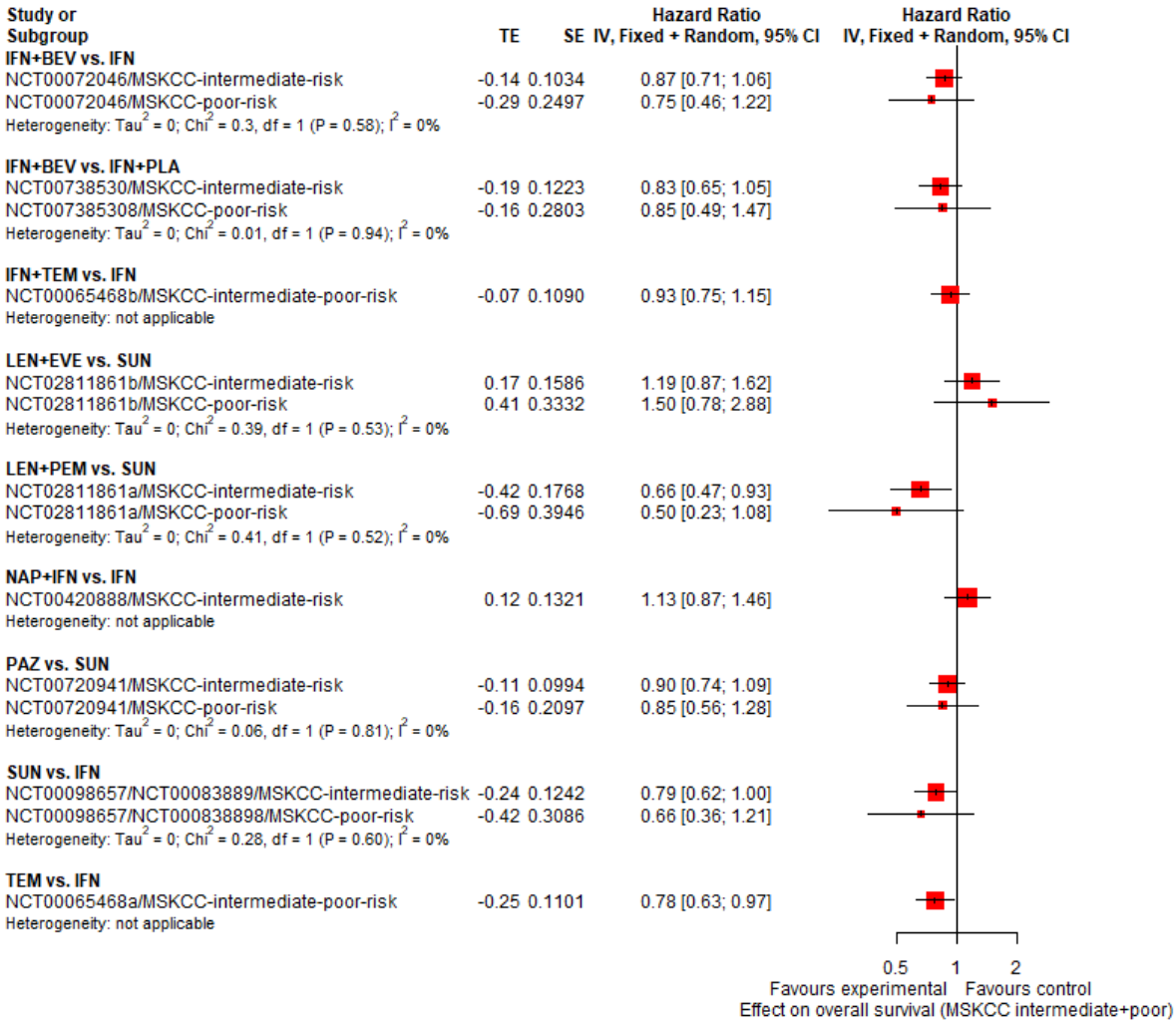
3. Pairwise comparison for OS (IMDC favourable): Figure 64

Figure 64. Pairwise comparison for OS (IMDC favourable)



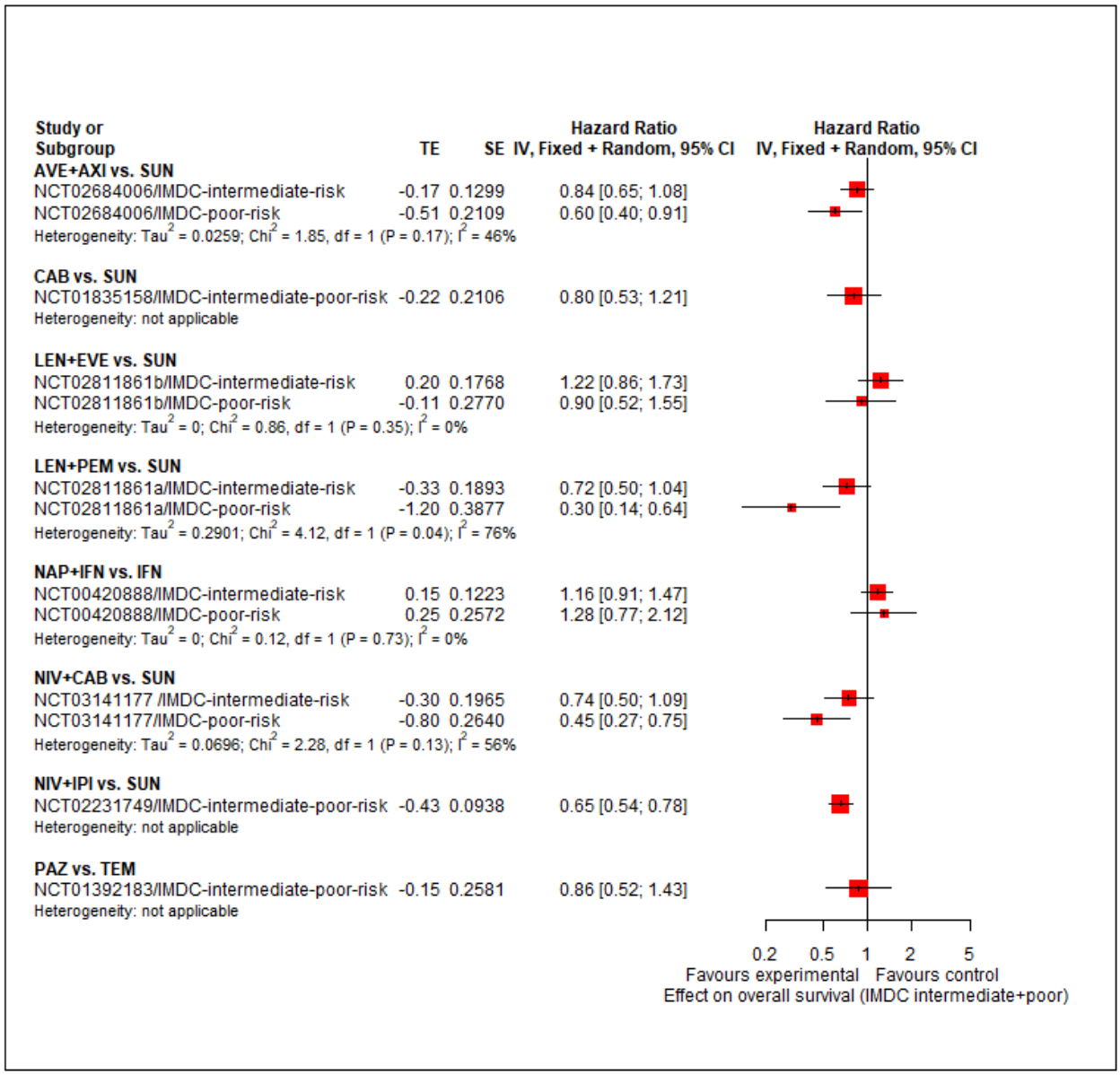
4. Pairwise comparison for OS (MSKCC intermediate, poor): Figure 65

Figure 65. Pairwise comparison for OS (MSKCC intermediate, poor)



5. Pairwise comparison for OS (IMDC intermediate, poor): Figure 66

Figure 66. Pairwise comparison for OS (IMDC intermediate, poor)



6. Pairwise comparison for SAEs (all risk groups combined): Figure 67

Figure 67. Pairwise comparison for SAEs (all risk groups combined)

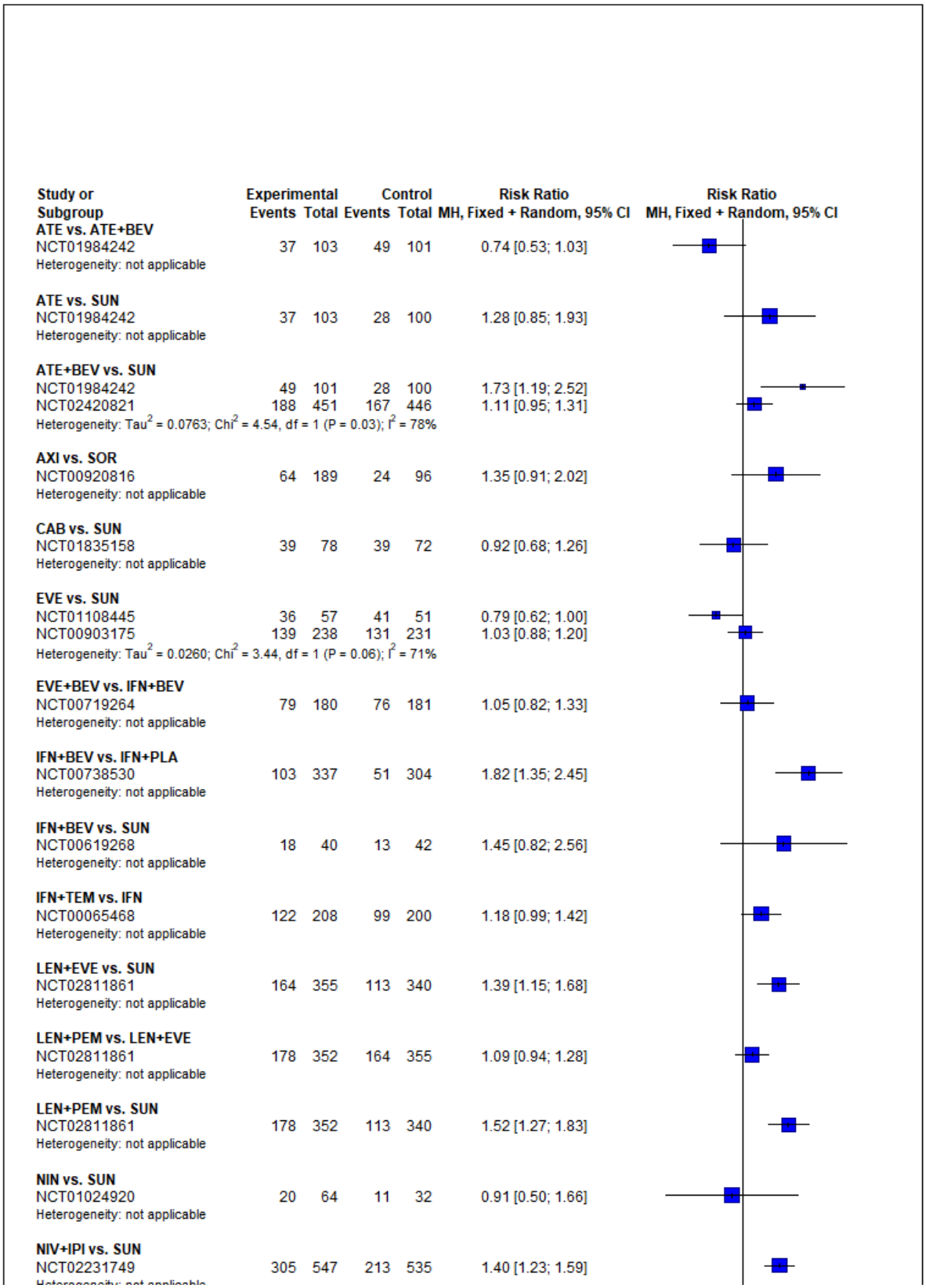
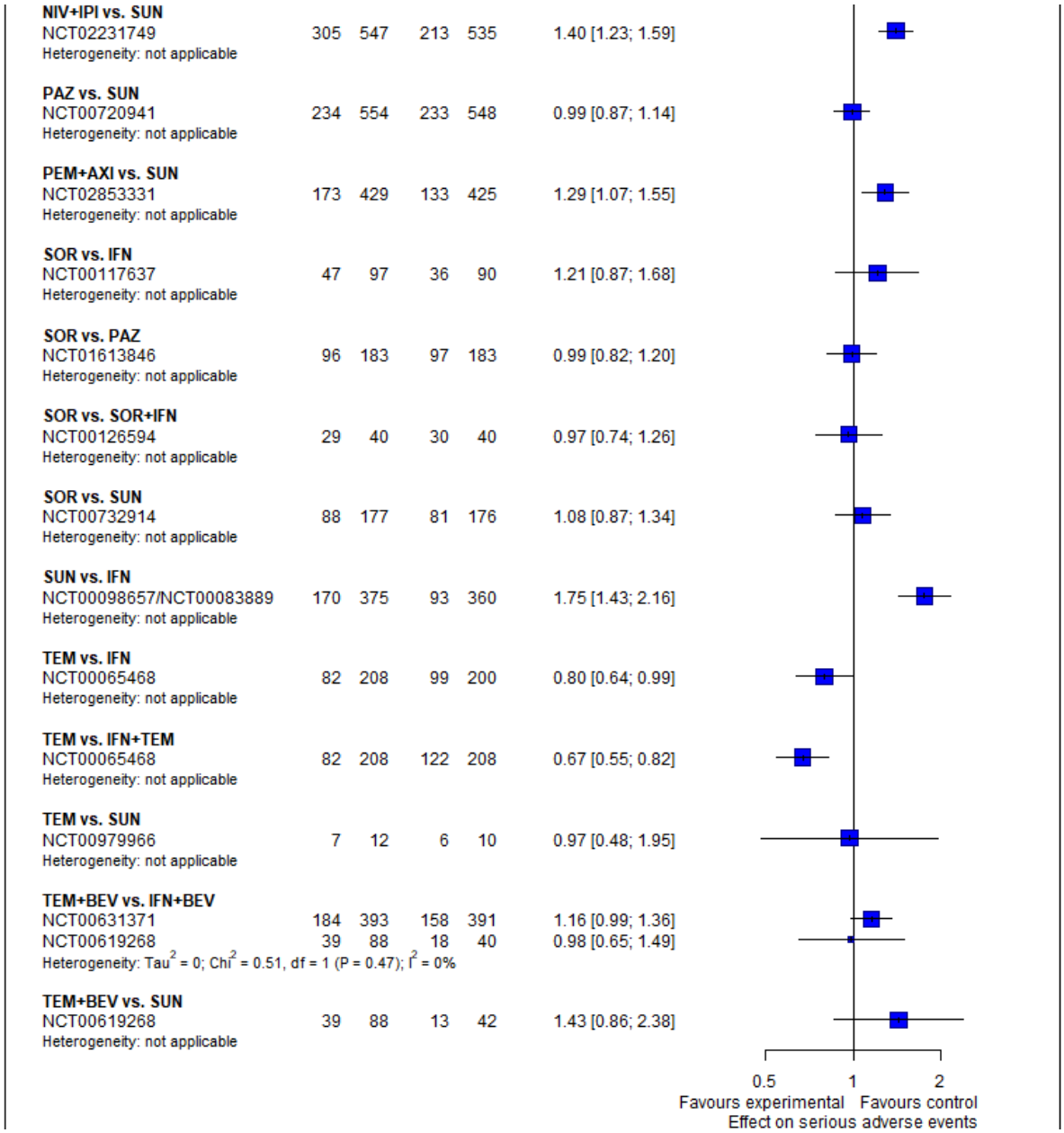


Figure 67. (Continued)



7. Forest plot of splitting direct and indirect evidence for SAE (all risk groups combined): Figure 68

Figure 68. Forest plot of splitting direct and indirect evidence for SAE (all risk groups combined)

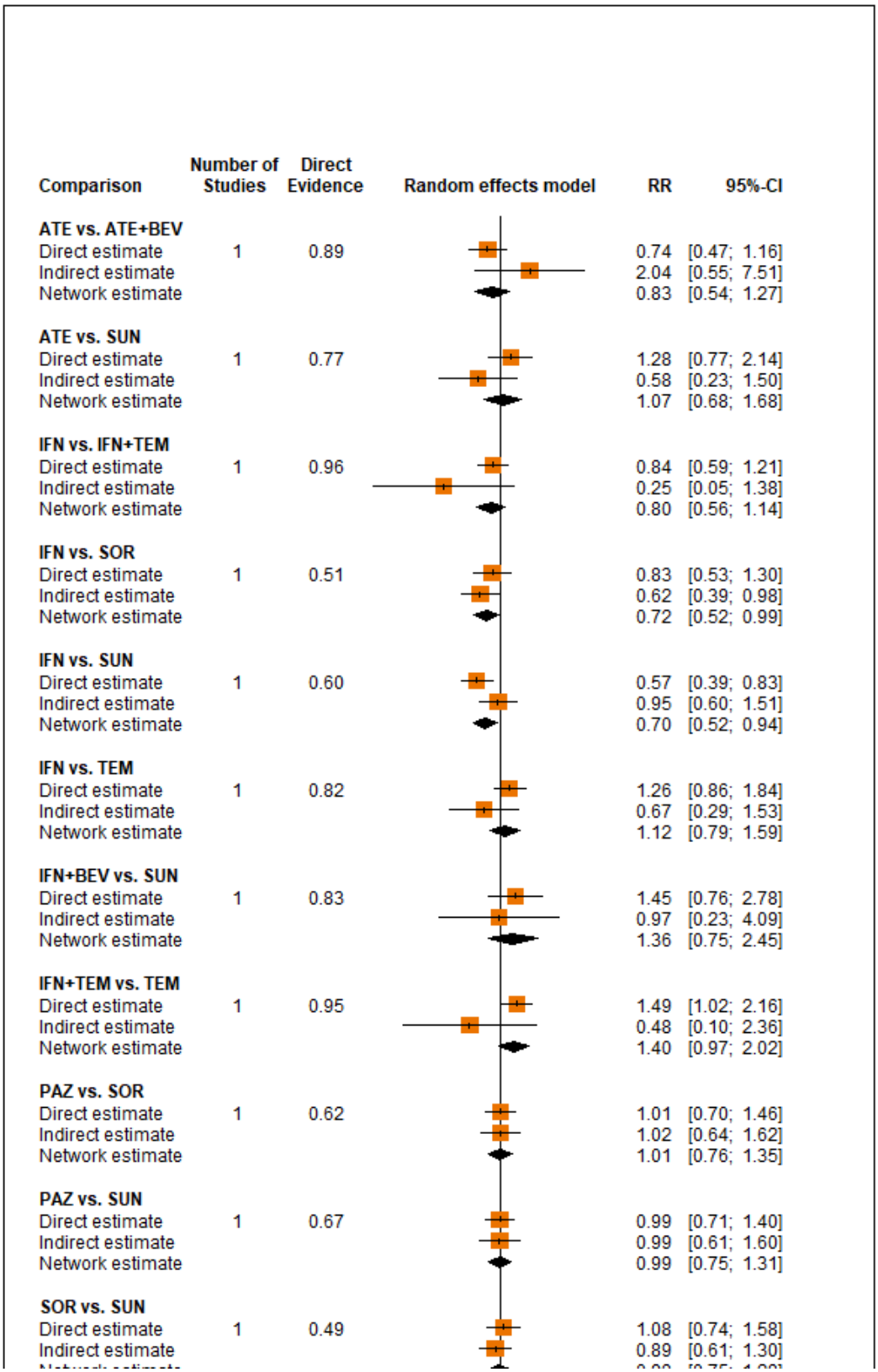
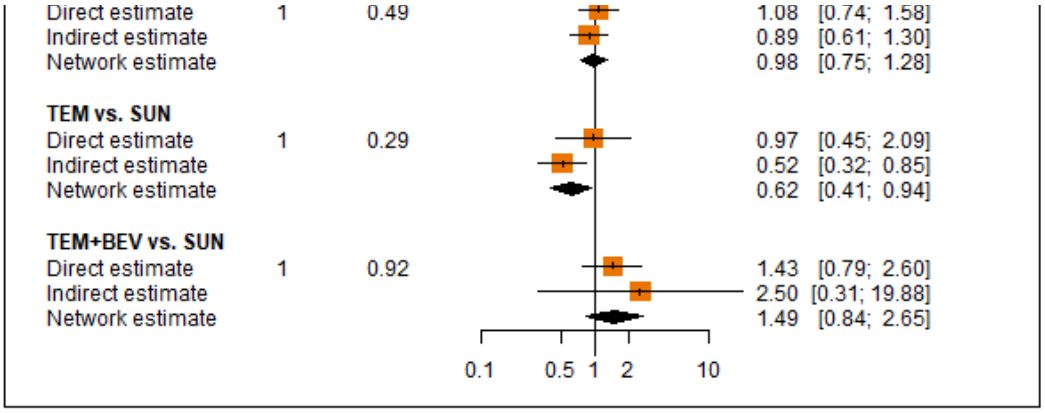
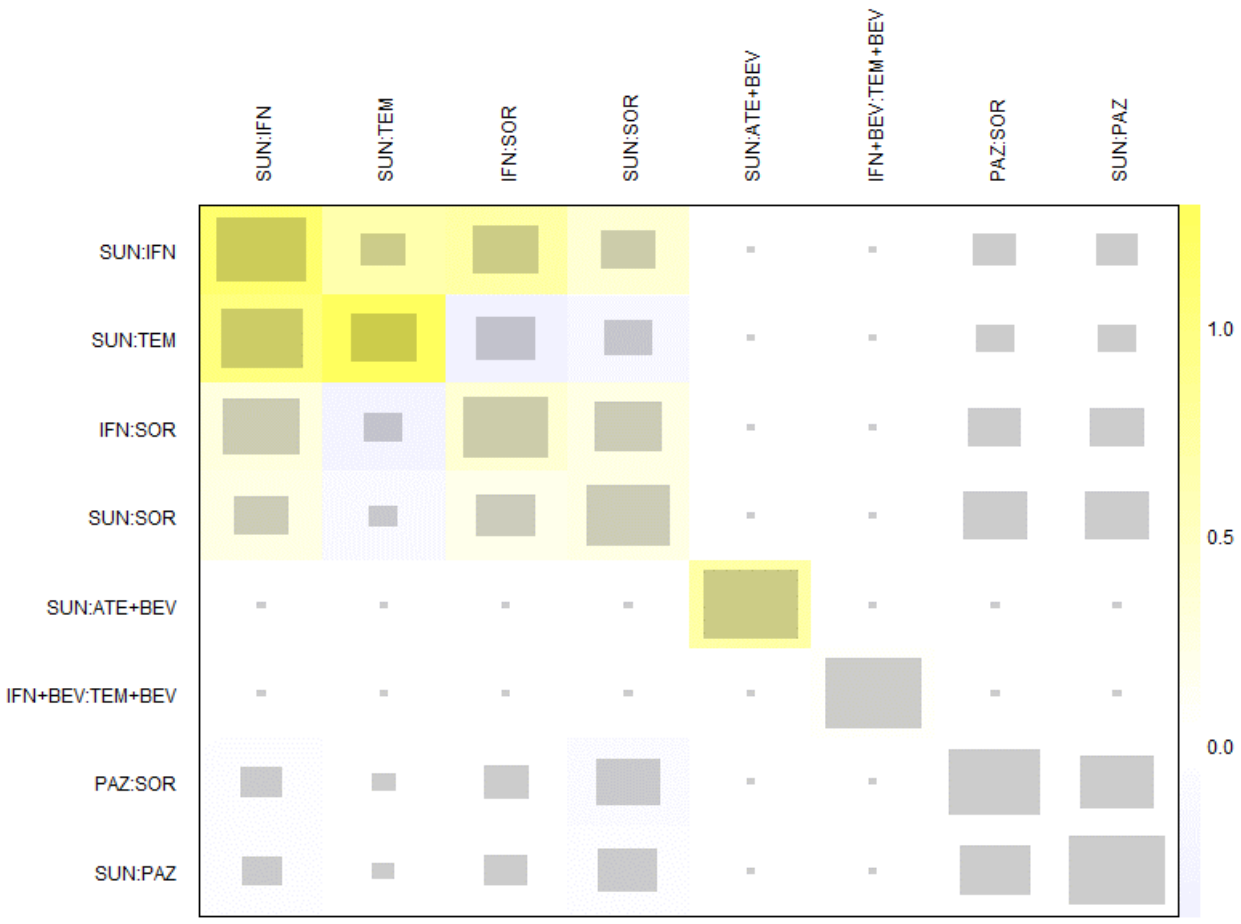


Figure 68. (Continued)



8. Net heat plot for SAEs (all risk groups combined) : Figure 69

Figure 69. Net heat plot for SAEs (all risk groups combined)



9. Pairwise comparison for PFS (all risk groups combined): Figure 70

Figure 70. Pairwise comparison for PFS (all risk groups combined)

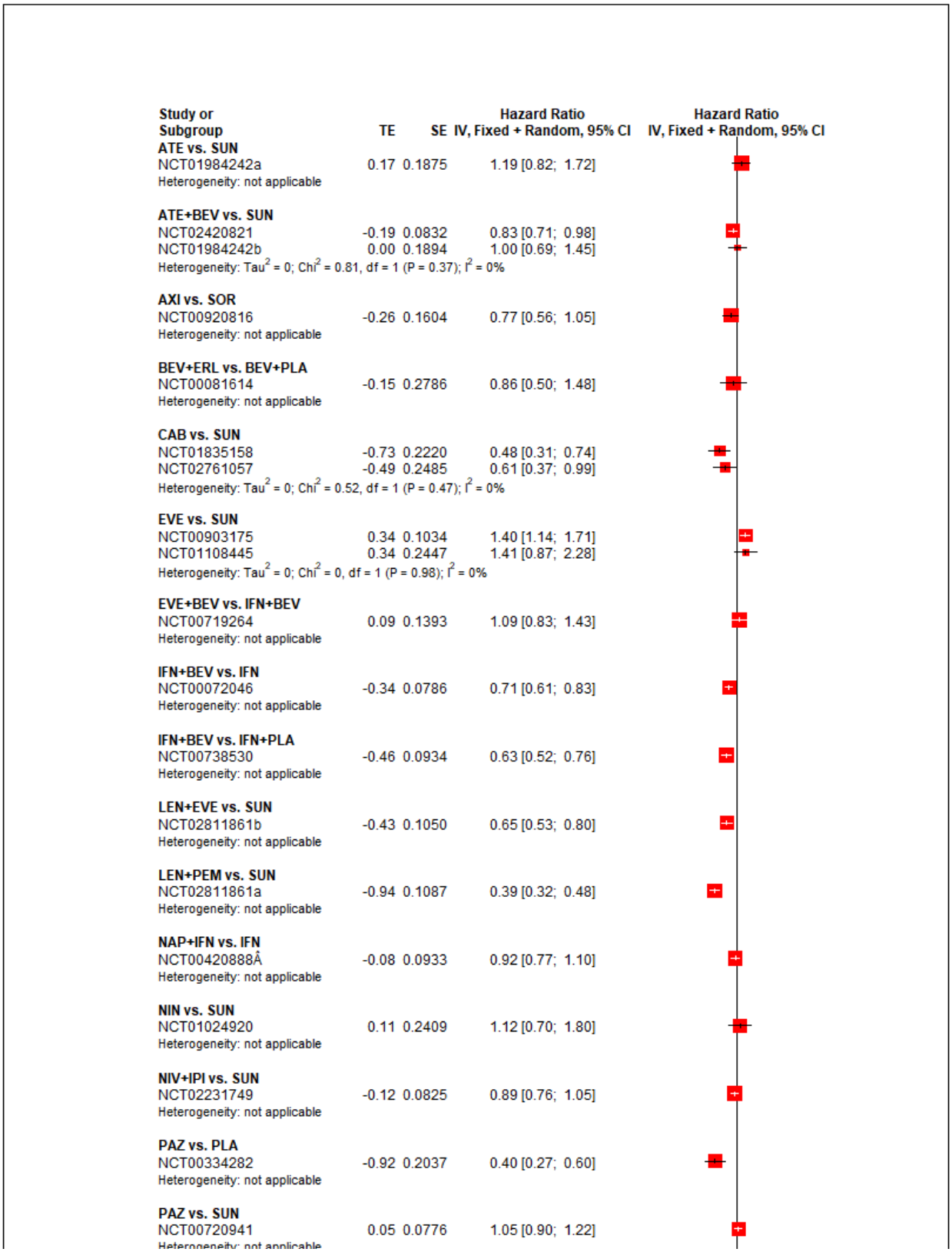
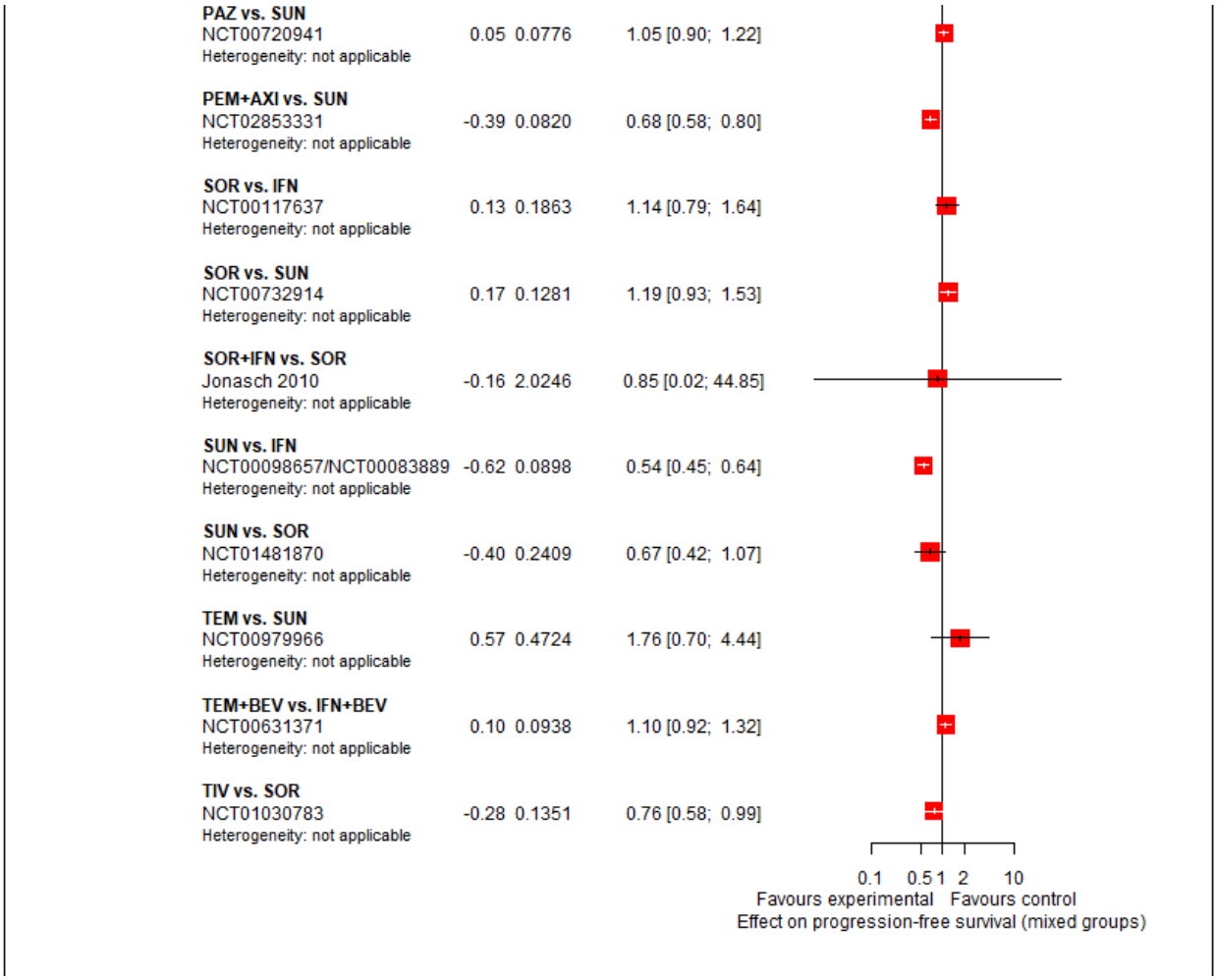
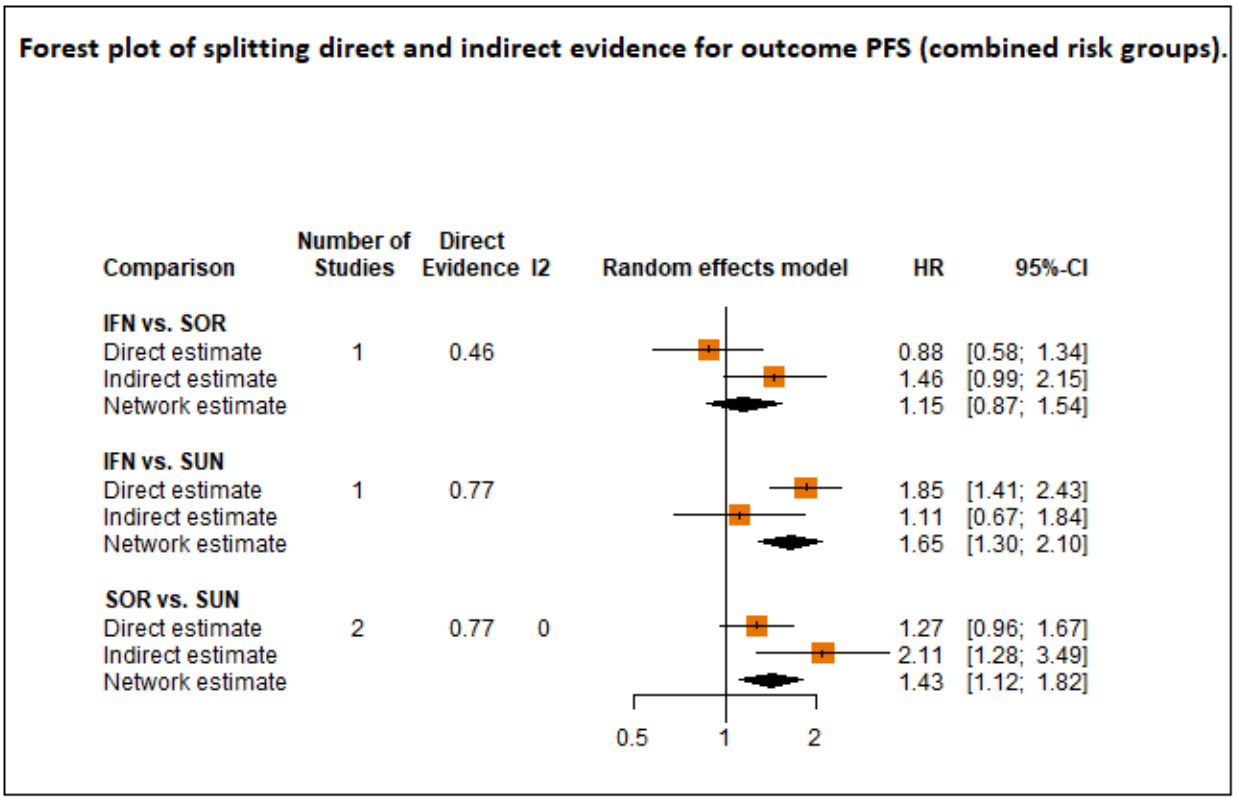


Figure 70. (Continued)



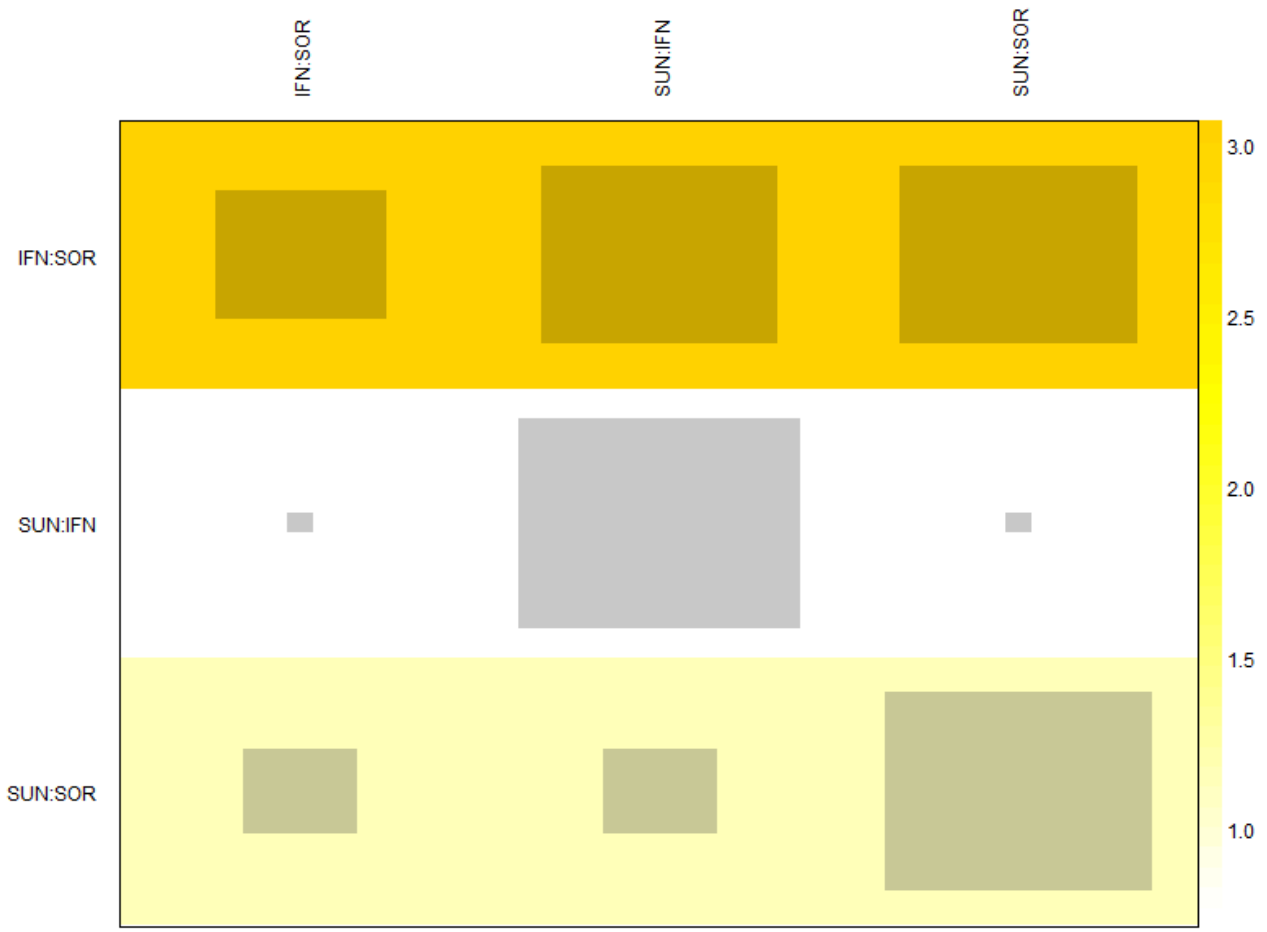
10. Forest plot of splitting direct and indirect evidence for PFS (all risk groups combined): Figure 71

Figure 71. Forest plot of splitting direct and indirect evidence for PFS (all rsik groups combined)



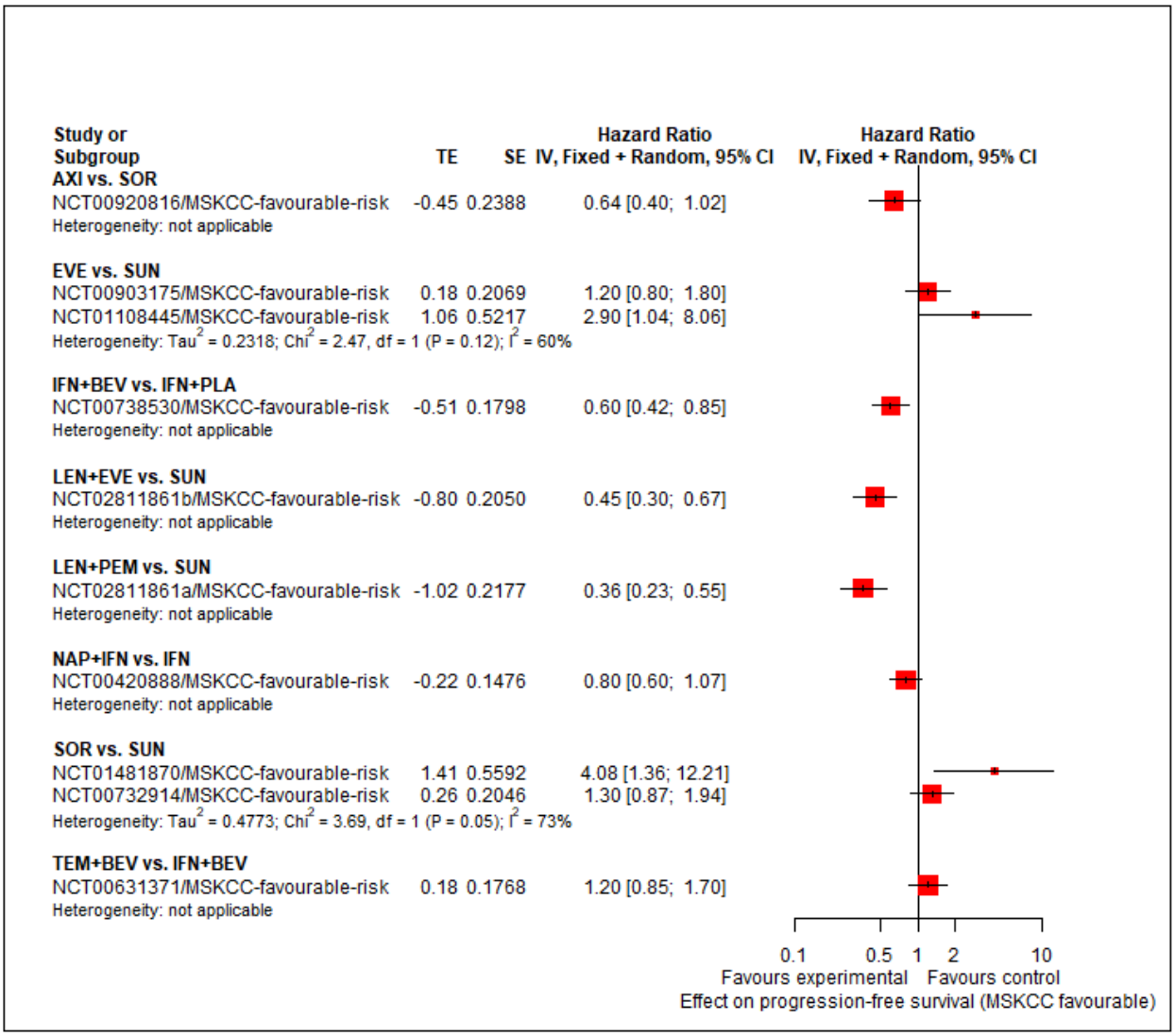
11. Net heat plot for PFS (all risk groups combined): Figure 72

Figure 72. Net heat plot for PFS (all risk groups combined)



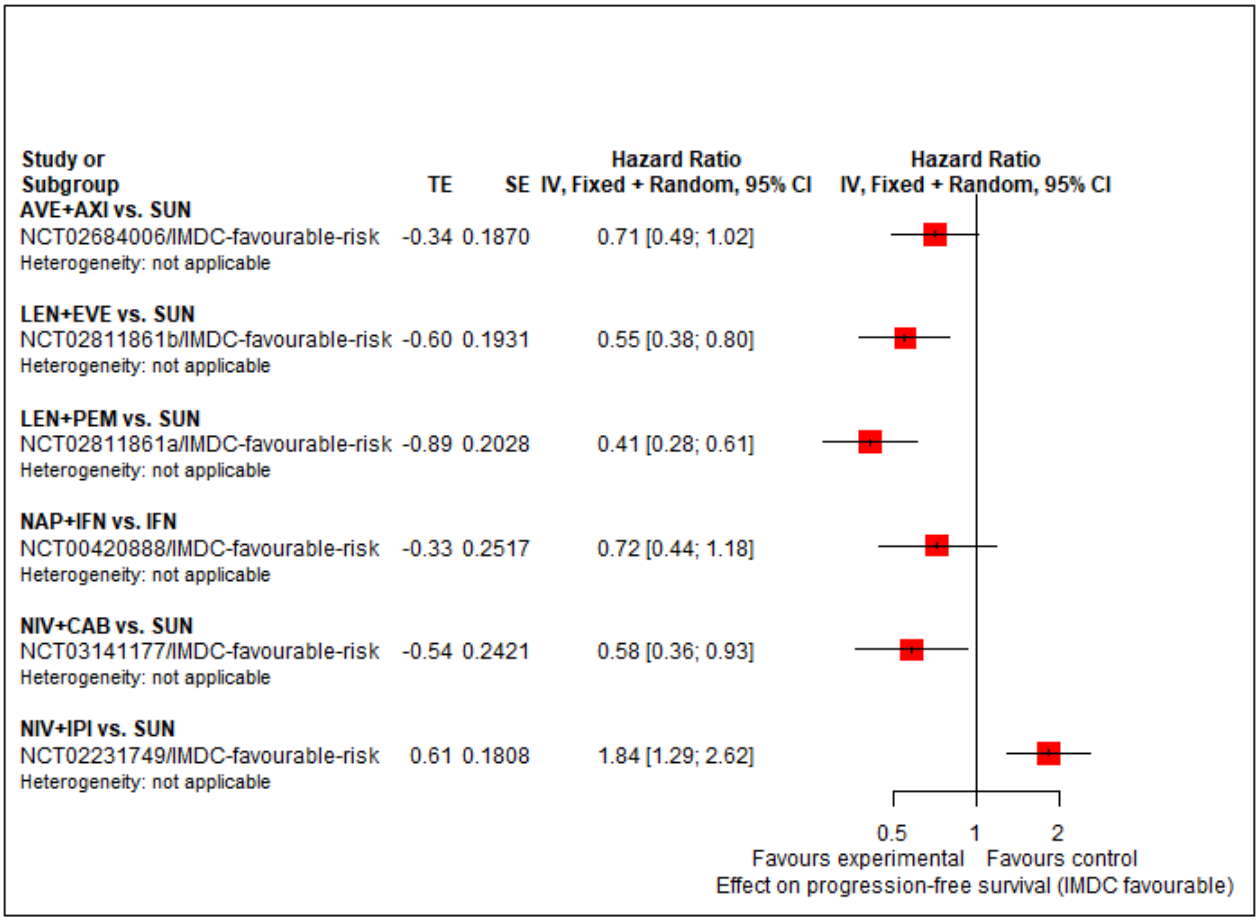
12. Pairwise comparison for PFS (MSKCC favourable): Figure 73

Figure 73. Pairwise comparison for PFS (MSKCC favourable)



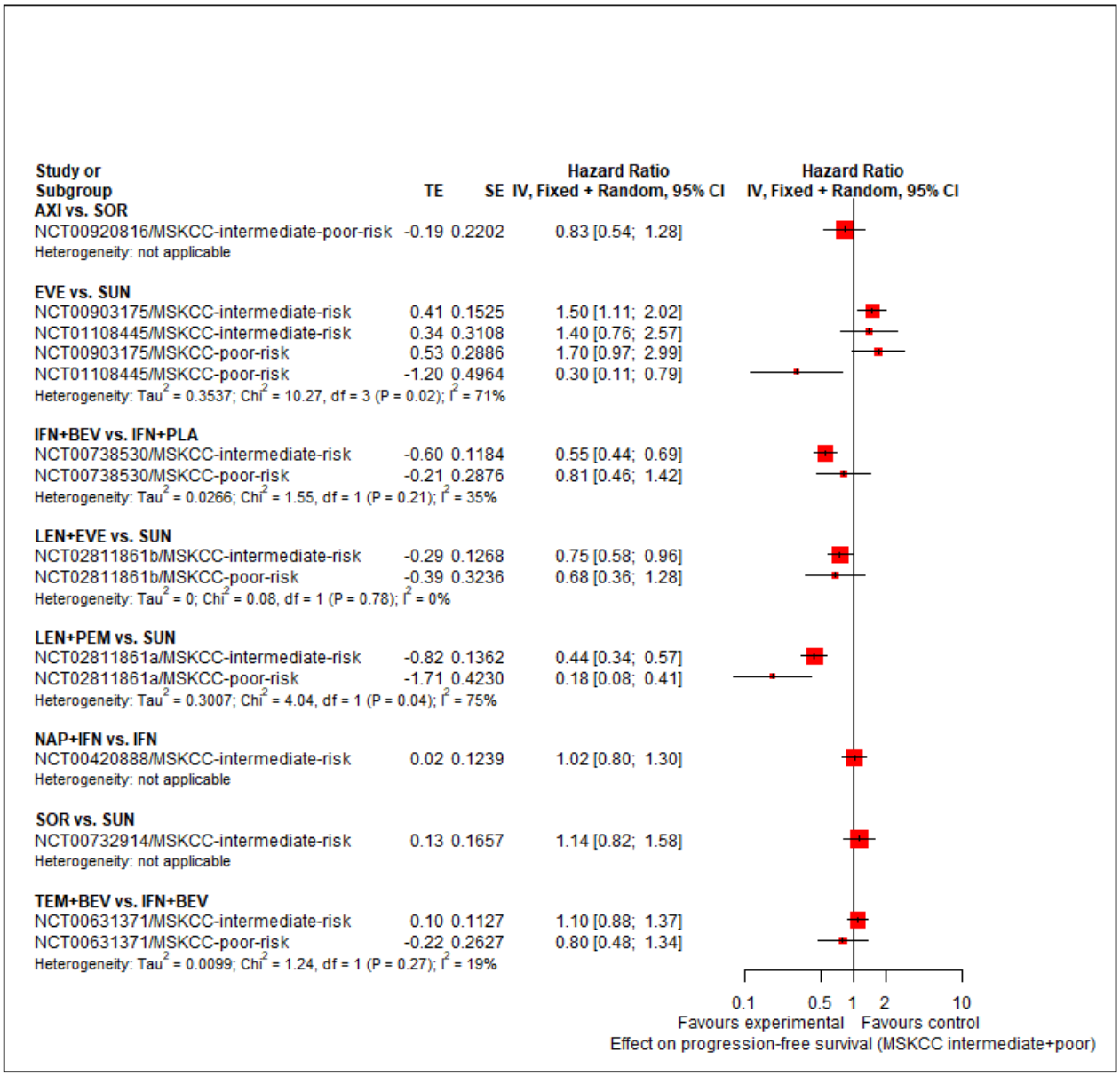
13. Pairwise comparison for PFS (IMDC favourable):Figure 74

Figure 74. Pairwise comparison for PFS (IMDC favourable)



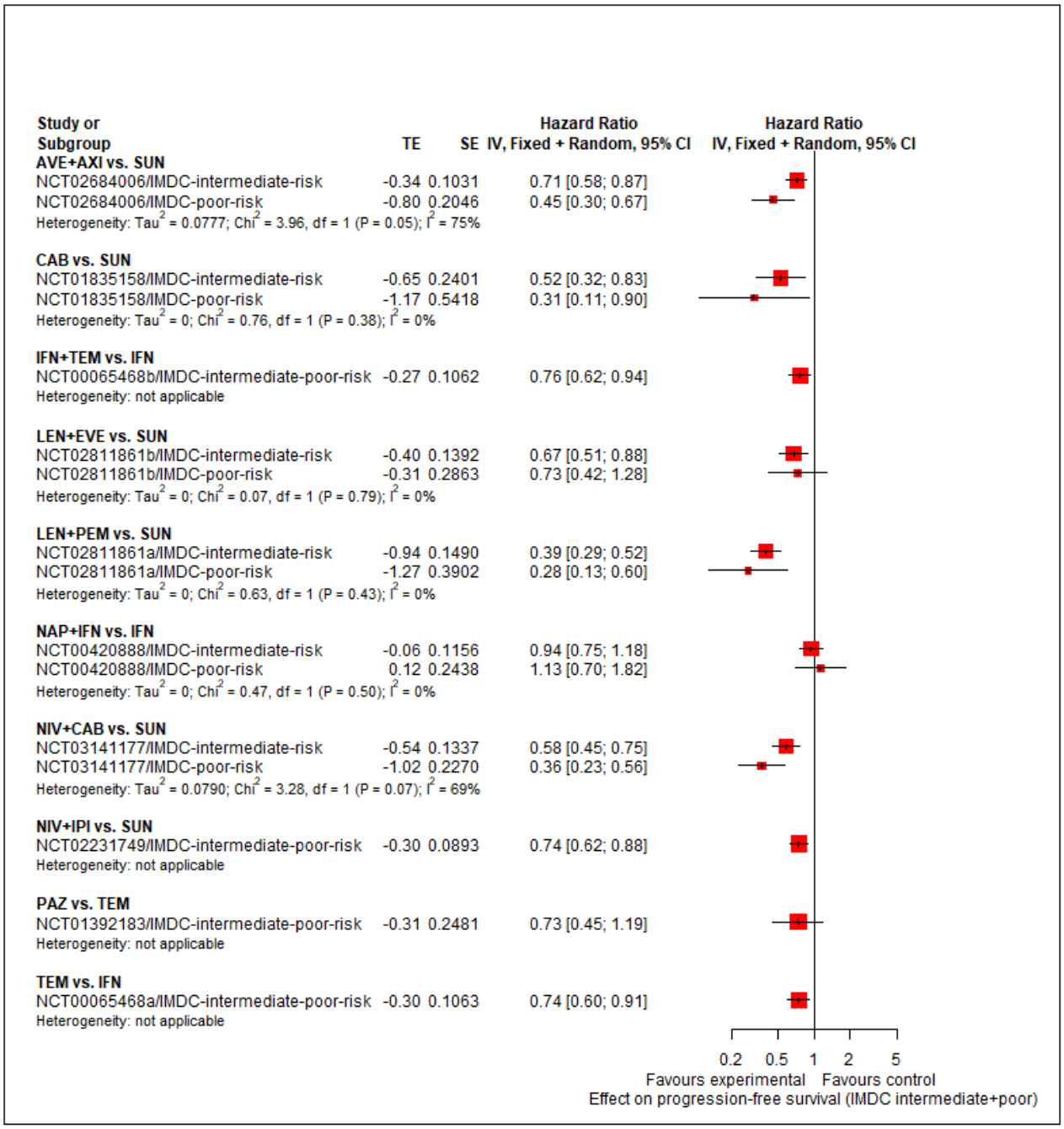
14. Pairwise comparison for PFS (MSKCC intermediate, poor): Figure 75

Figure 75. Pairwise comparison for PFS (MSKCC intermediate, poor)



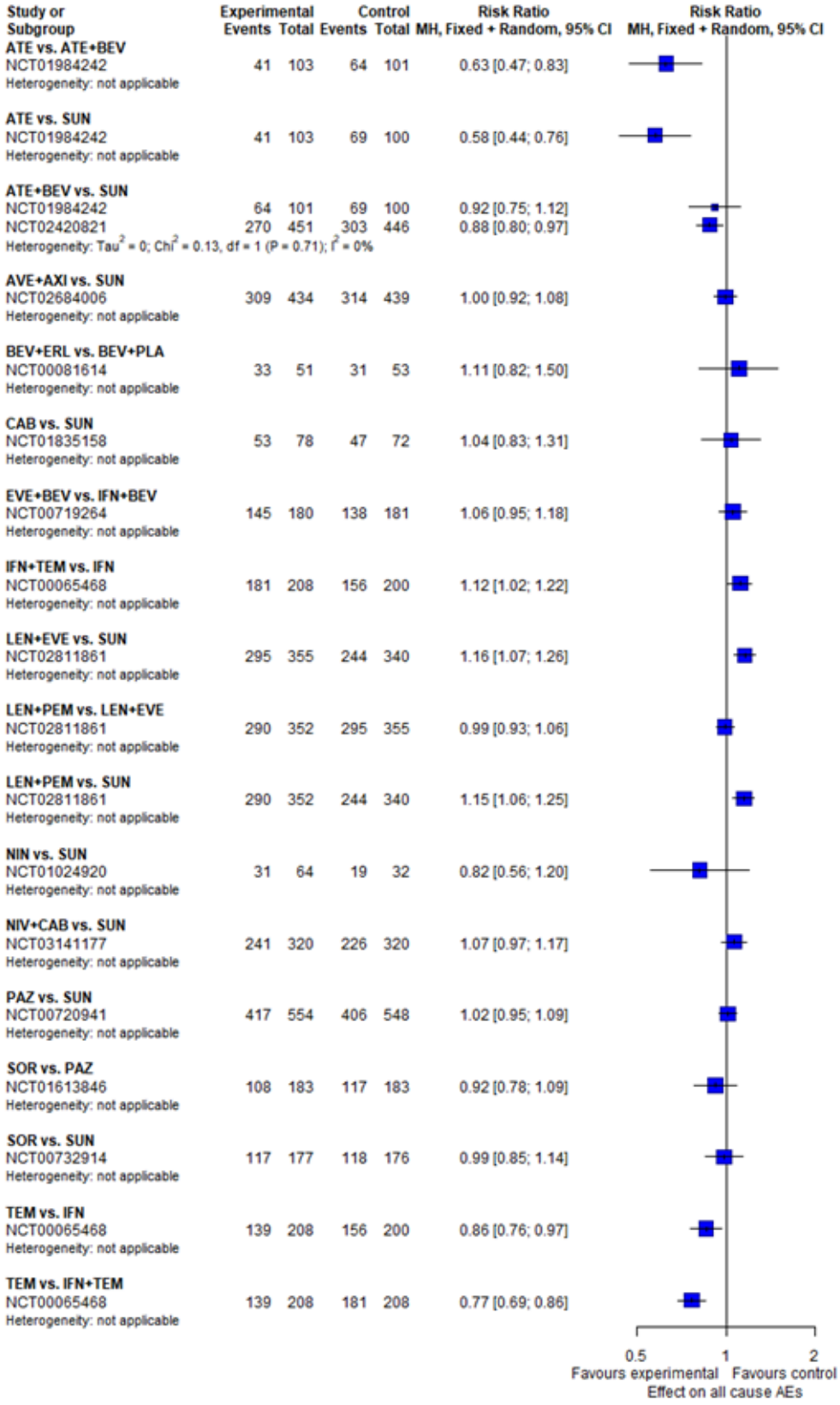
15. Pairwise comparison for PFS (IMDC intermediate, poor): Figure 76

Figure 76. Pairwise comparison for PFS (IMDC intermediate, poor)



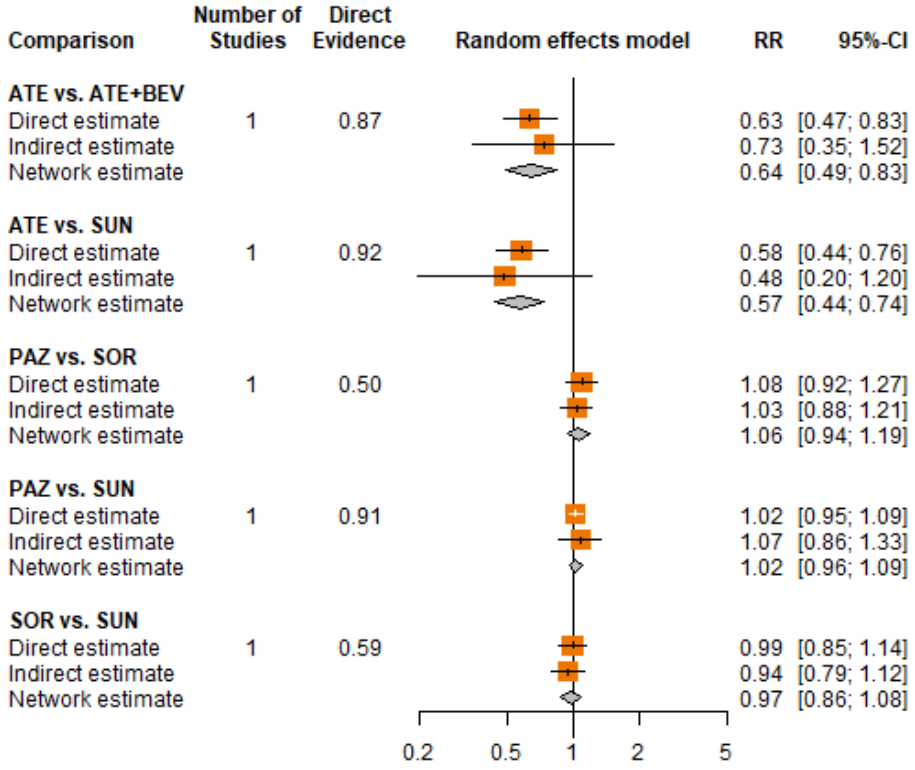
13. Pairwise comparison for all-cause AEs (all risk groups combined): Figure 77

Figure 77. Pairwise comparison for all-cause AEs (all risk groups combined)



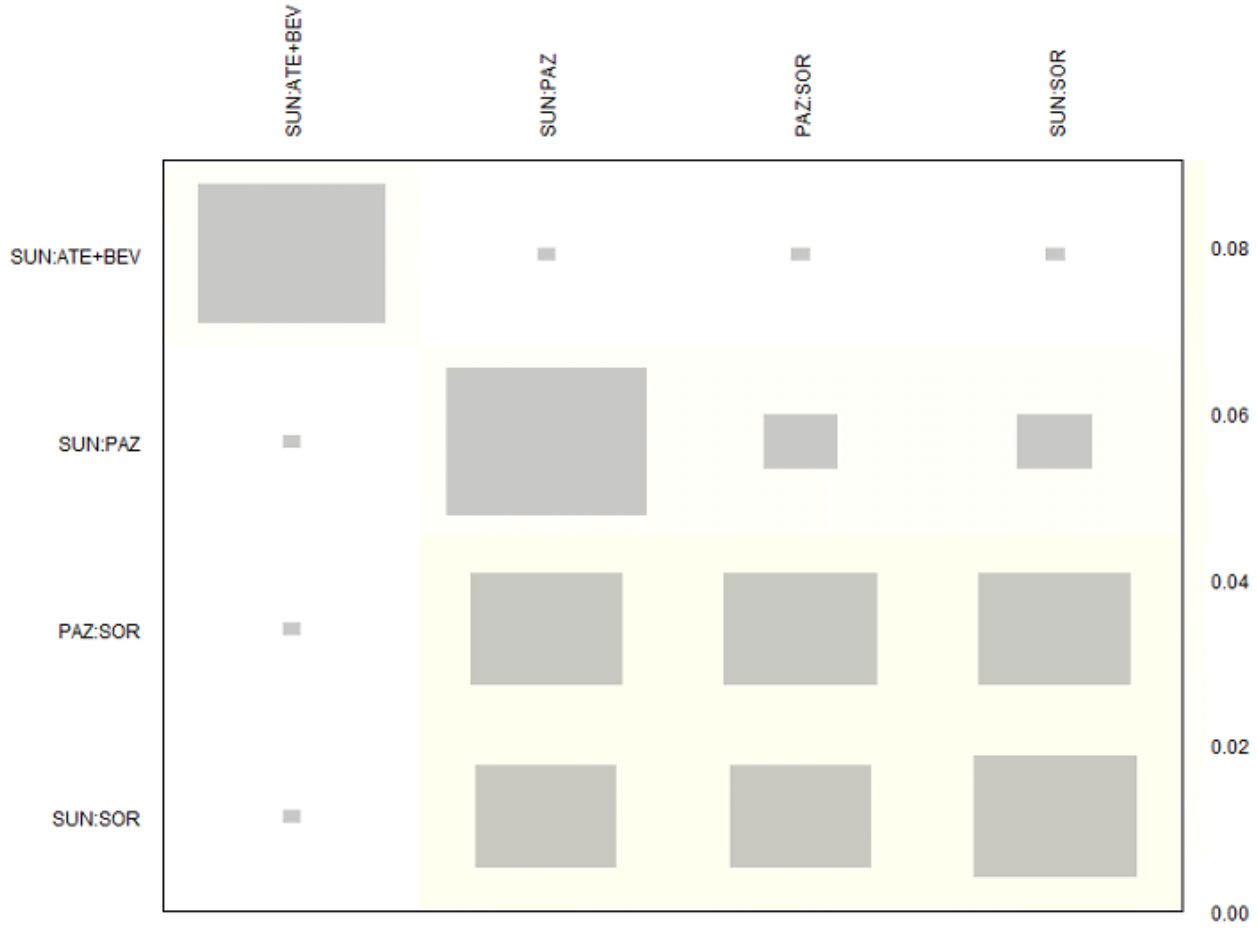
16. Forest plot of splitting direct and indirect evidence for all-cause grade 3or 4 AEs (all risk groups combined): Figure 78

Figure 78. Forest plot of splitting direct and indirect evidence for all-cause grade 3or 4 AEs (all risk groups combined)



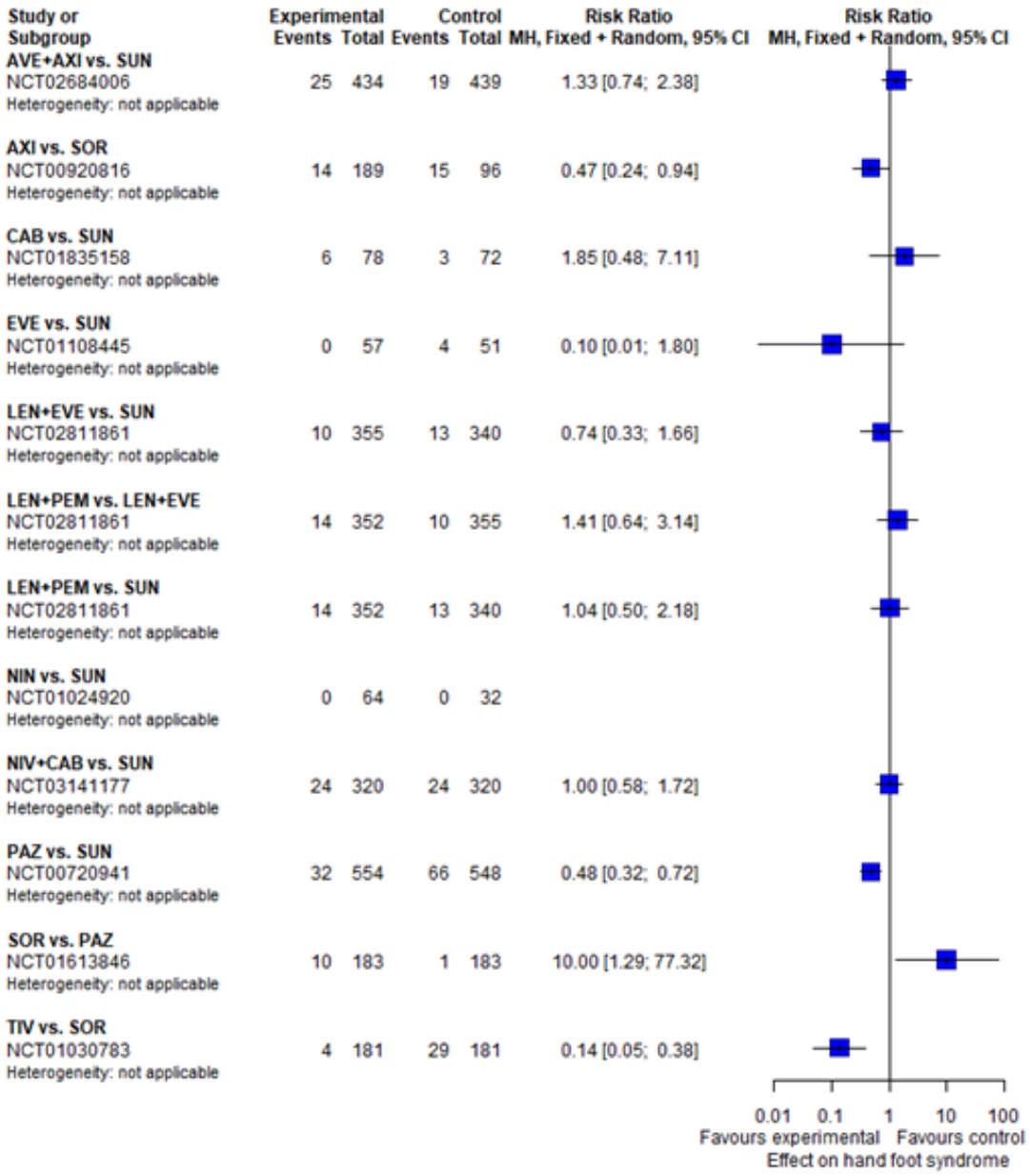
17. Net heat plot for all-cause AEs (all risk groups combined): Figure 79

Figure 79. Net heat plot for all-cause AEs (all risk groups combined)



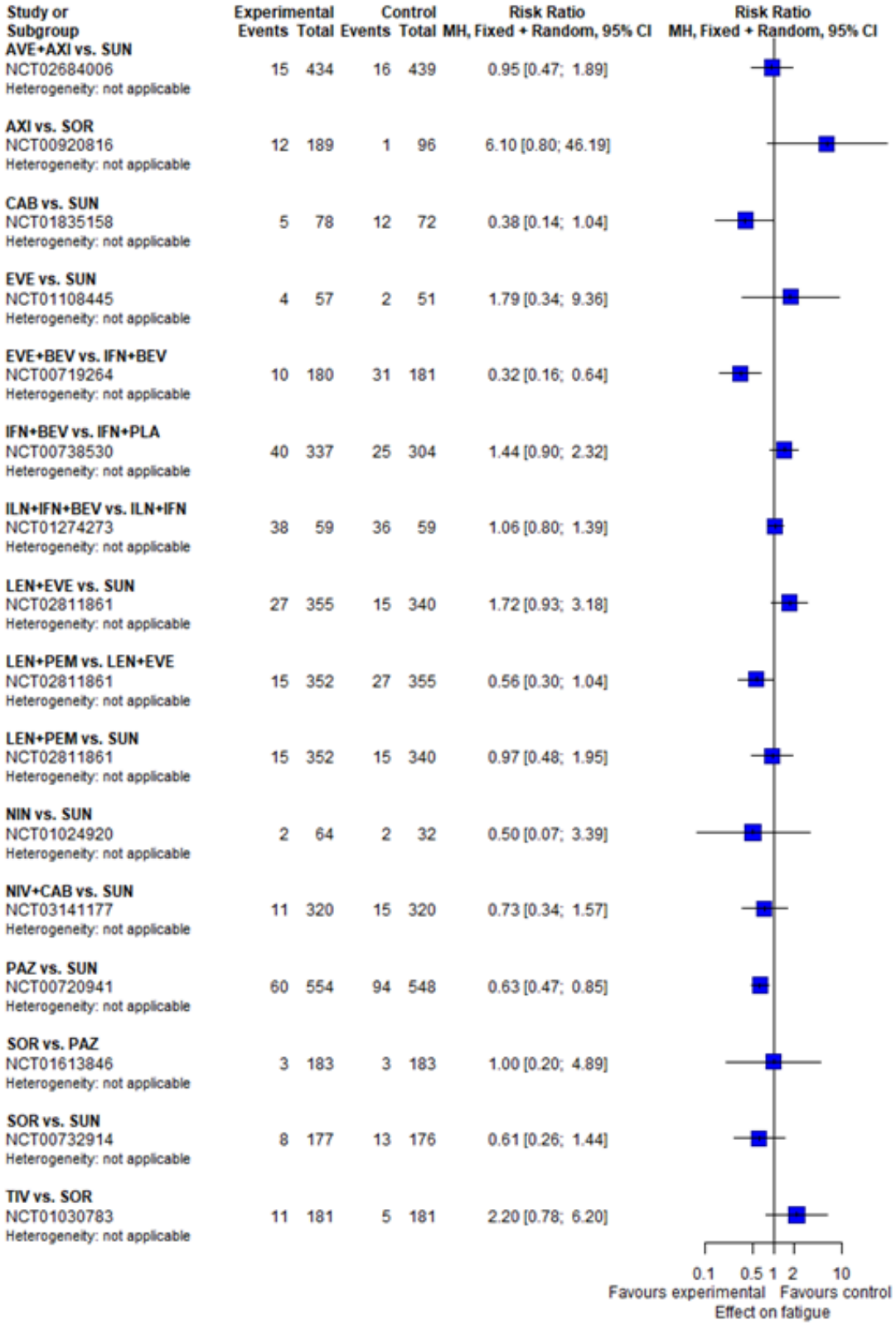
18. Pairwise comparison for AE hand-foot-syndrome (all risk groups combined): Figure 80

Figure 80. Pairwise comparison for AE hand-foot-syndrome (all risk groups combined)



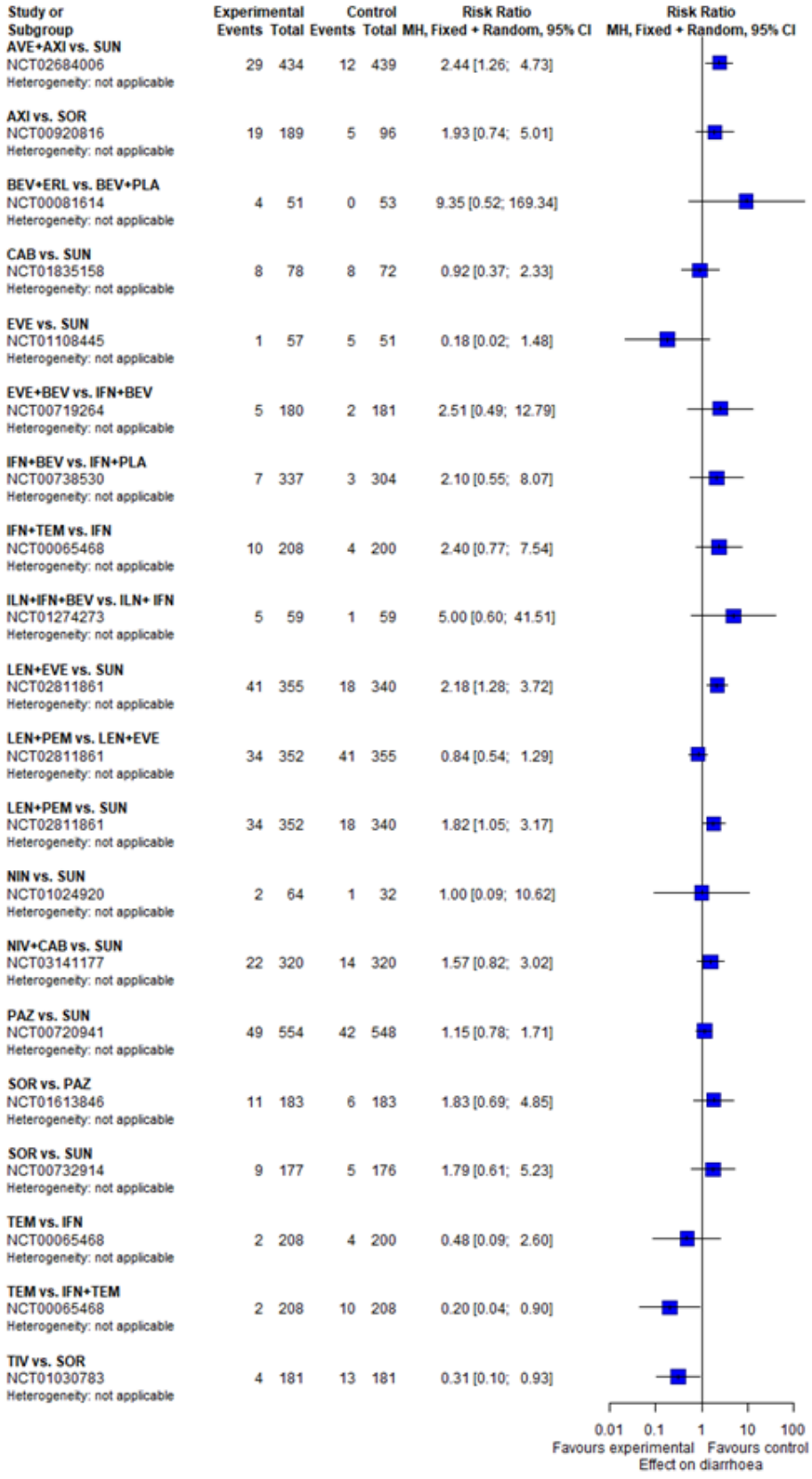
19. Pairwise comparison for AE fatigue (all risk groups combined): Figure 81

Figure 81. Pairwise comparison for AE fatigue (all risk groups combined)



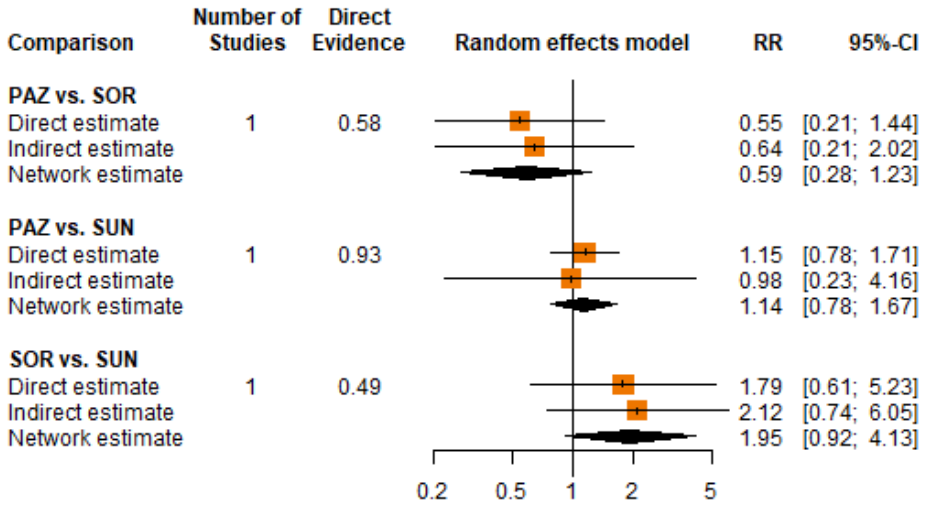
20. Pairwise comparison for AE diarrhoea (all risk groups combined): Figure 82

Figure 82. Pairwise comparison for AE diarrhoea (all risk groups combined)



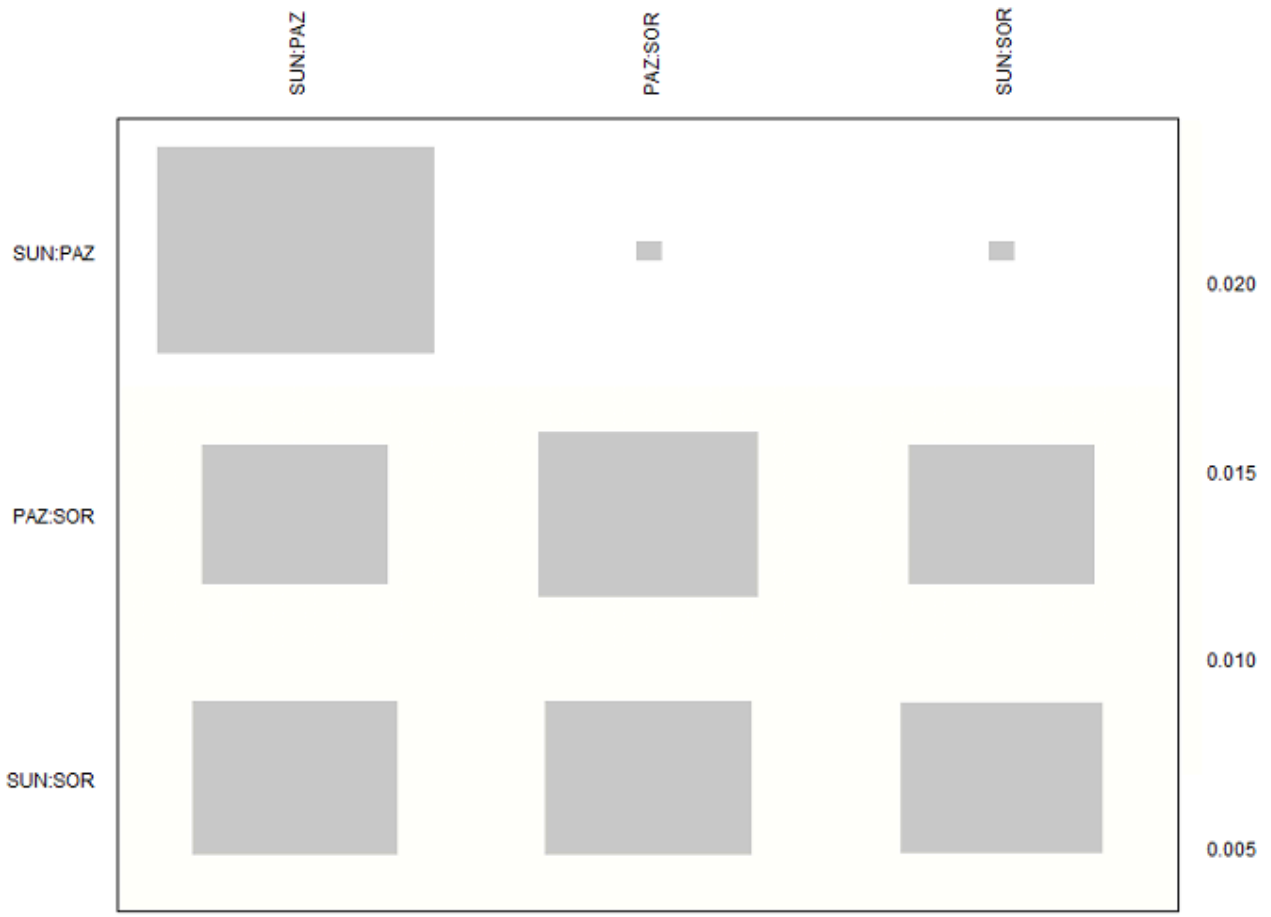
21.Forest plot of splitting direct and indirect evidence for AE diarrhoea (all risk groups combined): Figure 83

Figure 83. Forest plot of splitting direct and indirect evidence for AE diarrhoea (all risk groups combined)



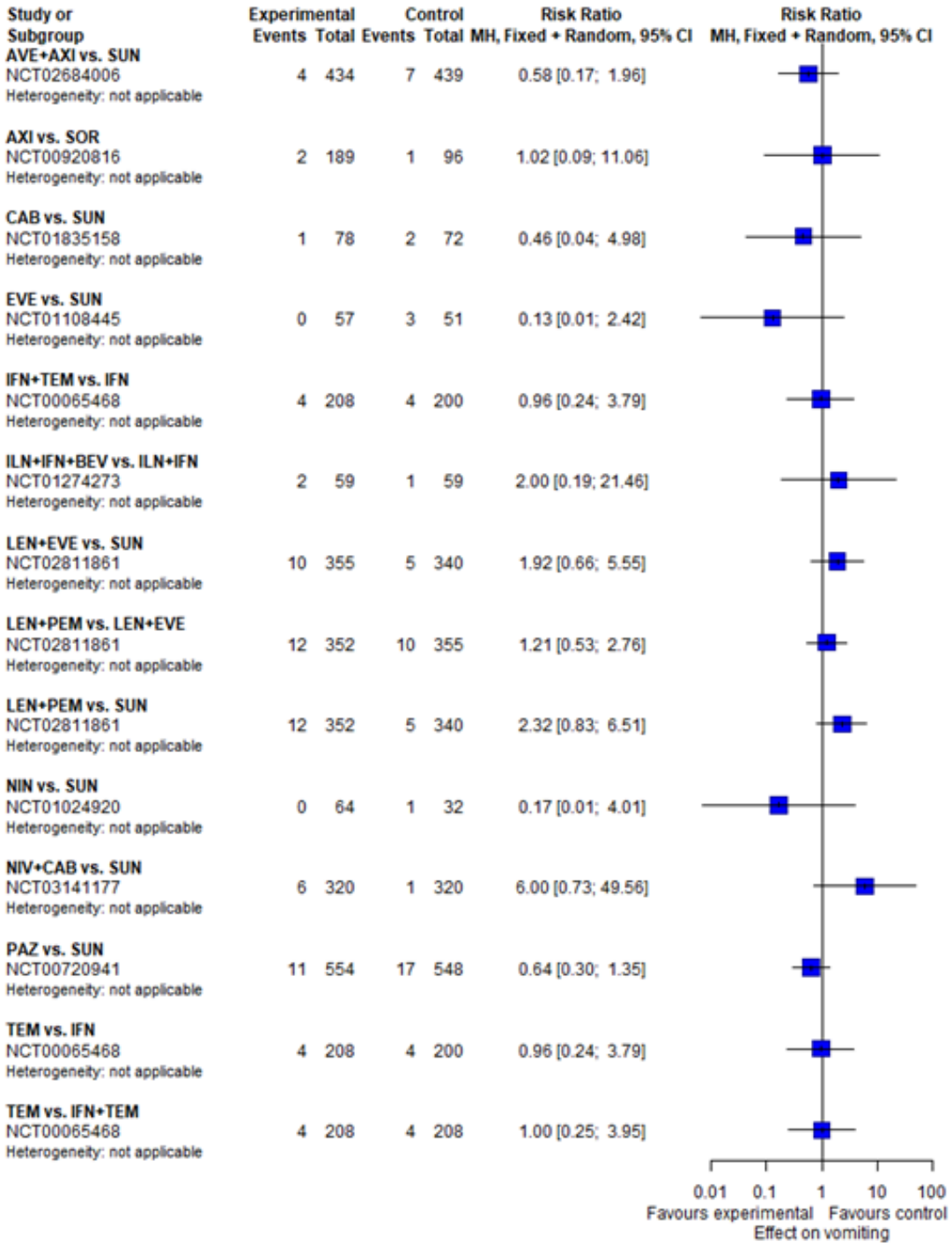
22. Net heat plot for AE diarrhoea (all risk groups combined): Figure 84

Figure 84. Net heat plot for AE diarrhoea (all risk groups combined)



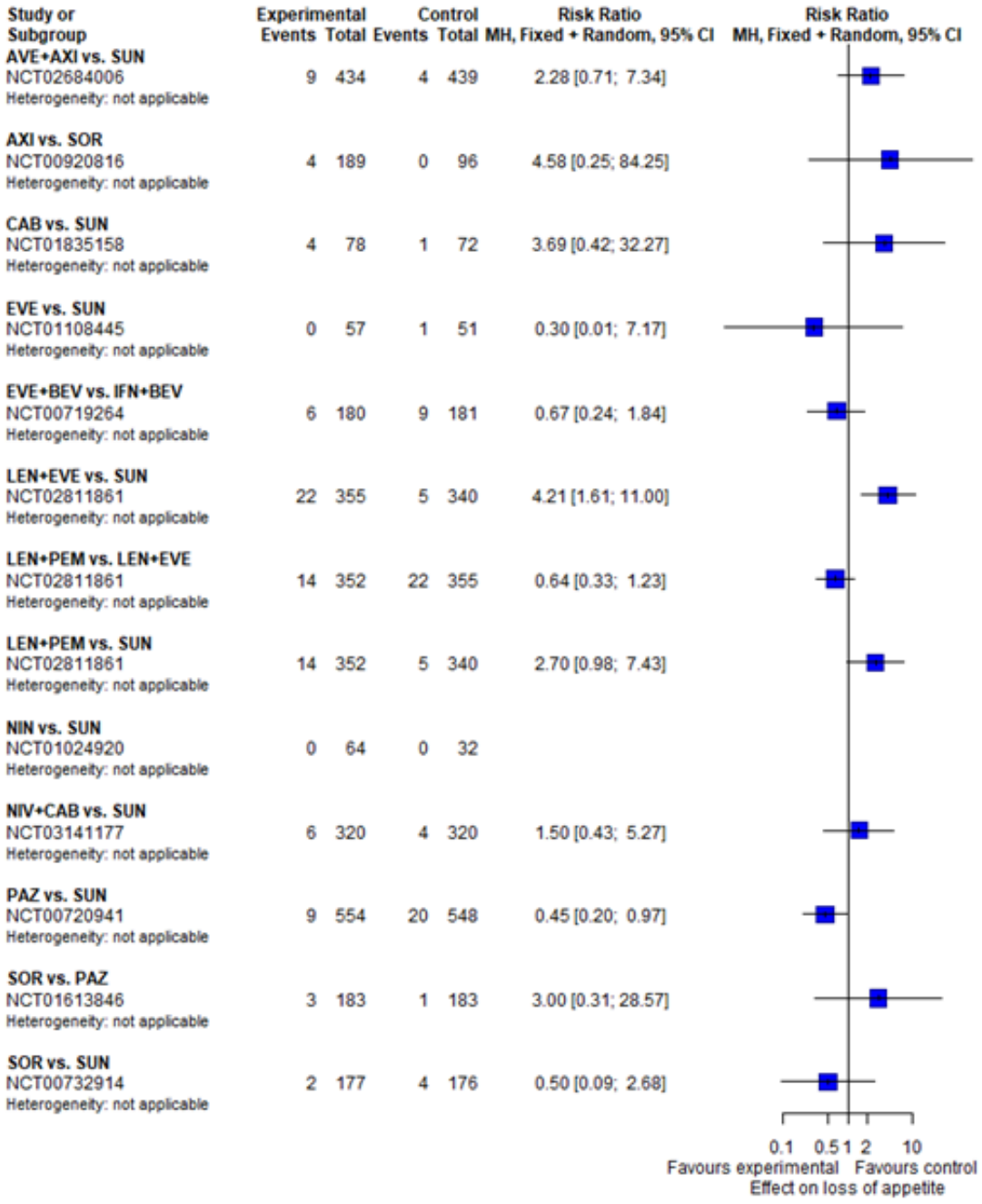
23. Pairwise comparison for AE vomiting (all risk groups combined): Figure 85

Figure 85. Pairwise comparison for AE vomiting (all risk groups combined)



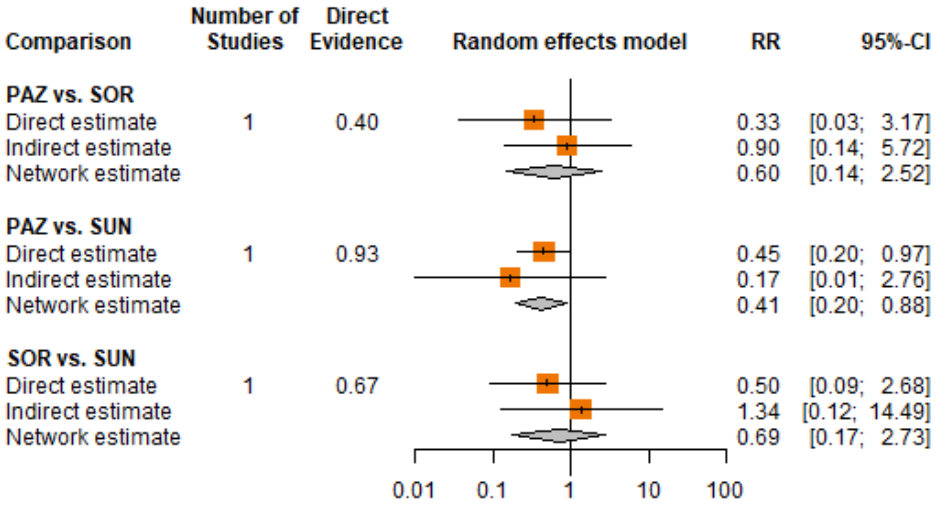
24. Pairwise comparison for AE loss of appetite (all risk groups combined): Figure 86

Figure 86. Pairwise comparison for AE loss of appetite (all risk groups combined)



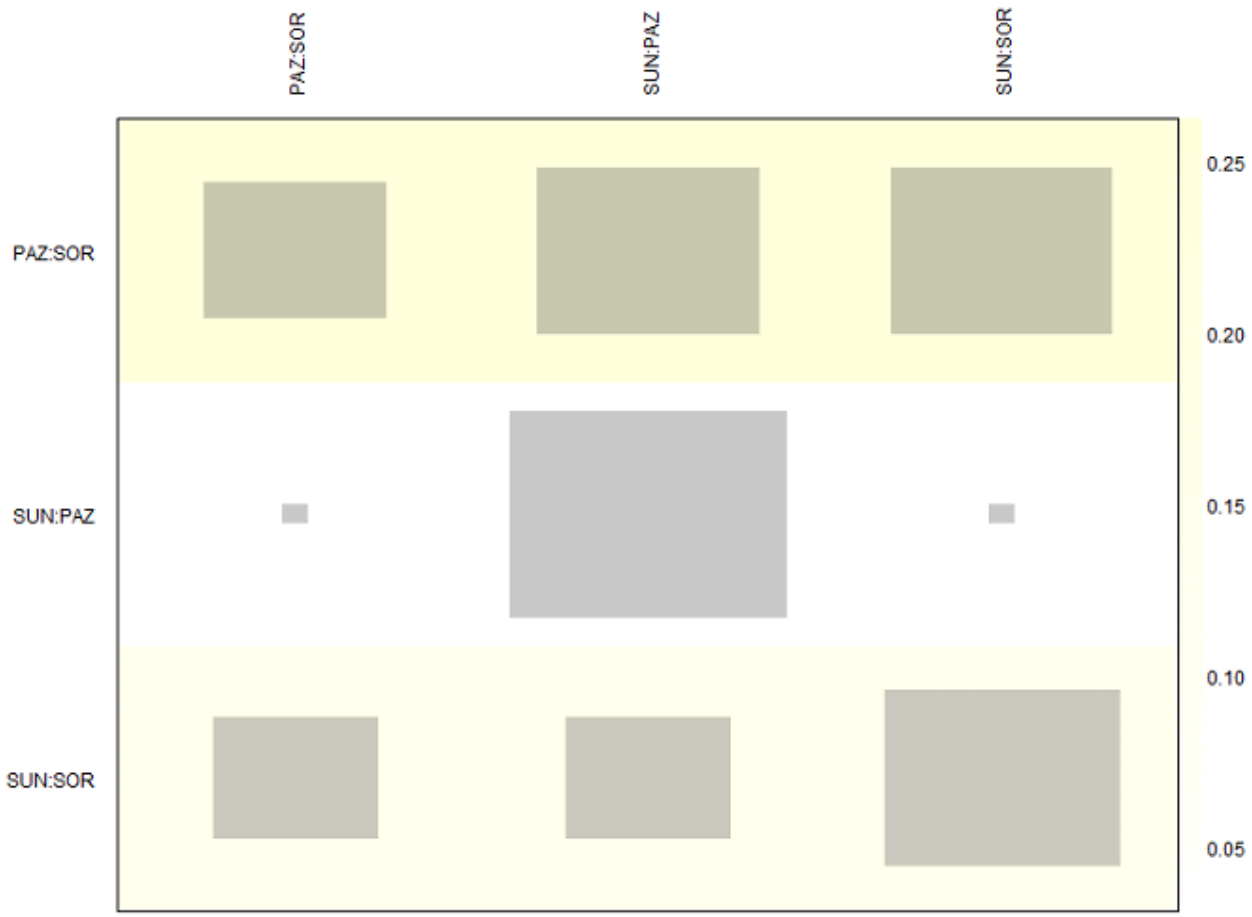
25. Forest plot of splitting direct and indirect evidence for AE loss of appetite (all risk groups combined): Figure 87

Figure 87. Forest plot of splitting direct and indirect evidencefor AE loss of appetite (all risk groups combined)



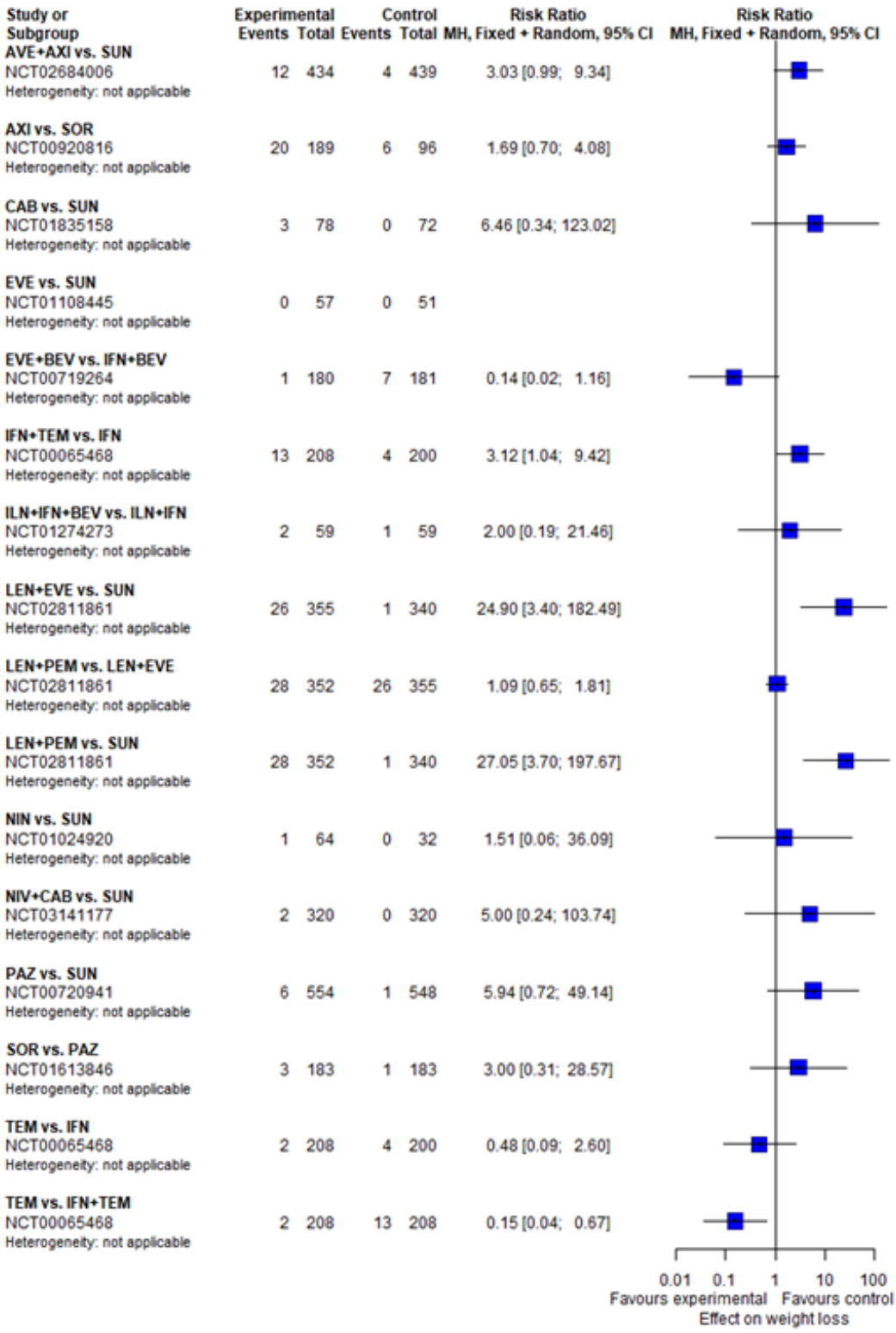
26. Net heat plot for AE loss of appetite (all risk groups combined): Figure 88

Figure 88. Net heat plot for AE loss of appetite (all risk groups combined)



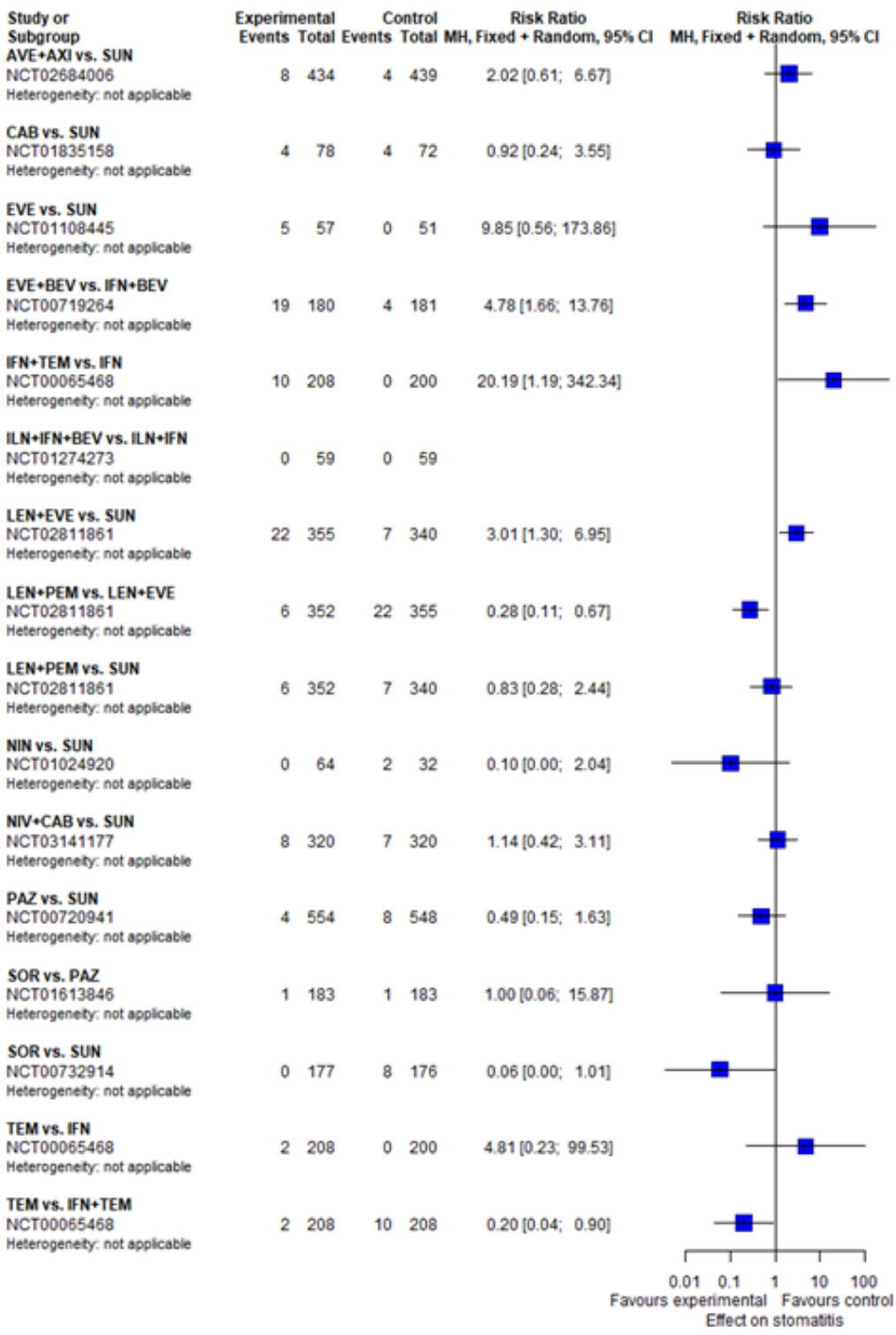
27. Pairwise comparison for AE weight loss (all risk groups combined): Figure 89

Figure 89. Pairwise comparison for AE weight loss (all risk groups combined)



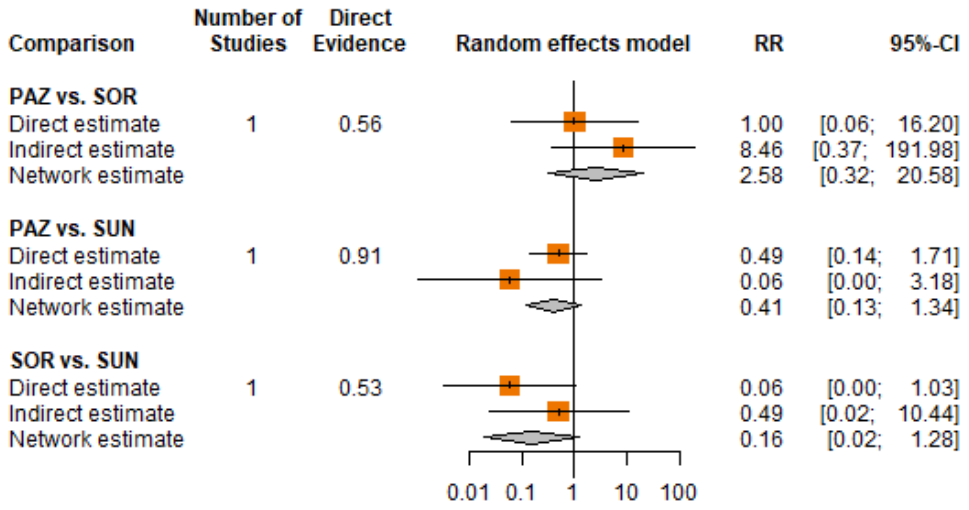
28. Pairwise comparison for AE stomatitis (all risk groups combined): Figure 90

Figure 90. Pairwise comparison for AE stomatitis (all risk groups combined)



29. Forest plot of splitting direct and indirect evidence for AE stomatitis (all risk groups combined): Figure 91

Figure 91. Forest plot of splitting direct and indirect evidence for AE stomatitis (all risk groups combined)

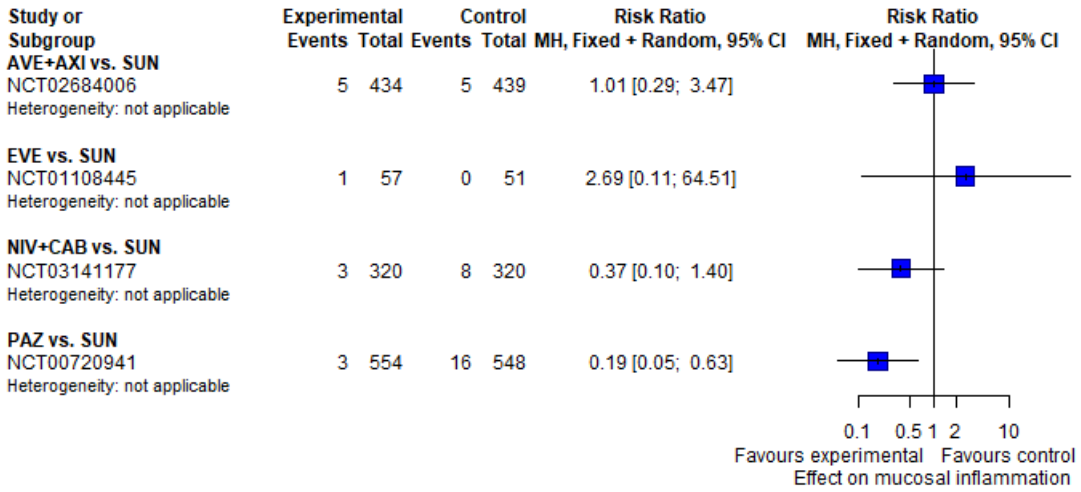


30. Net heat plot for AE stomatitis (all risk groups combined): Figure 92

Figure 92. Net heat plot for AE stomatitis (all risk groups combined)

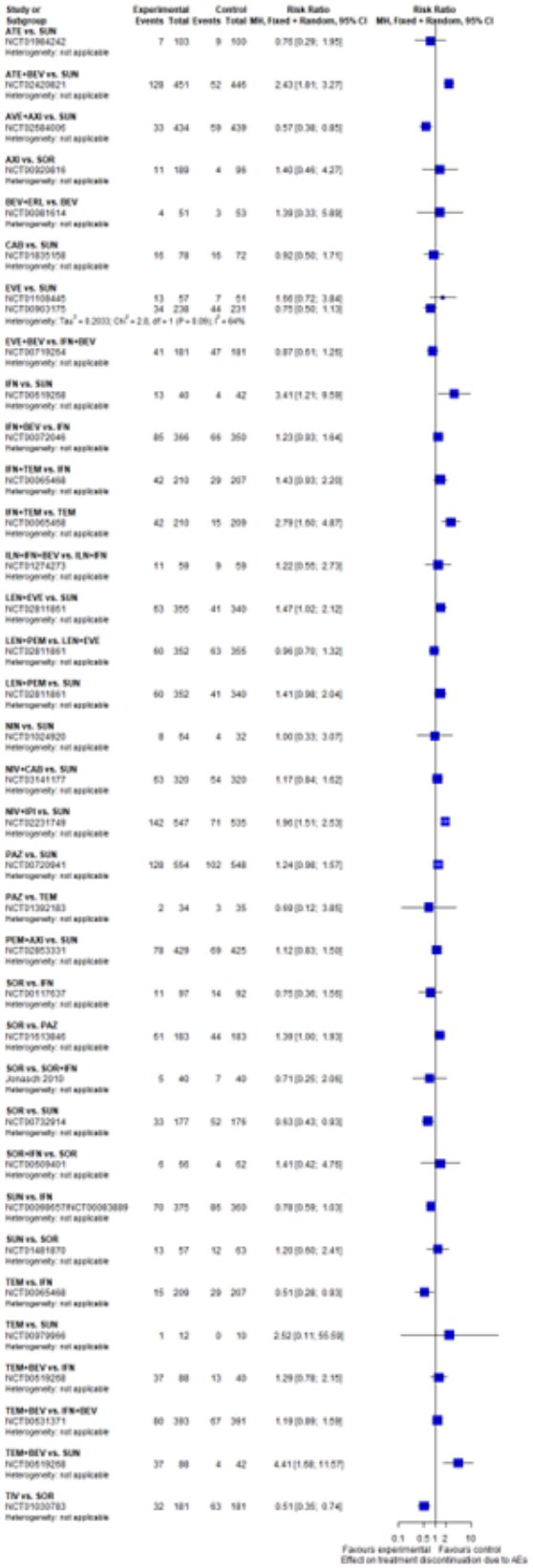
31. Pairwise comparison for AE mucosal inflammation (all risk groups combined): Figure 93

Figure 93. Pairwise comparison for AE mucosal inflammation (all risk groups combined)



32. Pairwise comparison for Number of participants who discontinued treatment due to an AE (all risk groups combined): Figure 94

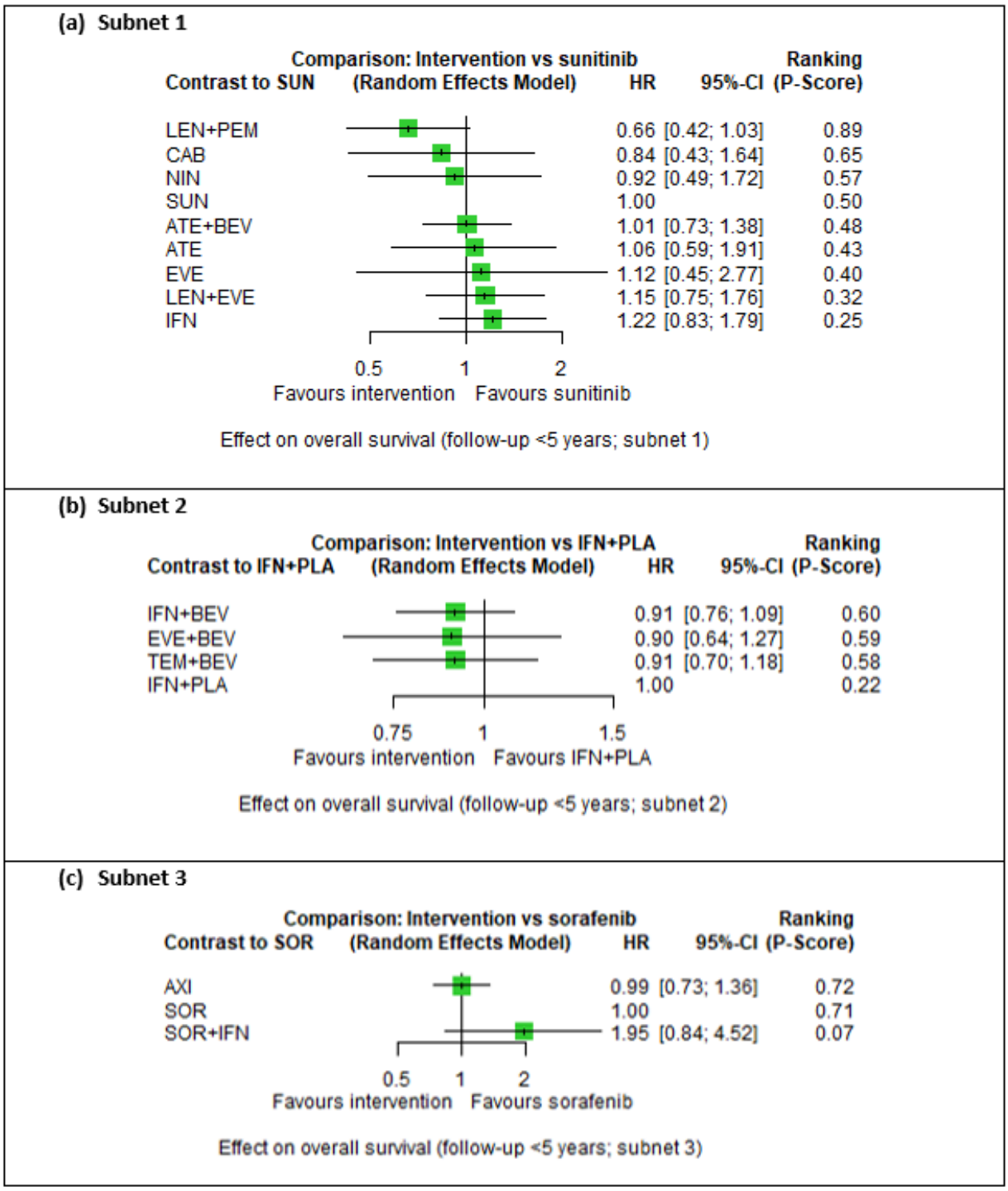
Figure 94. Pairwise comparison for the outcome Number of participants who discontinued treatment due to an AE (combined risk groups). Any two treatments are connected by a line when there is at least one trial comparing the two treatments.



Appendix 16. Additional figures (subgroup and sensitivity analyses)

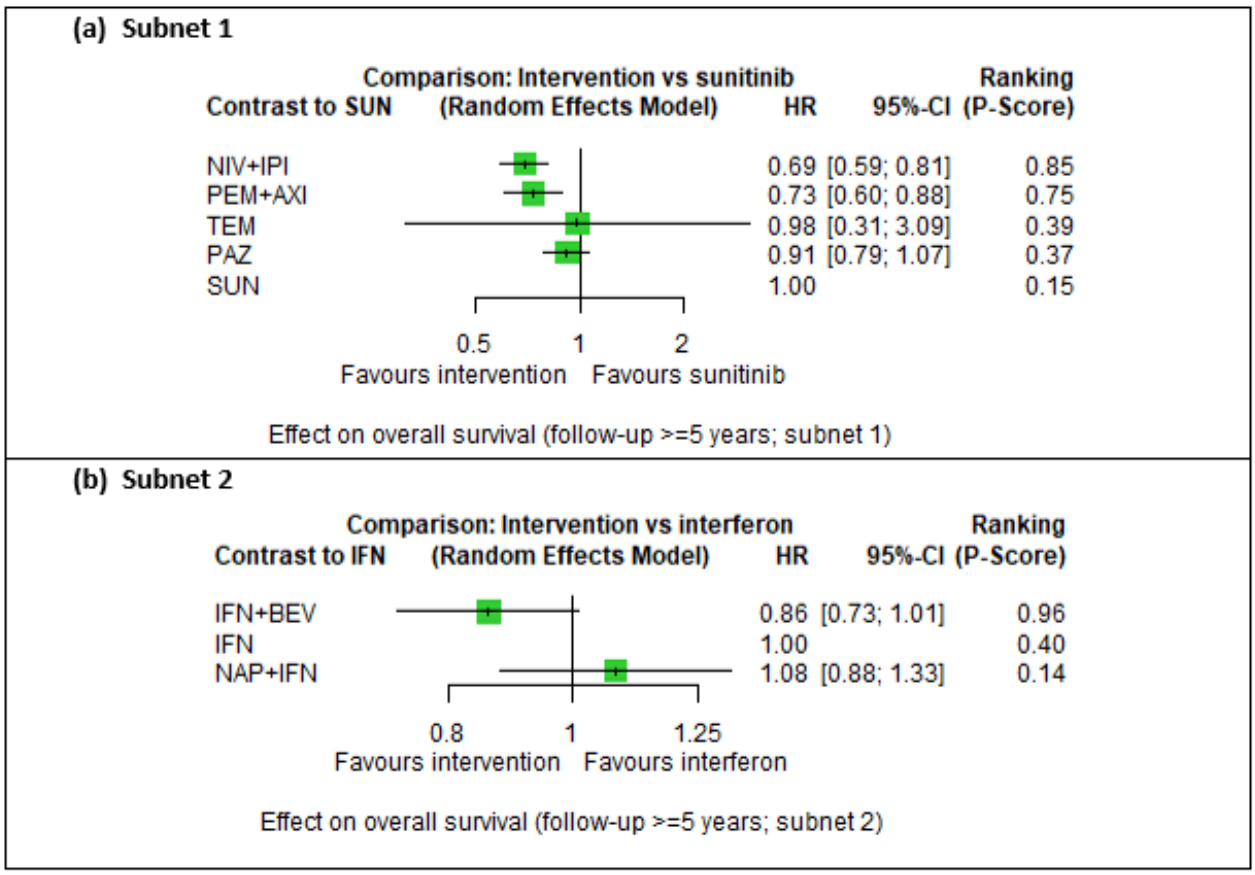
1. Follow-up time <5 years for OS (all risk groups combined): Figure 95

Figure 95. Follow-up time <5 years for OS (all risk groups combined)



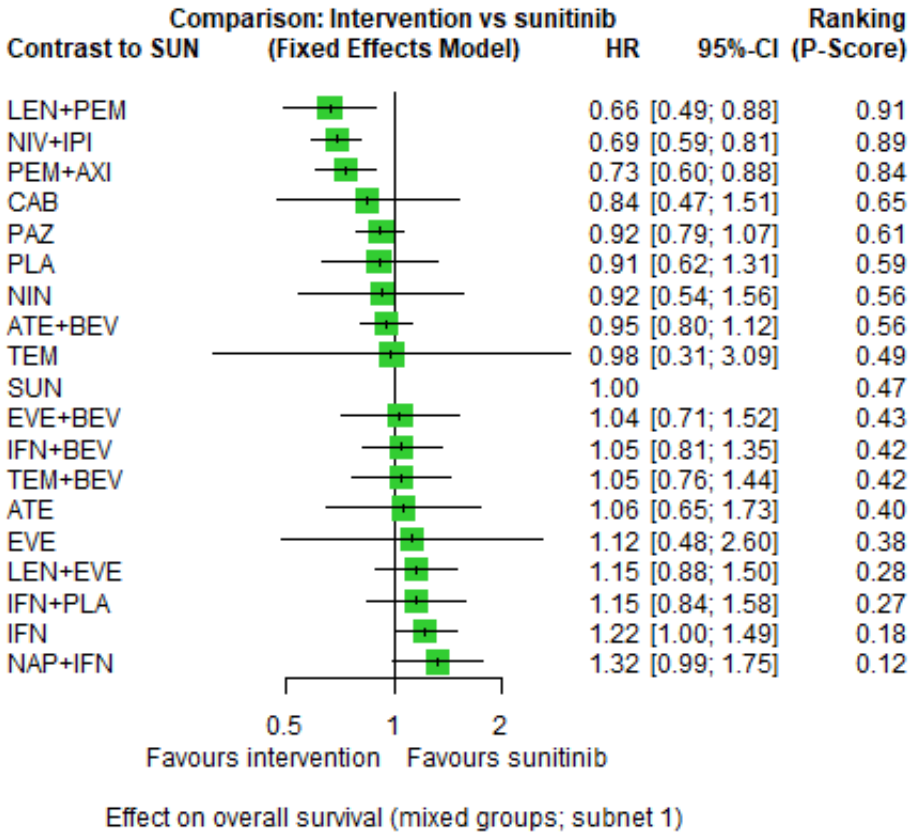
2. Follow-up time 5 years or more for OS (all risk groups combined): Figure 96

Figure 96. Follow-up time 5 years or more for OS (all risk groups combined)



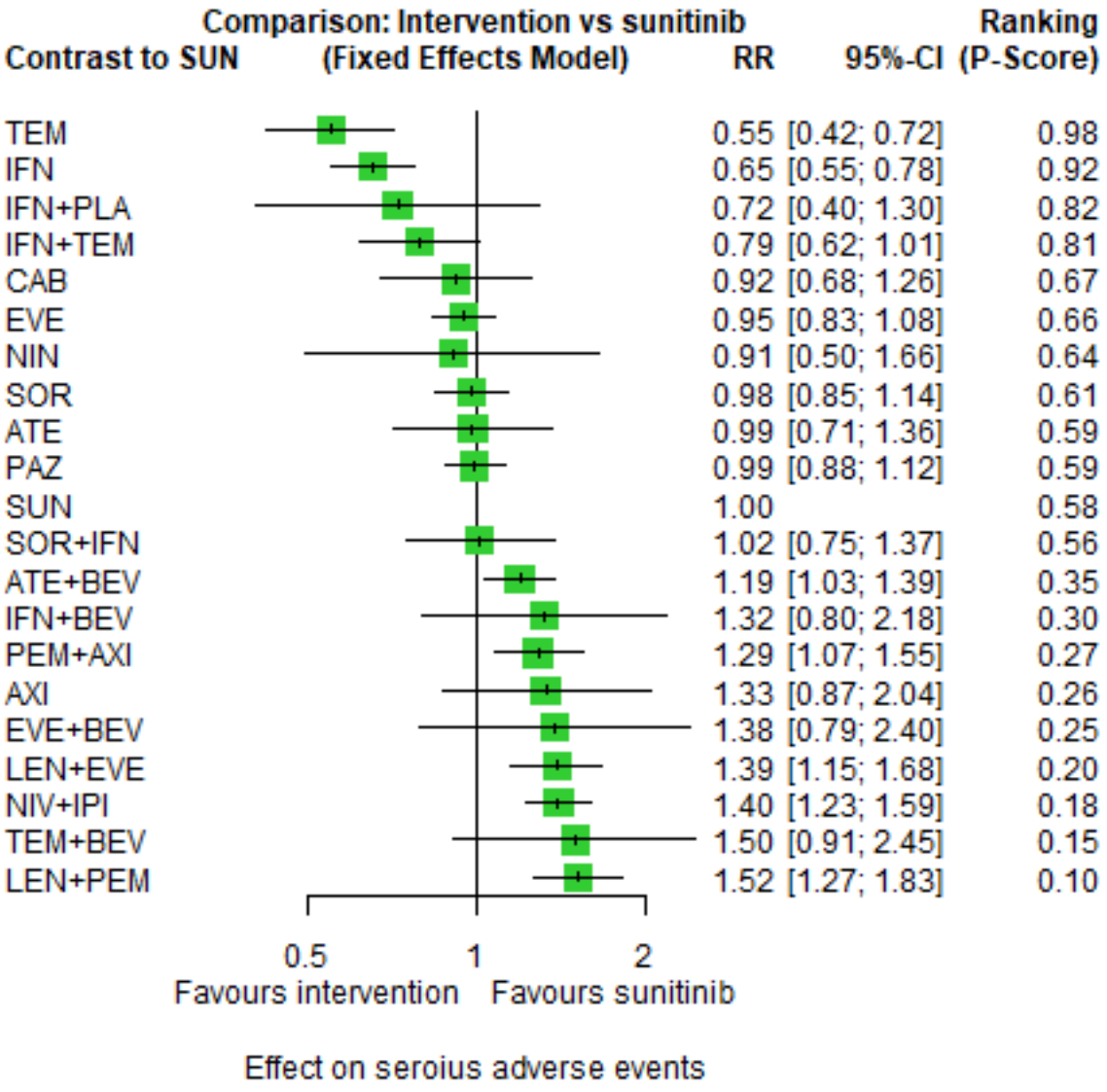
3. Fixed-effect model for OS (all risk groups combined): Figure 97

Figure 97. Fixed-effect model for OS (all risk groups combined)



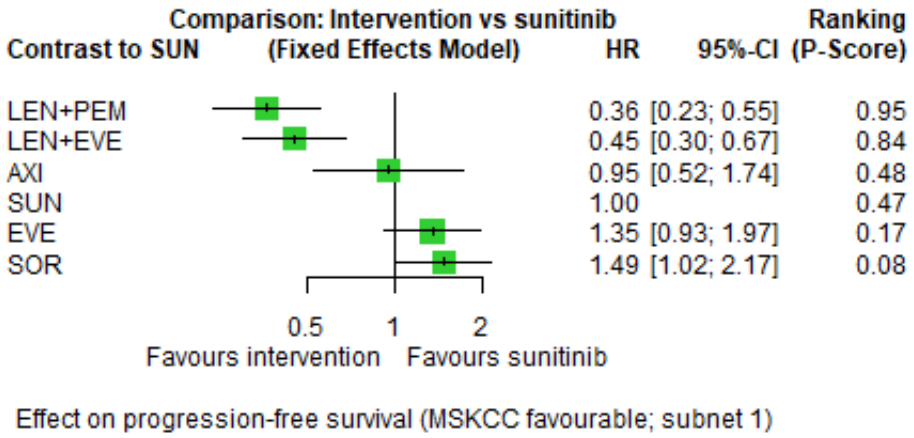
4. Fixed-effect model for SAE (all risk groups combined): Figure 98

Figure 98. Fixed-effect model for SAE (all risk groups combined)



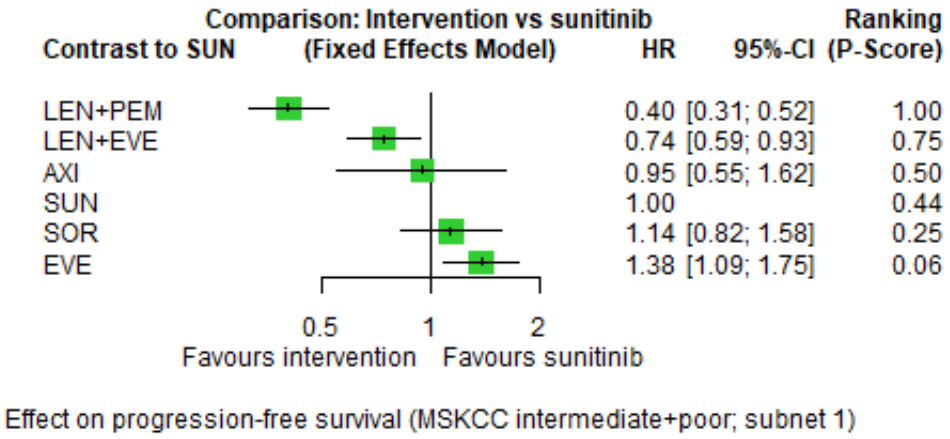
5. Fixed-effect model for PFS (MSKCC favourable - subnet 1): Figure 99

Figure 99. Fixed-effect model for PFS (MSKCC favourable - subnet 1)



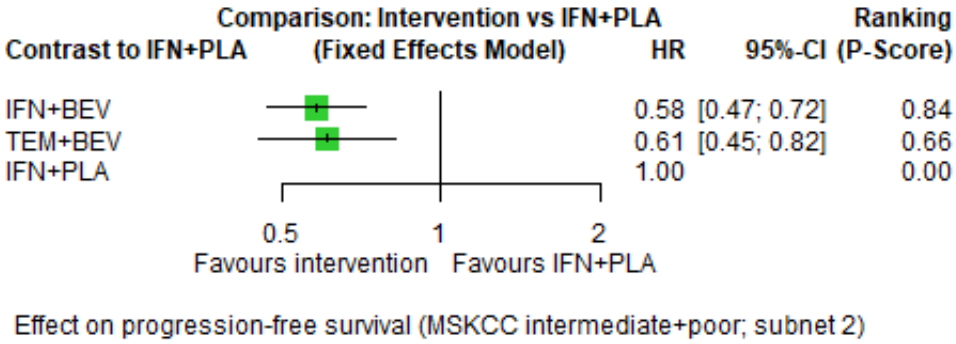
6. Fixed-effect model for PFS (MSKCC intermediate, poor - subnet 1): Figure 100

Figure 100. Fixed-effect model for PFS (MSKCC intermediate, poor - subnet 1)



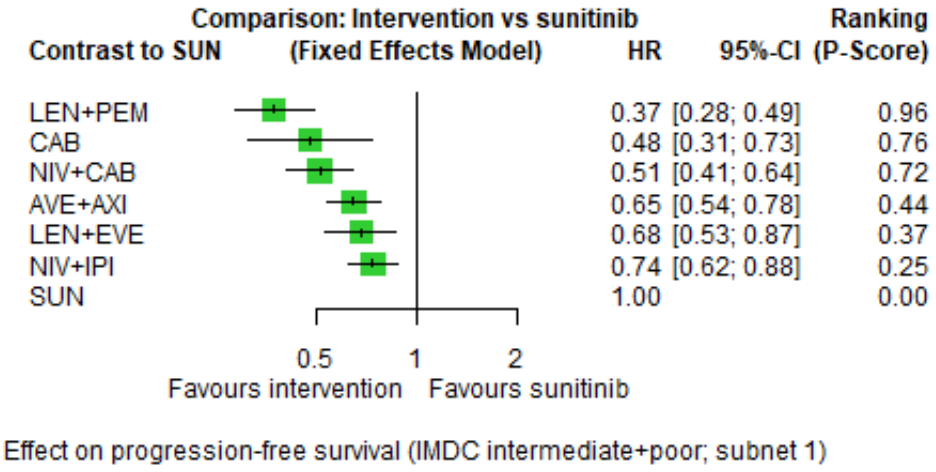
7. Fixed-effect model for PFS (MSKCC intermediate, poor - subnet 2): Figure 101

Figure 101. Fixed-effect model for PFS (MSKCC intermediate, poor - subnet 2)



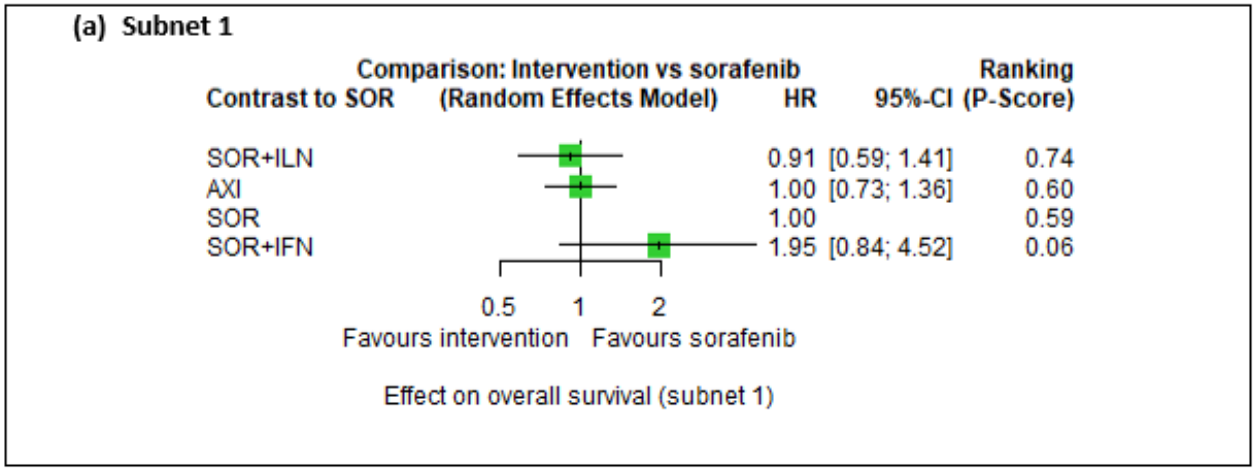
8. Fixed-effect model for PFS (IMDC intermediate, poor - subnet 1): Figure 102

Figure 102. Fixed-effect model for PFS (IMDC intermediate, poor - subnet 1)



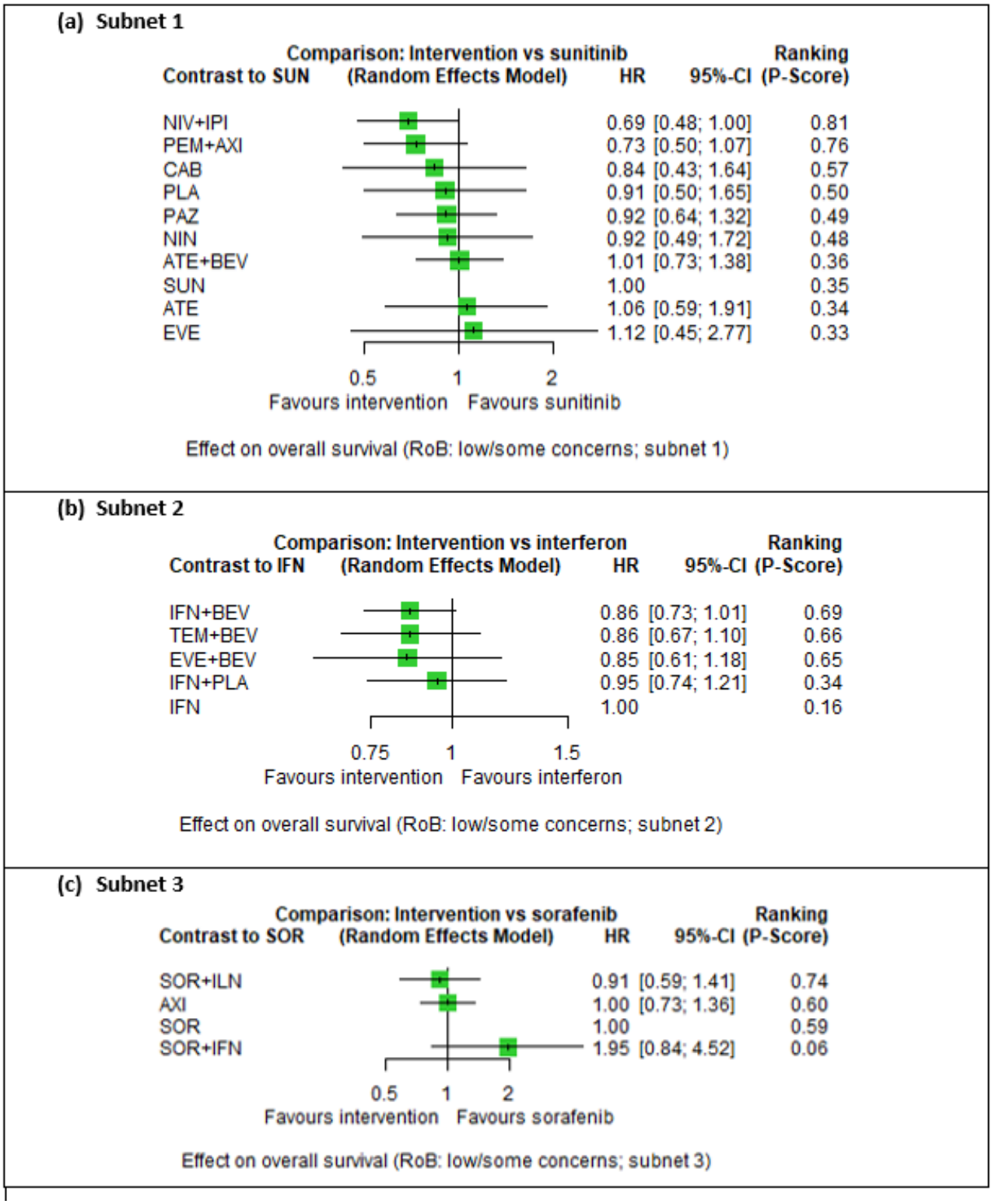
9. Validation of the PH assumption for the outcome OS (all risk groups combined): Figure 103

Figure 103. Validation of the PH assumption for the outcome OS (all risk groups combined)



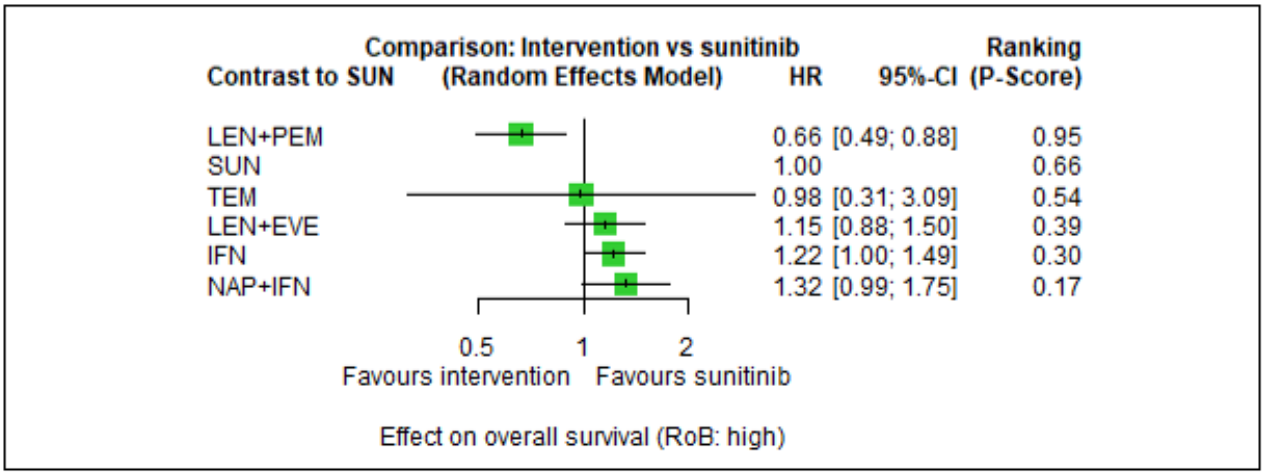
10. Sensitivity analysis for the outcome OS (all risk groups combined) with trials at 'low risk of bias' or 'some concerns': Figure 104

Figure 104. Sensitivity analysis for the outcome OS (all risk groups combined) with trials at 'low risk of bias' or 'some concerns'



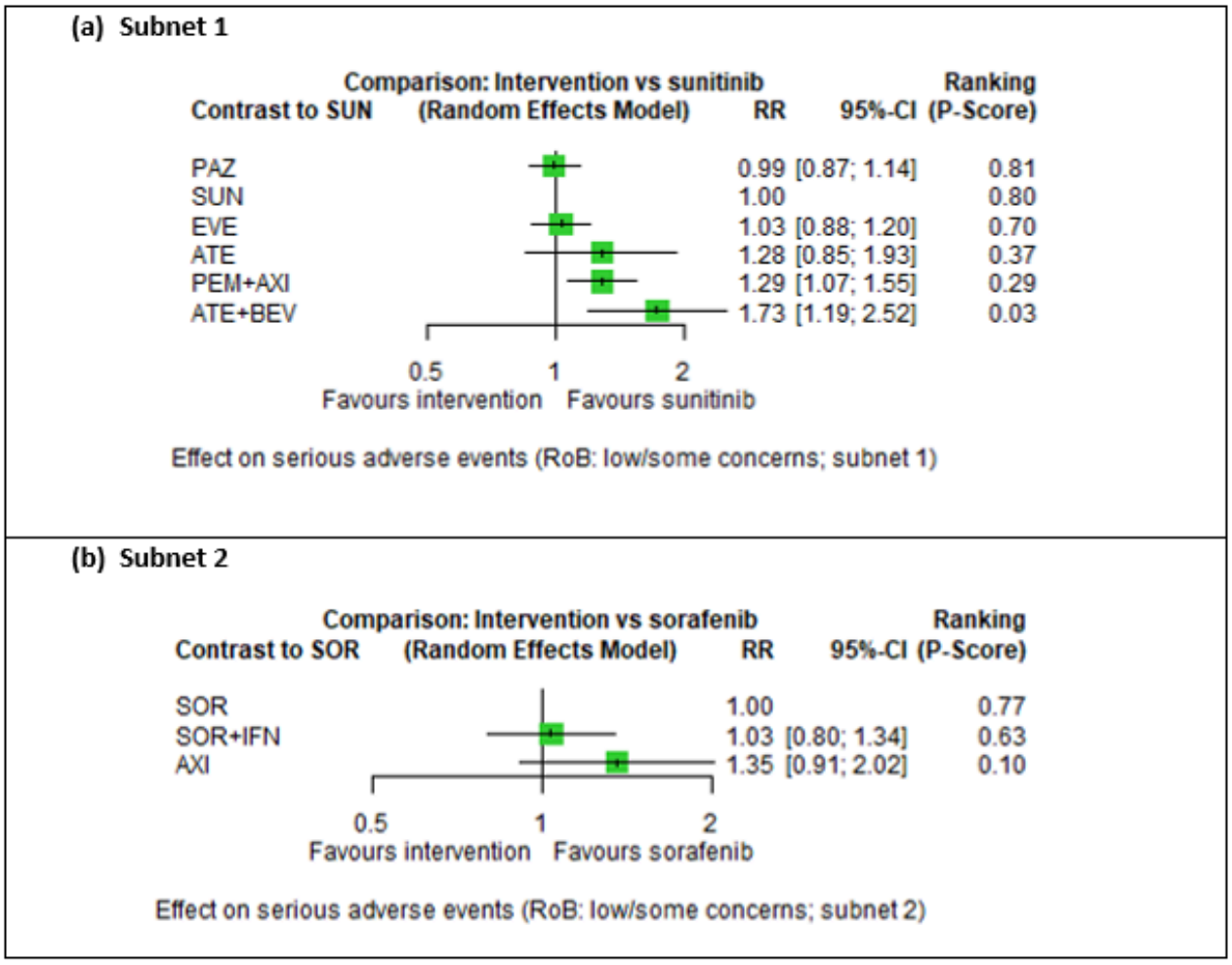
11. Sensitivity analysis for the outcome OS (all risk groups combined) with trials at 'high risk of bias': Figure 105

Figure 105. Sensitivity analysis for the outcome OS (all risk groups combined) with trials at 'high risk of bias'



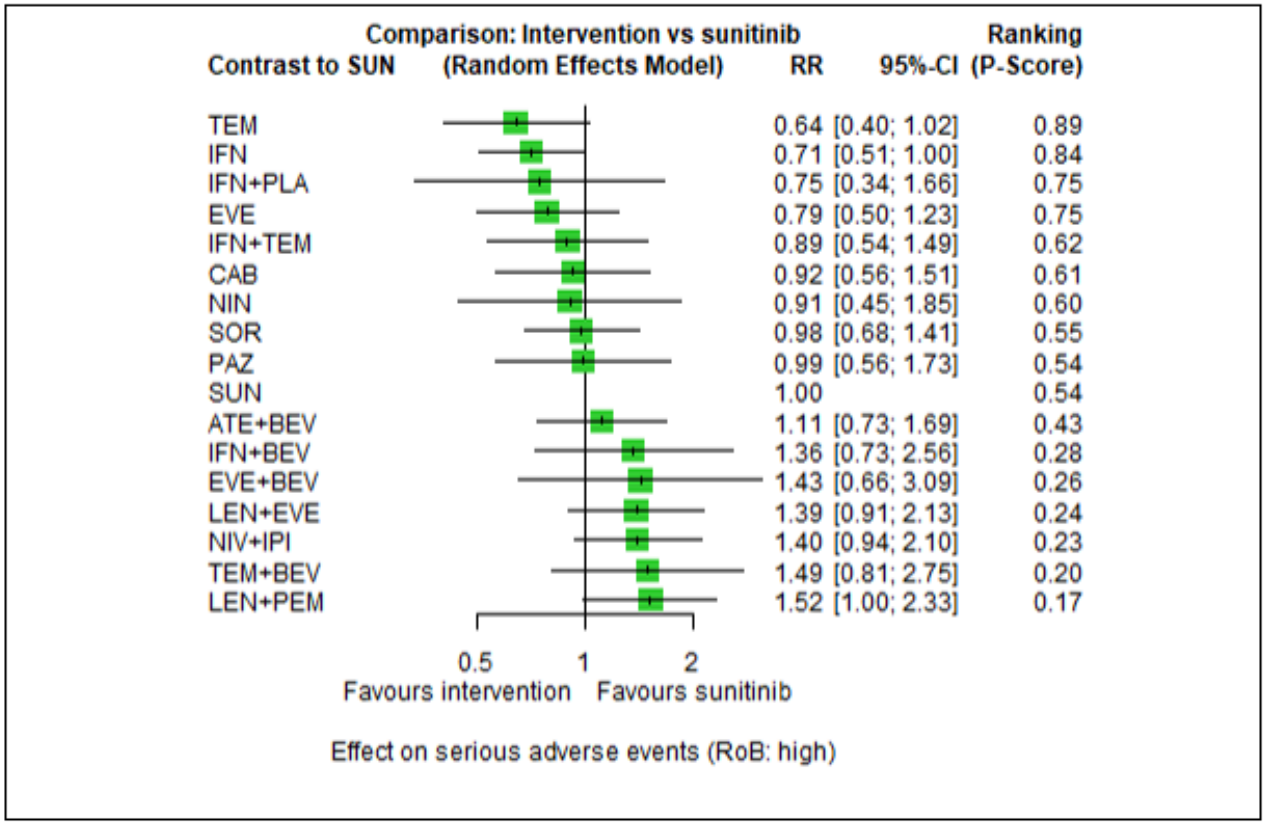
12. Sensitivity analysis for the outcome SAE (all risk groups combined) with trials at 'low risk of bias' or 'some concerns': Figure 106

Figure 106. Sensitivity analysis for the outcome SAE (all risk groups combined) with trials at 'low risk of bias' or 'some concerns'



13. Sensitivity analysis for the outcome SAE (all risk groups combined) with trials at 'high risk of bias': Figure 107

Figure 107. Sensitivity analysis for the outcome SAE (all risk groups combined) with trials at 'high risk of bias'



HISTORY

Protocol first published: Issue 12, 2020

CONTRIBUTIONS OF AUTHORS

- AAa: review development, including screening and study selection, data extraction, risk of bias assessment, GRADE assessment, interpretation and writing of results, developing the draft
- BB: review development, including screening and study selection, data extraction, risk of bias assessment, assisted in writing the draft (i.e. description of studies and bias assessments)
- AAb: conducted the statistical analyses, proofread the draft
- IM: designed the search strategies and conducted all searches, proofread the draft
- VP: screening and study selection, provided methodological expertise on NMA, assisted in the GRADE assessment, proofread the draft
- ET: data extraction, risk of bias assessment, proofread the draft
- CH: searching for CSRs, data extraction for the characteristics of included studies, proofread the draft
- ND: data extraction, risk of bias assessment, proofread the draft
- MG: screening and study selection, searching for CSRs, provided methodological expertise, proofread the draft
- PM: provided clinical expertise, proofread the draft
- PD: provided methodological and clinical expertise
- AH: provided clinical expertise
- NS: provided methodological and content expertise, proofread the draft

DECLARATIONS OF INTEREST

- AAa: The grant by the German Federal Ministry of Education and Research does not lead to a conflict of interest. She is editor at Cochrane, but was not involved in the editorial process for this review.
- BB: none known.
- AAb: The grant by the German Federal Ministry of Education and Research does not lead to a conflict of interest. She is editor at Cochrane, but was not involved in the editorial process for this review.
- IM: none known. She is information specialist for Cochrane Haematology, but was not involved in the editorial process for this review.
- VP: none known.
- ET: none known.
- CH: none known.
- ND: none known.
- PM: none known.
- PD: none known; he is Co-ordinating Editor of Cochrane Urology, but was not involved in the editorial process for this review.
- MG: none known.
- AH: speaker honorary and research grant for renal cell carcinoma by BMS; however, this does not lead to a conflict of interest.
- NS: none known; she is Co-ordinating Editor of Cochrane Haematology, but was not involved in the editorial process for this review.

SOURCES OF SUPPORT

Internal sources

- University Hospital Cologne, Department I of Internal Medicine, Center for Integrated Oncology Aachen Bonn Cologne Duesseldorf, Germany, Germany

  Provision of the offices, including technical equipment

External sources

- Bundesministerium für Bildung und Forschung (BMBF) (German Federal Ministry of of Education and Research), Germany

  Grant number: 01KG1901

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

We made some changes to the methods that were pre-specified in the protocol for this review (Goldkuhle 2020).

Criteria for considering studies in this review

Types of interventions

In the protocol for this review, we pre-specified that we would create two networks with interventions, one for the favourable risk group, and one combined for the intermediate and poor risk groups. During the conduct of this review we found that trials reported data for the different risk groups according to different criteria, i.e. risk groups either according to the IMDC or the MSKCC criteria. In some trials, data were even reported for both criteria. Together with the clinical experts on this review, we decided to analyse data not only separately for the different risk groups (i.e. favourable vs. intermediate and poor risk), but also separately by the different criteria (i.e. IMDC or MSKCC), as these should be looked at separately. In addition, as many trials reported both data for the separate risk groups and all risk groups combined (which we called the 'total trial population'), we also decided to conduct one overall analysis for each outcome with all risk groups combined. This was particularly the case for the safety outcomes, as no subgroup data according to risk group were reported for these. For the outcomes OS and PFS, we were able to extract data for both the total trial population (i.e., all risk groups combined) and separately according to the IMDC favourable/intermediate/poor risk groups and the MSKCC favourable/intermediate/poor risk groups.

Types of outcome measures

In the protocol we stated that we will extract all individual AEs reported in the trials as well as the frequency of the specific AEs. However, this was not feasible for this review, but will be considered in an update of this review.

We planned to analyse the outcome TFST as a time-to-event outcome. However, none of the included trials reported this outcome as a time-to-event outcome. We planned to analyse such outcomes as dichotomous outcomes if time-to-event analyses were not possible. Hence, we extracted the number of participants who received subsequent anticancer therapy after discontinuation of trial therapy. However, reporting between trials was heterogenous, for example in terms of the definition of this outcome and the timing of reporting (participants received different lengths of therapy, and it was unclear at which time point therapy stopped), so we refrained from pooling data and reported the results narratively in tabular form instead. Hence, we also reported this outcome narratively in the 'Summary of Findings' table.

Although we extracted and estimated some data for the outcome QoL, pooling data were not feasible. Even after combining all available data for the different time points, only one time point (long-term, 1 year after initiation of treatment) would have been feasible for network meta-analyses. However, for this time period ("long-term"), the individual time points also varied between two and four years, so pooling these different time points in a combined "long-term" analysis would not have produced meaningful results. For the other time points, comparisons including SUN were sparse (only two or three trials per time point) and only when combining different scales; so again, an analysis would have not delivered meaningful results. Conducting pairwise meta-analyses was also not possible because each comparison was reported by one trial only. Hence, we decided to report results for this outcome narratively in this review.

Data collection and analysis

Assessment of publication bias

Creating a 'comparison-adjusted' funnel plot was not feasible for this review because the individual effects are centred on the 1 to maintain the same reference line. That is, for each comparison, the pooled effect is used as the reference. This resulted in the effect being centred on the 1 for all comparisons consisting of only one trial. Since we often had one trial for each comparison in our analyses, there are barely any trials outside the reference line, so the funnel plot could not provide meaningful results.

Data synthesis

Certainty of the evidence

In the protocol we stated that we will use GRADEpro GDT for the GRADE assessment. However, as the software cannot be used for network meta-analyses, and because we did not conduct any traditional pairwise meta-analyses, we were not able to use this software. Instead, we created our own table in Excel where we conducted and recorded our GRADE assessments (judgements and explanations).

Summary of findings table

In the protocol we had stated that we will create two networks (one for the favourable risk group and one for the intermediate and poor risk groups) and thus present one summary of findings (SoF) table each. However, as we now have an additional network per outcome for the combined risk groups, we provided three SoF tables: one for the total trial population (i.e. all risk groups combined); one for the favourable risk group (results presented separately for IMDC and MSKCC); and one for the intermediate and poor risk groups (results presented separately for IMDC and MSKCC).

Furthermore, we stated that we will use CINeMA to present the SoF table. However, we decided to create a SoF table manually in a format applicable to the PICO of this review with network meta-analyses. Lastly, we did not calculate NNT/NNH as planned in the protocol of this review, for simplicity reasons, as we already present absolute effect numbers for every network treatment estimate in the SoF tables.

Living systematic review considerations

At protocol stage, we proposed an approach for updating this review. However, due to restricted funding there are currently no plans for an update.

We had proposed the following approach.

Following the approaches proposed by Cochrane, whenever new evidence (i.e. studies, data, or other information) relevant to the review is identified, we will extract the data and assess risk of bias as appropriate. We will wait until the accumulating evidence changes one or more of the following components of the review before incorporating it and republishing the review.

The findings for one or more of the primary outcomes change either in the size of the point estimate or the direction of effects, or both.

- The credibility (e.g. GRADE rating) of one or more primary outcomes.
- New settings, populations, interventions, comparisons, or outcomes are studied.

Furthermore, following these suggestions, we will not use formal sequential meta-analysis approaches for updated (network) meta-analyses.

In order to inform our readers on any changes in the review and its conclusions, including the search results and all additional evidence, we plan to update the status information of the review, for example when new searches are undertaken. Once we identify an additional trial or other substantial information with direct relevance to the review conclusions, we will republish the review with a new citation.

Subgroup analysis and investigation of heterogeneity

Subgroup analyses on administration routes and different dosages was not feasible because administration routes did not differ between trials for the individual drugs, and there were only little differences in the dosages for a few drugs administered in the trials. Particularly our main comparator, sunitinib, was administered the same way in all trials and the dosage was administered across trials. For more details see *Table 1. Interventions in the included trials* in Description of studies. Furthermore, subgroup analyses for a follow-up time of less than one year was also not possible as no trial had such a short follow-up time.

In addition, it was not possible to conduct the following pre-specified subgroup analyses for the outcomes OS, SAE and QoL, the main reasons being that either subgroup data were not available or because trials could not be grouped according to the pre-specified characteristics. Only for the outcome OS, we were indeed able to extract some subgroup data according to age, sex and prior nephrectomy. However, the analyses would have included no more than two or three trials, respectively, so no meaningful results would have been created.

- age (< 65 versus > 65), as all trials included participants over and under the age of 65, and insufficient subgroup data were reported.
- sex (male versus female), as all trials included both men and women, and insufficient subgroup data were reported.
- nephrectomy (yes versus no), as in all trials but one, most participants had previously received a nephrectomy (full or partial), or a nephrectomy was allowed based on the inclusion criteria and no further information was provided about how many participants actually had a prior nephrectomy.
- radiotherapy (yes versus no), because in most trials, most participants had received previous radiotherapy, or radiotherapy was "allowed" based on the inclusion criteria and no further information was provided about how many participants actually had previously received radiotherapy.
- histology type (clear cell type, papillary type, sarcomatoid type), because in most trials, only participants with clear cell RCC were included, followed by trials in which most participants had clear cell RCC. Only in three trials it was clearly indicated that mostly participants with non-clear cell (papillary type) RCC were included.
- site of metastases (lung, bone, liver), as most trials reported several metastatic sites, so we could not group these trials. Only few trials reported only one metastatic site, and few trials included advanced metastatic RCC but without reporting the sites of the metastases.

Sensitivity analysis

In addition to our primary outcomes (OS, SAE), we also conducted sensitivity analyses using the fixed-effect model for the outcome PFS. As we did not analyse data for the outcome QoL, there is no such sensitivity analysis for this outcome.

For time-to-event outcomes, we had planned to conduct sensitivity analyses to explore the robustness of findings in case we had to use variable techniques to reconstruct HR from primary trial reports. However, as reconstruction of the HR was either not necessary or not possible, such sensitivity analyses were not conducted.

We also planned to conduct sensitivity analyses on quality components (overall low risk of bias or some concerns versus overall high risk of bias). This was not possible for the outcome QoL as all trials were at high risk of bias for this outcome.

Sensitivity analyses for completed but not published trials was not possible because only one trial in the included trials for analyses was not published in a publication (except for a retrospective analysis of a subpopulation of the total trial population, which was not of interest for this review). Some outcome data were published on the trial registry, which we incorporated into this review.

Lastly, sensitivity analyses on the influence of trial design (blinded trial, unblinded (open-label)) was also not feasible, as 32 of 36 trials were open-label (i.e. non-blinded) trials. One trial did not report whether it was blinded or not, and of those three trials that were blinded, one was not included in any analyses, so no meaningful results would have been produced with a sensitivity analysis of the remaining two trials only.

Appendix ii (Systematic review II)

Aldin A, Umlauff L, Estcourt LJ, Collins G, Moons KG, Engert A, Kobe C, von Tresckow B, Haque M, Foroutan F, Kreuzberger N, Trivella M, Skoetz N. Interim PET-results for prognosis in adults with Hodgkin lymphoma: a systematic review and meta-analysis of prognostic factor studies. Cochrane Database of Systematic Reviews 2020, Issue 1. Art. No.: CD012643. DOI: 10.1002/14651858.CD012643.pub3.

Cochrane Library

Cochrane Database of Systematic Reviews

Interim PET-results for prognosis in adults with Hodgkin lymphoma: a systematic review and meta-analysis of prognostic factor studies (Review)

Aldin A, Umlauff L, Estcourt LJ, Collins G, Moons KGM, Engert A, Kobe C, von Tresckow B, Haque M, Foroutan F, Kreuzberger N, Trivella M, Skoetz N

Aldin A, Umlauff L, Estcourt LJ, Collins G, Moons KG, Engert A, Kobe C, von Tresckow B, Haque M, Foroutan F, Kreuzberger N, Trivella M, Skoetz N.
Interim PET-results for prognosis in adults with Hodgkin lymphoma: a systematic review and meta-analysis of prognostic factor studies.
Cochrane Database of Systematic Reviews 2020, Issue 1. Art. No.: CD012643.
DOI: 10.1002/14651858.CD012643.pub3.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Prognosis Review]

Interim PET-results for prognosis in adults with Hodgkin lymphoma: a systematic review and meta-analysis of prognostic factor studies

Angela Aldin[1], Lisa Umlauff[1], Lise J Estcourt[2], Gary Collins[3], Karel GM Moons[4], Andreas Engert[5], Carsten Kobe[6], Bastian von Tresckow[5], Madhuri Haque[1], Farid Foroutan[7], Nina Kreuzberger[1], Marialena Trivella[3a], Nicole Skoetz[8b]

[1]Cochrane Haematology, Department I of Internal Medicine, Center for Integrated Oncology Aachen Bonn Cologne Duesseldorf, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. [2]Haematology/Transfusion Medicine, NHS Blood and Transplant, Oxford, UK. [3]Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK. [4]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands. [5]Department I of Internal Medicine, Center for Integrated Oncology Aachen Bonn Cologne Duesseldorf, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. [6]University of Cologne, Faculty of Medicine and University Hospital Cologne, Department for Nuclear Medicine, Cologne, Germany. [7]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada. [8]Cochrane Cancer, Department I of Internal Medicine, Center for Integrated Oncology Aachen Bonn Cologne Duesseldorf, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany

[a]contributed equally. [b]contributed equally

Contact address: Angela Aldin, angela.aldin@uk-koeln.de.

Editorial group: Cochrane Haematology Group.
Publication status and date: Edited (no change to conclusions), published in Issue 8, 2020.

Citation: Aldin A, Umlauff L, Estcourt LJ, Collins G, Moons KG, Engert A, Kobe C, von Tresckow B, Haque M, Foroutan F, Kreuzberger N, Trivella M, Skoetz N. Interim PET-results for prognosis in adults with Hodgkin lymphoma: a systematic review and meta-analysis of prognostic factor studies. *Cochrane Database of Systematic Reviews* 2020, Issue 1. Art. No.: CD012643. DOI: 10.1002/14651858.CD012643.pub3.

ABSTRACT

Background

Hodgkin lymphoma (HL) is one of the most common haematological malignancies in young adults and, with cure rates of 90%, has become curable for the majority of individuals. Positron emission tomography (PET) is an imaging tool used to monitor a tumour's metabolic activity, stage and progression. Interim PET during chemotherapy has been posited as a prognostic factor in individuals with HL to distinguish between those with a poor prognosis and those with a better prognosis. This distinction is important to inform decision-making on the clinical pathway of individuals with HL.

Objectives

To determine whether in previously untreated adults with HL receiving first-line therapy, interim PET scan results can distinguish between those with a poor prognosis and those with a better prognosis, and thereby predict survival outcomes in each group.

Search methods

We searched MEDLINE, Embase, CENTRAL and conference proceedings up until April 2019. We also searched one trial registry (ClinicalTrials.gov).

Selection criteria

We included retrospective and prospective studies evaluating interim PET scans in a minimum of 10 individuals with HL (all stages) undergoing first-line therapy. Interim PET was defined as conducted during therapy (after one, two, three or four treatment cycles). The

minimum follow-up period was at least 12 months. We excluded studies if the trial design allowed treatment modification based on the interim PET scan results.

Data collection and analysis

We developed a data extraction form according to the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS). Two teams of two review authors independently screened the studies, extracted data on overall survival (OS), progression-free survival (PFS) and PET-associated adverse events (AEs), assessed risk of bias (per outcome) according to the Quality in Prognosis Studies (QUIPS) tool, and assessed the certainty of the evidence (GRADE). We contacted investigators to obtain missing information and data.

Main results

Our literature search yielded 11,277 results. In total, we included 23 studies (99 references) with 7335 newly-diagnosed individuals with classic HL (all stages).

Participants in 16 studies underwent (interim) PET combined with computed tomography (PET-CT), compared to PET only in the remaining seven studies. The standard chemotherapy regimen included ABVD (16) studies, compared to BEACOPP or other regimens (seven studies). Most studies (N = 21) conducted interim PET scans after two cycles (PET2) of chemotherapy, although PET1, PET3 and PET4 were also reported in some studies. In the meta-analyses, we used PET2 data if available as we wanted to ensure homogeneity between studies. In most studies interim PET scan results were evaluated according to the Deauville 5-point scale (N = 12).

Eight studies were not included in meta-analyses due to missing information and/or data; results were reported narratively. For the remaining studies, we pooled the unadjusted hazard ratio (HR). The timing of the outcome measurement was after two or three years (the median follow-up time ranged from 22 to 65 months) in the pooled studies.

Eight studies explored the independent prognostic ability of interim PET by adjusting for other established prognostic factors (e.g. disease stage, B symptoms). We did not pool the results because the multivariable analyses adjusted for a different set of factors in each study.

Overall survival

Twelve (out of 23) studies reported OS. Six of these were assessed as low risk of bias in all of the first four domains of QUIPS (study participation, study attrition, prognostic factor measurement and outcome measurement). The other six studies were assessed as unclear, moderate or high risk of bias in at least one of these four domains. Four studies were assessed as low risk, and eight studies as high risk of bias for the domain other prognostic factors (covariates). Nine studies were assessed as low risk, and three studies as high risk of bias for the domain 'statistical analysis and reporting'.

We pooled nine studies with 1802 participants. Participants with HL who have a negative interim PET scan result probably have a large advantage in OS compared to those with a positive interim PET scan result (unadjusted HR 5.09, 95% confidence interval (CI) 2.64 to 9.81, I^2 = 44%, moderate-certainty evidence). In absolute values, this means that 900 out of 1000 participants with a negative interim PET scan result will probably survive longer than three years compared to 585 (95% CI 356 to 757) out of 1000 participants with a positive result.

Adjusted results from two studies also indicate an independent prognostic value of interim PET scan results (moderate-certainty evidence).

Progression-free survival

Twenty-one studies reported PFS. Eleven out of 21 were assessed as low risk of bias in the first four domains. The remaining were assessed as unclear, moderate or high risk of bias in at least one of the four domains. Eleven studies were assessed as low risk, and ten studies as high risk of bias for the domain other prognostic factors (covariates). Eight studies were assessed as high risk, thirteen as low risk of bias for statistical analysis and reporting.

We pooled 14 studies with 2079 participants. Participants who have a negative interim PET scan result may have an advantage in PFS compared to those with a positive interim PET scan result, but the evidence is very uncertain (unadjusted HR 4.90, 95% CI 3.47 to 6.90, I^2 = 45%, very low-certainty evidence). This means that 850 out of 1000 participants with a negative interim PET scan result may be progression-free longer than three years compared to 451 (95% CI 326 to 569) out of 1000 participants with a positive result.

Adjusted results (not pooled) from eight studies also indicate that there may be an independent prognostic value of interim PET scan results (low-certainty evidence).

PET-associated adverse events

No study measured PET-associated AEs.

Authors' conclusions

This review provides moderate-certainty evidence that interim PET scan results predict OS, and very low-certainty evidence that interim PET scan results predict progression-free survival in treated individuals with HL. This evidence is primarily based on unadjusted data. More studies are needed to test the adjusted prognostic ability of interim PET against established prognostic factors.

PLAIN LANGUAGE SUMMARY

Imaging with positron emission tomography (PET) during chemotherapy to predict outcome in adults with Hodgkin lymphoma

Review question

This Cochrane Review aimed to find out whether the results of a positron emission tomography (PET) during therapy in people with Hodgkin lymphoma (HL) can help to distinguish between those with a poor prognosis and those with a better prognosis, and predict survival outcomes in each group.

Background

Hodgkin lymphoma is a cancer which affects the lymphoid system of the body. It is considered a relatively rare disease (two to three cases per 100,000 people every year in Western countries), that is most common in young adults in their twenties, but it can also occur in children and elderly people. As treatment options have improved, most people with HL can now be cured. It is important that individuals receive the treatment with the greatest efficacy and least toxicity possible. PET is an imaging tool for assessing the disease stage of an individual, and monitoring tumour activity. It has been suggested that PET performed during therapy (so-called interim PET, e.g. after two cycles of chemotherapy) can distinguish between people who respond well to therapy and those who do not respond well. The aim of this review was to demonstrate the prognostic ability to distinguish between these groups, and predict survival outcomes in each group, to help clinicians make an informed decision on the treatment pathway to improve long-term outcomes and safety for people with HL.

Study characteristics

We included 23 studies to explore the association between interim PET scan results after one to four cycles of chemotherapy and survival outcomes in adults with HL (all stages). We contacted 10 authors, and six provided us with relevant information and/or data.

Key results

In 16 included studies, participants received either ABVD chemotherapy or BEACOPP chemotherapy (four studies) only, with or without radiotherapy. In 16 studies, participants underwent an interim PET scan in combination with a computed tomography (CT) (PET-CT), which have higher accuracy in detecting primary and secondary cancers than a PET scan alone. In the remaining seven studies, PET-only was conducted. Twenty-one studies conducted interim PET scans after two cycles (PET2) of chemotherapy.

Eight studies did not report enough data on our outcomes or population of interest, so we reported the results from these studies narratively. We combined individual study results in meta-analyses to provide robust evidence for our outcomes of interest overall survival and progression-free survival. No study measured PET-associated adverse events (harms).

For overall survival, combined results from nine studies (1802 participants) show that there is probably a large advantage in overall survival for people with a negative interim PET scan compared to people with a positive interim PET scan. For progression-free survival, combined results from 14 studies (2079 participants) show that interim PET-negative people may have an advantage for progression-free survival, compared to interim PET-positive people, but we are uncertain about this result. These are unadjusted results, where interim PET was tested as the only prognostic factor.

Eight studies reported adjusted results, where the independent prognostic ability of interim PET was assessed against other established prognostic factors (e.g. disease stage, B symptoms). We could not combine individual study results because the studies did not include identical sets of covariates. Nevertheless, their results indicate a probable independent prognostic ability of interim PET to predict both outcomes.

Certainty of the evidence

Regarding the unadjusted results, we rated our certainty of the evidence as 'moderate' for overall survival. This means that the true effect is likely to be close to the estimated effect, but there is a possibility that it is substantially different. For progression-free survival, we rated our certainty of the evidence as 'very low', meaning that we have little confidence in the effect estimate, and that the true effect is likely to be substantially different from the estimated effect.

Regarding the adjusted results, we rated our certainty of the evidence as 'moderate' for overall survival, and 'low' for progression-free survival.

How up-to-date is this review?

We searched data bases up until 2 April 2019, and one trial registry on 25 January 2019.

Summary of findings 1. Comparison of interim PET-negative and interim PET-positive individuals with Hodgkin Lymphoma

Comparison of interim PET-positive and interim PET-negative participants with Hodgkin lymphoma

Population: Individuals with Hodgkin lymphoma
Setting: Eleven studies recruited participants from a total of 28 haemato-oncology treatment centres/hospitals in Brazil (N = 1), China (N = 1), Denmark (N = 4), France (N = 4), Italy (N = 3), Poland (N = 11), UK (N = 2) and the USA (N = 2). One study (Straus 2011) included participants from 29 institutions, but did not report the countries. One study (Simon 2016) reported the country (Hungary) but not the number of centres. One multi-centre study (Hutchings 2014) recruited participants from four countries (USA, Italy, Poland and Denmark). One RCT (Kobe 2018) included participants from 301 hospitals and private practices in Germany, Switzerland, Austria, the Netherlands, and the Czech Republic.

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with Interim PET-negative	Risk with Interim PET-positive				
Overall survival Follow up: 3 years	Low		HR 5.09 (2.64 to 9.81)	1802 (9 studies)	⊕⊕⊕⊖ MODERATE [2 3 4]	
	900 per 1.000 [1]	585 per 1.000 [1] (356 to 757)				
	High					
	980 per 1.000 [1]	902 per 1.000 [1] (820 to 948)				
Progression-free survival Follow up: 3 years	Low		HR 4.90 (3.47 to 6.90)	2079 (14 studies)	⊕⊖⊖⊖ VERY LOW [6 7 8]	
	850 per 1.000 [5]	451 per 1.000 [5] (326 to 569)				
	High					
	940 per 1.000 [5]	738 per 1.000 [5] (653 to 807)				
Adverse events associated with PET - not reported	No study measured PET-associated adverse events.		-	-	-	

Overall survival (adjusted effect estimate)	Two studies reported an adjusted effect estimate for overall survival after interim PET2: a hazard ratio of 3.2 (95% CI 1.3 to 8.4, P = 0.02) (Kobe 2018) and 11.51 (95% CI 3.14 to 42.86, P < 0.001) (Simon 2015) indicates the independent prognostic value of interim PET over and above other clinically relevant prognostic factors.	-	843 (2 studies)	⊕⊕⊕⊖ MODERATE [9]
Progression-free survival (adjusted effect estimate)	Eight studies conducted a multivariable analysis to test the independent prognostic value of interim PET over and above other clinically relevant prognostic factors. Four of these studies reported a hazard ratio as the adjusted effect estimate, of which the value ranges from 2.4 to 36.89, indicating the independent prognostic value of interim PET2.[10]	-	996 (4 studies)[10]	⊕⊕⊖⊖ LOW [11 12]

*__The survival in the PET-positive group__ (and its 95% confidence interval) is based on the assumed survival in the PET-negative group.

CI: Confidence interval; **HR:** Hazard ratio; **PET:** positron emission tomography

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

[1] The assumed event-free survival in the control group is based on the survival rate of the interim PET-negative participants at 3 years in the studies included (the lowest survival rate from Cerci 2010 and the highest survival rate from Kobe 2018).
[2] High risk of bias in seven studies for the domain 'other prognostic factors (covariates)', and high risk of bias in three studies for the domain 'statistical analysis and reporting'. Downgraded by 1 point for risk of bias.
[3] For one study we used the reported hazard ratio. For seven studies we had to estimate the hazard ratio and for one study we re-calculated it (Trivella 2006). Downgraded by 1 point for imprecision.
[4] Upgraded by one point due to the large effect showing the large difference between interim PET-negative and interim PET-positive participants (HR 5.09, CI 2.64 to 9.81).
[5] The assumed event-free survival in the control group is based on the survival rate of the interim PET-negative participants at 3 years in the studies included (the lowest survival rate from Rossi 2014 and the highest survival rate from Kobe 2018).
[6] High risk of bias in eight studies for the domain 'other prognostic factors (covariates)', and high risk of bias in six studies for the domain 'statistical analysis and reporting'. Downgraded by 1 point for risk of bias.
[7]The definition of PFS varied across studies, downgraded by 1 point for inconsistency
[8] For three studies we used the reported hazard ratio. For ten studies we had to estimate the value, and for one study we had to re-calculate it (Trivella 2006). Downgraded by 1 point for imprecision.
[9] High risk of bias for the domains 'other prognostic factors (covariates)' and statistical analysis and reporting for one study (Simon 2016). Downgraded by 1 point for risk of bias.
[10] Hutchings 2006; Kobe 2018; Mesguich 2016; Simon 2016.
[11] High risk of bias for the domains 'other prognostic factors (covariates)' and statistical analysis and reporting for one study (Simon 2016). Also high risk of bias for the domain study participation in one study (Hutchings 2006). Downgraded by 1 point for risk of bias.
[12] Studies included a heterogenous set of covariates in the adjusted analyses. Downgraded by 1 point for inconsistency.

BACKGROUND

Description of the condition

Hodgkin lymphoma (HL) is a cancer of the lymph nodes and the lymphoid system with possible involvement of other organs such as the liver, lung, bone or bone marrow (Lister 1989). With an annual incidence of approximately two to three per 100,000 inhabitants in Western countries, HL is a comparatively rare disease, but it is one of the most common malignancies in young adults (Howlader 2015). In industrialised countries, the age distribution of HL shows a first peak in the third decade and a second peak after the age of 50 (Thomas 2002).

The World Health Organization (WHO) Classification of Tumours of Haematopoietic and Lymphoid Tissues distinguishes between two types of HL: classical HL, representing about 95% of all HL; and lymphocyte-predominant HL, representing about 5% of all HL (Swerdlow 2008). Both types differ in morphology, phenotype and molecular features, and therefore in clinical behaviour and presentation (Re 2005).

The Ann Arbor Classification is used for staging and distinguishes between four different tumour stages. Stages one to three indicate the degree of lymph node and localised extranodal organ involvement, or both, and stage four includes disseminated organ involvement, which can be found in 20% of cases. Factors associated with a poor prognosis include a large mediastinal mass, three or more involved lymph node areas, a high erythrocyte sedimentation rate, extranodal lesions, B symptoms (weight loss > 10%, fever, drenching night sweats) and advanced age, but the factors considered as significant vary slightly between different study groups (German Study Hodgkin Lymphoma Study Group (GHSG); European Organization for Research and Treatment of Cancer (EORTC); National Cancer Institute of Canada (NCIC)). The Cotswold modification of the Ann Arbor Classification also takes into consideration the occurrence of bulky disease (largest tumour diameter greater than 10 cm) (Lister 1989). Hodgkin lymphoma is classified into early favourable, early unfavourable and advanced stage (Engert 2007; Klimm 2005). In Europe, the early favourable-stage group usually comprises Ann Arbor stages I and II without risk factors. The early unfavourable-stage group includes individuals with Ann Arbor stages I or II and one or more risk factors. Most individuals with stages IIB, III or IV disease are included in the advanced-stage risk group (Engert 2003).

With cure rates of up to 90%, HL is one of the most curable cancers worldwide (Engert 2010; Engert 2012; Rancea 2013a; von Tresckow 2012). A combination of adriamycin, bleomycin, vinblastine and dacarbazine (ABVD) is widely accepted as the standard chemotherapy regimen in early-stage HL (Bröckelmann 2018, Canellos 1992; Engert 2010). Individuals in this stage usually receive a combination of chemotherapy and involved-field radiation therapy (IF-RT) (Engert 2010; von Tresckow 2012), whereas those with advanced-stage disease receive an intensified regimen, such as BEACOPP (bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine and prednisone) (Skoetz 2017a; Borchmann 2011; Engert 2012; Skoetz 2013), or ABVD. A large randomised study showed that two cycles of ABVD followed by 20 Gy of IF-RT is sufficient for the treatment of early-favourable HL (Engert 2010), which is implemented into current standard treatment, whereas four cycles of chemotherapy followed by 30 Gy IF-RT is more suitable for individuals with early-unfavourable HL. Approximately 10% of people with HL will be refractory to initial treatment or will relapse; this is more common in people with advanced stage or bulky disease. These individuals can be treated with high-dose chemotherapy and autologous stem cell transplantation (Rancea 2013). Immunotherapy for relapsed HL as another possible approach is under active investigation (Moskowitz 2018).

The current treatment approach for HL aims to maximise progression-free and OS and to minimise acute and long-term toxicities like cardiac and pulmonary damage, infertility and secondary cancers. Development of a secondary cancer is one of the major causes of morbidity and mortality once the risk of progression and relapse of HL is over, i.e. from about five years after first-line treatment onwards. In a large systematic review based on individual patient data in people with HL, Franklin and colleagues demonstrated that treatment de-intensification by avoiding additional radiotherapy reduces the risk of a secondary cancer (Franklin 2005).

Description of the index (prognostic) factor

A prognostic factor is a characteristic of a patient or the disease (e.g. age, sex, co-morbidities, disease stage, blood or imaging results) that is likely to predict patient outcomes or health events, often related to OS and disease-free survival (Moons 2009; Riley 2013). Prognostic information ultimately provides a basis for the determination of treatment and also helps to stratify individuals for treatment according to their risk of future outcomes (Riley 2013). Established prognostic factors in HL include age, gender, B symptoms, Ann Arbor disease stage, bulky disease, albumin level, anaemia and white blood cell count, amongst others (Cuccaro 2014; Josting 2010; Kılıçkap 2013). Particularly male gender, advanced disease stage or age, and a low level of albumin, for example, are associated with worse prognosis and survival outcomes (Cuccaro 2014; Josting 2010).

The prognostic factor to be examined in this review is the tumour's metabolic activity, its stage, and progression as captured by [18F]-fluorodeoxy-D-glucose (FDG)-positron emission tomography (PET, also called PET scanning), which is an imaging tool. The principle of FDG-PET is based on a radio-labelled glucose analogue being a good indicator of the glucose metabolism of a tissue. It comprises two parts: a vector (2-deoxy-D-glucose) taken up by cells with a high metabolic rate, and 18F, a positron-emitting nuclide, which is detected by scintigraphy. FDG-PET scanning provides the opportunity to identify the state and degree of progression of FDG-avid tumours and has therefore become a standard imaging tool for various cancers (Boellaard 2010). Hodgkin lymphoma is a FDG-avid tumour; in a study of 233 people with HL, 100% were FDG-avid (Weigler-Sagie 2010). However, as the field of imaging continuously evolves, it is now widely accepted to use PET in combination with a computed tomography (CT), known as PET-CT (Barrington 2014). The combination of PET-CT is argued to provide clearer imaging and a more accurate measurement of nodal size (Cheson 2014). Nevertheless, in the studies included in this review, the use of PET or PET-CT varied.

Over the last few decades FDG-PET has been used more and more for staging, prognosis, treatment planning and response evaluation in individuals with HL, and is a widely accepted procedure (Barrington 2017a; Cheson 2014; Fitzgerald 2019; Kobe 2010a; Markova 2009; Meignan 2009; Radford 2015; Specht 2007).

FDG-PET is primarily used for the pretreatment assessment in order to determine the stage of the disease of an individual and thereby to decide on the appropriate treatment regimen (Cheson 2014; Meignan 2009). However, it is now argued that PET should also be conducted during first-line chemotherapy in individuals with HL, namely interim PET after a few cycles of chemotherapy (Barrington 2017a; Bröckelmann 2018; Meignan 2009). The result of the interim PET scan (positive or negative) is believed to be a good predictor of outcome, aiding the distinction between individuals with a poor prognosis from those with a better prognosis, while undergoing early treatment (Gallamini 2007; Kobe 2010; Markova 2012). Therapy adaptation based on interim PET results was introduced after detailed exploration of the FDG-PET procedure (Engert 2012; Kobe 2008a), the idea being to achieve maximum efficacy in terms of OS and progression-free survival (PFS).We will refer to the prognostic factor henceforth as 'interim PET'.

Why it is important to do this review

There is a need to systematically explore the prognostic ability of the factor (interim PET) in conditions where there is no treatment adaptation. The 'no treatment adaptation' clause is a rather important point in the prognostic exploration as adapting treatment based on interim PET results in daily practice when its prognostic ability is not yet proven is not desired. There is one systematic review on the prognostic value of interim PET without treatment adaptation in individuals with HL (Adams 2015a). However, this review looked at 'treatment failure' as an outcome of the interim PET scan, which is different to the outcomes the current review explored. Moreover, and despite the fact that it is entitled as a review of prognosis studies, the methodology used is akin to diagnostic test evaluation (with calculations of diagnostic odds ratio, specificity and sensitivity), rather than using established prognostic methodology and crucially, the confidence in the calculated estimates was not rated. Moreover, the review included studies published before December 2014 and, therefore, important research published since that time is not included.

One Cochrane Review on the role of PET-adapted treatment modification for people with HL found some evidence that PFS was decreased in people with early-stage HL and a negative PET scan receiving only chemotherapy (PET-adapted therapy) compared to those receiving radiotherapy in addition to chemotherapy (which is the standard therapy regimen) (Sickinger 2015). A similar result was found in another Cochrane Review (Blank 2017). The authors compared the effects of chemotherapy alone versus chemotherapy plus radiotherapy on outcome and safety for adults with early stage HL. They found moderate evidence that when individuals receive the same number of chemotherapy cycles, the addition of radiotherapy can improve PFS. However, both reviews were not able to give definite conclusions on the effect on OS. Another systematic review suggests the change of therapy after interim PET in advanced-stage individuals only (Amitai 2018). In the current German guideline for the treatment of HL, for example, it is recommended that patients with advanced HL receive an interim PET scan after two cycles of chemotherapy. The result of the interim PET scan can then be used to guide further treatment for patients in advanced stages of HL (Bröckelmann 2018). Hence, the disease stage is an additional key prognostic factor for patients with HL. Several randomised controlled trials (RCTs) have recently been published that investigated the consequences of treatment adaptation based on interim PET scan results on outcome and safety for individuals with HL (Andre 2017; Casasnovas 2019; Kobe 2018; Johnson 2016; Radford 2015).

Hence, the prognostic role of interim PET in individuals with HL undergoing first-line chemotherapy is very important and will strongly influence decision-making particularly regarding the choice of subsequent treatments. Therefore, we have summarised all available data from identified studies and included these in a meta-analysis when they were sufficiently homogeneous. Our aim was to produce robust evidence based on the improved power that a meta-analysis provides over the limitations of individual primary studies, and grade the evidence. A reliable answer to the question of the prognostic value of interim PET scan to predict survival outcomes in individuals with HL will strongly influence decision-making at a crucial point of an individual's treatment pathway. Moreover, grading the evidence on the prognostic value of interim PET will provide readers with an estimate of how much they can rely on the calculated results.

The aim of this systematic review was to determine whether in previously untreated adults with HL receiving first-line therapy, interim PET scan results can distinguish between those with a poor prognosis and those with a better prognosis, and whether it can predict survival outcomes in each group. Thereby, we assessed the prognostic value of interim PET scan results. Meta-analyses and grading of the evidence allow a conclusion of whether interim PET is a prognostic factor. This comprehensive overview will have a great impact on international guidelines and clinical pathways, and will contribute to a high-grade support in clinical decision-making for effective, supportive strategies for the individual patient.

OBJECTIVES

To determine whether in previously untreated adults with Hodgkin lymphoma (HL) receiving first-line therapy, interim positron emission tomography (PET) scan results can distinguish between those with a poor prognosis and those with a better prognosis, and thereby predict survival outcomes in each group.

Primary objective

To identify all studies evaluating interim PET scan results as a prognostic factor, describe the characteristics and risk of bias of included studies and meta-analyse results on the association between PET scan results and overall survival (OS), progression-free survival (PFS) and PET-associated adverse events.

PICOTS

We used the PICOTS (population, index, comparator, outcome(s), timing, setting) system to describe the key items for framing this review and its objective and methodology (Table 1) (Debray 2017; Riley 2019).

Table 1. PICOTS system

Population	Index (prognostic) factor	Comparator	Outcome(s)	Timing	Setting

• People with classic HL, at any stage of the disease • Newly diagnosed individuals undergoing first-line therapy • Adults, as defined in the studies	Interim PET scan results	Not applicable to this review	• Overall survival (OS) • Progression-free survival (PFS) • PET-associated adverse events (AEs) The outcome should be measured after a minimum follow-up of 12 months.	• Interim PET scan should be conducted during chemotherapy (after one, two, three or four cycles of chemotherapy)	Hospital/treatment centre

METHODS

This is a systematic review of prognostic factor studies.

Criteria for considering studies for this review

Types of studies

We included retrospective and prospective studies evaluating interim PET scan results in a minimum of 10 individuals with Hodgkin lymphoma (HL) undergoing first-line therapy.

We excluded studies that modified the treatment regimen based on the interim PET scan results in order to draw an unbiased conclusion of the ability of interim PET to predict the outcomes under study.

Participants

We included studies on adults with newly diagnosed classic HL receiving first-line therapy. If in a study a percentage of the included participants were adolescents but received adult treatment regimen and dosage, and the study considered them as adults, then we also accepted this 'adult' definition.

All participants received an interim PET scan during chemotherapy (e.g. after one, two, three and/or four cycles of chemotherapy), and continued with the planned chemotherapy regimen, without treatment adaptation due to the interim PET scan result.

Index (prognostic) factor

We included studies that assessed interim PET scan results as the index (prognostic) factor to predict survival outcomes. We expected the interim PET scan to be conducted during first-line treatment of adults with HL, and without interim PET-guided treatment adaptation, meaning participants should be treated in the same way regardless of the interim PET scan result. We accepted all studies that conducted a PET or PET-CT (see Background 'Description of index (prognostic) factor').

In the literature, it is generally recommended to use a five-point scale to assess the grade of uptake and report the PET scan result (Meignan 2009). Generally, scores 1-3 indicate PET-negativity, while scores 4-5 indicate PET-positivity (Barrington 2014). Most of the included studies used a validated scale, such as the 5-PS Deauville criteria (Meignan 2009), the Lugano classification (Cheson 2014), the Imaging Subcommittee of International Harmonization Project in Lymphoma criteria (Juweid 2007) or the joint Italian-Danish study criteria (Gallamini 2007).

Type of outcome measures

Primary outcome

- Overall survival (OS), defined as the time to death due to any cause.

We chose OS as our primary outcome because it has the greatest clinical relevance and is most important for individuals with HL. Furthermore, death due to any cause is an objective endpoint not susceptible to bias by the outcome assessor.

Secondary outcomes

- Progression-free survival (PFS), defined as the time to disease progression, relapse, death due to any cause or last follow-up.
- Adverse events (AEs), defined as any event associated with the index factor (e.g. radiation safety).

To report meaningful findings, the required minimum follow-up period was 12 months for each outcome.

Search methods for identification of studies

Electronic searches

Reporting and therefore retrieval of prognostic factor studies is very poor, as evaluation of guidelines on reporting of prognostic markers in cancer have shown (Altman 2012; Mallett 2010; McShane 2005). Moreover, no specific search filter exists for this new methodological approach, therefore published filters have to be combined for a sensitive search strategy (Geersing 2012). However, as PET scans often are not reported as a prognostic factor, we did not combine our search strategy with a filter for prognosis research. Therefore, the search strategy was not very specific and the results were screened independently and in detail by two teams of two review authors. Furthermore, we did not apply a language restriction in order to reduce the language bias, according to chapter six of the *Cochrane Handbook for Systematic Reviews of Interventions* (Lefebvre 2011).

We searched the following databases.

- Databases of medical literature
  - Cochrane Central Register of Controlled Trials (CENTRAL; 2 April 2019, Issue 11) (Appendix 1)
  - MEDLINE Ovid SP (1946 until 2 April 2019) (Appendix 2)
  - Embase (1990 until 2 April 2019) (Appendix 2)
- Conference proceedings of annual meetings of the following societies for abstracts (2000 to 2019)
  - American Society of Hematology

- European Hematology Association
- International Symposium on Hodgkin Lymphoma

- We searched ClinicalTrials.gov (on 25 January 2019 using the query PET and Hodgkin lymphoma) to identify clinical trials.

Searching other resources

- Handsearching of references
  * We searched the references of all identified studies, relevant review articles and current treatment guidelines for further literature to find other relevant studies and to identify associated articles.
- Personal contacts
  * We contacted 10 principal investigators of included studies for further information, of whom six replied and answered our questions for clarification. Two out of these six provided us also with relevant data to conduct our analyses.

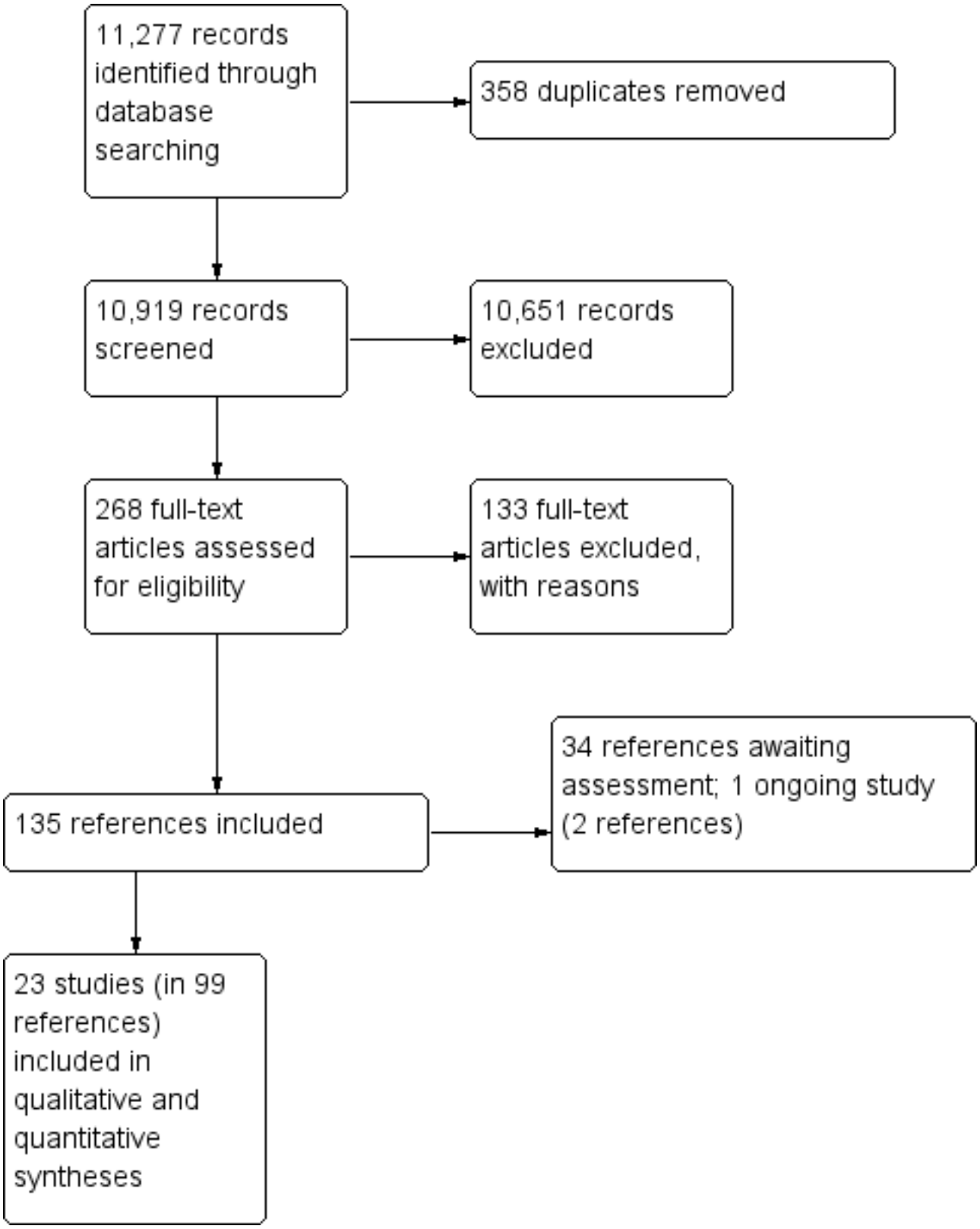
Data collection and analysis

Selection of studies

Two teams of two review authors (AA, LE, MHT, NS) independently screened the results of the search strategies to identify eligible studies by reading the titles and abstracts in Covidence (Covidence). In case of disagreements, consensus between the two review authors was reached by discussion of the full-text publication. When consensus could not be reached, a third review author was consulted for final decision (Higgins 2011).

We documented the study selection process in a flow chart as recommended in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher 2009), showing the total numbers of retrieved references and the numbers of included and excluded studies (Figure 1).

Figure 1. Study flow diagram according to PRISMA



Data extraction and management

We developed a data extraction form specific to studies of prognostic factors based on the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) (Moons 2014). The form was piloted using four of the included studies, and then further assessed during several teleconferences between the review authors to discuss required changes. After several amendments of the form, two teams of two review authors (AA, LE, MHT, NS) independently extracted all relevant data from the included studies. After data extraction, we contacted 10 principal investigators of included studies to request additional information.

Our form included the following items (in short).

- General information
  - i.e. Author, title, source, publication date, country, language, duplicate publications
- Source of data
  - i.e. Cohort, prospective planned study, randomised study participants, or registry data
- Participants
  - Participant eligibility and recruitment method (e.g. consecutive participants, location, number of centres, setting, inclusion and exclusion criteria)
  - Participant description (e.g. age, gender, stage of disease)
  - Details of treatments received
  - Study dates
- Prognostic factor
  - Definition and method for measurement of prognostic factor
  - Timing of prognostic factor measurement (number of chemotherapy cycles before and after measurement of the prognostic factor)
- Outcomes to be predicted
  - Definition and method for measurement of outcome
  - Was the same outcome definition (and method for measurement) used in all individuals?
  - Was the outcome assessed without knowledge of the prognostic factor (i.e. blinded)?
  - Time of outcome occurrence or summary of duration of follow-up
- Sample size
  - Number of participants and number of outcomes/events
- Missing data
  - Number of participants with any missing value (include predictors and outcomes)
  - Handling of missing data (e.g. complete-case analysis, imputation, or other methods)
- Reported results
  - Overall survival (OS) (including duration of follow-up)
  - Progression-free survival (PFS) (including duration of follow-up)
  - Adverse events (AEs) (including duration of follow-up)

Risk of bias

In the protocol for this review we prespecified that we will use the Quality in Prognostic Studies (QUIPS) tool (Hayden 2013) for the risk of bias assessment. However, recent methodological developments for the systematic review of prognostic factor studies (Riley 2019; Riley 2019b) led us to consider amending this tool. In the light of this we consulted the primary author (Hayden 2013) of the QUIPS tool and following discussions decided to add to the three bias ratings ('low', 'moderate' and 'high' risk of bias) a fourth 'unclear' option. This was necessary due to the inconsistent reporting of the included studies, when information was clearly missing, and hence, without an 'unclear' category, risk of bias assessment would not be feasible.

Following further discussions, we additionally decided to rename the fifth domain 'study confounding' to 'other prognostic factors (covariates)' in order to highlight the important distinction between confounding (the preferred term when seeking estimates of causal effect of a specific etiologic factor) and adjusting for other important prognostic factors, namely covariates (advocated when seeking the independent prognostic ability of index prognostic factors). As said, in the context of our review (adults with Hodgkin lymphoma), the disease stage is a key factor that is taken into account together with the interim PET scan result when decisions about treatment adaptation are made in daily clinical practice (Bröckelmann 2018). Hence, we assessed studies that only included participants within one disease stage (e.g. only early stages or only advanced stages of HL) as 'low' risk of bias, as such patient sampling can be considered as accounting for disease stage as another prognostic factor. Studies that included participants within all disease stages, but offered adjusted results including disease stage as another prognostic factor, were also assessed as 'low' risk of bias. Studies with participants of all disease stages, not accounting for disease stage, were assessed as 'high' risk of bias in this domain. This latter modification is also reflected in the GRADE assessment. Regardless of whether meta-analysis of adjusted or unadjusted (crude) effects of the prognostic factor of interest (interim PET scan results) was possible, we included this domain's risk of bias assessment in our GRADE judgement.

Two teams of two review authors (AA, LE, MHT, NS) independently assessed the risk of bias of the included studies according to the domains of the QUIPS tool. We judged each domain by taking into account the criteria listed for each domain in the QUIPS tool (Hayden 2013), and also provided a brief statement supporting our judgement.

We made the following judgements.

- **Low risk of bias**: the relationship between the prognostic factor and outcome is unlikely to be different for participants and eligible non-participants.
- **Moderate risk of bias**: the relationship between the prognostic factor and outcome may be different for participants and eligible non-participants.
- **High risk of bias**: the relationship between the prognostic factor and outcome is very likely to be different for participants and eligible non-participants.
- **Unclear risk of bias**: the study does not provide sufficient information that allows a clear judgement for this domain.

Furthermore, we decided to assess the risk of bias per outcome in each study because not all studies reported all of our outcomes of interest, and even studies reporting at least two of our outcomes showed differences in their outcome reporting.

We judged the following domains and criteria.

- Study participation
  - Adequate participation in the study by eligible persons
  - Description of the source population or population of interest
  - Description of the baseline study sample
  - Adequate description of the sampling frame and recruitment
  - Adequate description of the period and place of recruitment
  - Adequate description of inclusion and exclusion criteria
- Study attrition
  - Adequate response rate for study participants
  - Description of attempts to collect information on participants who dropped out
  - Reasons for loss to follow-up are provided
  - Adequate description of participants lost to follow-up
  - There are no important differences between participants who completed the study and those who did not
- Prognostic factor measurement
  - A clear definition or description of the prognostic factor is provided
  - Method of prognostic factor measurement is adequately valid and reliable
  - Continuous variables are reported or appropriate cut points are used
  - The method and setting of measurement of prognostic factor is the same for all study participants
  - Adequate proportion of the study sample has complete data for the prognostic factor
  - Appropriate methods of imputation are used for missing prognostic factor data
- Outcome measurement
  - A clear definition of the outcome is provided
  - Method of outcome measurement used is adequately valid and reliable
  - The method and setting of outcome measurement is the same for all study participants
- Other prognostic factors (covariates)
  - Other prognostic factors (covariates) are measured
  - Clear definitions of the important prognostic factors (covariates) measured are provided
  - Measurement of all important prognostic factors (covariates) is adequately valid and reliable
  - The method and setting of prognostic factor measurement are the same for all study participants
  - Appropriate methods are used if imputation is used for missing data
  - Important potential prognostic factors (covariates) are accounted for in the study design
  - Important potential prognostic factors (covariates) are accounted for in the analysis
- Statistical analysis and reporting
  - Sufficient presentation of data to assess the adequacy of the analytic strategy
  - Strategy for model building is appropriate and is based on a conceptual framework or model
  - The selected statistical model is adequate for the design of the study
  - There is no selective reporting of results

Reporting deficiencies

Methods and reporting in prognostic research often do not follow current methodological recommendations, limiting retrieval, reliability and applicability of these publications (Bouwmeester 2012; Peat 2014). There is evidence suggesting that prognosis research in cancer is cluttered with false-positive studies, which would not have been published if the results were negative (Kyzas 2005; Kyzas 2007; Sauerbrei 2005). Moreover, studies evaluating prognostic factors are usually not prospectively registered and no protocol is published (Peat 2014; Riley 2013), resulting in difficulties to identify all studies and to assess potential risks of publication bias. We used sensitive search filters for the disease (HL) and the prognostic factor (interim PET scan results) without any specific filter for research on prognosis in order to increase retrieval.

Due to the expected large effect of hazard ratios (HRs), tests for funnel plot asymmetry could result in publication bias being incorrectly indicated by the test (Macaskill 2010). Therefore, we decided not to evaluate the risk of publication bias by funnel plot asymmetry and describe reporting deficiencies instead.

Data synthesis

We performed analyses according to the recommendations of Cochrane, and the Cochrane Prognosis Methods Group in particular, and used the Cochrane statistical package Review Manager 5 (Deeks 2011; Review Manager 2014). We are aware that since the protocol development, the methodology on assessing studies of prognosis has evolved; hence, some differences between the published protocol and this full review may exist to account for the updated guidance. We have listed these in Differences between protocol and review.

We pooled unadjusted (crude) HRs for OS and PFS by applying meta-analysis using the RevMan's generic inverse variance methods random-effects model. Due to reporting inefficiencies and the expected heterogeneity between studies, we only combined studies that were sufficiently similar (e.g. most studies used ABVD as the main therapy regimen, or most studies conducted interim PET after two cycles of chemotherapy). Studies did not always provide an HR and associated standard error (SE), which are the parameters needed for meta-analysis. Where these values were not available, we estimated them from other available data where possible using an in-house calculator based on published methods for recovering survival data (Altman 1999; Parmar 1998; Tierney 2007). Recovered data included information and results reported in the text, tables, and Kaplan-Meier (K-M) curves. We also contacted 10 principal investigators of included studies to either ask for additional data, or to clarify issues regarding the studies.

As prespecified in the protocol, we would have also pooled adjusted HRs of the interim PET scan-result (the index prognostic factor) from multivariable analyses of the included studies as

adjusted prognostic effects (e.g. HRs) indicate the independent prognostic value of the prognostic factor over and above other clinically relevant prognostic factors (Riley 2019). However, pooling of adjusted estimates is recommended only if the same (largely) prognostic factors (covariates) are adjusted for in multivariable analyses (Riley 2019; Riley 2019b). As, said clinically relevant prognostic factors in individuals with HL particularly include the disease stage, as well as age, gender, and B symptoms (Cuccaro 2014). Regardless of whether pooling of adjusted or unadjusted effects of interim PET scan results was possible, we always assessed the risk of bias for all studies using the QUIPS tool, including the fifth domain 'other prognostic factors (covariates)', where we considered the disease stage as an important covariate to be taken into account.

Detailed description of the estimation of hazard ratios (HRs) and standard errors (SEs)

We used unadjusted HRs as the effect measure for OS and PFS. In cases where the HR and SE were not reported, we estimated them from available data using an in-house calculator (Trivella 2006), based on methods reported by Tierney 2007, Altman 1999 and Parmar 1998, or contacted authors to request additional data (Higgins 2011b). Recovered data included sample size, number of events, results such as the logrank P-value and confidence intervals (CIs), which were reported in the text, tables, and K-M curves. We kept detailed records of how the HR and SEs were calculated for each outcome in each included study. We identified the following six categories of HR precision.

1. HR was provided in the study, and the SE was either provided or easily estimated from reported CIs, and/or using the RevMan inbuilt calculator.
2. HR was provided but on checking while attempting to obtain the SE, there were errors and/or discrepancies with related provided data and we re-estimated the HR.
3. HR and SE were not provided but all necessary data for their estimation were available in the study.
4. HR and SE were not provided. Other necessary data were available but not an exact logrank P value, hence the nearest value was used in the estimation. For example, if they reported P < 0.001, then the nearest exact value was used, in this case P = 0.0009.
5. HR and SE were not provided. Other necessary data were available but the number of events was estimated from the K-M curves.
6. IPD data were available and HR and SE were accurately calculated.

We are aware that categories four and five are likely to over- or under-estimate the HR and associated SE. However, they were the best estimates we could obtain. We consider the remaining categories as precise. We explored the precision of the estimates in a post-hoc sensitivity analysis where the imprecise studies were temporarily removed to examine the robustness of the pooled result.

Grading the evidence

According to the recommendations of the GRADE working group, we rated and described the confidence in estimates for each outcome by assessing potential risk of bias, inconsistency, imprecision, indirectness and publication bias. We applied an approach that has been proposed for prognosis studies by the GRADE working group, suggesting that the starting point is one of high certainty of the evidence for observational studies (Iorio 2015).

Dealing with missing data

We dealt with missing data as suggested in Chapter 16 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011b). We contacted ten principal investigators of included studies to answer our questions regarding the studies and/or to provide us with additional data. Six principal investigators replied and answered our questions, of which two also provided us with additional data necessary to perform our analyses. One investigator kindly provided us with individual participant data for the whole data set. In some studies, the description of the methodology was rather unclear or relevant information was missing. In addition, some studies did not fully report their statistical analyses and data were missing, which complicated a full assessment of the study. We performed sensitivity analysis to assess how sensitive the results were to reasonable changes in the assumptions that were made, and addressed the potential impact of missing data on the findings of this review in the Discussion.

Furthermore, we noticed that most studies applied exclusion criteria on the baseline population (such as unavailability of interim PET or descriptive information) without providing a description of the size of this population and/or reasons for missing information. We treated this as a potential source of selection bias in the domain study participation of the QUIPS tool.

Investigation of heterogeneity

We investigated and discussed clinical and statistical heterogeneity and design aspects of included studies as mentioned in the section 'Data extraction and data management'. We assessed between-study heterogeneity using the I^2 statistic (an I^2 greater than 50% = moderate heterogeneity; an I^2 greater than 80% = considerable heterogeneity) (Deeks 2011). As most studies of prognosis are observational in nature, we are aware that they are prone to higher and/or inflated heterogeneity. Hence, we also assessed the Tau^2 values from the meta-analyses to be able to make a more robust judgment on the degree of statistical heterogeneity.

As specified in the protocol, we explored potential causes of heterogeneity by subgroup analysis. We considered the following parameters.

- Study design (e.g. prospective versus retrospective)
- Disease stage (e.g. early versus advanced stages)
- Type of chemotherapy (e.g. ABVD versus BEACOPP)
- Type of radiotherapy (e.g. involved field versus involved site)
- Type of PET measurement (e.g. PET versus PET-CT) *(post-hoc)*

In addition, we conducted a post hoc sensitivity analysis for the timing of the interim PET, as well as the availability/estimation of HR and SE to explore the robustness of the pooled results.

RESULTS

Results of the search

Our literature search in CENTRAL, MEDLINE and Embase (until 2 April 2019, see Appendix 1, Appendix 2 and Appendix 3, respectively) and one trial registry (ClinicalTrials.gov on 25 January

2019), identified 11,277 potentially relevant publications. After removal of 358 duplicates, we screened titles and abstracts of 10,919 references using inclusion and exclusion criteria defined at the protocol stage. These criteria led to the exclusion of 10,651 references, and 268 references were then included for full-text screening. Before starting full-text screening, we discussed and determined exclusion reasons. Full-text screening led to the exclusion of 133 references. Thirty-four references that were identified are still awaiting assessment (see Studies awaiting classification), and one study is still ongoing (see Ongoing studies). Hence, we finally included 23 studies (from 99 references) in this review. The overall number of publications screened, identified, selected and included in this review is shown in Figure 1

Description of studies

Included studies

See also Characteristics of included studies.

We included 23 studies in this review (Andre 2017; Annunziata 2016; Barnes 2011; Casasnovas 2019; Cerci 2010; Gallamini 2014; Gandikota 2015; Hutchings 2005; Hutchings 2006; Hutchings 2014; Kobe 2018; Markova 2012; Mesguich 2016; Oki 2014; Okosun 2012; Orlacchio 2012; Rossi 2014; Simon 2016; Straus 2011; Touati 2014; Ying 2014; Zaucha 2017; Zinzani 2012), which added up to a total of 99 references when secondary citations were included. To avoid duplication and overlapping of participant data in our analyses, we grouped those publications that assessed the same population (or groups from the same population). In such cases, we chose the publication with the greatest number of participants and/or most information as the primary publication. Duplicate or overlapping study populations were found for eight studies (Andre 2017; Barnes 2011; Gallamini 2014; Kobe 2018; Markova 2012; Simon 2016; Straus 2011; Zinzani 2012). Four studies did not report the duration of follow-up (Andre 2017; Annunziata 2016; Orlacchio 2012; Straus 2011). The earliest study recruited participants between 1993 and 2004 (Hutchings 2005), and the most recent between 2007 and 2014 (Annunziata 2016).

There was considerable heterogeneity between the included studies, particularly with regard to: stages of disease; treatment regimens; and the timing and criteria for evaluation of the interim PET scans, which are described in detail in the sections below. For meta-analyses, we only grouped studies that were homogenous enough in order to ensure comparability, and conducted subgroup analyses to explore the potential impact of heterogeneity on our results (see Methods 'Investigation of heterogeneity').

Study design

Of the 23 included studies, seven studies were retrospective single-centre studies (Annunziata 2016; Markova 2012; Oki 2014; Orlacchio 2012; Rossi 2014; Touati 2014; Ying 2014). Five studies were retrospective multi-centre studies (ranging between two to 17 centres) (Barnes 2011; Gallamini 2014; Mesguich 2016; Okosun 2012; Zinzani 2012). Two retrospective studies did not report the number of centres from which participants were recruited (Gandikota 2015; Simon 2016). Out of eight studies with a prospective study design, one study was a single-centre study (Cerci 2010), three were multi-centre studies (including between four and 11 centres, with Hutchings 2014 not reporting the number of study centres) (Hutchings 2006; Hutchings 2014; Zaucha 2017), and four were clinical trials (Andre 2017; Casasnovas 2019; Kobe 2018; Straus 2011). One study did not report the study design (Hutchings 2005).

For more details see Characteristics of included studies.

Sample size

The smallest study included 23 participants (Okosun 2012) and the largest study included 1945 participants (Kobe 2018).

Location

The included studies were conducted in a variety of countries, including Austria, Belgium, Brazil, Croatia, Czech Republic, Denmark, France, Germany, Hungary, Italy, the Netherlands, Poland, Slovakia, Switzerland, the United Kingdom (UK), the United States of America (USA), and the People's Republic of China. Four studies reported the country but not the study centre (Annunziata 2016; Hutchings 2014; Markova 2012; Simon 2016), and two studies reported neither country nor study centre (Gandikota 2015; Straus 2011).

Participants

This review included a total of 7335 male and female consecutive participants who were newly diagnosed with classic HL and received first-line therapy. Out of these, a total of 2205 participants were included in meta-analyses.

Follow-up

There were differences in the follow-up time between studies. Three studies did not report follow-up time (Annunziata 2016; Orlacchio 2012; Straus 2011). Two studies reported follow-up time per subgroup, i.e. surviving participants only (Kobe 2018; Zaucha 2017). The median follow-up time for the remaining 18 studies ranged from 23 to 66 months. The total raw range of follow-up time was between two to 195 months.

Stages of disease

Fifteen studies included all stages of the disease. Four studies included only early stages (Andre 2017; Barnes 2011; Gandikota 2015, Straus 2011) and four studies only advanced stages (Casasnovas 2019; Kobe 2018; Markova 2012; Okosun 2012).

Treatment/therapy

The following chemotherapy regimens were administered.

- ABVD (adriamycin/doxorubicin, bleomycin, vinblastine and dacarbazine) in 16 studies (Andre 2017; Annunziata 2016; Barnes 2011; Cerci 2010; Gallamini 2014; Hutchings 2005; Hutchings 2006; Hutchings 2014; Mesguich 2016; Oki 2014; Okosun 2012; Orlacchio 2012; Simon 2016; Touati 2014; Zaucha 2017; Zinzani 2012).
- Either ABVD or BEACOPP in one study (Ying 2014).
- $BEACOPP_{escalated}$ (bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine and prednisone in escalated doses) in one trial (Casasnovas 2019).
- $BEACOPP_{escalated}$ or $BEACOPP_{escalated}$ with rituximab in one trial (Kobe 2018).
- $BEACOPP_{escalated}$ or time-condensed $BEACOPP14_{baseline}$ (BEACOPP in standard, non-escalated doses repeated on day 15) in one study (Markova 2012).

- AVG (doxorubicin, vinblastine and gemcitabine) in one trial (Straus 2011).
- ABV/MOPP (adriamycin, bleomycin, vinblastine, mechlorethamine, vincristine, procarbazine and prednisone), ABVD/COPP (ABVD plus cyclophosphamide, vincristine, procarbazine and prednisone), eBEACOPP, or PVAG (prednisone, vinblastine, doxorubicin and gemcitabine) in subgroups of participants in three studies (Hutchings 2005; Hutchings 2006; Touati 2014).
- Anthracycline-based chemotherapy not further specified in one study (Rossi 2014).

The following number of chemotherapy cycles were administered.

- Two, three, four, six or eight cycles of chemotherapy alone or combined with radiotherapy in 15 studies (Andre 2017; Annunziata 2016; Barnes 2011; Casasnovas 2019; Cerci 2010; Gallamini 2014; Hutchings 2014; Markova 2012; Mesguich 2016; Orlacchio 2012; Rossi 2014; Simon 2016; Straus 2011; Zaucha 2017; Zinzani 2012). The number of cycles usually depended on the stage of the disease.
- Four, six or eight cycles of chemotherapy, depending on the interim PET scan results, in one trial (Kobe 2018). A protocol amendment during the trial introduced a reduction of standard therapy from eight to six cycles.
- Six cycles of chemotherapy combined with antiretroviral therapy due to HIV-positive study population in one study (Okosun 2012).

Six studies did not report the number of cycles (Gandikota 2015; Hutchings 2005; Hutchings 2006; Oki 2014; Touati 2014; Ying 2014).

The following radiotherapy techniques were used either in all or a subgroup of participants.

- Involved-field radiotherapy in eight studies (Barnes 2011; Gallamini 2014; Hutchings 2005; Hutchings 2006; Hutchings 2014; Mesguich 2016; Rossi 2014; Simon 2016), and either involved-field radiotherapy or extended-field radiotherapy in one study (Gandikota 2015).
- Involved-node radiotherapy in three studies (Andre 2017; Annunziata 2016; Zaucha 2017).
- Involved-site radiotherapy in two studies (Touati 2014; Zinzani 2012).
- Radiotherapy without further specification in five studies (Cerci 2010; Kobe 2018; Markova 2012; Orlacchio 2012; Ying 2014).
- No radiotherapy in three studies (Oki 2014; Okosun 2012; Straus 2011).

Stem cell transplantation was conducted in participants who relapsed after first-line therapy despite treatment escalation or salvage therapy.

- Autologous stem cell transplantation in eight studies (Cerci 2010; Gallamini 2014; Hutchings 2014; Mesguich 2016; Touati 2014; Ying 2014; Zaucha 2017).
- Autologous and/or allogeneic stem cell transplantation in one study (Zinzani 2012).
- Type of stem cell transplantation not specified in four studies (Hutchings 2005; Hutchings 2006; Markova 2012; Orlacchio 2012).
- No stem cell transplantation reported in 10 studies (Andre 2017; Annunziata 2016; Barnes 2011; Gandikota 2015; Kobe 2018; Oki 2014; Okosun 2012; Rossi 2014; Simon 2016; Straus 2011).

Index (prognostic) factor

Participants in 16 out of 23 studies underwent PET combined with computed tomography (CT), contrast enhanced CT, or multi detector CT (MDCT), compared to PET-only for participants in the other studies. Participants in 13 studies underwent PET-CT (Annunziata 2016; Cerci 2010; Gallamini 2014; Gandikota 2015; Hutchings 2014; Kobe 2018; Mesguich 2016; Okosun 2012; Rossi 2014; Simon 2016; Touati 2014; Ying 2014; Zaucha 2017). Participants in another study underwent either PET or PET-CT (Barnes 2011); participants in one study underwent PET with contrast-enhanced CT (Markova 2012); and participants in another study underwent PET/MDCT (Orlacchio 2012). In the remaining seven studies, participants underwent a PET scan only (Andre 2017; Casasnovas 2019; Hutchings 2005; Hutchings 2006; Oki 2014; Straus 2011; Zinzani 2012).

Timing of interim PET

The timing of interim PET imaging varied between studies. In most studies, participants underwent an interim PET scan after two cycles (PET2) of chemotherapy (Andre 2017; Casasnovas 2019; Cerci 2010; Gallamini 2014; Hutchings 2005; Hutchings 2006; Kobe 2018; Mesguich 2016; Oki 2014; Okosun 2012; Orlacchio 2012; Rossi 2014; Simon 2016; Straus 2011; Touati 2014; Zinzani 2012). In another study, participants underwent an interim PET scan after the first cycle (PET1) of chemotherapy only (Annunziata 2016). In one study, participants underwent interim PET scans after the first and second cycle of chemotherapy, but the study protocol was amended after interim analysis to limit PET2 scans to participants with positive results after PET1 (Zaucha 2017). In one multi-centre study, participants from two centres underwent both PET1 and PET2, whereas participants from the remaining two centres underwent PET2 only if PET1 was positive (Hutchings 2014). Three retrospective studies included participants who underwent interim PET after two to four cycles of chemotherapy (Barnes 2011; Gandikota 2015; Ying 2014), and in another study participants underwent interim PET after four cycles (PET4) of chemotherapy (Markova 2012). For meta-analyses, we used information at PET2 whenever available in order to ensure homogeneity across studies.

Evaluation of PET scans

In most studies, two nuclear medicine physicians evaluated the PET scans individually, and disagreements in scoring were solved in a consensus meeting (Annunziata 2016; Barnes 2011; Cerci 2010; Hutchings 2005; Hutchings 2006; Hutchings 2014; Mesguich 2016; Orlacchio 2012; Rossi 2014; Ying 2014; Zinzani 2012). Evaluation of PET scans was performed by only one expert in one study (Markova 2012); and by a panel consisting of three to six experts in eight studies (Andre 2017; Casasnovas 2019; Gallamini 2014; Kobe 2018; Oki 2014; Okosun 2012; Straus 2011; Zaucha 2017). Three studies did not report the number or qualification of persons who performed evaluation of PET scans (Gandikota 2015; Simon 2016; Touati 2014). Nine out of 13 multi-centre studies reported that evaluation of PET scans took place centrally (Andre 2017; Gallamini 2014; Hutchings 2006; Kobe 2018; Mesguich 2016; Okosun 2012; Straus 2011; Zaucha 2017; Zinzani 2012), and two studies did not report how reviewing of PET scans was performed across centres (Barnes 2011; Hutchings 2014).

In 11 studies, outcome assessors were blinded to the outcome (Kobe 2018; Gallamini 2014; Gandikota 2015; Hutchings 2006; Hutchings 2014; Mesguich 2016; Oki 2014; Rossi 2014; Straus 2011; Zaucha 2017; Zinzani 2012). The remaining studies did not report blinding.

Criteria for evaluation

Most studies reported the use of a standardised scale for the evaluation of the PET scans, but the scoring systems and cut-off points between studies varied.

- In 12 studies, the Deauville 5-point scoring system for evaluation of PET scans was used: in nine studies, Deauville scores 1 - 3 were considered as PET-negative, and Deauville scores 4 - 5 as PET-positive (cut-off ≥4) (Annunziata 2016; Casasnovas 2019; Gallamini 2014; Hutchings 2014; Oki 2014; Okosun 2012; Rossi 2014; Simon 2016; Zaucha 2017); in two studies, both cut-off points for evaluation of the PET scans were used by scoring each image twice, and comparing performance of interim PET between both scales (Kobe 2018; Mesguich 2016); and in one study, it was reported that the PET scans were re-interpreted retrospectively using the Deauville criteria, but it was not indicated which cut-off points were used (Touati 2014).
- In one study, the International Harmonization Project criteria were used: a PET scan was considered positive when the residual mass is ≥ 2 cm or, if less than 2 cm, positive if its activity is above that of the surrounding background (Andre 2017). A negative PET scan corresponds to Deauville score 1 (no uptake) and score 2 (uptake ≤ mediastinum).
- In two studies, the scoring systems were not specified, but similar scales and cut-off points as the Deauville scoring system were used: in one study, PET scans were reviewed using a 4-point scale (Barnes 2011), and in another study using a 5-point scale (Gandikota 2015).
- In three studies, other standardised scales for the evaluation of PET scans were used: one study used the Juweid criteria (Zinzani 2012), and two studies used the International Harmonization Project guidelines (Orlacchio 2012; Straus 2011).
- Two studies did not report how PET scans were evaluated (Hutchings 2005; Hutchings 2006); and four studies reported performance of visual evaluation but did not indicate the use of a standardised scoring system (Cerci 2010; Markova 2012; Touati 2014; Ying 2014).

Outcomes

Primary outcome

Overall survival (OS)

Univariable analyses

Twelve out of 23 included studies reported unadjusted results for our primary outcome OS (Barnes 2011; Casasnovas 2019; Cerci 2010; Gallamini 2014; Hutchings 2005; Hutchings 2006; Hutchings 2014; Kobe 2018; Simon 2016; Touati 2014; Zaucha 2017; Zinzani 2012). Of these, nine provided sufficient information and data to be included in meta-analysis. One study reported an HR that we used (Kobe 2018). Another study reported an HR, but we still re-calculated it due to discrepancies in values between the graph and table (Simon 2016). For the other seven studies, we estimated the HR using other available data from the publications (Barnes 2011; Cerci 2010; Hutchings 2005; Hutchings 2014; Touati 2014; Zaucha 2017; Zinzani 2012).

Multivariable analyses

Two studies reported adjusted results for OS (Kobe 2018; Simon 2016). Two additional studies planned, but did not conduct the analysis for different reasons (Gallamini 2014; Hutchings 2005).

Secondary outcomes

Progression-free survival (PFS)

Univariable analyses

Twenty-one out of 23 studies reported unadjusted results for PFS (Andre 2017; Annunziata 2016; Barnes 2011; Casasnovas 2019; Cerci 2010; Gallamini 2014; Hutchings 2005; Hutchings 2006; Hutchings 2014; Kobe 2018; Markova 2012; Mesguich 2016; Oki 2014; Okosun 2012; Rossi 2014; Simon 2016; Straus 2011; Touati 2014; Ying 2014; Zaucha 2017; Zinzani 2012). Of these, 15 provided sufficient information and data to be included in meta-analysis. Three studies provided an HR which we used (Annunziata 2016; Kobe 2018; Simon 2016). Another three studies reported an HR, but we still re-calculated it due to unclear description of the statistical methods used (Hutchings 2006), reporting discrepancies between graphs and tables (Mesguich 2016) or general uncertainties in the reported values (Rossi 2014). For eight studies we estimated the HR using other available data (Barnes 2011; Cerci 2010; Hutchings 2005; Straus 2011; Touati 2014; Ying 2014; Zaucha 2017; Zinzani 2012).

Multivariable analyses

Eight studies reported adjusted results for PFS (Casasnovas 2019; Gallamini 2014; Hutchings 2005; Hutchings 2006; Kobe 2018; Mesguich 2016; Rossi 2014; Simon 2016). Three studies took the importance of adjustment into account, but did not actually conduct a multivariable analysis (Annunziata 2016; Hutchings 2014; Oki 2014).

Definitions of Progression-free survival (PFS)

The definition of the progression outcome varied between studies. Four studies that reported PFS did not provide a definition (Hutchings 2014; Simon 2016; Straus 2011; Zaucha 2017). One study analysed event-free survival (Cerci 2010), which was identical with PFS and, therefore, included in the analysis. Table 2 presents an overview of definitions used for progression outcome. Studies with identical definitions were grouped.

Table 2. Definitions of progression outcomes

Study	Definition of progression outcome
Andre 2017	Progression-free survival, defined – from the date of random assignment to date of progression – as experiencing relapse after previous complete remission or progression after reaching partial remission (50% decrease and resolution of B symptoms and no new lesions); progressive dis-

	ease (50% increase from nadir of any previous partial remission lesions or appearance of new lesions) on CT scan measurements during protocol treatment; or death from any cause, whichever occurred first.
Casasnovas 2019	Progression-free survival defined as the time from randomisation to first progression, relapse, or death from any cause or last follow-up.
Annunziata 2016	The primary endpoint was PFS, with progression during treatment, lack of complete remission at the end of the first-line treatment, and relapse counted as adverse events.
Barnes 2011; Ying 2014; Zinzani 2012	Progression-free survival is defined as the time from diagnosis to progression or death from any cause.
Kobe 2018	Progression-free survival is defined as the time from completion of staging until progression, relapse, or death from any cause, or to the day when information was last received on the patient's disease status.
Cerci 2010	Three-year event-free survival was chosen as the endpoint and defined as the time from diagnosis to treatment failure or last follow-up. Treatment failure was defined as an incomplete response after first-line treatment, progression during therapy, relapse, or death.
Gallamini 2014; Markova 2012; Mesguich 2016; Oki 2014;	Progression-free survival is defined as the time from diagnosis to either disease progression or relapse, or to death as a result of any cause, whichever occurred first.
Hutchings 2005; Hutchings 2006	Progression-free survival is defined as the time from diagnosis to first evidence of progression or relapse, or to disease-related death.
Okosun 2012	Progression-free survival is defined as the time from diagnosis to disease progression or relapse or last follow-up.
Rossi 2014	Progression-free survival is defined as the time from the beginning of treatment until progression, relapse, or death from any cause or the date of last follow-up. Time-to-progression (TTP) is defined as the time from the date of the first course of chemotherapy to any treatment failure, including progression, relapse, or death related to lymphoma, or the date of the last follow-up.
Touati 2014	Progression-free survival is defined as the time from diagnosis to relapse or death.
Hutchings 2014; Simon 2016; Straus 2011; Zaucha 2017	Definition not reported.

Adverse events (AEs)

None of the included studies measured PET-associated AEs.

Conflict of interest

Two studies reported potential conflicts of interest (Andre 2017; Casasnovas 2019). Fourteen studies declared that the investigators had no conflict of interest (Annunziata 2016; Barnes 2011; Hutchings 2006; Hutchings 2014; Kobe 2018; Mesguich 2016; Oki 2014; Okosun 2012; Orlacchio 2012; Rossi 2014; Simon 2016; Straus 2011; Zaucha 2017; Zinzani 2012). Seven studies did not report investigators' disclosures of potential conflicts of interest (Cerci 2010; Gallamini 2014; Gandikota 2015; Hutchings 2005; Markova 2012; Touati 2014; Ying 2014).

Excluded studies

After screening titles and abstracts, we excluded 10651 references that did not match our inclusion criteria. In addition, we excluded a total of 133 references after full-text screening for the following reasons.

- Fifty-six references had a study design or publication type that did not match our inclusion criteria, i.e. letters and commentaries, case studies with a small sample size or

validation studies (Adams 2016; Adams 2017; Adams 2018; Adams 2018a; Adams 2018b; Adams 2019; Afanasyev 2017; Ansell 2016; Barrington 2017; Bar-Shalom 2003; Basu 2009; Becherer 2002; Bednaruk-Mlynski 2015; Biggi 2012; Bishop 2015; Bodet-Milin 2009; Boisson 2007; Borchmann 2016; Bucerius 2006; Cremerius 1999; D'Urso 2018; Dann 2018; deAndres-Galiana 2015; Diehl 2007; El-Galaly 2012; Evens 2014; Fanti 2008; Friedberg 2002; Friedberg 2004; Gallamini 2008; Gallamini 2018a; Gallowitsch 2008; Guidez 2016; Hagtvedt 2015; Hartmann 2012; Hartridge-Lambert 2013; Kobe 2008; Kobe 2014; Lowe 2002; Milgrom 2017; Mocikova 2010; NCT02292979; Pichler 2000; Reinhardt 2005; Rigacci 2002; Rigacci 2017; Rubello 2015; Sakr 2017; Specht 2007; Spinner 2018; Strigari 2016; Tirelli 2015; Xie 2018; Yasgur 2015; Zabrocka 2016; Zaucha 2009).

- Thirty-nine references adapted the treatment based on PET-results (Albano 2017; Albano 2018; Biggi 2017; Carras 2018; Ciammella 2016; Cuccaro 2016; Damlaj 2017; Damlaj 2019; Danilov 2017; Dann 2009; Dann 2010; Dann 2010a; Dann 2012; Dann 2013; Dann 2016; Dann 2017; Fornecker 2017; Gallamini 2017; Gallamini 2018; Greil 2018; Illidge 2015; Johnson 2015; Johnson 2016; Kamran 2016; Kamran 2018; Moskowitz 2015; NCT00784537; NCT00795613; NCT01358747; NCT01652261; Nguyen 2017; Paolini 2007; Pavlovsky 2019; Simontacchi 2015; Straus 2018; Torizuka 2004; Trotman 2017; Villa 2018; Zinzani 2016).
- Eighteen references also included participants with other types of lymphoma and did not report data for HL separately (Awan 2013; Blum 2002; Bodet-Milin 2008; Cremerius 2001; Filmont 2003; Freudenberg 2004; Fruchart 2006; Goldschmidt 2011; Haioun 2005; Honda 2014; Iagaru 2008; Kostakoglu 2006; Li 2013; Slaby 2002; Tomita 2015; Torizuka 2004; Zinzani 1999; Zinzani 2002).
- Ten references included participants who received treatment other than first-line therapy, i.e. second-line therapy for relapsed or refractory disease (Bjurberg 2006; Front 1999; Huic 2006; Mocikova 2010; Mocikova 2011; Schot 2007; Sucak 2011; Tseng 2012; Weidmann 1999; Yoshimi 2008).
- Eight references reported only end-of-chemotherapy PET-results (Advani 2007; Hueltenschmidt 2001; Hutchings 2007; Jerusalem 2003; Molnar 2010; Naumann 2001; Panizo 2004; Spaepen 2001).
- Two were duplicates (Freudenberg 2004; Kobe 2014).

These publications are described in Characteristics of excluded studies.

Risk of bias in included studies

We assessed the risk of bias at outcome level (OS and PFS) for each study using the QUIPS tool. No study reported PET-associated AE. The detailed assessment can be found in the 'Risk of bias (QUIPS)' section in the Characteristics of included studies.

Risk of bias in studies included in meta-analyses

The 'Risk of bias' summary (Figure 2) presents the combined judgement made by the review authors in a cross-tabulation. Studies included in meta-analysis are highlighted in bold.

Figure 2. 'Risk of bias' assessment according to QUIPS (Quality in Prognostic Studies) by outcome.

Outcome	Study	Quality in Prognostic Studies (QUIPS)					
		Study participation	Study attrition	Prognostic factor measurement	Outcome measurement	Other prognostic factors (covariates)	Statistical analysis and reporting
Overall survival	**Barnes 2011**	Unclear	Low	Moderate	High	Low	High
	Casasnovas 2019	Low	Low	Low	Low	Low	Low
	Cerci 2010	Low	Low	Low	Low	High	High
	Gallamini 2014	Low	Low	Low	Low	Low	Low
	Hutchings 2005	Unclear	Moderate	Low	Low	High	Low
	Hutchings 2006	High	Low	Low	Low	High	Low
	Hutchings 2014	Low	Low	Low	Low	High	Low
	Kobe 2018	Low	Low	Low	Low	Low	Low
	Simon 2016	Unclear	Low	Low	Low	High	High
	Touati 2014	Unclear	Low	Moderate	Low	High	Low
	Zaucha 2017	Low	Low	Moderate	Low	High	Low
	Zinzani 2011	Low	Low	Low	Low	High	Low
Progression-free survival	Andre 2017	Low	Low	Moderate	Low	Low	Low
	Annunziata 2016	Unclear	Unclear	Low	Low	High	High
	Barnes 2011	Unclear	Low	Moderate	High	Low	High
	Casasnovas 2019	Low	Low	Low	Low	Low	Low
	Cerci 2010	Low	Low	Low	Low	High	High
	Gallamini 2014	Low	Low	Low	Low	Low	Low
	Hutchings 2005	Unclear	Moderate	Low	Low	Low	Low
	Hutchings 2006	High	Low	Low	Low	Low	Low
	Hutchings 2014	Low	Low	Low	Low	High	Low
	Kobe 2018	Low	Low	Low	Low	Low	Low
	Markova 2012	Low	Low	Moderate	Low	Low	Low
	Mesguich 2016	Low	Low	Low	Low	Low	Low
	Oki 2014	Low	Low	Low	Low	High	High
	Okosun 2012	Low	Low	Low	Low	Low	High
	Rossi 2014	Low	Low	Low	Low	High	Low
	Simon 2016	Unclear	Low	Low	Low	High	High
	Straus 2011	Low	Low	Low	Low	Low	Low
	Touati 2014	Unclear	Low	Moderate	Low	High	Low
	Ying 2014	Low	Low	Moderate	Low	High	High
	Zaucha 2017	Low	Low	Moderate	High	High	High
	Zinzani 2011	Low	Low	Low	Low	High	Low

Overall survival (OS)

For our primary outcome OS, one out of nine studies included in meta-analysis was assessed as 'low' in all risk of bias domains (Kobe 2018). Four studies were assessed as 'unclear' for the domain study participation (Barnes 2011; Hutchings 2005; Simon 2016; Touati 2014), mostly due to a lack of information about the baseline population from which the study sample originated. Most studies had defined exclusion criteria to sample participants from the baseline population (e.g. unavailability of interim PET2) without providing a description of the original population or reasons for missing information. Considering this a potential source of selection bias, we assessed this domain as 'unclear' when information about the baseline population was missing. For the domains study attrition, prognostic factor measurement and outcome measurement, risk of bias was assessed as 'low' in most studies. Two studies did not report the use of standardised criteria for prognostic factor measurement, therefore we assessed the risk of bias as 'moderate' (Barnes 2011; Touati 2014). One study was assessed as 'moderate' risk because PET2 availability was dependent on PET1 result (Zaucha 2017). Due to inconsistency in reporting of the timing of the interim PET measurement, the risk of bias for outcome measurement was assessed as 'high' in one study (Barnes 2011), while the remaining studies were all assessed as 'low'. Two studies were assessed as 'low' risk of bias in the domain other prognostic factors (covariates) because they only included participants within one disease stage (e.g. early or advanced stages) (Barnes 2011; Kobe 2018), while the remaining seven studies were assessed as 'high' risk of bias for this domain because they included all disease stages without adjusting for stage (Cerci 2010; Hutchings 2005; Hutchings 2014; Simon 2016; Touati 2014; Zaucha 2017; Zinzani 2012). Six studies provided sufficient information about the methods used for univariable analysis (Hutchings 2005; Hutchings 2014; Kobe 2018; Touati 2014; Zaucha 2017; Zinzani 2012), therefore we assessed the risk of bias for statistical analysis and reporting as 'low'. The same domain was assessed as 'high' in three studies due to discrepancies between text and figures and/or tables (Barnes 2011; Cerci 2010; Simon 2016).

Progression-free survival (PFS)

For our secondary outcome PFS, two out of 14 studies included in meta-analysis were assessed as 'low' risk of bias in all domains (Casasnovas 2019; Mesguich 2016). Eight studies provided clear descriptions of study characteristics and participants (Cerci 2010; Kobe 2018; Mesguich 2016; Rossi 2014; Straus 2011; Ying 2014; Zaucha 2017; Zinzani 2012), so we assessed the risk of bias as 'low'. Five studies did not report inclusion and/or exclusion criteria (Annunziata 2016; Barnes 2011; Hutchings 2005; Simon 2016; Touati 2014), so we assessed the risk of bias for study participation as 'unclear'. One study reported a high number of participants with unavailable interim PET scans without further information (Hutchings 2006), so we assessed the risk of bias as 'high' in the same domain. Most studies had no loss to follow-up to report or

provided a clear description of how missing data were handled, so we assessed the risk of bias for study attrition as 'low' in the majority of studies. One study was assessed as 'unclear' due to a lack of information regarding loss to follow-up (Annunziata 2016); another study was assessed as 'moderate' because no explanation was provided as to why some participants were lost to follow-up (Hutchings 2005). The risk of bias for the domains prognostic factor measurement and outcome measurement was assessed as 'low' in most studies. Three studies did not report the use of standardised criteria for prognostic factor measurement, therefore we assessed the risk of bias as 'moderate' (Barnes 2011; Touati 2014; Ying 2014). A fourth study was assessed as 'moderate' risk because PET2 availability was dependent on PET1 result (Zaucha 2017). Due to lack of outcome definition or inconsistency in the reporting of the timing of the interim PET measurement, the risk of bias for outcome measurement was assessed as 'high' in one study (Barnes 2011). In another study, this domain was also assessed as 'high' because the outcome was not defined (Zaucha 2017). The remaining studies were all assessed as 'low' for the domain outcome measurement. For the domain other prognostic factors (covariates), six studies were assessed as 'low' risk of bias, because they either included participants within one disease stage only, or if all disease stages were included, the authors adjusted for disease stage (Barnes 2011; Hutchings 2005; Hutchings 2006; Kobe 2018; Mesguich 2016; Straus 2011). The remaining eight studies were assessed as 'high' risk of bias for this domain (Annunziata 2016; Cerci 2010; Rossi 2014; Simon 2016; Touati 2014; Ying 2014; Zaucha 2017; Zinzani 2012). Eight studies provided sufficient information about the methods used for univariable analysis (Hutchings 2005; Hutchings 2006; Kobe 2018; Mesguich 2016; Rossi 2014; Straus 2011; Touati 2014; Zinzani 2012), so we assessed the risk of bias for statistical analysis and reporting as 'low'. Five studies were assessed as 'high' for this domain because of the poor reporting of results (Annunziata 2016; Barnes 2011; Cerci 2010; Simon 2016; Ying 2014), including discrepancies between text and figures and/or tables in some studies. Another study was also assessed as 'high' because the method of analysis was not sufficiently described (Zaucha 2017).

Risk of bias in studies reported narratively

The risk of bias for all studies reported narratively is included in Figure 2.

Overall survival (OS)

The results for OS from three studies are reported narratively in this review (Casasnovas 2019; Gallamini 2014; Hutchings 2006). For two studies (Casasnovas 2019; Gallamini 2014) we assessed the risk of bias as 'low' in all six domains of the QUIPS tool. For Hutchings 2006, the first four domains were assessed as 'low' risk of bias. For the domain study participation, the study was assessed as 'high' risk because a great number of participants initially included in the study did not undergo an early interim PET. The study was also assessed as 'high' risk for the domain other prognostic factors (covariates) because participants within all disease stages were included.

Progression-free survival (PFS)

For PFS, the results from seven studies are reported narratively (Andre 2017; Casasnovas 2019; Gallamini 2014; Hutchings 2014; Markova 2012; Oki 2014; Okosun 2012). Out of these, two studies (Casasnovas 2019; Gallamini 2014) were assessed as 'low' risk of bias in all six domains of the QUIPS tool. From the remaining five studies, all were assessed as 'low' risk of bias for the domain study participation. For the domains study attrition, prognostic factor measurement and outcome measurement, three studies were assessed as a 'low' risk of bias (Hutchings 2014; Oki 2014; Okosun 2012). For the other two studies (Andre 2017; Markova 2012), the domain prognostic factor measurement was assessed as 'moderate' risk because the prognostic factor was measured differently in some participants. For the domain other prognostic factors (covariates), five studies were assessed as 'low' risk of bias (Andre 2017; Casasnovas 2019; Gallamini 2014; Markova 2012; Okosun 2012). The other two studies were assessed as 'high' risk for this domain because they included all disease stages without adjusting for disease stage (Hutchings 2014; Oki 2014). Regarding the domain statistical reporting and analysis, five studies were assessed as 'low' risk because they used appropriate methods for the planned analysis (Andre 2017; Casasnovas 2019; Gallamini 2014; Hutchings 2014; Markova 2012). The remaining two studies were assessed as 'high' risk due to inconsistent conduct and reporting of the analyses (Oki 2014; Okosun 2012).

Other potential sources of bias

Reporting deficiencies and selective reporting

We detected reporting deficiencies in some of the studies, particularly when not all analyses that were planned in the methods were actually conducted. In some cases, this was due to the low number of events (i.e. in PET-negative participants) that did not allow for further analyses. In other cases, it was unclear why certain analyses were performed and others not. This was particularly the case with regard to multivariable analyses, when studies planned to assess the independent prognostic ability of the interim PET in a prognostic model including other clinically relevant prognostic factors (covariates). Studies either did not perform such an analysis even though they initially planned to, or they did not consider adjustment. None of the studies stated clearly their rationale for the choice of covariates; in some cases, the choice was based on their significance in univariable analysis. For example, in studies that only included two or less covariates in the model in addition to interim PET, the interim PET was always independent in its performance. However, how interim PET possibly performed in comparison to other covariates remains unclear. Hence, it is particularly important to state why certain covariates were taken into account. Thus, we cannot be sure that studies did not only report certain positive ('significant') results, which can be an issue of selective reporting.

In addition, we detected discrepancies in the reporting of results within the texts of some studies, or between text and the corresponding graph(s) (i.e. in the reporting of the HR or number of events). In these cases, we tried to contact the corresponding principal investigator(s) for clarification in order to have a better understanding of the results.

Blinding of prognostic factor assessor

Eleven studies reported that the clinicians evaluating the interim PET scans were blinded to the outcome (Kobe 2018; Gallamini 2014; Gandikota 2015; Hutchings 2006; Hutchings 2014; Mesguich 2016; Oki 2014; Rossi 2014; Straus 2011; Zaucha 2017; Zinzani 2012).

Results of the analyses

Twenty-three studies evaluated interim PET as a prognostic factor in individuals with HL. Two studies did not report data for our outcomes of interest (Gandikota 2015; Orlacchio 2012) and we have not been able to either obtain or estimate any relevant data. None of the included studies reported PET-associated AEs. Fifteen studies were included in meta-analyses. Another six of the included studies in this review reported results for OS and/or PFS, but we were not able to pool results because, despite our approaches for possible estimation of missing data items, there was a lack of accurate information or data to do so (Andre 2017; Casasnovas 2019; Gallamini 2014; Markova 2012; Oki 2014; Okosun 2012). For all studies that were not included in meta-analyses, we reported the main results narratively in this review.

Overall survival (OS)

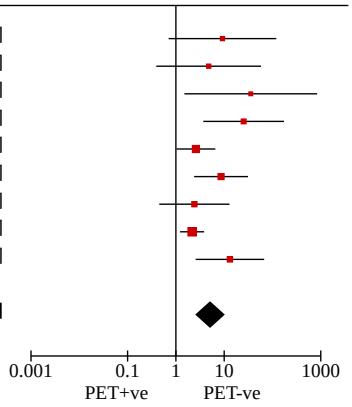
Meta-analysis of unadjusted results

We included nine studies with 1802 participants in meta-analysis for OS (Barnes 2011; Cerci 2010; Hutchings 2005; Hutchings 2014; Kobe 2018; Simon 2016; Touati 2014; Zaucha 2017; Zinzani 2012). There were 475 interim PET-positive and 1327 interim PET-negative participants. Meta-analysis shows a clear advantage in OS for participants with a negative interim PET scan compared to participants with a positive interim PET scan (HR 5.09, 95% CI 2.64 to 9.81, I^2 44%, moderate certainty of evidence) (Analysis 1.1) (Figure 3).

Figure 3. Forest plot of comparison: 1 Univariable comparison of PET+ve vs. PET-ve, outcome: 1.1 Overall survival

Study or Subgroup	log[Hazard Ratio]	SE	PET+ve Total	PET-ve Total	Weight	Hazard Ratio IV, Random, 95% CI
Barnes 2011	2.220623	1.309795	17	79	5.3%	9.21 [0.71 , 120.03]
Cerci 2010	1.563053	1.274372	30	74	5.6%	4.77 [0.39 , 58.02]
Hutchings 2005	3.570366	1.614442	22	63	3.7%	35.53 [1.50 , 841.02]
Hutchings 2014	3.231135	0.981951	37	89	8.3%	25.31 [3.69 , 173.42]
Kobe 2018	0.9555	0.4724	236	486	18.6%	2.60 [1.03 , 6.56]
Simon 2016	2.153725	0.654523	32	89	13.9%	8.62 [2.39 , 31.08]
Touati 2014	0.881751	0.854282	24	44	10.1%	2.42 [0.45 , 12.89]
Zaucha 2017	0.774	0.2928	24	152	24.1%	2.17 [1.22 , 3.85]
Zinzani 2012	2.575668	0.833487	53	251	10.4%	13.14 [2.57 , 67.31]
Total (95% CI)			**475**	**1327**	**100.0%**	**5.09 [2.64 , 9.81]**

Heterogeneity: $Tau^2 = 0.38$; $Chi^2 = 14.39$, df = 8 (P = 0.07); $I^2 = 44\%$
Test for overall effect: Z = 4.87 (P < 0.00001)
Test for subgroup differences: Not applicable



Subgroup analysis

We conducted subgroup analyses to explore the underlying clinical heterogeneity between the studies.

For subgroup analysis by radiotherapy, we found evidence on subgroup difference between the groups (P = 0.05, INRT/ISRT in three studies: N = 548, IFRT in four studies: N = 428, RT not further specified in two studies: N = 826). Results still show an advantage in OS for PET-negative participants, irrespective of the type of radiotherapy they received (Analysis 2.1).

For the remaining subgroups, there was no evidence of subgroup differences.

- Different study designs (P = 0.28; three prospective studies: N = 406, four retrospective studies: N = 589, one RCT: N = 722) (Analysis 2.2). One study (Hutchings 2005) was not included in this subgroup analysis because they did not explicitly state their study design.
- Different chemotherapy regimens (P = 0.33; ABVD in five studies: N = 801, ABVD and other in three studies: N = 279, BEACOPP in one study: N = 722) (Analysis 2.3). Chemotherapy-regimen in the included studies was mainly ABVD, with differentiating numbers of cycles, with or without radiotherapy (Barnes 2011; Cerci 2010; Hutchings 2014; Simon 2016; Zaucha 2017; Zinzani 2012). In Hutchings 2005, the majority of participants received ABVD, while the remaining received MOPP or MOPP/ABV, or another regimen which was not specified. Some participants also received additional radiotherapy. In Kobe 2018, all participants received eBEACOPP. In Touati 2014, the regimens included ABVD, MOPP/ABV hybrid or BEACOPP. If separate data had been available for each type of chemotherapy, we could have performed more specific subgroup analysis to test for differences between chemotherapies.
- PET-CT versus PET (P = 0.66; PET-CT in five studies: N = 595, PET only in three studies: N = 1111) (Analysis 2.4). One study (Barnes 2011) was not included in this subgroup analysis because they conducted PET in some participants and PET-CT in the other participants.
- Different stages of disease (P = 0.33; early stages with A or B symptoms in one study: N = 96, all stages in seven studies: N = 984, advanced stages in one study: N = 722) (Analysis 2.5). One study included disease stages IA, IB, IIA and IIB (Barnes 2011) and another study included advanced-stages only (Kobe 2018). The remaining seven studies included participants representing all disease stages of HL.

Sensitivity analysis

We conducted sensitivity analyses for the timing of interim PET (removing those that did not conduct a PET2), and the precision of

the estimated HR and SE (removing the studies with imprecise HR and SE estimation).

Regarding the timing of the interim PET, interim PET2 was conducted in six studies (N = 1495 participants in total) (Cerci 2010; Kobe 2018; Simon 2016; Touati 2014; Zinzani 2012; Zaucha 2017). In three studies (N = 307 participants in total), interim PET was conducted at other timings: in Barnes 2011, 41 participants received PET2 while the rest of the participants received PET3; in Hutchings 2005, 55 participants received PET2 and 35 participants received PET3; and in Hutchings 2014, PET1 was conducted for all participants (N = 126). Although 89 out of 126 also received a PET2, we used the data for PET1 as the publication provided us with the most information on PET1. At sensitivity analysis, temporarily removing studies that did not perform a PET2 slightly affected the pooled OS (overall: HR 5.09, 95% CI 2.64 to 9.81; sensitivity: HR 3.53, 95% CI 1.97 to 6.32) (Analysis 2.6). It seems that there was an overestimation of the HR for the studies that did not perform a PET2. However, the direction of the effect is firm and unchanged. This difference may also be partly explained by the very wide follow-up ranges within the studies. Hence, following the sensitivity analysis, we consider the overall OS to be robust.

Regarding the precision of the HR estimation, we were able to either obtain or estimate a precise HR and SE for seven studies (N = 1638 participants in total) (Cerci 2010; Hutchings 2005; Hutchings 2014; Kobe 2018; Simon 2016; Zaucha 2017; Zinzani 2012). For two studies (N = 164 participants in total) (Barnes 2011; Touati 2014), we were only able to provide imprecise estimations of the HR and SE. Temporarily removing the imprecise studies during sensitivity analysis barely affected the pooled results for OS, indicating that the measurements obtained from our imprecise method were quite accurate after all (overall: HR 5.09, 95% CI 2.64 to 9.81; sensitivity: HR 5.70, 95% CI 2.60 to 12.48) (Analysis 2.7). Hence, we concluded that the overall pooled OS is robust.

Narrative reporting of results

Univariable analyses

Three studies (Casasnovas 2019; Gallamini 2014; Hutchings 2006) that reported results for OS were not included in meta-analysis due to lack of adequate data for estimating the HR and associated SE (Table 3).

Table 3. Narrative reporting of results from univariable analysis for OS

Study	No. of participants + stages	Timing of interim PET scan	Unadjusted results for interim PET scan
Casasnovas 2019	Standard arm N = 413	PET2	PET2 results (N = 398) ***Intention-to-treat analysis*** 5-year OS for entire arm = 95·2% (95% CI 91·1 to 97·4), 13 events ***Per-procotol analysis*** N = 372 participants 5-year OS for entire arm = 95.6% (95% CI 91.2 to 97.8), 10 events **Comment:** Separate results for PET2-negative and PET2-positive participants in the standard arm were not reported for this outcome.
Gallamini 2014	260 (stages IIA - IVB)	PET2	PET-negative N = 215, 2 deaths, 3-year OS = 99% PET-positive N = 45, 6 deaths, 3-year OS = 87% **Comment:** Logrank test for difference between groups was not reported and could not be obtained.
Hutchings 2006	77 (all stages)	PET2 and PET4	PET2 results (N = 77) PET-negative N = 61, no deaths PET-positive N = 16, 2 deaths Logrank test for difference between groups: P < .01 PET4 results (N = 64) PET-negative N = 51, no deaths PET-positive N = 13, 2 deaths

Comment: Logrank test for difference between groups after PET4 was not reported and could not be obtained.

Multivariable analyses

Two studies (Kobe 2018; Simon 2016) reported adjusted effect estimates to test the prognostic ability of PET2 in addition to other prognostic factors. Table 4 displays a list of established prognostic factors (Cuccaro 2014; Josting 2010; Kılıçkap 2013), and shows which were considered as covariates in the final multivariable model. The selection of prognostic factors (covariates) for the final model was either based on the literature (Simon 2016), or on their significance in univariable analysis (Kobe 2018). However, pooling of adjusted data was not possible. In Simon 2016, only the results of those covariates that remained independent prognostic markers in multivariable analysis, namely LMR and PET2-positivity, were reported. It is unclear whether, or which other covariates were included in the final model. A full list of study-specific, candidate covariates can be found in the respective table for each study in the Characteristics of included studies.

The statistical methods used were Cox proportional hazards regression model and logistic regression model, which are the appropriate methods for a multivariable analysis.

Table 4. Adjusted results from final multivariable model for OS

Study	Prognostic factors								Adjusted results for interim PET
	Interim PET	Age	Gender	Disease stage	B symptoms	Bulky disease	IPS	Other study-specific factors	
Kobe 2018	x	-	-	-	-	-	x	x	Interim PET-positivity (DS 4) HR 3.2 (95% CI 1.3 to 8.4), P = 0.02 **Comment:** Adjusted results indicate an independent prognostic impact of PET2.
Simon 2016	x	-	-	-	-	-	-	x	Interim PET-positivity HR = 11.51 (95% CI 3.14 to 42.86), P < 0.001 **Comment:** Adjusted results indicate the independent prognostic impact of PET2.

x = prognostic factor considered for adjustment in the final model

- = prognostic factor was not considered in the final model

Progression-free survival (PFS)

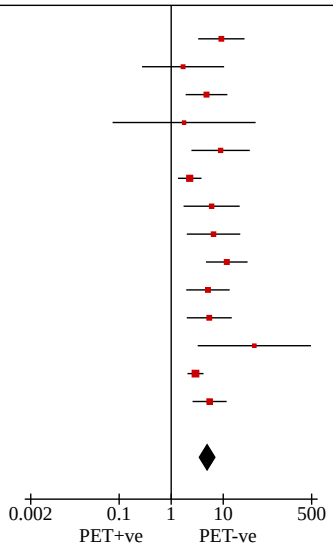
Meta-analysis of unadjusted results

We included 14 studies with 2079 participants in meta-analysis for PFS (Annunziata 2016; Barnes 2011; Cerci 2010; Hutchings 2005; Hutchings 2006; Kobe 2018; Mesguich 2016; Rossi 2014; Simon 2016; Straus 2011; Touati 2014; Ying 2014; Zaucha 2017; Zinzani 2012). There were 529 interim PET-positive and 1550 interim PET-negative participants. Meta-analysis shows a clear advantage in PFS for participants with a negative interim PET scan compared to participants with a positive interim PET scan (HR 4.90, 95% CI 3.47, 6.90, I^2 = 45%, very low certainty of evidence) (Analysis 1.2) (Figure 4).

Figure 4. Forest plot of comparison: 1 Univariable comparison of PET+ve vs. PET-ve, outcome: 1.2 Progression-free survival

Study or Subgroup	log[Hazard Ratio]	SE	PET+ve Total	PET-ve Total	Weight	Hazard Ratio IV, Random, 95% CI
Annunziata 2016	2.2192	0.5231	12	56	7.1%	9.20 [3.30 , 25.65]
Barnes 2011	0.5261	0.9261	17	79	3.0%	1.69 [0.28 , 10.39]
Cerci 2010	1.5624	0.4706	30	74	8.1%	4.77 [1.90 , 12.00]
Hutchings 2005	0.570366	1.6144	22	63	1.1%	1.77 [0.07 , 41.87]
Hutchings 2006	2.187	0.6587	16	61	5.2%	8.91 [2.45 , 32.40]
Kobe 2018	0.8198	0.2651	236	486	13.4%	2.27 [1.35 , 3.82]
Mesguich 2016	1.7891	0.6333	16	60	5.5%	5.98 [1.73 , 20.70]
Rossi 2014	1.873	0.6031	13	46	5.9%	6.51 [2.00 , 21.22]
Simon 2016	2.4596	0.4697	32	89	8.1%	11.70 [4.66 , 29.38]
Straus 2011	1.6268	0.49	24	64	7.7%	5.09 [1.95 , 13.29]
Touati 2014	1.685	0.5075	24	44	7.4%	5.39 [1.99 , 14.58]
Ying 2014	3.6783	1.278	10	25	1.7%	39.58 [3.23 , 484.51]
Zaucha 2017	1.0753	0.1812	24	152	16.0%	2.93 [2.05 , 4.18]
Zinzani 2012	1.6982	0.3845	53	251	10.0%	5.46 [2.57 , 11.61]
Total (95% CI)			**529**	**1550**	**100.0%**	**4.90 [3.47 , 6.90]**

Heterogeneity: Tau² = 0.16; Chi² = 23.52, df = 13 (P = 0.04); I² = 45%
Test for overall effect: Z = 9.07 (P < 0.00001)
Test for subgroup differences: Not applicable



Subgroup analysis

We conducted subgroup analyses to explore the underlying clinical heterogeneity between the studies.

Regarding the disease stage, we detected a significant difference between the groups (P = 0.02, early stages with A or B symptoms in two studies: N = 184, all stages in eleven studies: N = 1173, advanced stages in one study: N = 722). Results still showed an advantage for PFS in PET-negative participants in any stage of the disease (Analysis 3.4). Twelve studies included all disease stages, while one study included stages IA - IIB (Barnes 2011), and another study included advanced-stages only (Kobe 2018).

For the remaining subgroups, there was no evidence of subgroup differences.

- Different study designs (P = 0.29, three prospective studies: N = 357, eight retrospective studies: N = 827, two RCTs: N = 165) (Analysis 3.1). One study (Hutchings 2005) was not included in this subgroup analysis because they did not explicitly state their study design.
- Different chemotherapy regimen (P = 0.43; ABVD in seven studies: N = 945, ABVD and other chemotherapy in four studies: N = 265, other chemotherapies in three studies: N = 869) (Analysis 3.2). Chemotherapy-regimen was ABVD in seven studies, with or without radiotherapy (Annunziata 2016; Barnes 2011; Cerci 2010; Mesguich 2016; Simon 2016; Zaucha 2017; Zinzani 2012). In two studies, participants received either ABVD, ABV/MOPP, ABVD/COPP, BEACOPP esc., PVAG or radiotherapy only (Hutchings 2005; Hutchings 2006). In Touati 2014, the regimens included ABVD, MOPP/ABV hybrid or BEACOPP. In Ying 2014, participants received either ABVD or BEACOPP. In Kobe 2018, all participants received eBEACOPP. In Rossi 2014, all participants received anthracycline-based chemotherapy, and in Straus 2011, all participants received AVG.
- PET versus PET-CT (P = 0.30; PET-CT in eight studies: N = 707, PET only in five studies: N = 1276) (Analysis 3.3). One study (Barnes 2011) was not included in this analysis because they conducted PET in some participants and PET-CT in the other participants.
- Different radiotherapy (P = 0.29; INRT/ISRT in five studies: N = 651, IFRT in six studies: N = 514, RT not specified in two studies: N = 826, no RT given in one study: N = 88) (Analysis 3.5).

In addition, we detected variations between the studies with regard to the definition of PFS. However, all trials included in meta-analysis reported some progression endpoint such as treatment failure, progression or relapse. We have provided the exact reported definitions in Table 2.

Sensitivity analysis

Regarding the timing of interim PET, interim PET was conducted after two cycles of chemotherapy (PET2) in nine studies (N = 1677 participants in total) (Cerci 2010; Hutchings 2006; Kobe 2018; Rossi 2014; Simon 2016; Straus 2011; Touati 2014; Zaucha

2017; Zinzani 2012). In five studies (N = 402 participants in total), interim PET was conducted at other timings: in Annunziata 2016 all participants received PET1; in Barnes 2011 and Hutchings 2005 participants received either PET2 or PET3; and in Hutchings 2006 and Mesguich 2016 participants received either a PET2, PET3 or PET4. At sensitivity analysis, temporarily removing studies that did not perform a PET2 barely affected the results for PFS (overall: HR 4.90, 95% CI 3.47 to 6.90; sensitivity: HR 4.68, 95% CI 3.14 to 6.98) (Analysis 3.6). Hence, the timing of the interim PET measurement (when conducted at a time point other than PET2) did not affect the overall pooled result for PFS.

Regarding the precision of the HR estimation, we were able to provide a precise estimation of the HR and SE for nine studies (N = 1450 participants in total) (Annunziata 2016; Barnes 2011; Hutchings 2005; Kobe 2018; Rossi 2014; Simon 2016; Straus 2011; Ying 2014; Zaucha 2017). For five studies (N = 629 participants in total) we were only able to provide a slightly imprecise estimation of the HR and SE (Cerci 2010; Hutchings 2006; Mesguich 2016; Touati 2014; Zinzani 2012). However, at sensitivity analysis we found that the imprecise HRs did not significantly affect the pooled results. Temporarily removing the imprecise studies during sensitivity analysis barely affected the pooled results (overall: HR 4.90, 95% CI 3.47 to 6.90; sensitivity: HR 4.69, 95% CI 2.84 to 7.73) (Analysis 3.7). Hence, we concluded that the overall pooled PFS is robust and was not affected by our slightly imprecise method of HR and SE estimation.

Narrative reporting of results

Univariable analyses

Seven studies that reported results for PFS were not included in meta-analysis (Andre 2017; Casasnovas 2019; Gallamini 2014; Hutchings 2014; Markova 2012; Oki 2014; Okosun 2012). Table 5 presents the results from these studies narratively. We extracted all data that were available and relevant to us (i.e. number of interim PET-negative and interim PET-positive participants, number of events and percentages for PFS). Due to strong differences in the reporting between studies, the table presents more information for some studies compared to others.

Table 5. Narrative reporting of results from univariable analysis for PFS

Study	No. of participants analysed	Timing of interim PET scan	Unadjusted results for interim PET
Andre 2017	Favourable: N = 371 standard arm Unfavourable: N = 583 standard arm	PET2	*PET-negative Favourable group: N = 2 events (both relapses) in the ABVD + INRT arm, ITT 5-year PFS rate was 99.0% (95% CI 3.8 to 66.1) Unfavourable group: N = 22 events (16 relapses and 6 deaths not related to HL), ITT 5-year PFS rate was 92.1% (95% CI 88.0 to 94.8) *Results presented here are only for participants without interim PET adaptation (ABVD + INRT arm). Unclear how many of these participants were PET-positive or PET-negative. In total (all participants included in the study), there were 465 PET-negative participants and 361 PET-positive participants. *PET-positive N = 41 events (36 relapses and 5 deaths not related to HL) in the ABVD + INRT arm, ITT 5-year PFS rate was 77.4% (95% CI 70.4 to 82.9)
Casasnovas 2019	Standard arm N = 413	PET2	PET2 results (N = 398) ***Intention-to-treat analysis*** PET2-negative N = 349 participants (88%), 5-year PFS = 88.4% (95% CI 83.3 to 92) PET2-positive N = 49 participants (12%), 5-year PFS = 73.5% (95% CI 58.7 to 83.6) Results for entire standard arm 5-year PFS = 86.2% (95% CI 81.6 to 89.8) 41 participants relapsed or progressed, 14 deaths

			Comment: Logrank test for difference between groups in the standard arm after PET2 was not reported and could not be obtained. ***Per-protocol analysis*** N = 372 participants 5-year PFS = 86.7% (95% CI 81.9 to 90.3) for entire arm
Gallamini 2014	260 (stages IIA - IV)	PET2	PET-negative N = 215, 12 events (progression N = 7, relapse N = 5), 3-year PFS = 95% PET-positive N = 45, 33 events (progression N = 27, relapse N = 6), 3-year PFS = 28% Logrank test for difference between groups: P < 0.0001
Hutchings 2014	121 (all stages)	PET1 (N = 121) PET 2 (N = 89)	PET1 results (N = 126) PET-negative N = 89, 5 events (relapse), 2-year PFS = 94.1% PET-positive N = 37, 22 events (17 primary refractory disease, 5 relapses), 2-year PFS = 40.8% Log-rank test for difference between groups: P < 0.01 PET1 vs. PET2 results (N = 89) Participants scanned after PET1 and 2 PET1-negative 2-year PFS = 98.3% PET1-positive 2-year PFS = 38.5% PET2-negative 2-year PFS = 90.2% PET2-positive 2-year PFS = 23.1% 14 PET1-positive converted to a PET2-negative (6 progressed). All PET1-negative were also PET2-negative.
Markova 2012	69 (advanced stages)	PET4	PET-negative N = 51, 2 events (1 relapse and 1 death), % of PFS not reported PET-positive N = 18, 4 events (progression or relapse), % of PFS not reported Log-rank test for difference between groups: P = 0.016
Oki 2014	229 (all stages)	PET2	3-year PFS rates in PET2-negative versus PET–positive by disease subgroups Early stage favourable: 100% vs. 100% Early stage unfavourable: 91.5% vs. 56.3% (P < 0.0001) Early stage non-bulky: 95.9% vs. 76.9% (P = 0.0018) Stage II bulky: 83.3% vs. 20% (P = 0.017) Advanced stage with IPS≤2: 77.0% vs. 30.0% (P < 0.001) Advanced stage with IPS≥3: 71.0% vs. 44.4% (P = 0.155)
Okosun 2012	23 (stages II - IV)	PET2 or PET3	PET-negative: N = 21, no events, 2-year PFS = 100%

PET-positive: N = 2, 1 event (treatment failure), 2-year PFS = 50%

Log-rank test for difference between groups: P = 0.0012

Multivariable analyses

Eight studies reported adjusted effect estimates for PFS (Casasnovas 2019; Gallamini 2014; Hutchings 2005; Hutchings 2006; Kobe 2018; Mesguich 2016; Rossi 2014; Simon 2016). Table 6 shows which prognostic factors (covariates) were considered in the final multivariable model of the studies. In two studies, only the results of those covariates that remained independent prognostic factors in multivariable analysis were reported (Gallamini 2014; Simon 2016). It is unclear whether, or which other covariates were included in the final multivariable model. The selection of prognostic factors (covariates) for adjustment in the studies was either based on their significance in univariable analysis (Casasnovas 2019; Hutchings 2006; Kobe 2018), or on the literature (established prognostic factors) (Hutchings 2005; Rossi 2014; Simon 2016). In two studies, the rationale for the covariates was not clearly stated (Gallamini 2014; Mesguich 2016).

As there are no final models with an identical set of covariates, pooling of adjusted effect estimates was not feasible. A full list of study-specific, candidate covariates can be found in the respective table for each study in the Characteristics of included studies.

The statistical methods used were Cox proportional hazards regression model and logistic regression model.

Table 6. Adjusted results from final multivariable model for PFS

Study	Prognostic factors								Adjusted results for interim PET
	Interim PET	Age	Gender	Disease stage	B symptoms	Bulky disease	IPS	Other study-specific factors	
Casasnovas 2019	x	-	x	x	x	x	x	x	Multivariable analysis not reported separately for standard treatment group.
Gallamini 2014	x	-	-	-	-	-	-	x	PET2 HR N/A, P < 0.01 (Sig. 0.000), 95% CI 3.136 to 7.917 **Comment:** Adjusted results indicate the independent prognostic impact of interim PET2.
Hutchings 2005	x	-	-	x	-	-	-	x	Early interim PET Wald 19.05, HR N/A, P-value = 0.00007 **Comment:** Adjusted results indicate the independent prognostic impact of early interim PET.
Hutchings 2006	x	-	-	x	-	-	-	x	*Model 1 (interim PET2 + clinical stage + extranodal disease)* PET2 HR = 36.281 (95% CI 7.179 to 183.4), P < .001 *Model 2 (interim PET2 + extranodal disease)* PET2 HR = 36.887 (95% CI 7.338 to 185.4), P < .001 **Comment:** Adjusted results indicate the independent prognostic impact of interim PET2.
Kobe 2018	x	-	-	-	-	-	x	x	Interim PET-positivity (DS 4) HR 2.4 (95% CI 1.4 to 4.1), P = 0.002

									Comment: Adjusted results indicate an independent prognostic impact of PET2.
Mesguich 2016	**x**	-	-	**x**	-	**x**	-	-	*Model 1 (interim PET + disease stage)* Positive interim PET HR = 3.73 (95% CI 1.35 to 10.35), P = 0.0112 *Model 2 (interim PET + bulky disease)* Positive interim PET HR = 3.62 (95% CI 1.30 to 10.05), P = 0.0138 **Comment:** Adjusted results indicate the independent prognostic impact of interim PET.
Rossi 2014	**x**	-	-	-	-	-	-	**x**	SUVmax PET0-PET2 Relative risk = 7.9 (95% CI 2.9 to 22.9), P = 0.0001 **Comment:** Adjusted results indicate the independent prognostic impact of SUVmax PET0-PET2.
Simon 2016	**x**	-	-	-	-	-	-	**x**	Interim PET-positivity HR = 17.74, P < 0.001, 95% CI 6.61 to 47.57 **Comment:** Adjusted results indicate the independent prognostic impact of PET2.

x = prognostic factor considered for adjustment in the final model

- = prognostic factor was not considered in the final model

Adverse events (AEs)

None of the included studies measured PET-associated AE.

Studies not reporting our outcomes

Two studies (Gandikota 2015; Orlacchio 2012) did not report data for our outcomes of interest, but were still included in this review as they fit our inclusion criteria. Their investigated outcomes were very close to our review outcomes and potentially the authors could have measured them, but did not report them in their publication. However, it has not been possible to obtain the relevant information; therefore, they are reported narratively in this review. Table 7 presents the results from these studies narratively.

Table 7. Narrative reporting of results from studies not reporting our outcomes of interest

Study	No. of participants	Outcomes/comparison	Results
Gandikota 2015	77 (stages IIA - IIB)	• Analysis of imaging at different time points: Baseline imaging, imaging during (after two to four cycles of ABVD) and at the end of treatment, follow-up imaging • Need for surveillance imaging	Analysis of imaging at different time points *Baseline imaging* • 77 participants had baseline PET-CT scans, 1 had only chest X-ray due to pregnancy at baseline *Imaging during and at the end of treatment* • 77 participants had interim PET-CT during chemotherapy (N = 34) or after chemotherapy before initiation of radiotherapy (N = 43) • Out of 77, 4 remained PET-positive, scans after completion of radiotherapy showed a complete response in 2/4, inflammation in 1/4, resolution of all adenopathy in 1/4, 0/4 relapsed during follow-up *Follow-up imaging* • Median follow-up: 46 months (range 24 to 126) • Total of 466 scans in 78 participants (PET-CT in N = 42) • No relapses occurred in the entire cohort, N = 3 were diagnosed with a second primary malignancy by either imaging or clinical presentation, N = 6 had false-positive imaging findings (3/6 PET-CT) requiring further supplementary imaging or biopsy/surgery Need for surveillance imaging Quote: “No relapse of cHL was detected at a median follow-up of 46 months. […] Routine imaging (either CT or PET-CT) for the early detection of relapse does not appear necessary or justified in these participants.”
Orlacchio 2012	132 (all stages)	Interim PET2 vs. end PET (three months after the end of chemo- and radiotherapy).	Interim PET results • Negative interim PET2: 104 • Positive interim PET2: 28 End PET results *Negative interim PET2 group* • Negative final PET: 102/104 • Positive final PET: 2/104 *Positive interim PET2 group* • Negative final PET: 16/28 • Positive final PET: 12/28

Interim PET vs. end PET

Negative interim PET2 group

- Quote: “Final PET confirmed the negative results in 102 cases (98%) and revealed pathological uptake in the remaining two cases (2%).”

Positive interim PET2 group

- Of the 28 interim PET-positive participants, 19 showed a partial response and nine had disease stability or progression. Twelve of the 28 interim PET-positive participants had a positive final PET. Hence, the remaining 16 had a negative final PET.

NPV and PPV

- Quote: “Interim PET had a NPV of 98%, with 85.7% sensitivity, 86.4% specificity and 86.4% diagnostic accuracy.”
- Quote: “[In univariable analysis] the only independent predictor is the result of interim PET. […] PET had a PPV of 42%."

DISCUSSION

Summary of main results

In this systematic review, we summarised unadjusted data for interim positron emission tomography (PET) scan results as a prognostic factor in individuals with classic Hodgkin lymphoma (HL). The results of an interim PET scan during therapy, e.g. after two cycles of chemotherapy, has been suggested as a good predictor of outcome. Interim PET scan results have also been suggested as an indicator to guide further treatment in order to achieve the best possible outcome in those that have a poor prognosis and those that have a good prognosis, while also minimising adverse events due to the toxicity of the chemotherapy. The results of our review are summarised in the Summary of findings 1.

The findings emerging from meta-analyses are as follows.

- Unadjusted results for overall survival (OS) show a large advantage for participants with a negative interim PET scan result compared to participants with a positive interim PET scan result. We rated the certainty of the evidence as 'moderate'.
- Unadjusted results for progression-free survival (PFS) show an advantage for participants with a negative interim PET scan result compared to participants with a positive interim PET scan result, but the evidence is very uncertain. We rated the certainty of the evidence as 'very low'.

The findings of the adjusted results from multivariable analyses, reported narratively in this review, are as follows.

- Adjusted results for OS indicate an independent prognostic ability of interim PET beyond other associated factors. We rated our certainty of the evidence as 'moderate'.
- Adjusted results for PFS indicate that there may be an independent prognostic ability of interim PET beyond other associated factors. We rated our certainty of the evidence as 'low'.

No study measured adverse events (AEs) associated with PET.

Overall completeness and applicability of the evidence

The evidence in this review mostly applies to adults who were newly diagnosed with classic HL, and who receive a PET scan in combination with CT (PET-CT) after two cycles of chemotherapy (PET2). The studies included in this review addressed our research question in a total of 7335 male and female participants representing all stages of classic HL (Ann Arbor stages I - IV with A or B symptoms). Nine studies included individuals aged 18 years or older, while the remaining studies also included adolescents and young adults (the youngest being 13 years of age, although most studies started from the age of 16 and onwards). Overall, the findings from this review support the statement that in this group of individuals, interim PET scan results can predict OS and PFS. Most participants in the included studies received ABVD (adriamycin/doxorubicin, bleomycin, vinblastine and dacarbazine) chemotherapy, which is the standard treatment regimen for early-stage disease (Bröckelmann 2018; Engert 2010). However, as participants can have different therapy regimens, which is decided based on their disease stage and other clinical or individual characteristics, results should always be interpreted with caution for different patient groups, and this naturally restrains the applicability of the evidence for all people with classic HL. Twelve out of 23 studies reported our primary outcome of interest OS, while 21 studies reported PFS. No study reported PET-associated AE. As the main aim of the review was to identify the prognostic value of interim PET results to predict survival outcomes, it is unlikely that studies on prognosis will measure or report AE.

Heterogeneity between the studies was also found with regard to the evaluation of the interim PET scan, as studies used different criteria for the interpretation of the results. Most studies used the Deauville five-point scale (DS 1 - 5) for the evaluation of the PET scans. However, different cut-off values were used for PET-positivity. Most studies considered scores one to three (DS 1-3) for PET-negativity, and scores four to five (DS 4-5) for PET-positivity. In some studies, however, DS3 was also considered (or tested) for

PET-positivity. Results from these studies should be interpreted with caution, as using a score of ≥3 can have an important impact on the results and possibly introduce bias. Firstly, using this cut-off can lead to an increased number of false-positive results for interim PET (Casasnovas 2019). This can have a relevant impact for the individual, if treatment would be modified based on the interim PET scan results (such as in the studies by Andre 2017; Casasnovas 2019; Kobe 2018). Furthermore, using this cut-off can lead to an overestimation of the positive outcomes in the interim PET-positive group. In the study by Kobe 2018, in which cut-off DS3 and DS4 were tested for PET-positivity, the results showed no significant difference in DS1-2 compared to DS3, but a significant difference between DS1-3 and 4. Thereby, the authors argue for DS4 as the cut-off value for PET-positivity, which is interpreted as an [18F]-fluorodeoxy-D-glucose (FDG) uptake higher than in the liver, instead of an uptake higher than in the mediastinum (corresponding DS3) (Kobe 2018). Hence, the implementation of a commonly used cut-off in clinical practice is important in order to improve interobserver reliability and agreements between central reviewers, and is also highly crucial for the individual (Kobe 2018; Meignan 2009a). In the remaining studies included in this review, either different criteria were used (e.g. International Harmonization Project in Lymphoma criteria (Juweid 2007)), or no specific scale was indicated. However, in most studies, at least two nuclear medicine physicians independently interpreted the interim PET scan results.

One of the greatest issues regarding the prognostic factor studies in this review relates to the difficult reporting of their statistical analyses. Even when the methods of the statistical analyses were appropriate for the study design, the data were insufficiently reported in many of the included studies. We used hazard ratios (HRs) as the effect measure for time-to-event data in this review. We were able to pool data from only 15 studies, either because the HR and associated standard error (SE) were not reported, or because we did not have separate data for our participants or outcomes of interest. Out of these 15 studies, six studies reported an HR, but we still re-calculated the value for four of them for different reasons. For example, values were re-calculated either when we detected discrepancies between the text and corresponding graph(s) and table(s), or when they were simply not reported, while other relevant data were, helping us to estimate the HR and SE. For the remaining studies, we estimated the HR using other available data where possible (Altman 2012; Parmar 1998; Tierney 2007; Trivella 2006). For this reason, we contacted 10 principal investigators to clarify our questions and provide us with additional information or data, or both. This step was particularly helpful for deciding which data to pool.

We prespecified in our protocol that we would only pool adjusted associations of the index prognostic factor if analyses were based on an identical set of covariates. Although this was not feasible for our review, we suggest that future authors of systematic reviews of prognostic factor studies consider pre-specifying a core set of covariates (established prognostic factors) that are important to the disease under review, and should be investigated in the included studies (Riley 2019; Riley 2019b). In this way, authors may be able to pool adjusted effect estimates, if studies are homogenous enough in the adjustment set of the other prognostic factors. In addition, we have moderate between-study heterogeneity, which is reflected in the I^2 and wide confidence intervals (CIs). We took these issues around the reporting in the studies into account when we assessed risk of bias and GRADE.

Furthermore, the pooled estimates of the prognostic effect of the interim PET scan result in our analyses are based on crude HRs (no adjustment for covariates), therefore the reported results are at risk of overestimating the prognostic ability of the interim PET scan result. Hence, in light of the absence of adjustment for other prognostic factors, and considering the risk of bias assessment for the fifth domain of the QUIPS tool, we downgraded the strength of the evidence in our GRADE assessment. This is because it is widely acknowledged that adjusting the predictive effect of a specific prognostic factor for the contribution of other prognostic factors strengthens the robustness of the evidence on the clinically relevant prognostic ability of that factor (Riley 2019; Riley 2019b).

Lastly, although we did not conduct a test for funnel plot asymmetry as this type of test is not necessarily recommended for survival data due to issues of censoring (Debray 2018), we cannot exclude potential publication bias and the presence of small-study effects in our review (Riley 2019). Firstly, we assume that publication bias may be present in our review as most studies in our analyses have rather small sample sizes, of which all present positive results on the prognostic ability of interim PET scan results. Secondly, most studies included in this review are retrospective studies that have not been pre-registered, for example, in trial registries. Studies are also not always labelled or indexed as prognosis studies, and search filters for studies on prognosis are still under development, which is the main reason as to why we conducted a broad search with the disease (HL) and prognostic factor (PET) of interest. This led to a high number of search results that had to be screened. Thirdly, we identified a great number of conference abstracts on studies for which we could not find full-text publications (see Characteristics of studies awaiting classification). Hence, based on these experiences, we cannot preclude that more studies may exist that have either not been published, or not indexed properly.

Certainty of the evidence

Our certainty of the evidence is presented in the Summary of findings 1.

Unadjusted results

For our primary outcome OS, we judged the certainty of the evidence as 'moderate'. We included nine studies in the meta-analysis, of which eight were observational studies and one was a clinical trial. We used the data of participants from the standard arm (no treatment adaptation) of this trial. We judged the certainty of the evidence as 'moderate' due to some methodological issues. We downgraded by one point for risk of bias due to a high risk of bias in seven studies for the domain other prognostic factors (covariates), as well as a high risk of bias in three studies for the domain statistical analysis and reporting. In addition, we downgraded by one point for imprecision because the HR had to be estimated in seven studies, and re-calculated in one study. Hence, only one out of nine studies reported a HR that we used. Nevertheless, we upgraded by one point for a large effect showing the large difference in the OS between interim PET-positive and interim PET-negative participants (HR 5.09, CI 2.64 to 9.81).

For the outcome PFS, we judged the certainty of the evidence as 'very low'. We included 14 studies in the meta-analysis, of which 12

were observational studies and two were clinical trials (participants from the standard arms). For this outcome, we downgraded by one point for inconsistency because the definition of PFS varied across the studies. We also downgraded by one point for imprecision because the HR had to be estimated in 10 studies and re-calculated in one study. Hence, we were able to use a reported HR for only three out of 14 studies. In addition, we downgraded by one point for risk of bias, because of a high risk of bias in eight studies for the domain other prognostic factors (covariates), and high risk of bias in six studies for the domain 'statistical analysis and reporting'.

Adjusted results

For the outcome OS, two studies reported adjusted results from multivariable analyses including established prognostic factors (e.g. International Prognostic Score) in individuals with HL, and the results of both studies indicate the independent prognostic ability of interim PET to predict OS. We judged our certainty in the evidence as 'moderate' for this outcome due to some methodological issues. We downgraded by one point for risk of bias due to a high risk of bias in the domains other prognostic factors (covariates) and statistical analysis and reporting for one study.

For the outcome PFS, there were eight studies that reported adjusted results (adjusted for e.g. disease stage or B symptoms). All studies found that interim PET scan results have an independent prognostic ability to predict PFS. However, we rated our certainty in the evidence as 'low' for this outcome. We downgraded by one point for risk of bias due to a high risk of bias in the domain study participation in one study, as well as a high risk of bias in the domains other prognostic factors (covariates) and statistical analysis and reporting in a second study. Furthermore, we downgraded by one point for inconsistency because the studies included a heterogenous set of covariates in the multivariable analyses, which made the pooling of adjusted results not feasible.

Potential biases in the review process

To prevent bias in this review, two teams of two review authors independently performed all relevant processes (i.e. screening, data extraction, risk of bias and GRADE assessment). Due to the complexity of assessing bias in prognostic factor studies, as well as assessing the certainty of the evidence from these types of studies, we conducted several teleconferences with different experts in the field of prognosis to discuss our assessments. We consulted Jill Hayden (Hayden 2013) for the 'Risk of bias' assessment, and the GRADE for Prognosis working group for the GRADE assessment. In particular, the methods for grading the evidence from prognosis studies are still under development.

For the 'Risk of bias' assessment, we are aware that adding 'unclear' as a fourth possible rating, thereby setting an example for future authors, can lead to a potential bias in the assessment. However, for our assessment we only used 'unclear' when relevant information was evidently missing, thereby making it difficult to make a fair and transparent judgement for the respective study and domain. We felt that rating a domain as high risk of bias in such cases would be inappropriate. We clearly advise against the use of 'unclear' as a default option and want to recommend future authors of reviews of prognosis studies to use this fourth rating carefully (if the fourth rating will be included in an update of the QUIPS tool).

Our analyses included post-hoc subgroup analyses on the type of PET measurement (PET versus PET-CT), as well as post-hoc sensitivity analyses on the timing of the interim PET and the type of estimation (see Methods) used to estimate missing values. These analyses were necessary due to the heterogeneity between the studies. Results should be interpreted in light of differences that can exist when participants receive a PET-CT as compared to a PET scan only. Furthermore, the timing of the interim PET is crucial, as PET1 and PET2 may provide different results compared to PET3 and PET4.

Regarding the adjusted results, we refrained from pooling results because, although the studies looked at established prognostic factors, they did not include identical sets of covariates. As the studies are already very heterogeneous, pooling of the adjusted results was not feasible for our review, as the comparison and interpretation of these results may be problematic in this case. To avoid this in the future, we suggest pre-defining a core set of covariates in order to enable pooling of adjusted results (Riley 2019).

Agreements and disagreements with other studies or reviews

In our review, we included studies that have assessed the prognostic value of interim PET in HL participants without treatment modification. Overall, the findings from this review are in agreement with similar reviews and studies that have investigated the prognostic value of interim PET. Our results are also in agreement with the literature that interim PET can be used for disease and therapy monitoring (Barrington 2017a). Some reviews and studies have investigated this in participants in whom the treatment was changed based on the interim PET scan result, and have come to similar conclusions that interim PET can predict outcome in the different groups (PET-negative and PET-positive participants).

We are aware of three systematic reviews (Adams 2015a; Amitai 2018; Sickinger 2015) that have investigated interim PET as a prognostic factor. Adams 2015a included ten studies with limited-, intermediate- and advanced-stage HL participants in whom the treatment regimen was not modified based on the interim PET scan results. In fact, nine out of these 10 studies are also included in our review. One study was not included in our review because they only included children. The authors of this review concluded that a negative interim PET cannot exclude treatment failure, but that a positive interim PET can identify and predict treatment failure. The authors assessed the quality of the studies with the QUIPS tool (as we did in our review) and judged the overall methodological quality of the included studies as moderate. We have compared their QUIPS assessment with ours for each individual study, and identified that for the domains study participation and study attrition in particular, we found agreements between the authors and our review that there is a low risk of bias in the studies. Disagreement was found regarding the domain prognostic factor measurement, for which the authors judged the quality as moderate mainly due to the heterogeneity between the studies regarding the use of PET-CT versus PET only, which is an issue that we have also addressed in our review by subgroup analysis.

Comparison of interim PET with end PET

Nine of the included studies compared the performances of interim PET and end-of treatment PET (end PET) (Barnes 2011; Hutchings 2006; Hutchings 2014; Markova 2012; Mesguich 2016; Orlacchio

2012; Straus 2011; Ying 2014; Zinzani 2012), as omitting one of the two can have an impact on radiation safety for the patient. However, results between studies are rather contradictory. For example, in Barnes 2011 the authors could not detect a significant difference in OS and PFS between interim PET-negative and interim PET-positive participants. In their analyses, interim PET-positive participants that were negative at end PET had the same good outcomes as participants who were negative both at interim and end PET. In addition, after end PET, the difference between end PET-positive and end PET-negative participants was fairly high, with a greater four-year OS and PFS for end-PET-negative participants. In this study, 74 (end PET) out of 79 participants (interim PET) remained PET-negative, while nine (end PET) out of 17 (interim PET) participants remained PET-positive. The authors concluded that end-PET (after six cycles of chemotherapy) predicts outcome, rather than interim PET (after two or four cycles of chemotherapy). In Hutchings 2006, interim PET was conducted after two and four cycles of chemotherapy (total number of cycles was six to eight). Results show that PET2 and PET4 were similarly successful in predicting outcome in participants, but the authors of the study still argue that treatment modifications should be indicated as early as possible (e.g. after PET2) in order to achieve the best possible outcome. In the study by Mesguich 2016, interim PET was also lower in its predictive ability compared to end PET. Out of 60 interim PET-negative participants, seven converted to a positive end PET. Out of 16 interim PET-positive participants, seven converted to a negative end PET. In addition, treatment failure was most common in participants with a positive end PET as compared to participants with a positive interim PET. The sensitivity of interim PET was measured as 47% compared to 80% of end PET (Mesguich 2016).

Contrastingly, Orlacchio 2012 detected a very high negative predictive value (NPV) of 98% for interim PET2, with an overall diagnostic accuracy of 86.4%. Out of 104 interim PET-negative participants, 102 were still negative after end PET. Out of 28 interim PET-positive participants, however, 16 converted to a negative end PET. A high NPV for interim PET was also found in Hutchings 2005 (interim PET2/3) as interim PET-negative participants rarely relapsed. In Hutchings 2014, 89 participants had an interim PET1 and PET2, and both show a strong prognostic ability for predicting outcome. In this study, none of the participants in early stages that had a negative interim PET1 progressed or relapsed. Advanced-stage participants with a negative interim PET1 had a long-term PFS of more than 90%. The three-year PFS of interim PET1-positive participants was 30%. In total, 89 participants had both PET1 and PET2. Out of these, 62 were PET1-negative, and after treatment, 60 were in complete remission. Twenty-seven participants were PET1-positive, of which 15 were in complete remission. To compare, 76 participants were PET2-negative, of whom 70 were in complete remission. Thirteen participants were PET2-positive, of which five were in complete remission. The negative predictive value of PET1 was reported as 96.8%, while the positive predictive value was 44.4%. Zinzani 2012 also reported that interim PET after two cycles is highly predictive of OS and PFS. In their study, 92% of the interim PET-negative participants (n = 251) were in continuous complete remission as compared to 24.5% of the interim PET-positive participants (n = 53). These conclusions are supported by Ying 2014, although their sample size (n = 35) is too small to provide definite answers. Straus 2011 supported these statements particularly for participants in early stages (as included in their study), as participants with a negative interim PET2 result had a PFS of about 90%, compared to 50% for interim PET-positive participants, at two years. Markova 2012 reported similar findings for interim PET4, which had a high NPV of 98%. Out of 68 participants in total, 50 had a negative interim PET, but 59 a negative end PET. In other words, nine interim PET-positive participants were end PET-negative after chemotherapy. The other nine participants who were interim PET-positive were also end PET-positive. At both timings (PET4 and PET6/8) the authors found a significant difference in the survival between PET-positive and PET-negative participants. The high NPV of interim PET supports early de-escalation of chemotherapy, or omitting radiotherapy, in order to reduce the risk of toxicity and adverse events related to the harsh treatment.

Treatment adaptation based on interim PET

Although not an aim of our review, we considered it important to discuss some results from recently published randomised controlled trials (RCTs) in which the interim PET scan result was used to adapt the therapy for individuals with HL in order to improve outcomes (Andre 2017; Casasnovas 2019; Johnson 2016; Kobe 2018), based on the premise that interim PET scan results are indeed prognostic. For example, in the trial by Johnson 2016, the primary aim was to test the omission of bleomycin due to its toxic effects. All participants (N = 1214, advanced stages) started with ABVD chemotherapy. After interim PET2, PET-positive (DS4-5) participants (N = 182) were assigned to BEACOPP, and PET-negative (DS1-3) participants (N = 935) were randomised to receive either ABVD or AVD. Results show that three-year PFS was slightly better in the ABVD group compared to the AVD group (85.7% versus 84.4%, respectively). Regarding three-year OS, the ABVD group reached 97.2% compared to 97.6% in the AVD group. Hence, there were no significant subgroup differences. However, grade 3 and 4 AEs due to the chemotherapy were more common in the ABVD group. In the PET-positive group, which was escalated to BEACOPP chemotherapy, 3-year PFS was 67.5% and 3-year OS was 87.8%.

In another example by Casasnovas 2019, 823 advanced-stage HL participants were randomly assigned to standard treatment group or PET-driven treatment group. All participants received two cycles of $BEACOPP_{escalated}$ as the initial therapy and interim PET was conducted thereafter. PET-positive participants in both groups, as well as PET-negative participants in the standard group continued with the initial therapy after PET2. PET-negative participants in the experimental arm, however, were switched to two cycles of ABVD. Results of five-year PFS show a similar survival of PET-negative participants in the standard group and experimental group: 88.4% and 89.4%, respectively.

Several systematic reviews were also published that investigated treatment adaptation based on interim PET scan results. Amitai 2018 included 13 studies (of which four were RCTs) that investigated interim PET-adapted treatment in advanced-staged HL. Their findings support the statement that PET-adapted treatment is an appropriate strategy and that it should be considered as standard care for advanced HL (Amitai 2018). This finding is supported by a Phase II RCT (Carras 2018), which assessed interim PET-response adapted treatment strategy in advanced-stage HL. The authors concluded that early salvage therapy and high-dose chemotherapy or autologous stem cell transplant (ASCT) for PET2-positive participants is safe and can lead to similar positive outcomes as in PET2-negative participants (Carras 2018). To compare, Sickinger 2015 included studies in which the treatment was also modified, but concluded that PFS was shorter in individuals with early-stage HL and a negative PET

scan receiving chemotherapy only (PET-adapted therapy) than in those receiving additional RT (standard therapy). This finding was confirmed in another review by Blank 2017, showing improved PFS in early-stage participants receiving radiotherapy in addition to chemotherapy. However, the overall methodological quality of the included studies in both reviews was judged as moderate (for PFS) to very low (for OS). Constrasting evidence on the clinical and prognostic value of interim PET-adapted treatment was also found in non-systematic reviews, which particularly acknowledge the heterogeneity between available studies that makes it difficult to give definite conclusions (Adams 2016a; Berriolo-Riedinger 2018).

AUTHORS' CONCLUSIONS

Implications for practice

This review provides moderate-certainty evidence that interim positron emission tomography (PET) scan results predict overall survival (OS), and very low-certainty evidence that interim PET scan results predict progression-free survival (PFS) in individuals with Hodgkin lymphoma (HL) (evidence of the pooled, unadjusted results). The evidence on the ability of interim PET scan results to distinguish between individuals with a poor prognosis and individuals with a good prognosis can aid decision-making for clinicians and diagnosed individuals, and the evidence may be used in international treatment guidelines for individuals with HL.

Implications for research

Multivariable analyses and prognostic models

Thus far, the prognostic value of interim PET has mostly been assessed in univariable analyses, in which its prognostic ability of determining survival outcomes in individuals with HL has been shown. However, using one single factor is usually not sufficient to give a satisfactory prediction of an outcome, and clinicians, therefore, usually additional factors to give an accurate prediction of an individual's disease progression and health outcome (Moons 2009). Hence, it is important to assess the independent prognostic value of the prognostic factor of interest (in this case interim PET) against established prognostic factors such as disease stage, age, sex, B symptoms or other relevant clinical and individual factors in multivariable analyses as well (Moons 2009; Riley 2019). In such analyses, the independent prognostic ability of a factor, as well as its incremental value on top of other prognostic factors, can be assessed (Moons 2009). In a next step, prognostic models can be built that include multiple prognostic factors that have been proven to be predictive of outcome. Such models are built for risk adaptation and treatment stratification for participants who present those specific factors included in a prediction model for a specific disease, and thereby enables more individualised disease monitoring and treatment guidance. Using a combination of factors, rather than one factor only, allows for a more individual and accurate estimate of the risk of a patient to experience a certain health event (or outcome) within a specific period of time (Moons 2009; Steyerberg 2013).

With regard to our index prognostic factor, we could pool adjusted results in meta-analyses in an update of this review if new studies would adjust for the same set of prognostic factors (covariates). There is a number of different established clinical and individual prognostic factors that can be used to predict survival outcomes in individuals with HL (Cuccaro 2014; Josting 2010; Kılıçkap 2013). In order to enable pooling of adjusted results, future authors of systematic reviews of prognostic factor studies could define a core set of covariates a priori (Riley 2019).

Study design

There is some evidence from retrospective studies that interim PET scan results can predict outcome in individuals during chemotherapy. However, it is commonly agreed that the true prognostic value of this factor can best be assessed in randomised controlled trials (RCTs), in which participants are randomly assigned to a standard or an experimental arm. In the standard arm, participants continue with the planned therapy regimen independent of the interim PET scan result. In the experimental arm, however, different treatments are given according to the interim PET scan result, e.g. de-escalation of treatment in interim PET-negative participants. Hence, RCTs are the most suitable study design, with results from experimental arms in which participants receive therapy adaptation based on the interim PET scan result providing the most robust evidence on whether outcome can be approved, while treatment can be safer, by this strategy. Although assessing therapy modification was not an aim of our review, we judged it important to present and discuss some results of published trials that evaluated the impact of PET-adapted treatment on survival outcomes.

ACKNOWLEDGEMENTS

This review was published in collaboration with the Cochrane Fast-Track Service. We particularly thank Helen Wakeford (Managing Editor) and Heather Maxwell (copy-editor) for their support. The Fast-Track Service team made comments on the review and managed the editorial process.

We thank Nicola Köhler of the Cochrane Haematological Malignancies (CHM) Editorial Base as well as the editors Dr. Julia Bohlius and PD Dr. Sebastian Theurich and the consumer editor Anne Lyddiatt for commenting on the protocol of this review.

We thank Ina Monsef of CHM for developing the search strategy and conducting the search for this review, and Marius Goldkuhle and Vanessa Piechotta for commenting on the first submitted draft of this review.

We thank Jill Hayden for her help with the QUIPS tool, and particularly for agreeing to our adaptation of the tool for the purpose of this review.

We thank Yu-Tian Xiao for the translation of the Chinese publication (Ying 2014).

We thank all principal investigators who have replied to all of our questions and provided us with more information or data, thereby helping us to include as many studies as possible in our analysis. Hence, we want to thank Prof. Dr. Jan Maciej Zaucha, Prof. Dr. Andrea Gallamini, Prof. Dr. Peter Borchmann and his colleague Helen Görgen, Prof. Dr. Pier Luigi Zinzani, Dr. Martin Hutchings, Prof. Dr. Marc André and his colleague Dr. John Raemaekers.

We thank Kym Snell and Sarah Hodgkinson for commenting on the first submitted draft of this review, as well as Robin Featherstone for commenting on the search strategy.

We thank all external peer-reviewers who read and commented on this review. We thank Robert Wolff, Bob Phillips, Michel Meignan,

Tim Illidge, Julian Higgins and Ulrike Holtkamp who greatly helped to improve this review.

The research was supported by NHS Blood and Transplant and the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

REFERENCES

References to studies included in this review

Andre 2017 *{published data only}*

Adams HJ, Kwee TC. Interim fluorodeoxyglucose positron emission tomography-adapted therapy is not an efficient approach to improving outcome in early-stage Hodgkin lymphoma. *Journal of Clinical Oncology* 2017;**35**(24):2850-1.

* Andre ME, Girinsky T, Federico M, Reman O, Fortpied C, Gotti M, et al. Early positron emission tomography response-adapted treatment in stage I and II Hodgkin lymphoma: final results of the randomized EORTC/LYSA/FIL H10 trial (NCT00433433). *Journal of Clinical Oncology* 2017;**35**:1786-94.

Andre ME, Reman O, Federico M, Brice P, Brusamolino E, Girinski T, et al. First report on the H10 EORTC/GELA/IIL randomized intergroup trial on early FDG-PET scan guided treatment adaptation versus standard combined modality treatment in patients with supra-diaphragmatic stage I/II Hodgkin's lymphoma, for the Groupe d'Etude Des Lymphomes De l'Adulte (GELA), European Organisation for the Research and Treatment of Cancer (EORTC) Lymphoma Group and the Intergruppo Italiano Linfomi (IIL). *Blood* 2009;**114**(22):97.

Andre ME, Reman O, Federico M, Girinski T, Brice P, Brusamolino E, et al. Interim analysis of the randomized EORTC/LYSA/FIL intergroup H10 trial on early PET-scan driven treatment adaptation in stage I/II Hodgkin lymphoma. *Blood* 2012;**120**(21):549.

Cottereau AS, Versari A, Loft A, Casanovas O, Bellei M, Ricci R, et al. Prognostic value of baseline metabolic tumor volume in early-stage Hodgkin lymphoma in the standard arm of the H10 trial. *Journal of Nuclear Medicine* 2018;**131**(13):1456-63.

Fornecker LM, Lazarovici J, Aurer I, Casasnovas RO, Gac A C, Bonnet C, et al. PET-based response after 2 cycles of brentuximab vedotin in combination with AVD for first-line treatment of unfavorable early-stage Hodgkin lymphoma: First analysis of the primary endpoint of BREACH, a randomized phase II trial of LYSA-FIL-EORTC intergroup. *Blood* 2017;**130**(Suppl 1):736.

Hindie E, Mesguich C, Zanotti-Fregonara P. On the role of interim fluorine-18-labeled fluorodeoxyglucose positron emission tomography in early-stage favorable Hodgkin lymphoma. *Journal of Clinical Oncology* 2017;**35**(24):2851-2.

Raemaekers JM, Andre MP, Federico M, Girinsky T, Oumedaly R, Brusamolino E, et al. Omitting radiotherapy in early positron emission tomography-negative stage I/II Hodgkin lymphoma is associated with an increased risk of early relapse: Clinical results of the preplanned interim analysis of the randomized EORTC/LYSA/FIL H10 trial. *Journal of Clinical Oncology* 2014;**32**:1188-94.

Annunziata 2016 *{published data only}*

Annunziata S, Calcagni M, Rufini V, Cuccaro A, Massini G, Bartolomei F, et al. The prognostic role of interim FDG-PET/CT and CD68+ cells count in the treatment of Hodgkin lymphoma. *European Journal of Nuclear Medicine and Molecular Imaging* 2014;**41**(Suppl. 2):S182.

Annunziata S, Calcagni ML, Indovina L, Rufini V. Measurement uncertainty and clinical impact of target-to-background ratios derived by interim FDG-PET/CT in Hodgkin Lymphoma: reply to Laffon and Martan. *European Journal of Nuclear Medicine and Molecular Imaging* 2017;**44**(12):2140-1.

Annunziata S, Cuccaro A, Calcagni M, Indovina L, Hohaus S, Giordano A, et al. Could the ratio between lesion and liver SUVmax (rPET) be a prognostic factor in patients with Hodgkin lymphoma undergoing interim FDG-PET/CT? A retrospective study. In: European Journal of Nuclear Medicine and Molecular Imaging. Vol. 42. 2015:S686.

* Annunziata S, Cuccaro A, Calcagni ML, Hohaus S, Giordano A, Rufini V. Interim FDG-PET/CT in Hodgkin lymphoma: the prognostic role of the ratio between target lesion and liver SUVmax (rPET). *Annals of Nuclear Medicine* 2016;**30**:588-92.

Annunziata S, Cuccaro A, Hohaus S, Calcagni M L, Giordano A, Rufini V. Interim FDG-PET/CT in patients with Hodgkin lymphoma: The prognostic role of the ratio between target lesion and liver SUVmax (rPET). *Journal of Nuclear Medicine* 2016;**57**(Supp 2):651.

Annunziata S, Cuccaro A, Hohaus S, Giordano A, Rufini V. Semi-quantitative parameters could improve positive predictive value of interim FDG-PET/CT in Hodgkin lymphoma. *European Journal of Nuclear Medicine and Molecular Imaging* 2016;**43**(Suppl 1):S312.

Annunziata S, Cuccaro A, Hohaus S, Rufini V. Interim FDG-PET/CT in Hodgkin lymphoma: The prognostic value of ratios between target lesion and background SUV. *Clinical and Translational Imaging* 2017;**5**(Suppl 1):S44-5.

Annunziata S, Cuccaro A, Rufini V, Calcagni M, Giachelia M, Martini M, et al. The prognostic role of interim FDG-PET/CT, CD68 1 cells count and levels of plasma thymus and activation-regulated chemokine in the treatment of Hodgkin lymphoma. *Clinical and Translational Imaging* 2015;**3**(Suppl 1):S18-19.

Cuccaro A, Annunziata S, Rufini V, Calcagni M L, Giordano A, Hohaus S. Can the ratio between lesion and liver SUVmax improve outcome prediction in Hodgkin lymphoma with respect to the 5-point Deauville score? *Haematologica* 2015;**100**(Suppl 3):89.

Barnes 2011 *{published data only}*

* Barnes JA, LaCasce AS, Zukotynski K, Israel D, Feng Y, Neuberg D, et al. End-of-treatment but not interim PET scan predicts outcome in nonbulky limited-stage Hodgkin's lymphoma. *Annals of Oncology* 2011;**22**:910-5.

Sher DJ, Mauch PM, Van Den Abbeele A, LaCasce AS, Czerminski J, Ng AK. Prognostic significance of mid- and post-ABVD PET imaging in Hodgkin's lymphoma: the importance of involved-field radiotherapy. *Annals of Oncology* 2009;**20**:1848-53.

Casasnovas 2019 *{published data only}*

* Casasnovas O, Bouabdallah R, Brice P, Lazarovici J, Ghesquieres H, Stamatoullas A, et al. PET-adapted treatment for newly diagnosed advanced Hodgkin lymphoma (AHL2011): a randomised, multicentre, non-inferiority, phase 3 study. *Lancet Oncology* 2019;**20**(2):202-15.

Casasnovas O, Brice P, Bouabdallah R, Salles G, Stamatoulas A, Dupuis J, et al. Final analysis of the AHL2011 randomized phase III LYSA study comparing an early pet driven treatment de-escalation to a not pet-monitored strategy in patients with advanced stages Hodgkin lymphoma. *HemaSphere* 2018;**2**(Suppl. 2):7.

Casasnovas O, Brice P, Bouabdallah R, Salles GA, Stamatoulas A, Dupuis J, et al. Randomized phase III study comparing an early PET driven treatment de-escalation to a not PET-monitored strategy in patients with advanced stages Hodgkin lymphoma: final analysis of the AHL2011 LYSA study. *Journal of Clinical Oncology* 2018;**36**(15):7503.

Casasnovas O, Brice P, Bouabdallah R, Salles GA, Stamatoullas A, Dupuis J, et al. Randomized phase III study comparing an early pet driven treatment de-escalation to a not pet-monitored strategy in patients with advanced stages hodgkin lymphoma: interim analysis of the AHL2011 LYSA study. *Blood* 2015;**126**(23):577.

Casasnovas O, Kanoun S, Tal I, Cottereau AS, Edeline V, Brice P, et al. Baseline total metabolic volume (TMTV) to predict the outcome of patients with advanced Hodgkin lymphoma (HL) enrolled in the AHL2011 LYSA trial. *Journal of Clinical Oncology* 2016;**34**(15):7509.

Casasnovas O, Meignan M, Reman O, Gaillard I, Stamatoullas A, Brice P, et al. AHL 2011: A LYSA randomized phase III study of a treatment driven by early PET response compared to a standard treatment in patients with Ann Arbor stage III-IV or high-risk IIB Hodgkin lymphoma. *Journal of Clinical Oncology* 2013;**31**(Suppl. 15):8615.

Cerci 2010 *{published data only}*

Cerci JJ, Pracchia LF, Linardi CC, Pitella FA, Delbeke D, Izaki M, et al. 18F-FDG PET after 2 cycles of ABVD predicts event-free survival in early and advanced Hodgkin lymphoma. *Journal of Nuclear Medicine* 2010;**51**:1337-43.

Gallamini 2014 *{published data only}*

Adams HJ, Kwee TC. Prevention of large-scale implementation of unnecessary and expensive predictive tests in Hodgkin's lymphoma. *Lancet Haematology* 2017;**4**(2):e63-4.

Agostinelli C, Gallamini A, Stracqualursi L, Agati P, Tripodo C, Fuligni F, et al. The combined role of biomarkers and interim PET scan in prediction of treatment outcome in classical Hodgkin's lymphoma: a retrospective, European, multicentre cohort study. *Lancet Haematology* 2016;**3**(10):e467-79.

Biggi A, Bergesio F, Bianchi A, Menga M, Chauvie S, Fallanca F, et al. Semi-quantitative scan assessment improves the accuracy of the deauville 5-point scale? *Clinical and Translational Imaging* 2017;**5**(Suppl. 1):S43-4.

Biggi A, Gallamini A, Chauvie S, Hutchings M, Kostakoglu L, Gregianin M, et al. International validation study for interim PET in ABVD-treated, advanced-stage Hodgkin lymphoma: interpretation criteria and concordance rate among reviewers. *Journal of Nuclear Medicine* 2013;**54**:683-90.

Biggi A, Kostakoglu L, Barrington S, Chauvie S, Gregianin M, Meignan M, et al. How the threshold of positive results influence progression free survival (PFS) in advanced Hodgkin's lymphoma treated with ABVD: Experience from the International Validation Study. *European Journal of Nuclear Medicine and Molecular Imaging* 2012;**39**(Suppl. 2):S222.

* Gallamini A, Barrington SF, Biggi A, Chauvie S, Kostakoglu L, Gregianin M, et al. The predictive role of interim positron emission tomography for Hodgkin lymphoma treatment outcome is confirmed using the interpretation criteria of the Deauville five-point scale. *Haematologica* 2014;**99**:1107-13.

Gallamini A, Hutchings M, Rigacci L, Specht L, Merli F, Hansen M, et al. Early interim 2-[18F]fluoro-2-deoxy-D-glucose positron emission tomography is prognostically superior to international prognostic score in advanced-stage Hodgkin's lymphoma: a report from a joint Italian-Danish study. *Journal of Clinical Oncology* 2007;**25**:3746-52.

Gallamini A, Rigacci L, Merli F, Nassi L, Bosi A, Capodanno I, et al. The predictive value of positron emission tomography scanning performed after two courses of standard therapy on treatment outcome in advanced stage Hodgkin's disease. *Haematologica* 2006;**91**:475-81.

Meignan M. High-risk interim PET negative patients in Hodgkin's lymphoma. *Lancet Haematology* 2016;**3**(10):e449-50.

Rigacci L, Puccini B, Zinzani PL, Biggi A, Castagnoli A, Merli F, et al. The prognostic value of positron emission tomography performed after two courses (INTERIM-PET) of standard therapy on treatment outcome in early stage Hodgkin lymphoma: A multicentric study by the fondazione italiana linfomi (FIL). *American Journal of Hematology* 2015;**90**:499-503.

Gandikota 2015 *{published data only}*

Gandikota N, Hartridge-Lambert S, Migliacci JC, Yahalom J, Portlock CS, Schoder H. Very low utility of surveillance imaging in early-stage classic Hodgkin lymphoma treated with a combination of doxorubicin, bleomycin, vinblastine, and dacarbazine and radiation therapy. *Cancer* 2015;**121**:1985-92.

Hutchings 2005 *{published data only}*

Hutchings M, Mikhaeel NG, Fields PA, Nunan T, Timothy AR. Prognostic value of interim FDG-PET after two or three cycles of chemotherapy in Hodgkin lymphoma. *Annals of Oncology* 2005;**16**:1160-8.

Hutchings 2006 *{published data only}*

Hutchings M, Loft A, Hansen M, Pedersen LM, Buhl T, Jurlander J, et al. FDG-PET after two cycles of chemotherapy predicts treatment failure and progression-free survival in Hodgkin lymphoma. *Blood* 2006;**107**:52-9.

Hutchings 2014 *{published data only}*

Hutchings M, Kostakoglu L, Zaucha J M, Malkowski B, Biggi A, Danielewicz I, et al. Early determination of treatment sensitivity in Hodgkin lymphoma: FDG-PET/CT after one cycle of therapy has a higher negative predictive value than after two cycles of chemotherapy. *Annals of Oncology* 2011;**22**(Suppl. 4):138-9.

* Hutchings M, Kostakoglu L, Zaucha JM, Malkowski B, Biggi A, Danielewicz I, et al. In vivo treatment sensitivity testing with positron emission tomography/computed tomography after one cycle of chemotherapy for Hodgkin lymphoma. *Journal of Clinical Oncology* 2014;**32**:2705-11.

Kobe 2018 *{published data only}*

Adams HJ, Kwee TC. Interim FDG-PET does not predict outcome in advanced-stage Hodgkin lymphoma patients treated with BEACOPP. *British Journal of Haematology* 2018;**185**(4):758-60.

Adams HJ, Kwee TC. Interim FDG-PET/CT in Hodgkin lymphoma: what are we actually looking at? *Acta Oncologica* 2018;**57**(8):1128-30.

Borchmann P, Eichenauer DA, Pluetschow A, Haverkamp H, Kreissl S, Fuchs M, et al. Targeted BEACOPP variants in patients with newly diagnosed advanced stage classical Hodgkin lymphoma: Final analysis of a randomized phase II study. *Blood* 2015;**126**(23):580.

Borchmann P, Engert A. Reply to H.J.A. Adams et al, E.A. Hawkes et al, and C.F. Hess et al. *Journal of Clinical Oncology* 2017;**35**(3):375-6.

Borchmann P, Goergen H, Kobe C, Eichenauer D, Greil R, Lohri A, et al. EBEACOPP with or without rituximab in interim-PET-positive advanced-stage Hodgkin lymphoma: updated results of the international, randomized phase 3 GHSG HD18 trial. *Hematological Oncology* 2017;**35**(Suppl. 2):65.

Borchmann P, Goergen H, Kobe C, Fuchs M, Greil R, Zijlstra JM, et al. Treatment reduction in patients with advanced-stage Hodgkin lymphoma and negative interim pet: final results of the international, randomized phase 3 trial HD18 by the German Hodgkin study group. *Haematologica* 2017;**102**(S150):24-5.

Borchmann P, Goergen H, Kobe C, Lohri A, Greil R, Eichenauer DA, et al. Early interim PET in patients with advanced-stage Hodgkin's lymphoma treated within the phase 3 GHSG HD18 study. *Blood* 2017;**130**(Suppl. 1):653.

Borchmann P, Goergen H, Kobe C, Lohri A, Greil R, Eichenauer DA, et al. PET-guided treatment in patients with advanced-stage Hodgkin's lymphoma (HD18): final results of an open-label, international, randomised phase 3 trial by the German Hodgkin Study Group. *Lancet* 2018;**390**(10114):2790-802.

Borchmann P, Haverkamp H, Lohri A, Kreissl S, Greil R, Markova J, et al. Addition of rituximab to BEACOPPescalated to improve the outcome of early interim PET positive advanced stage hodgkin lymphoma patients: Second planned interim analysis of the HD18 study. *Blood* 2014;**124**(21):500.

Borchmann P, Haverkamp H, Lohri A, Mey U, Kreissl S, Greil R, et al. Progression-free survival of early interim PET-positive patients with advanced stage Hodgkin's lymphoma treated with BEACOPPescalated alone or in combination with rituximab (HD18): an open-label, international, randomised phase 3 study by the German Hodgkin Study Group. *Lancet Oncology* 2017;**18**(4):454-63.

Crump M. Predicting outcomes after a positive interim FDG-PET scan in advanced Hodgkin's lymphoma. *Lancet Oncology* 2017;**18**(4):416-8.

* Kobe C, Goergen H, Baues C, Kuhnert G, Voltin CA, Zijlstra J, et al. Outcome-based interpretation of early interim PET in advanced-stage Hodgkin lymphoma. *Blood* 2018;**132**(21):2273-9.

Kobe C, Goergen H, Fuchs M, Eich HT, Baues C, Diehl V, et al. Treatment reduction in patients with advanced-stage Hodgkin lymphoma and negative interim FDG-PET: final results of the international, randomized, phase 3 HD18 trial by the German Hodgkin Study Group. *European Journal of Nuclear Medicine and Molecular Imaging* 2017;**44**(Suppl. 1):S312.

Kreissl S, Goergen H, Kobe C, Fuchs M, Greil R, Huttmann A, et al. Treatment reduction in patients with advanced-stage Hodgkin-lymphoma and negative interim PET: final results of the international randomized phase 3 trial HD18 by the German Hodgkin Study Goup. *Oncology Research and Treatment* 2017;**40**(Suppl. 3):201.

Markova 2012 *{published data only}*

* Markova J, Kahraman D, Kobe C, Skopalova M, Mocikova H, Klaskova K, et al. Role of [18F]-fluoro-2-deoxy-D-glucose positron emission tomography in early and late therapy assessment of patients with advanced Hodgkin lymphoma treated with bleomycin, etoposide, adriamycin, cyclophosphamide, vincristine, procarbazine and prednisone. *Leukemia & Lymphoma* 2012;**53**:64-70.

Markova J, Kobe C, Skopalova M, Dedeckova K, Mocikova H, Klaskova K, et al. Response assessment after 4 cycles of BEACOPP using FDG-PET in patients with advanced-stage Hodgkin lymphoma. *Annals of Oncology* 2011;**22**(Suppl. 4):160.

Markova J, Kobe C, Skopalova M, Klaskova K, Dedeckova K, Plutschow A, et al. FDG-PET for assessment of early treatment response after four cycles of chemotherapy in patients with advanced-stage Hodgkin's lymphoma has a high negative predictive value. *Annals of Oncology* 2009;**20**:1270-4.

Markova J, Kobe C, Skopalova M, Zikavska L, Vernerova Z, Klaskova K, et al. Early and late response assessment with FDG-PET after BEACOPP-based chemotherapy in advanced-stage Hodgkin lymphoma patients has a high negative predictive value. In: Haematologica. Vol. 94. 2009:33.

Mesguich 2016 *{published data only}*

Adams HJ, Kwee TC. Hodgkin lymphoma: is there really a need for interim and end-of-treatment FDG-PET evaluations? *British Journal of Haematology* 2018;**181**(1):122-3.

Mesguich C, Cazeau AL, Bouabdallah K, Hindie E. Hodgkin lymphoma: is there really a need for interim and end-of-treatment FDG-PET evaluations? - Response to Adams & Kwee. *British Journal of Haematology* 2018;**181**(1):124-5.

* Mesguich C, Cazeau AL, Bouabdallah K, Soubeyran P, Guyot M, Milpied N, et al. Hodgkin lymphoma: a negative interim-PET cannot circumvent the need for end-of-treatment-PET evaluation. *British Journal of Haematology* 2016;**175**:652-60.

Mesguich C, Cazeau AL, Bouabdallah K, Soubeyran P, Guyot M, Milpied N, et al. Hodgkin's lymphoma: Interim FDG-PET result with a score <= 2 on the 5-point scale may obviate the need for end-of-treatment FDG-PET evaluation. *Journal of Nuclear Medicine* 2015;**56**(Suppl. 3):595.

Oki 2014 *{published data only}*

Oki Y, Chuang H, Chasen B, Jessop A, Pan T, Fanale M, et al. The prognostic value of interim positron emission tomography scan in patients with classical Hodgkin lymphoma. *British Journal of Haematology* 2014;**165**:112-6.

Okosun 2012 *{published data only}*

Okosun J, Shaw K, Montoto S, Marcus R, Fields P, Virchis A, et al. Interim FDG-PET scanning in patients with Hodgkin lymphoma and HIV predict response to ABVD chemotherapy. *Haematologica* 2011;**96**(Suppl. 2):322-3.

* Okosun J, Warbey V, Shaw K, Montoto S, Fields P, Marcus R, et al. Interim fluoro-2-deoxy-D-glucose-PET predicts response and progression-free survival in patients with Hodgkin lymphoma and HIV infection. *AIDS* 2012;**26**:861-5.

Orlacchio 2012 *{published data only}*

Orlacchio A, Schillaci O, Gaspari E, Della Gatta F, Danieli R, Bolacchi F, et al. Role of [18F]-FDG-PET/MDCT in evaluating early response in patients with Hodgkin's lymphoma. *La Radiologia Medica* 2012;**117**:1250-63.

Rossi 2014 *{published data only}*

Rossi C, Kanoun S, Berriolo-Riedinger A, Dygai-Cochet I, Humbert O, Legouge C, et al. Interim 18F-FDG PET SUVmax reduction is superior to visual analysis in predicting outcome early in Hodgkin lymphoma patients. *Journal of Nuclear Medicine* 2014;**55**:569-73.

Simon 2016 *{published data only}*

Barna S, Miltenyi Z, Simon Z, Jona A, Magyari F, Nagy Z, et al. Prognostical value of Interim and restaging PET/CT on Hodgkin lymphoma with CHEAP (Chemotherapy Effectiveness Assessed by PET/CT) long term observation. *European Journal of Nuclear Medicine and Molecular Imaging* 2014;**41**(Suppl. 2):S511.

Illes A, Magyari F, Barna S, Simon Z, Payer E, Garai I, et al. Experiences of the first two years with interim PET/CT in Hodgkin lymphoma - the Hungarian CHEAP study. *Haematologica* 2010;**95**(Suppl. 4):S45.

Miltenyi Z, Barna S, Garai I, Simon Z, Jona A, Magyari F, et al. Prognostic value of interim and restaging PET/CT in Hodgkin lymphoma. Results of the CHEAP (Chemotherapy Effectiveness Assessment by PET/CT) study - long term observation. *Neoplasma* 2015;**62**:627-34.

Miltenyi Z, Simon Z, Jona A, Magyari F, Barna S, Garai I, et al. Interim PET/CT in Hodgkin lymphoma - final results of the Hungarian CHEAP study (2007-2011). *Haematologica* 2013;**98**(Suppl. 2):42.

* Simon Z, Barna S, Miltenyi Z, Husi K, Magyari F, Jona A, et al. Combined prognostic value of absolute lymphocyte/monocyte ratio in peripheral blood and interim PET/CT results in Hodgkin lymphoma. *International Journal of Hematology* 2016;**103**(1):63-9.

Straus 2011 *{published data only}*

Kostakoglu L, Gandikota N, Hutchings M, Cotter R, Lamonica D, Nanni C, et al. Deauville criteria and post One-Cycle SUVmax decrease seem to predict progression free survival (PFS) better than metabolic tumor measurements in classical Hodgkin lymphoma (cHL). *European Journal of Nuclear Medicine and Molecular Imaging* 2012;**39**(Suppl. 2):S222.

Kostakoglu L, Schoder H, Hall N, Straus DJ, Johnson JL, Schwartz L, et al. Interim FDG PET imaging in CALGB 50203 trial of stage I/II non-bulky Hodgkin lymphoma: Would using combined PET and CT criteria better predict response than each test alone? *Blood* 2011;**118**(21):3644.

Kostakoglu L, Schoder H, Johnson JL, Hall NC, Schwartz LH, Straus DJ, et al. Interim [(18)F]fluorodeoxyglucose positron emission tomography imaging in stage I-II non-bulky Hodgkin lymphoma: would using combined positron emission tomography and computed tomography criteria better predict response than each test alone? *Leukemia & Lymphoma* 2012;**53**:2143-50.

* Straus DJ, Johnson JL, LaCasce AS, Bartlett NL, Kostakoglu L, Hsi ED, et al. Doxorubicin, vinblastine, and gemcitabine (CALGB 50203) for stage I/II nonbulky Hodgkin lymphoma: pretreatment prognostic factors and interim PET. *Blood* 2011;**117**:5314-20.

Touati 2014 *{published data only}*

Hutchings M, Kamper P. New clues to the prognostic challenge of Hodgkin Lymphoma. *Leukemia & Lymphoma* 2015;**56**(2):277-8.

* Touati M, Delage-Corre M, Monteil J, Abraham J, Moreau S, Remenieras L, et al. CD68-positive tumor-associated macrophages predict unfavorable treatment outcomes in classical Hodgkin lymphoma in correlation with interim fluorodeoxyglucose-positron emission tomography assessment. *Leukemia and Lymphoma* 2015;**56**:332-41.

Ying 2014 *{published data only}*

Ying Z, Wang X, Song Y, Zheng W, Wang X, Xie Y, et al. Prognostic value of 18F-FDG PET-CT in Hodgkin lymphoma. *Chinese Journal Of Hematology* 2014;**35**:325-7.

Zaucha 2017 *{published data only}*

Adams HJ, Kwee TC. The predictive value of interim FDG-PET in early-stage Hodgkin lymphoma is not well established. *Annals of Oncology* 2018;**29**(2):510-2.

Zaucha JM, Chauvie S, Malkowski B, Warszewska A, Biggi A, Kobylecka M, et al. The prognostic role of interim PET after first chemotherapy cycle in ABVD-treated Hodgkin lymphoma (HL) patients - Polish lymphoma research group (PLRG) observational study. *Haematologica* 2013;**98**(Suppl. 2):38.

* Zaucha JM, Malkowski B, Chauvie S, Subocz E, Tajer J, Kulikowski W, et al. The predictive role of interim PET after the first chemotherapy cycle and sequential evaluation of response to ABVD in Hodgkin lymphoma patients - the Polish Lymphoma Research Group (PLRG) Observational Study. *Annals of Oncology* 2017;**28**(12):3051-7.

Zaucha JM, Malkowski B, Subocz E, Chauvie S, Tajer J, Kulikowski W, et al. The prognostic role of interim PET after first chemotherapy cycle and PET sequential evaluation of response to ABVD in <hodgkin lymphoma patients - The Polish Lymphoma Research Group (PLRG) observational study. *Blood* 2015;**126**(23):3943.

Zinzani 2012 *{published data only}*

Adams HJ, Kwee TC. Does interim 18F-FDG-PET response-adapted therapy really benefit advanced-stage Hodgkin lymphoma patients? *Nuclear Medicine Communications* 2016;**37**(12):1333-4.

Puccini B, Rigacci L, Zinzani PL, Broccoli A, Gallamini A, Merli F, et al. Early positive FDG-PET scan do not confirm its prognostic impact in localized bulky disease Hodgkin lymphoma patients. *Haematologica* 2011;**96**(Suppl. 3):26-7.

Rigacci L, Puccini B, Gallamini A, Merli F, Stelitano C, Balzarotti M, et al. Early FDG-PET scan confirms its prognostic impact also in localized stage, ABVD treated Hodgkin lymphoma patients. *Haematologica* 2009;**94**(Suppl. 2):34.

Rigacci L, Puccini B, Zinzani PL, Kovalchuk S, Broccoli A, Evangelista A, et al. Clinical characteristics of patients with negative interim-PET and positive final PET: Data from the prospective PET-oriented HD0801 study by Fondazione Italiana linfomi (FIL). *Hematological Oncology* 2017;**35**(Suppl. 2):38.

Rigacci L, Zinzani PL, Puccini B, Broccoli A, Gallamini A, Merli F, et al. Early FDG-PET scan confirms its prognostic impact also in localized stage, ABVD treated Hodgkin lymphoma patients. *Haematologica* 2010;**95**(Suppl. 2):474.

Rigacci L, Zinzani PL, Puccini B, Broccoli A, Gallamini A, Merli F, et al. Early FDG-PET scan confirms its prognostic impact also in localized stage, ABVD treated Hodgkin lymphoma patients. In: Haematologica. Vol. 96. 2010:S13.

Rigacci L, Zinzani PL, Puccini B, Broccoli A, Gallamini A, Merli F, et al. Early positive FDG-PET scan do not confirm its prognostic impact in bulky disease Hodgkin lymphoma patients. *Haematologica* 2011;**96**(Suppl. 2):320-1.

Rigacci L, Zinzani PL, Puccini B, Broccoli A, Gallamini A, Merli F, et al. IHP interpretation criteria of interim-PET scan confirms prognostic impact in early stage Hodgkin lymphoma patients without bulky disease. *Blood* 2010;**116**(21):3890.

Stefoni V, Broccoli A, Alinari L, Ambrosini V, Derenzini E, Fanti S, et al. Predictive role of early interim FDG-PET in Hodgkin lymphoma. *Blood* 2009;**114**(22):1659.

Stefoni V, Pellegrini C, Rigacci L, Broccoli A, Puccini B, Fanti S, et al. Interim PET 2: Is it a real prognostic factor in selecting two different subsets of Hodgkin disease patients? *Haematologica* 2011;**96**(Suppl. 3):123.

* Zinzani PL, Rigacci L, Stefoni V, Broccoli A, Puccini B, Castagnoli A, et al. Early interim 18F-FDG PET in Hodgkin's lymphoma: evaluation on 304 patients. *European Journal of Nuclear Medicine and Molecular Imaging* 2012;**39**:4-12.

Zinzani PL, Tani M, Fanti S, Alinari L, Musuraca G, Marchi E, et al. Early positron emission tomography (PET) restaging: a predictive final response in Hodgkin's disease patients. *Annals of Oncology* 2006;**17**:1296-300.

References to studies excluded from this review

Adams 2016 *{published data only}*

Adams HJ, Kwee TC. RAPID trial demonstrates low positive predictive value of interim FDG-PET in early-stage Hodgkin lymphoma after three cycles of ABVD. *Journal of Pediatric Hematology/Oncology* 2016;**38**(2):165.

Adams 2017 *{published data only}*

Adams HJ, Kwee TC. Predictive value of interim [18F]fluorodeoxyglucose-positron emission tomography in advanced-stage Hodgkin lymphoma is not well established. Journal of Clinical Oncology 2017;**35**(3):370-1.

Adams 2018 *{published data only}*

Adams HJ, Kwee TC. Interim FDG-PET has no value in selecting patients who require treatment modification in both early- and advanced-stage Hodgkin lymphoma. British Journal of Haematology 2018;**183**(1):129-31.

Adams 2018a *{published data only}*

Adams HJ, Kwee TC. No evidence to promote interim FDG-PET adapted therapy in the NCCN guidelines for Hodgkin lymphoma. Journal of the National Comprehensive Cancer Network 2018;**16**(3):226-7.

Adams 2018b *{published data only}*

Adams HJ, Kwee TC. Strikingly heterogeneous results among studies on interim fluorodeoxyglucose-positron emission tomography-adapted treatment in advanced-stage Hodgkin lymphoma. Journal of Clinical Oncology 2018;**36**(20):2123-4.

Adams 2019 *{published data only}*

Adams HJ, Kwee TC. Post-ABVD biopsy results, and not post-ABVD FDG-PET results, predict outcome in early-stage Hodgkin lymphoma. British Journal of Haematology 2019;**184**(2):290-2.

Advani 2007 *{published data only}*

Advani R, Maeda L, Lavori P, Quon A, Hoppe R, Breslin S, et al. Impact of positive positron emission tomography on prediction of freedom from progression after Stanford V chemotherapy in Hodgkin's disease. *Journal of Clinical Oncology* 2007;**25**:3902-7.

Afanasyev 2017 *{published data only}*

Afanasyev BV, Moiseev IS, Alekseev SM, Mikhailova NB, Kondakova EV, Ilyin NV, et al. Multicenter prospective escalation-deescalation PET-guided clinical study in classical type Hodgkin disease in the north-west of Russian federation (RNWOHG-HD1): Rationale and design. Cellular Therapy and Transplantation 2017;**6**(4):76-81.

Albano 2017 *{published data only}*

Albano D, Patti C, Lagalla R, Midiri M, Galia M. Whole-body MRI, FDG-PET/CT, and bone marrow biopsy, for the assessment of bone marrow involvement in patients with newly diagnosed lymphoma. *Journal of Magnetic Resonance Imaging* 2017;**45**(4):1082-9.

Albano 2018 *{published data only}*

Albano D, Patti C, Matranga D, Lagalla R, Midiri M, Galia M. Whole-body diffusion-weighted MR and FDG-PET/CT in Hodgkin Lymphoma: Predictive role before treatment and early assessment after two courses of ABVD. *European Journal of Radiology* 2018;**103**:90-8.

Altamirano 2008 *{published data only}*

Altamirano J, Esparza JR, de la Garza Salazar J, Calvo PS, Vera SR, Chalapud Revelo JR, et al. Staging, response to therapy, and restaging of lymphomas with18F-FDG PET. *Archives of Medical Research* 2008;**39**(1):69-77.

Ansell 2016 *{published data only}*

Ansell SM. Hodgkin lymphoma: 2016 update on diagnosis, risk-stratification, and management. *American Journal of Hematology* 2016;**91**(4):434-42.

Awan 2013 *{published data only}*

Awan UE, Siddiqui N, SaadUllah M, Bashir H, Farooqui ZS, Muzaffar N, et al. FDG-PET scan in assessing lymphomas and the application of Deauville Criteria. *Journal of the Pakistan Medical Association* 2013;**63**(6):725-30.

Barrington 2011a *{published data only}*

Barrington SF, O'Doherty MJ, Pike L, Franceschetto A, Cucca M, Brun E, et al. Standardised PET-CT reporting for an international multicentre trial in lymphoma (RATHL). *European Journal of Nuclear Medicine and Molecular Imaging* 2011;**38 (SUPPL. 2)**:S115.

Barrington 2017 *{published data only}*

Barrington SF, Johnson PW. <18F-FDG PET/CT in lymphoma: Has imaging-directed personalized medicine become a reality? *Journal of Nuclear Medicine* 2017;**58**(10):1539-44.

Bar-Shalom 2003 *{published data only}*

Bar-Shalom R, Yefremov N, Haim N, Dann EJ, Epelbaum R, Keidar Z, et al. Camera-based FDG pet and 67Ga SPECT in evaluation of lymphoma: Comparative study. *Radiology* 2003;**227**(2):353-60.

Basu 2009 *{published data only}*

Basu S. Early FDG-PET response-adapted risk stratification and further therapeutic decision-making in lymphoma: Will this replace the established prognostic indices and be the standard-of-care in clinical management? *European Journal of Nuclear Medicine and Molecular Imaging* 2009;**36**(12):2089-90.

Becherer 2002 *{published data only}*

Becherer A, Mitterbauer M, Jaeger U, Kalhs P, Greinix HT, Karanikas G, et al. Positron emission tomography with 2-fluoro-D-2-deoxyglucose (FDG-PET) predicts relapse of malignant lymphoma after high-dose therapy with stem cell transplantation. *Leukemia* 2002;**16**(2):260-7.

Bednaruk-Mlynski 2015 *{published data only}*

Bednaruk-Mlynski E, Pienkowska J, Skorzak A, Malkowski B, Kulikowski W, Subocz E, et al. Comparison of positron emission tomography/computed tomography with classical contrast-enhanced computed tomography in the initial staging of Hodgkin lymphoma. *Leukemia and Lymphoma* 2015;**56**(2):377-82.

Biggi 2012 *{published data only}*

Biggi A, Kostakoglu L, Barrington S, Chauvie S, Gregianin M, Meignan M, et al. How the threshold of positive results influence progression free survival (PFS) in advanced Hodgkin's lymphoma treated with ABVD: Experience from the International Validation Study. *European Journal of Nuclear Medicine and Molecular Imaging* 2012;**39**(2):S222.

Biggi 2017 *{published data only}*

Biggi A, Bergesio F, Menga M, Bianchi A, Fallanca F, Chauvie S, et al. Diagnostic accuracy of FDG PET/CT at the of treatment of Hodgkin lymphoma in the HD0607 trial. *Clinical and Translational Imaging* 2017;**5**:S44.

Bishop 2015 *{published data only}*

Bishop G. PET-directed therapy for Hodgkin's lymphoma. *New England Journal of Medicine* 2015;**373**(4):392.

Bjurberg 2006 *{published data only}*

Bjurberg M, Gustavsson A, Ohlsson T, Brun E. FDG-PET in the detection of residual disease and relapse in patients with Hodgkin's lymphoma. Experience from a Swedish centre. *Acta Oncologica* 2006;**45**(6):743-9.

Blum 2002 *{published data only}*

Blum R, Prince HM, Hicks RJ, Patrikeos A, Seymour J. Discordant response to chemotherapy detected by PET scanning: Unveiling of a second primary cancer. *American Journal of Clinical Oncology: Cancer Clinical Trials* 2002;**25**(4):368-70.

Bodet-Milin 2008 *{published data only}*

Bodet-Milin C, Kraeber-Bodere F, Moreau P, Campion L, Dupas B, Le Gouill S. Investigation of FDG-PET/CT imaging to guide biopsies in the detection of histological transformation of indolent lymphoma. *Haematologica* 2008;**93**(3):471-2.

Bodet-Milin 2009 *{published data only}*

Bodet-Milin C, Salaun PY, Crespin C, Vuillez JP, Kraeber-Bodere F. FDG-PET scanning in managing patients with lymphoma. *Medecine Nucleaire* 2009;**33**(8):486-90.

Boisson 2007 *{published data only}*

Boisson N, Cachin F, Kelly A, De Freitas D, Isnardi V, Mestas D, et al. PET-CT in the Hodgkin's disease. *Medecine Nucleaire* 2007;**31**(10):562-7.

Borchmann 2016 *{published data only}*

Borchmann S, von Tresckow B, Engert A. Current developments in the treatment of early-stage classical Hodgkin lymphoma. *Current Opinion in Oncology* 2016;**28**(5):377-83.

Bucerius 2006 *{published data only}*

Bucerius J, Herkel C, Joe AY, Altehoefer C, Finke J, Moser E, et al. ¹⁸F-FDG PET and conventional imaging for assessment of Hodgkin's disease and non Hodgkin's lymphoma: An analysis of 193 patient studies. *Nuklearmedizin* 2006;**45**(3):105-10.

Carras 2018 *{published data only}*

Carras S, Dubois B, Senecal D, Jais JP, Peoc'h M, Quittet P, et al. Interim PET response-adapted strategy in untreated advanced stage Hodgkin lymphoma: Results of GOELAMS LH 2007 phase 2 multicentric trial. *Clinical Lymphoma, Myeloma & Leukemia* 2018;**18**(3):191-8.

Ciammella 2016 *{published data only}*

Ciammella P, Filippi AR, Simontacchi G, Buglione M, Botto B, Mangoni M, et al. Post-ABVD/pre-radiotherapy 18F-FDG-PET provides additional prognostic information for early-stage Hodgkin lymphoma: A retrospective analysis on 165 patients. *British Journal of Radiology* 2016;**89**(1061):20150983.

Cremerius 1999 *{published data only}*

Cremerius U, Fabry U, Kroll U, Zimny M, Neuerburg J, Osieka R, et al. [Clinical value of FDG PET for therapy monitoring of malignant lymphoma--results of a retrospective study in 72 patients]. *Clinic for Nuclear Medicine (Stuttgart)* 1999;**38**(1):24-30.

Cremerius 2001 *{published data only}*

Cremerius U, Fabry U, Neuerburg J, Zimny M, Bares R, Osieka R, et al. Prognostic significance of positron emission tomography using fluorine-18-fluorodeoxyglucose in patients treated for malignant lymphoma. *Clinic for Nuclear Medicine (Stuttgart)* 2001;**40**(1):23-30.

Cuccaro 2016 *{published data only}*

Cuccaro A, Annunziata S, Cupelli E, Martini M, Calcagni ML, Rufini V, et al. CD68+ cell count, early evaluation with PET and plasma TARC levels predict response in Hodgkin lymphoma. *Cancer Medicine* 2016;**5**(3):398-406.

D'Urso 2018 *{published data only}*

D'Urso D, Stefano A, Romano A, Russo G, Cosentino S, Fallanca F, et al. Analysis of metabolic parameters coming from basal and interim PET in Hodgkin lymphoma. *Current Medical Imaging Reviews* 2018;**14**(4):533-44.

Damlaj 2017 *{published data only}*

Damlaj M, Al-Zahrani M, Syed G, Gmati G, Pasha T, Abuelgasim K, et al. Escalation from ABVD following positive interim functional imaging improves progression free survival but not overall survival in advanced classical Hodgkin lymphoma-a real world analysis. *Blood* 2017;**130**(Suppl. 1):2799.

Damlaj 2019 *{published data only}*

Damlaj M, Al-Zahrani M, Syed G, Gmati G, Alahmari B, Pasha T, et al. Interim functional imaging is an independent predictor of progression-free survival in advanced classical Hodgkin lymphoma - A real-world analysis. *Clinical Lymphoma, Myeloma and Leukemia* 2019;**19**(1):e71-9.

Danilov 2017 *{published data only}*

Danilov AV, Li H, Press OW, Shapira I, Swinnen LJ, Noy A, et al. Feasibility of interim positron emission tomography (PET)-adapted therapy in HIV-positive patients with advanced Hodgkin lymphoma (HL): a sub-analysis of SWOG S0816 Phase 2 trial. *Leukemia & Lymphoma* 2017;**58**(2):461-5.

Dann 2009 *{published data only}*

Dann EJ, Bar-Shalom R, Tamir A, Ben-Shachar M, Avivi I, Zuckerman T, et al. For standard and high-risk patients with Hodgkin lymphoma six cycles of tailored BEACOPP, based on interim scintigraphy, are effective and female fertility is preserved. *Blood* 2009;**114**(22):1552.

Dann 2010 *{published data only}*

Dann EJ, Bar-Shalom R, Tamir A, Epelbaum R, Avivi I, Ben-Shachar M, et al. A functional dynamic scoring model to elucidate the significance of post-induction interim fluorine-18-fluorodeoxyglucose positron emission tomography findings in patients with Hodgkin's lymphoma. *Haematologica* 2010;**95**(7):1198-206.

Dann 2010a *{published data only}*

Dann EJ, Bairey O, Bar-Shalom R, Izak M, Korenberg A, Akria L, et al. Tailored therapy in Hodgkin lymphoma, based on predefined risk factors and early interim PET/CT, can lead to modification and safe reduction in therapy: Results of 134 patients on the Israel National Hodgkin Study. *Blood* 2010;**116**(21):2809.

Dann 2012 *{published data only}*

Dann EJ, Bairey O, Bar-Shalom R, Mashiach T, Izak M, Korenberg A, et al. Early Hodgkin lymphoma therapy, based on predefined risk factors and early interim PET/CT. Israeli H2 protocol: Preliminary report. *Haematologica* 2012;**97**(Suppl. 1):85.

Dann 2013 *{published data only}*

Dann EJ, Bairey O, Bar-Shalom R, Izak M, Korenberg A, Akria L, et al. Tailored therapy in Hodgkin lymphoma, based on predefined risk factors and early interim PET/CT, Israeli H2 protocol: Preliminary report on 317 patients. *Haematologica* 2013;**98**(Suppl. 2):37.

Dann 2016 *{published data only}*

Dann EJ, Bairey O, Bar-Shalom R, Mashiach T, Barzilai E, Korenberg A, et al. Adjustment of therapy for Hodgkin lymphoma based on interim PET is beneficial and radiotherapy may be substituted with chemotherapy in patients with negative interim study: Final results of H2 trial. European Haematology Association Open Access Library 2016;(S107):4-5.

Dann 2017 *{published data only}*

Dann EJ, Bairey O, Bar-Shalom R, Mashiach T, Barzilai E, Kornberg A, et al. Modification of initial therapy in early and advanced Hodgkin lymphoma, based on interim PET/CT is beneficial: a prospective multicentre trial of 355 patients. *British Journal of Haematology* 2017;**178**:709-718.

Dann 2018 *{published data only}*

Dann EJ, Paltiel O. Interim FDG-PET has no value in selecting patients who require treatment modification in both early- and advanced-stage Hodgkin lymphoma: Response to Adams and Kwee. *British Journal of Haematology* 2018;**183**(1):131-3.

deAndres-Galiana 2015 *{published data only}*

deAndres-Galiana EJ, Fernandez-Martinez JL, Luaces O, del Coz JJ, Fernandez R, Solano J, et al. On the prediction of Hodgkin lymphoma treatment response. *Clinical and Translational Oncology* 2015;**17**(8):612-9.

Diehl 2007 *{published data only}*

Diehl V, Kobe C, Haverkamp H, Dietlein M, Engert A. FDG-PET for assessment of residual tissue after completion of chemotherapy in Hodgkin lymphoma report on the 2nd interim analysis of the PET investigation in the trial HD15 of the GHSG. In: Blood. The American Society of Hematology. 2007.

El-Galaly 2012 *{published data only}*

El-Galaly TC, Mylam KJ, Brown P, Specht L, Christiansen I, Munksgaard L, et al. Positron emission tomography/computed tomography surveillance in patients with Hodgkin lymphoma in first remission has a low positive predictive value and high costs. *Haematologica* 2012;**97**(6):931-6.

Evens 2014 *{published data only}*

Evens AM, Kostakoglu L. The role of FDG-PET in defining prognosis of Hodgkin lymphoma for early-stage disease. *Blood* 2014;**124**(23):3356-64.

Fanti 2008 *{published data only}*

Fanti S, Castellucci P, Stefoni V, Nanni C, Tani M, Rubello D, et al. Early relapse in a patient with Hodgkin's disease and negative interim FDG-PET. *Annals of Nuclear Medicine* 2008;**22**(5):429-32.

Filmont 2003 *{published data only}*

Filmont J E, Czernin J, Yap C, Silverman DH, Quon A, Phelps ME, et al. Value of F-18 fluorodeoxyglucose positron emission tomography for predicting the clinical outcome of patients with aggressive lymphoma prior to and after autologous stem-cell transplantation. *Chest* 2003;**124**(2):608-13.

Fornecker 2017 *{published data only}*

Fornecker LM, Lazarovici J, Aurer I, Casasnovas RO, Gac AC, Bonnet C, et al. Pet-based response after 2 cycles of brentuximab vedotin in combination with avd for first-line treatment of unfavorable early-stage Hodgkin Lymphoma: first analysis of the primary endpoint of breach, a randomized phase II trial of LYSA-FIL-EORTC intergroup. *Blood* 2017;**130**(Suppl. 1):736.

Freudenberg 2004 *{published data only}*

Freudenberg LS, Antoch G, Schutt P, Beyer T, Jentzen W, Muller SP, et al. FDG-PET/CT in re-staging of patients with lymphoma. *European Journal of Nuclear Medicine and Molecular Imaging* 2004;**31**(3):325-9.

Friedberg 2002 *{published data only}*

Friedberg JW, Fischman A, Neuberg D, Kim H, Takvorian T, Mauch PM, et al. FDG-PET is superior to gallium scintigraphy in the staging and follow-up of patients with de novo Hodgkin's disease: a prospective, blinded comparison. Leukemia & Lymphoma 2004;**45**(1):85-92.

Friedberg 2004 *{published data only}*

Friedberg JW, Fischman A, Neuberg D, Kim H, Takvorian T, Ng AK, et al. FDG-PET is superior to gallium scintigraphy in staging and more sensitive in the follow-up of patients with de novo Hodgkin lymphoma: a blinded comparison. *Leukemia and Lymphoma* 2004;**45**(1):85-92.

Front 1999 *{published data only}*

Front D, Bar-Shalom R, Mor M, Haim N, Epelbaum R, Frenkel A, et al. Hodgkin disease: prediction of outcome with 67Ga scintigraphy after one cycle of chemotherapy. *Radiology* 1999;**210**(2):487-91.

Fruchart 2006 *{published data only}*

Fruchart C, Reman O, Le Stang N, Musafiri D, Cheze S, Macro M, et al. Prognostic value of early 18 fluorodeoxyglucose positron emission tomography and gallium-67 scintigraphy in aggressive lymphoma: A prospective comparative study. *Leukemia and Lymphoma* 2006;**47**(12):2547-57.

Gallamini 2008 *{published data only}*

Gallamini A, Hutchings M, Avigdor A, Polliack A. Early interim PET scan in Hodgkin lymphoma: Where do we stand? *Leukemia and Lymphoma* 2008;**49**(4):659-62.

Gallamini 2017 *{published data only}*

Gallamini A, Tarella C, Viviani S, Rossi A, Patti C, Mulé A, et al. Early chemotherapy intensification with escalated BEACOPP in advanced-stage Hodgkin lymphoma with a positive interim PET-CT after 2 ABVD cycles: Long-term results of the GITIL/FIL HD 0607 trial. *Journal of Clinical Oncology* 2018;**36**(5):454-462.

Gallamini 2018 *{published data only}*

Gallamini A, Tarella C, Viviani S, Rossi A, Patti C, Mule A, et al. Early chemotherapy intensification with escalated BEACOPP in patients with advanced-stage Hodgkin lymphoma with a positive interim positron emission tomography/computed tomography scan after two ABVD cycles: Long-term results of the GITIL/FIL HD 0607 trial. *Journal of Clinical Oncology* 2018;**36**(5):454-62.

Gallamini 2018a *{published data only}*

Gallamini A, Viviani S, Pavoni C, Rambaldi A. Reply to H.J.A. Adams et al and C. Mesguich et al. *Journal of Clinical Oncology* 2018;**36**(20):2127-8.

Gallowitsch 2008 *{published data only}*

Gallowitsch HJ, Igerc I, Kohlfurst S, Lind P. The incremental value of F-18 FDG PET and PET/CT in malignant lymphoma. *Imaging Decisions MRI* 2008;**12**(4):2-6.

Goldschmidt 2011 *{published data only}*

Goldschmidt N, Or O, Klein M, Savitsky B, Paltiel O. The role of routine imaging procedures in the detection of relapse of patients with Hodgkin lymphoma and aggressive non-Hodgkin lymphoma. *Annals of Hematology* 2011;**90**(2):165-71.

Greil 2018 *{published data only}*

Greil R. Treatment optimisation trial in the first-line treatment of advanced stage Hodgkin lymphoma; comparison of 4-6 cycles of escalated BEACOPP with 4-6 cycles of BrECADD. *Memo - Magazine of European Medical Oncology.* 2018;**11**(1):1-28.

Guidez 2016 *{published data only}*

Guidez S, Delwail V, Brice P. PET-scan in management of Hodgkin's lymphoma. *Hematologie* 2016;**22**(6):389-91.

Hagtvedt 2015 *{published data only}*

Hagtvedt T, Seierstad T, Lund KV, Løndalen AM, Bogsrud TV, Smith HJ, et al. Diffusion-weighted MRI compared to FDG PET/CT for assessment of early treatment response in lymphoma. *Acta Radiologica* 2015;**56**(2):152-8.

Haioun 2005 *{published data only}*

Haioun C, Itti E, Rahmouni A, Brice P, Rain JD, Belhadj K, et al. [18F]fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) in aggressive lymphoma: an early prognostic tool for predicting patient outcome. *Blood* 2005;**106**(4):1376-81.

Hartmann 2012 *{published data only}*

Hartmann S, Agostinelli C, Diener J, Doring C, Fanti S, Zinzani PL, et al. GLUT1 expression patterns in different Hodgkin lymphoma subtypes and progressively transformed germinal centers. *BMC Cancer* 2012;**12**:586.

Hartridge-Lambert 2013 *{published data only}*

Hartridge-Lambert SK, Schoder H, Lim RC, Maragulia JC, Portlock CS. ABVD alone and a PET scan complete remission negates the need for radiologic surveillance in early-stage, nonbulky Hodgkin lymphoma. *Cancer* 2013;**119**(6):1203-9.

Honda 2014 *{published data only}*

Honda A, Nakamura F, Nannya Y, Shintani Y, Fukayama M, Ichikawa M, et al. Pulmonary lymphocyte-rich classical Hodgkin lymphoma with early response to ABVD therapy. *Annals of Hematology* 2014;**93**(6):1073-4.

Hueltenschmidt 2001 *{published data only}*

Hueltenschmidt B, Sautter-Bihl ML, Lang O, Maul FD, Fischer J, Mergenthaler HG, et al. Whole body positron emission tomography in the treatment of Hodgkin disease. *Cancer* 2001;**91**(2):302-10.

Huic 2006 *{published data only}*

Huic D, Mutvar A, Radman I, Grosev D, Labar B, Zuvic M, et al. The value of F-18 FDG triple-head coincidence PET in the posttreatment evaluation of patients with lymphoma. *Clinical Nuclear Medicine* 2006;**31**(5):275-8.

Hutchings 2007 *{published data only}*

Hutchings M, Loft A, Hansen M, Berthelsen AK, Specht L. Clinical impact of FDG-PET/CT in the planning of radiotherapy for early-stage Hodgkin lymphoma. *European Journal of Haematology* 2007;**78**(3):206-12.

Iagaru 2008 *{published data only}*

Iagaru A, Wang Y, Mari C, Quon A, Goris ML, Horning S, et al. (18)F-FDG-PET/CT evaluation of response to treatment in lymphoma: when is the optimal time for the first re-evaluation scan? *Hellenic Society of Nuclear Medicine* 2008;**11**(3):153-6.

Illidge 2015 *{published data only}*

Illidge T. Personalised approach to treating early Hodgkin's lymphoma. *BMJ* 2015;**350**:h2927.

Jerusalem 2003 *{published data only}*

Jerusalem G, Beguin Y, Fassotte MF, Belhocine T, Hustinx R, Rigo P, et al. Early detection of relapse by whole-body positron emission tomography in the follow-up of patients with Hodgkin's disease. *Annals of Oncology* 2003;**14**(1):123-30.

Johnson 2015 *{published data only}*

Johnson PW, Federico M, Fossa A, O'Doherty M, Roberts T, Stevens L, et al. Response-adapted therapy based on interim FDG-PET scans in advanced Hodgkin lymphoma: first analysis of the safety of deescalation and efficacy of escalation in the international RATHL study (CRUK/07/033). *Clinical Advances in Hematology and Oncology* 2015;**13**(8 Suppl. 9):6-7.

Johnson 2016 *{published data only}*

Johnson P, Federico M, Kirkwood A, Fossa A, Berkahn L, Carella A, et al. Adapted treatment guided by interim PET-CT scan in advanced Hodgkin's lymphoma. *New England Journal of Medicine* 2016;**374**(25):2419-29.

Kamran 2016 *{published data only}*

Kamran SC, Jacene HA, Chen YH, Mauch PM, Ng AK. Clinical outcome of patients with early stage favorable Hodgkin lymphoma treated with ABVDX2 cycles followed by PET/CT restaging and 20 Gy of involved-site radiotherapy. *Haematologica* 2016;**101**(Suppl. 5):14.

Kamran 2018 *{published data only}*

Kamran SC, Jacene HA, Chen YH, Mauch PM, Ng AK. Clinical outcome of patients with early stage favorable Hodgkin lymphoma treated with ABVD x two cycles followed by FDG-PET/CT restaging and 20 Gy of involved-site radiotherapy. *Leukemia and Lymphoma* 2018;**59**(6):1384-90.

Kobe 2008 *{published data only}*

Kobe C, Dietlein M, Franklin J, Markova J, Lohri A, Amthauer H, et al. Positron emission tomography has a high negative predictive value for progression or early relapse for patients with residual disease after first-line chemotherapy in advanced-stage Hodgkin lymphoma. *Blood* 2008;**112**(10):3989-94.

Kobe 2014 *{published data only}*

Kobe C, Kuhnert G, Kahraman D, Haverkamp H, Eich HT, Franke M, et al. Assessment of tumor size reduction improves outcome prediction of positron emission tomography/computed tomography after chemotherapy in advanced-stage Hodgkin lymphoma. *Journal of Clinical Oncology* 2014;**32**(17):1776-81.

Kostakoglu 2006 *{published data only}*

Kostakoglu L, Goldsmith SJ, Leonard JP, Christos P, Furman RR, Atasever T, et al. FDG-PET after 1 cycle of therapy predicts outcome in diffuse large cell lymphoma and classic Hodgkin disease. *Cancer* 2006;**107**:2678-87.

Li 2013 *{published data only}*

Li YJ, Li ZM, Xia XY, Huang HQ, Xia ZJ, Lin TY, et al. Prognostic value of interim and posttherapy 18F-FDG PET/CT in patients with mature T-cell and natural killer cell lymphomas. *Journal of Nuclear Medicine* 2013;**54**(4):507-15.

Lowe 2002 *{published data only}*

Lowe VJ, Wiseman GA. Assessment of Lymphoma Therapy Using (18)F-FDG PET. *Journal of Nuclear Medicine* 2002;**43**(8):1028-30.

Milgrom 2017 *{published data only}*

Milgrom SA, Pinnix CC, Chuang H, Oki Y, Akhtari M, Mawlawi O, et al. Early-stage Hodgkin lymphoma outcomes after combined modality therapy according to the post-chemotherapy 5-point score: can residual pet-positive disease be cured with radiotherapy alone? *British Journal of Haematology* 2017;**179**(3):488-96.

Mocikova 2010 *{published data only}*

Mocikova H, Obrtlikova P, Vackova B, Trneny M. Positron emission tomography at the end of first-line therapy and during follow-up in patients with Hodgkin lymphoma: a retrospective study. *Annals of Oncology* 2010;**21**(6):1222-7.

Mocikova 2011 *{published data only}*

Mocikova H, Pytlik R, Markova J, Steinerova K, Kral Z, Belada D, et al. Pre-transplant positron emission tomography in patients with relapsed Hodgkin lymphoma. *Leukemia & Lymphoma* 2011;**52**(9):1668-74.

Molnar 2010 *{published data only}*

Molnar Z, Simon Z, Borbenyi Z, Deak B, Galuska L, Keresztes K, et al. Prognostic value of FDG-PET in Hodgkin lymphoma for posttreatment evaluation. Long term follow-up results. *Neoplasma* 2010;**57**(4):349-54.

Moskowitz 2015 *{published data only}*

Moskowitz CH. Early FDG-PET adapted treatment improves the outcome of early FDG-PET-positive patients with stages I/II hodgkin lymphoma (HL): Final results of the randomized intergroup EORTC/LYSA/FIL H10 trial. *Clinical advances in Hematology & Oncology: H&O* 2015;**13**(8 Supplement 9):16-7.

Naumann 2001 *{published data only}*

Naumann R, Vaic A, Beuthien-Baumann B, Bredow J, Kropp J, Kittner T, et al. Prognostic value of positron emission tomography in the evaluation of post-treatment residual mass in patients with Hodgkin's disease and non-Hodgkin's lymphoma. *British Journal of Haematology* 2001;**115**(4):793-800.

NCT00784537 *{published data only}*

NCT00784537. High-dose chemotherapy and stem cell transplantation, in patients PET-2 positive, after 2 courses of ABVD and comparison of RT versus no RT in PET-2 negative patients. clinicaltrials.gov/show/nct00784537 (first received 4 November 2008).

NCT00795613 *{published data only}*

NCT00795613. Positron Emission Tomography (PET)-adapted chemotherapy In advanced Hodgkin Lymphoma (HL). clinicaltrials.gov/show/nct00795613 (first received 21 November 2008).

NCT01358747 *{published data only}*

NCT01358747. Study of a treatment driven by early PET response to a treatment not monitored by early PET in patients with AA stage 3-4 or 2B HL. clinicaltrials.gov/show/nct01358747 (first received 24 May 2011).

NCT01652261 *{published data only}*

NCT01652261. Very early FDG-PET/CT-response adapted therapy for advanced Hodgkin Lymphoma (H11). clinicaltrials.gov/show/nct01652261 (first received 30 July 2012).

NCT02292979 *{published data only}*

NCT02292979. Brentuximab vedotin associated with chemotherapy in untreated patients with Hodgkin Lymphoma. clinicaltrials.gov/show/nct02292979 (first received 18 November 2014).

Nguyen 2017 *{published data only}*

Nguyen VT, Pophali PA, Tsai J P, Jagadeesh D, Dean RM, Pohlman B, et al. Early stage, bulky Hodgkin lymphoma patients have a favorable outcome when treated with or without consolidative radiotherapy: potential role of PET scan in treatment planning. *British Journal of Haematology* 2017;**179**(4):674-6.

Panizo 2004 *{published data only}*

Panizo C, Perez-Salazar M, Bendandi M, Rodriguez-Calvillo M, Boan JF, Garcia-Velloso MJ, et al. Positron emission tomography using 18F-fluorodeoxyglucose for the evaluation of residual Hodgkin's disease mediastinal masses. *Leukemia and Lymphoma* 2004;**45**(9):1829-33.

Paolini 2007 *{published data only}*

Paolini R, Rampin L, Rodella E, Ramazzina E, Banti E, Al-Nahhas A, et al. The prognostic value of 18F-FDG PET-CT in the management of Hodgkin's lymphoma: Preliminary results of a prospective study. *Nuclear Medicine Review* 2007;**10**(2):87-90.

Pavlovsky 2019 *{published data only}*

Pavlovsky A, Fernandez I, Kurgansky N, Prates V, Zoppegno L, Negri P, et al. PET-adapted therapy after three cycles of ABVD for all stages of Hodgkin lymphoma: results of the GATLA LH-05 trial. *British Journal of Haematology* 2019;**185**(5):865-73.

Pichler 2000 *{published data only}*

Pichler R, Maschek W, Hatzl-Griesenhofer M, Huber H, Wimmer G, Wahl G, et al. Clinical value of FDG hybrid-PET in staging and restaging of malignant lymphoma - compared with conventional diagnostic methods. *NuklearMedizin* 2000;**39**(6):166-73.

Reinhardt 2005 *{published data only}*

Reinhardt MJ, Herkel C, Altehoefer C, Finke J, Moser E. Computed tomography and 18F-FDG positron emission tomography for therapy control of Hodgkin's and non-Hodgkin's lymphoma patients: when do we really need FDG-PET? *Annals of Oncology* 2005;**16**(9):1524-9.

Rigacci 2002 *{published data only}*

Rigacci L, Castagnoli A, Carpaneto A, Carrai V, Vaggelli L, Matteini M. Can (18)F-FDG PET after first cycle chemotherapy predict the efficacy of therapy in Hodgkin's disease? *Haematologica* 2002;**87**(5):ELT24.

Rigacci 2017 *{published data only}*

Rigacci L, Puccini B, Zinzani P, Kovalchuk S, Broccoli A, Evangelista A, et al. Clinical characteristics of patients with negative interim-pet and positive final PET: data from the prospective PET-oriented HD0801 study by Fondazione Italiana linfomi (FIL). *Hematological Oncology* 2017;**35**(Suppl. 2):38.

Rubello 2015 *{published data only}*

Rubello D, Gordien P, Morliere C, Guyot M, Bordenave L, Colletti PM, et al. Variability of hepatic 18F-FDG uptake at interim PET in patients with Hodgkin lymphoma. *Clinical Nuclear Medicine* 2015;**40**:e405-10.

Sakr 2017 *{published data only}*

Sakr R, Massoud M, Kerbage F, Rached L, Zeghondy J, Akoury E, et al. Real-life Experience for Integration of PET-CT in the Treatment of Hodgkin Lymphoma in Lebanon. *Clinical Lymphoma, Myeloma & Leukemia* 2017;**17S**:S92-5.

Schot 2007 *{published data only}*

Schot BW, Zijlstra JM, Sluiter WJ, van Imhoff GW, Pruim J, Vaalburg W, et al. Early FDG-PET assessment in combination with clinical risk scores determines prognosis in recurring lymphoma. *Blood* 2007;**109**(2):486-91.

Simontacchi 2015 *{published data only}*

Simontacchi G, Filippi AR, Ciammella P, Buglione M, Saieva C, Magrini SM, et al. Interim PET after two ABVD cycles in early-stage Hodgkin lymphoma: Outcomes following the continuation of chemotherapy plus radiotherapy. *International Journal of Radiation Oncology Biology Physics* 2015;**92**:1077-83.

Slaby 2002 *{published data only}*

Slaby J, Belohlavek O, Taborska K, Prochazka M, Trneny M, Klener P. [Predictive features of positron emission tomography after two cycles of induction therapy in malignant lymphoma]. *Casopis Lékaru Ceských [Journal of Czech Physicians]* 2002;**141**(10):312-5.

Spaepen 2001 *{published data only}*

Spaepen K, Stroobants S, Dupont P, Thomas J, Vandenberghe P, Balzarini J, et al. Can positron emission tomography with [(18)F]-fluorodeoxyglucose after first-line treatment distinguish Hodgkin's disease patients who need additional therapy from others in whom additional therapy would mean avoidable toxicity? *British Journal of Haematology* 2001;**115**:272-8.

Specht 2007 *{published data only}*

Specht L. FDG-PET scan and treatment planning for early stage Hodgkin lymphoma. *Radiotherapy and Oncology* 2007;**85**(2):176-7.

Spinner 2018 *{published data only}*

Spinner MA, Advani RH. Risk-adapted therapy for advanced-stage Hodgkin lymphoma. *Hematology* 2018;**1**:200-6.

Straus 2018 *{published data only}*

Straus DJ, Jung SH, Pitcher B, Kostakoglu L, Grecula JC, Hsi ED, et al. CALGB 50604: risk-adapted treatment of nonbulky early-stage Hodgkin lymphoma based on interim PET. *Blood* 2018;**132**(10):1013-21.

Strigari 2016 *{published data only}*

Strigari L, Attili A, Duggento A, Chiaravalloti A, Schillaci O, Guerrisi MG. Quantitative analysis of basal and interim PET/CT images for predicting tumor recurrence in patients with Hodgkin's lymphoma. *Nuclear Medicine Communications* 2016;**37**:16-22.

Sucak 2011 *{published data only}*

Sucak GT, Ozkurt ZN, Suyani E, Yasar DG, Akdemir OU, Aki Z, et al. Early post-transplantation positron emission tomography in patients with Hodgkin lymphoma is an independent prognostic factor with an impact on overall survival. *Annals of Hematology* 2011;**90**(11):1329-36.

Tirelli 2015 *{published data only}*

Tirelli U, Spina M. PET-adapted salvage therapy in Hodgkin's lymphoma. *Lancet Oncology* 2015;**16**(3):239-40.

Tomita 2015 *{published data only}*

Tomita N, Hattori Y, Fujisawa S, Hashimoto C, Taguchi J, Takasaki H, et al. Post-therapy 18F-fluorodeoxyglucose positron emission tomography for predicting outcome in patients with peripheral T cell lymphoma. *Annals of Hematology* 2015;**94**(3):431-6.

Torizuka 2004 *{published data only}*

Torizuka T, Nakamura F, Kanno T, Futatsubashi M, Yoshikawa E, Okada H, et al. Early therapy monitoring with FDG-PET in aggressive non-Hodgkin's lymphoma and Hodgkin's lymphoma. *European Journal of Nuclear Medicine and Molecular Imaging* 2004;**31**(1):22-8.

Trotman 2017 *{published data only}*

Trotman J, Fossa A, Federico M, Stevens L, Kirkwood A, Clifton-Hadley L, et al. Response-adjusted therapy for advanced Hodgkin lymphoma (RATHL) trial: longer follow up confirms efficacy of de-escalation after a negative interim PET scan (CRUK/07/033). *Hematological Oncology* 2017;**35**:65-7.

Tseng 2012 *{published data only}*

Tseng D, Rachakonda LP, Su Z, Advani R, Horning S, Hoppe RT, et al. Interim-treatment quantitative PET parameters predict progression and death among patients with Hodgkin's disease. *Radiation Oncology* 2012;**7**:5.

Villa 2018 *{published data only}*

Villa D, Sehn LH, Aquino-Parsons C, Tonseth P, Scott D W, Gerrie AS, et al. Interim PET-directed therapy in limited stage Hodgkin lymphoma initially treated with ABVD. *Haematologica* 2018;**12**:12.

Weidmann 1999 *{published data only}*

Weidmann E, Baican B, Hertel A, Baum RP, Chow KU, Knupp B, et al. Positron emission tomography (PET) for staging and evaluation of response to treatment in patients with Hodgkin's disease. *Leukemia and Lymphoma* 1999;**34**(5-6):545-51.

Wilson 2018 *{published data only}*

Wilson D, Benard F, Gascoyne RD, Slack GW, Farinha P, Morris J, et al. Interim PET-directed therapy in limited-stage Hodgkin lymphoma initially treated with ABVD. *Haematologica* 2018;**103**(12):e590-3.

Xie 2018 *{published data only}*

Xie W, Jiang X F, Zhao W L, Wang L. Prognostic evaluation of different PET/CT reading methods in Hodgkin lymphoma and diffused large B-cell lymphoma. *Journal of Shanghai Jiaotong University* 2018;**38**(8):954-9.

Yasgur 2015 *{published data only}*

Yasgur BS. Interim PET results guide ongoing therapy in Hodgkin lymphoma. *Oncology Report* 2015;**11**(8):available at: https://www.mdedge.com/hematologynews/nhlhub/article/101136/indolent-lymphoma/interim-pet-results-guide-ongoing-therapy.

Yoshimi 2008 *{published data only}*

Yoshimi A, Izutsu K, Takahashi M, Kako S, Oshima K, Kanda Y, et al. Conventional allogeneic hematopoietic stem cell transplantation for lymphoma may overcome the poor prognosis associated with a positive FDG-PET scan before transplantation. *American Journal of Hematology* 2008;**83**(6):477-81.

Zabrocka 2016 *{published data only}*

Zabrocka E, Sierko E, Wojtukiewicz MZ. Positron emission tomography scanning in the management of Hodgkin lymphoma patients: a single-institution experience. *Advances in Clinical and Experimental Medicine* 2016;**25**(6):1185-92.

Zaucha 2009 *{published data only}*

Zaucha J, Danielewicz I, Malkowski B, Zaucha R, Lesniewski-Kmak K. The role of PET for interim response assessment in patients with Hodgkin's lymphoma. *Wspolczesna Onkologia* 2009;**13**(4):161-6.

Zinzani 1999 *{published data only}*

Zinzani PL, Magagnoli M, Chierichetti F, Zompatori M, Garraffa G, Bendandi M, et al. The role of positron emission tomography (PET) in the management of lymphoma patients. *Annals of Oncology* 1999;**10**(10):1181-4.

Zinzani 2002 *{published data only}*

Zinzani PL, Chierichetti F, Zompatori M, Tani M, Stefoni V, Garraffa G, et al. Advantages of positron emission tomography (PET) with respect to computed tomography in the follow-up of lymphoma patients with abdominal presentation. *Leukemia and Lymphoma* 2002;**43**(6):1239-43.

Zinzani 2016 *{published data only}*

Zinzani PL, Broccoli A, Gioia DM, Castagnoli A, Ciccone G, Evangelista A, et al. Interim positron emission tomography response-adapted therapy in advanced-stage hodgkin lymphoma: final results of the phase II part of the HD0801 study. Journal of Clinical Oncology 2016;**34**(12):1376-85.

References to studies awaiting assessment

Abramson 2010 *{published data only}*

Abramson JS, Barnes JA, LaCasce AS, Zukotynski K, Israel D, Feng Y, et al. End of treatment but not interim PET scan predicts outcome in non-bulky limited stage Hodgkin lymphoma. *Haematologica* 2010;**95**(Suppl. 4):S16.

Algrin 2010 *{published data only}*

Algrin C, Berenger N, Chevret S, Vercellino L, De Bazelaire C, Brice P, et al. Interim-positron emission tomography with [18F]fluorodeoxyglucose (interim-PET) evaluation in mediastinal lymphoma including Hodgkin lymphoma (HL) and primary mediastinal large B-cell lymphoma (PMBL). *Blood* 2010;**116**(21):2860.

Arce-Calisaya 2013 *{published data only}*

Arce-Calisaya P, Scarsbrook A, Thygesen H, Chowdhury F, Patel C. Interim FDG PET/CT in Hodgkin's lymphoma - Does binary response assessment criteria have any prognostic value? *European Journal of Nuclear Medicine and Molecular Imaging* 2013;**40**(Suppl. 2):S476.

Baratto 2015 *{published data only}*

Baratto L, Guerra L, Elisei F, Crivellaro C, De Ponti E, Bolis S, et al. Interim-PET in Hodgkin lymphoma: Deauville criteria and metabolic parameters as prognostic factors. *Clinical and Translational Imaging* 2015;**3**(Suppl. 1):S16-7.

Barna 2011 *{published data only}*

Barna S, Fedinecz N, Magyari F, Varga J, Illes A, Garai I. Prognostic value of interim 18FDG-PET/CT in patients with Hodgkin's lymphoma using different 5-point visual scales for interpretation. *European Journal of Nuclear Medicine and Molecular Imaging* 2011;**38**(Suppl. 2):S377.

Barrington 2011 *{published data only}*

Barrington SF, Kostakoglu L, Hutchings M, Meignan M, Biggi A, Gregianin M, et al. Are the Deauville criteria a reliable tool for assessment of interim PET in Hodgkin lymphoma? In: European Journal of Nuclear Medicine and Molecular Imaging. Vol. 38. 2011:S164.

Bentur 2017 *{unpublished data only}*

Bentur OS, Eldad DJ, Paran E, Lavie D, Nachmias B, Dally N, et al. The predictive value of interim PET-CT in elderly patients with Hodgkin lymphoma. In: Haematologica. Conference: 22th Congress of the European Hematology Association. Spain. Vol. 102. 2017:460-1.

Berenger 2010 *{published data only}*

Berenger N, Vercellino L, Algrin C, Groheux D, Hindie E, Lussato D, et al. Prognostic value of interim 18F-FDG PET/CT in mediastinal bulky Hodgkin lymphoma. *European Journal of Nuclear Medicine and Molecular Imaging* 2010;**37**(Suppl. 2):S436.

Bhatwadekar 2017 *{published data only}*

Bhatwadekar S, Deshpande S, Khadse S, Shah B, Desai D, Kachchhi U, et al. Excellent outcome in Hodgkin lymphoma with ABVD and CMT: A single-centre retrospective analysis. *Hematological Oncology* 2017;**35**(Suppl. 2):316-7.

Cimino 2014 *{published data only}*

Cimino G, Zaucha JM, Cirillo S, Saviolo C, Hutchings M, El-Galaly TC, et al. The complementary prognostic role of baseline and interim PET in predicting treatment outcome in advanced-stage Hodgkin lymphoma. *Blood* 2014;**124**(21):4405.

Cocorocchio 2009 *{published data only}*

Cocorocchio E, Vanazzi A, Botteri E, Alietti A, Negri M, Bassi S, et al. Prognostic role of interim 18FDG-PET in Hodgkin lymphoma: A single-center experience. *Journal of Clinical Oncology* 2009;**27**(15 suppl. 1):e19520.

Cocorocchio 2011 *{published data only}*

Cocorocchio E, Botteri E, Gigli F, Bassi S, Bertazzoni P, Sammassimo S, et al. Evaluation of interim 18FDG-PET in advanced Hodgkin lymphoma (HL) patients (PTS) treated with ChlVPP/ABVVP regimen. In: Annals of Oncology. Vol. 22. 2011:216.

Copeland 2010 *{published data only}*

Copeland A, Fanale M, Chuang H, Macapinlac H, Faria SC, Siegmund B, et al. Single institution experience with interim PET evaluation in newly diagnosed CHL receiving ABVD chemotherapy: Need for standardization. *Haematologica* 2010;**95**(Suppl. 4):S50.

Cuzzocrea 2015 *{published data only}*

Cuzzocrea M, Guerra L, Elisei F, Crivellaro C, De Ponti E, Bolis S, et al. The Deauville criteria and metabolic parameters as prognostic factors in interim PET in Hodgkin lymphoma: A single centre experience. *European Journal of Nuclear Medicine and Molecular Imaging* 2015;**42**(1 suppl. 1):S152-3.

De Rueda 2013 *{published data only}*

De Rueda B, Costilla L, Catalina S, Grasa J, Rubio D, Giraldo P. Prognostic value of 18F-FDG PET/CT in Hodgkin lymphoma. *Haematologica* 2013;**98**(Suppl. 1):572-3.

Fabbri 2011 *{published data only}*

Fabbri A, Rigacci L, Lazzi S, Di Lollo S, Pietrini A, Puccini B, et al. 'Early FDG-PET' predicts clinical course of Hodgkin's lymphoma although does not correlate with macrophages infiltration in diagnostic specimens. *Haematologica* 2011;**96**(Suppl. 2):321-2.

Fiore 2010 *{published data only}*

Fiore F, Viviani S, Luminari S, Levis A, Di Raimondo F, Merli F, et al. Early interim FDG-PET during intensified BEACOPP therapy for advanced-stage Hodgkin disease shows a lower positive predictive value than during ABVD. *Haematologica* 2010;**95**(Suppl. 4):S19.

Gallegos 2012 *{published data only}*

Gallegos C, De Rueda B, Grasa JM, Banso A, Giraldo P. The importance of PET/CT as method of evaluation of early response to treatment in HL. *Haematologica* 2012;**97**(Suppl. 1):566.

Hohaus 2015 *{published data only}*

Hohaus S, Cuccaro A, Annunziata S, Martini M, D'Alo F, Calcagni ML, et al. The risk of progression of Hodgkin lymphoma in patients with negative interim PET: A role for the number of tumor-infiltrating macrophages (CD68+ cell counts) and B symptoms. *Haematologica* 2015;**100**(Suppl. 3):7.

Hutchings 2010 *{published data only}*

Hutchings M, Kostakoglu L, Loft A, Coleman M, Specht L. Correlation of FDG-PET results after one cycle and after two cycles of chemotherapy in Hodgkin lymphoma. *Journal of Clinical Oncology* 2010;**28**(15):8061.

Knight-Greenfield 2013 *{published data only}*

Knight-Greenfield A, Marshall RA, Hutchings M, Doucette J, Stern J, Coleman M, et al. Interim FDG PET/CT to predict progression-free survival (PFS) better than clinical and baseline metabolic measurements in Hodgkin lymphoma (cHL). *Journal of Clinical Oncology* 2013;**31**(15):8555.

Leontjeva 2016 *{published data only}*

Leontjeva A, Demina E, Ryabukhina J, Tumyan G, Trofimova O, Sotnikov V, et al. Significance of early interim PET results in advanced Hodgkin lymphoma treated intensive program EACOPP-14. *British Journal of Haematology* 2016;**173**(Suppl. 1):107.

Luminari 2010 *{published data only}*

Luminari S, Cesaretti M, Tomasello C, Guida A, Bagni B, Merli F, et al. The use of FDG positron emission tomography (FDG-PET) in patients with Hodgkin lymphoma (HL) in the "real world": A population based study from northern Italy. *Haematologica* 2010;**95**(Suppl. 4):S42-3.

Luminari 2011 *{published data only}*10.3109/10428194.2011.580475

Luminari S, Cesaretti M, Tomasello C, Guida A, Bagni B, Merli F, et al. Use of 2-[18F]fluoro-2-deoxy-D-glucose positron emission tomography in patients with Hodgkin lymphoma in daily practice: a population-based study from Northern Italy. *Leukemia and Lymphoma* 2011;**52**(9):1689-96.

Medvedovskaya 2016 *{published data only}*

Medvedovskaya E, Leontjeva A, Demina E, Ryabukhina J, Tumyan G, Kokosadze N, et al. The impact of outcome of

interim PET/CT on advanced Hodgkin lymphoma treated with EACOPP-14. *Haematologica* 2016;**101**(Suppl. 5):26.

Molnar 2011 *{published data only}*

Molnar Z, Deak B, Kajary K, Lengyel Z, Molnar P, Rosta A, et al. The value of interim 18F-FDG PET/CT in Hodgkin lymphoma. *European Journal of Nuclear Medicine and Molecular Imaging* 2011;**38**(Suppl. 2):S165.

Molnar 2011a *{published data only}*

Molnar Z, Deak B, Kajary K, Lengyel Z, Molnar P, Rosta A, et al. Interim FDG PET/CT examinations in advanced stage Hodgkin lymphoma. *Nuclear Medicine Review* 2011;**14**(Suppl. A):A4.

Moreira 2013 *{published data only}*

Moreira C, Fidalgo P, Queiroz L, Domingues N, Moreira I, Oliveira I, et al. Prognostic value of interim vs. end-of-treatment PET scan in Hodgkin's lymphoma. *Hematological Oncology* 2013;**31**(Suppl. 1):257.

Perrone 2009 *{published data only}*

Perrone T, Gaudio F, Giordano A, De Risi C, Spina A, Curci P, et al. Role of positron emission tomography (PET) after 2 and 4 courses of chemotherapy in patients with Hodgkin's lymphoma: A single center experience. *Haematologica* 2009;**94**(Suppl. 4):216.

Pophali 2014 *{published data only}*

Pophali PA, Rybicki LA, Fenner KB, Jagadeesh D, Dean RM, Pohlman B, et al. Bulky disease does not adversely affect overall survival in early stage Hodgkin lymphoma: Role of interim PET and possible omission of radiotherapy in select patients. *Blood* 2014;**124**:21.

Rusconi 2010 *{published data only}*

Rusconi C, Ravelli E, Gabutti C, Zilioli V, Meli E, Nichelatti M, et al. Baseline and dynamic prognostic factors in newly diagnosed classical Hodgkin's lymphoma. *Haematologica* 2010;**95**(Suppl. 4):S43.

Spallino 2017 *{published data only}*

Spallino M, Guerra L, Cuzzocrea M, Elisei F, Crivellaro C, De Ponti E, et al. The Deauville criteria and QPET as prognostic factors in interim PET in adult Hodgkin lymphoma: A single centre experience. *Clinical and Translational Imaging* 2017;**5**(Suppl. 1):S43.

Yaghmour 2012 *{published data only}*

Yaghmour G, Farhat M, Valdivieso BS, Janakiraman N. PET-negative at 2, 3 or 4 cycles of ABVD in Hodgkin's lymphoma is still good. *Journal of General Internal Medicine* 2012;**27**(Suppl. 2):S258-9.

Zanoni 2011 *{published data only}*

Zanoni L, Agostinelli C, Gallamini A, Rigacci L, Sista M, Piccaluga P, et al. The predictive value of interim PET and immunohistochemical markers in Hodgkin lymphoma (HL). *European Journal of Nuclear Medicine and Molecular Imaging* 2011;**38**(Suppl. 2):S165.

References to ongoing studies

NCT00736320 *{unpublished data only}*

Kobe C, Dietlein M, Kuhnert G, Holstein A, Kahraman D, Haverkamp H, et al. Recruitment and PET interpretation in the HD16 trial for early stage Hodgkin lymphoma. Treatment stratification by FDG-PET. In: NuklearMedizin. Vol. 50. Jahrestagung der Deutschen Gesellschaft für Nuklearmedizin. Bremen, Germany. 2012:A39-40.

NCT00736320. HD16 for early stages - treatment optimization trial in the first-line treatment of early stage Hodgkin lymphoma; treatment stratification by means of FDG-PET. clinicaltrials.gov/ct2/show/nct00736320 (first received 15 August 2008).

Additional references

Adams 2015a

Adams HJ, Nievelstein RA, Kwee TC. Prognostic value of interim FDG-PET in Hodgkin lymphoma: systematic review and meta-analysis. *British Journal of Haematology* 2015;**170**(3):356-66.

Adams 2016a

Adams HJ, Kwee TC. Controversies on the prognostic value of interim FDG-PET in advanced-stage Hodgkin lymphoma. *European Journal of Haematology* 2016;**97**(6):491-8.

Altman 1999

Altman DG. Practical Statistics for Medical Research. 1 edition. London: Chapman & Hall, 1991.

Altman 2012

Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. *PLOS Medicine* 2012;**9**(5):e1001216. [PMID: 22675273]

Amitai 2018

Amitai I, Gurion R, Vidal L, Dann EJ, Raanani P, Gafter-Gvili A. PET-adapted therapy for advanced Hodgkin lymphoma – systematic review. *Acta Oncologica* 2018;**57**(6):765-72.

Barrington 2014

Barrington SF, Mikhaeel NG, Kostakoglu L, Meignan M, Hutchings M, Mueller SP, et al. Role of imaging in the staging and response assessment of lymphoma: consensus of the International Conference on Malignant Lymphomas Imaging Working Group. *Journal of Clinical Oncology* 2014;**32**(27):3048-58.

Barrington 2017a

Barrington SF, Kluge R. FDG PET for therapy monitoring in Hodgkin and non-Hodgkin lymphomas. *European Journal of Nuclear Medicine and Molecular Imaging* 2017;**44**(1):97-110.

Berriolo-Riedinger 2018

Berriolo-Riedinger A, Becker S, Casasnovas O, Vander Borght T, Edeline V. Role of FDG PET-CT in the treatment management of Hodgkin lymphoma. *Cancer Radiotherapy* 2018;**22**(5):393-400.

Blank 2017

Blank O, von Tresckow B, Monsef I, Specht L, Engert A, Skoetz N. Chemotherapy alone versus chemotherapy plus radiotherapy for adults with early stage Hodgkin lymphoma. *Cochrane Database of Systematic Reviews* 2017, Issue 4. Art. No: CD007110. [DOI: 10.1002/14651858.CD007110.pub3]

Boellaard 2010

Boellaard R, O'Doherty MJ, Weber WA, Mottaghy FM, Lonsdale MN, Stroobants SG, et al. FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. *European Journal of Nuclear Medicine and Molecular Imaging* 2010;**37**(1):181-200. [PMID: 19915839]

Borchmann 2011

Borchmann P, Haverkamp H, Diehl V, Cerny T, Markova J, Ho AD, et al. Eight cycles of escalated-dose BEACOPP compared with four cycles of escalated-dose BEACOPP followed by four cycles of baseline-dose BEACOPP with or without radiotherapy in patients with advanced-stage Hodgkin's lymphoma: final analysis of the HD12 trial of the German Hodgkin Study Group. *Journal of Clinical Oncology* 2011;**29**(32):4234-42. [PMID: 21990399]

Bouwmeester 2012

Bouwmeester W, Zuithoff NP, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. Reporting and methods in clinical prediction research: a systematic review. *PLOS Medicine* 2012;**9**(5):1-12.

Bröckelmann 2018

Bröckelmann PJ, Eichenauer DA, Jakob T, Follmann M, Engert A, Skoetz N. Clinical practice guideline: Hodgkin lymphoma in adults—diagnosis, treatment, and follow-up. *Deutsches Ärzteblatt International* 2018;**115**:535-40.

Canellos 1992

Canellos GP, Anderson JR, Propert KJ, Nissen N, Cooper MR, Henderson ES, et al. Chemotherapy of advanced Hodgkin's disease with MOPP, ABVD, or MOPP alternating with ABVD. *New England Journal of Medicine* 1992;**327**(21):1478-84. [PMID: 1383821]

Cheson 2014

Cheson BD, Fisher RI, Barrington SF, Cavalli F, Schwartz LH, Zucca E, et al. Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification. *Journal of Clinical Oncology* 2014;**32**(27):3059-68. [PMID: 25113753]

Covidence [Computer program]

Veritas Health Innovation Covidence. Version accessed 4 March 2019. Melbourne, Australia.: Veritas Health Innovation. Available at www.covidence.org.

Cuccaro 2014

Cuccaro A, Bartolomei F, Cupelli E, Galli E, Giachelia M, Hohaus S. Prognostic factors in hodgkin lymphoma. *Mediterranean Journal of Hematology and Infectious Diseases* 2014;**6**(1):e2014053.

Debray 2017

Debray T, Damen J, Snell K, Ensor J, Hooft L, Reitsma J, et al. A guide to systematic review and meta-analysis of prediction model performance. *BMJ* 2017;**356**:i6460.

Debray 2018

Debray T, Moons K, Riley R. Detecting small-study effects and funnel plot asymmetry in meta-analysis of survival data: A comparison of new and existing tests. *Research Synthesis Methods* 2018;**9**(1):41-50.

Deeks 2011

Deeks JJ, Higgins JP, Altman DG. Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S (editors), Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Engert 2003

Engert A, Schiller P, Josting A, Herrmann R, Koch P, Sieber M, et al. Involved-field radiotherapy is equally effective and less toxic compared with extended-field radiotherapy after four cycles of chemotherapy in patients with early-stage unfavorable Hodgkin's lymphoma: results of the HD8 trial of the German Hodgkin's Lymphoma Study Group. *Journal of Clinical Oncology* 2003;**21**(19):3601-8.

Engert 2007

Engert A, Franklin J, Eich HT, Brillant C, Sehlen S, Cartoni C, et al. Two cycles of doxorubicin, bleomycin, vinblastine, and dacarbazine plus extended-field radiotherapy is superior to radiotherapy alone in early favorable Hodgkin's lymphoma: final results of the GHSG HD7 trial. *Journal of Clinical Oncology* 2007;**25**(23):3495-502.

Engert 2010

Engert A, Plutschow A, Eich HT, Lohri A, Dorken B, Borchmann P, et al. Reduced treatment intensity in patients with early-stage Hodgkin's lymphoma. *New England Journal of Medicine* 2010;**363**(7):640-52. [PMID: 20818855]

Engert 2012

Engert A, Haverkamp H, Kobe C, Markova J, Renner C, Ho A, et al. Reduced-intensity chemotherapy and PET-guided radiotherapy in patients with advanced stage Hodgkin's lymphoma (HD15 trial): a randomised, open-label, phase 3 non-inferiority trial. *Lancet* 2012;**379**(9828):1791-9. [PMID: 22480758]

Fitzgerald 2019

Fitzgerald TJ. Optimisation of adaptive therapy for advanced Hodgkin lymphoma. *Lancet Oncology* 2019;**20**(2):167-8.

Franklin 2005

Franklin JG, Paus MD, Pluetschow A, Specht L. Chemotherapy, radiotherapy and combined modality for Hodgkin's disease, with emphasis on second cancer risk. *Cochrane Database of Systematic Reviews* 2005, Issue 4. Art. No: CD003187. [DOI: 10.1002/14651858.CD003187.pub2] [PMID: 16235316]

Gallamini 2007

Gallamini A, Hutchings M, Rigacci L, Specht L, Merli F, Hansen M, et al. Early interim 2-[18F]fluoro-2-deoxy-D-glucose positron emission tomography is prognostically superior to international prognostic score in advanced-stage Hodgkin's lymphoma: a report from a joint Italian-Danish study. *Journal of Clinical Oncology* 2007;**25**(24):3746-52. [PMID: 17646666]

Geersing 2012

Geersing GJ, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons KG. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. *PLOS One* 2012;**7**(7):e32844.

Hayden 2013

Hayden JA, van der Windt DA, Cartwright JL, Cote P, Bombardier C. Assessing bias in studies of prognostic factors. *Annals of Internal Medicine* 2013;**158**(4):280-6. [PMID: 23420236]

Higgins 2011

Higgins JP, Deeks JJ (editors). Chapter 7: Selecting studies and collecting data. In: Higgins JP, Green S (editors), Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Higgins 2011b

Higgins JP, Deeks JJ, Altman DG. Chapter 16: Special topics in statistics. In: Higgins JP, Green S, editors(s). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011.

Howlader 2015

Howlader N, Noone AM, Krapcho M, Garshell J, Neyman N, Altekruse SF, et al. SEER Cancer Statistics Review, 1975-2010. National Cancer Institute. Bethesda, MD. http://seer.cancer.gov/csr/1975_2010/ based on November 2012 SEER data submission, posted to the SEER web site, April 2013 (accessed 23 June 2015).

Iorio 2015

Iorio A, Spencer FA, Falavigna M, Alba C, Lang E, Burnand B, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. *BMJ* 2015;**350**:1-8.

Josting 2010

Josting A. Prognostic factors in Hodgkin lymphoma. *Expert Review of Hematology* 2010;**3**(5):583-92.

Juweid 2007

Juweid ME, Stroobants S, Hoekstra OS, Mottaghy FM, Dietlein M, Guermazi A, et al. Use of positron emission tomography for response assessment of lymphoma: consensus of the Imaging Subcommittee of International Harmonization Project in Lymphoma. *Journal of Clinical Oncology* 2007;**25**:571-8.

Klimm 2005

Klimm B, Engert A, Diehl V. First-line treatment of Hodgkin's lymphoma. *Current Hematology Reports* 2005;**4**(1):15-22.

Kobe 2008a

Kobe C, Dietlein M, Franklin J, Markova J, Lohri A, Amthauer H, et al. Positron emission tomography has a high negative predictive value for progression or early relapse for patients with residual disease after first-line chemotherapy in advanced-stage Hodgkin lymphoma. *Blood* 2008;**112**(10):3989-94. [PMID: 18757777]

Kobe 2010

Kobe C, Dietlein M, Fuchs M. Interpretation and validation of interim positron emission tomography in Hodgkin lymphoma. Leukemia and Lymphoma 2010;**51**(3):552-3. [PMID: 20141443]

Kobe 2010a

Kobe C, Dietlein M, Kriz J, Furth C, Fuchs M, Borchmann P, et al. The role of PET in Hodgkin's lymphoma and its impact on radiation oncology. *Expert Review of Anticancer Therapy* 2010;**10**(9):1419-28. [PMID: 20836677]

Kyzas 2005

Kyzas PA, Loizou KT, Ioannidis JP. Selective reporting biases in cancer prognostic factor studies. *Journal of the National Cancer Institute* 2005;**97**(14):1043-55. [PMID: 16030302]

Kyzas 2007

Kyzas PA, Denaxa-Kyza D, Ioannidis JP. Quality of reporting of cancer prognostic marker studies: association with reported prognostic effect. *Journal of the National Cancer Institute* 2007;**99**(3):236-43. [PMID: 17284718]

Kılıçkap 2013

Kılıçkap S, Barışta I, Ulger S, Celik I, Selek U, Yıldız F, et al. Clinical Features and Prognostic Factors of Hodgkin's Lymphoma: A Single Center Experience. *Balkan Medical Journal* 2013;**30**(2):178-85.

Lefebvre 2011

Lefebvre C, Manheimer E, Glanville J. Chapter 6: Searching for studies. In: Higgins JP, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Lister 1989

Lister TA, Crowther D, Sutcliffe SB, Glatstein E, Canellos GP, Young RC, et al. Report of a committee convened to discuss the evaluation and staging of patients with Hodgkin's disease: Cotswolds meeting. *Journal of Clinical Oncology* 1989;**7**(11):1630-6.

Macaskill 2010

Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y. Chapter 10 Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C, editors(s). Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Version 1.0. The Cochrane Collaboration, 2010. Available from: http://srdta.cochrane.org/.

Mallett 2010

Mallett S, Timmer A, Sauerbrei W, Altman DG. Reporting of prognostic studies of tumour markers: a review of published

articles in relation to REMARK guidelines. *British Journal of Cancer* 2010;**102**(1):173-80. [PMID: 19997101]

Markova 2009

Markova J, Kobe C, Skopalova M, Klaskova K, Dedeckova K, Plutschow A, et al. FDG-PET for assessment of early treatment response after four cycles of chemotherapy in patients with advanced-stage Hodgkin's lymphoma has a high negative predictive value. *Annals of Oncology* 2009;**20**(7):1270-4. [PMID: 19228806]

McShane 2005

McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting recommendations for tumour MARKer prognostic studies (REMARK). *British Journal of Cancer* 2005;**93**(4):387-91.

Meignan 2009

Meignan M, Gallamini A, Meignan M, Gallamini A, Haioun C. Report on the First International Workshop on Interim-PET-Scan in Lymphoma. *Leukemia and Lymphoma* 2009;**50**(8):1257-60.

Meignan 2009a

Meignan M, Itti E, Bardet S, Lumbroso J, Edeline V, Olivier P, et al. Development and application of a real-time on-line blinded independent central review of interim PET scans to determine treatment allocation in lymphoma trials. *Journal of Clinical Oncology* 2009;**27**(16):2739-41.

Moher 2009

Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Journal of Clinical Epidemiology* 2009;**62**(10):1006-12.

Moons 2009

Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ* 2009;**338**:b375.

Moons 2014

Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Medicine* 2014;**11**:e1001744.

Moskowitz 2018

Moskowitz CH, Walewski J, Nademanee A, Masszi T, Agura E, Holowiecki J, et al. Five-year PFS from the AETHERA trial of brentuximab vedotin for Hodgkin lymphoma at high risk of progression or relapse. *Blood Journal* 2018;**132**(25):2639.

Parmar 1998

Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Statistics in Medicine* 1998;**17**(24):2815-34.

Peat 2014

Peat G, Riley RD, Croft P, Morley KI, Kyzas PA, Moons KG, et al. Improving the transparency of prognosis research: the role of reporting, data sharing, registration, and protocols. *PLOS Medicine* 2014;**11**(7):e1001671.

Radford 2015

Radford J, Illidge T, Counsell N, Hancock B, Pettengell R, Johnson P, et al. Results of a trial of PET-directed therapy for early-stage Hodgkin's lymphoma. *New England Journal of Medicine* 2015;**372**(1):1598-607.

Rancea 2013

Rancea M, Monsef I, von Tresckow B, Engert A, Skoetz N. High-dose chemotherapy followed by autologous stem cell transplantation for patients with relapsed/refractory Hodgkin lymphoma. *Cochrane Database of Systematic Reviews* 2013, Issue 6. Art. No: CD009411. [DOI: 10.1002/14651858.CD009411.pub2] [PMID: 23784872]

Rancea 2013a

Rancea M, Engert A, von Tresckow B, Halbsguth T, Behringer K, Skoetz N. Hodgkin's lymphoma in adults: diagnosis, treatment and follow-up. *Deutsches Arzteblatt International* 2013;**110**(11):177-83, 183e1-3. [PMID: 23555321]

Re 2005

Re D, Thomas RK, Behringer K, Diehl V. From Hodgkin disease to Hodgkin lymphoma: biologic insights and therapeutic potential. *Blood* 2005;**105**(12):4553-60. [PMID: 15728122]

Review Manager 2014 [Computer program]

Nordic Cochrane Centre, The Cochrane Collaboration Review Manager 5 (RevMan 5). Version 5.3. Copenhagen: Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Riley 2013

Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLOS Medicine* 2013;**10**(2):e1001380. [PMID: 23393429]

Riley 2019

Riley RD, Moons KG, Snell KE, Ensor J, Hooft L, Altman DG, et al. A guide to systematic review and meta-analysis of prognostic factor studies. *BMJ* 2019;**364**:k4597.

Riley 2019b

Riley RD, van der Windt D, Croft P, Moons KGM. Prognosis Research in Healthcare. Concepts, Methods, and Impact. Oxford University Press, 2019.

Sauerbrei 2005

Sauerbrei W. Prognostic factors. Confusion caused by bad quality design, analysis and reporting of many studies. *Advances in Oto-Rhino-Laryngology* 2005;**62**:184-200. [PMID: 15608428]

Sickinger 2015

Sickinger MT, von Tresckow B, Kobe C, Engert A, Borchmann P, Skoetz N. Positron emission tomography-adapted therapy for first-line treatment in individuals with Hodgkin lymphoma. *Cochrane Database of Systematic Reviews* 2015, Issue 1. Art. No: CD010533. [DOI: 10.1002/14651858.CD010533.pub2]

Skoetz 2013

Skoetz N, Trelle S, Rancea M, Haverkamp H, Diehl V, Engert A, et al. Effect of initial treatment strategy on survival of patients with advanced-stage Hodgkin's lymphoma: a systematic review and network meta-analysis. *Lancet. Oncology* 2013;**14**(10):943-52. [PMID: 23948348]

Skoetz 2017a

Skoetz N, Will A, Monsef I, Brillant C, Engert A, von Tresckow B. Comparison of first-line chemotherapy including escalated BEACOPP versus chemotherapy including ABVD for people with early unfavourable or advanced stage Hodgkin lymphoma. *Cochrane Database of Systematic Reviews* 2017, Issue 5. Art. No: CD007941. [DOI: 10.1002/14651858.CD007941.pub3]

Steyerberg 2013

Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLOS Medicine* 2013;**10**(2):e1001381.

Swerdlow 2008

Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H, et al. WHO Classification of Tumours of the Haematopoietic and Lymphoid Tissues. Fourth edition. Vol. **2**. Lyon: WHO Press, 1211 Geneva 27, Switzerland, 2008. [ISBN-13: 9789283224310]

Thomas 2002

Thomas RK, Re D, Zander T, Wolf J, Diehl V. Epidemiology and etiology of Hodgkin's lymphoma. *Annals of Oncology* 2002;**13**:147-52. [PMID: 12401681]

Tierney 2007

Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials* 2007;**8**:16.

Trivella 2006

Trivella M. Systematic Reviews of Prognostic Factor Studies (Section: Estimating the Hazard Ratio) [DPhil]. University of Oxford (Altman DG, thesis advisor) June 10th 2006:219.

von Tresckow 2012

von Tresckow B, Plutschow A, Fuchs M, Klimm B, Markova J, Lohri A, et al. Dose-intensification in early unfavorable Hodgkin's lymphoma: final analysis of the German Hodgkin study group HD14 trial. *Journal of Clinical Oncology* 2012;**30**(9):907-13. [PMID: 22271480]

Weigler-Sagie 2010

Weiler-Sagie M, Bushelev O, Epelbaum R, Dann EJ, Haim N, Avivi I, et al. (18)F-FDG avidity in lymphoma readdressed: a study of 766 patients. *Journal of Nuclear Medicine* 2010;**51**(1):25-30.

References to other published versions of this review

Skoetz 2017

Skoetz N, Collins G, Moons K, Estcourt LJ, Engert A, Kobe C, et al. Interim PET for prognosis in adults with Hodgkin lymphoma: a prognostic factor exemplar review. *Cochrane Database of Systematic Reviews* 2017, Issue 4. Art. No: CD012643. [DOI: 10.1002/14651858.CD012643]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Andre 2017

Study characteristics

Methods	Secondary citation(s) • Raemarkers 2014, Cottereau 2018 Language of publication • English Study design • Prospective, multi-centre, phase III randomised trial Study centre(s) • Various Countries • Belgium, Croatia, Denmark, France, Italy, the Netherlands, Slovakia, Switzerland

Andre 2017 *(Continued)*

	Median follow-up time (range) • 55 months
Participants	Number of included participants • Total: 1925 • Randomised to standard treatment without change in protocol because of interim PET: 954 Inclusion criteria • Previously untreated • Classic supradiaphragmic stage I and II HL • Age 15 to 70 years Exclusion criteria • Previous laparotomy • Concomitant or previous cancer other than basal-cell carcinoma of the skin or in situ carcinoma of the cervix • Concomitant severe illness that would reduce life expectancy • Social circumstances not allowing for proper treatment and follow-up • Positivity for the human immunodeficiency virus (exclusion criteria reported in Fermé 2007[1]) Consent • Yes; written informed consent Recruitment period • November 2006 to June 2011 Age (range, in years) • Favourable, standard treatment group median: 31 (15-49) • Unfavourable, standard treatment group median: 32 (15-70) Ethnic group(s) • Not reported Stages of disease • Early stages (I and II) Comorbidities • Not reported, except for the exclusion criteria Therapy regimen • ABVD and radiotherapy depending on treatment arm, favourable/unfavourable disease, and early PET (ePET) positivity
Prognostic factor(s)	Prognostic factor(s) • Early PET (ePET) Definition of prognostic factor(s) • Not reported

Andre 2017 *(Continued)*

	Timing of prognostic factor measurement • After 2 ABVD cycles Method for measurement (use of specific scale and cut-off) • International Harmonization Project criteria. According to these criteria: PET-negative corresponds to Deauville score 1 (no uptake) and score 2 (uptake ≤ mediastinum) • Central review performed online (up to 6 experts, and one local expert) Was the same definition and method for measurement used in all participants? • Central review started later for 2 centres in Italy due to technical difficulties, only 75% of ePET were centrally reviewed Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • Not reported
Outcome(s)	Primary outcome(s) and definition(s) • Progression-free survival (PFS), defined as time from random assignment to date of progression (as experiencing relapse after previous complete remission, progressive disease, or death from any cause) Secondary outcome(s) and definition(s) • Overall survival (OS), not defined Timing of outcome measurement • At 5 years Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • Not applicable
Analysis	**Univariable analysis: Yes** Total number of participants included inunivariateanalysis for each outcome • PFS: all • OS: all Statistical method • Kaplan-Meier method • HR (95% CI) • Randomised arms were compared using the log-rank test stratified by Ann Arbor stage and availability of a baseline FDG-PET scan How was the prognostic factor treated? • Binary

Andre 2017 *(Continued)*

	Multivariable analysis: No
Risk of bias (QUIPS)	Study participation • Low risk • Clear description of participants and study characteristics. Study attrition • Low risk • Length of follow-up reported. Exclusion of participants due to safety amendment during the study. Prognostic factor measurement • Moderate risk • Adequate measurement and description. Central review only for 75% of scans and delayed in the case of 2 centres due to technical difficulties. **Outcome: Overall survival** *Not reported* **Outcome: Progression-free survival** Outcome measurement • Low risk • No definition of outcome. Outcome measured the same way for all participants. 'Other prognostic factors (covariates)' • Low risk Statistical analysis and reporting • Low risk • Statistical method appropriate for the data. **Outcome: Adverse events** *Not reported*
Notes	Conflict of interest • Casasnovas O: honoraria received from Genentech, Takeda, Gilead Sciences, Sanofi; consulting or advisory role at Genentech, Takeda, Gilead Sciences; research funding received from Genentech; travel, accomodation, expenses received from Genentech, Takeda, Gilead Sciences • Brice P: research funding received from Merck Sharp & Dohme Oncology, Takeda; travel, accomodation, expenses received from Takeda • Specht L: consulting or advisory role at Takeda; research funding received from Varian Medical Systems; travel, accomodation, expenses received from Takeda • Delarue R: honoraria received from Servier, Gilead Sciences, Roche, Celgene, Takeda; consulting or advisory role at Gilead Sciences, Roche; Speakers' Bureau at Karyopharm Therapeutics; travel, accomodation, expenses received from Roche, Takeda, Celgene • Hutchings M: consulting or advisory role at Takeda, Genentech, Celgene, Bayer; research funding received from Takeda, Janssen-Cilag, Genentech, Celgene; travel, accomodation, expenses received from Takeda, Bristol-Myers, Squibb, Janssen-Cilag Funding • Supported by European Organisation for Research and Treatment of Cancer (Belgium), LYmphoma Study Association (France), Fondazione Italiana Limfomi (Italy), Fondation Belge Contre le Cancer (Belgium), Dutch Cancer Society (the Netherlands), Institut National du Cancer (France), Assistance

Andre 2017 *(Continued)*

Publique des Hopitaux de Paris (France), Societe Française de Medecine Nucleaire et Imagerie Moleculaire (France), Associazone Angela Serra (Italy), van Vlissingen Lymfoom Fonds (the Netherlands), and Chugai Pharmaceutical (Japan).

[1] Fermé C, Eghbali H, Meerwaldt JH, Rieux C, Bosq J, Berger F, et al. Chemotherapy plus involved-field radiation in early-stage Hodgkin's disease. New England Journal of Medicine 2007;357:1916-1927

Annunziata 2016

Study characteristics

Methods	Secondary citation(s) • NA Language of publication • English Study design • Retrospective, single-centre study Study centre(s) • Not reported Country • Italy Median follow-up time (range) • Not reported
Participants	Number of included participants • 68 Inclusion criteria • HL diagnosis Exclusion criteria • Not reported Consent • Not reported Recruitment period • January 2007 to December 2014 Age (range, in years) • 39 (16-72) Ethnic group(s) • Not reported

Annunziata 2016 *(Continued)*

	Stages of disease • All stages Comorbidities • Not reported Therapy regimen • ABVD according to the presence of risk factors defined by the European Organisation for Research and Treatment of Cancer (EORTC) • Favourable group (age < 50 years with ≤ 3 involved nodal areas, absence of mediastinal bulk (mediastinum-to-thorax ratio < 0.35), and erythrocyte sedimentation rate (ESR) < 50 mm without B symptoms or ESR < 30 mm with B symptoms): 3 cycles ABVD followed by radiotherapy, or 4 cycles ABVD without radiotherapy • Unfavourable group (age ≥ 50 years, > 4 involved nodal areas, presence of mediastinal bulk (mediastinum to-thorax ratio ≥ 0.35), or ESR ≥ 50 mm without B symptoms or ESR ≥ 30 mm with B symptoms): 4 cycles ABVD followed by radiotherapy, or 6 cycles ABVD without radiotherapy (therapy regimen reported in Raemaekers 2014[1])
Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Half-body PET scan (base of the skull to mid-thigh) Timing of prognostic factor measurement • Around day 25 (mean, range 22-27) after cycle 1 of ABVD Method for measurement (use of specific scale and cut-off) • Deauville 5-point scoring system • Scores of 1-3 considered negative, scores of 4-5 considered positive • 2 nuclear medicine physicians interpreted all scans Was the same definition and method for measurement used in all participants? • Yes Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • Not reported
Outcome(s)	Primary outcome(s) and definition(s) • Progression-free survival (PFS), with progression during treatment, lack of complete remission at the end of first-line treatment, and relapse counted as adverse events (AE) Secondary outcome(s) and definition(s) • None Timing of outcome measurement • At 2 years Was the same definition and method for measurement used in all participants? • Yes

Annunziata 2016 *(Continued)*

	Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • Not applicable
Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • PFS: all Statistical method • Receiver operating characteristic (ROC) approach • Kaplan-Meier (survival analysis) • Log-rank (differences between groups) • Cox proportional hazards model How was the prognostic factor treated? • Binary **Multivariable analysis: No**
Risk of bias (QUIPS)	Study participation • Unclear • Clear description of participants and study characteristics. No inclusion and exclusion criteria provided. Study attrition • Unclear risk • No loss to follow-up reported. No length of follow-up reported. Prognostic factor measurement • Low risk • Adequate measurement and description. Prognostic factor measured the same way for all participants. **Outcome: Overall survival** *Not reported* **Outcome: Progression-free survival** Outcome measurement • Low risk • Clear definition. Outcome measured the same way for all participants. 'Other prognostic factors (covariates)' • High risk • Stated in methods section that multiple factors were taken into account for analysis, but unclear which variables and how adjustment was conducted. Disease stage not accounted for.

Annunziata 2016 *(Continued)*

	Statistical analysis and reporting • High risk • Poorly reported. Unclear whether multivariable analysis was reported. **Outcome: Adverse events** *Not reported*
Notes	Conflict of interest • The authors declare that they have no conflict of interest. Funding • Not reported [1] Raemaekers JM, André MP, Federico M, Girinsky T, Oumedaly R, Brusamolino E, et al. Omitting radiotherapy in early positron emission tomography–negative stage I/II Hodgkin lymphoma is associated with an increased risk of early relapse: clinical results of the preplanned interim analysis of the randomised EORTC/LYSA/FIL H10 trial. Journal of Clinical Oncology 2014;32(12):1188-1194

Barnes 2011

Study characteristics	
Methods	Secondary citation(s) • Sher 2009 Language of publication • English Study design • Retrospective, multi-centre study (2 centres) Study centre(s) • Massachusetts General Hospital Cancer Center and Dana-Farber Cancer Institute, Massachusetts, USA Country • USA Median follow-up time (range) • 46 months
Participants	Number of included participants • 96 Inclusion criteria • Diagnosed with classic, histology-proven HL • Adults • Limited-stage non-bulky disease (mass < 10 cm) • ABVD chemotherapy • Availability of interim PET and end-of-treatment PET

Barnes 2011 *(Continued)*

Exclusion criteria

- Nodular lymphocyte predominant HL

Consent

- Not reported

Recruitment period

- January 2000 to December 2008

Age (range, in years)

- 34 (18-77)

Ethnic group(s)

- Not reported

Stages of disease

- Early stages (I to IIB)

Comorbidities

- Not reported

Therapy regimen

- 4 or 6 cycles of ABVD with or without IFRT

Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Whole-body PET scan (base of the skull to mid-thighs) Timing of prognostic factor measurement • After 2 to 4 treatment cycles Method for measurement (use of specific scale and cut-off) • 2 nuclear medicine physicians interpreted all scans, final result based on consensus • Grading on a 4-point scale with scores 0 or 1 considered negative and scores 2 to 4 considered positive Was the same definition and method for measurement used in all participants? • Yes Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • Not reported
Outcome(s)	Primary outcome(s) and definition(s) • Overall survival (OS), defined as the time from initial pathological diagnosis to death from any cause • Progression-free survival (PFS), defined as time from diagnosis to progression or death from any cause Secondary outcome(s) and definition(s) • Overall response rate (ORR), defined as number of subjects with either complete response (CR) or partial response (PR)

Barnes 2011 *(Continued)*

	• Primary refractory disease, defined as progressive disease on treatment or relapse within 3 months of completing therapy Timing of outcome measurement • Unclear: 4 years reported in text, 10 years reported in figure Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • Not reported If yes, how were missing data handled? • Not applicable
Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • OS: all • PFS: all Statistical method • Kaplan-Meier (survival analysis) • Log-rank test • Fisher's exact test (CR) How was the prognostic factor treated? • Binary **Multivariable analysis: No**
Risk of bias (QUIPS)	Study participation • Unclear risk • Description of participants provided. Missing interim and end-of-treatment PET was part of the exclusion criteria. No comparison of baseline study sample (n = 155) and participants (n = 96) included. No reasons for missing scans provided. Study attrition • Low risk • No loss to follow-up. Prognostic factor measurement • Moderate risk • Adequate measurement and description. No standardised criteria, but description of scoring system used. Prognostic factor measured the same way for all participants. Blinding not reported. **Outcome: Overall survival** Outcome measurement

Barnes 2011 *(Continued)*

- High risk
- Clear definition. Reporting of timing inconsistent (4 vs. 10 years).

'Other prognostic factors (covariates)'

- Low risk

Statistical analysis and reporting

- High risk
- Statistical method appropriate for the data, but discrepancies between text and graphs detected.

Outcome: Progression-free survival

Outcome measurement

- High risk
- Clear definition. Reporting of timing inconsistent (4 vs. 10 years).

'Other prognostic factors (covariates)'

- Low risk

Statistical analysis and reporting

- High risk
- Statistical method appropriate for the data, but discrepancies between text and graphs detected.

Outcome: Adverse events

Not reported

Notes	Conflict of interest • There are no relevant conflicts of interests to disclose. Funding • Not reported

Casasnovas 2019

Study characteristics	
Methods	Secondary citation(s) • Casasnovas 2018 Language of publication • English Study design • Open-label, randomised phase 3 trial Study centre(s) • Multicentre (90 centres) Countries • Belgium, France

Casasnovas 2019 *(Continued)*

	Median follow-up time (range) • 50.4 months (IQR: 42.9-59.3) for all participants, not reported separately for standard treatment group
Participants	Number of included participants • 823 in total • 413 in standard treatment group Inclusion criteria • Age 16-60 years • Newly diagnosed HL • ECOG performance status score < 3 • Minimum life expectancy of 3 months • Ann Arbor disease stage III, IV, or IIB with a mediastinum-to-thorax ratio of 0.33 or greater or extranodal localisation • No previous treatment for HL • Baseline PET (PET0) with at least one hypermetabolic lesion • Negative HIV, hepatitis C virus, and human T-lymphotropic serology • Normal liver, renal, and haematological functions except for abnormalities related to HL Exclusion criteria • Nodular lymphocyte predominant subtype • Severe cardiopulmonary or metabolic disease Consent • Written, informed consent Recruitment period • 19 May 2011 to 29 April, 2014 Age (range, in years) • 31 (IQR; ranges 23 - 41) Ethnic group(s) • Not reported Stages of disease • II to IV, with B symptoms Comorbidities • Not reported Therapy regimen • Standard treatment group: 4 cycles of $BEACOPP_{escalated}$, irrespective of PET2 result. After PET4: If PET4-negative: 2 further cycles of $BEACOPP_{escalated}$, if PET4-positive: salvage therapy.
Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Whole-body PET scan (groin to head)

Casasnovas 2019 *(Continued)*

Timing of prognostic factor measurement

- 2 to 4 weeks after completion of cycles 2 and 4 of chemotherapy

Method for measurement (use of specific scale and cut-off)

- Deauville criteria, with scores 1 to 3 considered negative, and scores 4 or 5 considered positive; Independent central review by 3 expert reviewers, final decision was based on at least two concordant responses

Was the same definition and method for measurement used in all participants?

- Yes; participants were scanned on the same camera for all PET scans

Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)?

- Presumably yes, but not explicitly mentioned

Outcome(s)

Primary outcome(s) and definition(s)

- Progression-free survival (PFS), defined as the time from randomisation to first progression, relapse, or death from any cause or last follow-up

Secondary outcome(s) and definition(s)

- Safety, not defined
- Overall response, not defined
- Event-free survival, defined as the time from randomisation to the first documented disease progression, relapse, start of a new anti-lymphoma therapy, death from any cause, or last follow-up
- Disease-free survival, defined as the time that complete response was recorded to the date of first documented disease progression, relapse or death related to lymphoma, toxicity from the study treatment (including treatment-related secondary cancer), unknown cause or last follow-up
- Overall survival, defined as the time from randomisation to death from any cause or last follow-up

Timing of outcome measurement

- PFS: at 5 years

Was the same definition and method for measurement used in all participants?

- Yes

Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)?

- Not reported

Missing data

Participants with any missing value?

- Yes; N = 11 stopped treatment before PET2, and further N = 14 stopped treatment before PET4

If yes, how were missing data handled?

- All 413 participants included in ITT analysis, N = 412 included in safety analysis, N = 372 included in per-protocol analysis

Analysis

Univariable analysis: Yes

Total number of participants included in univariable analysis for each outcome

- PFS, OS: N = 413 in ITT analysis, N = 372 in per-protocol analysis

Statistical method

- Kaplan-Meier (survival analysis)

Casasnovas 2019 *(Continued)*

- Log-rank test
- Cox proportional hazard regression models

How was the prognostic factor treated?

- Binary

Multivariable analysis: Yes

Total number of participants included in multivariable analysis for each outcome

- PFS: 759 (all participants that had reviewed PET2 and PET4 scans; not reported separately for standard group after PET2 without treatment modification)
- OS: not reported

Statistical method

- Cox proportional hazards regression model

How was the prognostic factor treated?

- Binary

Number of candidate covariates

- 8

List of all candidate covariates

- PET assessment (PET2 and PET4)
- Sex
- Age
- Eastern Cooperative Oncology Group score
- B symptoms
- Ann Arbor disease stage
- Bulky disease
- International Prognosis Score

Risk of bias (QUIPS)

Study participation

- Low risk
- Adequate description of study population and recruitment. Detailed inclusion criteria.

Study attrition

- Low risk
- Reasons for loss to follow-up provided for most participants with missing data.

Prognostic factor measurement

- Low risk
- Adequate measurement and description. Prognostic factor measured the same way for all participants.

Outcome: Overall survival

Outcome measurement

- Low risk
- Clear definition. Outcome measured the same way for all participants.

'Other prognostic factors (covariates)'

Casasnovas 2019 *(Continued)*

- Low risk
- Only advanced stages included.

Statistical analysis and reporting

- Low risk
- Statistical methods appropriate and analysis fully reported.

Outcome: Progression-free survival

Outcome measurement

- Low risk
- Clear definition. Outcome determined based on investigator assessment.

'Other prognostic factors (covariates)'

- Low risk
- Only advanced stages included. Multivariable analysis conducted.

Statistical analysis and reporting

- Low risk
- Statistical methods appropriate and analysis fully reported.

Outcome: Adverse events

Not reported

Notes	Conflict of interest • R-OC has received grants, personal fees and non-financial support from Gilead, Roche, and Takeda, personal fees and non-financial support from Bristol-Myers Squibb, Celegne, and Merck Sharpe & Dohme, and personal fees from Abbvie. PB has received personal fees from Bristol-Myers Squibb, Merck Sharpe & Dohme, and Takeda, grants from Takeda Millenium, and non-financial support from Roche. AS has received personal fees from Takeda. EN-V has received personal fees from Keocyt and Sanofi. FM has received personal fees from Celegne, Gilead, Janssen, and Roche/Genentech. RD has received personal fees from Bristol-Myers Squibb, Celegne, Gilead, Janssen, Karyopharm, Roche, Sanofi, and Takeda. MM has received personal fees from Roche China. The other authors declare no competing interests. Funding • Programme Hospitalier de Recherche Clinique

Cerci 2010

Study characteristics	
Methods	Secondary citation(s) • NA Language of publication • English Study design • Prospective, single-centre study

Cerci 2010 *(Continued)*

	Study centre(s) • São Paulo University Clinics Hospital, Brazil Country • Brazil Median follow-up time (range) • 36 months (32-40)
Participants	Number of included participants • 104 Inclusion criteria • Newly diagnosed, biopsy-proven, classic HL Exclusion criteria • Pregnancy Consent • Yes; written Recruitment period • August 2005 to December 2007 Age (range, in years) • 28 (13-82) Ethnic group(s) • Not reported Stages of disease • All stages Comorbidities • Not reported Therapy regimen • ABVD 4-6 cycles (stage I and II), 6-8 cycles (stage III), 8 cycles (stage IV) • Radiation therapy (stage I or II with no adverse risk factors and treated with 4 cycles ABVD; participants with bulky disease)
Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Whole-body PET scan Timing of prognostic factor measurement • After 2 cycles of ABVD, as late as possible within the week before start of cycle 3 Method for measurement (use of specific scale and cut-off)

Cerci 2010 *(Continued)*

	• No specific scale indicated • 2 board-certified nuclear medicine physicians interpreted all scans • PET-negative defined as no pathologic 18F-FDG uptake at any site; PET-positive defined as presence of focal 18F-FDG uptake not attributed to physiologic biodistribution Was the same definition and method for measurement used in all participants? • Yes Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • Not reported
Outcome(s)	Primary outcome(s) and definition(s) • 3-year event-free survival (EFS), defined as the time from diagnosis to treatment failure (incomplete response after first-line treatment, progression during therapy, relapse or death) or last follow-up Secondary outcome(s) and definition(s) • 3-year overall survival (OS) Timing of outcome measurement • At 3 years Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • Not applicable
Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • EFS: all • OS: all Statistical method • Log-rank (probability of treatment failure) • Kaplan-Meier (survival curves) How was the prognostic factor treated? • Binary **Multivariable analysis: No**
Risk of bias (QUIPS)	Study participation • Low risk

Cerci 2010 *(Continued)*

- Clear description of participants and study characteristics, consecutive sampling and no participants excluded based on interim-PET availability.

Study attrition

- Low risk
- Loss to follow-up reported.

Prognostic factor measurement

- Low risk
- Adequate measurement and description. Prognostic factor measured the same way for all participants.

Outcome: Overall survival

Outcome measurement

- Low risk
- No definition. Outcome measured the same way for all participants.

'Other prognostic factors (covariates)'

- High risk
- Univariable analysis for multiple prognostic factors showed significance of factor of interest, but no multivariable analysis performed. Disease stage not accounted for.

Statistical analysis and reporting

- High risk
- Statistical method in univariable analysis appropriate for the data, but no figures, only table with prognostic values, sensitivity and specificity. Discrepancies detected between text and graphs.

Outcome: Event-free survival

Outcome measurement

- Low risk
- Clear definition. Outcome measured the same way for all participants.

'Other prognostic factors (covariates)'

- High risk
- Univariable analysis for multiple prognostic factors showed significance of factor of interest, but no multivariable analysis performed. Disease stage not accounted for.

Statistical analysis and reporting

- High risk
- Statistical method in univariable analysis appropriate for the data, but no figures, only table with prognostic values, sensitivity and specificity. Discrepancies detected between text and graphs.

Outcome: Adverse events

Not reported

Notes	Conflict of interest • Not reported Funding • Not reported

Gallamini 2014

Study characteristics	
Methods	Secondary citation(s) • Agostinelli 2016, Biggi 2013, Gallamini 2006, Gallamini 2007 Language of publication • English Study design • Retrospective, international, multi-centre study (17 centres) Study centre(s) • 17 academic institutions worldwide Countries • Various Median follow-up time (range) • 37 months (2-110)
Participants	Number of included participants • 260 Inclusion criteria • HL participants with early stage unfavourable disease (IIA with adverse prognostic factors) or advanced stage disease (IIB – IVB) • Staging with PET-CT at baseline and after 2 courses of ABVD • No change of treatment according to PET2 • Minimum follow-up of 1 year after completion of first treatment Exclusion criteria • Missing CT data, baseline PET, interim PET, CT or PET slices; poor quality PET images; miscellaneous reasons (n=9) Consent • No; due to retrospective study design Recruitment period • January 2002 to December 2009 Age (range, in years) • 37.3 (14-82) Ethnic group(s) • Not reported Stages of disease • Early stage unfavourable (IIA HL with adverse prognostic factors) • Advanced stages (IIB – IVB)

Gallamini 2014 *(Continued)*

	Comorbidities • Not reported Therapy regimen • 4-8 cycles ABVD with or without involved-field radiotherapy or consolidation radiotherapy
Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Not reported Timing of prognostic factor measurement • A median of 12.3 days (range, 7-22) after cycle 2 of ABVD Method for measurement (use of specific scale and cut-off) • Deauville 5-point scoring system • PET negative defined as scores 1-3, PET positive defined as scores 4 or 5 • 6 reviewers interpreted all scans independently Was the same definition and method for measurement used in all participants? • Yes Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • Yes
Outcome(s)	Primary outcome(s) and definition(s) • Disease progression, defined as new disease within 6 months of first-line treatment • Relapse, defined as disease occurring 6 months or longer after achieving complete remission • Progression-free survival (PFS), defined as time from diagnosis to either disease progression or relapse, or to death as a result of any cause, whichever occurred first • Overall survival (OS), defined as the time from diagnosis to death from any cause Secondary outcome(s) and definition(s) • Inter-observer agreement using the 5-PS for PET2 interpretation Timing of outcome measurement • At 3 years Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Yes
Missing data	Participants with any missing value? • No If yes, how were missing data handled?

Gallamini 2014 *(Continued)*

- Not applicable

Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • PFS: 260 Statistical method • Kaplan Meier survival curves with Mantel-Haenszel, log-rank, Wilcoxon and Tarone-Ware tests • Univariable regression analyses How was the prognostic factor treated? • Binary **Multivariable analysis: Yes** Total number of participants included in multivariable analysis for each outcome • PFS: 260 Statistical method • Cox proportional hazards regression model How was the prognostic factor treated? • Binary Number of candidate covariates • 9 List of all candidate covariates • Bulky disease • Lymphocyte • Albumin • White blood cells • IPS (0-2 vs. ≥3) • Continued complete remission (CR) vs. no CR • Lactate dehydrogenase • Bone marrow involvement • PET2
Risk of bias (QUIPS)	Study participation • Low risk • Clear description of participants and study characteristics. Study attrition • Low risk • Length of follow-up reported. Prognostic factor measurement • Low risk • Adequate measurement and description. Blinding not reported. **Outcome: Overall survival**

Gallamini 2014 *(Continued)*

Outcome measurement

- Low risk
- Clear definition. Outcome measured the same way for all participants. Blinding not reported.

'Other prognostic factors (covariates)'

- Low risk
- Only unfavourable and advances stages included.

Statistical analysis and reporting

- Low risk
- Statistical method appropriate for the data.

Outcome: Progression-free survival

Outcome measurement

- Low risk
- Clear definition. Outcome measured the same way for all participants. Blinding not reported.

'Other prognostic factors (covariates)'

- Low risk
- Only unfavourable and advances stages included.

Statistical analysis and reporting

- Low risk
- Statistical method appropriate for the data.

Outcome: Adverse events

Not reported

Notes	Conflict of interest • Not reported Funding • Reporting incomplete • The authors would like to thank: ... Keosys company for providing the Positoscope (R) network to distribute images to reviewers.

Gandikota 2015

Study characteristics	
Methods	Secondary citation(s) • NA Language of publication • English Study design • Retrospective study

Gandikota 2015 *(Continued)*

Study centre(s)

- Not reported

Country/Countries

- Not reported

Median follow-up time (range)

- 46 months (24-126)

Participants	Number of included participants • 78 Inclusion criteria • Biopsy-proven, early-stage (IA to IIB) classic HL of any subtype with or without bulky disease • Age > 18 years • Completion of planned ABVD and radiation therapy • At least 24 months of follow-up or until proven relapse if earlier Exclusion criteria • None Consent • No Recruitment period • January 2000 to December 2012 Age (range, in years) • 43 (median; 22-86) Ethnic group(s) • Not reported Stages of disease • Early stages (IA to IIB) Comorbidities • Not reported Therapy regimen • ABVD (number of cycles based on risk factors and institutional guidelines) followed by involved-field or extended-field radiotherapy
Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • PET-CT scan (from base of the skull to upper thigh) Timing of prognostic factor measurement

Gandikota 2015 *(Continued)*

	• After ABVD cycle 2 to 4 or at the end of chemotherapy Method for measurement (use of specific scale and cut-off) • 5-point scale • PET negative defined as a score ≤ 3 • Staff physicians who were unaware of patient outcomes reviewed all scans Was the same definition and method for measurement used in all participants? • Yes Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • Yes
Outcome(s)	Primary outcome(s) and definition(s) • Outcomes relevant to this review were not explored in the study Secondary outcome(s) and definition(s) • Not applicable Timing of outcome measurement • Not applicable Was the same definition and method for measurement used in all participants? • Not applicable Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not applicable
Missing data	Participants with any missing value? • Yes: one patient without baseline PET due to pregnancy; one patient without detectable disease on the baseline scan (excision of single site disease) If yes, how were missing data handled? • One patient without detectable disease on the baseline scan did not receive follow-up PET since not considered necessary
Analysis	**Univariable analysis: No** **Multivariable analysis: No**
Risk of bias (QUIPS)	*No risk of bias assessment, since outcomes relevant to this review were not explored in the study.*
Notes	Conflict of interest • The authors made no disclosure. Funding • No specific funding was disclosed.

Hutchings 2005

Study characteristics	
Methods	Secondary citation(s) • NA Language of publication • English Study design • Not reported Study centre(s) • Guy's and St. Thomas' Hospital, London Country • UK Median follow-up time (range) • 40.2 months (6-125)
Participants	Number of included participants • 85 Inclusion criteria • Histologically-confirmed HL • Early interim FDG-PET scans Exclusion criteria • None Consent • Not reported Recruitment period • May 1993 to January 2004 Age (range, in years) • 36.7 (15-73) Ethnic group(s) • Not reported Stages of disease • All stages Comorbidities • Not reported Therapy regimen

Hutchings 2005 *(Continued)*

- According to departmental protocols: mainly ABVD, number of cycles not reported; additional radiotherapy depending on stage and site of HL
- Alternative therapy for participants without satisfactory remission during initial chemotherapy

Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Half-body PET scan (mid-brain to upper thigh) Timing of prognostic factor measurement • After 2 or 3 cycles of chemotherapy, within the second week of the interval between cycles or as late as possible before administration of the next cycle Method for measurement (use of specific scale and cut-off) • No specific scale indicated • 2 experienced nuclear medicine physicians interpreted all scans, differences decided by consensus • PET-negative defined as no evidence of disease; PET-positive defined as increased uptake suspicious for malignant disease; Minimal residual uptake (MRU) defined as low-grade uptake not likely to represent malignancy Was the same definition and method for measurement used in all participants? • Yes Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • Not reported
Outcome(s)	Primary outcome(s) and definition(s) • Progression-free survival (PFS), defined as the time from diagnosis to first evidence of progression or relapse, or to disease-related death • Overall survival (OS), defined as the time from diagnosis to death from any cause Secondary outcome(s) and definition(s) • None Timing of outcome measurement • At 2 and 5 years Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • NA
Analysis	**Univariable analysis: Yes**

Hutchings 2005 *(Continued)*

Total number of participants included in univariable analysis for each outcome

- PFS: all
- OS: not reported

Statistical method

- Kaplan-Meier (survival curves)
- Log-rank (differences between groups)
- Proportional hazards Cox regression analysis

How was the prognostic factor treated?

- Binary

Multivariable analysis: Yes

Total number of participants included in multivariable analysis for each outcome

- PFS: all
- OS: none

Statistical method

- Kaplan-Meier (survival curves)
- Log-rank (differences between groups)
- Proportional hazards Cox regression analysis

How was the prognostic factor treated?

- Binary

Number of candidate covariates

- 4

List of all candidate covariates

- Early interim PET
- Ann Arbor stage
- PET-MRU vs. PET-negative
- PET-positive vs. PET-negative

Risk of bias (QUIPS)

Study participation

- Unclear risk
- All eligible participants included. Clear description of participants and study characteristics. No inclusion / exclusion criteria provided. No comparison to baseline population, and no explanation of missing scans provided.

Study attrition

- Moderate risk
- Loss to follow-up (8 participants), but reasons not provided.

Prognostic factor measurement

- Low risk
- Adequate description. PET results separated into negative, positive and low MRU, which sometimes was considered negative (clearly stated in these cases). No clear cut-off in numbers.

Outcome: Overall survival

Hutchings 2005 *(Continued)*

Outcome measurement

- Low risk
- Clear definition. Outcome measured the same way for all participants. Participants lost to follow-up were still included in analysis.

'Other prognostic factors (covariates)'

- High risk
- Disease stage not accounted for.

Statistical analysis and reporting

- Low risk
- Statistical method in univariable analysis appropriate for the data.

Outcome: Progression-free survival

Outcome measurement

- Low risk
- Clear definition. Outcome measured the same way for all participants. Participants lost to follow-up were still included in analysis.

'Other prognostic factors (covariates)'

- Low risk
- Adjusted for disease stage.

Statistical analysis and reporting

- Low risk
- Statistical method appropriate for the data.

Outcome: Adverse events

Not reported

Notes	Conflict of interest • Not reported Funding • Not reported

Hutchings 2006

Study characteristics	
Methods	Secondary citation(s) • NA Language of publication • English Study design • Prospective, multi-centre study (4 centres)

Hutchings 2006 *(Continued)*

	Study centre(s) • Copenhagen University Hospital, Rigshospitalet, Herlev Hospital, Aarhus University Hospital Country • Denmark Median follow-up time (range) • 22.8 months (6.1-40.8)
Participants	Number of included participants • Total: 99 • With Interim-PET: 77 Inclusion criteria • Newly diagnosed HL • Adults (≥18 years of age) Exclusion criteria • Diabetes mellitus • Pregnancy Consent • Yes; written Recruitment period • November 2001 to June 2004 Age (range, in years) • 36.2 (18.6 – 74.0) Ethnic group(s) • Not reported Stages of disease • All stages Comorbidities • Not reported Therapy regimen • Various therapy regimens: ABVD (91%), ABV/MOPP (3%), ABVD/COPP (3%), BEACOPPesc. (3%), PVAG (1%)
Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Half-body PET scan (mid-brain to upper thigh) Timing of prognostic factor measurement

Hutchings 2006 *(Continued)*

- Within the last week before start of cycle 3 (PET2) and before cycle 5 (PET4)

Method for measurement (use of specific scale and cut-off)

- No specific scale indicated
- 2 experienced nuclear medicine physicians interpreted all scans, differences in interpretation decided by consensus
- Definitions for PET-positive and PET–negative not reported

Was the same definition and method for measurement used in all participants?

- Yes

Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)?

- Yes; nuclear medicine physicians were blinded from all clinical information except diagnosis

Outcome(s)	Primary outcome(s) and definition(s) • Progression-free survival (PFS), defined as the time from diagnosis to first evidence of progression or relapse, or to disease-related death • Overall survival (OS), defined as the time from diagnosis to death from any cause Secondary outcome(s) and definition(s) • None Timing of outcome measurement • At 2 years Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Yes; clinicians were blinded from the results of PET
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • Not applicable
Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • PFS: all • OS: not reported Statistical method • Kaplan-Meier (survival curves) • Log-rank (differences between groups) • Proportional hazards Cox regression analysis How was the prognostic factor treated? • Binary **Multivariable analysis: Yes**

Hutchings 2006 *(Continued)*

Total number of participants included in multivariable analysis for each outcome

- PFS: all
- OS: not reported

Statistical method

- Kaplan-Meier (survival curves)
- Log-rank (differences between groups)
- Proportional hazards Cox regression analysis

How was the prognostic factor treated?

- Binary

Number of candidate covariates

- 3

List of all candidate covariates

- Interim PET
- Clinical stage
- Extranodal disease

Risk of bias (QUIPS)	Study participation • High risk • Significant number of participants without PET (n = 22 out of total n = 99). Imbalance between groups with or without PET scan regarding stage of disease. Study attrition • Low risk • Lack of compliance in a small number of participants (n = 7 out of n = 99), but not in the subjects included in PET2 analysis. Prognostic factor measurement • Low risk • Adequate measurement and description. Significant number of participants without PET (n = 22 out of n = 99). **Outcome: Overall survival** Outcome measurement • Low risk • Clear definition of outcome. Outcome measured the same way for all participants. 'Other prognostic factors (covariates)' • High risk • Disease stage not accounted for. Statistical analysis and reporting • Low risk • Statistical method appropriate for the data. **Outcome: Progression-free survival** Outcome measurement

Hutchings 2006 *(Continued)*

- Low risk
- Clear definition of outcome. Outcome measured the same way for all participants.

'Other prognostic factors (covariates)'

- Low risk
- Adjusted for disease stage.

Statistical analysis and reporting

- Low risk
- Statistical method appropriate for the data.

Outcome: Adverse events

Not reported

Notes	Conflict of interest • The authors have no financial interests in products studied in this work. Funding • Not reported

Hutchings 2014

Study characteristics	
Methods	Secondary citation(s) • NA Language of publication • English Study design • Prospective, multi-centre study Study centre(s) • Not reported Countries • USA, Italy, Poland, Denmark Median follow-up time (range) • 29 months
Participants	Number of included participants • 126* *Potential overlap of Danish participants with those included in Hutchings 2006 Inclusion criteria • Newly diagnosed classic HL

Hutchings 2014 *(Continued)*

Exclusion criteria

- None

Consent

- Yes; written

Recruitment period

- Not reported

Age (range, in years)

- 34.1 (median, 16.8-76.7)

Ethnic group(s)

- Not reported

Stages of disease

- All stages

Comorbidities

- Not reported

Therapy regimen

- Early-stage disease: 2-4 cycles ABVD followed by radiotherapy, or 6 cycles ABVD
- Advanced-stage disease: 6-8 cycles ABVD with or without consolidation radiotherapy, with exceptions (5 Danish participants treated with BEACOPPesc)

Prognostic factor(s)

Prognostic factor(s)

- Interim PET

Definition of prognostic factor(s)

- Whole-body PET scan

Timing of prognostic factor measurement

- Within the last 5 days of cycle 1 (PET1) and cycle 2 (PET2) (US and Italian participants had PET2 only if PET1 was positive)

Method for measurement (use of specific scale and cut-off)

- Deauville 5-point scoring system
- Scores of 1-3 considered negative, scores of 4-5 considered positive
- Baseline interpretation by an expert with access to clinical information, second interpretation by an independent expert from another country blinded to clinical information

Was the same definition and method for measurement used in all participants?

- No; not all participants received PET2

Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)?

- Yes; experts in both stages blinded to clinical outcome, baseline experts also blinded to clinical information

---

Outcome(s)

Primary outcome(s) and definition(s)

- Progression-free survival (PFS), not defined

Hutchings 2014 *(Continued)*

- Overall survival (OS), not defined

Secondary outcome(s) and definition(s)

- None

Timing of outcome measurement

- At 2 and 3 years

Was the same definition and method for measurement used in all participants?

- Not reported; unclear due to multi-national study design

Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)?

- Not reported

Missing data	Participants with any missing value? • No If yes, how were missing data handled? • Not applicable
Analysis	**Univariable analysis: Yes** Total number of participants included in univariate analysis for each outcome • PFS: all • OS: all Statistical method • Kaplan-Meier (survival analysis) • Log-rank (differences between groups) How was the prognostic factor treated? • Binary **Multivariable analysis: Yes** Total number of participants included in multivariable analysis for each outcome • PFS: all • OS: none Statistical method • Kaplan-Meier (survival analysis) • Log-rank (differences between groups) How was the prognostic factor treated? • Binary Number of candidate covariates • 3 List of all candidate covariates • Interim PET (positive or negative)

Hutchings 2014 *(Continued)*

- Extranodal involvement
- Disease stage (early or advanced stage)

Risk of bias (QUIPS)	Study participation • Low risk • Description of participants and study characteristics given. No inclusion and exclusion criteria. Consecutive sampling and no exclusion based on interim PET availability. Detailed description of treatment regimen. Study attrition • Low risk • No loss to follow-up. Prognostic factor measurement • Low risk • Adequate measurement and description. Prognostic factor measured the same way for all participants. **Outcome: Overall survival** Outcome measurement • Low risk • Adequate measurement and description. Prognostic factor measured the same way for all participants. 'Other prognostic factors (covariates)' • High risk • Disease stage not accounted for. Statistical analysis and reporting • Low risk • Statistical method in univariable analysis appropriate for the data. **Outcome: Progression-free survival** Outcome measurement • Low risk • No definition of outcome. Outcome measured the same way for all participants. 'Other prognostic factors (covariates)' • High risk • Disease stage not accounted for. Statistical analysis and reporting • Low risk • Statistical method in univariable analysis appropriate for the data. **Outcome: Adverse events** *Not reported*
Notes	Conflict of interest • The author(s) indicated no potential conflicts of interest.

Hutchings 2014 *(Continued)*

Funding

- Not reported

Kobe 2018

Study characteristics

Methods	Secondary citation(s) • Borchmann 2017 Language of publication • English Study design • Open-label, international, randomised phase 3 trial Study centre(s) • 301 hospitals and private practices in five European countries Countries • Germany, Switzerland, Austria, the Netherlands, Czech Republic Median follow-up time (range) • Not reported for entire study population
Participants	Number of included participants • Total: 2101 • Qualified for randomisation: 1945 Inclusion criteria • Histologically proven primary diagnosis of HL • Advanced stages: stage IIB with one or both of the risk factors large mediastinal mass and extranodal lesions, or stage III or IV • No previous treatment for HL • Age 18-60 years at inclusion • Normal organ function, except for HL-related impairments • Negative HIV test • Negative pregnancy test • Life expectancy > 3 months Exclusion criteria • Incomplete diagnosis of the disease stage • Prior or concurrent disease that prevents treatment according to protocol • HL as part of a composite lymphoma • Prior chemotherapy or radiation • Malignant disease within the last 5 years (exceptions: basalioma, carcinoma in situ of the cervix uteri, completely resected melanoma TNMpT1) • Pregnancy, lactation • Eastern Cooperative Oncology Group (ECOG) performance status > 2

Kobe 2018 *(Continued)*

- Long-term ingestion of corticosteroids or antineoplastic drugs
- Patient's lack of accountability, inability to appreciate the nature, meaning and consequences of the trial and to formulate his/her own wishes correspondingly
- Noncompliance: refusal of blood products during treatment, epilepsy, drug dependency, change of residence to abroad, prior cerebral injury or similar circumstances that appear to make protocol treatment or long-term follow-up impossible
- Antiepileptic treatment
- General intolerance of any protocol medication
- Unsafe contraceptive methods
- Relationship of dependence or employer-employee relationship to the sponsor or the investigator
- Commitment to an institution on judicial or official order
- Participation in another interventional trial that could interact with this trial

Consent

- Yes; written, including consent to participate in the trial and to storage of data and tissue samples

Recruitment period

- 14 May, 2008 to 18 July 2014

Age (range, in years)

- Not reported for entire study population (Borchmann 2017)

Ethnic group(s)

- Not reported

Stages of disease

- Advanced stages: stage III-IV, or stage II with B symptoms and one or both risk factors of large mediastinal mass

Comorbidities

- None, due to exclusion criteria

Therapy regimen

- 6 or 8 cycles of eBEACOPP (standard arm)
- 4 cycles of eBEACOPP or 8 cycles of eBEACOPP with rituximab (experimental arm)

Prognostic factor(s)

Prognostic factor(s)

- Interim PET

Definition of prognostic factor(s)

- Not reported

Timing of prognostic factor measurement

- Between day 17 and day 21 of cycle 2 of chemotherapy

Method for measurement (use of specific scale and cut-off)

- Deauville 5-point scoring system
- PET negative defined as scores 1 or 2, PET positive defined as scores 3 to 5
- A multidisciplinary panel of experts centrally interpreted all scans

Was the same definition and method for measurement used in all participants?

Kobe 2018 *(Continued)*

- Yes

Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)?

- No; assessors who were masked to local findings, centrally reviewed PET-2 and CT scans as well as x-rays and clinical information (Borchmann 2017)

Outcome(s)	Primary outcome(s) and definition(s) • Progression-free survival (PFS), defined as the time from completion of staging until progression, relapse, or death from any cause Secondary outcome(s) and definition(s) • Overall survival (OS), defined as time from completion of staging until death from any cause Timing of outcome measurement • At 3 years Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • Participants with progressive disease, denoted by DS5 (Deauville score 5), were taken off protocol • 505 participants treated before the protocol amendment in June 2011 were excluded from survival analysis If yes, how were missing data handled? • Participants with missing data were excluded from analysis
Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • OS: 722 • PFS: 722 Statistical method • Kaplan-Meier (survival analysis) • Cox regression analysis (hazard ratios) How was the prognostic factor treated? • Binary **Multivariable analysis: Yes** Total number of participants included in multivariable analysis for each outcome • OS: 722 • PFS: 722 Statistical method • Kaplan-Meier (survival analysis) • Cox regression analysis (hazard ratios)

Kobe 2018 *(Continued)*

How was the prognostic factor treated?

- Binary

Number of candidate covariates

- 9

List of all candidate covariates

- Clinical stage
- B symptoms
- Large mediastinal mass
- Extra-nodal involvement
- Involvement of 3 or more nodal areas
- Elevated erythrocyte sedimentation rate
- International Prognosis Score
- HL subtype
- PET positivity (DS4 vs. 1-3)

Risk of bias (QUIPS)

Study participation

- Low risk
- Clear description of participants and study characteristics.

Study attrition

- Low risk
- Length of follow-up reported. Exclusion of participants due to safety amendment during the study.

Prognostic factor measurement

- Low risk
- Adequate measurement and description.

Outcome: Overall survival

Outcome measurement

- Low risk
- Clear definition. Outcome measured the same way for all participants.

'Other prognostic factors (covariates)'

- Low risk
- Only advanced stages included.

Statistical analysis and reporting

- Low risk
- Statistical method appropriate for the data.

Outcome: Progression-free survival

Outcome measurement

- Low risk
- Clear definition. Outcome measured the same way for all participants.

'Other prognostic factors (covariates)'

- Low risk

Kobe 2018 *(Continued)*

- Only advanced stages included.

Statistical analysis and reporting

- Low risk
- Statistical method appropriate for the data.

Outcome: Adverse events

Not reported

Notes

Conflict of interest

- We declare no competing interests.

Funding

- The HD18 trial was funded by the Deutsche Krebshilfe (No. 107957 and 110617) and the Swiss State Secretariat for Education, Research and Innovation (SERI), and supported by Roche Pharma AG (No. ML-21683).

Markova 2012

Study characteristics

Methods

Secondary citation(s)

- Markova 2009

Language of publication

- English

Study design

- Retrospective, single-centre study

Study centre(s)

- Prague, institution not reported

Country

- Czech Republic

Median follow-up time (range)

- 52 months

Participants

Number of included participants

- 69

Inclusion criteria

- Newly diagnosed, histologically proven HL
- Clinical stage IIB with large mediastinal mass and/or extranodal disease, stage III or IV
- Age 18-60 years

Exclusion criteria

- Presence of any concurrent disease precluding protocol treatment

Markova 2012 *(Continued)*

- Composite lymphoma
- Previous malignancy
- Previous chemo- or radiotherapy
- Pregnancy or lactation
- Diabetes mellitus and elevated fasting blood sugar level >130 mg/dl (exclusion from PET)

Consent

- Not reported

Recruitment period

- January 2004 to February 2008

Age (range, in years)

- 30.7 (± 8.4)

Ethnic group(s)

- Not reported

Stages of disease

- IIB to IVB

Comorbidities

- None, due to exclusion criteria

Therapy regimen

- Treatment according to the HD15 trial of the German Hodgkin Study Group (GHSG) randomly assigned to either 8 cycles of BEACOPPescalated, 6 cycles of BEACOPPescalated or 8 cycles of time-condensed BEACOPP14baseline
- Local radiotherapy for participants with partial remission with residual mass ≥2.5cm and positive PET scan after chemotherapy

---

Prognostic factor(s)

Prognostic factor(s)

- Interim PET

Definition of prognostic factor(s)

- Not reported

Timing of prognostic factor measurement

- After cycle 4 of chemotherapy, as close as possible to cycle 5

Method for measurement (use of specific scale and cut-off)

- A local nuclear medicine physician interpreted all interim-PET scans
- PET-positive defined as focal or diffuse uptake above background in a location incompatible with normal anatomy or physiology, without a specific standardised uptake cut-off value; PET-negative defined as no uptake, or increased uptake at the site of residual mass with an intensity lower or equal to the mediastinal blood pool

Was the same definition and method for measurement used in all participants?

- Yes

Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)?

Markova 2012 *(Continued)*

	• Not reported
Outcome(s)	Primary outcome(s) and definition(s) • Progression-free survival (PFS), defined as the time from diagnosis to the first evidence of progression or relapse, or death from any cause Secondary outcome(s) and definition(s) • None Timing of outcome measurement • After cycle 4, 6/8 and 3 months after completion of chemotherapy Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • Not applicable
Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • PFS: all Statistical method • Kaplan-Meier (survival analysis) • Log-rank test (comparison between groups) How was the prognostic factor treated? • Binary **Multivariable analysis: No**
Risk of bias (QUIPS)	Study participation • Low risk • All eligible participants included. Clear description of participants and study characteristics. Consecutive sampling. Inclusion and exclusion criteria provided. Study attrition • Low risk • No loss to follow-up. Prognostic factor measurement • Moderate risk • Prognostic factor measured differently: PET4 scans reviewed locally (at the centre) by one physician, whereas PET6/8 assessment included central review.

Markova 2012 *(Continued)*

Outcome: Overall survival

Not reported

Outcome: Progression-free survival

Outcome measurement

- Low risk
- Outcome measured the same way for all participants.

'Other prognostic factors (covariates)'

- Low risk
- Only advanced stages included.

Statistical analysis and reporting

- Low risk
- Statistical method in univariable analysis appropriate for the data.

Outcome: Adverse events

Not reported

Notes	Conflict of interest • Not reported Funding • Not reported

Mesguich 2016

Study characteristics	
Methods	Secondary citation(s) • NA Language of publication • English Study design • Retrospective, multi-centre study (2 centres) Study centre(s) • Haut-Lévêque Hospital and Bergonié Institute, Bordeaux, France Country • France Median follow-up time (range) • 58.9 months
Participants	Number of included participants

Mesguich 2016 *(Continued)*

- 76

Inclusion criteria

- Biopsy-proven, classic HL
- Availability of baseline, interim and end-of-treatment PET-CT

Exclusion criteria

- Treatment with chemotherapy different than ABVD
- Planned treatment modification following int-PET results
- End-PET performance > 6 months after end of treatment

Consent

- No; waived because of retrospective design

Recruitment period

- December 2005 to April 2011

Age (range, in years)

- 37 (median; 14-67)

Ethnic group(s)

- Not reported

Stages of disease

- All stages

Comorbidities

- Not reported

Therapy regimen

- Various therapy regimens: 3, 4, 6 or 8 cycles of ABVD with or without radiotherapy

Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Not reported Timing of prognostic factor measurement • After 2, 3 or 4 treatment cycles Method for measurement (use of specific scale and cut-off) • Deauville 5-point scoring system • Consensual reading of two nuclear medicine physicians • Two cut-offs for interim PET positivity tested and compared: either scores 4 to 5 considered PET positive, or scores 3 to 5 considered PET positive Was the same definition and method for measurement used in all participants? • Yes Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)?

Mesguich 2016 *(Continued)*

	• Yes
Outcome(s)	Primary outcome(s) and definition(s) • Progression-free survival (PFS), defined as the time from diagnosis to either failure of first-line treatment, relapse or death Secondary outcome(s) and definition(s) • None Timing of outcome measurement • At 5 years Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • NA
Analysis	**Univariable analysis: Yes** Total number of participants included in univariate analysis for each outcome • PFS: all Statistical method • Kaplan Meier analysis curve • Log-rank test How was the prognostic factor treated? • Binary **Multivariable analysis: Yes** Total number of participants included in multivariable analysis for each outcome • PFS: all Statistical method • Cox proportional hazard models How was the prognostic factor treated? • Binary Number of candidate covariates • 3 List of all candidate covariates

Mesguich 2016 *(Continued)*

- Interim PET
- Disease stage*
- Bulky disease*

*2 separate models, each adjusted for one of the 2 covariates other than interim PET

Risk of bias (QUIPS)	Study participation • Low risk • Clear description of participants and study characteristics. Study attrition • Low risk • No loss to follow-up. Prognostic factor measurement • Low risk • Adequate measurement and description. Prognostic factor measured the same way for all participants. **Outcome: Overall survival** *Not reported* **Outcome: Progression-free survival** Outcome measurement • Low risk • Clear definition. Outcome measured the same way for all participants. Blinding not reported. 'Other prognostic factors (covariates)' • Low risk • Adjusted for disease stage. Statistical analysis and reporting • Low risk • Statistical method appropriate for the data. **Outcome: Adverse events** *Not reported*
Notes	Conflict of interest • None declared. Funding • No funding was sought or received for this study.

Oki 2014

Study characteristics	
Methods	Secondary citation(s)

Oki 2014 *(Continued)*

- NA

Language of publication

- English

Study design

- Retrospective, single-centre study

Study centre(s)

- MD Anderson Cancer Center, Houston, Texas, USA

Country

- USA

Median follow-up time (range)

- 45 months

Participants	Number of included participants • Total: 325 • 229 participants with PET2 analysed • 96 participants with PET3 excluded post-hoc Inclusion criteria • Classic HL • Treatment with ABVD • Availability of interim PET scan Exclusion criteria • Additional treatment (e.g. with brentuximab vedotin or rituximab) except for radiotherapy Consent • Not reported Recruitment period • January 2001 to May 2011 Age (range, in years) • Group I (early-stage non-bulky): 32 (median, 18-77) • Group II (stage II bulky): 36 (20-60) • Group III (advanced stage IPS ≤ 2): 30 (19-79) • Group IV (advanced stage IPS ≥ 3): 49 (19-84) Ethnic group(s) • Not reported Stages of disease • All stages Comorbidities • Not reported

Oki 2014 *(Continued)*

	Therapy regimen • ABVD with or without radiotherapy
Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Not reported Timing of prognostic factor measurement • After 2 or 3 cycles of ABVD Method for measurement (use of specific scale and cut-off) • Deauville 5-point scoring system • Scores of 1-3 considered negative, scores of 4-5 considered positive • Independent assessment by 3 nuclear medicine physicians Was the same definition and method for measurement used in all participants? • No, 10 participants had only PET without CT scan Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • Yes
Outcome(s)	Primary outcome(s) and definition(s) • Progression-free survival (PFS), defined as the time from diagnosis to disease progression, relapse or death from any cause Secondary outcome(s) and definition(s) • None Timing of outcome measurement • At 3 years Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • No
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • Not applicable
Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • PFS: all Statistical method

Oki 2014 *(Continued)*

- Kaplan-Meier survival curves with log-rank test per subgroup
- Univariable Cox proportional hazard models

How was the prognostic factor treated?

- Binary

Multivariable analysis: No

Risk of bias (QUIPS)

Study participation

- Low risk
- Clear description of participants and study characteristics.

Study attrition

- Low risk
- No loss to follow-up.

Prognostic factor measurement

- Low risk
- Adequate measurement and description. Prognostic factor measured the same way for all participants.

Outcome: Overall survival

Not reported

Outcome: Progression-free survival

Outcome measurement

- Low risk
- Clear definition. Outcome measured the same way for all participants.

'Other prognostic factors (covariates)'

- High risk
- Disease stage not accounted for.

Statistical analysis and reporting

- High risk
- Exclusion of participants with PET3 during analysis due to lack of prognostic value. Stratification according to disease stage resulted in small sample sizes per subgroup.

Outcome: Adverse events

Not reported

Notes

Conflict of interest

- No conflict of interest to disclose for the study.

Funding

- Not reported

Okosun 2012

Study characteristics	
Methods	Secondary citation(s) • NA Language of publication • English Study design • Retrospective, multi-centre study (6 centres) Study centre(s) • 6 centres in London, UK Country • UK Median follow-up time (range) • 27 months
Participants	Number of included participants • 23 Inclusion criteria • Newly diagnosed, histologically confirmed classic HL • Advanced stage • HIV positivity Exclusion criteria • None Consent • Not reported Recruitment period • June 2007 to August 2010 Age (range, in years) • 42 (median, 32-60) Ethnic group(s) • Not reported Stages of disease • Advanced stages: stage III –IV or stage IIB with at least one adverse prognostic factor Comorbidities • HIV positive participants only Therapy regimen

Okosun 2012 *(Continued)*

- Treatment for HL: standard ABVD therapy
- Treatment for HIV: HAART (two NRTIs in combination with either a non-NRTI or a boosted protease inhibitor) antiretroviral therapy; G-CSF per centre protocol; prophylaxis for Pneumocystis jiroveci

Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Half-body PET-CT scan Timing of prognostic factor measurement • After 2-3 cycles of ABVD, within the week before start of the next cycle Method for measurement (use of specific scale and cut-off) • Deauville 5-point scoring system • Scores 1-3 considered negative, scores 4-5 considered positive • Assessed at 3 established PET centres by own nuclear medicine physician and central review by nuclear medicine expert Was the same definition and method for measurement used in all participants? • Yes Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • Not reported
Outcome(s)	Primary outcome(s) and definition(s) • Progression-free survival (PFS), defined as the time from diagnosis to disease progression or relapse or last follow-up Secondary outcome(s) and definition(s) • Overall survival (OS), defined as the time from diagnosis to death from any cause • Complete remission, defined as the disappearance of all disease manifestations at the end of therapy Timing of outcome measurement • At 2 years Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • Not applicable
Analysis	**Univariable analysis: Yes** Total number of participants included in univariate analysis for each outcome

Okosun 2012 *(Continued)*

- PFS: all
- OS: not applicable, since no participants died

Statistical method

- Kaplan-Meier survival curves with log-rank test

How was the prognostic factor treated?

- Binary

Multivariable analysis: No

Risk of bias (QUIPS)	Study participation • Low risk • Clear description of participants and study characteristics. Three participants did not have a staging PET, no reasons for missing PET provided. Study attrition • Low risk • No loss to follow-up. Length of follow-up reported. Participants without interim PET (n = 11) excluded. Prognostic factor measurement • Low risk • Adequate measurement and description. Prognostic factor measured the same way for all participants. **Outcome: Overall survival** *Not reported* **Outcome: Progression-free survival** Outcome measurement • Low risk • Clear definition. Outcome measured the same way for all participants. 'Other prognostic factors (covariates)' • Low risk • Only unfavourable and advanced stages included. Statistical analysis and reporting • High risk • Small sample size for some events (only two participants with positive interim PET result). **Outcome: Adverse events** *Not reported*
Notes	Conflict of interest • All authors have no conflicts of interest or disclaimers to declare. Funding • Not reported

Orlacchio 2012

Study characteristics	
Methods	Secondary citation(s) • NA Language of publication • English Study design • Retrospective, single-centre study Study centre(s) • Policlinico Universitario TorVergata, Rome, Italy Country • Italy Median follow-up time (range) • Not reported
Participants	Number of included participants • 132 Inclusion criteria • HL diagnosis based on biochemical tests and bone marrow biopsy • PET-MDCT staging examination, interim PET-MDCT and end of treatment PET-MDCT performed Exclusion criteria • None Consent • Not reported Recruitment period • January 2005 to June 2010 Age (range, in years) • 34 (mean, 16-74) Ethnic group(s) • Not reported Stages of disease • All stages Comorbidities • Not reported Therapy regimen

Orlacchio 2012 *(Continued)*

- ABVD dose dependent on disease stage: stages I-IIA 4x ABVD with radiotherapy; stages IIB-IV 6-8x ABVD with radiotherapy

Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • PET scan from pelvis to head Timing of prognostic factor measurement • At the end of the second ABVD cycle Method for measurement (use of specific scale and cut-off) • International Harmonization Project guidelines • Rated by a radiologist and nuclear medicine specialist, confirmation by semi-quantitative analysis Was the same definition and method for measurement used in all participants? • Yes Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • Not reported
Outcome(s)	Primary outcome(s) and definition(s) • Complete remission, defined as the disappearance of symptoms and metabolic activity at any nodal or extranodal site with negative bone marrow biopsy • Partial remission, defined as persistence of significant metabolic activity at one site only, with at least 50% reduction in volume of the nodal masses or parenchymal nodular formations and persistence of disease at bone marrow level • Stable disease, defined as unchanged metabolic findings • Disease progression, defined as the appearance of new sites of pathological uptake and as a 50% increase in volume of nodal masses or previously detected parenchymal localisations Secondary outcome(s) and definition(s) • None Timing of outcome measurement • At the end of treatment Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • NA
Analysis	**Univariable analysis: Yes**

Orlacchio 2012 *(Continued)*

	Total number of participants included in univariate analysis for each outcome • Outcomes selected for univariable analysis unclear Statistical method • Sensitivity, specificity, PPV, NPV How was the prognostic factor treated? • Binary **Multivariable analysis: No**
Risk of bias (QUIPS)	*No risk of bias assessment, since outcomes relevant to this review were not explored in this study.*
Notes	Conflict of interest • None Funding • Not reported

Rossi 2014

Study characteristics	
Methods	Secondary citation(s) • NA Language of publication • English Study design • Retrospective, single-centre study Study centre(s) • Hospital of Dijon, France Country • France Median follow-up time (range) • 50 months (22-71)
Participants	Number of included participants • 59 Inclusion criteria • First diagnosis of classic HL Exclusion criteria • Positive serology for HIV

Rossi 2014 *(Continued)*

	Consent • Yes; written informed consent Recruitment period • January 2007 to January 2010 Age (range, in years) • 35.5 (16-76) Ethnic group(s) • Not reported Stages of disease • All stages Comorbidities • Not reported, except for exclusion of HIV positive participants Therapy regimen • Anthracycline-based chemotherapy dependent on disease stage: stages I-II 4-6x chemotherapy with radiotherapy; stages III-IV 8x chemotherapy
Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Whole-body PET-CT scan Timing of prognostic factor measurement • After 2 cycles of chemotherapy Method for measurement (use of specific scale and cut-off) • Deauville 5-point scoring system • Scores 1-3 considered negative, scores 4-5 considered positive • ΔSUVmax (PET0-PET2) dichotomized by applying the ROC approach • Independent review by 2 nuclear medicine physicians Was the same definition and method for measurement used in all participants? • Different scanner used for 4 participants Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • Yes
Outcome(s)	Primary outcome(s) and definition(s) • Progression-free survival (PFS), defined as the time from the beginning of treatment until progression, relapse, or death from any cause or the date of last follow-up • Time to progression (TTP), defined as time from the date of the first course of chemotherapy to any treatment failure, including progression, relapse, or death related to lymphoma, or the date of last follow-up (participants with death from other cause were censored at the time of death) Secondary outcome(s) and definition(s)

Rossi 2014 *(Continued)*

	• None Timing of outcome measurement • At 4 years Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • Not applicable
Analysis	**Univariable analysis: Yes** Total number of participants included in univariate analysis for each outcome • PFS: all Statistical method • Kaplan-Meier product limit method with log-rank test How was the prognostic factor treated? • Binary **Multivariable analysis: Yes** Total number of participants included in multivariable analysis for each outcome • PFS: all Statistical method • Cox proportional hazards regression models per outcome How was the prognostic factor treated? • Binary Number of candidate covariates • 2 List of all candidate covariates • ΔSUVmax (PET0-PET2) • International prognosis score (IPS)
Risk of bias (QUIPS)	Study participation • Low risk • Clear description of participants and study characteristics. Study attrition

Rossi 2014 *(Continued)*

- Low risk
- No loss to follow-up.

Prognostic factor measurement

- Low risk
- Adequate measurement and description. Prognostic factor measured the same way for all participants.

Outcome: Overall survival

Not reported

Outcome: Progression-free survival

Outcome measurement

- Low risk
- Clear definition. Outcome measured the same way for all participants. Blinding not reported.

'Other prognostic factors (covariates)'

- High risk
- Disease stage not accounted for.

Statistical analysis and reporting

- Low risk
- Statistical method appropriate for the data.

Outcome: Adverse events

Not reported

Notes	Conflict of interest • The costs of publication of this article were defrayed in part by the payment of page charges. Therefore, and solely to indicate this fact, this article is hereby marked “advertisement” in accordance with 18 USC section 1734. No potential conflict of interest relevant to this article was reported. Funding • Not reported

Simon 2016

Study characteristics	
Methods	Secondary citation(s) • Miltenyi 2015 Language of publication • English Study design • Retrospective study Study centre(s)

Simon 2016 *(Continued)*

- Not reported

Country

- Hungary

Median follow-up time (range)

- 47.52 months (11-80)

Participants	Number of included participants • 121 Inclusion criteria • Newly diagnosed HL • No previous treatment Exclusion criteria • Immunosuppressive medications • Immunodeficiency Consent • No Recruitment period • 2007 to 2013 Age (range, in years) • 36.7 (mean, 17-79) Ethnic group(s) • Not reported Stages of disease • All stages Comorbidities • None due to exclusion criteria Therapy regimen • ABVD dependent on disease stage: 6 or 8 cycles of ABVD, or 4 or 6 cycles of ABVD combined with radiotherapy
Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Not reported Timing of prognostic factor measurement • After cycle 2 of ABVD between days 11 and 14 Method for measurement (use of specific scale and cut-off)

Simon 2016 *(Continued)*

- Deauville 5-point scoring system
- Scores 1-3 considered negative, scores 4-5 considered positive
- Person(s) interpreting the scans not reported

Was the same definition and method for measurement used in all participants?

- Yes

Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)?

- Not reported

Outcome(s)	Primary outcome(s) and definition(s) • Overall survival (OS), not defined • Progression-free survival (PFS), not defined Secondary outcome(s) and definition(s) • None Timing of outcome measurement • At 5 years after diagnosis Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • NA
Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • OS: all • PFS: all Statistical method • Kaplan-Meier (survival analysis) • Log-rank test (comparison between groups) • Cox proportional hazard model (effect of variants on survival) How was the prognostic factor treated? • Binary **Multivariable analysis: Yes** Total number of participants included in multivariable analysis for each outcome • OS: all • PFS: all

Simon 2016 *(Continued)*

Statistical method

- Kaplan-Meier (survival analysis)
- Log-rank test (comparison between groups)
- Cox proportional hazard model (effect of variants on survival

How was the prognostic factor treated?

- Binary

Number of candidate covariates

- 8

List of all candidate covariates

- Age
- Disease stage
- Gender
- B symptoms
- Bulky disease
- Treatment
- PET2 positivity
- Lymphocyte/monocyte ratio (LMR)

Risk of bias (QUIPS)

Study participation

- Unclear risk
- Description of participants provided, but no in- and exclusion criteria provided. Not clear how many participants were sampled and included from the baseline sample.

Study attrition

- Low risk
- No dropouts.

Prognostic factor measurement

- Low risk
- Adequate measurement and description. Prognostic factor measured the same way for all participants.

Outcome: Overall survival

Outcome measurement

- Low risk
- Outcome measured the same way for all participants.

'Other prognostic factors (covariates)'

- High risk
- Disease stage not accounted for.

Statistical analysis and reporting

- High risk
- Statistical analysis appropriate for the data. All primary outcomes reported, but discrepancies between text and graphs/tables detected.

Outcome: Progression-free survival

Simon 2016 *(Continued)*

Outcome measurement

- Low risk
- Outcome measured the same way for all participants.

'Other prognostic factors (covariates)'

- High risk
- Disease stage not accounted for.

Statistical analysis and reporting

- High risk
- Statistical analysis appropriate for the data and all primary outcomes reported. However, discrepancies between text and graphs/tables detected.

Outcome: Adverse events

Not reported

Notes

Conflict of interest

- None of the authors have any competing interest in the manuscript.

Funding

- Not reported

Straus 2011

Study characteristics

Methods

Secondary citation(s)

- Kostakoglu 2012

Language of publication

- English

Study design

- Prospective phase 2, multi-centre (29 centres), clinical trial

Study centre(s)

- 29 Cancer and Leukemia Group B (CALGB) institutions

Country/Countries

- Not reported

Median follow-up time (range)

- Not reported

Participants

Number of included participants

- Total: 99
- With interim-PET: 88

Inclusion criteria

Straus 2011 *(Continued)*

- Previously untreated, histologically confirmed, classic HL with clinical stages I or II, measurable through physical examination or imaging studies

Exclusion criteria

- Bulky disease

Consent

- Yes; written

Recruitment period

- 15 May 2004 to 29 September 2006

Age (range, in years)

- 37 (18-80)

Ethnic group(s)

- Not reported

Stages of disease

- Stages I - IIB

Comorbidities

- Not reported

Therapy regimen

- 6 cycles of AVG administered on days 1 and 15 per cycle

Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Not reported Timing of prognostic factor measurement • 1 to 2 weeks after completion of cycle 2 of AVG Method for measurement (use of specific scale and cut-off) • Visual assessment was performed using International Harmonization Project criteria • Central review by 2 independent reviewers and an adjudicator Was the same definition and method for measurement used in all participants? • Yes Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • Yes
Outcome(s)	Primary outcome(s) and definition(s) • Complete response, defined as complete remission or complete remission unconfirmed after 6 cycles of chemotherapy Secondary outcome(s) and definition(s)

Straus 2011 *(Continued)*

- Progression-free survival (PFS), measured from study entry until relapse
- Adverse events (AEs), defined as toxicity including grade 3 or greater myelosuppression

Timing of outcome measurement

- At 3 years

Was the same definition and method for measurement used in all participants?

- Yes

Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)?

- Not reported

Missing data	Participants with any missing value? • No If yes, how were missing data handled? • None
Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • Complete response: none • PFS: 88 Statistical method • Kaplan-Meier (survival analysis) • Log-rank test (comparison between groups) How was the prognostic factor treated? • Binary **Multivariable analysis: No**
Risk of bias (QUIPS)	Study participation • Low risk • Clear description of participants and study characteristics. Study attrition • Low risk • Loss to follow-up reported (n = 2). Prognostic factor measurement • Low risk • Adequate measurement and description. PET2 available for n = 88 out of a total of n = 99 participants. **Outcome: Overall survival** *Not reported* **Outcome: Progression-free survival** Outcome measurement • Low risk

Straus 2011 *(Continued)*

- Clear definition. Outcome measured the same way for all participants.

'Other prognostic factors (covariates)'

- Low risk
- Only stages I - IIB included.

Statistical analysis and reporting

- Low risk
- Statistical method in univariable analysis appropriate for the data.

Notes

Conflict of interest

- The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked ‘‘advertisement’’ in accordance with 18 USC section 1734.The authors declare no competing financial interests.

Funding

- This work was supported by the National Cancer Institute: CA77651 (D.J.S., H.S.), CA33601 (J.L.J.), CA32291 (A.S.L., G.P.C.), CA77440 (N.L.B.), CA04457 (L.K.), CA77658 (N.C.H., S.-H.J.), CA32291 (R.W.T.), CA47642 (M.E.J.), and CA77597 (B.D.C.). This work was supported in part by the Lymphoma Foundation, Adam Spector Fund for Hodgkin Research, the Ernest & Jeanette Dicker Charitable Foundation, and Mr Daniel Moon and Family (for D.J.S.). This work was also supported by CALGB (National Cancer Institute) with partial support by Eli Lilly and Company. The research for CALGB 50203 was supported in part by grants from the National Cancer Institute (CA31946) to the CALGB (Dr Monica M. Bertagnolli, Chair) and to the CALGB Statistical Center (Dr Stephen George, CA33601). The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute.

Touati 2014

Study characteristics

Methods

Secondary citation(s)

- NA

Language of publication

- English

Study design

- Retrospective, single-centre study

Study centre(s)

- University Hospital of Limoges, France

Country

- France

Median follow-up time (range)

- 65.8 months (2.2-194.5)

Participants

Number of included participants

Touati 2014 *(Continued)*

- Total: 158
- With interim-PET: 68

Inclusion criteria

- Histologically proven, classic HL

Exclusion criteria

- Nodular lymphocyte predominant HL

Consent

- Not reported

Recruitment period

- February 1995 to July 2011

Age (range, in years)

- 38 (16-85)

Ethnic group(s)

- Not reported

Stages of disease

- All stages

Comorbidities

- Not reported

Therapy regimen

- According to the standard of care at the time of diagnosis therapy regimens included ABVD, MOPP/ABV hybrid or BEACOPP; number of cycles not reported

Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Not reported Timing of prognostic factor measurement • After cycle 2 of chemotherapy Method for measurement (use of specific scale and cut-off) • Visual evaluation • PET-positive if focal or diffuse accumulation of FDG in lesions higher than in surrounding tissue • FDG-PET-CT data (2005 and later) retrospectively reinterpreted using the Deauville 5-point scoring system Was the same definition and method for measurement used in all participants? • Different PET imaging techniques over time (dual-head coincidence until 2005, then FDG-PET-CT), quality assurance and quality control program to ensure comparability of methods Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)?

Touati 2014 *(Continued)*

	• Not reported
Outcome(s)	Primary outcome(s) and definition(s) • Progression-free survival (PFS), defined as time from date of diagnosis until relapse or death • Overall survival (OS), defined as time from first day of diagnosis until death from any cause Secondary outcome(s) and definition(s) • None Timing of outcome measurement • At 5 years Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • Not applicable
Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • PFS: 68 • OS: 68 Statistical method • Kaplan-Meier (survival analysis) • Chi-squared test or t-test (differences between groups) • ANOVA (comparison of means) How was the prognostic factor treated? • Binary **Multivariable analysis: No**
Risk of bias (QUIPS)	Study participation • Unclear risk • Availability of interim PET as part of inclusion criteria, but not clear why less than 50% of participants had interim PET data. No comparison of baseline study sample (n = 357) with included participants (n = 158). Study attrition • Low risk • All participants with available interim PET included. Prognostic factor measurement

Touati 2014 *(Continued)*

- Moderate risk
- Retrospective reinterpretation of PET scans using the Deauville criteria. Method described, but unclear whether assessors were blinded to initial interpretation.

Outcome: Overall survival

Outcome measurement

- Low risk
- Clear definition. Outcome measured the same way for all participants.

'Other prognostic factors (covariates)'

- High risk
- Disease stage not accounted for.

Statistical analysis and reporting

- Low risk
- Statistical method appropriate for the data.

Outcome: Progression-free survival

Outcome measurement

- Low risk
- Clear definition. Outcome measured the same way for all participants.

'Other prognostic factors (covariates)'

- High risk
- Disease stage not accounted for.

Statistical analysis and reporting

- Low risk
- Statistical method appropriate for the data.

Outcome: Adverse events

Not reported

Notes	Conflict of interest • Not reported Funding • This work was supported by the University Hospital of Limoges, CHU Limoges, F-87042 France.

Ying 2014

Study characteristics	
Methods	Secondary citation(s) • NA Language of publication • Chinese, translated to English

Ying 2014 *(Continued)*

	Study design • Retrospective study Study centre(s) • Peking University Cancer Hospital Country • People's Republic of China Median follow-up time (range) • 29.4 months (12.2-52.4)* *For the whole population (n = 50), but only 35 participants underwent interim PET
Participants	Number of included participants • Total: 50 • With interim PET: 35 Inclusion criteria • Newly diagnosed HL according to the 2008 WHO Hematopoietic and Lymphoid Tissue Classification Exclusion criteria • Not reported Consent • Not reported Recruitment period • September 2009 to December 2012 Age (range, in years) • 33 (14-74) Ethnic group(s) • Not reported Stages of disease • All stages Comorbidities • Not reported Therapy regimen • ABVD or BEACOPP with or without radiotherapy
Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • From the top of the head to the middle thigh, the entire lower extremity was scanned if necessary

Ying 2014 *(Continued)*

	Timing of prognostic factor measurement • After 2 to 4 cycles of treatment Method for measurement (use of specific scale and cut-off) • Interpretation of scans by 2 experienced PET-CT physicians • Scale and cut-off not reported Was the same definition and method for measurement used in all participants? • Not reported Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • Not reported
Outcome(s)	Primary outcome(s) and definition(s) • Progression-free survival (PFS), defined as the interval from diagnosis to first signs of tumour progression, patient death, or end of follow-up Secondary outcome(s) and definition(s) • None Timing of outcome measurement • At 3 years Was the same definition and method for measurement used in all participants? • Unclear, follow-up was conducted via telephone and/or outpatient visits Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • Only 35/50 participants underwent interim PET If yes, how were missing data handled? • Not reported
Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • PFS: 35 Statistical method • Kaplan-Meier curves and life tables (survival analysis) • Log-rank tests (comparison between groups) How was the prognostic factor treated? • Binary **Multivariable analysis: No**
Risk of bias (QUIPS)	Study participation • Low risk

Ying 2014 *(Continued)*

- Clear description of participants and study characteristics.

Study attrition

- Low risk
- No loss to follow-up. Length of follow-up reported.

Prognostic factor measurement

- Moderate risk
- Adequate measurement and description, but no standardised criteria for PET scan evaluation.

Outcome: Overall survival

Not reported

Outcome: Progression-free survival

Outcome measurement

- Low risk
- Clear definition. Outcome assessed differently for some participants (via telephone and/or outpatient visits).

'Other prognostic factors (covariates)'

- High risk
- Disease stage not accounted for.

Statistical analysis and reporting

- High risk
- Poor reporting of univariable analysis.

Outcome: Adverse events

Not reported

Notes	Translated from Chinese to English by Yu-Tian Xiao. Conflict of interest • Not reported Funding • This study was funded by Natural Science Foundation of China (NSFC) grant no. 81470328 and Youth Fund of NSFC grant no. 81600162, 81600130.

Zaucha 2017

Study characteristics	
Methods	Secondary citation(s) • NA Language of publication • English Study design

Zaucha 2017 *(Continued)*

- Prospective, observational, multi-centre study (11 centres)

Study centre(s)

- 11 haemato-oncology centres

Country

- Poland

Median follow-up time (range)

- 44.7 months (12.7–90.2)*

*Data for surviving participants only

Participants	Number of included participants • 310 registered participants, out of which 24 were excluded from analysis due to treatment intensification based on PET1 and/or clinical symptoms of active HL Inclusion criteria • Newly diagnosed with classic HL Exclusion criteria • Absent/poor-quality PET-CT images Consent • Yes; written informed consent Recruitment period • January 2008 to October 2014 Age (range, in years) • 30.8 (median, 18–80) Ethnic group(s) • Not reported Stages of disease • All stages Comorbidities • Not reported Therapy regimen • ABVD dependent on disease stage: stages I-IIA 2-4x ABVD with radiotherapy or 6x ABVD; stages IIB-IV 6-8x ABVD with or without radiotherapy
Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Whole-body scan (mandibular angle to one third upper femur) Timing of prognostic factor measurement

Zaucha 2017 *(Continued)*

- 11-13 days after end of ABVD cycle 1 (PET1)
- Additional scan after ABVD cycle 2 for participants with a PET1 score of 3-5 (PET2)

Method for measurement (use of specific scale and cut-off)

- Deauville 5-point scoring system
- Scores 1-3 considered negative, scores 4-5 considered positive
- 6 reviewers interpreted all scans using the blinded independent central review method, disagreements were resolved in a joint session

Was the same definition and method for measurement used in all participants?

- No; PET2 only administered to participants with a PET1 score of 3-5

Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)?

- Yes

Outcome(s)	Primary outcome(s) and definition(s) • Progression-free survival (PFS), not defined Secondary outcome(s) and definition(s) • Kinetics of response Timing of outcome measurement • At 3 years Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Yes
Missing data	Participants with any missing value? • Yes; only 198 participants had PET2 scans If yes, how were missing data handled? • Not reported
Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • 286 (PET1) / 198 (PET2) Statistical method • Kaplan-Meier (survival analysis) • Cox proportional hazard regression analysis (HR between treatment groups) How was the prognostic factor treated? • Binary **Multivariable analysis: No**
Risk of bias (QUIPS)	Study participation

Zaucha 2017 *(Continued)*

- Low risk
- Clear description of participants and study characteristics.

Study attrition

- Low risk

Prognostic factor measurement

- Moderate risk
- Adequate measurement and description. Prognostic factor measured the same way for all participants. However, while PET1 scans were available for all participants, the availability of PET2 scans was dependent on the result of PET1. No further scans were performed if PET1 was negative

Outcome: Overall survival

Not reported as a primary endpoint in the publication. IPD data were available and used to calculate the HR and SE for this outcome.

Outcome: Progression-free survival

Outcome measurement

- High risk
- No definition of outcome.

'Other prognostic factors (covariates)'

- High risk
- Disease stage not accounted for.

Statistical analysis and reporting

- High risk
- No detailed description of analysis.

Outcome: Adverse events

Not reported

Notes

Conflict of interest

- The authors have declared no conflicts of interest.

Funding

- No funders to report.

Zinzani 2012

Study characteristics

Methods

Secondary citation(s)

- Zinzani 2006

Language of publication

- English

Study design

Zinzani 2012 *(Continued)*

- Retrospective, multi-centre study (2 centres)

Study centre(s)

- Bologna and Florence, Italy

Country

- Italy

Median follow-up time (range)

- 45 months (6-100)

Participants	Number of included participants • 304 Inclusion criteria • Diagnosed with HL Exclusion criteria • Other treatment regimens than ABVD • Secondary lymphomas • Continuation of therapy during data analysis Consent • Yes; written informed consent Recruitment period • June 1997 to June 2009 Age (range, in years) • 32 (13-78) Ethnic group(s) • Not reported Stages of disease • All stages Comorbidities • Assessed, but not reported Therapy regimen • ABVD dependent on disease stage: early stages 6x ABVD or 4x ABVD with radiotherapy; advanced stages 6x ABVD
Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Not reported Timing of prognostic factor measurement

Zinzani 2012 *(Continued)*

	• After cycle 2 of ABVD Method for measurement (use of specific scale and cut-off) • Juweid criteria • PET positive considered if focal FDG uptake that could not be attributed to physiological biodistribution, benign uptake or normal anatomy, with clearly increased activity relative to the background, excluding participants with minimal residual uptake • 2 experienced board-certified nuclear medicine physicians interpreted all scans Was the same definition and method for measurement used in all participants? • Yes Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • No
Outcome(s)	Primary outcome(s) and definition(s) • Response at the end of first-line treatment and at follow-up Secondary outcome(s) and definition(s) • Progression-free survival (PFS), defined as time from diagnosis to first observation of progressive disease or death from any cause • Overall survival (OS), defined as time from diagnosis to time of most recent visit or death Timing of outcome measurement • At 9 years (for PFS and OS) Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • No
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • Not applicable
Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • PFS: all • OS: all Statistical method • Kaplan-Meier (survival analysis) • Log-rank test (comparison between groups) How was the prognostic factor treated? • Binary

Zinzani 2012 *(Continued)*

	Multivariable analysis: No
Risk of bias (QUIPS)	Study participation • Low risk • Clear description of participants and study characteristics. Study attrition • Low risk • No loss to follow-up. Prognostic factor measurement • Low risk • Adequate measurement and description. Prognostic factor measured the same way for all participants. No blinding of assessors. **Outcome: Overall survival** Outcome measurement • Low risk • Clear definition of outcome. Outcome measured the same way for all participants. 'Other prognostic factors (covariates)' • High risk • Disease stage not accounted for. Statistical analysis and reporting • Low risk • Statistical method appropriate for the data. **Outcome: Progression-free survival** Outcome measurement • Low risk • Clear definition of outcome. Outcome measured the same way for all participants. 'Other prognostic factors (covariates)' • High risk • Disease stage not accounted for. Statistical analysis and reporting • Low risk • Statistical method appropriate for the data. **Outcome: Adverse events** *Not reported*
Notes	Conflict of interest • None Funding • This work was partially supported by BolognAIL (Bologna, Italy).

ABVD: adriamycin/doxorubicin, bleomycin, vinblastine and dacarbazine; **BEACOPP:** bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine and prednisone; **ePET:** early positron emission tomography; **FDG:** [18F]-fluorodeoxy-D-glucose**; HL:** Hodgkin lymphoma; HR: hazard ratio; **IF-RT:** involved-field radiation therapy; **ITT:** intention-to-treat**; IQR:** interquartile range; **NPV:** negative predictive value**; OS:** overall survival; **PET:** positron emission tomography; **PET-CT:** positron emission tomography computed tomography; **PFS:** progression-free survival**; PPV:** positive predictive value.

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Adams 2016	Wrong publication type. Letter to the editor.
Adams 2017	Wrong publication type. Letter to the editor.
Adams 2018	Wrong publication type. Letter to the editor.
Adams 2018a	Wrong publication type. Letter to the editor.
Adams 2018b	Wrong publication type. Letter to the editor.
Adams 2019	Wrong publication type. Letter to the editor.
Advani 2007	Reported only end-of-chemotherapy PET scan results.
Afanasyev 2017	Wrong publication type. Protocol.
Albano 2017	PET-adapted outcomes.
Albano 2018	PET-adapted outcomes. Treatment was modified according to PET2 results.
Altamirano 2008	Wrong study population. Includes non-Hodgkin lymphoma patients.
Ansell 2016	Wrong publication type. Article.
Awan 2013	Wrong patient population. Patients with any lymphoma were included; no separate data for Hodgkin lymphoma patients.
Bar-Shalom 2003	Wrong study design. Comparison FDG PET and 67Ga scintigraphy.
Barrington 2011a	Wrong publication type. Meeting abstract.
Barrington 2017	Wrong publication type. Commentary.
Basu 2009	Wrong publication type. Commentary.
Becherer 2002	Wrong study design. End-of-chemotherapy PET. Includes non-Hodgkin lymphoma.
Bednaruk-Mlynski 2015	Wrong study design. Role of baseline PET/CT.
Biggi 2012	Wrong publication type. Conference abstract.
Biggi 2017	PET-adapted outcomes. Treatment was modified according to PET2 results.
Bishop 2015	Wrong publication type. Commentary.
Bjurberg 2006	Wrong treatment. Retrospective study of patients with residual tumour or suspected relapse after therapy.

Study	Reason for exclusion
Blum 2002	Wrong patient population. Non-Hodgkin lymphoma patients
Bodet-Milin 2008	Wrong patient population. Non-Hodgkin lymphoma patients.
Bodet-Milin 2009	Wrong publication type. Article.
Boisson 2007	Wrong publication type. Article.
Borchmann 2016	Wrong study design. Literature review.
Bucerius 2006	Wrong publication type. Conference abstract.
Carras 2018	PET-adapted outcomes. Treatment was modified according to PET2 results.
Ciammella 2016	PET-adapted outcomes.
Cremerius 1999	Wrong study design. Retrospective study to validate the clinical value of FDG-PET for therapy control.
Cremerius 2001	Wrong patient population. Patients with any lymphoma were included; no separate data for Hodgkin lymphoma patients.
Cuccaro 2016	PET-adapted outcomes. Positive interim PET results led to change in therapy in three patients; data from these patients was not reported separately from the study population in analysis.
D'Urso 2018	Wrong study design. Analysis of metabolic parameters.
Damlaj 2017	PET-adapted outcomes. Treatment was modified according to PET2 results.
Damlaj 2019	Pet-adapted outcomes.
Danilov 2017	PET-adapted outcomes. Treatment was modified according to interim PET results.
Dann 2009	PET-adapted outcomes.
Dann 2010	PET-adapted outcomes. Treatment was modified according to interim PET results.
Dann 2010a	PET-adapted outcomes.Treatment was modified according to PET2 results.
Dann 2012	PET-adapted outcomes.Treatment was modified according to PET2 results.
Dann 2013	PET-adapted outcomes.Treatment was modified according to PET2 results.
Dann 2016	PET-adapted outcomes. Treatment was modified according to PET2 results.
Dann 2017	PET-adapted outcomes. Treatment was modified according to PET2 results.
Dann 2018	Wrong publication type. Response to letter.
deAndres-Galiana 2015	Wrong study design. Prognostic factor identification study.
Diehl 2007	Wrong study design. The aim was to specify the negative predictive value of PET in patients with residual tumour mass after chemotherapy.

Study	Reason for exclusion
El-Galaly 2012	Wrong study design. The study evaluated the utility of PET scans for post-therapy routine surveillance imaging.
Evens 2014	Wrong publication type. Article.
Fanti 2008	Wrong study design. Case study.
Filmont 2003	Wrong patient population. Patients with aggressive lymphoma undergoing salvage therapy.
Fornecker 2017	PET-adapted outcomes. Treatment was modified according to interim PET results.
Freudenberg 2004	Wrong patient population. Patients with any lymphoma were included; no separate data for Hodgkin lymphoma patients.
Friedberg 2002	Wrong study design. The study intended to compare FDG-PET to gallium scintigraphy in the staging and follow-up of newly diagnosed patients with Hodgkin lymphoma.
Friedberg 2004	Wrong study design. The study intended to compare FDG-PET to gallium scintigraphy in the staging and follow-up of newly diagnosed patients with Hodgkin lymphoma.
Front 1999	Wrong treatment. The study investigated the utility of gallium scintigraphy performed early during treatment as a means to predict outcome and optimise treatment in Hodgkin lymphoma patients.
Fruchart 2006	Wrong patient population. Patients with B-cell lymphoma.
Gallamini 2008	Wrong publication type. Article.
Gallamini 2017	PET-adapted outcomes. Treatment was modified according to interim PET results.
Gallamini 2018	PET-adapted outcomes. Treatment was modified according to interim PET results.
Gallamini 2018a	Wrong publication type. Reply to letter.
Gallowitsch 2008	Wrong publication type. Commentary.
Goldschmidt 2011	Wrong patient population. Relapsed, aggressive non-Hodgkin lymphoma.
Greil 2018	PET-adapted outcomes. Treatment was modified according to interim PET results.
Guidez 2016	Wrong publication type. No abstract or full text.
Hagtvedt 2015	Wrong study design. Comparison between FDG-PET and diffusion-weighted magnetic resonance imaging for assessment of early treatment response in lymphoma.
Haioun 2005	Wrong patient population. Patients with any lymphoma were included; no separate data for Hodgkin lymphoma patients.
Hartmann 2012	Wrong study design. The study investigated protein expression patterns in different Hodgkin lymphoma subtypes.
Hartridge-Lambert 2013	Wrong study design. The study evaluated the risk of disease recurrence and the value of radiologic surveillance in patients treated with ABVD alone who achieved a complete remission according to post-treatment PET. PET was not treated as a prognostic factor.
Honda 2014	Wrong patient population. Letter to the editor, presenting the case of one patient with pulmonary Hodgkin lymphoma.

Study	Reason for exclusion
Hueltenschmidt 2001	Baseline and end-of-chemotherapy PET results.
Huic 2006	Wrong treatment. Patients within three months after completion of conventional initial therapy or salvage therapy with high-dose chemotherapy were included in the study population; no subgroup analysis was reported.
Hutchings 2007	End-of-chemotherapy PET.
Iagaru 2008	Wrong patient population. Patients with any lymphoma were included; no separate data for Hodgkin lymphoma patients.
Illidge 2015	PET-adapted outcomes. Commentary on a research news article about PET-adapted treatment in Hodgkin lymphoma patients.
Jerusalem 2003	End-of-chemotherapy PET.
Johnson 2015	PET-adapted outcomes.
Johnson 2016	PET-adapted outcomes.
Kamran 2016	PET-adapted outcomes.
Kamran 2018	PET-adapted outcomes. Treatment was modified according to PET2 results. Data was not reported separately for PET-positive and PET-negative patients.
Kobe 2008	Wrong study design. The study evaluated the negative predictive value of PET scans in advanced-stage Hodgkin lymphoma patients.
Kobe 2014	Wrong study design. The study evaluated how computed tomography might help improve the positive predictive value of PET in identifying potential high-risk patients.
Kostakoglu 2006	Wrong patient population. Patients with either diffuse large cell lymphoma or Hodgkin lymphoma were included; no separate data for Hodgkin lymphoma patients.
Li 2013	Wrong patient population. The study population consisted of patients with mature T-cell and natural killer cell lymphomas.
Lowe 2002	Wrong study design. Commentary.
Milgrom 2017	Wrong study design. The study population consisted mostly of PET-positive patients. The study compared data from PET-positive patients who received salvage chemotherapy or autologous stem cell transplantation with patients who received radiotherapy only.
Mocikova 2010	Wrong study design. The study evaluated the routine use of PET scans in Hodgkin lymphoma patients during follow-up and in cases of suspected relapse.
Mocikova 2011	Wrong treatment. The study evaluated the prognostic significance of pre-transplant PET scans after salvage chemotherapy before autologous stem cell transplant in patients with relapsed or refractory Hodgkin lymphoma.
Molnar 2010	End-of-chemotherapy PET.
Moskowitz 2015	PET-adapted outcomes. Treatment was modified according to interim PET results.
Naumann 2001	End-of-chemotherapy PET.

Study	Reason for exclusion
NCT00784537	PET-adapted outcomes. Treatment was modified according to interim PET results.
NCT00795613	PET-adapted outcomes. Treatment was modified according to interim PET results.
NCT01358747	PET-adapted outcomes. Treatment was modified according to interim PET results.
NCT01652261	PET-adapted outcomes. Study closed due to lack of recruitment.
NCT02292979	Wrong study design.
Nguyen 2017	PET-adapted outcomes. Treatment was modified according to interim PET results.
Panizo 2004	End of chemotherapy PET.
Paolini 2007	PET-adapted outcomes.
Pavlovsky 2019	PET-adapted outcomes.
Pichler 2000	Wrong study design. Comparison of FDG-Hybrid-PET scans.
Reinhardt 2005	Wrong study design. The study evaluated the accuracy of computed tomography and FDG-PET for prediction of progression-free survival of Hodgkin lymphoma and non-Hodgkin lymphoma patients after completion of therapy.
Rigacci 2002	Wrong study design. Letter.
Rigacci 2017	Wrong study design. Letter.
Rubello 2015	Wrong study design. The study evaluated the variability of FDG liver uptake in patients with Hodgkin lymphoma.
Sakr 2017	Wrong study design.
Schot 2007	Wrong treatment. The study population included patients with recurring lymphoma who were treated with second-line chemotherapy followed by autologous stem cell transplantation.
Simontacchi 2015	PET-adapted outcomes. Treatment was modified according to interim PET results.
Slaby 2002	Wrong patient population. Patients with any lymphoma were included; no separate data for Hodgkin lymphoma patients.
Spaepen 2001	Reported only end-of-chemotherapy PET scan results.
Specht 2007	Wrong publication type. Article.
Spinner 2018	Wrong publication type. Article.
Straus 2018	PET-adapted outcomes. Treatment was modified according to interim PET results.
Strigari 2016	Wrong study design. The aim of the study was to present a novel quantitative tool to refine the risk-class assessment of the Deauville criteria.
Sucak 2011	Wrong treatment. The study population included patients with relapsed or refractory lymphoma post-autologous stem cell transplantation.

Study	Reason for exclusion
Tirelli 2015	Wrong publication type. Article.
Tomita 2015	Wrong patient population. The study population consisted of patients with peripheral T cell lymphoma.
Torizuka 2004	PET-adapted outcomes. Includes non-Hodgkin lymphoma patients.
Trotman 2017	PET-adapted outcomes. Treatment was modified according to interim PET results.
Tseng 2012	Wrong patient population. The study population included relapsed patients.
Villa 2018	PET-adapted outcomes. Treatment was modified according to interim PET results.
Weidmann 1999	Wrong patient population. Includes relapsed patients.
Wilson 2018	Wrong publication type. Commentary.
Xie 2018	Wrong publication type. Review.
Yasgur 2015	Wrong publication type. Commentary.
Yoshimi 2008	Wrong treatment. The study population included lymphoma patients with a poor prognosis who had received FDG-PET scans within one month before allogeneic stem cell transplantation.
Zabrocka 2016	Wrong study design. The study evaluated the current usage of PET scans and its clinical usefulness at different points in Hodgkin lymphoma management based on a single-institution experience.
Zaucha 2009	Wrong publication type. Review.
Zinzani 1999	Wrong patient population. Includes non-Hodgkin lymphoma patients.
Zinzani 2002	Wrong patient population. Includes non-Hodgkin lymphoma patients.
Zinzani 2016	PET-adapted outcomes. Treatment was modified according to interim PET results.

ABVD: adriamycin/doxorubicin, bleomycin, vinblastine and dacarbazine; **FDG:** [18F]-fluorodeoxy-D-glucose**; PET:** positron emission tomography; **PET-CT:** positron emission tomography computed tomography.

Characteristics of studies awaiting classification *[ordered by study ID]*

Abramson 2010

Notes	Title: End of treatment but not interim PET scan predicts outcome in non-bulky limited stage Hodgkin lymphoma. (Conference abstract) Aim • To establish the prognostic value of interim PET scans in limited stage patients with non-bulky disease Study design • Retrospective Country/treatment centre(s)

Abramson 2010 *(Continued)*

- USA (Massachusetts General Hospital, Dana-Farber Cancer Institute, Harvard School of Public Health, Boston MA)

Number of included participants

- 96

Inclusion criteria

- Non-bulky limited stage cHL treated at the institutions between 2000 and 2008; Bulk was defined as a mass >=10 cm or >=1/3 of the intrathoracic diameter.

Exclusion criteria

- None

Treatment

- 4 to 6 cycles of ABVD with or without IFRT

Primary outcome measure(s)

- Overall survival
- Progression-free survival

Algrin 2010

Notes

Title: Interim-positron emission tomography with [18F]fluorodeoxyglucose (interim-PET) evaluation in mediastinal lymphoma including Hodgkin lymphoma (HL) and primary mediastinal large B-cell lymphoma (PMBL).

(Conference abstract)

Aim

- To investigate the prognostic value of qualitative and semi-quantitative evaluations of interim-PET in mediastinal lymphoma

Study design

- Retrospective

Country/treatment centre(s)

- Not reported

Number of included participants

- 48

Inclusion criteria

- Previously untreated, age under 60 at diagnosis and at least one interim-PET evaluation available

Exclusion criteria

- Individuals with sub-diaphragmatic or medullar localisations of lymphoma

Treatment

- Not reported

Primary outcome measure(s)

Algrin 2010 *(Continued)*

- Event-free survival

Arce-Calisaya 2013

Notes

Title: Interim FDG PET-CT in Hodgkin's lymphoma - Does binary response assessment criteria have any prognostic value?

(Conference abstract)

Aim

- To evaluate whether binary response assessment criteria (positive or negative) has any prognostic significance after 2 cycles of ABVD therapy

Study design

- Retrospective

Country/treatment centre(s)

- UK

Number of included participants

- 99

Inclusion criteria

- Newly diagnosed adults with advanced-stage HL undergoing baseline and interim (post-2 cycles ABVD) 18F-FDG PET-CT

Exclusion criteria

- None

Treatment

- ABVD

Primary outcome measure(s)

- Recurrence-free survival after 1 year

Baratto 2015

Notes

Title: Interim-PET in Hodgkin lymphoma: Deauville criteria and metabolic parameters as prognostic factors.

(Conference abstract)

Aim

- To explore the prognostic role of i-PET in individuals with HL

Study design

- Retrospective

Country/treatment centre(s)

- Italy

Baratto 2015 *(Continued)*

Number of included participants

- 83

Inclusion criteria

- Newly diagnosed HL, stage I-IV disease

Exclusion criteria

- None

Treatment

- Not reported

Primary outcome measure(s)

- Overall survival
- Disease-free survival

Barna 2011

Notes

Title: Prognostic value of interim 18FDG-PET-CT in patients with Hodgkin's lymphoma using different 5-point visual scales for interpretation.

(Conference abstract)

Aim

- To compare the effect on prognosis of the currently applied MRU definitions

Study design

- Prospective

Country/treatment centre(s)

- Hungary

Number of included participants

- 82

Inclusion criteria

- Newly-diagnosed HL

Exclusion criteria

- None

Treatment

- 6 courses of ABVB/EBVD, additional radiotherapy according to the protocol

Primary outcome measure(s)

- Overall survival
- Progression-free survival

Barrington 2011

Notes	Title: Are the Deauville criteria a reliable tool for assessment of interim PET in Hodgkin lymphoma? (Conference abstract) Aim • To measure agreement between experienced reporters reading interim PET-CT scans from an international cohort of patients according to the Deauville criteria • To measure progression-free survival in advanced HL according to interim PET Study design • Not reported Country/treatment centre(s) • International study Number of included participants • 262 Inclusion criteria • Individuals diagnosed with stage IIB-IV HL Exclusion criteria • None Treatment • ABVD Primary outcome measure(s) • Progression-free survival

Bentur 2017

Notes	Title: The predictive value of interim PET-CT in elderly patients with Hodgkin lymphoma. This is an abstract only and a lot of relevant information is missing. A full-text has not been published yet. It is particularly unclear whether participants have received treatment adaptation based on the interim PET result. Authors need to be contacted for more information. Aim • To evaluate the significance of iPET in elderly individuals with HL Study design • Retrospective study (1998 to 2016) Country/treatment centre(s) • Unclear, multicentre study (5 centres) Number of included participants • 95

Bentur 2017 *(Continued)*

Inclusion criteria

- Individuals diagnosed with HL between 1998 to 2016
- Older adults (>= 60 years)

Exclusion criteria

- Not reported

Treatment

- Fifty-nine participants received first-line treatment with ABVD, in 13 participants chemotherapy was followed by IVRT (treatment unclear for the remaining participants)

Primary outcome measure(s)

- Overall survival
- Progression-free survival
- Time frame: five years

Berenger 2010

Notes

Title: Prognostic value of interim 18F-FDG PET-CT in mediastinal bulky Hodgkin lymphoma.

(Conference abstract)

Aim

- To determine if Negative Predictive Value (NPV) remains high in individuals who present with mediastinal bulky disease

Study design

- Retrospective

Country/treatment centre(s)

- France

Number of included participants

- 38

Inclusion criteria

- Previously untreated individuals with HL, with localiSed mediastinal bulky disease

Exclusion criteria

- None

Treatment

- Chemotherapy with or without additional radiotherapy

Primary outcome measure(s)

- Progression-free survival
- NPV and PPV of iPET

Bhatwadekar 2017

Notes	Title: Excellent outcome in Hodgkin lymphoma with ABVD and CMT: A single-centre retrospective analysis. (Conference abstract) Aim • To evaluate the outcome of individuals with HL receiving ABVD alone or in combination with RT Study design • Retrospective Country/treatment centre(s) • India (Haemato Oncology Care Centre, Vadodara) Number of included participants • 63 Inclusion criteria • Not reported Exclusion criteria • Not reported Treatment • ABVD alone or in combination with RT Primary outcome measure(s) • Overall survival • Progression-free survival

Cimino 2014

Notes	Title: The complementary prognostic role of baseline and interim PET in predicting treatment outcome in advanced-stage Hodgkin lymphoma. (Conference abstract) Aim • To evaluate the contribution of PET combined with computed tomography (PET-CT) and contrast enhanced computed tomography (ceCT) in the staging and in the prognostication of untreated advanced HL Study design • Retrospective Country/treatment centre(s) • Italy, Poland, Denmark (multicentre) Number of included participants • 162

Cimino 2014 *(Continued)*

Inclusion criteria

- Not reported

Exclusion criteria

- Not reported

Treatment

- ABVD with or without RT

Primary outcome measure(s)

- Overall survival
- Event-free survival

Cocorocchio 2009

Notes

Title: Prognostic role of interim 18FDG-PET in Hodgkin lymphoma: A single-center experience.

(Conference abstract)

Aim

- Single-centre experience with using 18FDG-PET as a prognostic factor for long term complete remission (CR)

Study design

- Retrospective

Country/treatment centre(s)

- Italy

Number of included participants

- 65

Inclusion criteria

- Newly diagnosed with HL

Exclusion criteria

- Not reported

Treatment

- VBM or ChlVPP/ABVVP followed by IFRT

Primary outcome measure(s)

- Complete remission
- Freedom from treatment failure

Cocorocchio 2011

Notes	Title: Evaluation of interim 18FDG-PET in advanced Hodgkin lymphoma (HL) patients (PTS) treated with ChlVPP/ABVVP regimen. (Conference abstract) Aim • To evaluate the prognostic value of interim 18 FDG-PET in advanced HL patients treated with intensified ChlVPP/ABVVP Study design • Not reported Country/treatment centre(s) • Italy Number of included participants • 70 Inclusion criteria • Not reported Exclusion criteria • Not reported Treatment • 6 cycles of ChlVPP/ABVVP Primary outcome measure(s) • Overall survival • Freedom from treatment failure

Copeland 2010

Notes	Title: Single institution experience with interim PET evaluation in newly diagnosed CHL receiving ABVD chemotherapy: Need for standardization. (Conference abstract) Aim • To evaluate the use of interim PET for the identification of individuals with classic HL, who are at risk for relapse after first-line therapy Study design • Retrospective Country/treatment centre(s) • USA (MD Anderson Cancer Center, Houston TX) Number of included participants • 57

Copeland 2010 *(Continued)*

Inclusion criteria

- Newly diagnosed cHL

Exclusion criteria

- Not reported

Treatment

- ABVD

Primary outcome measure(s)

- Event-free survival

Cuzzocrea 2015

Notes

Title: The Deauville criteria and metabolic parameters as prognostic factors in interim PET in Hodgkin lymphoma: A single centre experience.

(Conference abstract)

Aim

- To explore the prognostic role of i-PET in individuals with HL

Study design

- Retrospective

Country/treatment centre(s)

- Italy

Number of included participants

- 83

Inclusion criteria

- Newly diagnosed HL, stage I-IV disease

Exclusion criteria

- Not reported

Treatment

- Not reported

Primary outcome measure(s)

- Overall survival
- Disease-free survival

De Rueda 2013

Notes

Title: Prognostic value of 18F-FDG PET-CT in Hodgkin lymphoma.

(Conference abstract)

De Rueda 2013 *(Continued)*

Aim

- To determine the value of 18F-FDG PET-CT after the second and sixth cycle of first line therapy with ABVD or BEACOPP in the outcome of individuals with HL

Study design

- Retrospective, January 2007 to December 2012

Country/treatment centre(s)

- Spain

Number of included participants

- 79

Inclusion criteria

- HL diagnosis

Exclusion criteria

- Not reported

Treatment

- ABVD or BEACOPP

Primary outcome measure(s)

- Progression-free survival

Fabbri 2011

Notes

Title: 'Early FDG-PET' predicts clinical course of Hodgkin's lymphoma although does not correlate with macrophages infiltration in diagnostic specimens.

(Conference abstract)

Aim

- To verify the prognostic role both of "early-FDG PET" and of macrophagic infiltration, and to test if "early-FDG PET" positivity could correlate with high macrophagic infiltration in diagnostic specimens

Study design

- Retrospective, February 2007 to July 2010

Country/treatment centre(s)

- Italy (Siena and Florence haematology departments)

Number of included participants

- 52

Inclusion criteria

- Diagnosed HL

Exclusion criteria

Fabbri 2011 *(Continued)*

- Not reported

Treatment

- 4 to 6 cycles of ABVD with or without IFRT

Primary outcome measure(s)

- Complete remission
- CD68 expression

Fiore 2010

Notes

Title: Early interim FDG-PET during intensified BEACOPP therapy for advanced-stage Hodgkin disease shows a lower positive predictive value than during ABVD.

(Conference abstract)

Aim

- To examine the predictive role on treatment outcome of early interim FDG-PET in individuals with HL, treated with BEACOPP (4 escalated + 4 baseline cycles)

Study design

- Retrospective

Country/treatment centre(s)

- Italy (8 haematological institutions)

Number of included participants

- 44

Inclusion criteria

- Diagnosed HL, advanced stage (IIB to IVB, or IIA with adverse prognostic factors)

Exclusion criteria

- Not reported

Treatment

- BEACOPP

Primary outcome measure(s)

- Complete remission
- Failure-free survival

Gallegos 2012

Notes

Title: The importance of PET-CT as method of evaluation of early response to treatment in HL.

(Conference abstract)

Aim

Gallegos 2012 *(Continued)*

- To assess the importance of PET-CT as method of evaluation of early response to treatment in HL

Study design

- Retrospective, 2002 to 2011

Country/treatment centre(s)

- Spain (The Miguel Servet's Hospital, Zaragoza)

Number of included participants

- 61

Inclusion criteria

- Diagnosed HL, first-line therapy

Exclusion criteria

- Not reported

Treatment

- ABVD or BEACOPP

Primary outcome measure(s)

- Progression-free survival

Hohaus 2015

Notes

Title: The risk of progression of Hodgkin lymphoma in patients with negative interim PET: A role for the number of tumor-infiltrating macrophages (CD68+ cell counts) and B symptoms.

(Conference abstract)

Aim

- To evaluate if integration of the response evaluation with iPET with parameters available at diagnosis could add prognostic information, allowing a better risk-stratification of individuals with HL

Study design

- Retrospective, 2007 to 2014

Country/treatment centre(s)

- Italy (Università Cattolica del Sacro Cuore, Rome)

Number of included participants

- 102

Inclusion criteria

- Diagnosed classic HL

Exclusion criteria

- Not reported

Treatment

Hohaus 2015 *(Continued)*

- ABVD

Primary outcome measure(s)

- Progression-free survival

Hutchings 2010

Notes

Title: correlation of FDG-PET results after one cycle and after two cycles of chemotherapy in Hodgkin lymphoma.

(Conference abstract)

Aim

- To study the correlation of PET results after one cycle and after two cycles of chemotherapy, and to investigate if the high predictive value of PET after two cycles is obtainable already after one cycle of chemotherapy

Study design

- Prospective trial

Country/treatment centre(s)

- Denmark (Copenhagen), USA (New York)

Number of included participants

- 36

Inclusion criteria

- Diagnosed HL

Exclusion criteria

- Not reported

Treatment

- ABVD or BEACOPPesc

Primary outcome measure(s)

- Negative predictive value

Knight-Greenfield 2013

Notes

Title: Interim FDG PET-CT to predict progression-free survival (PFS) better than clinical and baseline metabolic measurements in Hodgkin lymphoma (cHL).

(Conference abstract)

Aim

- To determine the best predictor of PFS among various variables of tumour metabolic measurements at baseline and at interim PET-CT compared to conventional methods in individuals with classic HL

Study design

Knight-Greenfield 2013 *(Continued)*

- Retrospective

Country/treatment centre(s)

- Not reported

Number of included participants

- 58

Inclusion criteria

- Diagnosed classic HL, ABVD therapy, minimal follow-up of 2 years

Exclusion criteria

- Not reported

Treatment

- ABVD

Primary outcome measure(s)

- Progression-free survival

Leontjeva 2016

Notes

Title: Significance of early interim PET results in advanced Hodgkin lymphoma treated intensive program EACOPP-14.

(Conference abstract)

Aim

- To evaluate the use of interim PET to guide treatment in advanced stage individuals with classic HL

Study design

- Not reported, December 2009 to December 2013

Country/treatment centre(s)

- Russia

Number of included participants

- 36

Inclusion criteria

- Newly diagnosed classic HL (stages IIB to IV, or IIA with bulk), adults

Exclusion criteria

- Not reported

Treatment

- EACOPP-14 with or without RT

Primary outcome measure(s)

- Progression-free survival

Luminari 2010

Notes	Title: The use of FDG positron emission tomography (FDG-PET) in patients with Hodgkin lymphoma (HL) in the "real world": A population based study from northern Italy. (Conference abstract) Aim • To assess how FDG-PET is currently used in individuals with HL Study design • Unclear, 2006 to 2008 Country/treatment centre(s) • Italy (Cancer Registries in Modena, Ferrara, Parma and Reggio Emilia) Number of included participants • 136 Inclusion criteria • Diagnosed HL, adults (18 to 75 years), HIV negative Exclusion criteria • None Treatment • Not reported Primary outcome measure(s) • Overall survival • Relapse-free survival • Failure-free survival

Luminari 2011

Notes	Title: Use of 2-[^{18}F]fluoro-2-deoxy-D-glucose positron emission tomography in patients with Hodgkin lymphoma in daily practice: a population-based study from Northern Italy Authors need to be contacted to clarify whether the treatment has been adapted based on the interim PET results. Aim • To investigate how PET is currently used in daily practice and whether results obtained in clinical trials and retrospective series can be generalised to all individuals with HL. Study design • Retrospective Country/treatment centre(s) • Italy • Participants were identified from archives of four population-based Italian cancer registries

Luminari 2011 *(Continued)*

Number of included participants

- Total: 136

Inclusion criteria

- Registration in one of the four population-based Italian cancer registries (Modena, Reggio Emilia, Parma, Ferrara)
- Histologicallyconfirmed diagnosis of HL between 1 January 2006 and 31 December 2008, age between 18 and 75 years, and human immunodeficiency virus (HIV) negativity

Exclusion criteria

- Missing data

Treatment

- N = 116 (85%) participants received ABVD chemotherapy, 11 participants (8%) received intensified regimens such as BEACOPP or COPP/EBV/CAD, six participants (4%) received chemotherapy without anthracycline such as VBM or MOPP, and three participants (3%) received other therapies such as radiotherapy alone

Primary outcome measure(s)

- Failure-free survival
- Overall survival

Medvedovskaya 2016

Notes

Title: The impact of outcome of interim PET-CT on advanced Hodgkin lymphoma treated with EA-COPP-14.

(Conference abstract)

Aim

- To assess the role of interim PET-CT and compare it with PET-CT results after the end of treatment in individuals with advanced stage classic HL

Study design

- Not reported

Country/treatment centre(s)

- Russia

Number of included participants

- 114

Inclusion criteria

- Newly diagnosed classic HL

Exclusion criteria

- None

Treatment

- 6 cycles of EACOPP-14 with or without RT

Medvedovskaya 2016 *(Continued)*

Primary outcome measure(s)

- Complete metabolic response

Molnar 2011

Notes	Title: The value of interim 18F-FDG PET-CT in Hodgkin lymphoma. (Conference abstract) Aim • To summarise our experience with 18F-FDG PET-CT in HL Study design • Retrospective, November 2006 to January 2010 Country/treatment centre(s) • Hungary (National Institute of Oncology, Budapest) Number of included participants • 60 Inclusion criteria • Not reported Exclusion criteria • Not reported Treatment • ABVD or BEACOPPesc, with or without RT Primary outcome measure(s) • Prognostic value

Molnar 2011a

Notes	Title: Interim FDG PET-CT examinations in advanced stage Hodgkin lymphoma. (Conference abstract) Aim • To summarise our experience with 18F-FDG PET-CT in interim staging Study design • Retrospective, November 2007 to January 2010 Country/treatment centre(s) • Hungary (National Institute of Oncology, Budapest) Number of included participants

Molnar 2011a *(Continued)*

- 19

Inclusion criteria

- Not reported

Exclusion criteria

- Not reported

Treatment

- ABVD or BEACOPPesc, with or without RT

Primary outcome measure(s)

- Prognostic value

Moreira 2013

Notes

Title: Prognostic value of interim vs. end-of-treatment PET scan in Hodgkin's lymphoma.

(Confernece abstract)

Aim

- To evaluate the prognostic value of interim PET scan (PET2) and end-of-treatment PET (PET6) in the outcome of individuals with HL

Study design

- Retrospective, January 2004 to December 2011

Country/treatment centre(s)

- Portugal (Porto, single-centre)

Number of included participants

- 261

Inclusion criteria

- Diagnosed HL

Exclusion criteria

- PET-guided treatment adaptation

Treatment

- ABVD, BEACOPPesc or CVP/CEB

Primary outcome measure(s)

- Complete remission
- Overall survival
- Progression-free survival

Perrone 2009

Notes	Title: Role of positron emission tomography (PET) after 2 and 4 courses of chemotherapy in patients with Hodgkin's lymphoma: A single center experience. (Conference abstract) Aim • To investigate the value of PET performed after 2 (PET2) and 4 (PET4) cycles of therapy for the management of patients with HL Study design • Not reported, September 2006 to September 2008 Country/treatment centre(s) • Italy (University of Bari) Number of included participants • 26 Inclusion criteria • Newly diagnosed HL Exclusion criteria • None Treatment • ABVD Primary outcome measure(s) • Complete remission • Partial remission • Progression-free survival

Pophali 2014

Notes	Title: Bulky disease does not adversely affect overall survival in early stage Hodgkin lymphoma: Role of interim PET and possible omission of radiotherapy in select patients. (Conference abstract) Aim • To assess the impact of disease bulk, interim PET and treatment modality on outcomes Study design • Retrospective, 1995 to 2011 Country/treatment centre(s) • USA (Cleveland Clinic, Cleveland OH) Number of included participants • 121

Pophali 2014 *(Continued)*

Inclusion criteria

- Previously untreated HL, early stages (I and II)

Exclusion criteria

- Missing clinical data

Treatment

- ABVD or other chemotherapy (not specified)

Primary outcome measure(s)

- Overall survival
- Progression-free survival

Rusconi 2010

Notes

Title: Baseline and dynamic prognostic factors in newly diagnosed classical Hodgkin's lymphoma.

(Conference abstract)

Aim

- To identify characteristics, both at baseline and during therapy, predictive for survival outcomes in HL

Study design

- Retrospective

Country/treatment centre(s)

- Italy (Niguarda Hospital, Milan)

Number of included participants

- 105

Inclusion criteria

- Diagnosed HL

Exclusion criteria

- None

Treatment

- 3 to 8 cycles of ABVD with or without IFRT

Primary outcome measure(s)

- Overall survival
- Event-free survival
- Relapse-free survival

Spallino 2017

Notes	Title: The Deauville criteria and QPET as prognostic factors in interim PET in adult Hodgkin lymphoma: A single centre experience. (Conference abstract) Aim • To explore the prognostic role of iPET in individuals with HL by correlating Deauville criteria and qPET to DFS and OS Study design • Retrospective Country/treatment centre(s) • Italy Number of included participants • 131 Inclusion criteria • Newly diagnosed HL, disease stages I to IV Exclusion criteria • None Treatment • Not reported Primary outcome measure(s) • Overall survival • Disease-free survival

Yaghmour 2012

Notes	Title: PET-negative at 2, 3 or 4 cycles of ABVD in Hodgkin's lymphoma is still good. (Conference abstract) Aim • To assess the prognostic value of anytime negative PET scan in the course of first line treatment in individuals with HL receiving ABVD Study design • Retrospective Country/treatment centre(s) • USA (Henry Ford Health System, Detroit MI) Number of included participants • 32 Inclusion criteria

Yaghmour 2012 *(Continued)*

- Newly diagnosed HL

Exclusion criteria

- Not reported

Treatment

- ABVD

Primary outcome measure(s)

- Overall survival

Zanoni 2011

Notes

Title: The predictive value of interim PET and immunohistochemical markers in Hodgkin lymphoma (HL).

(Conference abstract)

Aim

- To compare iPET with a series of histological and immunohistochemical parameters obtained on tissue-micro-arrays as possible predictive factors

Study design

- Retrospective

Country/treatment centre(s)

- Italy (Bologna)

Number of included participants

- 209

Inclusion criteria

- Biopsy-proven HL, complete clinical and iPET data

Exclusion criteria

- None

Treatment

- Not reported

Primary outcome measure(s)

- Overall survival
- Progression-free survival

ABVD: adriamycin/doxorubicin, bleomycin, vinblastine and dacarbazine; **BEACOPP:** bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine and prednisone; **EBVD:** epirubicin, bleomycin, vinblastine and dacarbazine; **FDG:** [18F]-fluorodeoxy-D-glucose**; HL:** Hodgkin lymphoma; **IF-RT:** involved-field radiation therapy; **iPET:** interim positron emission tomography**; MOPP:** mustargen, Oncovin, procarbazine and prednisone; **NPV:** negative predictive value**; PET:** positron emission tomography; **PET-CT:** positron emission tomography computed tomography; **RT:** radiotherapy.

Characteristics of ongoing studies *[ordered by study ID]*

NCT00736320

Study name	HD16 for Early Stages - Treatment optimisation trial in the first-line treatment of early stage Hodgkin lymphoma; treatment stratification by means of FDG-PET
Starting date	November 2009
Contact information	Prof. Dr. Andreas Engert, University of Cologne, Germany
Notes	Study design • Randomised clinical trial (phase III) including 1150 participants with HL Country/treatment centre • 1st Department of Medicine, Cologne University Hospital, Cologne, Germany Number of included participants • Total: 1150 Inclusion criteria • Hodgkin lymphoma • Adults (18 to 75 years) • CS I and II without risk factors (large mediastinal mass (> 1/3 of maximum transverse thorax diameter), extranodal involvement, elevated ESR, three or more involved nodal areas) • Written informed consent Exclusion criteria • Leukocytes < 3000/µl • Platelets < 100000/µl • Hodgkin lymphoma as composite lymphoma • Activity index (WHO) > 2 Arms and interventions • Active comparator (A): two cycles ABVD followed by 20 Gy IF-RT, irrespective of FDG-PET results after chemotherapy • Expertimental (B): two cycles ABVD followed by 20 Gy IF-RT if FDG-PET is positive after chemotherapy; 2 cycles ABVD and treatment stop if FDG-PET is negative after chemotherapy Primary outcome measure(s) • Progression-free survival • Time frame: five years Secondary outcome measure(s) • Overall survival • Acute toxicity • Late toxicity • Complete response rate • Time frame: five years Estimated study completion date • May 2020

ESR: erythrocyte sedimentation rate; **FDG:** [18F]-fluorodeoxy-D-glucose; **HL:** Hodgkin lymphoma; **IF-RT** :involved-field radiation therapy; **MDCT:** multi detector computed tomography; **WHO:** World Health Organization

DATA AND ANALYSES

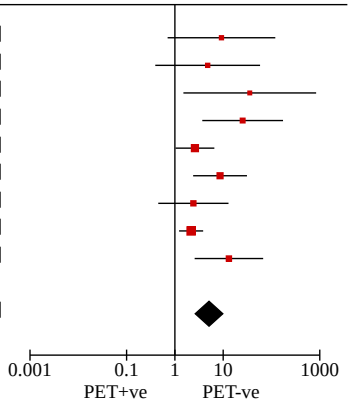
Comparison 1. Univariable comparison of PET+ve vs. PET-ve

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.1 Overall survival	9	1802	Hazard Ratio (IV, Random, 95% CI)	5.09 [2.64, 9.81]
1.2 Progression-free survival	14	2079	Hazard Ratio (IV, Random, 95% CI)	4.90 [3.47, 6.90]

Analysis 1.1. Comparison 1: Univariable comparison of PET+ve vs. PET-ve, Outcome 1: Overall survival

Study or Subgroup	log[Hazard Ratio]	SE	PET+ve Total	PET-ve Total	Weight	Hazard Ratio IV, Random, 95% CI
Barnes 2011	2.220623	1.309795	17	79	5.3%	9.21 [0.71 , 120.03]
Cerci 2010	1.563053	1.274372	30	74	5.6%	4.77 [0.39 , 58.02]
Hutchings 2005	3.570366	1.614442	22	63	3.7%	35.53 [1.50 , 841.02]
Hutchings 2014	3.231135	0.981951	37	89	8.3%	25.31 [3.69 , 173.42]
Kobe 2018	0.9555	0.4724	236	486	18.6%	2.60 [1.03 , 6.56]
Simon 2016	2.153725	0.654523	32	89	13.9%	8.62 [2.39 , 31.08]
Touati 2014	0.881751	0.854282	24	44	10.1%	2.42 [0.45 , 12.89]
Zaucha 2017	0.774	0.2928	24	152	24.1%	2.17 [1.22 , 3.85]
Zinzani 2012	2.575668	0.833487	53	251	10.4%	13.14 [2.57 , 67.31]
Total (95% CI)			**475**	**1327**	**100.0%**	**5.09 [2.64 , 9.81]**

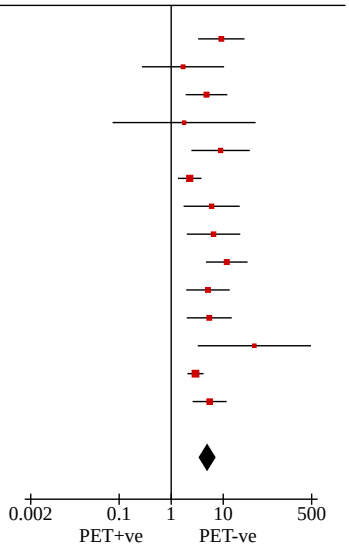
Heterogeneity: Tau² = 0.38; Chi² = 14.39, df = 8 (P = 0.07); I² = 44%
Test for overall effect: Z = 4.87 (P < 0.00001)
Test for subgroup differences: Not applicable



Analysis 1.2. Comparison 1: Univariable comparison of PET+ve vs. PET-ve, Outcome 2: Progression-free survival

Study or Subgroup	log[Hazard Ratio]	SE	PET+ve Total	PET-ve Total	Weight	Hazard Ratio IV, Random, 95% CI
Annunziata 2016	2.2192	0.5231	12	56	7.1%	9.20 [3.30 , 25.65]
Barnes 2011	0.5261	0.9261	17	79	3.0%	1.69 [0.28 , 10.39]
Cerci 2010	1.5624	0.4706	30	74	8.1%	4.77 [1.90 , 12.00]
Hutchings 2005	0.570366	1.6144	22	63	1.1%	1.77 [0.07 , 41.87]
Hutchings 2006	2.187	0.6587	16	61	5.2%	8.91 [2.45 , 32.40]
Kobe 2018	0.8198	0.2651	236	486	13.4%	2.27 [1.35 , 3.82]
Mesguich 2016	1.7891	0.6333	16	60	5.5%	5.98 [1.73 , 20.70]
Rossi 2014	1.873	0.6031	13	46	5.9%	6.51 [2.00 , 21.22]
Simon 2016	2.4596	0.4697	32	89	8.1%	11.70 [4.66 , 29.38]
Straus 2011	1.6268	0.49	24	64	7.7%	5.09 [1.95 , 13.29]
Touati 2014	1.685	0.5075	24	44	7.4%	5.39 [1.99 , 14.58]
Ying 2014	3.6783	1.278	10	25	1.7%	39.58 [3.23 , 484.51]
Zaucha 2017	1.0753	0.1812	24	152	16.0%	2.93 [2.05 , 4.18]
Zinzani 2012	1.6982	0.3845	53	251	10.0%	5.46 [2.57 , 11.61]
Total (95% CI)			**529**	**1550**	**100.0%**	**4.90 [3.47 , 6.90]**

Heterogeneity: Tau² = 0.16; Chi² = 23.52, df = 13 (P = 0.04); I² = 45%
Test for overall effect: Z = 9.07 (P < 0.00001)
Test for subgroup differences: Not applicable



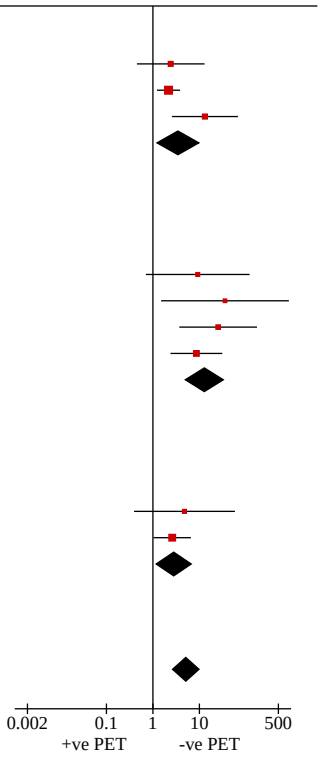
Comparison 2. Subgroups in univariable comparison of OS: PET+ve vs. PET-ve

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
2.1 OS by radiotherapy	9	1802	Hazard Ratio (IV, Random, 95% CI)	5.09 [2.64, 9.81]
2.1.1 Involved node and/or site	3	548	Hazard Ratio (IV, Random, 95% CI)	3.45 [1.22, 9.72]
2.1.2 involved field	4	428	Hazard Ratio (IV, Random, 95% CI)	12.75 [4.98, 32.68]
2.1.3 not specified	2	826	Hazard Ratio (IV, Random, 95% CI)	2.80 [1.17, 6.67]
2.2 OS by study design	8	1717	Hazard Ratio (IV, Random, 95% CI)	4.63 [2.43, 8.80]
2.2.1 Prospective	3	406	Hazard Ratio (IV, Random, 95% CI)	5.35 [1.07, 26.68]
2.2.2 Retrospective	4	589	Hazard Ratio (IV, Random, 95% CI)	7.12 [3.14, 16.14]
2.2.3 RCT	1	722	Hazard Ratio (IV, Random, 95% CI)	2.60 [1.03, 6.56]
2.3 OS by chemotherapy	9	1802	Hazard Ratio (IV, Random, 95% CI)	5.09 [2.64, 9.81]
2.3.1 ABVD	5	801	Hazard Ratio (IV, Random, 95% CI)	5.19 [2.11, 12.72]
2.3.2 ABVD and/or other	3	279	Hazard Ratio (IV, Random, 95% CI)	10.30 [1.71, 62.13]
2.3.3 BEACOPP	1	722	Hazard Ratio (IV, Random, 95% CI)	2.60 [1.03, 6.56]
2.4 OS for PET/CT vs PET	8	1706	Hazard Ratio (IV, Random, 95% CI)	5.01 [2.50, 10.02]
2.4.1 PET/CT	5	595	Hazard Ratio (IV, Random, 95% CI)	4.70 [1.86, 11.86]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
2.4.2 PET only	3	1111	Hazard Ratio (IV, Random, 95% CI)	6.99 [1.58, 30.90]
2.5 OS by disease stage	9	1802	Odds Ratio (IV, Random, 95% CI)	5.09 [2.64, 9.81]
2.5.1 Stages I and II with A and B symptoms	1	96	Odds Ratio (IV, Random, 95% CI)	9.21 [0.71, 120.03]
2.5.2 All stages	7	984	Odds Ratio (IV, Random, 95% CI)	6.28 [2.62, 15.05]
2.5.3 Advanced	1	722	Odds Ratio (IV, Random, 95% CI)	2.60 [1.03, 6.56]
2.6 Timing of interim PET	9	1802	Hazard Ratio (IV, Random, 95% CI)	5.09 [2.64, 9.81]
2.6.1 PET2	6	1495	Hazard Ratio (IV, Random, 95% CI)	3.53 [1.97, 6.32]
2.6.2 Other (including mixed)	3	307	Hazard Ratio (IV, Random, 95% CI)	20.13 [5.04, 80.38]
2.7 OS by HR type of estimation	9	1802	Hazard Ratio (IV, Random, 95% CI)	5.09 [2.64, 9.81]
2.7.1 precise	7	1638	Hazard Ratio (IV, Random, 95% CI)	5.70 [2.60, 12.48]
2.7.2 Imprecise	2	164	Hazard Ratio (IV, Random, 95% CI)	3.60 [0.89, 14.64]

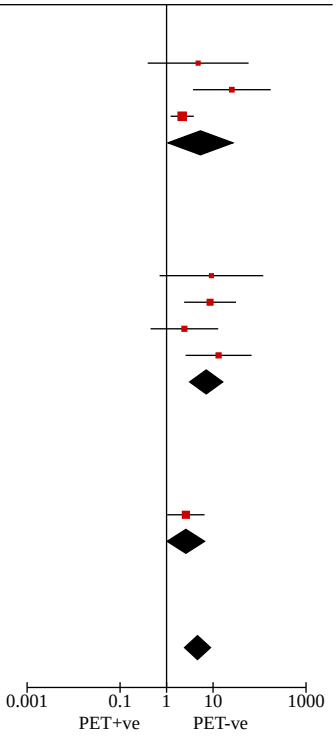
Analysis 2.1. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 1: OS by radiotherapy

Study or Subgroup	log[Hazard Ratio]	SE	Experimental Total	Control Total	Weight	Hazard Ratio IV, Random, 95% CI
2.1.1 Involved node and/or site						
Touati 2014	0.881751	0.8542821	24	44	10.1%	2.42 [0.45 , 12.89]
Zaucha 2017	0.774	0.2928	24	152	24.1%	2.17 [1.22 , 3.85]
Zinzani 2012	2.575668	0.8334869	53	251	10.4%	13.14 [2.57 , 67.31]
Subtotal (95% CI)			**101**	**447**	**44.6%**	**3.45 [1.22 , 9.72]**
Heterogeneity: Tau² = 0.45; Chi² = 4.17, df = 2 (P = 0.12); I² = 52%						
Test for overall effect: Z = 2.34 (P = 0.02)						
2.1.2 involved field						
Barnes 2011	2.220623	1.3097947	17	79	5.3%	9.21 [0.71 , 120.03]
Hutchings 2005	3.570366	1.6144415	22	63	3.7%	35.53 [1.50 , 841.02]
Hutchings 2014	3.231135	0.9819507	37	89	8.3%	25.31 [3.69 , 173.42]
Simon 2016	2.153725	0.6545228	32	89	13.9%	8.62 [2.39 , 31.08]
Subtotal (95% CI)			**108**	**320**	**31.3%**	**12.75 [4.98 , 32.68]**
Heterogeneity: Tau² = 0.00; Chi² = 1.31, df = 3 (P = 0.73); I² = 0%						
Test for overall effect: Z = 5.30 (P < 0.00001)						
2.1.3 not specified						
Cerci 2010	1.563053	1.274372	30	74	5.6%	4.77 [0.39 , 58.02]
Kobe 2018	0.9555	0.4724	236	486	18.6%	2.60 [1.03 , 6.56]
Subtotal (95% CI)			**266**	**560**	**24.2%**	**2.80 [1.17 , 6.67]**
Heterogeneity: Tau² = 0.00; Chi² = 0.20, df = 1 (P = 0.65); I² = 0%						
Test for overall effect: Z = 2.32 (P = 0.02)						
Total (95% CI)			**475**	**1327**	**100.0%**	**5.09 [2.64 , 9.81]**
Heterogeneity: Tau² = 0.38; Chi² = 14.39, df = 8 (P = 0.07); I² = 44%						
Test for overall effect: Z = 4.87 (P < 0.00001)						
Test for subgroup differences: Chi² = 6.01, df = 2 (P = 0.05), I² = 66.7%						

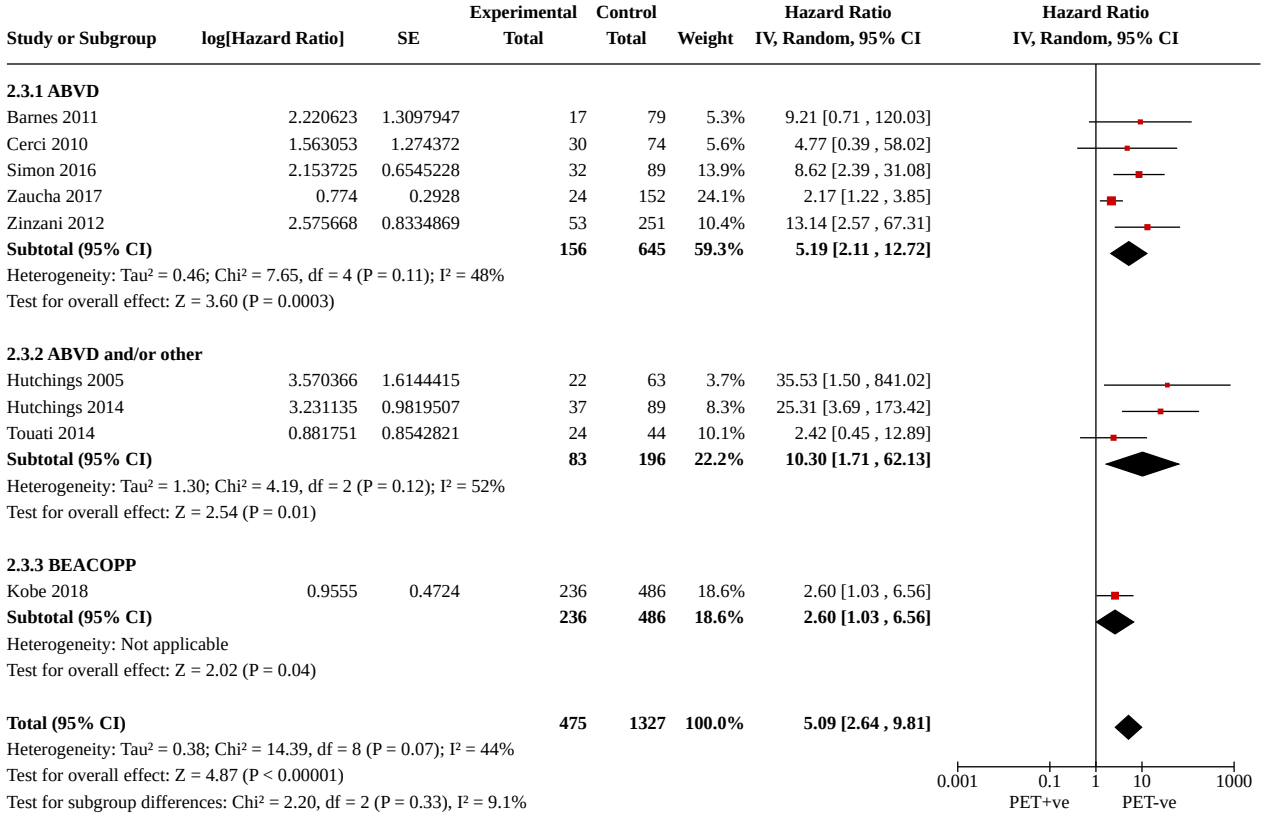


Analysis 2.2. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 2: OS by study design

Study or Subgroup	log[Hazard Ratio]	SE	Experimental Total	Control Total	Weight	Hazard Ratio IV, Random, 95% CI
2.2.1 Prospective						
Cerci 2010	1.563053	1.274372	30	74	5.5%	4.77 [0.39 , 58.02]
Hutchings 2014	3.231135	0.981951	37	89	8.3%	25.31 [3.69 , 173.42]
Zaucha 2017	0.774	0.2928	24	152	26.2%	2.17 [1.22 , 3.85]
Subtotal (95% CI)			**91**	**315**	**40.1%**	**5.35 [1.07 , 26.68]**
Heterogeneity: Tau² = 1.32; Chi² = 5.95, df = 2 (P = 0.05); I² = 66%						
Test for overall effect: Z = 2.04 (P = 0.04)						
2.2.2 Retrospective						
Barnes 2011	2.220623	1.309795	17	79	5.3%	9.21 [0.71 , 120.03]
Simon 2016	2.153725	0.654523	32	89	14.3%	8.62 [2.39 , 31.08]
Touati 2014	0.881751	0.854282	24	44	10.2%	2.42 [0.45 , 12.89]
Zinzani 2012	2.575668	0.833487	53	251	10.5%	13.14 [2.57 , 67.31]
Subtotal (95% CI)			**126**	**463**	**40.3%**	**7.12 [3.14 , 16.14]**
Heterogeneity: Tau² = 0.00; Chi² = 2.27, df = 3 (P = 0.52); I² = 0%						
Test for overall effect: Z = 4.70 (P < 0.00001)						
2.2.3 RCT						
Kobe 2018	0.9555	0.4724	236	486	19.6%	2.60 [1.03 , 6.56]
Subtotal (95% CI)			**236**	**486**	**19.6%**	**2.60 [1.03 , 6.56]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.02 (P = 0.04)						
Total (95% CI)			**453**	**1264**	**100.0%**	**4.63 [2.43 , 8.80]**
Heterogeneity: Tau² = 0.32; Chi² = 12.32, df = 7 (P = 0.09); I² = 43%						
Test for overall effect: Z = 4.67 (P < 0.00001)						
Test for subgroup differences: Chi² = 2.58, df = 2 (P = 0.28), I² = 22.5%						

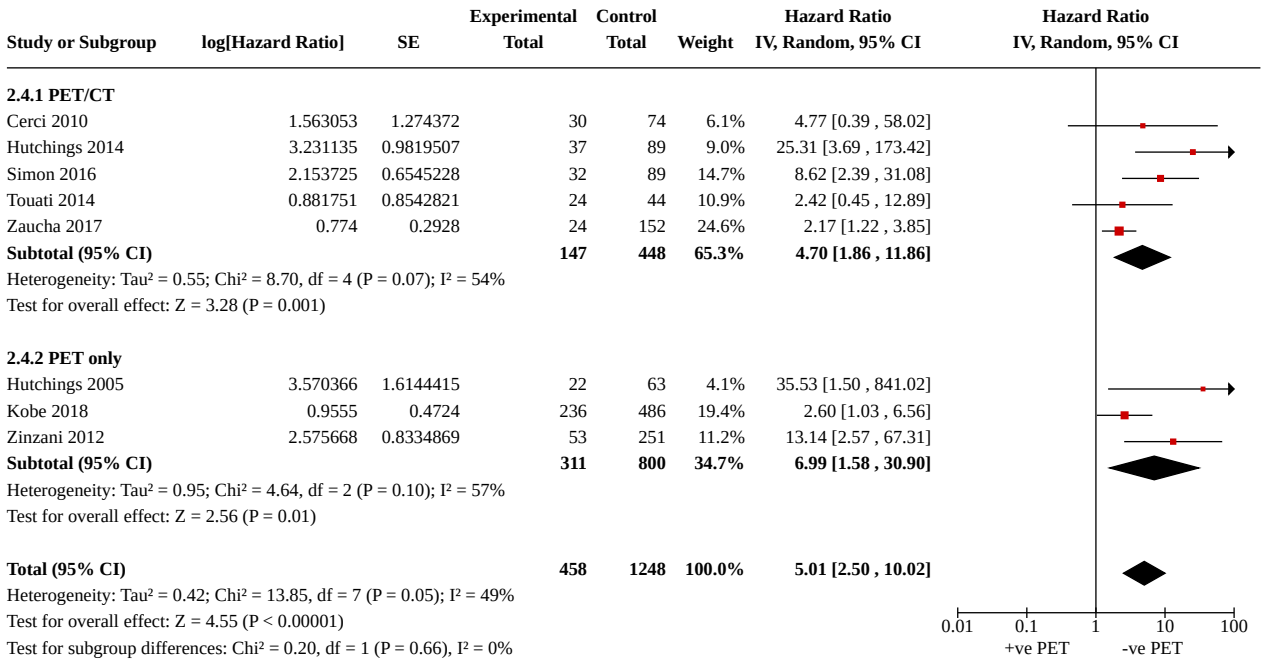


Analysis 2.3. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 3: OS by chemotherapy



Study or Subgroup	log[Hazard Ratio]	SE	Experimental Total	Control Total	Weight	Hazard Ratio IV, Random, 95% CI
2.3.1 ABVD						
Barnes 2011	2.220623	1.3097947	17	79	5.3%	9.21 [0.71 , 120.03]
Cerci 2010	1.563053	1.274372	30	74	5.6%	4.77 [0.39 , 58.02]
Simon 2016	2.153725	0.6545228	32	89	13.9%	8.62 [2.39 , 31.08]
Zaucha 2017	0.774	0.2928	24	152	24.1%	2.17 [1.22 , 3.85]
Zinzani 2012	2.575668	0.8334869	53	251	10.4%	13.14 [2.57 , 67.31]
Subtotal (95% CI)			**156**	**645**	**59.3%**	**5.19 [2.11 , 12.72]**
Heterogeneity: Tau² = 0.46; Chi² = 7.65, df = 4 (P = 0.11); I² = 48%						
Test for overall effect: Z = 3.60 (P = 0.0003)						
2.3.2 ABVD and/or other						
Hutchings 2005	3.570366	1.6144415	22	63	3.7%	35.53 [1.50 , 841.02]
Hutchings 2014	3.231135	0.9819507	37	89	8.3%	25.31 [3.69 , 173.42]
Touati 2014	0.881751	0.8542821	24	44	10.1%	2.42 [0.45 , 12.89]
Subtotal (95% CI)			**83**	**196**	**22.2%**	**10.30 [1.71 , 62.13]**
Heterogeneity: Tau² = 1.30; Chi² = 4.19, df = 2 (P = 0.12); I² = 52%						
Test for overall effect: Z = 2.54 (P = 0.01)						
2.3.3 BEACOPP						
Kobe 2018	0.9555	0.4724	236	486	18.6%	2.60 [1.03 , 6.56]
Subtotal (95% CI)			**236**	**486**	**18.6%**	**2.60 [1.03 , 6.56]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.02 (P = 0.04)						
Total (95% CI)			**475**	**1327**	**100.0%**	**5.09 [2.64 , 9.81]**
Heterogeneity: Tau² = 0.38; Chi² = 14.39, df = 8 (P = 0.07); I² = 44%						
Test for overall effect: Z = 4.87 (P < 0.00001)						
Test for subgroup differences: Chi² = 2.20, df = 2 (P = 0.33), I² = 9.1%						

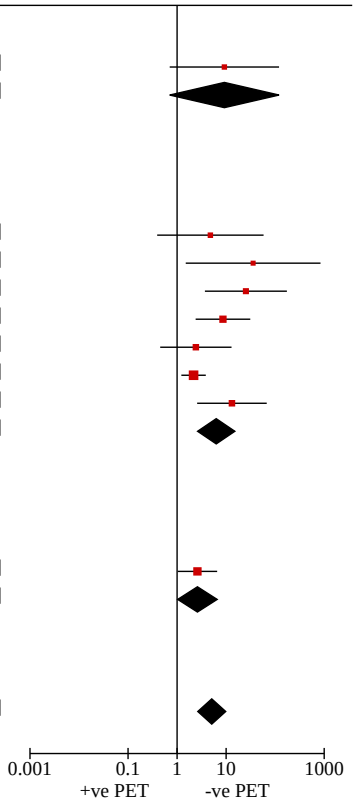
Analysis 2.4. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 4: OS for PET/CT vs PET



Study or Subgroup	log[Hazard Ratio]	SE	Experimental Total	Control Total	Weight	Hazard Ratio IV, Random, 95% CI
2.4.1 PET/CT						
Cerci 2010	1.563053	1.274372	30	74	6.1%	4.77 [0.39 , 58.02]
Hutchings 2014	3.231135	0.9819507	37	89	9.0%	25.31 [3.69 , 173.42]
Simon 2016	2.153725	0.6545228	32	89	14.7%	8.62 [2.39 , 31.08]
Touati 2014	0.881751	0.8542821	24	44	10.9%	2.42 [0.45 , 12.89]
Zaucha 2017	0.774	0.2928	24	152	24.6%	2.17 [1.22 , 3.85]
Subtotal (95% CI)			**147**	**448**	**65.3%**	**4.70 [1.86 , 11.86]**
Heterogeneity: Tau² = 0.55; Chi² = 8.70, df = 4 (P = 0.07); I² = 54%						
Test for overall effect: Z = 3.28 (P = 0.001)						
2.4.2 PET only						
Hutchings 2005	3.570366	1.6144415	22	63	4.1%	35.53 [1.50 , 841.02]
Kobe 2018	0.9555	0.4724	236	486	19.4%	2.60 [1.03 , 6.56]
Zinzani 2012	2.575668	0.8334869	53	251	11.2%	13.14 [2.57 , 67.31]
Subtotal (95% CI)			**311**	**800**	**34.7%**	**6.99 [1.58 , 30.90]**
Heterogeneity: Tau² = 0.95; Chi² = 4.64, df = 2 (P = 0.10); I² = 57%						
Test for overall effect: Z = 2.56 (P = 0.01)						
Total (95% CI)			**458**	**1248**	**100.0%**	**5.01 [2.50 , 10.02]**
Heterogeneity: Tau² = 0.42; Chi² = 13.85, df = 7 (P = 0.05); I² = 49%						
Test for overall effect: Z = 4.55 (P < 0.00001)						
Test for subgroup differences: Chi² = 0.20, df = 1 (P = 0.66), I² = 0%						

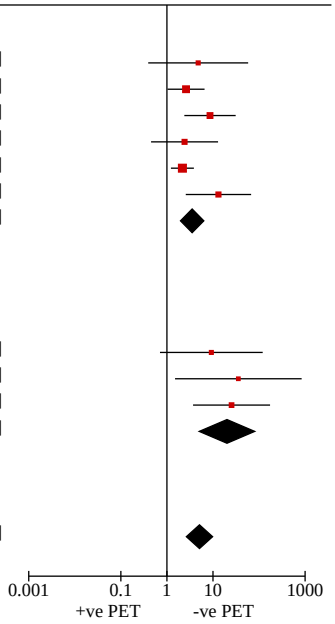
Analysis 2.5. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 5: OS by disease stage

Study or Subgroup	log[OR]	SE	Experimental Total	Control Total	Weight	Odds Ratio IV, Random, 95% CI
2.5.1 Stages I and II with A and B symptoms						
Barnes 2011	2.220623	1.3097947	17	79	5.3%	9.21 [0.71 , 120.03]
Subtotal (95% CI)			**17**	**79**	**5.3%**	**9.21 [0.71 , 120.03]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.70 (P = 0.09)						
2.5.2 All stages						
Cerci 2010	1.563053	1.274372	30	74	5.6%	4.77 [0.39 , 58.02]
Hutchings 2005	3.570366	1.6144415	22	63	3.7%	35.53 [1.50 , 841.02]
Hutchings 2014	3.231135	0.9819507	37	89	8.3%	25.31 [3.69 , 173.42]
Simon 2016	2.153725	0.6545228	32	89	13.9%	8.62 [2.39 , 31.08]
Touati 2014	0.881751	0.8542821	24	44	10.1%	2.42 [0.45 , 12.89]
Zaucha 2017	0.774	0.2928	24	152	24.1%	2.17 [1.22 , 3.85]
Zinzani 2012	2.575668	0.8334869	53	251	10.4%	13.14 [2.57 , 67.31]
Subtotal (95% CI)			**222**	**762**	**76.1%**	**6.28 [2.62 , 15.05]**
Heterogeneity: $Tau^2 = 0.67$; $Chi^2 = 13.40$, df = 6 (P = 0.04); $I^2 = 55\%$						
Test for overall effect: Z = 4.12 (P < 0.0001)						
2.5.3 Advanced						
Kobe 2018	0.9555	0.4724	236	486	18.6%	2.60 [1.03 , 6.56]
Subtotal (95% CI)			**236**	**486**	**18.6%**	**2.60 [1.03 , 6.56]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.02 (P = 0.04)						
Total (95% CI)			**475**	**1327**	**100.0%**	**5.09 [2.64 , 9.81]**
Heterogeneity: $Tau^2 = 0.38$; $Chi^2 = 14.39$, df = 8 (P = 0.07); $I^2 = 44\%$						
Test for overall effect: Z = 4.87 (P < 0.00001)						
Test for subgroup differences: $Chi^2 = 2.20$, df = 2 (P = 0.33), $I^2 = 8.9\%$						



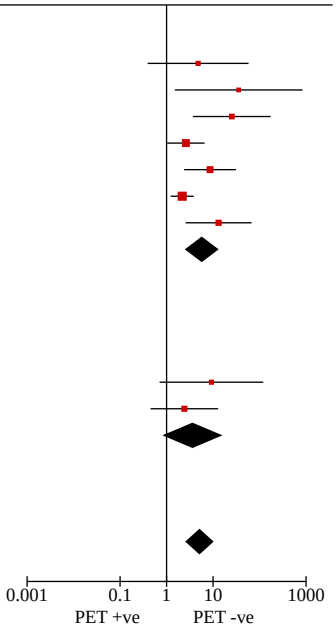
Analysis 2.6. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 6: Timing of interim PET

Study or Subgroup	log[Hazard Ratio]	SE	Experimental Total	Control Total	Weight	Hazard Ratio IV, Random, 95% CI
2.6.1 PET2						
Cerci 2010	1.563053	1.274372	30	74	5.6%	4.77 [0.39 , 58.02]
Kobe 2018	0.9555	0.4724	236	486	18.6%	2.60 [1.03 , 6.56]
Simon 2016	2.153725	0.6545228	32	89	13.9%	8.62 [2.39 , 31.08]
Touati 2014	0.881751	0.8542821	24	44	10.1%	2.42 [0.45 , 12.89]
Zaucha 2017	0.774	0.2928	24	152	24.1%	2.17 [1.22 , 3.85]
Zinzani 2012	2.575668	0.8334869	53	251	10.4%	13.14 [2.57 , 67.31]
Subtotal (95% CI)			**399**	**1096**	**82.6%**	**3.53 [1.97 , 6.32]**
Heterogeneity: Tau² = 0.16; Chi² = 7.26, df = 5 (P = 0.20); I² = 31%						
Test for overall effect: Z = 4.24 (P < 0.0001)						
2.6.2 Other (including mixed)						
Barnes 2011	2.220623	1.3097947	17	79	5.3%	9.21 [0.71 , 120.03]
Hutchings 2005	3.570366	1.6144415	22	63	3.7%	35.53 [1.50 , 841.02]
Hutchings 2014	3.231135	0.9819507	37	89	8.3%	25.31 [3.69 , 173.42]
Subtotal (95% CI)			**76**	**231**	**17.4%**	**20.13 [5.04 , 80.38]**
Heterogeneity: Tau² = 0.00; Chi² = 0.53, df = 2 (P = 0.77); I² = 0%						
Test for overall effect: Z = 4.25 (P < 0.0001)						
Total (95% CI)			**475**	**1327**	**100.0%**	**5.09 [2.64 , 9.81]**
Heterogeneity: Tau² = 0.38; Chi² = 14.39, df = 8 (P = 0.07); I² = 44%						
Test for overall effect: Z = 4.87 (P < 0.00001)						
Test for subgroup differences: Chi² = 5.16, df = 1 (P = 0.02), I² = 80.6%						



Analysis 2.7. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 7: OS by HR type of estimation

Study or Subgroup	log[Hazard Ratio]	SE	Experimental Total	Control Total	Weight	Hazard Ratio IV, Random, 95% CI
2.7.1 precise						
Cerci 2010	1.563053	1.274372	30	74	5.6%	4.77 [0.39 , 58.02]
Hutchings 2005	3.570366	1.614442	22	63	3.7%	35.53 [1.50 , 841.02]
Hutchings 2014	3.231135	0.981951	37	89	8.3%	25.31 [3.69 , 173.42]
Kobe 2018	0.9555	0.4724	236	486	18.6%	2.60 [1.03 , 6.56]
Simon 2016	2.153725	0.654523	32	89	13.9%	8.62 [2.39 , 31.08]
Zaucha 2017	0.774	0.2928	24	152	24.1%	2.17 [1.22 , 3.85]
Zinzani 2012	2.575668	0.833487	53	251	10.4%	13.14 [2.57 , 67.31]
Subtotal (95% CI)			**434**	**1204**	**84.6%**	**5.70 [2.60 , 12.48]**
Heterogeneity: Tau² = 0.53; Chi² = 13.66, df = 6 (P = 0.03); I² = 56%						
Test for overall effect: Z = 4.35 (P < 0.0001)						
2.7.2 Imprecise						
Barnes 2011	2.220623	1.309795	17	79	5.3%	9.21 [0.71 , 120.03]
Touati 2014	0.881751	0.854282	24	44	10.1%	2.42 [0.45 , 12.89]
Subtotal (95% CI)			**41**	**123**	**15.4%**	**3.60 [0.89 , 14.64]**
Heterogeneity: Tau² = 0.00; Chi² = 0.73, df = 1 (P = 0.39); I² = 0%						
Test for overall effect: Z = 1.79 (P = 0.07)						
Total (95% CI)			**475**	**1327**	**100.0%**	**5.09 [2.64 , 9.81]**
Heterogeneity: Tau² = 0.38; Chi² = 14.39, df = 8 (P = 0.07); I² = 44%						
Test for overall effect: Z = 4.87 (P < 0.00001)						
Test for subgroup differences: Chi² = 0.31, df = 1 (P = 0.58), I² = 0%						



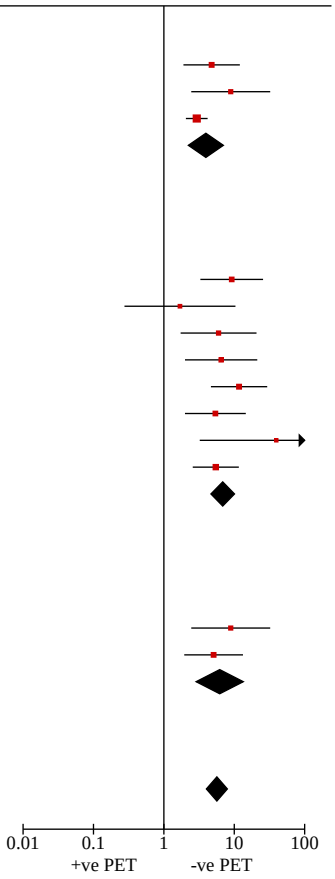
Comparison 3. Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
3.1 PFS by study design	13	1349	Hazard Ratio (IV, Random, 95% CI)	5.66 [4.02, 7.97]
3.1.1 prospective	3	357	Hazard Ratio (IV, Random, 95% CI)	3.95 [2.23, 7.00]
3.1.2 retrospective	8	827	Hazard Ratio (IV, Random, 95% CI)	6.85 [4.66, 10.08]
3.1.3 RCT	2	165	Hazard Ratio (IV, Random, 95% CI)	6.21 [2.87, 13.42]
3.2 PFS by chemotherapy	14	2079	Hazard Ratio (IV, Random, 95% CI)	4.90 [3.47, 6.90]
3.2.1 ABVD	7	945	Hazard Ratio (IV, Random, 95% CI)	5.13 [3.18, 8.27]
3.2.2 ABVD and/or other	4	265	Hazard Ratio (IV, Random, 95% CI)	7.07 [3.40, 14.70]
3.2.3 other NON-ABVD chemo	3	869	Hazard Ratio (IV, Random, 95% CI)	3.64 [1.83, 7.24]
3.3 PFS for PET/CT vs PET	13	1983	Hazard Ratio (IV, Random, 95% CI)	5.08 [3.57, 7.21]
3.3.1 PET/CT	8	707	Hazard Ratio (IV, Random, 95% CI)	6.03 [3.68, 9.90]
3.3.2 PET only	5	1276	Hazard Ratio (IV, Random, 95% CI)	4.06 [2.33, 7.08]
3.4 PFS by disease stage	14	2079	Hazard Ratio (IV, Random, 95% CI)	4.90 [3.47, 6.90]
3.4.1 Stages I and II with A and B symptoms	2	184	Hazard Ratio (IV, Random, 95% CI)	3.88 [1.54, 9.83]
3.4.2 All stages	11	1173	Hazard Ratio (IV, Random, 95% CI)	5.81 [3.93, 8.57]
3.4.3 Advanced	1	722	Hazard Ratio (IV, Random, 95% CI)	2.27 [1.35, 3.82]
3.5 PFS by radiotherapy	14	2079	Hazard Ratio (IV, Random, 95% CI)	4.90 [3.47, 6.90]
3.5.1 Involved node and/or site	5	651	Hazard Ratio (IV, Random, 95% CI)	5.35 [2.94, 9.75]
3.5.2 Involved field	6	514	Hazard Ratio (IV, Random, 95% CI)	7.06 [4.15, 12.00]
3.5.3 Not specified	2	826	Hazard Ratio (IV, Random, 95% CI)	2.97 [1.48, 5.98]
3.5.4 None	1	88	Hazard Ratio (IV, Random, 95% CI)	5.09 [1.95, 13.29]
3.6 Timing of interim PET	14	2079	Hazard Ratio (IV, Random, 95% CI)	4.90 [3.47, 6.90]
3.6.1 PET2	9	1677	Hazard Ratio (IV, Random, 95% CI)	4.68 [3.14, 6.98]
3.6.2 Other (including mixed)	5	402	Hazard Ratio (IV, Random, 95% CI)	6.32 [3.40, 11.75]
3.7 PFS by HR type of estimation	14	2079	Hazard Ratio (IV, Random, 95% CI)	4.90 [3.47, 6.90]

Outcome or subgroup title	No. of studies	No. of partici-pants	Statistical method	Effect size
3.7.1 precise	9	1450	Hazard Ratio (IV, Random, 95% CI)	4.69 [2.84, 7.73]
3.7.2 Imprecise	5	629	Hazard Ratio (IV, Random, 95% CI)	5.66 [3.65, 8.77]

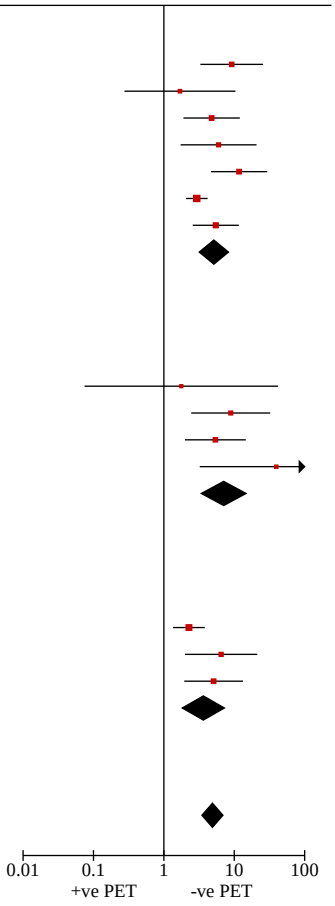
Analysis 3.1. Comparison 3: Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve, Outcome 1: PFS by study design

Study or Subgroup	log[Hazard Ratio]	SE	Experimental Total	Control Total	Weight	Hazard Ratio IV, Random, 95% CI
3.1.1 prospective						
Cerci 2010	1.5624	0.4706	30	74	8.8%	4.77 [1.90 , 12.00]
Hutchings 2006	2.187	0.6587	16	61	5.5%	8.91 [2.45 , 32.40]
Zaucha 2017	1.0753	0.1812	24	152	19.3%	2.93 [2.05 , 4.18]
Subtotal (95% CI)			**70**	**287**	**33.6%**	**3.95 [2.23 , 7.00]**
Heterogeneity: Tau² = 0.11; Chi² = 3.31, df = 2 (P = 0.19); I² = 40%						
Test for overall effect: Z = 4.70 (P < 0.00001)						
3.1.2 retrospective						
Annunziata 2016	2.2192	0.5231	12	56	7.7%	9.20 [3.30 , 25.65]
Barnes 2011	0.5261	0.9261	17	79	3.1%	1.69 [0.28 , 10.39]
Mesguich 2016	1.7891	0.6333	16	60	5.8%	5.98 [1.73 , 20.70]
Rossi 2014	1.873	0.6031	13	46	6.2%	6.51 [2.00 , 21.22]
Simon 2016	2.4596	0.4697	32	89	8.8%	11.70 [4.66 , 29.38]
Touati 2014	1.685	0.5075	24	44	8.0%	5.39 [1.99 , 14.58]
Ying 2014	3.6783	1.278	10	25	1.7%	39.58 [3.23 , 484.51]
Zinzani 2012	1.6982	0.3845	53	251	11.2%	5.46 [2.57 , 11.61]
Subtotal (95% CI)			**177**	**650**	**52.6%**	**6.85 [4.66 , 10.08]**
Heterogeneity: Tau² = 0.00; Chi² = 6.40, df = 7 (P = 0.49); I² = 0%						
Test for overall effect: Z = 9.79 (P < 0.00001)						
3.1.3 RCT						
Kobe 2018	2.187	0.6587	16	61	5.5%	8.91 [2.45 , 32.40]
Straus 2011	1.6268	0.49	24	64	8.4%	5.09 [1.95 , 13.29]
Subtotal (95% CI)			**40**	**125**	**13.8%**	**6.21 [2.87 , 13.42]**
Heterogeneity: Tau² = 0.00; Chi² = 0.47, df = 1 (P = 0.50); I² = 0%						
Test for overall effect: Z = 4.65 (P < 0.00001)						
Total (95% CI)			**287**	**1062**	**100.0%**	**5.66 [4.02 , 7.97]**
Heterogeneity: Tau² = 0.13; Chi² = 18.76, df = 12 (P = 0.09); I² = 36%						
Test for overall effect: Z = 9.92 (P < 0.00001)						
Test for subgroup differences: Chi² = 2.48, df = 2 (P = 0.29), I² = 19.5%						



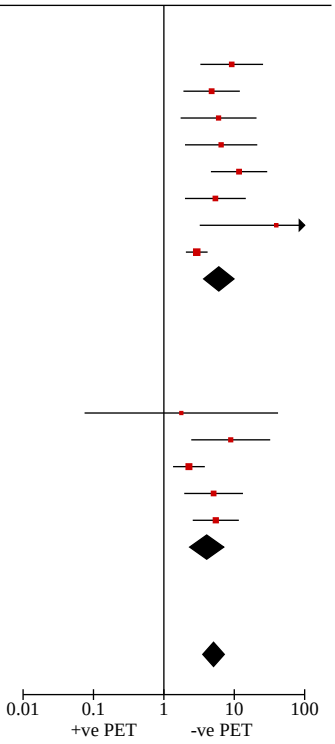
Analysis 3.2. Comparison 3: Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve, Outcome 2: PFS by chemotherapy

Study or Subgroup	log[Hazard Ratio]	SE	Experimental Total	Control Total	Weight	Hazard Ratio IV, Random, 95% CI
3.2.1 ABVD						
Annunziata 2016	2.2192	0.5231	12	56	7.1%	9.20 [3.30 , 25.65]
Barnes 2011	0.5261	0.9261	17	79	3.0%	1.69 [0.28 , 10.39]
Cerci 2010	1.5624	0.4706	30	74	8.1%	4.77 [1.90 , 12.00]
Mesguich 2016	1.7891	0.6333	16	60	5.5%	5.98 [1.73 , 20.70]
Simon 2016	2.4596	0.4697	32	89	8.1%	11.70 [4.66 , 29.38]
Zaucha 2017	1.0753	0.1812	24	152	16.0%	2.93 [2.05 , 4.18]
Zinzani 2012	1.6982	0.3845	53	251	10.0%	5.46 [2.57 , 11.61]
Subtotal (95% CI)			**184**	**761**	**57.7%**	**5.13 [3.18 , 8.27]**
Heterogeneity: Tau² = 0.20; Chi² = 12.73, df = 6 (P = 0.05); I² = 53%						
Test for overall effect: Z = 6.70 (P < 0.00001)						
3.2.2 ABVD and/or other						
Hutchings 2005	0.570366	1.6144	22	63	1.1%	1.77 [0.07 , 41.87]
Hutchings 2006	2.187	0.6587	16	61	5.2%	8.91 [2.45 , 32.40]
Touati 2014	1.685	0.5075	24	44	7.4%	5.39 [1.99 , 14.58]
Ying 2014	3.6783	1.278	10	25	1.7%	39.58 [3.23 , 484.51]
Subtotal (95% CI)			**72**	**193**	**15.4%**	**7.07 [3.40 , 14.70]**
Heterogeneity: Tau² = 0.00; Chi² = 2.96, df = 3 (P = 0.40); I² = 0%						
Test for overall effect: Z = 5.24 (P < 0.00001)						
3.2.3 other NON-ABVD chemo						
Kobe 2018	0.8198	0.2651	236	486	13.4%	2.27 [1.35 , 3.82]
Rossi 2014	1.873	0.6031	13	46	5.9%	6.51 [2.00 , 21.22]
Straus 2011	1.6268	0.49	24	64	7.7%	5.09 [1.95 , 13.29]
Subtotal (95% CI)			**273**	**596**	**26.9%**	**3.64 [1.83 , 7.24]**
Heterogeneity: Tau² = 0.18; Chi² = 3.91, df = 2 (P = 0.14); I² = 49%						
Test for overall effect: Z = 3.68 (P = 0.0002)						
Total (95% CI)			**529**	**1550**	**100.0%**	**4.90 [3.47 , 6.90]**
Heterogeneity: Tau² = 0.16; Chi² = 23.52, df = 13 (P = 0.04); I² = 45%						
Test for overall effect: Z = 9.07 (P < 0.00001)						
Test for subgroup differences: Chi² = 1.69, df = 2 (P = 0.43), I² = 0%						



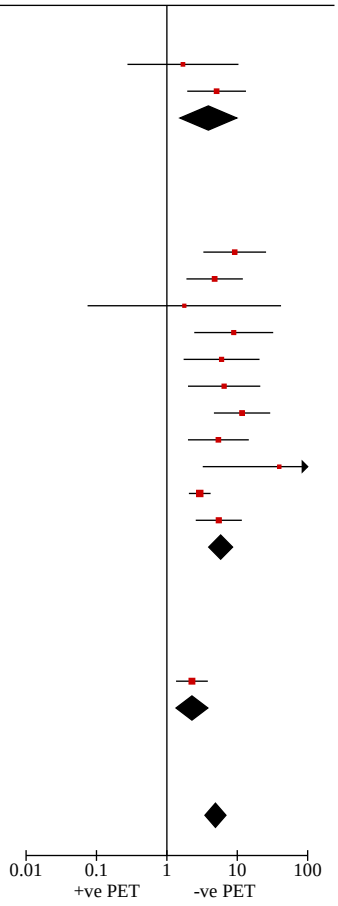
Analysis 3.3. Comparison 3: Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve, Outcome 3: PFS for PET/CT vs PET

Study or Subgroup	log[Hazard Ratio]	SE	Experimental Total	Control Total	Weight	Hazard Ratio IV, Random, 95% CI
3.3.1 PET/CT						
Annunziata 2016	2.2192	0.5231	12	56	7.3%	9.20 [3.30 , 25.65]
Cerci 2010	1.5624	0.4706	30	74	8.3%	4.77 [1.90 , 12.00]
Mesguich 2016	1.7891	0.6333	16	60	5.7%	5.98 [1.73 , 20.70]
Rossi 2014	1.873	0.6031	13	46	6.1%	6.51 [2.00 , 21.22]
Simon 2016	2.4596	0.4697	32	89	8.4%	11.70 [4.66 , 29.38]
Touati 2014	1.685	0.5075	24	44	7.6%	5.39 [1.99 , 14.58]
Ying 2014	3.6783	1.278	10	25	1.8%	39.58 [3.23 , 484.51]
Zaucha 2017	1.0753	0.1812	24	152	16.3%	2.93 [2.05 , 4.18]
Subtotal (95% CI)			**161**	**546**	**61.5%**	**6.03 [3.68 , 9.90]**
Heterogeneity: Tau² = 0.24; Chi² = 15.14, df = 7 (P = 0.03); I² = 54%						
Test for overall effect: Z = 7.12 (P < 0.00001)						
3.3.2 PET only						
Hutchings 2005	0.570366	1.6144	22	63	1.2%	1.77 [0.07 , 41.87]
Hutchings 2006	2.187	0.6587	16	61	5.4%	8.91 [2.45 , 32.40]
Kobe 2018	0.8198	0.2651	236	486	13.7%	2.27 [1.35 , 3.82]
Straus 2011	1.6268	0.49	24	64	7.9%	5.09 [1.95 , 13.29]
Zinzani 2012	1.6982	0.3845	53	251	10.3%	5.46 [2.57 , 11.61]
Subtotal (95% CI)			**351**	**925**	**38.5%**	**4.06 [2.33 , 7.08]**
Heterogeneity: Tau² = 0.15; Chi² = 6.78, df = 4 (P = 0.15); I² = 41%						
Test for overall effect: Z = 4.93 (P < 0.00001)						
Total (95% CI)			**512**	**1471**	**100.0%**	**5.08 [3.57 , 7.21]**
Heterogeneity: Tau² = 0.16; Chi² = 22.65, df = 12 (P = 0.03); I² = 47%						
Test for overall effect: Z = 9.06 (P < 0.00001)						
Test for subgroup differences: Chi² = 1.09, df = 1 (P = 0.30), I² = 8.0%						



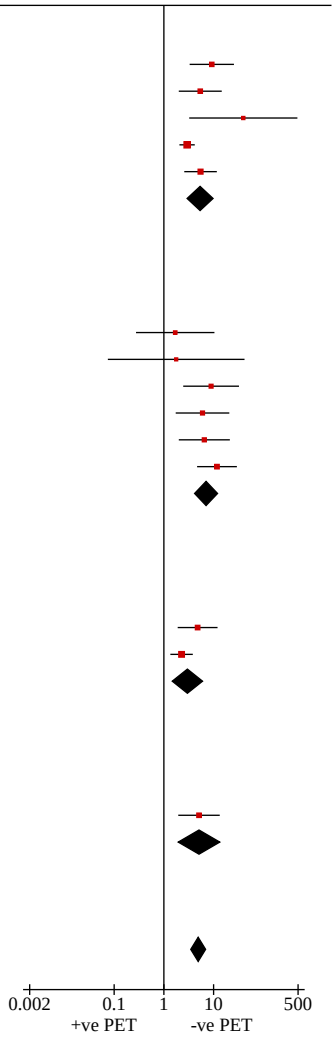
Analysis 3.4. Comparison 3: Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve, Outcome 4: PFS by disease stage

Study or Subgroup	log[Hazard Ratio]	SE	Experimental Total	Control Total	Weight	Hazard Ratio IV, Random, 95% CI
3.4.1 Stages I and II with A and B symptoms						
Barnes 2011	0.5261	0.9261	17	79	3.0%	1.69 [0.28 , 10.39]
Straus 2011	1.6268	0.49	24	64	7.7%	5.09 [1.95 , 13.29]
Subtotal (95% CI)			**41**	**143**	**10.7%**	**3.88 [1.54 , 9.83]**
Heterogeneity: Tau² = 0.06; Chi² = 1.10, df = 1 (P = 0.29); I² = 9%						
Test for overall effect: Z = 2.87 (P = 0.004)						
3.4.2 All stages						
Annunziata 2016	2.2192	0.5231	12	56	7.1%	9.20 [3.30 , 25.65]
Cerci 2010	1.5624	0.4706	30	74	8.1%	4.77 [1.90 , 12.00]
Hutchings 2005	0.570366	1.6144	22	63	1.1%	1.77 [0.07 , 41.87]
Hutchings 2006	2.187	0.6587	16	61	5.2%	8.91 [2.45 , 32.40]
Mesguich 2016	1.7891	0.6333	16	60	5.5%	5.98 [1.73 , 20.70]
Rossi 2014	1.873	0.6031	13	46	5.9%	6.51 [2.00 , 21.22]
Simon 2016	2.4596	0.4697	32	89	8.1%	11.70 [4.66 , 29.38]
Touati 2014	1.685	0.5075	24	44	7.4%	5.39 [1.99 , 14.58]
Ying 2014	3.6783	1.278	10	25	1.7%	39.58 [3.23 , 484.51]
Zaucha 2017	1.0753	0.1812	24	152	16.0%	2.93 [2.05 , 4.18]
Zinzani 2012	1.6982	0.3845	53	251	10.0%	5.46 [2.57 , 11.61]
Subtotal (95% CI)			**252**	**921**	**75.9%**	**5.81 [3.93 , 8.57]**
Heterogeneity: Tau² = 0.15; Chi² = 16.88, df = 10 (P = 0.08); I² = 41%						
Test for overall effect: Z = 8.85 (P < 0.00001)						
3.4.3 Advanced						
Kobe 2018	0.8198	0.2651	236	486	13.4%	2.27 [1.35 , 3.82]
Subtotal (95% CI)			**236**	**486**	**13.4%**	**2.27 [1.35 , 3.82]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 3.09 (P = 0.002)						
Total (95% CI)			**529**	**1550**	**100.0%**	**4.90 [3.47 , 6.90]**
Heterogeneity: Tau² = 0.16; Chi² = 23.52, df = 13 (P = 0.04); I² = 45%						
Test for overall effect: Z = 9.07 (P < 0.00001)						
Test for subgroup differences: Chi² = 8.05, df = 2 (P = 0.02), I² = 75.2%						



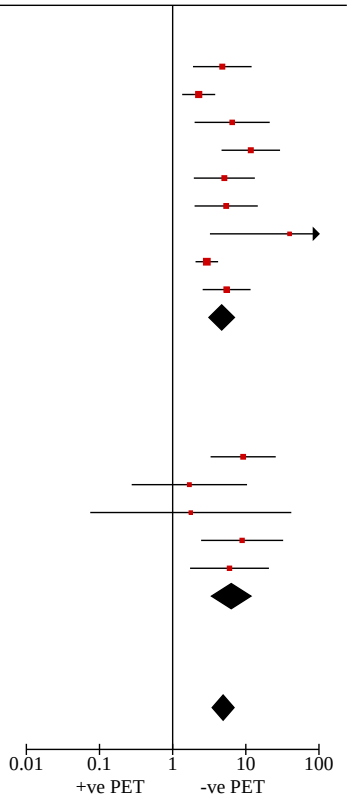
Analysis 3.5. Comparison 3: Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve, Outcome 5: PFS by radiotherapy

Study or Subgroup	log[Hazard Ratio]	SE	Experimental Total	Control Total	Weight	Hazard Ratio IV, Random, 95% CI
3.5.1 Involved node and/or site						
Annunziata 2016	2.2192	0.5231	12	56	7.1%	9.20 [3.30 , 25.65]
Touati 2014	1.685	0.5075	24	44	7.4%	5.39 [1.99 , 14.58]
Ying 2014	3.6783	1.278	10	25	1.7%	39.58 [3.23 , 484.51]
Zaucha 2017	1.0753	0.1812	24	152	16.0%	2.93 [2.05 , 4.18]
Zinzani 2012	1.6982	0.3845	53	251	10.0%	5.46 [2.57 , 11.61]
Subtotal (95% CI)			**123**	**528**	**42.2%**	**5.35 [2.94 , 9.75]**
Heterogeneity: Tau² = 0.24; Chi² = 9.63, df = 4 (P = 0.05); I² = 58%						
Test for overall effect: Z = 5.48 (P < 0.00001)						
3.5.2 Involved field						
Barnes 2011	0.5261	0.9261	17	79	3.0%	1.69 [0.28 , 10.39]
Hutchings 2005	0.570366	1.6144	22	63	1.1%	1.77 [0.07 , 41.87]
Hutchings 2006	2.187	0.6587	16	61	5.2%	8.91 [2.45 , 32.40]
Mesguich 2016	1.7891	0.6333	16	60	5.5%	5.98 [1.73 , 20.70]
Rossi 2014	1.873	0.6031	13	46	5.9%	6.51 [2.00 , 21.22]
Simon 2016	2.4596	0.4697	32	89	8.1%	11.70 [4.66 , 29.38]
Subtotal (95% CI)			**116**	**398**	**28.7%**	**7.06 [4.15 , 12.00]**
Heterogeneity: Tau² = 0.00; Chi² = 4.48, df = 5 (P = 0.48); I² = 0%						
Test for overall effect: Z = 7.21 (P < 0.00001)						
3.5.3 Not specified						
Cerci 2010	1.5624	0.4706	30	74	8.1%	4.77 [1.90 , 12.00]
Kobe 2018	0.8198	0.2651	236	486	13.4%	2.27 [1.35 , 3.82]
Subtotal (95% CI)			**266**	**560**	**21.4%**	**2.97 [1.48 , 5.98]**
Heterogeneity: Tau² = 0.13; Chi² = 1.89, df = 1 (P = 0.17); I² = 47%						
Test for overall effect: Z = 3.05 (P = 0.002)						
3.5.4 None						
Straus 2011	1.6268	0.49	24	64	7.7%	5.09 [1.95 , 13.29]
Subtotal (95% CI)			**24**	**64**	**7.7%**	**5.09 [1.95 , 13.29]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 3.32 (P = 0.0009)						
Total (95% CI)			**529**	**1550**	**100.0%**	**4.90 [3.47 , 6.90]**
Heterogeneity: Tau² = 0.16; Chi² = 23.52, df = 13 (P = 0.04); I² = 45%						
Test for overall effect: Z = 9.07 (P < 0.00001)						
Test for subgroup differences: Chi² = 3.73, df = 3 (P = 0.29), I² = 19.7%						

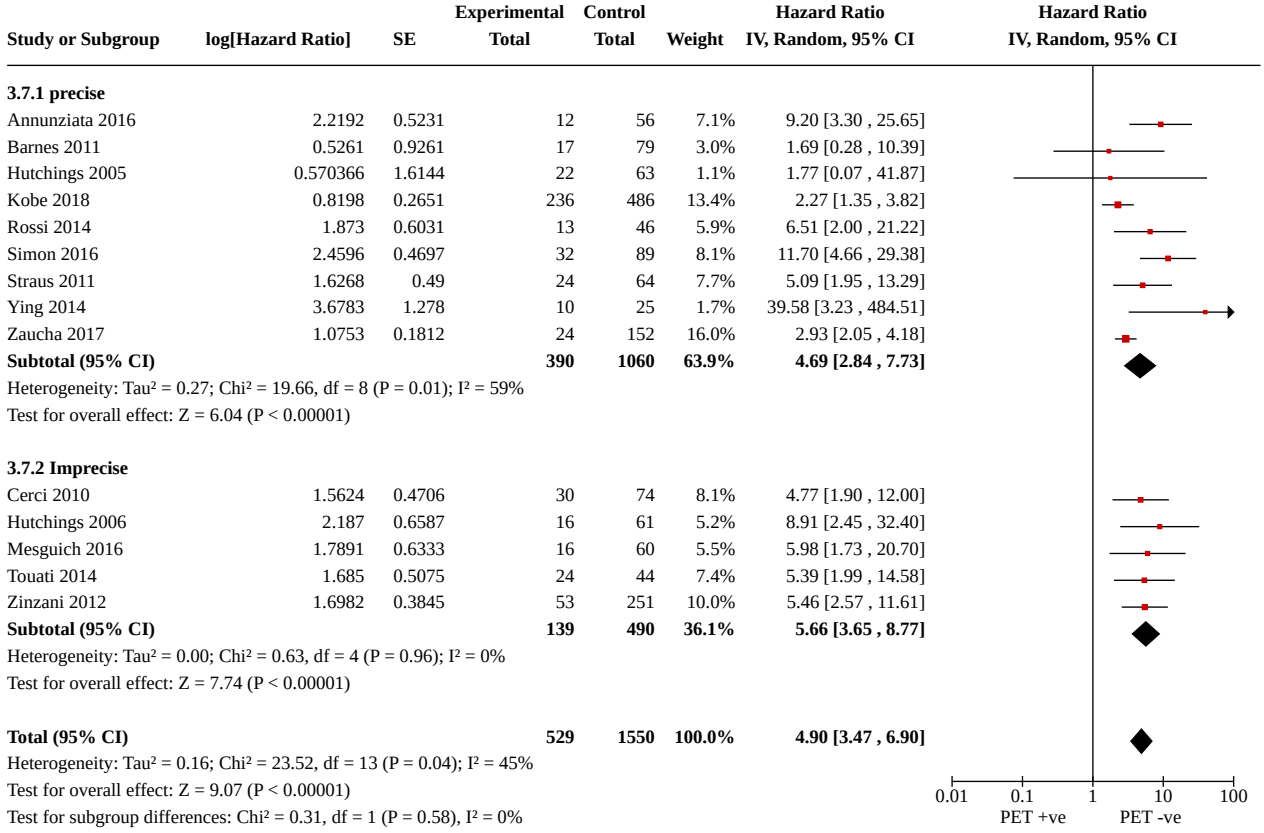


Analysis 3.6. Comparison 3: Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve, Outcome 6: Timing of interim PET

Study or Subgroup	log[Hazard Ratio]	SE	PET +ve Total	PET -ve Total	Weight	Hazard Ratio IV, Random, 95% CI
3.6.1 PET2						
Cerci 2010	1.5624	0.4706	30	74	8.1%	4.77 [1.90 , 12.00]
Kobe 2018	0.8198	0.2651	236	486	13.4%	2.27 [1.35 , 3.82]
Rossi 2014	1.873	0.6031	13	46	5.9%	6.51 [2.00 , 21.22]
Simon 2016	2.4596	0.4697	32	89	8.1%	11.70 [4.66 , 29.38]
Straus 2011	1.6268	0.49	24	64	7.7%	5.09 [1.95 , 13.29]
Touati 2014	1.685	0.5075	24	44	7.4%	5.39 [1.99 , 14.58]
Ying 2014	3.6783	1.278	10	25	1.7%	39.58 [3.23 , 484.51]
Zaucha 2017	1.0753	0.1812	24	152	16.0%	2.93 [2.05 , 4.18]
Zinzani 2012	1.6982	0.3845	53	251	10.0%	5.46 [2.57 , 11.61]
Subtotal (95% CI)			**446**	**1231**	**78.1%**	**4.68 [3.14 , 6.98]**
Heterogeneity: Tau² = 0.18; Chi² = 17.66, df = 8 (P = 0.02); I² = 55%						
Test for overall effect: Z = 7.56 (P < 0.00001)						
3.6.2 Other (including mixed)						
Annunziata 2016	2.2192	0.5231	12	56	7.1%	9.20 [3.30 , 25.65]
Barnes 2011	0.5261	0.9261	17	79	3.0%	1.69 [0.28 , 10.39]
Hutchings 2005	0.570366	1.6144	22	63	1.1%	1.77 [0.07 , 41.87]
Hutchings 2006	2.187	0.6587	16	61	5.2%	8.91 [2.45 , 32.40]
Mesguich 2016	1.7891	0.6333	16	60	5.5%	5.98 [1.73 , 20.70]
Subtotal (95% CI)			**83**	**319**	**21.9%**	**6.32 [3.40 , 11.75]**
Heterogeneity: Tau² = 0.00; Chi² = 3.44, df = 4 (P = 0.49); I² = 0%						
Test for overall effect: Z = 5.83 (P < 0.00001)						
Total (95% CI)			**529**	**1550**	**100.0%**	**4.90 [3.47 , 6.90]**
Heterogeneity: Tau² = 0.16; Chi² = 23.52, df = 13 (P = 0.04); I² = 45%						
Test for overall effect: Z = 9.07 (P < 0.00001)						
Test for subgroup differences: Chi² = 0.64, df = 1 (P = 0.42), I² = 0%						



Analysis 3.7. Comparison 3: Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve, Outcome 7: PFS by HR type of estimation



Study or Subgroup	log[Hazard Ratio]	SE	Experimental Total	Control Total	Weight	Hazard Ratio IV, Random, 95% CI
3.7.1 precise						
Annunziata 2016	2.2192	0.5231	12	56	7.1%	9.20 [3.30 , 25.65]
Barnes 2011	0.5261	0.9261	17	79	3.0%	1.69 [0.28 , 10.39]
Hutchings 2005	0.570366	1.6144	22	63	1.1%	1.77 [0.07 , 41.87]
Kobe 2018	0.8198	0.2651	236	486	13.4%	2.27 [1.35 , 3.82]
Rossi 2014	1.873	0.6031	13	46	5.9%	6.51 [2.00 , 21.22]
Simon 2016	2.4596	0.4697	32	89	8.1%	11.70 [4.66 , 29.38]
Straus 2011	1.6268	0.49	24	64	7.7%	5.09 [1.95 , 13.29]
Ying 2014	3.6783	1.278	10	25	1.7%	39.58 [3.23 , 484.51]
Zaucha 2017	1.0753	0.1812	24	152	16.0%	2.93 [2.05 , 4.18]
Subtotal (95% CI)			**390**	**1060**	**63.9%**	**4.69 [2.84 , 7.73]**
Heterogeneity: Tau² = 0.27; Chi² = 19.66, df = 8 (P = 0.01); I² = 59%						
Test for overall effect: Z = 6.04 (P < 0.00001)						
3.7.2 Imprecise						
Cerci 2010	1.5624	0.4706	30	74	8.1%	4.77 [1.90 , 12.00]
Hutchings 2006	2.187	0.6587	16	61	5.2%	8.91 [2.45 , 32.40]
Mesguich 2016	1.7891	0.6333	16	60	5.5%	5.98 [1.73 , 20.70]
Touati 2014	1.685	0.5075	24	44	7.4%	5.39 [1.99 , 14.58]
Zinzani 2012	1.6982	0.3845	53	251	10.0%	5.46 [2.57 , 11.61]
Subtotal (95% CI)			**139**	**490**	**36.1%**	**5.66 [3.65 , 8.77]**
Heterogeneity: Tau² = 0.00; Chi² = 0.63, df = 4 (P = 0.96); I² = 0%						
Test for overall effect: Z = 7.74 (P < 0.00001)						
Total (95% CI)			**529**	**1550**	**100.0%**	**4.90 [3.47 , 6.90]**
Heterogeneity: Tau² = 0.16; Chi² = 23.52, df = 13 (P = 0.04); I² = 45%						
Test for overall effect: Z = 9.07 (P < 0.00001)						
Test for subgroup differences: Chi² = 0.31, df = 1 (P = 0.58), I² = 0%						

APPENDICES

Appendix 1. Cochrane Central Register of Controlled Trials search strategy

Searches until 07/02/2016

ID Search

#1 MeSH descriptor: [Lymphoma] this term only

#2 MeSH descriptor: [Hodgkin Disease] explode all trees

#3 Germinoblastom*

#4 Reticulolymphosarcom*

#5 Hodgkin*

#6 (malignan* near/2 (lymphogranulom* or granulom*))

#7 #1 or #2 or #3 or #4 or #5 or #6

#8 MeSH descriptor: [Positron-Emission Tomography] explode all trees

#9 (pet* or petscan*)

#10 tomograph*

#11 emission*

#12 MeSH descriptor: [Deoxyglucose] explode all trees

#13 MeSH descriptor: [Fluorodeoxyglucose F18] explode all trees

#14 (deoxyglucose* or desoxyglucose* or deoxy-glucose* or desoxy-glucose* or deoxy-d-glucose* or desoxy-d-glucose* or 2deoxyglucose* or 2deoxy-d-glucose* or fluorodeoxyglucose* or fluorodesoxyglucose* or fludeoxyglucose* or fluordeoxyglucose* or fluordesoxyglucose* or 18fluorodeoxyglucose* or 18fluorodesoxyglucose* or 18fluordeoxyglucose* or fdg* or 18fdg* or 18f-dg*)

#15 (fluor* or 2fluor* or fluoro* or fluorodeoxy* or fludeoxy* or fluorine* or 18f* or 18flu*)

#16 glucose*

#17 #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16

#18 #7 and #17 in trials

Searches from 08/02/2016 - 13/07/2017

ID Search

#1 MeSH descriptor: [Lymphoma] this term only

#2 MeSH descriptor: [Hodgkin Disease] explode all trees

#3 Germinoblastom*

#4 Reticulolymphosarcom*

#5 Hodgkin*

#6 (malignan* near/2 (lymphogranulom* or granulom*))

#7 #1 or #2 or #3 or #4 or #5 or #6

#8 MeSH descriptor: [Positron-Emission Tomography] explode all trees

#9 (pet* or petscan*)

#10 tomograph*

#11 emission*

#12 MeSH descriptor: [Deoxyglucose] explode all trees

#13 MeSH descriptor: [Fluorodeoxyglucose F18] explode all trees

#14 (deoxyglucose* or desoxyglucose* or deoxy-glucose* or desoxy-glucose* or deoxy-d-glucose* or desoxy-d-glucose* or 2deoxyglucose* or 2deoxy-d-glucose* or fluorodeoxyglucose* or fluorodesoxyglucose* or fludeoxyglucose* or fluordeoxyglucose* or fluordesoxyglucose* or 18fluorodeoxyglucose* or 18fluorodesoxyglucose* or 18fluordeoxyglucose* or fdg* or 18fdg* or 18f-dg*)

#15 (fluor* or 2fluor* or fluoro* or fluorodeoxy* or fludeoxy* or fluorine* or 18f* or 18flu*)

#16 glucose*

#17 #15 and #16

#18 #8 or #9 or #10 or #11 or #12 or #13 or #14 or #17

#19 #7 and #18 in Trials

#20 #19 Publication Year from 2016 to 2017

Searches from 12/07/2017 - 12/11/2018

Cochrane Central Register of Controlled Trials (Cochrane Library Issue 11, 2018)

ID Search

#1 MeSH descriptor: [Lymphoma] this term only

#2 MeSH descriptor: [Hodgkin Disease] this term only

#3 germinoblastom*

#4 reticulolymphosarcom*

#5 hodgkin*

#6 (malignan* near/2 (lymphogranulom* or granulom*))

#7 #1 or #2 or #3 or #4 or #5 or #6

#8 MeSH descriptor: [Positron-Emission Tomography] explode all trees

#9 (pet*)

#10 tomograph*

#11 emission*

#12 MeSH descriptor: [Deoxyglucose] explode all trees

#13 MeSH descriptor: [Fluorodeoxyglucose F18] this term only

#14 (deoxyglucose* or desoxyglucose* or deoxy-glucose* or desoxy-glucose* or deoxy-d-glucose* or desoxy-d-glucose* or 2deoxyglucose* or 2deoxy-d-glucose* or fluorodeoxyglucose* or fluorodesoxyglucose* or fludeoxyglucose* or fluordeoxyglucose* or fluordesoxyglucose* or 18fluorodeoxyglucose* or 18fluorodesoxyglucose* or 18fluordeoxyglucose* or fdg* or 18fdg* or 18f-dg*)

#15 (glucose* and (fluor* or 2fluor* or fludeoxy* or 18f*))

#16 #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15

#17 #7and #16 in Trials

key: *: truncation, near/#: adjacent within # number of words

Searches from 12/11/2018 - 02/04/2019

ID Search

#1 MeSH descriptor: [Lymphoma] this term only

#2 MeSH descriptor: [Hodgkin Disease] explode all trees

#3 Germinoblastom*

#4 Reticulolymphosarcom*

#5 Hodgkin*

#6 (malignan* near/2 (lymphogranulom* or granulom*))

#7 #1 or #2 or #3 or #4 or #5 or #6

#8 MeSH descriptor: [Positron-Emission Tomography] explode all trees

#9 (pet* or petscan*)

#10 tomograph*

#11 emission*

#12 MeSH descriptor: [Deoxyglucose] explode all trees

#13 MeSH descriptor: [Fluorodeoxyglucose F18] explode all trees

#14 (deoxyglucose* or desoxyglucose* or deoxy-glucose* or desoxy-glucose* or deoxy-d-glucose* or desoxy-d-glucose* or 2deoxyglucose* or 2deoxy-d-glucose* or fluorodeoxyglucose* or fluorodesoxyglucose* or fludeoxyglucose* or fluordeoxyglucose* or fluordesoxyglucose* or 18fluorodeoxyglucose* or 18fluorodesoxyglucose* or 18fluordeoxyglucose* or fdg* or 18fdg* or 18f-dg*)

#15 (fluor* or 2fluor* or fluoro* or fluorodeoxy* or fludeoxy* or fluorine* or 18f* or 18flu*)

#16 glucose*

#17 #15 and #16

#18 #8 or #9 or #10 or #11 or #12 or #13 or #14 or #17

#19 #7 and #18 in Trials

Appendix 2. MEDLINE Ovid search strategy

#	Searches until 02/02/2016
1	*LYMPHOMA/
2	exp HODGKIN DISEASE/
3	Germinoblastom$.tw,kf,ot.
4	Reticulolymphosarcom$.tw,kf,ot.
5	Hodgkin$.tw,kf,ot.
6	(malignan$ adj2 (lymphogranulom$ or granulom$)).tw,kf,ot.
7	or/1-6
8	POSITRON-EMISSION TOMOGRAPHY/
9	(pet$ or petscan$).tw,kf,ot.
10	tomograph$.tw,kf,ot.
11	emission$.tw,kf,ot.
12	exp DEOXYGLUCOSE/
13	FLUORODEOXYGLUCOSE F18/
14	(deoxyglucose$ or desoxyglucose$ or deoxy-glucose$ or desoxy-glucose$ or deoxy-d-glucose$ or desoxy-d-glucose$ or 2deoxyglucose$ or 2deoxy-d-glucose$ or fluorodeoxyglucose$ or fluorodesoxyglucose$ or fludeoxyglucose$ or fluordeoxyglucose$ or fluordesoxyglucose$ or 18fluorodeoxyglucose$ or 18fluorodesoxyglucose$ or 18fluordeoxyglucose$ or fdg$ or 18fdg$ or 18f-dg$).tw.
15	(fluor$ or 2fluor$ or fluoro$ or fluorodeoxy$ or fludeoxy$ or fluorine$ or 18f$ or 18flu$).tw.
16	glucose$.tw.
17	or/8-16
18	7 and 17
19	ANIMALS/ not HUMANS/
20	18 not 19

#	Searches from 03/02/2016 – 12/07/2017
1	*LYMPHOMA/
2	exp HODGKIN DISEASE/
3	Germinoblastom$.tw,kf,ot.
4	Reticulolymphosarcom$.tw,kf,ot.
5	Hodgkin$.tw,kf,ot.
6	(malignan$ adj2 (lymphogranulom$ or granulom$)).tw,kf,ot.
7	or/1-6
8	exp POSITRON-EMISSION TOMOGRAPHY/
9	(pet$ or petscan$).tw,kf,ot.
10	tomograph$.tw,kf,ot.
11	emission$.tw,kf,ot.
12	exp DEOXYGLUCOSE/
13	FLUORODEOXYGLUCOSE F18/
14	(deoxyglucose$ or desoxyglucose$ or deoxy-glucose$ or desoxy-glucose$ or deoxy-d-glucose$ or desoxy-d-glucose$ or 2deoxyglucose$ or 2deoxy-d-glucose$ or fluorodeoxyglucose$ or fluorodesoxyglucose$ or fludeoxyglucose$ or fluordeoxyglucose$ or fluordesoxyglucose$ or 18fluorodeoxyglucose$ or 18fluorodesoxyglucose$ or 18fluordeoxyglucose$ or fdg$ or 18fdg$ or 18f-dg$).tw.
15	(fluor$ or 2fluor$ or fluoro$ or fluorodeoxy$ or fludeoxy$ or fluorine$ or 18f$ or 18flu$).tw.
16	glucose$.tw.
17	15 and 16
18	or/8-14
19	17 or 18
20	7 and 19
21	ANIMALS/ not HUMANS/
22	20 not 21
23	limit 22 to ed=20160203-20170712

#	Searches from 12/07/2017 - 12/11/2018
1	*LYMPHOMA/
2	HODGKIN DISEASE/
3	germinoblastom$.tw,kf,ot.
4	reticulolymphosarcom$.tw,kf,ot.
5	hodgkin$.tw,kf,ot.
6	(malignan$ adj2 (lymphogranulom$ or granulom$)).tw,kf,ot.
7	or/1-6
8	exp POSITRON-EMISSION TOMOGRAPHY/
9	(pet$).tw,kf,ot.
10	tomograph$.tw,kf,ot.
11	emission$.tw,kf,ot.
12	exp DEOXYGLUCOSE/
13	FLUORODEOXYGLUCOSE F18/
14	(deoxyglucose$ or desoxyglucose$ or deoxy-glucose$ or desoxy-glucose$ or deoxy-d-glucose$ or desoxy-d-glucose$ or 2deoxyglucose$ or 2deoxy-d-glucose$ or fluorodeoxyglucose$ or fluorodesoxyglucose$ or fludeoxyglucose$ or fluordeoxyglucose$ or fluordesoxyglucose$ or 18fluorodeoxyglucose$ or 18fluorodesoxyglucose$ or 18fluordeoxyglucose$ or fdg$ or 18fdg$ or 18f-dg$).tw.
15	(glucose$ and (fluor$ or 2fluor$ or fludeoxy$ or 18f$)).tw.
16	or/8-15
17	7 and 16
18	exp ANIMALS/ not HUMANS/
19	17 not 18
20	limit 19 to ed=20160203-20170712
21	limit 19 to ed=20170712-20181112
22	ANIMALS/ not HUMANS/
23	21 not 22
24	limit 23 to ed=20160203-20170712
25	limit 23 to ed=20170712-20181112

#	Searches from 12/11/2018 - 02/04/2019
1	*LYMPHOMA/
2	HODGKIN DISEASE/
3	Germinoblastom$.tw,kf,ot.
4	Reticulolymphosarcom$.tw,kf,ot.
5	Hodgkin$.tw,kf,ot.
6	(malignan$ adj2 (lymphogranulom$ or granulom$)).tw,kf,ot.
7	or/1-6
8	exp POSITRON-EMISSION TOMOGRAPHY/
9	(pet$ or petscan$).tw,kf,ot.
10	tomograph$.tw,kf,ot.
11	emission$.tw,kf,ot.
12	exp DEOXYGLUCOSE/
13	Fluorodeoxyglucose F18/
14	(deoxyglucose$ or desoxyglucose$ or deoxy-glucose$ or desoxy-glucose$ or deoxy-d-glucose$ or desoxy-d-glucose$ or 2deoxyglucose$ or 2deoxy-d-glucose$ or fluorodeoxyglucose$ or fluorodesoxyglucose$ or fludeoxyglucose$ or fluordeoxyglucose$ or fluordesoxyglucose$ or 18fluorodeoxyglucose$ or 18fluorodesoxyglucose$ or 18fluordeoxyglucose$ or fdg$ or 18fdg$ or 18f-dg $).tw.
15	(glucose$ and (fluor$ or 2fluor$ or fludeoxy$ or 18f$)).tw.
16	or/8-15
17	7 and 16
18	exp ANIMALS/ not HUMANS/
19	17 not 18
20	limit 19 to ed=20160203-20170712
21	limit 19 to ed=20170712-20181112
22	limit 19 to ed=20181112-20190402

key: exp # /: explode # MeSH subject heading, tw: text word, kf: keyword heading word, ot: original title, ti: title, $: truncation, adj#: adjacent within # number of words

Appendix 3. Embase/Ovid search strategy

#	Searches
1	exp CLASSICAL HODGKIN LYMPHOMA/
2	*HODGKIN DISEASE/
3	germinoblastom*.tw,kw.
4	reticulolymphosarcom*.tw,kw.
5	hodgkin*.tw,kw.
6	(malignan* adj2 (lymphogranulom* or granulom*)).tw,kw.
7	or/1-6
8	exp POSITRON EMISSION TOMOGRAPHY/
9	pet*.tw,kw.
10	tomograph*.tw,kw.
11	emission*.tw,kw.
12	exp DEOXYGLUCOSE/
13	FLUORODEOXYGLUCOSE F 18/
14	(deoxyglucose* or desoxyglucose* or deoxy-glucose* or desoxy-glucose* or deoxy-d-glucose* or desoxy-d-glucose* or 2deoxyglucose* or 2deoxy-d-glucose* or fluorodeoxyglucose* or fluorodesoxyglucose* or fludeoxyglucose* or fluordeoxyglucose* or fluordesoxyglucose* or 18fluorodeoxyglucose* or 18fluorodesoxyglucose* or 18fluordeoxyglucose* or fdg* or 18fdg* or 18f-dg*).tw.
15	(glucose* and (fluor* or 2fluor* or fludeoxy* or 18f*)).tw.
16	or/8-15
17	7 and 16
18	exp ANIMAL/ not HUMAN/
19	17 not 18
20	limit 19 to yr="1990 -Current"
21	meta-analys:.mp. or search:.tw. or review.pt.
22	(child* or p?ediatric*).ti.
23	20 not (21 or 22)
24	limit 23 to embase

(Continued)

25	limit 23 to conference abstracts
26	24 or 25

key: exp # /: explode # EMTREE term, * # /: focus # EMTREE term, /: EMTREE term, tw: text word, kw: keyword, ti: title, mp: multiple purpose, pt: publication type, *: truncation, ?: wildcard

search line #21: Review Embase search filter - best balance of sensitivity and specificity https://hiru.mcmaster.ca/hiru/HIRU_Hedges_EMBASE_Strategies.aspx

WHAT'S NEW

Date	Event	Description
14 August 2020	Amended	Following correspondence between the editorial base and the funding institution of one of the authors, the internal sources of support and the acknowledging statement was updated.

HISTORY

Protocol first published: Issue 4, 2017
Review first published: Issue 9, 2019

Date	Event	Description
20 December 2019	New citation required but conclusions have not changed	Following correspondence between the authors and one of the peer reviewers post-publication, the authors have revised some of the risk of bias judgements. Some terminology around confounding has also been changed.
20 December 2019	Amended	Following correspondence between the authors and one of the peer reviewers post-publication, the authors have revised some of the risk of bias judgements. Some terminology around confounding has also been changed.

CONTRIBUTIONS OF AUTHORS

Angela Aldin: screening and selection of studies, development of data extraction form, data extraction, 'Risk of bias' assessment, GRADE assessment, data analysis interpretation, 'Summary of findings' tables, writing and drafting of the review, communication with and between authors.

Lisa Umlauff: 'Risk of bias' assessment, characteristics of included and excluded studies (texts and tables), abstract and Plain language summary, proofread and commented on the draft.

Karel Moons: methodological input on reviews of prognosis studies.

Lise J Estcourt: screening and selection of studies, data extraction, risk of bias assessment, clinical and methodological input.

Andreas Engert: medical and content input, particularly on the clinical comparability of studies and subgroup analysis.

Carsten Kobe: nuclear medical input on PET-CT.

Bastian von Tresckow: clinical input, particularly on the clinical comparability of studies and subgroup analysis.

Gary Collins: methodological input on reviews of prognostic studies.

Madhuri Haque: screening and selection of studies.

Farid Foroutan: input on risk of bias and GRADE assessment of prognostic factor studies.

Nina Kreuzberger: 'Risk of bias' assessment, proofread and commented on the review draft.

Marialena Trivella: screening and selection of studies, data extraction, risk of bias assessment, statistical analysis, proofread and commented on the review draft.

Nicole Skoetz: protocol development, screening and selection of studies, data extraction, risk of bias assessment, GRADE assessment, proofread and commented on the review draft.

DECLARATIONS OF INTEREST

Angela Aldin: award of the grant by Federal Ministry of Education and Research for the University Hospital of Cologne to perform this systematic review does not lead to a conflict of interest.

Lisa Umlauff: award of the grant by Federal Ministry of Education and Research for the University Hospital of Cologne to perform this systematic review does not lead to a conflict of interest.

Karel Moons: none known.

Lise J Estcourt: award of the grant by Federal Ministry of Education and Research to the University of Oxford to perform this systematic review does not lead to a conflict of interest.

Andreas Engert: award of the grant by Federal Ministry of Education and Research for the University Hospital of Cologne to perform this systematic review does not lead to a conflict of interest. Principal investigator of the HD18 trial, does not lead to a conflict of interest. Received funds from Takeda Pharma GmbH, BMS and MSD for consultancy and educational presentations, but these were not related to the intervention in this review. No competing interests.

Carsten Kobe: award of the grant by Federal Ministry of Education and Research for the University Hospital of Cologne to perform this systematic review does not lead to a conflict of interest.

Bastian von Tresckow: award of the grant by Federal Ministry of Education and Research for the University Hospital of Cologne to perform this systematic review does not lead to a conflict of interest. Received funds from Novartis Pharma GmbH, Takeda Pharma GmbH and MSD for consultancy and educational presentations, but these were not related to the intervention in this review. No competing interests.

Gary Collins: supported by the NIHR Biomedical Research Centre, Oxford, and Cancer Research UK (programme grant: C49297/A27294). No conflict of interest.

Madhuri Haque: award of the grant by Federal Ministry of Education and Research for the University Hospital of Cologne to perform this systematic review does not lead to a conflict of interest.

Farid Foroutan: none known.

Nina Kreuzberger: award of the grant by Federal Ministry of Education and Research for the University Hospital of Cologne to perform this systematic review does not lead to a conflict of interest.

Marialena Trivella: part of the grant from the Federal Ministry of Education and Research to the University Hospital of Cologne, was paid to the University of Oxford for author time spent working on this review. However, the funder played no part in the design and execution of the project and it does not constitute a conflict of interest.

Nicole Skoetz: award of the grant by Federal Ministry of Education and Research for the University Hospital of Cologne to perform this systematic review does not lead to a conflict of interest.

SOURCES OF SUPPORT

Internal sources

- University Hospital of Cologne, Germany

  Cochrane Haematological Malignancies, Department 1 of Internal Medicine
- NHS Blood and Transplant, UK

External sources

- Federal Ministry of Education and Research, Germany

  Funding Number 01KG1709

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

We included studies that evaluated both adult and adolescent participants (the youngest being 13 years old), as opposed to including adult participants (≥ 18 years of age) only as stated in the protocol of this review. Hodgkin lymphoma is a disease with a typical onset in adolescence to mid-adulthood, with little physiological differences between adolescents and adults. In the studies included in this review, participants under the age of 18 were treated in the same clinic and received the same treatment as participants over ≥ 18 years of age. We believe that the results regarding interim PET are equally relevant to adolescents as they are to adult participants and, therefore, did not see reasons against the inclusion of studies including both younger and older adults. Nevertheless, we did not include studies that evaluated solely paediatric participants. In studies where only paediatric participants are included, it is more likely they will be treated in paediatric clinics and receive a different treatment regimen than adult participants.

We used an amended version of the Quality in Prognostic Factor Studies (QUIPS) tool to assess the risk of bias of the included studies. In consultation with Hayden and colleagues (Hayden 2013), we adapted the QUIPS tool by adding 'unclear (no information)' as a fourth judgement in the tool. In addition, we renamed the fifth domain of the tool, originally named 'study confounding', into 'other prognostic factors (covariates)', to highlight the importance of adjusting for other prognostic factors and distinguish it from confounding. Lastly, we assessed all six domains (study participation, study attrition, prognostic factor measurement, outcome measurement, other prognostic factors (covariates), statistical analysis and reporting) per outcome (OS and PFS) in each study. The first three domains ended up always receiving the same judgement as they are indeed to be considered at study level. With regard to the outcomes, however, we identified differences in analysis and reporting within studies.

With regard to data extraction, we developed our own data extraction form specific to prognostic factor studies (particularly those that are included in this review), which includes more items than stipulated in the protocol of this review.

Lastly, we searched Embase as an additional database, as well as one trial registry (ClinicalTrials.gov).

INDEX TERMS

Medical Subject Headings (MeSH)

Antineoplastic Combined Chemotherapy Protocols [*therapeutic use]; Chemoradiotherapy; Decision Making; Disease Progression; Disease-Free Survival; Hodgkin Disease [*drug therapy]; Positron Emission Tomography Computed Tomography [*methods]; Prognosis

MeSH check words

Humans; Young Adult

Appendix III: Declaration of an oath (dt. Eidesstattliche Versicherung)

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen - einschließlich Tabellen, Karten und Abbildungen -, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, sind in jedem Einzelfall als Entlehnung kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertationsschrift noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss der Promotion nicht ohne Genehmigung der / des Vorsitzenden des IPHS-Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Frau Prof. Dr. Nicole Skoetz betreut worden.

Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen.

Übersicht der Publikationen:

1. **Aldin A**, Besiroglu B, Adams A, Monsef I, Piechotta V, Tomlinson E, Hornbach C, Dressen N, Goldkuhle M, Maisch P, Dahm P, Heidenreich A, Skoetz N. First-line therapy for adults with advanced renal cell carcinoma: a systematic review and network meta-analysis. Cochrane Database Syst Rev. 2023 May 4;5(5):CD013798. doi: 10.1002/14651858.CD013798.pub2. PMID: 37146227; PMCID: PMC10158799.
2. **Aldin A**, Umlauff L, Estcourt LJ, Collins G, Moons KG, Engert A, Kobe C, von Tresckow B, Haque M, Foroutan F, Kreuzberger N, Trivella M, Skoetz N. Interim PET-results for prognosis in adults with Hodgkin lymphoma: a systematic review and meta-analysis of prognostic factor studies. Cochrane Database of Systematic Reviews 2020, Issue 1. Art. No.: CD012643. DOI: 10.1002/14651858.CD012643.pub3.

Ich versichere, dass ich alle Angaben wahrheitsgemäß nach bestem Wissen und Gewissen gemacht habe und verpflichte mich, jedmögliche, die obigen Angaben betreffenden Veränderungen, dem IPHS-Promotionsausschuss unverzüglich mitzuteilen.

27.06.2023

....

Datum Unterschrift