

Systematic Reviews of Prognosis in Haematological Malignancies

Inaugural Dissertation

zur

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

vorgelegt von

Nina Kreuzberger
aus Dinslaken/Wesel

Köln, 2025

Betreuerin:	Prof. Dr. Nicole Skoetz
Gutachterin / Gutachter:	Prof. Dr. Kerstin Rhiem Prof. Dr. Bastian von Tresckow
Datum der Mündlichen Prüfung:	19. August 2025

Table of Content

Summary	IV
Background	IV
Objective	IV
Methods	IV
Results	V
Conclusion	V
Zusammenfassung	VI
Hintergrund	VI
Zielsetzung	VI
Methoden	VII
Ergebnisse	VII
Schlussfolgerung	VIII
List of abbreviations.....	IX
Introduction.....	1
Systematic Reviews	1
Prognosis	2
Overall Prognosis	3
Prognostic Factors	4
Prognostic Models	6
Stratified medicine	11
Prognosis Systematic Reviews	12
Defining the question	12
Searching	13
Screening	14
Data extraction	14
Risk of Bias assessment	15
Meta-analysis	16
Grades of Recommendation, Assessment, Development, and Evaluation (GRADE)	18
Prognosis in Haematological Malignancies	19

Aim and Objectives	22
PROGRESS theme 2 – prognostic factors	22
PROGRESS theme 3 – prognostic models	22
PROGRESS theme 4 – stratified medicine	23
Results: Dissertation Project Synopses	25
Publication 1: Interim PET-results for prognosis in adults with Hodgkin lymphoma: a systematic review and meta-analysis of prognostic factor studies	25
Summary	25
Visual abstract	29
Publication 2: Prognostic models for newly-diagnosed chronic lymphocytic leukaemia in adults: a systematic review and meta-analysis	30
Summary	30
Visual abstract	40
Publication 3: Positron emission tomography-adapted therapy for first-line treatment in adults with Hodgkin lymphoma	41
Summary	41
Visual abstract	47
Discussion.....	48
Summary	48
Strengths and Limitations of the Methodology	49
Strengths and Limitations of the included Literature, and resulting implications for primary prognosis studies	49
Implications for Systematic Reviews of Prognosis Studies	51
Implications for Clinical practice	52
Conclusion	54
References.....	55
Acknowledgements	64
Supplementary information	65
Attributions of Icons and Software	65
Funding	66
Conflicts of Interest	66

Appendices	1
Appendix I: Scientific Contribution of Authors	1
Publication 1: Interim PET-results for prognosis in adults with Hodgkin lymphoma: a systematic review and meta-analysis of prognostic factor studies	1
Publication 2: Prognostic models for newly-diagnosed chronic lymphocytic leukaemia in adults: a systematic review and meta-analysis	2
Publication 3: Positron emission tomography-adapted therapy for first-line treatment in adults with Hodgkin lymphoma	3
Appendix II: Publication 1	1
Appendix III: Publication 2	1
Appendix IV: Publication 3	1
Appendix V: Declaration of an Oath	1

Summary

Background

Systematic reviews provide a summary of all available empirical evidence for a well-defined research question, whose methodology can also be applied to prognosis research. Prognosis is the forecasting of a future health state or the course of a disease based on a defined baseline health state. Prognostic research questions can be divided into four themes according to the PROGnosis RESearch Strategy (PROGRESS) framework. While overall prognosis describes the course of a disease under standard of care, prognostic factor research focuses on individual characteristics that can distinguish between people with a good versus a worse prognosis. Prognostic models are algorithms that combine at least two factors to provide individual risks to people regarding a condition in question. The fourth theme, stratified medicine, can be applied both to prognostic factors and models, resulting in treatment adaptations based on individual characteristics or factor combinations. Biomarkers may also serve as targets for treatment. Although the amount of prognostic primary literature is increasing, the methodology for systematic reviews on prognosis questions within Cochrane was not yet established.

Objective

The aim of this dissertation was to cover three out of four prognosis themes and develop exemplary work regarding methodology of prognosis systematic reviews based on clinically relevant questions in the field of haematology. The following systematic reviews were conducted as part of this dissertation:

- Interim PET-results for prognosis in adults with Hodgkin lymphoma (HL): a systematic review and meta-analysis of prognostic factor studies
- Prognostic models for newly-diagnosed chronic lymphocytic leukaemia (CLL) in adults: a systematic review and meta-analysis
- Positron emission tomography-adapted therapy for first-line treatment in adults with HL.

Methods

The three Cochrane systematic reviews were conducted using the main steps according to the internationally recognised gold standard by Cochrane: protocol development, search strategy, screening of the literature, risk of bias assessment, data extraction and synthesis, and rating of the certainty in the evidence. Within each step, the methodology was adapted according to the type of research question: the risk of bias tools, extracted items, and meta-analytical approaches varied between the reviews. The methodology for rating the certainty in the evidence was just being developed for prognostic factors and had not yet been developed for prognostic models.

Results

The systematic review investigating interim-PET as a prognostic factor for the course of previously untreated HL showed that across stages, the interim-PET status is probably a good predictor for overall survival and may also be a predictor for progression-free survival. Adverse events related to PET were not reported. There are serious limitations regarding the conduct and reporting of prognostic factor studies. For instance, most studies were rated as being at high risk of bias due to a lack of adjustment, due to lack of reporting, or inconsistencies regarding their analysis.

The systematic review on prognostic models identified 52 developed models. Of those, only 12 were externally validated, and only three were validated at least three times, which allowed meta-analysis. These models are the Chronic lymphocytic leukaemia international prognostic index (CLL-IPI) based on five factors, the MDACC 2007 score consisting of six factors, and the Barcelona-Bрно score combining three factors. The CLL-IPI seems to discriminate best between low and high-risk individuals. As calibration was not reported in a way that allowed meta-analyses, we pooled survival per risk group. Risk of bias summarised for validation studies was rated as high due to frequent inappropriate handling of missing data, low number of events and lack of reporting, in addition to problems in other domains.

The review on interim-PET based treatment adaptation in HL shows that in early-stage HL with a negative interim-PET, de-escalation cannot be recommended due to uncertainty. In intermediate-stage HL, de-intensification might be an option, as it does not seem to affect survival and progression and secondary malignancies. In advanced-stage HL, de-escalation probably results in a benefit for patients regarding overall survival and secondary malignancies without affecting progression-free survival. In interim-PET positive HL, treatment intensification may increase overall and progression-free survival in early-to-intermediate stage. Adding rituximab to standard therapy in advanced-stage HL has probably no benefit for overall and progression-free survival but increases harms.

Conclusion

Prognosis systematic reviews are a feasible study design for Cochrane. The primary study landscape still suffers from limitations such as lack of indexing, very serious reporting deficiencies and studies at high risk of bias. This series of prognosis reviews shows that knowledge on a good prognostic factor can result in this factor being explored as a target for treatment adaptation. However, the differences of PET-based treatment adaptation across early, intermediate, and advanced stages also show that such factor does not always translate into good stratification. The precise situations of benefit have to be carefully disentangled. The prognostic model literature seems to lean unproportionally towards developing models as compared to externally validating or impact testing these models, leading to an abundance of models without clinical applicability.

Zusammenfassung

Hintergrund

Systematische Reviews bieten eine Übersicht aller verfügbaren Evidenz zu einer klar definierten Forschungsfrage. Die Methodik ist auch auf die Prognoseforschung anwendbar. Unter Prognose versteht man die Vorhersage eines zukünftigen Gesundheitszustandes oder eines Krankheitsverlaufes auf der Grundlage eines definierten Ausgangszustandes. Prognostische Forschungsfragen können gemäß dem PROGnosis RESearch Strategy (PROGRESS)-Rahmen in vier Themenbereiche unterteilt werden. Während die Gesamtprognose den Verlauf einer Krankheit unter Standardbehandlung beschreibt, konzentriert sich die Forschung zu prognostischen Faktoren auf einzelne Merkmale, die zwischen Menschen mit einer guten und einer schlechten Prognose unterscheiden können. Prognosemodelle sind Algorithmen, die mindestens zwei Faktoren kombinieren, um individuelle Risiken für Menschen in Bezug auf ein bestimmtes Krankheitsbild zu ermitteln. Das vierte Thema, die stratifizierte Medizin, kann sowohl auf prognostische Faktoren als auch auf Modelle angewendet werden. Sie kann zu einer Anpassung der Behandlung auf der Grundlage individueller Merkmale oder Faktorenkombinationen führen. Verschiedene Biomarker können auch als Target für Behandlungsansätze dienen. Obwohl die Menge der prognostischen Primärliteratur steigt, wurde die Methodik für systematische Reviews zu Prognosefragen innerhalb von Cochrane noch nicht festgelegt.

Zielsetzung

Ziel dieser Dissertation war es, exemplarische prognostischen systematischen Übersichtsarbeiten in drei der vier Themenbereiche abzudecken und auf Basis klinisch relevanter Fragestellungen im Bereich der Hämatologie weiterzuentwickeln. Die folgenden systematischen Übersichten wurden im Rahmen dieser Dissertation durchgeführt:

- Interim-PET-Befunde zur Prognose bei Erwachsenen mit Hodgkin-Lymphom (HL): ein systematisches Review und Meta-Analyse von Studien zu prognostischen Faktoren
- Prognosemodelle für neu diagnostizierte chronische lymphatische Leukämie (CLL) bei Erwachsenen: ein systematisches Review und Meta-Analyse
- Positronen-Emissions-Tomographie-angepasste Therapie für die Erstlinienbehandlung bei Erwachsenen mit HL.

Methoden

Die drei systematischen Reviews wurden gemäß dem international anerkannten Cochrane-Goldstandard durchgeführt. Die Hauptschritte waren jeweils folgende: Erstellung des Protokolls, Suchstrategie, Screening der Literatur, Bewertung des Verzerrungsrisikos, Datenextraktion und -synthese sowie Bewertung der Evidenzsicherheit. In jedem Arbeitsschritt wurde die Methodik bezüglich der Bewertung des Verzerrungsrisikos, der extrahierten Elemente und der meta-analytischen Ansätze an die Art der Forschungsfrage angepasst. Die Methodik zur Bewertung des Vertrauens in die Evidenz wurde für prognostische Faktoren entwickelt und war zum Zeitpunkt der Durchführung der Modellüberprüfung noch nicht für prognostische Modelle verfügbar.

Ergebnisse

Das systematische Review, in der die Interim-PET als prognostischer Faktor für den Verlauf von zuvor unbehandeltem HL untersucht wurde, zeigte, dass der Interim-PET-Status über alle Stadien hinweg wahrscheinlich ein guter Prädiktor für das Gesamtüberleben ist und möglicherweise auch ein Prädiktor für das progressionsfreie Überleben ist. Unerwünschte Ereignisse im Zusammenhang mit der PET wurden nicht berichtet. Bei der Durchführung und Berichterstattung von Studien zu prognostischen Faktoren gibt es erhebliche Mängel. So wurde die Mehrzahl der Studien als stark verzerrungsgefährdet eingestuft, weil sie nicht angepasst wurden, weil sie nicht berichtet wurden oder weil ihre Auswertung inkonsistent mit dem Protokoll war.

Im systematischen Review zu prognostischen Modellen wurden 52 Modelle identifiziert. Von diesen wurden nur 12 extern validiert, drei davon wurden mindestens dreimal validiert, was eine Meta-Analyse ermöglichte. Bei diesen Modellen handelt es sich um den auf fünf Faktoren basierenden Chronic lymphocytic leukaemia international prognostic index (CLL-IPI), den aus sechs Faktoren bestehenden MDACC 2007 Score und den Barcelona-Brno Score, der drei Faktoren kombiniert. Der CLL-IPI scheint am besten zwischen Personen mit niedrigem und hohem Sterberisiko zu differenzieren. Da die Kalibrierung nicht in einer Weise berichtet wurde, die Meta-Analysen erlaubte, wurde das Überleben pro Risikogruppe meta-analysiert. Das für die Validierungsstudien zusammengefasste Verzerrungsrisiko wurde als hoch eingestuft, da häufig mit fehlenden Daten unsachgemäß umgegangen wurde, die Zahl der Ereignisse gering war, und Studiendetails nicht ausreichend berichtet wurden, zusätzlich zu Schwierigkeiten in anderen Bereichen.

Das Review zu einer auf Interim-PET basierenden Therapieanpassung beim HL zeigt, dass im Frühstadium mit einem negativen Interim-PET eine Deeskalation aufgrund der Unsicherheit nicht empfohlen werden kann. Bei HL im intermediären Stadium könnte eine De-Intensivierung eine Option sein, da sie sich nicht auf das Überleben, die Progression und sekundäre Tumore auswirkt.

Beim fortgeschrittenen HL führt die Deeskalation wahrscheinlich zu einem Vorteil hinsichtlich des Gesamtüberlebens und der sekundären Tumore, ohne das progressionsfreie Überleben zu beeinträchtigen. Bei Interim-PET-positivem HL kann eine Intensivierung der Behandlung das Gesamtüberleben und das progressionsfreie Überleben im frühen bis mittleren Stadium erhöhen. Die Zugabe von Rituximab zur Standardtherapie beim HL im fortgeschrittenen Stadium hat wahrscheinlich keinen Nutzen für das Gesamtüberleben und das progressionsfreie Überleben, erhöht jedoch negative Effekte.

Schlussfolgerung

Systematische Reviews zu prognostischen Fragestellungen sind ein geeignetes Studiendesign für Cochrane. Die Primärstudienlandschaft unterliegt nach wie vor Beschränkungen, wie z. B. fehlende Indexierung, gravierende Berichtsmängel und Studien mit hohem Verzerrungsrisiko. Diese Serie von Prognosereviews verdeutlicht, dass die Feststellung eines guten prognostischen Faktors dazu führen kann, dass dieser Faktor als Grundlage für die Anpassung der Behandlung untersucht wird. Die stadienabhängigen Unterschiede bei der PET-basierten Behandlungsanpassung zeigt jedoch auch, dass ein solcher Faktor nicht in jeder Situation zu einer guten Stratifizierung führt. Die genauen Situationen, in denen ein Vorteil besteht, sind sorgfältig zu entschlüsseln. Die Literatur zu prognostischen Modellen tendiert unverhältnismäßig stark zur Entwicklung von Modellen im Vergleich zur externen Validierung oder Wirkungsprüfung dieser Modelle, was zu einer Vielzahl von Modellen ohne klinische Anwendbarkeit führt.

List of abbreviations

Abbreviation	Definition
ABVD	Adriamycin, bleomycin, vinblastine and dacarbazine (chemotherapy regimen)
BEACOPP	Bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine and prednisone
BMBF	Bundesministerium für Bildung und Forschung
CCA	Creative Commons Attribution
CENTRAL	Cochrane Central Register of Controlled Trials
CHARMS	CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies
CI	Confidence interval
CLL	Chronic lymphocytic leukaemia
DLBCL	Diffuse large B-cell lymphoma
GRADE	Grading of Recommendations, Assessment, Development and Evaluation
Gy	Gray (unit of ionizing radiation dose)
HL	Hodgkin lymphoma
HR	Hazard ratio
IgHV	Immunoglobulin heavy chain variable segment
LDH	Lactate dehydrogenase
FDG-PET	Fluorodeoxyglucose - positron emission tomography
FISH	Fluorescence in situ hybridization
OS	Overall survival
PICO	Population, Intervention, Comparator, Outcome
PFS	Progression-free survival
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
PROBAST	Prediction Model Risk of Bias Assessment Tool
PROGRESS	PROGnosis RESearch Strategy
QoL	Quality of life
QUIPS	Quality in Prognostic Studies

RCT	Randomised controlled trial
RoB	Risk of Bias
RT	Radiotherapy
SAE	Serious adverse event
SoC	Standard of care

Introduction

Systematic Reviews

A systematic review is a summary of all available empirical evidence regarding an explicitly formulated research question that can help clinicians and policy-makers take the best health-related decisions. They are conducted using preplanned, reproducible, and transparent methodology. This includes an a priori registered protocol specifying the scope and eligibility of included studies, a literature search strategy for medical databases, the process for study selection, data extraction, risk of bias (RoB) assessment, and meta-analytic approaches (1). A step-by-step description of the process is also depicted in Figure 1.

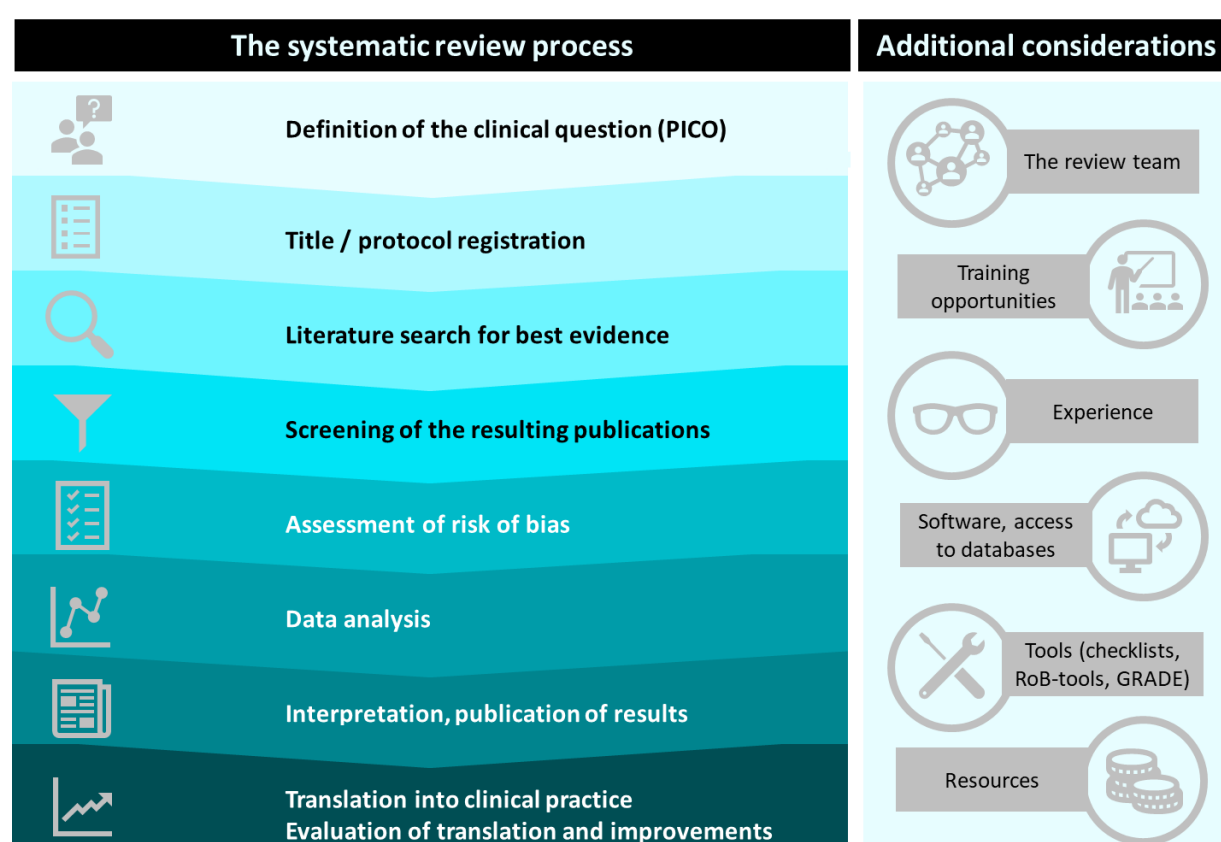


Figure 1: The systematic review process, from defining the clinically relevant research question to interpretation and publication of its results, ideally including a translation into clinical practice and the evaluation of the intervention's impact. In the right panel, additional considerations that, ideally, should be integrated at the planning stage, latest when writing the protocol.

Before starting a systematic review, the researchers should assemble a multidisciplinary team with methodological as well as the relevant clinical expertise of the question. At least some authors should be experienced in systematic review methodology. Additionally, they have to consider training options for those who have not yet contributed to a systematic review, identify the most helpful

software to use, determine whether databases for literature can be accessed, and calculate the resources in personnel and consumables needed (1).

Any kind of question may be answered with help of a systematic review; most commonly, intervention questions following the PICO-framework (Population, Intervention, Comparator, Outcome) are answered, however, other research question formats have emerged, such as diagnostic systematic reviews (2), systematic reviews of qualitative research (3), and systematic reviews of prognostic questions. The main steps outlined in Figure 1 can be applied to any type of question and study design to be included. However, the methodology to conduct each step differs, i.e., different study designs and questions may vary in their filters for study identification, items for data extraction, effect estimates, risk of bias assessment, and meta-analytic approach.

Prognosis

In the medical field, the term prognosis describes the forecast of a future health state, the likely course of a disease, or the expectancy of a specific outcome in the future from knowing the current health state of a person, also called the baseline health state or starting point of prediction (4). This starting point can be any time point of interest that is to be used for a prediction of the future outcome, i.e., the incidence of diseases calculated from a polygenic risk score at birth (5), the prediction of the risk of cardiovascular disease from a general population of children based varying levels of risk factors (6), or the probability of survival at 5 years given a certain diagnosis such as lung cancer based on national averages (7).

Accurate knowledge on prognosis is essential for various reasons depending on the context. For instance, information on the incidence and outcome of certain conditions can be used for prioritization of resources and policy making (4), such as screening and prevention programs, improved hygiene standards, reconsideration of access to care, etc. Prognostic information based on a defined diagnosis of a specific disease is especially important for the process of informed decision making between patient and clinician regarding the further treatment of the disease and to potentially target additional risk factors such as weight or smoking status. In addition, prognostic information may help set up clinical trials and identify those individuals potentially benefitting most from a specific trial, i.e., if a person shows specific molecular markers that a new therapy was developed for (4).

Prognosis research can be divided into four different themes with different aims and methodological challenges according to the PROGRESS framework (4, 8-10):

- Overall prognosis, also called fundamental prognosis,
- Prognostic factors,

- Prognostic models,
- Stratified medicine.

Despite the emergence of numerous reporting guidelines such as CHARMS-PF and REMARK for prognostic factors (11, 12), or CHARMS and TRIPOD for prognostic model studies (13, 14), studies on prognosis are yet far from being well-conducted and well-reported (15, 16). For instance, the median adherence to reporting items of TRIPOD across 152 developed machine learning models was only 41% (range 10% to 67%); each publication had at least one item flagged for inadequate reporting (15).

Overall Prognosis

Overall prognosis research lays the foundation for the other interrelated themes by providing an average future risk or average score of an outcome under current clinical practice in a defined clinical setting and population. Research on overall prognosis is broad and descriptive in nature. The theme may cover the natural history of a disease in the absence of care, a description of diagnosis and care, access to care and respective outcomes without the aim of comparative effect estimates. It is possible that differences in outcomes may be described based on patient clusters, i.e. based on country or hospital level, entity, disease stage, and similar, but without a comparative effect estimate. Figure 2A shows questions that a policy maker, clinician or patient themselves may have regarding a disease. Figure 2B and 1C illustrate how overall prognosis data can be presented. Often, simple descriptive graphs such as boxplots are used and entities or groups descriptively compared (1B), but data can also be mapped if related to location as the average predicted life expectancy per country based on data from the World Health Organization from 2019 in 1C.

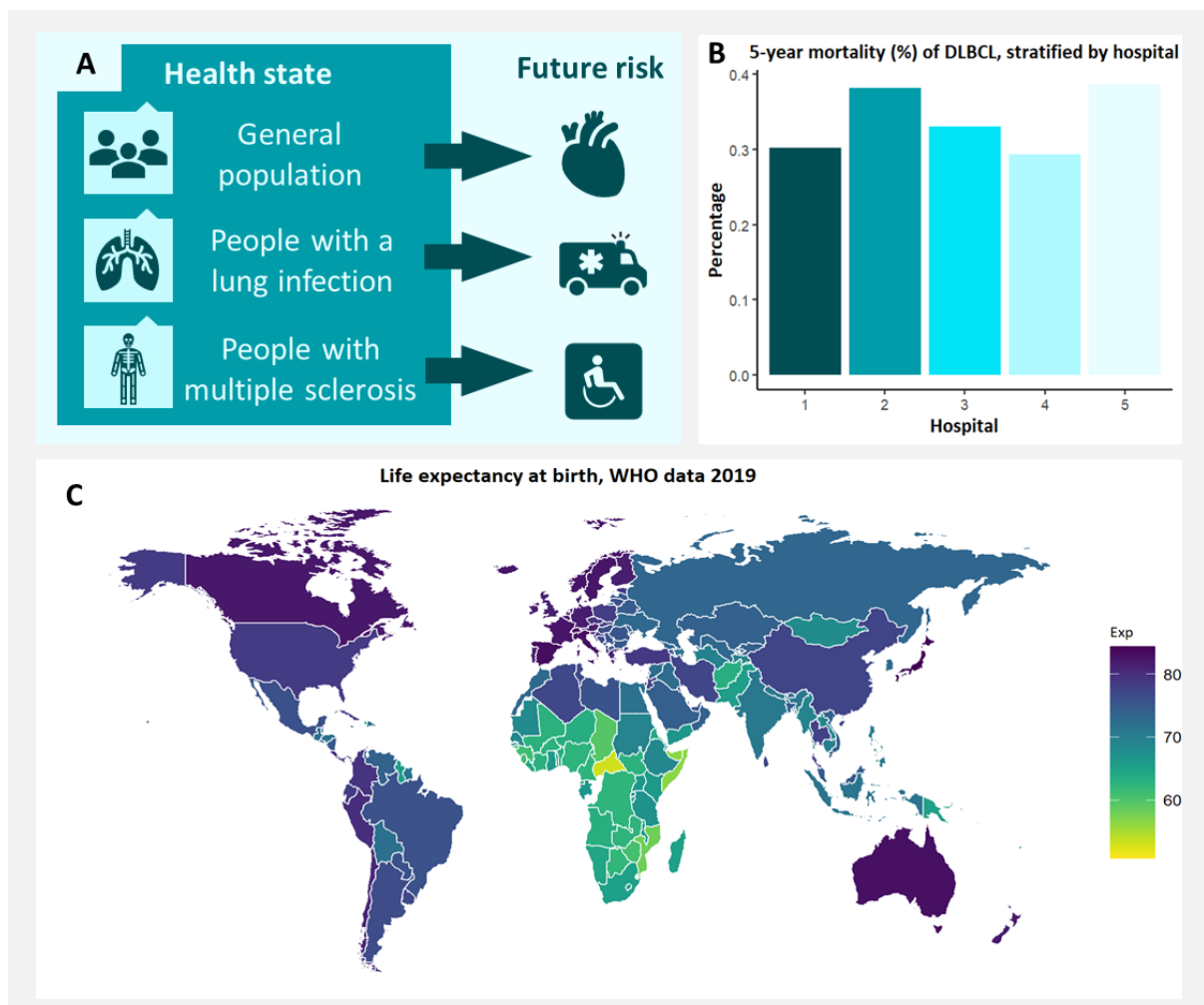


Figure 2: Overall prognosis research. **A** Exemplary overall prognosis questions, i.e., What is the risk of cardiovascular disease in the general population? How likely is the need for hospitalisation in a person with a viral lung infection? What is the risk of a person with multiple sclerosis to need a wheelchair in the future? **B** The risk of 5-year mortality for people with diffuse large B-cell lymphoma, presented for five different hospitals (benchmarking; fictional data for illustrative purpose only). **C** Life expectancy at birth in years for the world population per country, based on WHO data from 2019 (17).

Prognostic Factors

A prognostic factor is a characteristic that can distinguish between groups with different average future outcomes within a defined health state. Any characteristic can be considered a prognostic factor if it has a relationship with the outcome in question. Common examples include clinical data such as symptoms of a disease, pathological variables, medical imaging, and a variety of biomarkers, but also behavioural characteristics such as smoking, psychosocial or environmental factors such as pain perception. Even the built environment can be considered as prognostic factors (9). Those factors that are modifiable, such as behavioural and environmental factors, may serve as a basis for future interventions, while biomarkers may serve as targets for new therapeutics to be developed for specific antigens or to impact functions to prevent or treat diseases (9). Other uses of prognostic

factor information are their effect on refining diagnostic criteria, i.e., changes in staging systems over time, they may inform treatment strategies, may be used as parts of a prognostic model, or used for monitoring if changes in the prognostic factor are related to changes in the risk of the outcome (9).

Figure 3 shows how in a population with DLBCL, as in example 1B, a prognostic factor, lactate dehydrogenase (LDH), that has already been suspected to predict overall survival, is used and presented in prognostic factor research (18, 19). Figure 3A shows the process. At a starting point, for example at diagnosis, serum LDH levels are determined. The population sample is then divided into an elevated and a normal range LDH group and followed up over time. For specific time points, a risk ratio or odds ratio and its 95% confidence intervals can be calculated, while Figure 3B shows the presentation including consideration of those who drop out (also called censored) over time with help of a Kaplan-Meier curve and the hazard ratio (HR) as effect estimate (the underlying data in Figure 3 were simulated for illustrative purpose only, and are not related to real patient data).

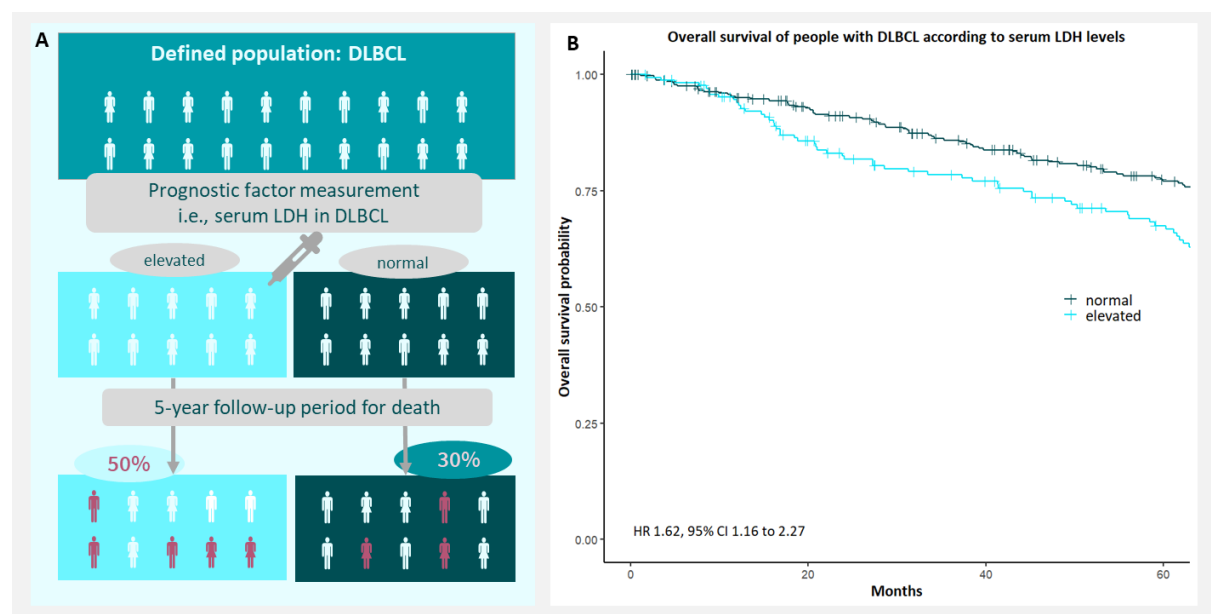


Figure 3: Prognostic factor research. A Exemplary process of confirmatory prognostic factor research with a defined health state, the diagnosis of diffuse large B-cell lymphoma. The prognostic factor to be examined is the serum lactate hydrogenase value at diagnosis. The cut-off value was chosen according to the literature, patients grouped according to their LDH level, and the sample followed up for 5 years. At 5 years, 50% of participants in the elevated group and 30% of the participants in the normal range experienced an event, in this case death, resulting in a risk ratio of 1.66. **B** Another presentation that is often chosen for long-term outcomes is a Kaplan-Meier curve, with each line representing one factor level, i.e., the dark line represents the normal range, the lighter line represents the elevated LDH group (fictional data for illustrative purpose only).

The question of prognostic factor research is merely whether an association between a factor and an outcome is present, the notion of causality is secondary. Research ranges from being exploratory, hypothesis-free testing of a range of factors, or even hundreds of genes, related to an outcome, to being confirmative, replicating the relationship in a number of independent studies (9). The

difference to stratified overall prognosis research similar to data presented in Figure 1B and 1C is that prognostic factor research aims to be comparative.

A particular challenge in prognostic factor research is its high potential for publication bias (9), which is a bias usually caused by a higher likelihood of positive factor – outcome relationships to get published than studies showing no relationship between a factor and an outcome (20). Often, prognostic factors are tested as a byproduct of a cohort study or clinical trial; they are rarely pre-registered for the purpose of prognosis and report selectively those factors that have tested as “significant” (also called reporting bias). Additionally, frequently, numerous factors are tested, resulting in an increase for potential false positive results. As a result, seemingly large effects of prognostic factors may be overestimated due to negative findings remaining unreported (9).

Another challenge that affects continuous prognostic factors when used individually as well as within a prognostic model is the choice to dichotomise or categorise a factor at an arbitrary value such as, for example, the median cohort value, or the assumption of a certain form of the factor-outcome relationship, i.e., linearity. These approaches, depending on the original form of the factor-outcome relationship, are inefficient and result in a loss of information, ultimately leading to a lower predictive performance both in the development and external validation cohorts (21). In addition, factors may be measured on different scales or with different measurements across settings.

Prognostic Models

Reality is usually more complex than an individual prognostic factor predicting an outcome. To forecast the outcome more accurately, multiple prognostic factors can be formally combined into an algorithm that provides individual risk scores or absolute risk probabilities for a specific outcome. Some famous prognostic models are the Apgar score (22) and the Framingham risk score.

- The Apgar score is a score with five factors – breathing effort, heart rate, muscle tone, reflexes, and skin colour - that is performed at one and five minutes after birth. Each factor can obtain a score between 0 and 2, totalling 10, with higher scores indicating a better health status (22). The score has been correlated to a multitude of outcomes such as cerebral palsy and epilepsy (23), academic performance (24), and mortality (25), amongst others.
- The Framingham risk score (Framingham Wilson) (26) consists of seven factors – sex, age, total cholesterol, HDL, systolic blood pressure, smoking status and diabetes and predicts the risk of cardiovascular disease at 10 years from the moment of prediction. This model has shown to overestimate the risk if it is not adapted to the baseline risk of the population it is used for (27).

Figure 4A illustrates the revised international prognostic index for DLBCL patients receiving R-CHOP as therapy that predicts the 4-year overall and progression-free survival, including the previously discussed prognostic factor LDH (28). The score is based on reshaping risk groups calculated from a previous model, the international prognostic index for Non-Hodgkin lymphoma (NHL), which was based on a broader population and generated using univariable and multivariable Cox proportional hazards regression analysis (29). Figure 4B provides the 4-year estimates (28), and Figure 4C & D, adapted from Alamer et al., 2020 (30), show the OS and PFS per risk group in an independent cohort.

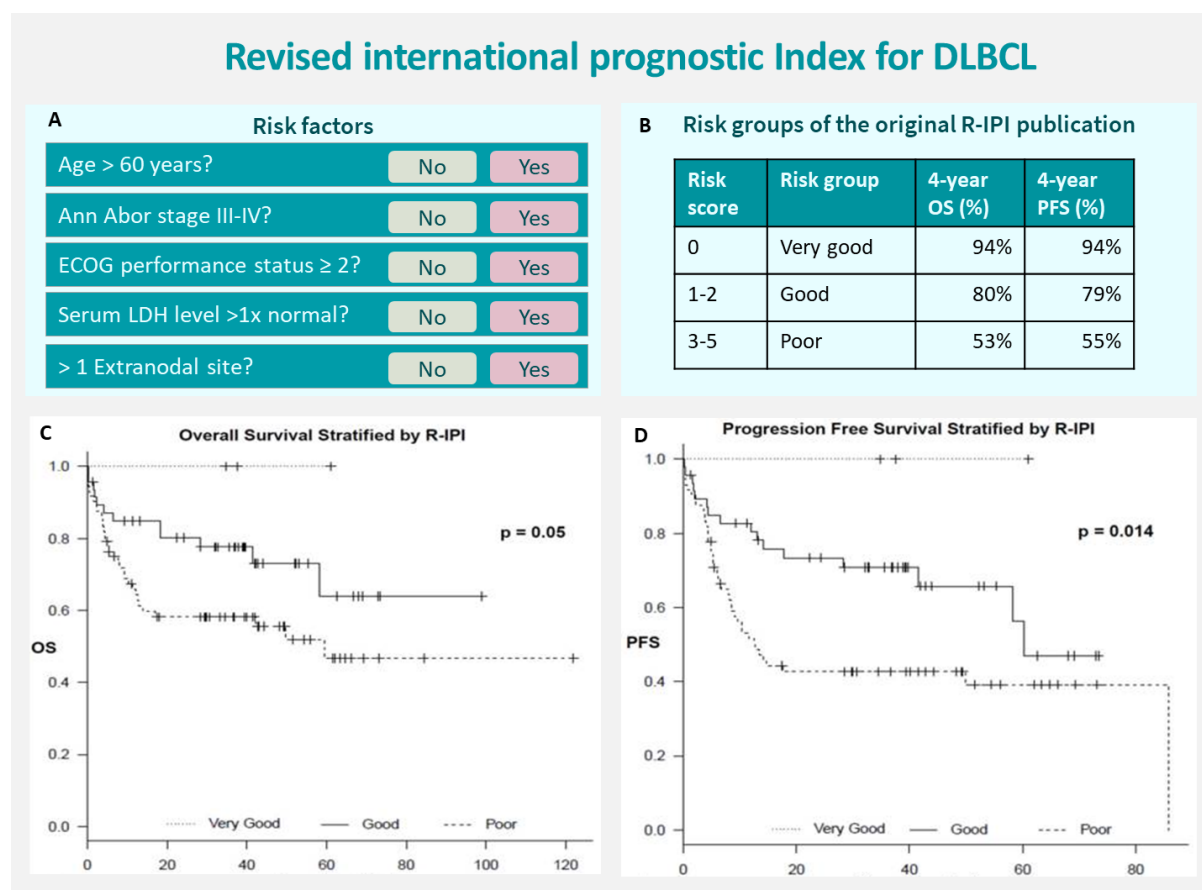


Figure 4: **A** The factors included in the revised international prognostic index for diffuse large B-cell lymphoma receiving R-CHOP – age, stage, ECOG performance status, serum LDH level, and extranodal sites (28). **B** The 4 year risk of OS and PFS per risk group after summing up the number of risk factors from **A**. **C & D** OS and PFS by revised international prognostic index risk group in an independent sample of DLBCL patients. These graphs were published as figure 2 in Alamer et al., 2020 (30) under CC BY 4.0 / Cropped from original.

In current medical literature, an abundance of prognostic models, also sometimes called prognostic index, prognostic rule, nomogram, risk algorithm, or similar, can be identified (10). To be clinically accurate and useful, prognostic models undergo three stages (Figure 5):

- a) model development,
- b) external validation,
- c) research on clinical impact.

The initial step of prognostic model research is the development of the model based on clinical necessity, which means the derivation of an algorithm from a large, reliable and representable dataset for the target population the model will be used for (10). Algorithms that can be used for model development range from simpler techniques such as multivariable regression analysis including logistic or Cox proportional hazard regression techniques (10, 16) to more complex machine learning algorithms such as neural networks, random forests, and more (16, 31). A common pitfall of prognostic models is overfitting to its original dataset, which means the model has learnt the patterns in the data so well that it can predict the outcome of interest extremely accurately in the development dataset. In contrast, it is not generalisable to other datasets due to failure in distinguishing between the general patterns across the entire target population and the random fluctuations in the data of this specific sample (32). To reduce the risk of overfitting, internal validation techniques have been proposed such as splitting the development cohort data randomly into two, using only one part for development and the other for an estimate of the model's precision (split-sample internal validation). Cross-validation is an extension of the split-sample technique where a dataset is randomly split multiple times and a model is trained and then tested on each of the split sets, resulting in an average model and model performance estimate. A more efficient approach is the bootstrapping technique, where samples of the same sample size as the original dataset are generated by replacing a percentage of the data. Then a model is developed on each bootstrap sample and averaged across (33).

The split or replaced data is usually still more similar to the model development data than the patients whom the model should be used for in the future, which may be differing in baseline risk, baseline characteristics, and setting, for example, the healthcare system. Therefore, to know whether a prognostic model is applicable to other populations or settings than those defining the model development cohort, the performance of a model has to be evaluated in independent samples, which is called external validation of a model (10, 34, 35). A model can be applied as it was presented, or be adapted to the current sample (model updating, i.e., based on the baseline risk of the new population).

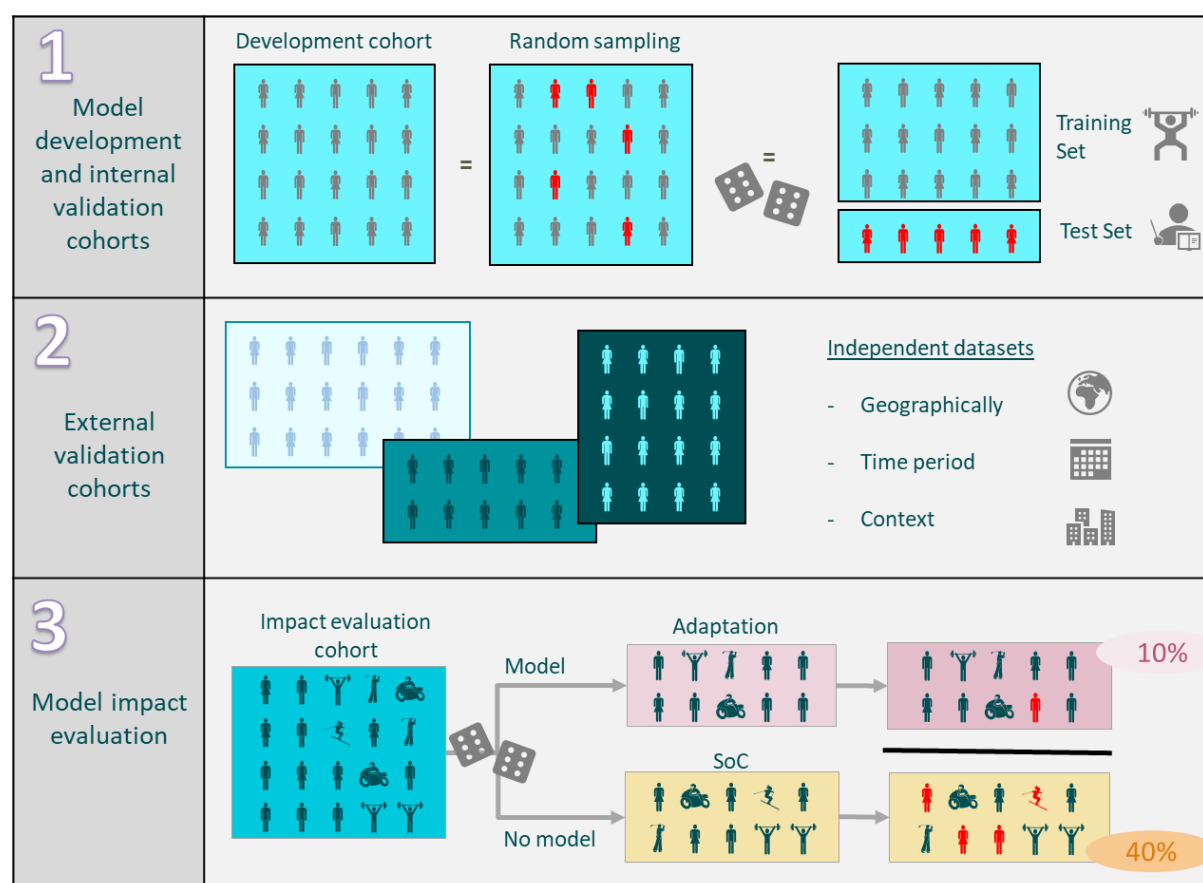


Figure 5: The three phases of prognostic model development and evaluation. First, a model is developed, preferably on a large, reliable, and representable cohort for the target population. Often, to reduce the risk of overfitting, an internal validation technique is used – here, the simplest version, split-sample validation, is illustrated. In a second step, a prognostic model should be evaluated in multiple cohorts representing the target population but different in geography, enrolment period, and context, i.e., hospital setting vs. ambulant, city vs. country side, etc. Thirdly, to finally evaluate if the prognostic model carries additional value for clinical practice, a comparison should be made between using the model (i.e., for treatment adaptation) and not using the model.

The performance of prognostic models, both in the development and the external validation cohorts they are tested with, is often represented as calibration and discrimination measures. The calibration of a model describes the agreement between a model's predictions and the observed outcomes in a cohort. It can be illustrated using a scatter plot with predictions on the x-axis and observed values on the y-axis, or a calibration plot with an intercept which represents systematic over- or underprediction, and a slope (36, 37), as presented in Figure 6. Numeric presentations are the ratios between observed and expected number of events or outcome frequencies (O:E ratios) or calibration tables (34).

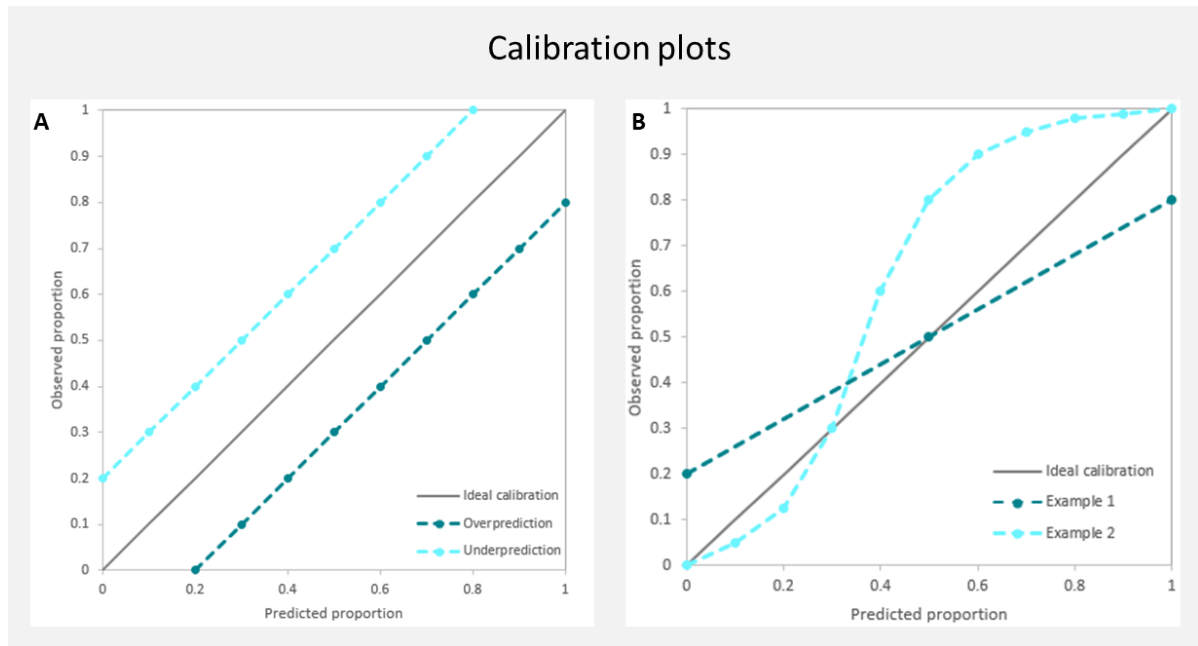


Figure 6: Examples for calibration plots. The x-axis represents the proportion of a cohort predicted with an outcome, while the y-axis represents the observed outcome frequency. **A** Systematic over- and underprediction compared to ideal calibration (grey line), with difference in intercept only. Perfect calibration has an intercept of 0, the overprediction line has an intercept of -0.2 and the underprediction line an intercept of 0.2. **B** In example 1, both the intercept and the slope differ from ideal calibration. On the left-hand side, the model underpredicts the observed frequencies, while at the higher-risk end, it overpredicts. In example 2, the intercept is 0, however, the pattern of the slope does not align with ideal prediction.

Discrimination refers to the model's ability to classify the individuals in the dataset correctly as being with or without the outcome. In dichotomous outcomes, ideally, the model predicts all individuals with an outcome as being of higher risk than any of those individuals without the outcome, which would result in perfect discrimination between the two groups. The most commonly used discrimination measure is the concordance (c-)statistic, which is a rank order statistic that is insensitive to the calibration of a model (36). It can be interpreted as the probability that in any randomly selected pair of individuals from the cohort with one experiencing the outcome and one not, the model had assigned a lower risk to the individual without the outcome (38). It is thus a statistic that can range from 0 to 1, where 0.5 stands for model predictions equal to chance and 1 for perfect discrimination. For models with binary outcomes, it is identical to the area under the receiver-operating characteristic (ROC) curve. Alternative calculations have been proposed for time-to-event outcomes (36).

When a model has shown to be generalisable to external populations, its actual impact in the clinical setting should be evaluated. It is essential to assess whether the prognostic model is able to affect the clinical management of patients, the patient outcome, or the cost effectiveness in an advantageous way, or whether it has adverse effects such as under- or overtreatment of incorrectly classified individuals, which may result in fatal effects. Testing of impact needs a comparison of a

study group that uses the model outcome and is stratified according to its risk score for the reception of treatment, and a study group without this stratification (usual care). Ideally, the impact evaluation would be conducted as a randomised controlled trial, however, it can also be tested using a before-after design (10).

One of the main issues affecting prognostic model research is the disequilibrium between model development studies, external validation studies, and model impact assessments. Currently, there are many developed prognostic models for similar questions based on questionable data sets and inappropriate data analyses. In contrast, there are less studies providing predictive performance measures of these models on external cohorts, and even fewer impact assessments, which are both necessary for a model to be transferred to clinical practice (10). Of those few validation studies, there are rarely direct comparisons of multiple models for the same population and outcome, increasing the difficulty of choosing the best model for one's setting.

A systematic review on prognostic models for COVID-19 showed that many models are developed using machine learning methods on available data sets that are not well-characterised, at high risk of bias, difficult to be applied to other data sets, and not externally validated (16). For authors wishing to externally validate a model, the lack of reporting impedes applicability assessment and proper application of the predictor weights. For systematic review authors, this means, on the one hand, less external validation studies to rate the performance of a model, and difficulties in the assessment of model's applicability to one's PICOTS question, risk of bias assessment, and meta-analysis.

Over time, even established prognostic models may lose their predictive performance due to changes in diagnoses and treatment options; additional biomarkers may be detected. In these cases, updating those existing models, i.e., recalibrating or adding a biomarker, could improve the model (10).

Stratified medicine

The fourth theme of the PROGRESS series, stratified medicine, describes the use of a person's absolute risk for stratification, or specific prognostic characteristics such as biomarkers as target for treatments (8), which relates to the impact assessment of a single prognostic factor or a prognostic model described above. On the one hand, treatment may be adapted according to the predicted risk of an individual for an outcome, i.e., individuals with the highest risk for cardiovascular diseases, and thus the largest absolute effect of the treatment, may be offered lipid lowering therapy. On the other hand, some treatments may be more effective in individuals with specific characteristics. A factor or biomarker may modify the relative effect of a treatment, which is statistically speaking an interaction between the characteristic and the treatment. It is essential to reliably identify or diagnose the

presence or absence of this factor that predicts the differential treatment response, i.e., the HER-2 status in women with breast cancer which, if positive, are prescribed the antibody trastuzumab. Both ways to stratify can be evaluated by randomised controlled trials. Furthermore, information on prognosis, the test to identify a factor, and the relative effect of the differential treatment response can serve as input for cost effectiveness decision models and healthcare services research (8).

A major difficulty of stratified medicine is that usually, clinical trials are not powered for the detection of a predictive factor, which may lead to the incorrect conclusion that a factor (or model) is not useful for stratification. At the same time, if many factors (and maybe even multiple cut points per factor) are tested, the rate of potential false positive results increases, leading to spurious results (8).

Prognosis Systematic Reviews

Similar to clinical trials of interventions, a systematic summary of all available evidence regarding a research question, be it overall prognosis, prognostic factors, models or stratified medicine, provides a more complete picture on the applicability of prognostic information than individual studies alone.

The methodology for the conduct and reporting of prognosis systematic reviews is, up to now, not yet completely established. The main steps of conducting a systematic review as outlined in Figure 1 remain the same, however, the tools used to accomplish each step change depending on the defined research question.

Defining the question

In evidence-based medicine, each study or systematic review begins with a question; this also applied to systematic reviews of prognosis studies. Just as for individual studies, the format for a prognosis question for a systematic review follows the PICOTS framework illustrated in Figure 7 (14, 39), or variations of this. The review question will define all following steps, including the search strategy, in- and exclusion criteria for screening, data extraction, the RoB assessment tool, and the meta-analytic approach.

P		Target population
I		Index factor / index model
C		Competing factor / model
O		Outcome
T		Timing
S		Setting

Figure 7: The PICOTS framework for formulating a prognosis research question (14, 34).

As in intervention research questions, the < P > for population should be clearly defined, and may also be defined as a healthy population. In contrast, the < I > stands for the index factor or model, which means the factor or model to be explored. The < C > in prognosis research is used less frequently and may involve a factor or model that the index is directly compared to. For overall prognosis, < I > and < C > are not applicable. The outcome is essential in prognosis research, as well as a predefined starting point (i.e., at diagnosis) and prediction horizon, which indicates the future time point or duration of prognostication. An additional item is the specification of the setting (14, 34).

Searching

Searches are conducted in topic-related (medical) databases such as PubMed, Embase, and domain-specific databases such as PsycINFO for psychological or psychiatric topics and CINAHL for nursing-related topics. A lack of homogeneous terminology and indexing of prognosis studies and a variety of possible study designs and data sources underlying prognosis research render the identification of prognosis studies challenging. The search strategies are usually based on the PICOTS terms but stay relatively broad, resulting in a large number of records to screen (40). Some search filters have been proposed for overall prognosis (41), prognostic factor and prognostic model research (42) to limit the number of hits. However, the resulting number of records to screen is high even after applying these filters, and supplementary search strategies such as hand searching of reference lists, forward citation

tracking, and recommended publications from experts in the field are indispensable (43). The process of searching, screening, and representation of the results is depicted in Figure 8.

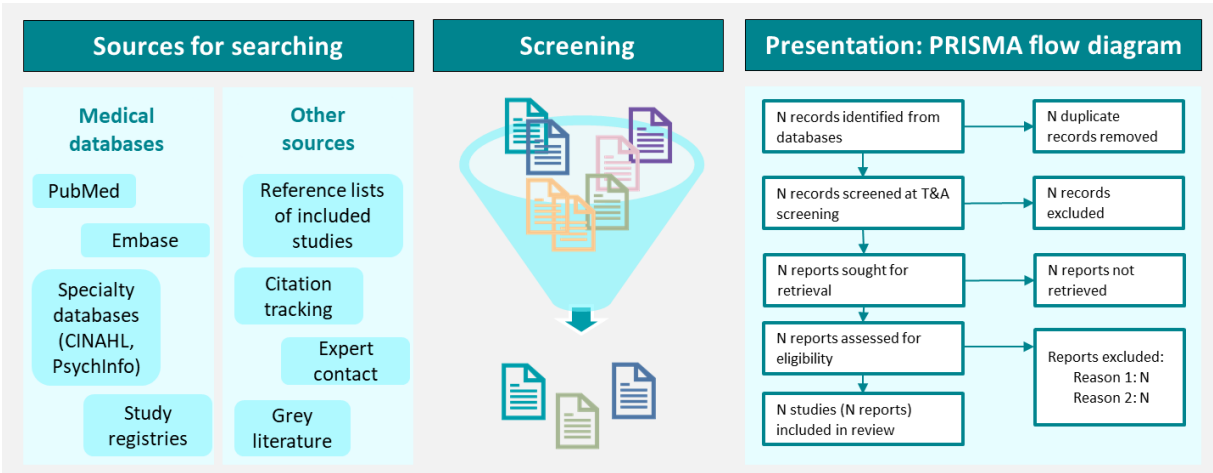


Figure 8: Sources for identifying relevant publications, screening, and presentation of the search and screening process based on the PRISMA statement (44).

Screening

Like systematic reviews of interventions, ideally, at least two review authors screen the literature independently in a two-step approach, first based on titles and abstracts followed by examination of the selected full-texts. The lack of common terminology and study designs results in challenges identifying eligible publications from titles and abstracts alone. Another difficulty is that prognostic factors may be reported as a secondary objective of a publication or in appendices only. The line between a prognostic factor and prognostic model may be thin and only be clarified seeing the full manuscript. Consequently, a higher proportion of records needs to be screened during full-text screening (40), requiring resources to be planned carefully.

Data extraction

Data extraction is the process of obtaining relevant information from the selected and included studies. This step is also done by multiple review authors independently and in duplicate, follows the PICOTS scheme and includes additional items regarding study design, data source, sample size, statistical analyses, handling of missing data, predictor selection, and performance of factors and models according to the CHARMS checklist (in an adapted version for overall prognosis and prognostic factors) (14, 39). For prognostic models, this includes information on the type of study (development or external validation) and discrimination and calibration of the model in the respective cohort. If data are missing, primary study authors may be contacted for further information (39). If multiple records belong to one study (i.e., one model), the record providing the

highest information density (i.e., longest follow-up, larger sample size, way of reporting) should be chosen for extraction.

Risk of Bias assessment

Just as systematic reviews of interventions, reviews of prognosis studies should aim to assess the limitations of each included study regarding design or analyses, which is ideally done by (at least) two review authors independently. For overall prognosis studies, no specific instrument has been developed yet for the risk of bias assessment - frequently, the tool for prognostic factors (45) or models (46) is being adapted, other general tools such as the Newcastle-Ottawa scale used (i.e., in (47), or authors created their own list of items (48, 49).

Prognostic factor studies are evaluated by the Quality In Prognosis Studies tool (QUIPS) tool (45). It contains prompting items in the six domains study participation (whether the study population is representative), study attrition (availability of study data), prognostic factor measurement, outcome measurement, study confounding, and statistical analysis and reporting, resulting in ratings of either low, moderate or high risk of bias (45).

The risk of bias instrument for prognostic model studies, applicable to both model development and external validation studies, is the Prediction model Risk Of Bias ASsessment Tool (PROBAST) (46). This instrument consists of four domains – participants, predictors, outcomes, and analysis - with multiple signalling questions per domain that can be rated as yes, probably yes, no, probably no, or no information. Each domain is rated as low, high, or unclear risk of bias (46). An exemplary representation is provided in Figure 9.

The risk of bias assessments of the individual studies of the review may be used to shape the analysis, i.e., sensitivity analysis based on risk of bias ratings, and the discussion and interpretation of the results, including the rating of the certainty in the evidence (GRADE) (see 0).

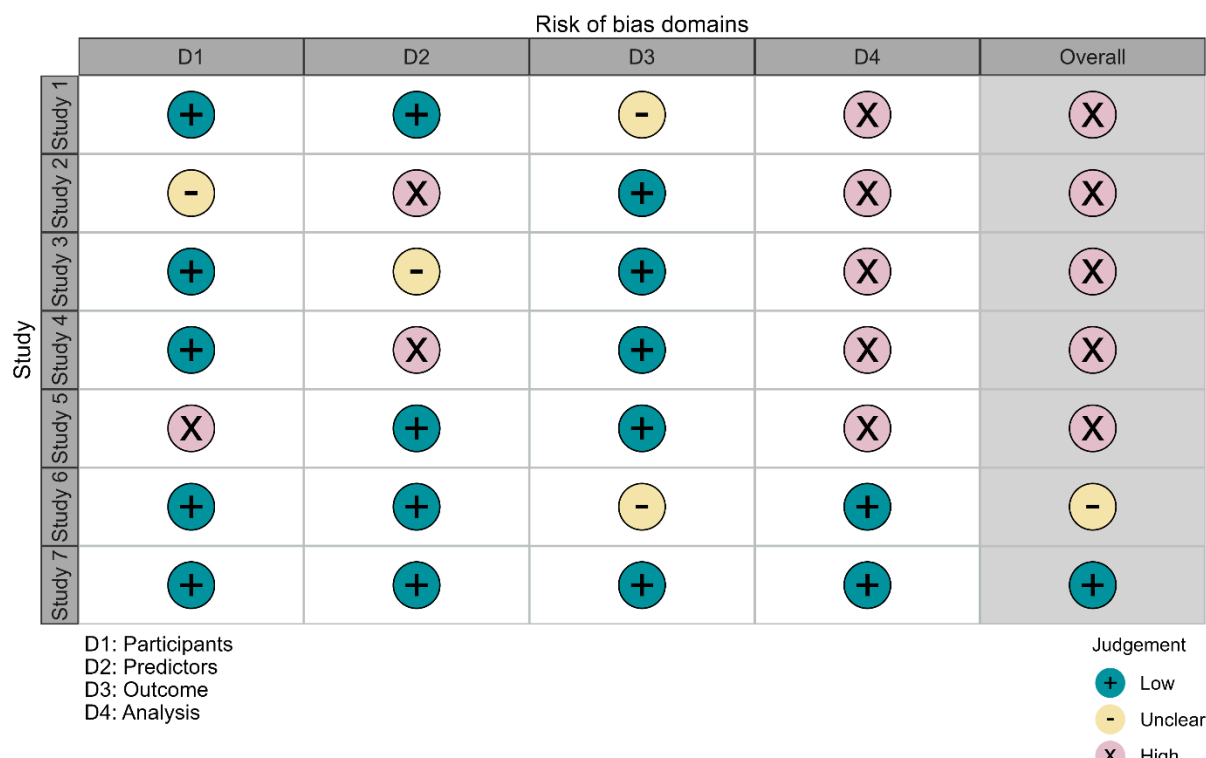


Figure 9: Typical presentation of risk of bias assessments (hypothetical data only; graph created using robvis for R). The domains in this risk of bias graph represent those of PROBAST (46), the RoB-tool for prognostic model development and external validation studies. Rows represent studies, D1 to D4 are the domains participants, predictors, outcome and analysis. Green dots indicate low, yellow unclear, and red high risk of bias. The evaluation is performed depending on the aim of the systematic review, i.e., per included prognostic model, or per external validation of one / each included prognostic model. An overall rating is based on the worst rating over all domains.

Meta-analysis

In meta-analysis, the results of multiple studies are statistically combined. For all types of prognosis studies, the random-effects model meta-analysis as compared to the fixed effects model is usually more appropriate because differences in study population, study design and conduct are unlikely to result in one underlying true effect estimate; meta-analysis should thus account for heterogeneity by estimating the between-study standard deviation (11, 34). Figure 10 illustrates a typical meta-analysis of interventions, with additional radiotherapy versus no radiotherapy after chemotherapy in early-stage Hodgkin lymphoma (fictional data, for illustration only).

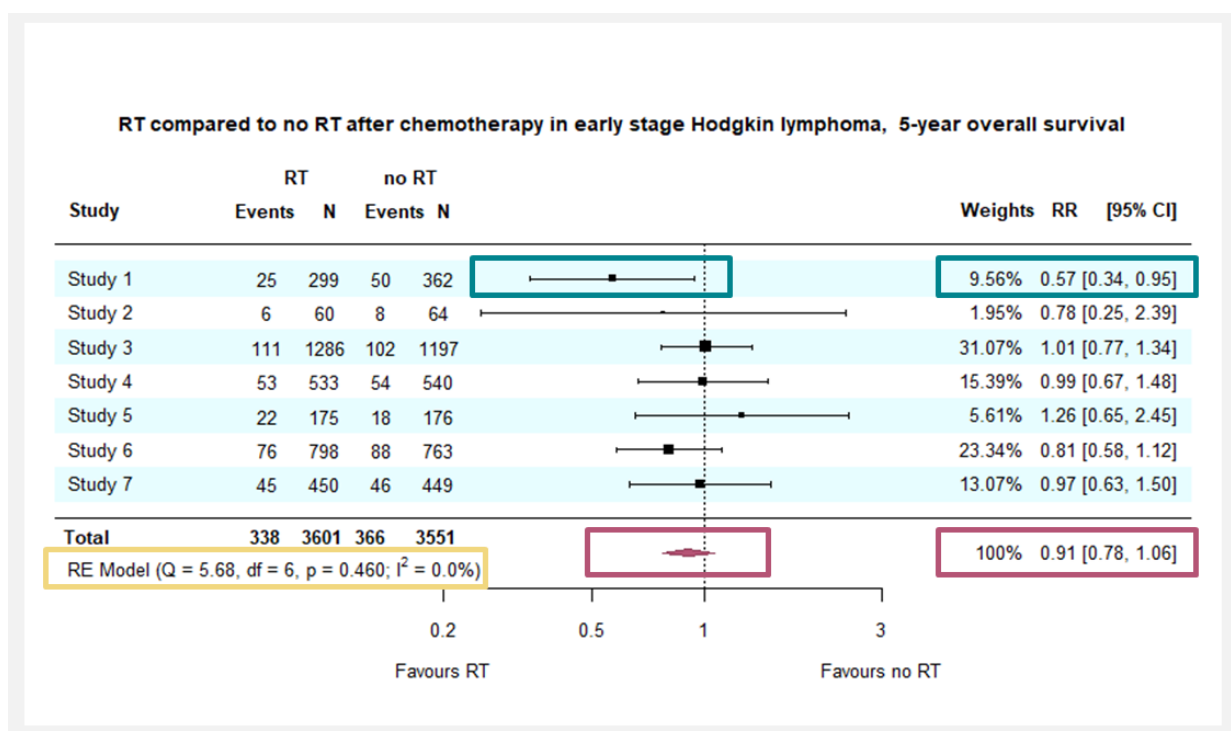


Figure 10: Forest plot with pooled effect estimate from aggregate data (fictional data, for illustration purpose only). The dark green box represents the study effect estimate (risk ratio) and 95% confidence interval of study 1, in the letter columns, study weight, risk ratio and 95% confidence interval are provided numerically. The pink boxes show the pooled effect estimate and its 95% confidence interval. The yellow box provides details regarding the model: RE: random effects model, Q: Cochran's Q test result, a test for heterogeneity; df: degrees of freedom; p: p-value; I²: indicating the percentage of variability between studies.

For overall prognosis studies, the relevant effect estimate for pooling would usually be an incidence (or absolute risk) (4). These can be pooled using the inverse-variance method and ideally the above-mentioned random-effects model. For stability of the variance, study data is arcsine or logit-transformed before meta-analysis, and transformed back for the presentation of the results as absolute risks (50).

The effect estimates and methodology for meta-analysis of prognostic factor studies is similar to those used in intervention research – (adjusted) comparative effect estimates such as risk ratios and odds ratios for dichotomous outcomes, hazard ratios for time-to-event analyses, and mean differences for continuous outcomes of primary studies can be pooled with the standard methodology if sufficiently homogeneous regarding the PICOTS, i.e., factor cut points, timings, etc. (11). Meta-analysis of prognostic factor studies would thus look like the meta-analysis in Figure 10, depending on the relevant effect estimate.

For meta-analysis of prognostic models, the predictive performance measures of external validation studies are summarised per prognostic model. This means two separate meta-analyses per prognostic model under evaluation, one for the discrimination of the model by combining the c-statistics, and one for the calibration of the model using the intercept and slope of the calibration

plots. If these are not available, an observed versus expected ratio can be calculated from the number of observed and expected events or frequencies as a presentation of a model's overall calibration in each external cohort, which can subsequently be meta-analysed by using the logit transformation (34, 51).

Grades of Recommendation, Assessment, Development, and Evaluation (GRADE)

Besides providing a summary of the evidence regarding a defined research question, the task of a systematic review is to judge whether the results indicated by the identified body of evidence are trustable, in other words, whether we are confident the results present the true effect estimate, or in this case, the true prognosis (52). The Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) framework provides a transparent framework consisting of five domains to rate the confidence in one's results - risk of bias, imprecision, inconsistency, indirectness, and publication bias - resulting in a rating of high, moderate, low, or very low certainty in the evidence, as illustrated in Figure 11. These domains can also be applied to prognosis research (52). While guidance has been developed for overall prognosis and prognostic factor research (52, 53), GRADE for prognostic models is currently being developed and still in the conceptual phase (54, 55).

In contrast to intervention studies, the ideal study design of overall prognosis, prognostic factor, and prognostic model research is a longitudinal, prospective cohort study. The ratings thus start at high confidence for observational studies, and are rated down for concerns in each domain. Three reasons may mandate rating the evidence up: a plausible dose-response relationship, large effect, or plausible confounding (52, 53). The rating in the risk of bias domain would depend on the respective risk of bias assessment tool outlined in paragraph 0. If risk of bias is high for all included studies, one would rate down for risk of bias. Mixed ratings warrant a sensitivity analysis where possible; if differences arise, the body of evidence at low risk of bias can be used without downgrading, similar to the standard GRADE approach (52).

Inconsistency is based on a judgment of unexplained heterogeneity between studies, taking into consideration both point estimates and 95% confidence intervals, while the domain imprecision is rated based on the width and position of the 95% confidence interval of the pooled estimate. In prognosis, one would downgrade if the clinical course or treatment of a patient would be different while having a prognosis at the lower bound of the confidence interval as compared to a prognosis close to the upper bound of the confidence interval (52), or when a predefined decision threshold around an absolute risk difference has been crossed by the point estimate or confidence interval (53).

Indirectness describes the applicability of a study population, factor or model, outcome, timing and setting of the study with the question of the systematic review (53, 56). A study population may be more selective, a factor or factors in a model measured in different ways, outcome definitions or

measurements may have changed, or the timing varies. In addition, changes in setting, such as the general healthcare context including insurance policies and differences in clinical care may affect the baseline risks and prognosis of the outcome in question.

The fifth domain, publication bias, is challenging to assess for prognosis research. A funnel plot may help to detect small-study effect for prognostic factors, which can be caused by publication bias (53). However, Huguet et al. suggest to assume publication bias except evidence that shows contrary, i.e. large number of studies that have intentionally collected data to examine the prognostic factor in question (20). Publication bias may be less problematic for prognostic model research, whose GRADE guidance is not yet finalised.

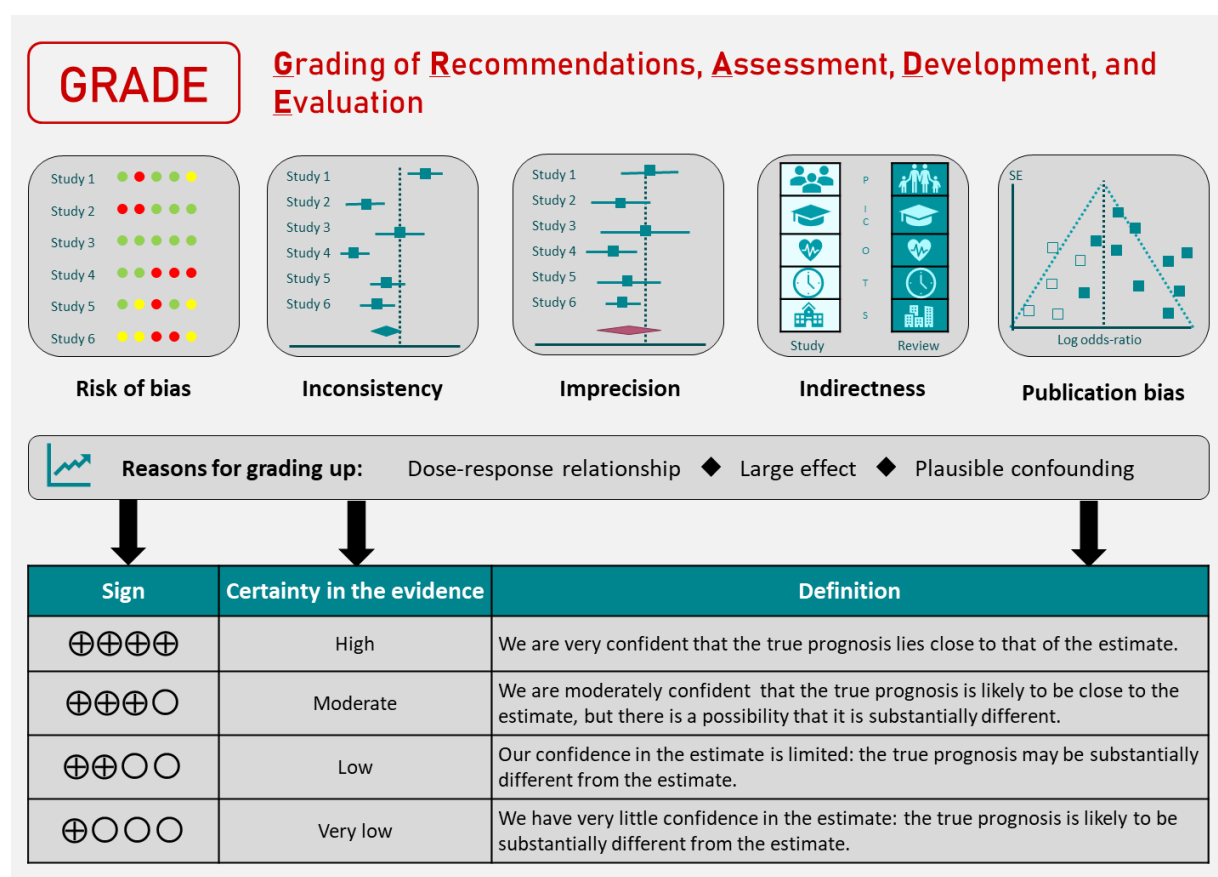


Figure 11: The Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) framework, illustrated for prognosis research adapted from Balshem et al. 2011 (57) and Iorio et al. 2015 (52).

Prognosis in Haematological Malignancies

Despite recent improvements of diagnostic procedures and advancements in treatments of cancer, the course of haematological malignancies can, depending on the type of malignancy, be highly variable. Even relapsed or refractory malignancies may be cured, while sometimes, there is a quick decline in the condition of the patient, resulting in a fast transition to palliative care (58). To many

patients, knowing one's prognosis is important and forms the basis for informed decision making. It allows balancing the harms of a treatment option that may be highly toxic against its benefits, considering to the patient's own preferences regarding extension of life versus quality of life. Although some patients prefer to not know their prognosis, for haematologists, prognostic information is essential to define the patient's medical care (58).

Hodgkin lymphoma (HL) is a rare malignancy of the lymph system, the B-cells in particular, with an incidence between 2 to 3 per 100,000 inhabitants in high-income countries (59). It is staged into four categories using the Ann Arbor classification based on lymph node involvement and organ involvement (60). Its cure rates are with up to 90% high (61). However, multiple factors for a poor prognosis such as large mediastinal mass, extra nodal lesions, three or more involved areas, and bulky disease, have been identified that can be used to classify HL into early favourable, intermediate (early unfavourable) and advanced stage disease (62). Early-stage disease is currently being treated by the chemotherapy regimen adriamycin, bleomycin, vinblastine and dacarbazine (ABVD) plus involved-field radiotherapy. An intensified regimen, BEACOPP (bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine and prednisone), is available for advanced stage HL (Borchmann 2011, Engert 2012). High dose chemotherapy followed by autologous stem cell transplantation is the first treatment option for relapsed or refractory HL (63).

One prognostic factor in addition to those that has been used for the classification above is interim positron emission tomography (PET) during chemotherapy. PET is an imaging tool that marks the tumour's metabolic activity using a radio-labelled glucose analogue, [18F]-fluorodesoxy-D-glucose (FDG), which emits positrons that are detected by scintigraphy (64). While it has been used for diagnosis and staging, numerous studies have also evaluated FDG-PET as a prognostic factor (65-67). Subsequently, randomised controlled trials were set up to test treatment adaptation, for example the study H10 (68). The current dissertation project aimed to systematically review both interim FDG-PET's role as prognostic factor in HL (PROGRESS theme 2, publication 1) and its potential to stratify patient's treatment course according to interim FDG-PET results (PROGRESS theme 4, publication 3).

Another haematological malignancy, chronic lymphocytic leukaemia (CLL), has a high uncertainty in its clinical course. This malignancy of the lymphatic system affects mostly the elderly and shows as an accumulation of mature B cells in blood, bone marrow, lymph nodes and spleen (69). Some people do not show any or only few symptoms for many years and reach a comparable length of life as their healthy counterparts. Others progress shortly after diagnosis and may need aggressive treatment (REF). The initial staging systems that distinguish between early, intermediate, and advanced stage were developed by Binet and Ray in 1981 and 1975 (70, 71). However, those staging

systems are not sufficient to reliably prognosticate the course of the disease, as survival times can vary largely between patients within the same disease stage.

Recently, multiple prognostic factors have been identified that may predict the outcome of CLL patients, such as gene abnormalities that affect the protein production of p53, mutational status of the immunoglobulin heavy chain variable segment (IgHV), and more (72). However, an individual factor cannot determine the treatment course. It is therefore more promising to combine multiple of those prognostic factors in an algorithm for more precise predictions of the disease course and choice of treatment. In recent literature, several prognostic models have emerged to predict a variety of outcomes in patients with CLL. The current dissertation project aimed to systematically collect all prognostic models for newly diagnosed CLL patients, and to meta-analyse their respective external validation studies to know their performance and transferability to other settings.

Aim and Objectives

The aim of this doctoral project was to explore the methodology and develop exemplary work regarding systematic reviews of prognosis in the field of haematology within Cochrane as international gold standard for systematic review development. This cumulative dissertation covers three of the four outlined themes of the PROGRESS framework - the contribution to a systematic review on prognostic factors, the development and publication of the first prognostic model review in the Cochrane library, and a systematic review on the application of a prognostic factor for stratified medicine using randomised controlled trials as study design.

PROGRESS theme 2 – prognostic factors

Publication 1: Interim PET-results for prognosis in adults with Hodgkin lymphoma: a systematic review and meta-analysis of prognostic factor studies

The dissertation project started by participation as a co-author in one of the first Cochrane systematic reviews on prognostic factors with the clinical question whether interim positron emission tomography (PET) during chemotherapy is a valuable prognostic factor for predicting the overall survival and progression-free survival in patients with newly diagnosed Hodgkin lymphoma with a time horizon of at least 12 months. As this was one of the first systematic reviews on prognostic factors within the Cochrane library, it was conducted in collaboration with the Cochrane Prognosis Methods group.

A summary of the systematic review, including a visual abstract, is presented in the results section of this manuscript, while the full publication is attached as Appendix II: Publication 1.

PROGRESS theme 3 – prognostic models

Publication 2: Prognostic models for newly-diagnosed chronic lymphocytic leukaemia in adults: a systematic review and meta-analysis

The second publication that forms part of this dissertation project was the conduct of the first Cochrane systematic review on prognostic models. The scientific question was to collect all prognostic models available for previously untreated chronic lymphocytic leukaemia and their respective external validation studies. Eligible outcomes were overall survival, progression-free survival and treatment-free survival, external validation studies had to validate the model for the same outcome a model was developed for. There was no restriction regarding timing or setting.

A summary of the systematic review, including a visual abstract, is presented in the results section of this manuscript, while the full publication is attached as Appendix III: Publication 2.

PROGRESS theme 4 – stratified medicine**Publication 3: Positron emission tomography-adapted therapy for first-line treatment in adults with Hodgkin lymphoma**

The third publication covers a topic fitting into the framework of stratified medicine. The scientific question was whether the prognostic factor interim-PET after multiple cycles of chemotherapy, examined in publication 1 of this dissertation project, can be utilised to stratify Hodgkin lymphoma patients for an adapted treatment course, i.e., de-escalating treatment for lower risk patients, or intensifying (escalating) treatment for higher risk patients. It was restricted to randomised controlled trial data. The systematic review was an update of a Cochrane systematic review first published with only three included studies in 2015.

A summary of the systematic review, including a visual abstract, is presented in the results section of this manuscript, while the full publication is attached as Appendix IV: Publication 3.

Systematic Reviews of Prognosis in Haematological Malignancies - Project Overview

Background



Prognosis describes the prediction of a future health state or course of disease, which is essential for patients, health care providers and policy makers to determine further treatment options and prioritise research and implementation. Prognosis research questions are divided into four themes: overall prognosis, prognostic factors, prognostic models, and stratified medicine. Prognosis research, similar to intervention research, can be summarized using systematic reviews. However, the methodology for prognosis systematic reviews is not yet fully developed.

Objective

To develop exemplary work on prognosis systematic reviews in the specialty of haematology within Cochrane, covering three of the four themes.

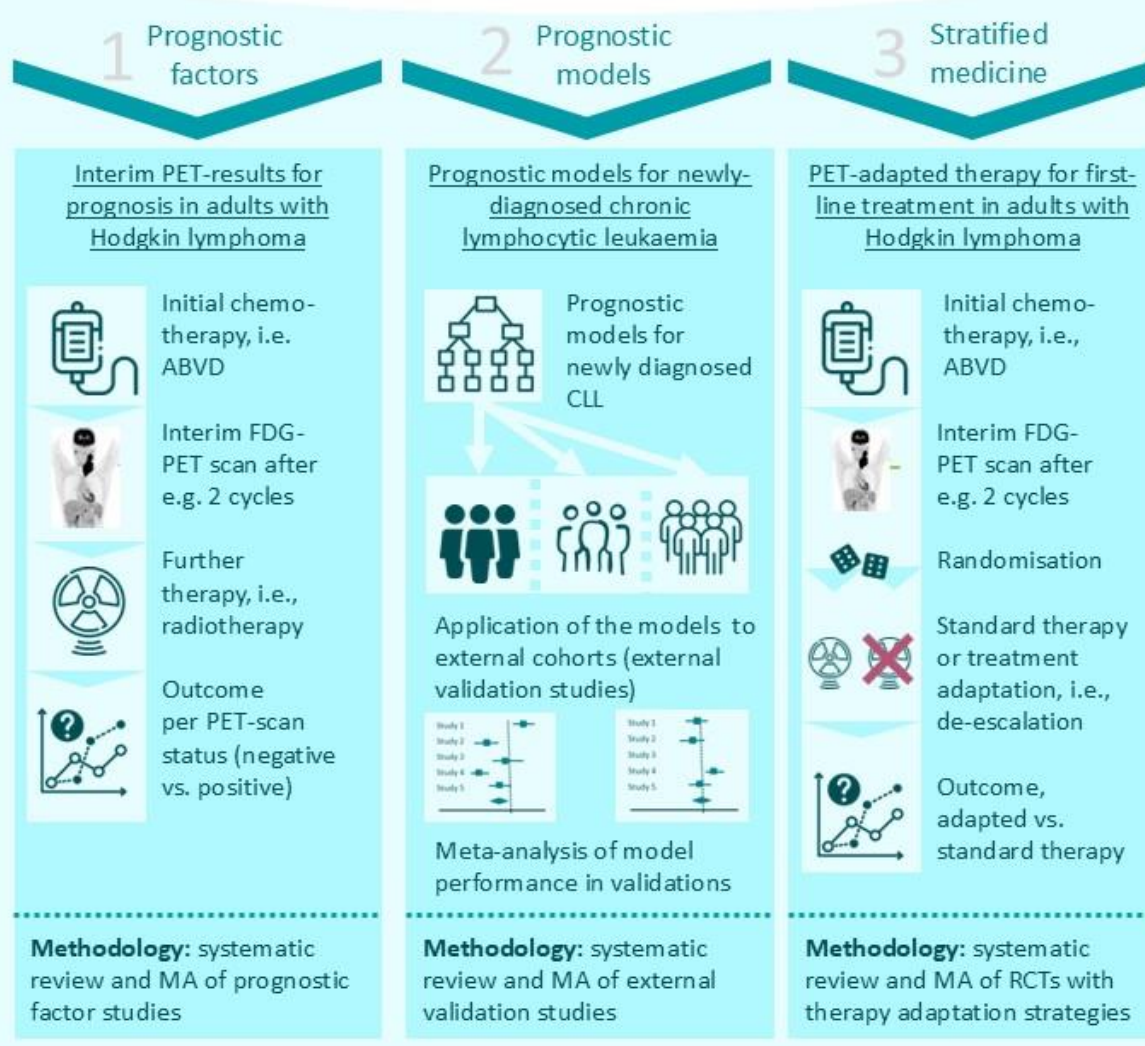


Figure 12: Graphical overview of the objectives of the three publications included in this dissertation project (PET-scan in publication 1 on prognostic factors was adapted from figure 3b in Picardi et al., 2021 (66) [published under CCA 4.0 International License]. Icons by Freepik, more attributions can be found in the section "Attributions of Icons")

Results: Dissertation Project Synopses

Publication 1: Interim PET-results for prognosis in adults with Hodgkin lymphoma: a systematic review and meta-analysis of prognostic factor studies

This systematic review of interim-PET as prognostic factor for overall and progression-free survival in HL was published in January 2020 in the Cochrane Library of Systematic Reviews. The dissertation manuscript provides a comprehensive summary and a visual abstract of this systematic review in the main text. The published manuscript, due to its length, is attached as Appendix II: Publication 1.

- 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
- Co-authorship
- 2-year Impact Factor of the Cochrane Library of Systematic Reviews (2020): 9.28
- This review received direct funding for its conduct from the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, Grant no. 01KG1709), Germany.

Summary

Note: The summary contains excerpts from the published manuscript, which is provided in the appendix.

Introduction

Hodgkin lymphoma is a rare but highly curable cancer (61, 73) of the lymphoid system that can involve other organs such as liver, lungs, bone, or bone marrow (60). HL is divided into four tumour stages by the Ann Arbor classification, where stage one to three describes the lymph node and localised organ involvement, while stage four indicates disseminated organ involvement (74). Standard treatment consists of a combination regimen of chemotherapy (ABVD) and radiotherapy for early stages, and an intensified chemotherapy regimen (BEACOPP) for advanced stages (75).

FDG-PET is an imaging tool that can indicate the metabolic activity of a tumour. The vector 2 deoxy-D-glucose, that carries the positron-emitting nuclide ¹⁸F, is taken up by the cells and can thus be detected by scintigraphy (64). Previously, PET and PET in combination with computed tomography (CT) was mostly used for staging, but has become more prominent in also determining the end-of-treatment as well as interim tumour activity during therapy (76). This interim PET has been evaluated in multiple individual studies as prognostic factor for survival outcomes (77, 78), and some

prospective studies and RCTs had been set up to explore its potential for treatment adaptation (systematically summarised in Sickinger et al, 2015 (79) and its update, see publication 3 of this thesis). Interim PET after multiple cycles of chemotherapy as first-line therapy of HL as prognostic factor had not yet been systematically summarised using methodology of prognosis systematic reviews.

Objectives

To determine whether in adults with 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.

Methods

Eligibility criteria

Eligible study designs were any retrospective or prospective cohort studies with more than 10 individuals, including RCTs, as long as they did not modify treatment based on interim PET.

P	- Confirmed diagnosis of classic HL - Any stage, all subtypes - First-line treatment / not previously treated
I	Interim FDG-PET scan (negative, positive)
C	Not applicable, no comparative factor examined
O	- Overall survival (OS) - Progression-free survival - Adverse events
T	- Prognostic factor: after one, two, three, or four cycles of chemotherapy - Outcome: prediction horizon minimum 12 months
S	Hospital, treatment centre

Table 1: Eligibility criteria for publication 1 in PICOTS format: interim PET as prognostic factor for Hodgkin lymphoma

Search methods

MEDLINE, Embase, the Cochrane Central Register of Controlled Trials (CENTRAL), and clinicaltrials.gov were searched up to 25 January 2019. No specific search filters for prognosis research were used, resulting in a sensitive but unspecific search strategy. Reference lists of included studies and relevant review articles were checked for relevance.

Data collection

Two review authors independently screened the resulting records, extracted data into a piloted data extraction form, rated the risk of bias according to the QUIPS tool for prognostic factor research (45),

and used the GRADE framework for prognostic factors to rate the certainty in the evidence (53). Any discrepancies at any step were discussed between the two review authors, and if necessary, a third author was consulted. Where data was missing, primary study authors were contacted for more information.

Data analysis

Meta-analysis of multiple studies was undertaken where studies were clinically sufficiently homogeneous, i.e., provided similar treatment regimen etc. As covariates were variable across studies, only unadjusted hazard ratios were meta-analysed using random effects models with the inverse variance methods. If HRs were not reported, it was attempted to recalculate these from available data using the methods proposed by Tierney et al (80).

Main results

The systematic review identified 23 studies as eligible, including 7335 participants of which 2205 could be meta-analysed. Fifteen studies included all stages of HL, while four included only early stages and another four included only advanced stages. The majority of studies provided ABVD as first-line treatment (N=16) and conducted the interim PET scan after two cycles of chemotherapy (N=21). Sixteen studies used combined PET-CT. Median follow-up times across all studies ranged from 22 to 65 months.

Twelve studies reported OS. Of those, nine studies including 1802 participants were sufficiently homogeneous for pooling. The evidence suggests that participants with a negative interim PET or PET-CT scan have probably a large survival benefit compared to those with a positive interim PET or PET-CT scan (unadjusted HR 5.09, 95% CI 2.64 to 9.81; moderate-certainty evidence). The evidence was downgraded by one level due to serious risk of bias (mostly due to potential confounding or statistical analysis) and one level for imprecision due to re-estimation of the effect estimates. The large effect justified upgrading by one level. Adjusted effect estimates from two studies confirm the unadjusted pooled result.

Overall, 21 studies reported PFS. Of these, 14 studies with 2079 participants were pooled. The evidence suggests that those with a negative interim PET or PET-CT scan may have increased PFS compared to those with a positive interim PET result; however, the evidence is very uncertain (unadjusted HR 4.90, 95% CI 3.47 to 6.90, very low-certainty evidence). The evidence was graded down by one level for serious risk of bias (mostly due to potential confounding or statistical analysis), one level for serious inconsistency due to varying definitions of PFS, and one level for imprecision due to re-estimation of effect estimates in nine studies. Multivariable effect estimates from eight studies confirm the potential role of interim PET scan as prognostic factor for PFS (low certainty evidence).

Data on PET-associated adverse events were not reported.

ConclusionsImplications for practice

Interim PET or PET-CT scan is probably a good prognostic factor for overall survival, although evidence is less certain for progression-free survival. Interim PET scans might thus serve or prognosis and, after further evaluation, potentially decide the further treatment course.

Implications for research

Although evidence suggests that interim PET-CT scan is probably a prognostic factor for OS in first-line treated HL, this has not been shown in conjunct with other prognostic factors. Its independent role has still to be established. Interim PET might also be used as part of a prognostic model to stratify participants. Furthermore, final results of RCTs with treatment adaptation based on interim PET might show its value in changing clinical practice.

Visual abstract

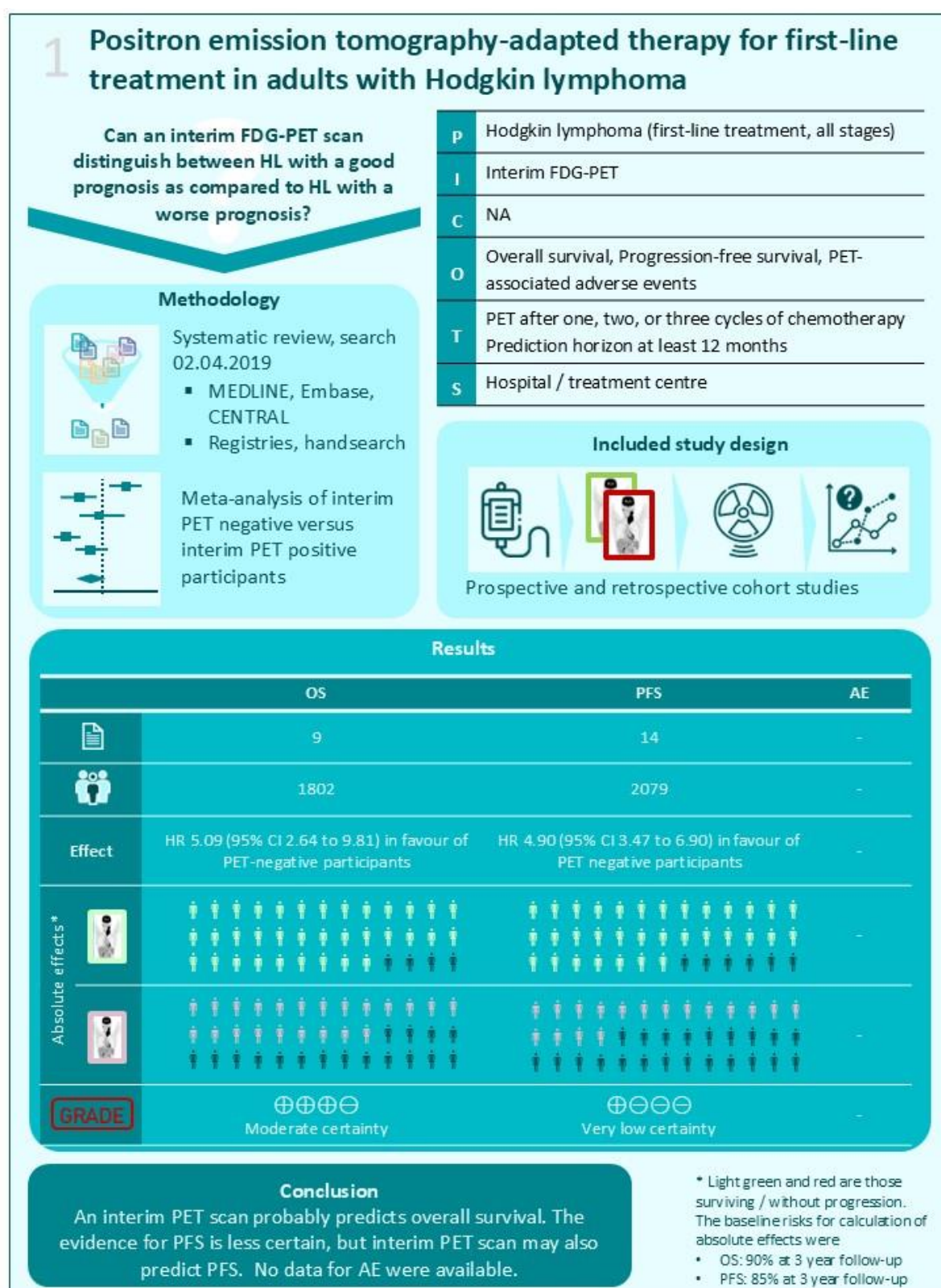


Figure 13: Visual abstract for the publication Aldin et al., 2020 (81) – “Interim PET-results for prognosis in adults with Hodgkin lymphoma: a systematic review and meta-analysis of prognostic factor studies” (the PET-scan was adapted from figure 3b in Picardi et al., 2021 [66] [published under CCA 4.0 International License]. Icons by Freepik, more specific attributions can be found in the section “Attributions of Icons”)

Publication 2: Prognostic models for newly-diagnosed chronic lymphocytic leukaemia in adults: a systematic review and meta-analysis

This systematic review on prognostic model development and external validation studies in chronic lymphocytic leukaemia was the first prognostic model review published in the Cochrane Library of Systematic Reviews in July 2020. This manuscript presents a comprehensive summary and a visual abstract in the main text. The published manuscript, due to its length, is attached as Appendix III: Publication 2.

- Kreuzberger N, Damen JAAG, Trivella M, Estcourt LJ, Aldin A, Umlauff L, Vazquez-Montes MDLA, Wolff R, Moons KG, Monsef I, Foroutan F, Kreuzer K-A, Skoetz N. Prognostic models for newly-diagnosed chronic lymphocytic leukaemia in adults: a systematic review and meta-analysis. Cochrane Database of Systematic Reviews 2020, Issue 7. Art. No.: CD012022. DOI: 10.1002/14651858.CD012022.pub2.
- First authorship
- 2-year Impact Factor of the Cochrane Library of Systematic Reviews (2019): 7.89
- This review received direct funding for its conduct from the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, Grant no. 01KG1711), Germany.

Summary**Introduction**

In CLL, mature B cells accumulate in blood, bone marrow, lymph nodes, and spleen (REF). The diagnosis is based on blood counts, differential counts, blood smears, and immunophenotyping (Hallek 2017). It constitutes 25% of all leukaemia and affects mainly in the elderly (Chiorazzi 2005). CLL can be staged using either the Rai (Rai 1975) or Binet (Binet 1981) staging systems, which include the number of lymphocytes in the peripheral blood, presence of enlarged lymph nodes, anaemia or thrombocytopenia, and enlarged liver or spleen. However, the prognostic value of the two staging systems is limited, survival times within a stage can vary significantly. While some individuals with CLL have no or only few symptoms, and have a life expectancy comparable to a healthy person without the need of treatment, others experience a more aggressive course and die within a few years despite treatment.

In the past years, numerous prognostic factors have been identified for CLL, including genomic aberrations, gene abnormalities, mutational status of IgHV, or surrogate markers for these factors (Döhner 2000, El Roubi 1993, Kay 2007). This opens the opportunity to develop prognostic models by combining those factors in a weighted manner. Up to now, the use of single prognostic factor or

prognostic models is controversial, and no systematic review to summarise all available models and their performance in external validation studies had been conducted.

Objectives

One objective of this systematic review was to identify, describe and appraise all prognostic models for the prediction of OS, PFS and treatment-free survival (TFS) in newly diagnosed CLL. Secondly, each model's predictive performance was to be examined by meta-analysing their respective external validation studies.

Methods

Eligibility criteria

This systematic review included both development and external validation studies of prognostic models, including publications that present both development and external validation(s). A prognostic model was defined as any mathematical function with two or more independent prognostic factors that aimed to calculate individualised predictions, in contrast to studies that aim to establish whether one factor is an independent prognostic factor among others in a multivariable model.

P	Individuals with confirmed chronic lymphocytic leukaemia, untreated
I	Any model developed for the population, and their corresponding external validation studies (validated for the same outcome as the original model)
C	Not applicable, no direct comparison of models planned
O	- Overall survival - Progression-free survival - Treatment-free survival
T	- At diagnosis of CLL - Time horizon any time after diagnosis
S	Not specified

Table 2: Eligibility criteria in PICOTS format for publication 2: prognostic models for CLL.

Search methods

MEDLINE (via Ovid) and Embase were searched up to 24 June 2019, and clinicaltrials.gov and the WHO ICTRP were searched up to 5 March 2020. We used a search filter by Geersing 2012 (42). We did not search conference proceedings due to their limited reporting. Reference lists of included studies and relevant review articles were checked for relevance.

Data collection

Each resulting record was screened by at least two review authors independently in a two-step approach (titles and abstracts, full-texts). Due to the complexity of the screening, we organised regular author team meetings online. Main discussion points can be found in the published manuscript (Appendix III: Publication 2: Prognostic models for newly-diagnosed chronic lymphocytic leukaemia in adults: a systematic review and meta-analysis). At least two review authors independently extracted data into a piloted, standardised data extraction form, and assessed the risk of bias with the PROBAST tool for prognostic models (46). Any disagreements were solved by group discussion.

Data synthesis

We analysed the external validation studies of each model that had at least three validations for the outcome it was developed for. Random-effects models were applied to pool the logit transformed c-statistic using R (Debray 2014, Debray 20018b, Vichtbauer 2010). The model development study performance was not included in this meta-analysis, as it would overestimate the performance. It was planned to meta-analyse the calibration as observed – expected ratios. However, these were rarely reported. Therefore, in collaboration with the Cochrane prognosis methods group, illustrations of survival or progression-free survival per risk category were calculated and graphically presented.

Main results

The review includes 52 prognostic model development studies; twelve of these were externally validated for their primary outcome in 35 external validation studies. Forty-one models were thus not at all externally validated and are described in more detail in the appendix of the main manuscript in Appendix III: Publication 2. An overview of the twelve models with any validation studies is provided in Table 3.

Model	N° of predictors	Predictors	Primary outcome	External validation studies
CLL-IPI (82)	5	- TP53 status - IgHV mutational status - Serum B2-microglobulin - Clinical stage (Rai or Binet) - Age	OS	11 cohorts 10 validations for OS 8 validations for TFS
Barcelona-Brno model (83)	3	- IgHV mutational status - del(17p) - del(11q)	OS	6 cohorts 5 for OS 6 for TFS
MDACC 2007 (84)	6	- Age - Absolute lymphocyte count	OS	10 cohorts 8 for OS

		- Serum B2-microglobulin - Nodal groups - Rai stage - Gender		6 for TFS
GCLLSG model (85)	8	- Gender - Age - ECOG PS - del17p - del11q - IgHV mutational status - Serum thymidine kinase - Serum B2-microglobulin	OS	3 cohorts 1 for OS 2 for TFS
Rossi 2013 (86)	5	- TP53 - BIRC3 DIS - SF3B1 - NOTCH1 d - el11q22-q23	OS	2 cohorts
Stephens 2015 (87)	3	- ECOG PS - Age - Lactate dehydrogenase	OS	1 cohort
Stephens 2015 (87)	5	- ECOG PS - Rai stage - Age - white blood cell count - del11q22.3	TFS	1 cohort
Baliakas 2019 (88)	3	- TP53 abnormality - Trisomy 12 - Stereotyped subset #2 defined as IGHV3-21/IGLV3-21 BcR IG	TFS	1 cohort
GIMEMA model (89)	4	- Lymphocyte doubling time - Absolute peripheral blood lymphocytosis - Rai stage - Gender	PFS	1 cohort
MDACC 2011 (90)	6	- IgHV mutational status - diameter of largest palpated lymph node - FISH category (del11q or del17p versus none) - number of involved lymph node sites - lactate dehydrogenase - IgHV mutational status and LDH interaction term	TFS	1 cohort
Morabito (91)	3	- IgHV mutational status - CD38 - ZAP-70	TFS	1 cohort
O-CLL1 model (92)	4	- Rai stage, - Absolute lymphocyte count - serum B2-microglobulin	TFS	2 cohorts

- IgHV mutational status

Abbreviations: ECOG PS: Eastern Cooperative Oncology Group performance status; IgHV: Immunoglobulin Heavy Chain Variable; LDH: lactate dehydrogenase

Table 3: Summary of prognostic models for CLL developed and published up to June 2019 that were externally validated at least once.

The most common issues in those twelve development studies were missing data handling, the selection of predictors (based on univariable analysis), dichotomisation of factors, a lack of correction for overfitting in the apparent performance indicators, and simplified model weights, resulting in a loss of information.

Three models were validated three times or more, and their performance was meta-analysed. Information on all models can be found in the main manuscript; this synopsis will describe those three models in more detail.

CLL-IPI

The CLL-IPI (82) was based on a stepwise proportional hazards model with univariable factor preselection and dichotomisation of prognostic factors. It is presented as point score (see Table 4). Risk of bias for the model development was rated as low for the domains participants, predictors, and outcomes, and high for the analysis domain due to univariable factor selection, dichotomisation of factors, and the handling of missing data.

Prognostic factor	Point distribution
Age (≥ 65 years)	1
Clinical stage (Rai I-IV or Binet B-C)	1
IgHV mutational status (unmutated)	2
B2-microglobulin (> 3.5 mg/L)	2
TP53 status (deleted or mutated)	4

Table 4: Point distribution of the CLL-IPI per prognostic factor (simplified model)

Ten studies externally validated the CLL-IPI. Risk of bias across those was mixed in the participant domain, the predictor and outcome domains were rated predominantly as low, and the domain analysis predominantly as high. The Calibration of the model was not reported in the external validation studies. Thus, survival per risk group was pooled and presented with the original survival percentages (Table 5) based on eight studies with 4891 individuals.

Risk group	Risk score	Development study: percentage of persons surviving at 5 years (95% CI)	Pooled percentage of persons surviving at 5 years (95% CI)	95% prediction interval (PI)
Low risk	0-1	93.2% (90.5% to 96.0%)	92.5% (89.2% to 94.8%)	82.5% to 97.0%
Intermediate risk	2-3	79.3% (75.5% to 83.2%)	85.0% (79.7% to 89.1%)	68.8% to 93.6%
High risk	4-6	63.3% (57.9% to 68.8%)	64.9% (56.4% to 72.6%)	41.9% to 82.6%
Very high risk	7-10	23.3% (12.5% to 34.1%)	40.4% (29.3% to 52.6%)	13.2% to 72.4%

Table 5: Five-year survival per risk group as defined by the CLL-IPI development study in the training cohort, and pooled 5-year survival from eight external validation studies.

Seven external validations including 3307 participants and 337 events studies reported the discrimination of the CLL-IPI. The pooled c-statistic was 0.72 (95% CI 0.67 to 0.77), with some degree of heterogeneity (between-study standard deviation, $\tau = 0.21$). The 95% PI ranged from 0.59 to 0.83 (Meta-analysis provided in Figure 14).

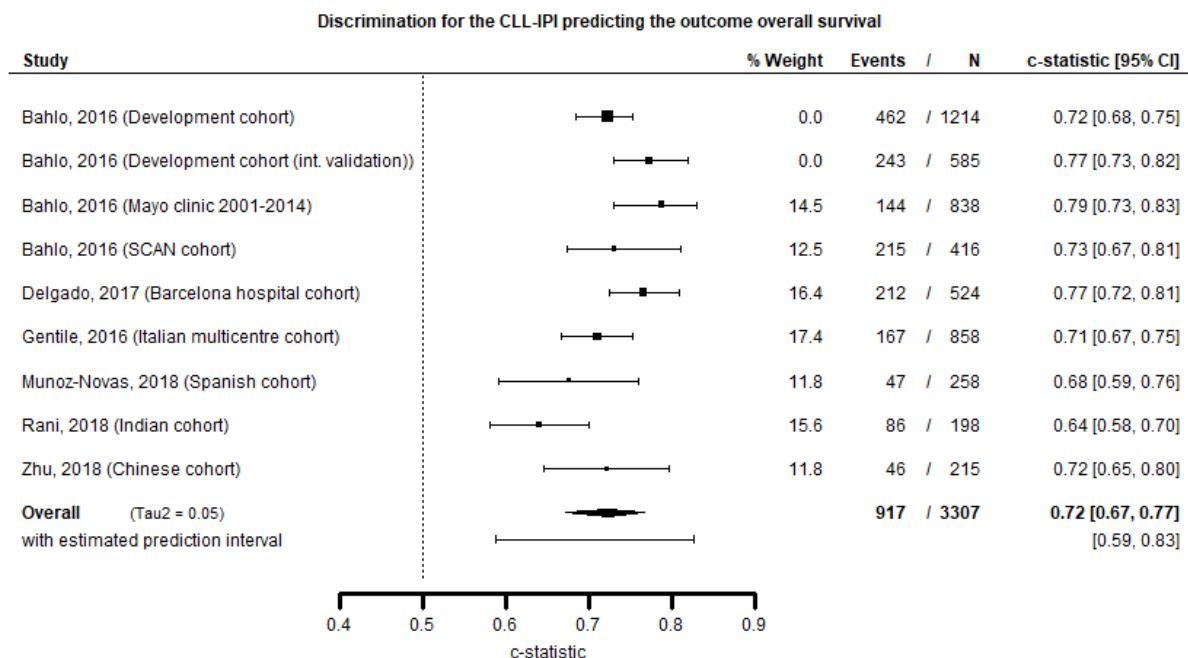


Figure 14: Meta-analysis of the c-statistic of external validation studies for the CLL-IPI model, taken from Kreuzberger et al. 2019 (93), figure 22.

Barcelona-Brno score

The Barcelona-Brno score (83) was based on the best fit of factor combinations from CLL-IPI model described previously. The score classified individuals into three risk groups: persons without del(11q) or del(17p) and mutated IgHV status were classified as low risk; persons with del(11q) or del(17p) and unmutated IgHV status were classified as high risk; all other persons were classified as intermediate risk. The risk of bias for the development of the Barcelona-Brno model was rated as unclear for the domain participants, low for participants and outcome, and high for the analysis domain.

Five studies externally validated the Barcelona-Brno model. Across studies, the risk of bias in the participants domain was rated as low or unclear, the predictor and outcome domains were rated predominantly as low, and the analysis domain was rated mostly as high risk of bias in all studies. Calibration of the model was not reported in the external validation studies. Thus, survival per risk group was pooled and presented with the original survival percentages (Table 6).

Risk group	Risk group	Development study: percentage of persons surviving at 5 years (95% CI)	Pooled percentage of persons surviving at 5 years (95% CI)	95% prediction interval (PI)
Low risk	No del(11q) or del(17p) and mutated IgHV status	93.4% (90.4% to 96.5%)	90.5% (85.1% to 94.0%)	80.4% to 95.7%
Intermediate risk	All other	83.8% (78.4% to 89.6%)	79.7% (70.7% to 86.5%)	63.1% to 90.0%
High risk	Del(11q) or del(17p) and unmutated IgHV status	70.0% (59.3% to 82.7%)	62.5% (49.3% to 74.1%)	41.3% to 79.7%

Table 6: Five-year survival per risk group as defined by the Barcelona-Brno development study in the training cohort, and pooled 5-year survival from three external validation studies.

Four external validation studies reported the discrimination of the Barcelona-Brno model in 1755 individuals experiencing 516 events. The pooled c-statistic of the four studies was 0.64 (95% CI 0.60 to 0.67), with no heterogeneity between studies ($\tau = 0.00$). The 95% PI ranged from 0.59 to 0.68 (Meta-analysis provided in Figure 15).

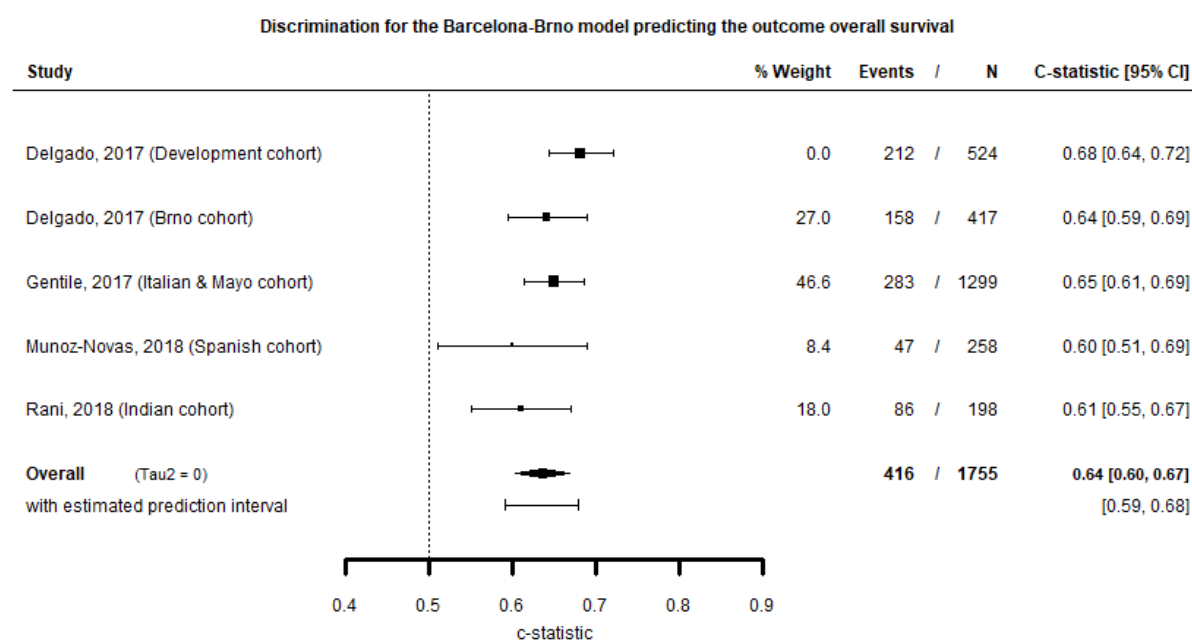


Figure 15: Meta-analysis of the c-statistic of external validation studies for the CLL-IPI model, taken from Kreuzberger et al. 2019 (93), figure 27.

MDACC 2007 model

The MDACC2007 model (84) was based on a proportional hazards model with univariable factor preselection. Bootstrapping accounted for optimistic apparent performance measures. The precise formula, a point score and a graphical translation in form of a nomogram were reported that translate into 10-year survival probabilities (point score presented in Table 7:Table 7). Risk of bias of the development study of the MDACC 2007 model was rated as low for all domains.

Prognostic factor	Point distribution			
	0	1	2	3
Age (years)	< 50		50-65	> 65
β-2 microglobulin (mg/L)	< ULN	1-2 × ULN	> 2 × ULN	
ALC (× 10 ⁹ /L)	< 20	20-50	> 50	
Gender	Female	Male		
Rai stage	0-II	III-IV		
No. of involved nodal groups	≤ 2	3		

Table 7: Point score of the MDACC 2007 prognostic model (simplified model).

There were nine external validation studies. The participants domain obtained mixed ratings, while the predictor and outcome domain were mostly rated as low risk of bias. The analysis domain was

rated in all except one study as high. The Calibration of the model was not reported in the external validation studies. Thus, survival per risk group was pooled and presented with the original survival percentages (Table 8) based on five studies including 3786 individuals.

Risk group	Index score	Development study: Percentage of persons surviving at 5 years (95% CI)	Pooled percentage of persons surviving at 5 years (95% CI)	95% prediction interval (PI)
Low risk	1-3	97.0% (95.0% to 99.0%)	97.0% (94.3% to 98.4%)	90.9% to 99.0%
Intermediate risk	4-7	80.0% (78.0% to 82.0%)	82.3% (74.6% to 88.0%)	61.5% to 93.1%
High risk	≥ 8	55.0% (47.2% to 62.8%)	45.6% (31.3% to 60.5%)	21.2% to 72.3%

Table 8: Five-year survival per risk group as defined by the MDACC 2007 model in the training cohort, and pooled 5-year survival from five external validation studies.

Seven external validation studies including 5127 participants with 994 events reported the discrimination of the MDACC 2007 index score. The pooled estimate of the seven studies was 0.65 (95% CI 0.60 to 0.70), with some degree of heterogeneity (between-study standard deviation, tau = 0.21). The prediction interval ranged from 0.51 to 0.77.

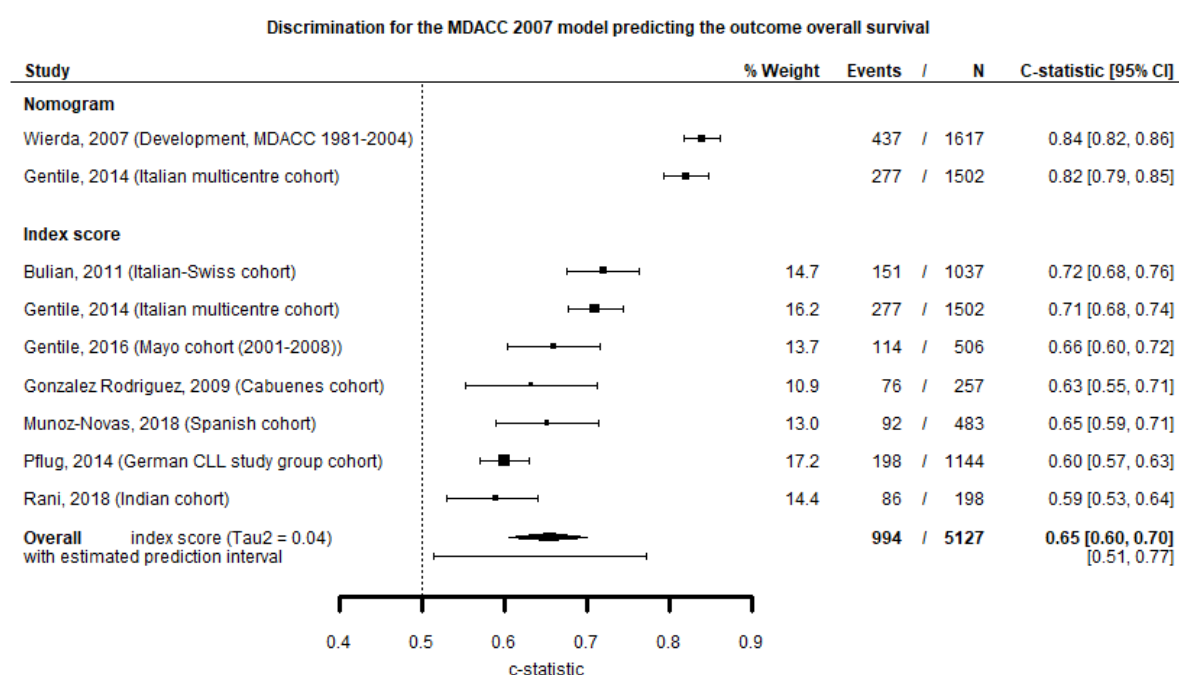


Figure 16: Meta-analysis of the c-statistic of external validation studies for the MDACC 2007 score, taken from Kreuzberger et al. 2019 (93), figure 28.

Conclusions**Implications for practice**

Only few models have been externally validated, of which the CLL-IPI might perform best according to the c-statistics, although the models have not been compared directly to each other. In light of rapidly changing treatment options in CLL, the models might need to be reevaluated regarding both calibration and discrimination.

Implications for research

One of the main points for research is a strong recommendation to adhere to reporting guidelines for prognostic model studies so that researchers and clinicians can sufficiently estimate whether both model development and external validation cohorts represent their patients, whether the treatments are those provided in their clinical context, etc. In addition, the review shows that the proportion of development studies is high in contrast to validation studies, although many more validation studies are needed for prognostic models to be used in clinical practice. We have not identified any studies assessing the impact of the prognostic models included, which would be a next step on the road towards implementation of a model.

This review was the first published prognostic model review in the Cochrane library. Its insights will also be used for the Cochrane handbook for prognosis systematic reviews which is currently under development.

Visual abstract

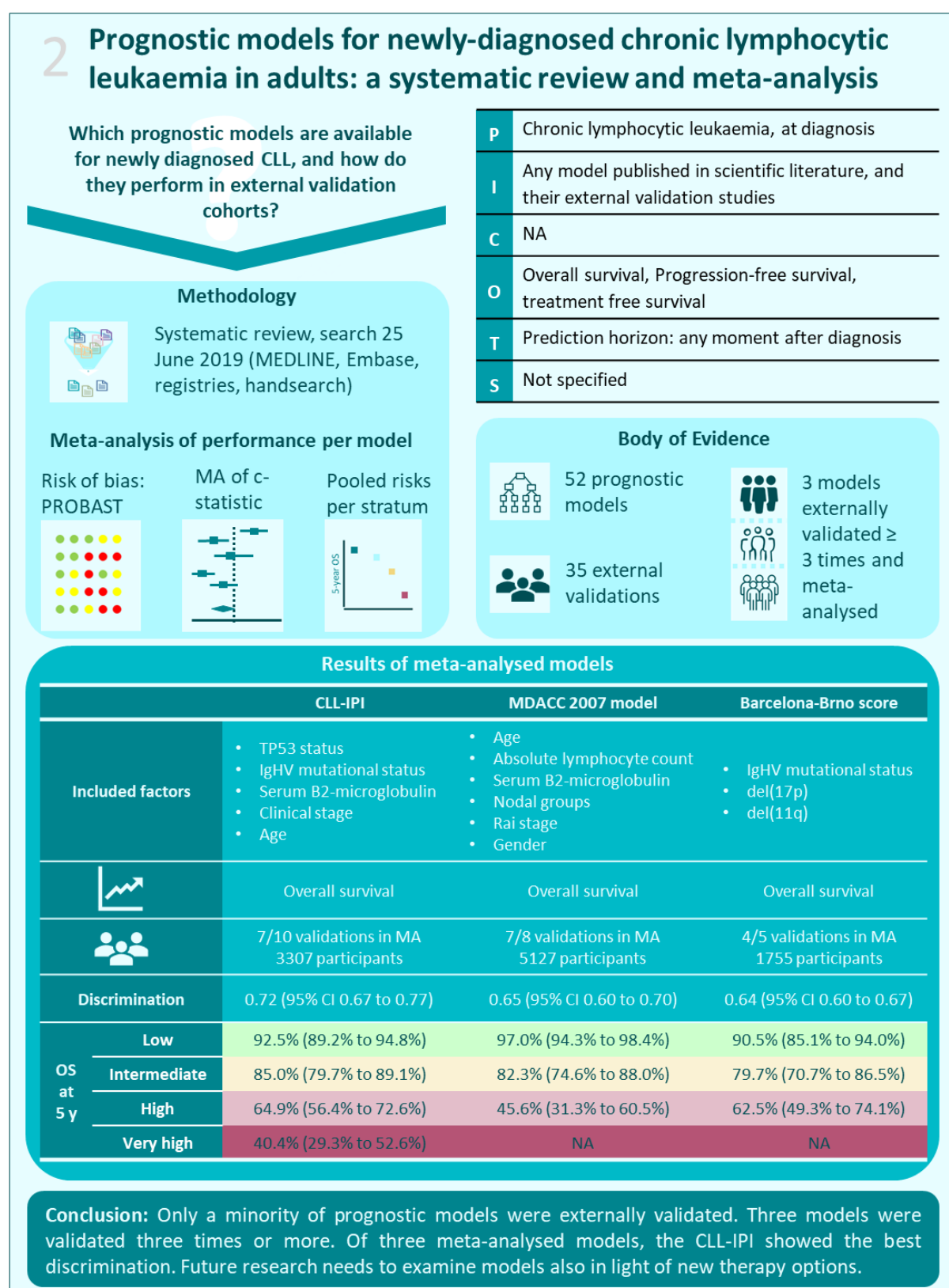


Figure 17: Visual abstract for the publication Kreuzberger et al, 2020 (93). Icons by Freepik, more attributions can be found in the section "Attributions of Icons and Software")

Publication 3: Positron emission tomography-adapted therapy for first-line treatment in adults with Hodgkin lymphoma

This systematic review and meta-analysis of randomised controlled trials examined FDG-PET based treatment adaptation strategies in Hodgkin lymphoma. This review is an update of a previous Cochrane review published in January 2015 (79) that included three studies. The current update was published in March 2025 in the Cochrane Library of Systematic Reviews. The published manuscript, due to its length, is attached as Appendix IV: Publication 3.

- Kreuzberger N, Goldkuhle M, von Tresckow B, Kobe C, Sickinger MT, Monsef I, Skoetz N. Positron emission tomography-adapted therapy for first-line treatment in adults with Hodgkin lymphoma. Cochrane Database of Systematic Reviews 2025, Issue 3. Art. No.: CD010533. DOI: 10.1002/14651858.CD010533.pub3
- First authorship
- 2-year Impact Factor of the Cochrane Library of Systematic Reviews (2023): 8.8
- This project did not receive any financial support.

Summary

Note: The summary contains excerpts from the published manuscript, which is provided in the appendix.

Introduction

Hodgkin lymphoma is a rare cancer that involves the lymph nodes, usually in the upper part of the body, and sometimes the spleen and other organs (60). It affects the B cells and is characterised by a small proportion of malignant tumour cells called Hodgkin and Reed-Sternberg cells (94). Around 40% of those affected experience typical B-symptoms such as fever, night sweats, and unintended weight loss (60). For staging of HL, the Ann Arbor classification is used. Stages one to three indicate the degree of lymph node involvement, whereas stage four indicates disseminated organ involvement (74). HL is with rates of up to 90% highly curable (61, 73). The current standard treatment consists of a combination regimen of chemotherapy (ABVD) and radiotherapy for early stages, and an intensified chemotherapy regimen (BEACOPP) for advanced stages (75).

Due to possible long-term adverse events such as secondary malignancies and cardiac toxicities, a treatment de-intensification of those individuals with a good prognosis has been suggested. Similarly, those with a worse prognosis might benefit from more intensive treatment (treatment escalation). As interim FDG-PET after multiple cycles of chemotherapy is a good prognostic factor for HL (81), it has been suggested as basis for treatment adaptation (79). Since the

last version of this systematic review, additional studies and data of included studies have been published.

Objectives

To assess the effects of interim [18F]-FDG-PET imaging treatment modification in previously untreated individuals with HL.

Methods

All methods were applied according to the Cochrane Handbook for Systematic Reviews of Interventions (95).

Eligibility criteria

We included randomised controlled trials that met the following eligibility criteria:

P	▪ Confirmed diagnosis of HL ▪ Any stage, all subtypes (analysed as separate comparisons) ▪ First-line treatment / not previously treated
I	▪ PET-adapted treatment strategies ○ Treatment modification / de-escalation after interim-PET negative HL ○ Treatment modification / escalation after interim PET positive HL
C	▪ Standard treatment regimen
O	▪ Overall survival (OS) ▪ Progression-free survival (PFS) ▪ Response rate ▪ Treatment-related mortality (TRM) ▪ Adverse events • Number of participants with any AE • Number of participants with grade 3 to 4 AE • Number of participants with serious adverse events (SAE) • Incidence of secondary malignancy ▪ Quality of life (QoL)

Table 9: Eligibility criteria in PICO format for publication 3: PET-based treatment adaptation in newly diagnosed HL.

Search methods

MEDLINE, Embase, the Cochrane Central Register of Controlled Trials (CENTRAL), clinicaltrials.gov and WHO ICTRP were searched for the time of last review update (2014) to 17 November 2023. Reference lists of included studies and relevant review articles were checked for relevance.

Data collection

Two reviewers independently screened identified records in a two-step approach for eligibility, extracted predefined items into a standardised data extraction form, and assessed the risk of bias with the risk of bias 1 tool (96) (to stay aligned with the previous review version).

Data analysis

Early, intermediate, and advanced stage HL were analysed separately, similar to PET-negative as compared to PET-positive participants (treatment de-escalation versus escalation). The most informative effect estimate available per comparison was analysed (i.e., time-to-event analyses where possible). In meta-analysis, the random-effects model was applied, using the Mantel-Haenszel method for dichotomous outcomes and inverse variance for time-to-event outcomes. The GRADE framework was applied to rate the certainty in the evidence and for expressing the results (60, 97).

Main Results

The current update of the review includes 10 RCTs covering early (1 RCTs, 667 participants), intermediate (1 RCTs, 651 participants), early-to-intermediate (3 RCTs, 1639 participants) and advanced-stage HL (5 RCTs; 3629 participants). There are no further ongoing studies.

PET-negative participants

Early-stage HL

This comparison included two studies, H10 (favourable HL participants) and HD16 with a total of 1132 participants. Both studies omitted radiotherapy from the standard regimen of ABVD plus radiotherapy for de-escalation, although H10 added one cycle of chemotherapy for those randomised to treatment de-escalation without radiotherapy. Follow-up time was between 66 months and 9.5 years.

It is uncertain whether PET-based treatment adaptation has an effect on OS, PFS, or TRM (HR 0.84, 95% CI 0.13 to 5.32; HR 4.52, 95% CI 0.72 to 28.41, and TRM not estimable; very low-certainty evidence). Results were downgraded due to very serious imprecision (very wide confidence intervals) and serious risk of bias (attrition bias and lack of blinding). There may be little to no effect on secondary malignancies (RR 0.84, 95% CI 0.46 to 1.49; low-certainty evidence). No data on QoL was available yet, further AEs could not be used for analysis.

Intermediate-stage HL

The comparison on intermediate stage HL includes two studies, H10 (unfavourable HL participants) and HD17 with a total of 1245 participants. The trials differ in their initial therapy; H10 starts with two cycles ABVD before interim PET, and continues with either two cycles of ABVD plus radiotherapy, or the treatment adaptation with four cycles of ABVD. HD17 initiates with two cycles of BEACOPP escalated and 2 cycles ABVD followed by interim PET, then continuing with either RT or no RT. Median study follow-up ranged between 46 and 114 months.

PET-based treatment adaptation in intermediate stage HL may have little to no effect on OS (HR 0.91, 95% CI 0.42 to 1.96), PFS (HR 1.59, 95% CI 0.95 to 2.67), SAE (RR 1.01, 95% CI 0.84 to 1.21) and secondary malignancy (RR 1.01, 95% CI 0.57 to 1.82). These outcomes were rated as low certainty

evidence due to serious risk of bias (attrition bias in H10 and lack of blinding in all trials), serious imprecision, and potential publication bias for SAE. The effect on TRM is uncertain (RR 2.94, 95% CI 0.31 to 28.20, very low certainty due to very serious imprecision and serious risk of bias). Data on QoL was not reported.

Advanced-stage HL

For advanced stage HL, there were two sets of treatment strategies: Three trials, AHL2011, HD18 and RATHL with a total of 2635 participants, varied the number or intensity of chemotherapy after interim PET, while two trials (HD0607, HD0801 with 412 participants) considered radiotherapy after chemotherapy as standard therapy and de-escalated by omitting RT.

Adapting the intensity of chemotherapy likely increases OS (HR 0.65, 95% CI 0.40 to 1.07) while probably having no effect on PFS (HR 0.98, 95% CI 0.78 to 1.25) compared to standard of care. The evidence is of moderate certainty and was downgraded by one level for serious imprecision or risk of bias. It may reduce TRM (HR 0.21, 95% CI 0.06 to 0.73) and secondary malignancies (HR 0.87, 95% CI 0.60 to 1.26) and probably slightly reduces the number of participants with grade 3 to 4 adverse events (RR 0.94, 95% CI 0.92 to 0.97). These data were at low certainty of evidence due to serious imprecision and serious risk of bias. Data on QoL are not yet available.

Omitting radiotherapy from the alternative standard regimen may increase OS (HR 0.63, 95% CI 0.11 to 3.69) and PFS (HR 0.78, 95% CI 0.43 to 1.43). The evidence was downgraded by two levels due to very serious imprecision. It may reduce the incidence of secondary malignancy (RR 0.41, 95% CI 0.08 to 2.09, low certainty evidence due to serious risk of bias and serious imprecision). The evidence was of low certainty due to serious risk of bias and serious imprecision. Data for TRM, AE, and QoL were not reported.

PET-positive participants

Early-to-intermediate stage HL

One RCT, H10, examined treatment escalation after a positive interim PET result in 361 early-to-intermediate stage participants. Multiple manuscripts were published, here, data from the 10-year update is reported for OS and PFS. Participants started with two cycles of ABVD, followed by interim PET. Those randomised to standard regimen received 2 cycles of ABVD plus radiotherapy, while intensified treatment consisted of two cycles of BEACOPP_{escalated} plus radiotherapy.

Treatment escalation may increase OS (HR 0.92, 95% CI 0.43 to 1.97) and PFS (HR 0.67, 95% CI 0.29 to 1.55), for both estimates, we have low certainty in the evidence; it was downgraded for very serious imprecision for OS and serious imprecision and serious risk of bias (lack of blinding) for PFS. It is uncertain whether treatment escalation has an effect on the incidence of secondary malignancy (RR 1.23, 95% CI 0.43 to 3.55, very low certainty evidence due to very serious imprecision and serious risk of bias). Data on AE, SAE, and QoL were not reported.

Advanced-stage HL

There were two studies including 582 participants that examined treatment escalation with rituximab after a positive interim PET, HD0607 and HD18. HD0608 started with two cycles of ABVD, followed by interim PET and four cycles BEACOPP and BEACOPP_{escalated}, either with or without rituximab. HD18 initiated with two cycles of BEACOPP_{escalated}, followed by interim PET and six additional cycles of BEACOPP_{escalated} with or without rituximab.

Treatment intensification with rituximab does probably not increase OS (HR 1.39, 95% CI 0.74 to 2.63, moderate certainty, downgraded for serious imprecision) or PFS (HR 1.03 95% CI 0.68 to 1.54, moderate certainty, downgraded for serious risk of bias), while it may increase TRM (RR 4.00, 95% CI 0.45 to 35.50; very low certainty, downgraded for very serious imprecision and serious risk of bias), the number of people with AE (RR 1.16, 95% CI 0.89 to 1.51, low certainty evidence) and SAE (RR 1.61, 95% CI 1.00 to 2.60, moderate certainty, downgraded for serious risk of bias). It may have no effect on grade 3 to 4 AE (RR 1.06, 95% CI 0.77 to 1.45, low certainty) and secondary malignancy (RR 0.67, 95% CI 0.28 to 1.60, low certainty). Data on QoL were not reported.

Other outcomes, comparisons, and subgroup data

Information on further outcomes such as fertility, treatment discontinuation, and adverse events, as well as subgroup data, can be found in the text and tables of the published Cochrane review provided in Appendix IV: Publication 3.

Conclusions

Implications for practice

In early-stage interim PET negative HL, standard treatment would be, except through individual consideration, the preferred option based on the serious uncertainty of the evidence. In intermediate stage PET negative HL, treatment de-escalation by omission of radiotherapy might be a feasible treatment option, as there is no negative effect on survival and progression. This contrast between early and intermediate stages might be caused by the variation in initial treatment used, which is more intensive in intermediate stage HL. In advanced stage HL, treatment de-escalation by deintensification of chemotherapy seems beneficial regarding survival and secondary malignancies without affecting progression. Similarly, if using an alternative standard regimen including radiotherapy, de-escalation might be beneficial for survival and progression.

In early and intermediate stage HL, treatment escalation with intensified chemotherapy might improve the prognosis. Intensification of treatment by adding rituximab to standard treatment of advanced stage HL is not implicated, as there is probably no beneficial effect on survival and progression, while adverse events are increased.

Implications for research

This systematic review has identified ten studies on both de-intensification of PET-negative and intensification PET-positive participants. The examined treatment regimen assumed standard therapy consisting of combined chemo- and radiotherapy in early- and intermediate stage HL, and intensive chemotherapy in advanced stage HL. From included trials, some outcomes, especially quality of life and fertility data, have not yet been published. There are still some gaps that are not addressed by the current studies, i.e., the role of PET based treatment adaptation for subgroups such as pregnant and elderly participants, which might come from non-randomised trials. In addition, the role of interim PET scans must be reevaluated in light of new developments, for example immunotherapies.

Visual abstract

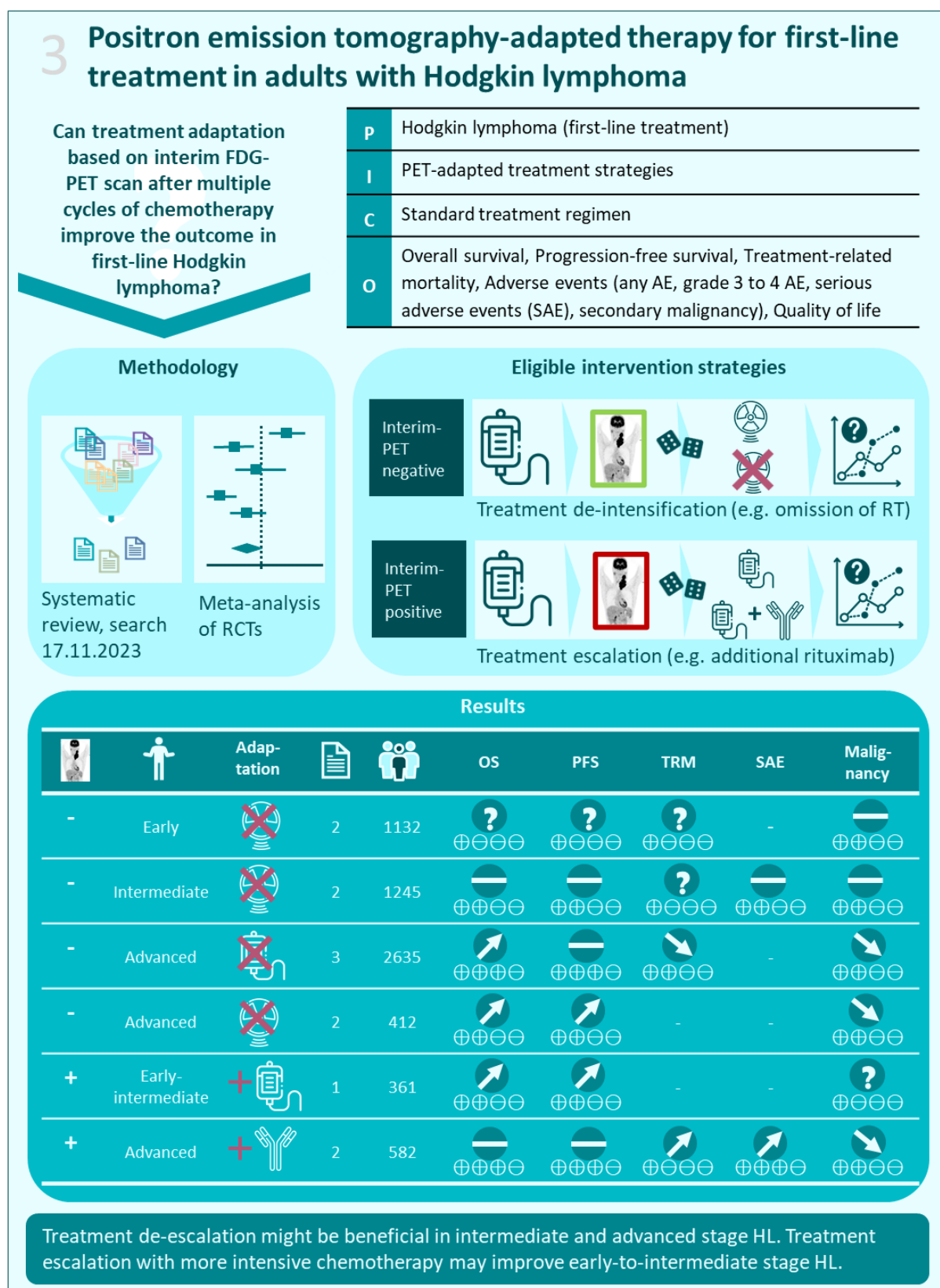


Figure 18: Visual abstract for the publication Kreuzberger et al, 2025 (98) (the PET-scan was adapted from figure 3b in Picardi et al., 2021 [66] [published under CCA 4.0 International License]. Icons by Freepik, more detailed attributions can be found in the section "Attributions of Icons")

Discussion

Summary

The aim of this dissertation project was to develop and apply the currently evolving methodology of prognosis systematic reviews by means of clinically relevant research questions related to the field of haematology. This aim was achieved by covering three out of four prognosis themes according to the PROGRESS framework (4): prognostic factors, prognostic models, and stratified medicine systematic reviews.

The first systematic review on interim-PET scans as a prognostic factor for OS, PFS and adverse events in untreated HL was one of the first Cochrane reviews published in the *Cochrane Library of Systematic Reviews*, showing the feasibility of applying methodology developed with the Cochrane Prognosis Methods group. Studies of various designs were included and rated using the QUIPS tool (45) designed for prognostic factor studies. Meta-analyses of unadjusted effect estimates were conducted which showed that interim-PET is probably a good predictor for OS, while less certain data was available for PFS. Participation in the GRADE prognosis working group allowed the application of the GRADE framework, which had just been developed for prognostic factors (53) at the time of conducting the review.

Similarly, in close collaboration with the Cochrane Prognosis Methods Group, the systematic review on prognostic models for previously untreated CLL and their external validation studies was published as the first of its kind in the *Cochrane Library of Systematic Reviews*. This review used PROBAST (46), the newly developed risk of bias tool for prognostic models, and applied the most recent meta-analytical strategies (34, 51). The main result of this review is that from three prognostic models that were sufficiently externally validated to allow met-analysis, the CLL-IPI seems to perform best regarding calibration of the model. However, the interpretation comes with serious limitations, as the calibration could not be assessed; instead, the survival rates per risk group were pooled.

The third systematic review that forms part of this dissertation project evaluated the previously identified prognostic factor interim-PET as basis for treatment adaptation in HL. Following standard Cochrane methodology for randomised controlled trials (2), this review's meta-analyses were divided into clinical stages in line with current evidence-based clinical guidelines (99) (and in contrast to a previous version of the review (79)). It concludes that whether treatment adaptation is advantageous depends on the stage and interim-PET status: While in early-stage interim-PET-negative HL, the evidence on treatment de-escalation by omission of radiotherapy is uncertain, there may be little to no difference in intermediate-stage HL. In advanced-stage interim-PET-negative HL, treatment de-escalation is probably beneficial. In early-to-intermediate stage, PET-positive HL,

intensification of the chemotherapy regimen may improve outcomes, while adding rituximab to standard of care in advanced stage HL probably improves harms.

Altogether, the three clinically relevant prognosis systematic reviews that advanced the recent methodological developments are associated with several strengths and limitations as outlined below.

Strengths and Limitations of the Methodology

All three systematic reviews were conducted based on an a-priori registered protocol and followed the high standard of Cochrane reviews wherever possible, and reporting according to PRISMA 2020 (44). In case of deviations from the standard methodology due to the new systematic review type in prognosis, the methodology was chosen in accordance with suggestions by the Cochrane prognosis methods group and the respective expert was involved. For instance, for the prognostic model review, a working version of the new risk of bias tool for prognostic models, PROBAST, was used up to its final publications, and the lead authors of PROBAST were involved in the review process for meaningful interpretation and application of the items, as the explanation and elaboration paper had not yet been published (100). Similarly, GRADE for prognostic factors was just under development during conduct of the prognostic factor review (53), therefore, the respective expert from the GRADE working group was involved. A limitation of the prognostic model review is that GRADE for this type of evidence was not yet available, and is, even now, in a conceptual phase (54, 55) without guidance for its application. As the PhD candidate is member of the prognosis methods group, she was involved in those developments.

Whenever there were differences between a priori defined methodology and applied methodology, this was transparently noted per systematic review in the respective section (for the details, consult manuscripts attached as appendices).

While a strength of the included systematic reviews is the collaboration between methodology and clinical expertise, a limitation is the general lack of patient or patient representative involvement. Patient involvement has increasingly come to the fore recently, which is evident from its inclusion in both grant proposals and as a specific heading in new Cochrane reviews. The reviews in this dissertation have been planned and conducted before this shift in format and might have benefitted from patient input.

Strengths and Limitations of the included Literature, and resulting implications for primary prognosis studies

One major limitation of the included literature is related to reporting of primary prognosis studies. For prognostic factor reviews, any factor – outcome association can be of interest, which might be

presented in RCTs or cohort studies as exploratory analyses in form of a short sentence without further details, or hidden in appendices. Primary study investigators might not aware the variables included in one of their multivariable models might be a factor of interest for another research team. Due to limited word count, often, only the “significant” factors are mentioned and further examined in multivariable analysis, leading to an enormous publication bias especially for systematic reviews that only consider multivariable results. As a consequence, similar to previous work (101), the studies included in prognosis reviews show serious limitations in conduct and reporting. The prognostic factor systematic review showed that reporting deficiencies existed especially in the domain of analysis; methods varied between initial study protocols and reported analyses. Often, adjusted analyses were omitted, and if reported, the rationale for covariates was lacking. It would be desirable if effect estimates and their respective 95% confidence intervals were reported in supplemental tables for all tested factor – outcome relationships of a study instead of abbreviation to a sentence of non-significance, with clear description of all included co-variables. However, due to current journal policies, this might not always be possible.

The literature search of the prognosis reviews was very sensitive, and the records were screened independently by two systematic reviewers. Additionally, we tracked the citing publications for the prognostic model review. We are therefore confident to have identified all relevant studies up to the search date. Regarding the prognostic model review, the resulting pattern of studies conducted confirms the previously detected disproportion between developed models (N = 52) and external validations, without any implementation testing. An abundance of non-validated prognostic models of varying quality may be confusing for clinical practice as it is unclear how they perform, and whether they can change the course of a patient. More useful would be few well-validated and implemented models whose predictions can be adapted according to setting, and whose implementation shows some benefit, either in informing patients or in adapting treatment regimen accordingly. In future, the most commonly cited or used models might be externally validated on the same cohorts to provide a directly comparable performance estimate. Then, if one of the models shows potential based on discrimination and calibration, its ease of use and ability to change clinical practice should be evaluated before implementation.

It is also evident that the calibration of models both in development and in external validation cohorts has not yet been normalised, although this has been part of the CHARMS checklist and TRIPOD statement for prognostic models that were published in 2014 and 2015 (13, 14). Although we attempted to contact primary trial investigators for missing information, we could not obtain additional data for all relevant studies.

The review on interim-PET based treatment adaptation for HL showed that the main outcomes such as OS and PFS are often well-reported, while data on secondary outcomes such as quality of life

and fertility outcomes are omitted from the reports. Adverse events are reported very heterogeneously; sometimes they are reported as counts of adverse events, sometimes as persons with any adverse event. The way of summarising across events and event categories also renders meta-analyses challenging; events cannot be summed up by systematic review authors without further information on overlapping participants. This has been consistent over time and resulted in an update of the CONSORT HARMS statement (102) that primary study investigators could follow.

Finally, as each scientific report, the systematic reviews present just a snapshot over the recent past. The reviews were conducted during a period of clinical developments regarding new therapies. For example, the prognostic model review showed that many of the development and external validation studies did not report the treatment regimen that the included CLL patients received during the course of the study (only the number of those treated at all). Those that do report regimen used data from patients treated overwhelmingly with chemotherapy. In contrast, currently, different combinations of BCL-2-Inhibitor (venetoclax), bruton tyrosine kinase (BTK) inhibitors (e.g., ibrutinib; acalabrutinib) and anti-CD20 monoclonal antibodies (rituximab; obinutuzumab) (103). This likely increases the prognosis in general, and models may not be accurate in these patients. External validation studies need to be conducted in populations treated with the current standard of care to show whether models are still accurate.

Implications for Systematic Reviews of Prognosis Studies

The current dissertation project includes two exemplary Cochrane reviews on prognosis, conducted in close collaboration with the methodological experts in the field. For other Cochrane authors, these may serve as an example both regarding structure and methodology, by keeping an eye on the most recent developments of the Prognosis Methods Group and the GRADE prognosis working group. The reviews show that although there are more and more prognosis questions and systematic reviews, their conduct is currently still more time-consuming and challenging due to difficulties in multiple steps. Starting with the search strategy and number of hits, review authors have to carefully consider the restrictions they want to enter into the search strategy versus the resources they have to put into screening and selecting studies. Data extraction forms have to be adapted to the specific type of prognosis studies the review authors are interested in, and authors have to carefully consider the unit of analysis (i.e., multiple reports on the same or overlapping data with same factor-outcome associations or external validations, counting of one model per cohort or multiple models).

The risk of bias assessment differs between the types of prognosis reviews. No specific tool is yet available for overall prognosis, which is currently under development (104). QUIPS is the tool to choose to assess prognostic factor studies (45), although this tool may need adaptations such as an additional unclear category, and ideally a priori defined tailored items to avoid arbitrary decisions

and reduce the mix of incomplete reporting and risk of bias. While PROBAST for prognostic models (46) prompts both rating the risk of bias and the applicability of a model, review authors may carefully consider if structured applicability ratings are necessary for their research question, especially when eligibility criteria are broad and allow a variety of studies to be included.

Meta-analytic approaches are different than those for intervention reviews. As review authors are often confronted with missing data, they may think upfront what to do in such cases. In first instance, of course, the primary study investigators would be contacted, but this may not always resolve the challenge. Therefore, authors need to think whether they would do alternative analyses, for example if only unadjusted instead of adjusted effect estimates, or both types of effect estimates, in prognostic factor studies are reported. Similarly, prognostic models may have multiple versions, for example an exact regression formula and a simplified score, that can be validated, or multiple outcomes that were tested in addition to the outcome used for model development.

Lastly, the interpretation of the results of prognosis systematic reviews might not be as straightforward as for interventions. In overall prognosis, only an incidence is rated without an effect estimate. The interpretation of prognostic factor pooled estimates is similar to that of intervention studies; however, confounding plays a larger role than in RCTs. The largest challenge is the interpretation of pooled discrimination of the external validation studies of a prognostic model, as there is currently no recommended cut-off. The GRADE prognosis working group has published two concept papers on the rating of certainty in the body of evidence of prognostic models since the completion of the prognostic model systematic review in this dissertation project (105, 106). They suggest three types of thresholds as target for rating the certainty: random classification of the patients (c-statistic of 0.5), clinician intuition (which may not always be easy to observe, and carry its own variability between contexts), and an alternative model that is established in the field (106). More guidance is currently under development.

Implications for Clinical practice

This series of systematic reviews shows that a good prognostic factor such as an interim-PET after several cycles of chemotherapy can be further explored to serve as basis for treatment stratification. In the case of Hodgkin lymphoma, the role of treatment stratification based on interim-PET in Hodgkin lymphoma differs between early, intermediate, and advanced stages, although interim-PET status was prognostic across stages (81). This may be due to the different standard of care regimen depending on the stage. In early-stage HL, which is originally treated with two cycles of ABVD plus 20 Gy involved-site RT (62, 107), de-escalation in PET-negative patients in the HD16 study meant omission of radiotherapy (108), while H10 omitted RT but added one cycle of ABVD (109). Deviation from standard of care cannot be concluded. In contrast, intermediate stages were treated with a

more intensive chemotherapy regimen of either 2 cycles of BEACOPP_{escalated} plus two cycles of ABVD (73) or four cycles of ABVD (107), both plus 30 Gy involved-site radiotherapy. With the more intensive treatment regimen, omission of RT may be a tangible option (110). In advanced stage HL, treatment de-escalation based on a negative interim-PET has shown to benefit patients based on the studies HD18 (111) and RATHL (112), was thus recommended and is already standard of care (99, 113). In other cases, a prognostic factor or prognostic biomarker that defines prognosis for a certain disease may be identified and serve as target for treatment development.

The prognostic model systematic review involved another haematological malignancy and is thus clinically not directly linked to the other two reviews. However, this review shows the difficulty clinicians may be faced with in choosing a relevant model for prognosis. First of all, the implementation and duration of applying of a model in clinical practice has to be feasible, thus quick and based on available data. In the models validated for CLL, for example, del(17p) and/or TP53 mutations are commonly included because they have shown to be good prognostic factors (114), and are nowadays usually measured as part of diagnostic procedures by fluorescence in situ hybridization (FISH). Secondly, this model should then be superior to clinician intuition, which in many cases cannot be determined. And thirdly, it should result in some advantage for the patient, either in terms of feeling more informed, or even in his or her treatment course. This implies that a prognostic model has to undergo a hierarchy of validation and implementation before it can be routinely recommended, which is rarely undertaken, including the examined prognostic models for CLL (93).

Conclusion

Although the methodology of prognosis reviews is currently still under development, the three reviews included in this dissertation project show that prognosis systematic reviews are a feasible study design that can already be implemented for Cochrane reviews. They also flag that the primary study landscape still suffers from serious limitations such as lack of indexing, very serious reporting deficiencies and studies at high risk of bias that would largely benefit from applying available checklists and guidance. The prognostic model literature seems to lean unproportionally towards developing models as compared to externally validating or impact testing these models, leading to an abundance of models without clinical applicability.

It is evident from this series that a good prognostic factor can result in it being explored as a target for treatment adaptation. However, the differences of PET-based treatment adaptation across early, intermediate, and advanced stages also show that such factor does not always translate to be a good stratification factor, consequently leading to changes in clinical practice. The precise situations of benefit have to be carefully disentangled.

References

1. Lasserson TJ, Thomas J, Higgins JPT. Chapter 1: Starting 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 64 (updated August 2023): Cochrane; 2023.
2. Deeks JJ, Bossuyt PM, M. LM, Takwoingi Y. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. Version 2.0 (updated July 2023). Chichester (UK): John Wiley & Sons; 2023.
3. Noyes J, Harden A. *Cochrane-Campbell Handbook for Qualitative Evidence Synthesis*, Version 1. London: Cochrane; 2023.
4. 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. *BMJ*. 2013;346:e5595.
5. Jermy B, Läll K, Wolford BN, Wang Y, Zguro K, Cheng Y, et al. A unified framework for estimating country-specific cumulative incidence for 18 diseases stratified by polygenic risk. *Nature Communications*. 2024;15(1):5007.
6. Kartiosuo N, Raitakari OT, Juonala M, Viikari JSA, Sinaiko AR, Venn AJ, et al. Cardiovascular risk factors in childhood and adulthood and cardiovascular disease in middle age. *JAMA Network Open*. 2024;7(6):e2418148-e.
7. National Cancer Institute. *Cancer Trends Progress Report: National Cancer Institute, NIH*; 2024 [Available from: <https://progressreport.cancer.gov>].
8. Hingorani AD, Windt DAVd, Riley RD, Abrams K, Moons KGM, Steyerberg EW, et al. Prognosis research strategy (PROGRESS) 4: Stratified medicine research. *BMJ*. 2013;346:e5793.
9. Riley RD, Hayden JA, Steyerberg EW, Moons KGM, Abrams K, Kyzas PA, et al. Prognosis Research Strategy (PROGRESS) 2: Prognostic Factor Research. *PLOS Medicine*. 2013;10(2):e1001380.
10. 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.
11. 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.
12. Sauerbrei W, Taube SE, McShane LM, Cavenagh MM, Altman DG. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): An Abridged Explanation and Elaboration. *Journal of the National Cancer Institute*. 2018;110(8):803-11.
13. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Medicine*. 2015;13(1):1.
14. Moons KGM, de Groot JAH, 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(10):e1001744.
15. Dhiman P, Ma J, Navarro CA, Speich B, Bullock G, Damen JAA, et al. Reporting of prognostic clinical prediction models based on machine learning methods in oncology needs to be improved. *Journal of Clinical Epidemiology*. 2021;138:60-72.

16. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. 2020;369:m1328.
17. World Health Organization. Global Health Estimates: Life expectancy and leading causes of death and disability; GHE: Life expectancy and healthy life expectancy: World Health Organization; 2024 [Available from: <https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-life-expectancy-and-healthy-life-expectancy>].
18. Kim SJ, Hong JS, Chang MH, Kim JA, Kwak JY, Kim JS, et al. Highly elevated serum lactate dehydrogenase is associated with central nervous system relapse in patients with diffuse large B-cell lymphoma: Results of a multicenter prospective cohort study. *Oncotarget*. 2016;7(44):72033-43.
19. Wu W, Miao L, Zhao L, Zhu Y, Mao J, Cai Z, et al. Prognostic value of lactate dehydrogenase, serum albumin and the lactate dehydrogenase/albumin ratio in patients with diffuse large B-cell lymphoma. *Hematology*. 2024;29(1):2293514.
20. Hugueta A, Hayden JA, Stinson J, McGrath PJ, Chambers CT, Tougas ME, Wozney L. Judging the quality of evidence in reviews of prognostic factor research: adapting the GRADE framework. *Systematic reviews*. 2013;2:71.
21. Collins GS, Ogundimu EO, Cook JA, Manach YL, Altman DG. Quantifying the impact of different approaches for handling continuous predictors on the performance of a prognostic model. *Statistics in medicine*. 2016;35(23):4124-35.
22. Apgar V. A Proposal for a new method of evaluation of the newborn infant. Originally published in July 1953, volume 32, pages 250-259. *Anesthesia and Analgesia*. 2015;120(5):1056-9.
23. Persson M, Razaz N, Tedroff K, Joseph KS, Cnattingius S. Five and 10 minute Apgar scores and risks of cerebral palsy and epilepsy: population based cohort study in Sweden. 2018;360:k207.
24. Ladelund AK, Bruun FJ, Slavensky JA, Ladelund S, Kesmodel US. Association of Apgar score at 5 minutes with academic performance and intelligence in youth: A cohort study. *Acta Obstet Gynecol Scand*. 2022;101(3):303-12.
25. Hong J, Crawford K, Jarrett K, Triggs T, Kumar S. Five-minute Apgar score and risk of neonatal mortality, severe neurological morbidity and severe non-neurological morbidity in term infants, 2013; an Australian population-based cohort study. *The Lancet Regional Health – Western Pacific*. 2024;44.
26. Wilson PWF, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation*. 1998;97(18):1837-47.
27. Damen JA, Pajouheshnia R, Heus P, Moons KGM, Reitsma JB, Scholten RJPM, et al. Performance of the Framingham risk models and pooled cohort equations for predicting 10-year risk of cardiovascular disease: a systematic review and meta-analysis. *BMC Medicine*. 2019;17(1):109.
28. Sehn LH, Berry B, Chhanabhai M, Fitzgerald C, Gill K, Hoskins P, et al. The revised International Prognostic Index (R-IPI) is a better predictor of outcome than the standard IPI for patients with diffuse large B-cell lymphoma treated with R-CHOP. *Blood*. 2007;109(5):1857-61.
29. International Non-Hodgkin's Lymphoma Prognostic Factors Project. A predictive model for aggressive Non-Hodgkin's lymphoma. *New England Journal of Medicine*. 1993;329(14):987-94.
30. Alamer F, Alamir A, Alqahtani A, Alkabli AM, Alshabib H, Damraj M. Validation of the International Prognostic Index and subsequent revisions for diffuse large B-cell lymphoma in patients from the Middle East and North Africa Region. *Cureus*. 2020;12(8):e9620.

31. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open*. 2021;11(7):e048008.
32. Montesinos López OA, Montesinos López A, Crossa J. Overfitting, model tuning, and evaluation of prediction performance. *Multivariate statistical machine learning methods for genomic prediction*. Cham: Springer International Publishing; 2022. p. 109-39.
33. Steyerberg EW, Harrell FE, Borsboom GJJM, Eijkemans MJC, Vergouwe Y, Habbema JDF. Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *Journal of Clinical Epidemiology*. 2001;54(8):774-81.
34. Debray TPA, Damen JAAG, Snell KIE, Ensor J, Hooft L, Reitsma JB, et al. A guide to systematic review and meta-analysis of prediction model performance. *BMJ*. 2017;356:i6460.
35. Debray TPA, de Jong VMT, Moons KGM, Riley RD. Evidence synthesis in prognosis research. *Diagnostic and prognostic research*. BMC. 2019;3:13.
36. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology (Cambridge, Mass)*. 2010;21(1):128-38.
37. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Bossuyt P, et al. Calibration: the Achilles heel of predictive analytics. *BMC Medicine*. 2019;17(1):230.
38. Caetano SJ, Sonpavde G, Pond GR. C-statistic: A brief explanation of its construction, interpretation and limitations. *European Journal of Cancer (Oxford, England : 1990)*. 2018;90:130-2.
39. Damen JAA, Moons KGM, van Smeden M, Hooft L. How to conduct a systematic review and meta-analysis of prognostic model studies. *Clinical Microbiology and Infection*. 2023;29(4):434-40.
40. Rayner DG, Kim B, Foroutan F. A brief step-by-step guide on conducting a systematic review and meta-analysis of prognostic model studies. *Journal of Clinical Epidemiology*. 2024;170:111360.
41. Kavanagh PL, Frater F, Navarro T, LaVita P, Parrish R, Iorio A. Optimizing a literature surveillance strategy to retrieve sound overall prognosis and risk assessment model papers. *Journal of the American Medical Informatics Association*. 2021;28(4):766-71.
42. Geersing G-J, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons K. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. *PLOS ONE*. 2012;7(2):e32844.
43. Boulos L, Ogilvie R, Hayden JA. Search methods for prognostic factor systematic reviews: a methodologic investigation. *Journal of the Medical Library Association*. 2021;109(1):23-32.
44. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.
45. Hayden JA, van der Windt DA, Cartwright JL, Côté P. Assessing Bias in Studies of Prognostic Factors. *Annals of Internal Medicine*. 2013;158(4):280-6.
46. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Annals of Internal Medicine*. 2019;170(1):51-8.

47. Rutter M, Bowley J, Lanyon PC, Grainge MJ, Pearce FA. A systematic review and meta-analysis of the incidence rate of Takayasu arteritis. *Rheumatology (Oxford, England)*. 2021;60(11):4982-90.
48. Authority Health Information and Quality. Evidence summary of reinfection and the duration of antibody responses following SARS-CoV-2 infection. HIQA; 2020.
49. O Murchu E, Byrne P, Walsh KA, Carty PG, Connolly M, De Gascun C, et al. Immune response following infection with SARS-CoV-2 and other coronaviruses: A rapid review. *Reviews in Medical Virology*. 2021;31(2):e2162.
50. Schwarzer G, Chemaitelly H, Abu-Raddad LJ, Rücker G. Seriously misleading results using inverse of Freeman-Tukey double arcsine transformation in meta-analysis of single proportions. *Research Synthesis Methods*. 2019;10(3):476-83.
51. Debray TP, Damen JA, Riley RD, Snell K, Reitsma JB, Hooft L, et al. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. *Statistical Methods in Medical Research*. 2019;28(9):2768-86.
52. 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.
53. Foroutan F, Guyatt G, Zuk V, Vandvik PO, Alba AC, Mustafa R, et al. GRADE Guidelines 28: Use of GRADE for the assessment of evidence about prognostic factors: rating certainty in identification of groups of patients with different absolute risks. *Journal of Clinical Epidemiology*. 2020;121:62-70.
54. Foroutan F, Guyatt G, Trivella M, Kreuzberger N, Skoetz N, Riley RD, et al. GRADE concept paper 2: Concepts for judging certainty on the calibration of prognostic models in a body of validation studies. *Journal of Clinical Epidemiology*. 2022;143:202-11.
55. Foroutan F, Mayer M, Guyatt G, Riley RD, Mustafa R, Kreuzberger N, et al. GRADE Concept Paper 8: Judging the certainty of discrimination performance estimates of prognostic models in a body of validation studies. *Journal of Clinical Epidemiology*. 2024;111344.
56. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence--indirectness. *Journal of Clinical Epidemiology*. 2011;64(12):1303-10.
57. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. *Journal of Clinical Epidemiology*. 2011;64(4):401-6.
58. Gray TF, Temel JS, El-Jawahri A. Illness and prognostic understanding in patients with hematologic malignancies. *Blood Reviews*. 2021;45:100692.
59. SEER. SEER*Explorer: An interactive website for SEER cancer statistics [Internet]. Data source(s): SEER Incidence Data, November 2022 Submission (1975-2020), SEER 22 registries.: Surveillance Research Program, National Cancer Institute; Available from: <https://seer.cancer.gov/statistics-network/explorer/>; 2022 [
60. Swerdlow S, Camp E, Harris N, Jaffe E, Pileri S, Stein H, Thiele J. WHO classification of tumours of haematopoietic and lymphoid tissues. Lyon, France: International Agency for Research on Cancer (IARC); 2017.
61. 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* (London, England). 2012;379(9828):1791-9.
62. 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.
63. Eichenauer DA, Aleman BMP, André M, Federico M, Hutchings M, Illidge T, et al. Hodgkin lymphoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up††FootnotesApproved by the ESMO Guidelines Committee: August 2002, last update December 2017. This publication supersedes the previously published version—*Annals of Oncology* 2014; 25(Suppl 3): iii70–iii75. *Annals of Oncology*. 2018;29:iv19-iv29.
64. 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.
65. 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. *The Lancet Haematology*. 2016;3(10):e467-e79.
66. 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(7):1160-8.
67. Adams HJA, Kwee TC. Interim FDG-PET does not predict outcome in advanced-stage Hodgkin lymphoma patients treated with BEACOPP. *British Journal of Haematology*. 2019;185(4):758-60.
68. 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 randomized EORTC/LYSA/FIL H10 trial. *Journal of Clinical Oncology*. 2014;32(12):1188-94.
69. Herishanu Y, Katz BZ, Lipsky A, Wiestner A. Biology of chronic lymphocytic leukemia in different microenvironments: clinical and therapeutic implications. *Hematology/oncology Clinics of North America*. 2013;27(2):173-206.
70. Binet JL, Auquier A, Dighiero G, Chastang C, Piguet H, Goasguen J, et al. A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. *Cancer*. 1981;48(1):198-206.
71. Rai KR, Sawitsky A, Cronkite EP, Chanana AD, Levy RN, Pasternack BS. Clinical staging of chronic lymphocytic leukemia. *Blood*. 1975;46(2):219-34.
72. Eichhorst B, Hallek M. Prognostication of chronic lymphocytic leukemia in the era of new agents. *Hematology American Society of Hematology Education Program*. 2016;2016(1):149-55.
73. von Tresckow B, Plütschow 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.

74. 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.
75. 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. *The Cochrane Database of Systematic Reviews*. 2017;5(5):Cd007941.
76. Meignan M, Gallamini A, Meignan M, Gallamini A, Haioun C. Report on the First International Workshop on interim-PET scan in lymphoma. *Leukemia & Lymphoma*. 2009;50(8):1257-60.
77. Andrea G, Sally FB, Alberto B, Stephane C, Lale K, Michele G, 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(6):1107-13.
78. 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(1):64-70.
79. 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. *The Cochrane Database of Systematic Reviews*. 2015;1(1):Cd010533.
80. 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.
81. 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. *Cochrane Database of Systematic Reviews*. 2020(1).
82. The International CLL-IPI working group. An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. *The Lancet Oncology*. 2016;17(6):779-90.
83. Delgado J, Doubek M, Baumann T, Kotaskova J, Molica S, Mozas P, et al. Chronic lymphocytic leukemia: A prognostic model comprising only two biomarkers (mutational status and FISH cytogenetics) separates patients with different outcome and simplifies the CLL-IPI. *American Journal of Hematology*. 2017;92(4):375-80.
84. Wierda WG, O'Brien S, Wang X, Faderl S, Ferrajoli A, Do KA, et al. Prognostic nomogram and index for overall survival in previously untreated patients with chronic lymphocytic leukemia. *Blood*. 2007;109(11):4679-85.
85. Pflug N, Bahlo J, Shanafelt TD, Eichhorst BF, Bergmann MA, Elter T, et al. Development of a comprehensive prognostic index for patients with chronic lymphocytic leukemia. *Blood*. 2014;124(1):49-62.
86. Rossi D, Rasi S, Spina V, Bruscaggin A, Monti S, Ciardullo C, et al. Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia. *Blood*. 2013;121(8):1403-12.
87. Stephens DM, Ruppert AS, Weirda WG, Jones JA, Woyach JA, Maddocks K, et al. Externally validated predictive clinical model for untreated del(17p13.1) chronic lymphocytic leukemia patients. *American Journal of Hematology*. 2015;90(11):967-9.

88. Baliakas P, Moysiadis T, Hadzidimitriou A, Xochelli A, Jeromin S, Agathangelidis A, et al. Tailored approaches grounded on immunogenetic features for refined prognostication in chronic lymphocytic leukemia. *Haematologica*. 2019;104(2):360-9.
89. Molica S, Mauro FR, Callea V, Gentile M, Giannarelli D, Lopez M, et al. A gender-based score system predicts the clinical outcome of patients with early B-cell chronic lymphocytic leukemia. *Leukemia & Lymphoma*. 2005;46(4):553-60.
90. Wierda WG, O'Brien S, Wang X, Faderl S, Ferrajoli A, Do K-A, et al. Multivariable model for time to first treatment in patients with chronic lymphocytic leukemia. *Journal of Clinical Oncology*. 2011;29(31):4088-95.
91. Morabito F, Cutrona G, Gentile M, Matis S, Todoerti K, Colombo M, et al. Definition of progression risk based on combinations of cellular and molecular markers in patients with Binet stage A chronic lymphocytic leukaemia. *British Journal of Haematology*. 2009;146(1):44-53.
92. Gentile M, Shanafelt TD, Cutrona G, Molica S, Tripepi G, Alvarez I, et al. A progression-risk score to predict treatment-free survival for early stage chronic lymphocytic leukemia patients. *Leukemia*. 2016;30(6):1440-3.
93. Kreuzberger N, Damen JAAG, Trivella M, Estcourt LJ, Aldin A, Umlauff L, et al. Prognostic models for newly-diagnosed chronic lymphocytic leukaemia in adults: a systematic review and meta-analysis. *Cochrane Database of Systematic Reviews*. 2020(7).
94. Fraga M, Forteza J. Diagnosis of Hodgkin's disease: an update on histopathological and immunophenotypical features. *Histology and Histopathology*. 2007;22(8):923-35.
95. Cochrane Handbook for Systematic Reviews of Interventions: Cochrane; 2024.
96. Higgins J, Altman D, Sterne J. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT GS, editor. *Cochrane Handbook for Systematic Reviews of Interventions Version 510* (updated March 2011): The Cochrane Collaboration; 2011.
97. Santesso N, Glenton C, Dahm P, Garner P, Akl EA, Alper B, et al. GRADE guidelines 26: informative statements to communicate the findings of systematic reviews of interventions. *Journal of Clinical Epidemiology*. 2020;119:126-35.
98. Kreuzberger N, Goldkuhle M, von Tresckow B, Kobe C, Sickinger MT, Monsef I, Skoetz N. Positron emission tomography-adapted therapy for first-line treatment in adults with Hodgkin lymphoma. *Cochrane Database of Systematic Reviews*. 2025(3).
99. Leitlinienprogramm Onkologie (Deutsche Krebsgesellschaft DK, AWMF),,. S3-Leitlinie Diagnostik, Therapie und Nachsorge des Hodgkin Lymphoms bei erwachsenen Patienten, Langversion 3.2, AWMF Registernummer: 018/029 OL. In: *Leitlinienprogramm Onkologie der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. (AWMF) dDKeVDudSDKD*, editor. 2022.
100. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Annals of Internal Medicine*. 2019;170(1):W1-w33.
101. Sauerbrei W, Haeussler T, Balmford J, Huebner M. Structured reporting to improve transparency of analyses in prognostic marker studies. *BMC Medicine*. 2022;20(1):184.
102. Junqueira DR, Zorzela L, Golder S, Loke Y, Gagnier JJ, Julious SA, et al. CONSORT Harms 2022 statement, explanation, and elaboration: updated guideline for the reporting of harms in randomized trials. *Journal of Clinical Epidemiology*. 2023;158:149-65.

103. Leitlinienprogramm Onkologie (Deutsche Krebsgesellschaft DK, AWMF). Konsultationsfassung, S3-Leitlinie Diagnostik, Therapie und Nachsorge für Patient*innen mit einer chronischen lymphatischen Leukämie (CLL). Version 2.01 – August 2024. AWMF-Registernummer: 018-032OL. 2024.
104. Dorando E, Piechotta V, Skoetz N, Moons KG, Riley RD, Wolff R, Kreuzberger N. Protocol for the development of a tool for assessing risk of bias in overall prognosis studies (ROB-OPS). Open Science Framework. 2021.
105. Foroutan F, Guyatt G, Trivella M, Kreuzberger N, Skoetz N, Riley RD, et al. GRADE concept paper 2: Concepts for judging certainty on the calibration of prognostic models in a body of validation studies. *Journal of Clinical Epidemiology*. 2021.
106. Foroutan F, Mayer M, Guyatt G, Riley RD, Mustafa R, Kreuzberger N, et al. GRADE concept paper 8: judging the certainty of discrimination performance estimates of prognostic models in a body of validation studies. *Journal of Clinical Epidemiology*. 2024;170:111344.
107. 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. *The New England Journal of Medicine*. 2007;357(19):1916-27.
108. Fuchs M, Jacob AS, Kaul H, Kobe C, Kuhnert G, Pabst T, et al. Follow-up of the GHSG HD16 trial of PET-guided treatment in early-stage favorable Hodgkin lymphoma. *Leukemia*. 2024;38(1):160-7.
109. Federico M, Fortpied C, Stepanishyna Y, Gotti M, van der Maazen R, Cristinelli C, et al. Long-term follow-Up of the response-adapted intergroup EORTC/LYSA/FIL H10 trial for localized Hodgkin lymphoma. *Journal of Clinical Oncology*. 2024;42(1):19-25.
110. Borchmann P, Plütschow A, Kobe C, Greil R, Meissner J, Topp MS, et al. PET-guided omission of radiotherapy in early-stage unfavourable Hodgkin lymphoma (GHSG HD17): a multicentre, open-label, randomised, phase 3 trial. *The Lancet Oncology*. 2021;22(2):223-34.
111. Kreissl S, Goergen H, Buehnen I, Kobe C, Moccia A, Greil R, et al. PET-guided eBEACOPP treatment of advanced-stage Hodgkin lymphoma (HD18): follow-up analysis of an international, open-label, randomised, phase 3 trial. *The Lancet Haematology*. 2021;8(6):e398-e409.
112. Johnson P, Federico M, Kirkwood A, Fosså A, Berkahn L, Carella A, et al. Adapted treatment guided by interim PET-CT scan in advanced Hodgkin's lymphoma. *The New England Journal of Medicine*. 2016;374(25):2419-29.
113. Eichenauer DA, Aleman BMP, André M, Federico M, Hutchings M, Illidge T, et al. Hodgkin lymphoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*. 2018;29:iv19-iv29.
114. Mkhwanazi ZA, Nyambuya TM, Mfusi SA, Nkambule BB. Prognostic markers in patients with chronic lymphocytic leukaemia on targeted therapy, chemoimmunotherapy with anti-CD20 monoclonal antibody: a systematic review and meta-analysis of prognostic factors. *BMC Cancer*. 2022;22(1):1218.
115. Goldkuhle M, Bender R, Akl EA, van Dalen EC, Nevitt S, Mustafa RA, et al. GRADE Guidelines: 29. Rating the certainty in time-to-event outcomes—Study limitations due to censoring of participants with missing data in intervention studies. *Journal of Clinical Epidemiology*. 2021;129:126-37.

116. Iannizzi C, Chai KL, Piechotta V, Valk SJ, Kimber C, Monsef I, et al. Convalescent plasma for people with COVID-19: a living systematic review. *Cochrane Database of Systematic Reviews*. 2023(2).

Acknowledgements

I would like to express my sincerest gratitude towards everyone who supported and encouraged me during the completion of this doctoral thesis project.

First, without the guidance and support of my supervisor Prof. Dr. Nicole Skoetz, this project would not have been possible. As my first full-time placement after studies, she has shaped both my academic and personal understanding of good scientific practice and a rigorous work ethic. With her permanent encouragement for development, and the numerous possibilities she facilitated regarding networking, conference attendances, and seminars on leadership, she has contributed tremendously to my professional and personal growth. I'd like to express a special thanks for making it easy to balance career development and family. I would also like to thank my tutors Prof. Dr. Bastian von Tresckow and Prof. Dr. Kerstin Rhiem for their continued support during the development of this manuscript.

A huge thanks to all my former and current colleagues at the Institute of Public Health (former Evidence-based Oncology working group and Cochrane Haematology). It's a great pleasure to work with this highly motivated and thoughtful team, to have methodological exchanges, and to see new exciting ideas developing in all possible health-related directions. In addition, I'd like to thank the members of the Cochrane Prognosis Methods group, and the GRADE prognosis group whose methodological discussions and support taught me a lot over the years. A huge thanks also to all co-authors who contributed to the reviews of this thesis, the staff at Cochrane who conducted the editorial process, and all peer reviewers for their invaluable input.

Lastly, I'd also like to thank my family. Especially my husband, for his unlimited positivity, for teaching me patience and many more essential life lessons, and for his tremendous support during the process.

Supplementary information

Attributions of Icons and Software

The meta-analyses in the publications were conducted as described in the respective publication, using either Review Manager, Review Manager Web, or R (packages metamisc and metafor for meta-analysis).

The fictional data and figures for this doctoral thesis were created using the software R (robvis, meta, metafor, survsim, viridis) and Microsoft Excel, and combined using Microsoft PowerPoint. Software and the source of individual images and icons are listed per illustration below. Multiple icons were adapted and downloaded from Freepik.

- Figure 1: Created in Microsoft PowerPoint. Icons included in Microsoft PowerPoint.
- Figure 2: Figure 2A created in Microsoft PowerPoint, icons from Microsoft PowerPoint. Figure 2B created with R. Data simulated with the package survsim. Figure 2C based on WHO data from 2019 (17) and represented with the R package viridis.
- Figure 3: Figure 3A created in Microsoft PowerPoint, icons from Microsoft PowerPoint. Figure 3B created with R. Data simulated with the package survsim.
- Figure 4: Figure 4A and B created in Microsoft PowerPoint. Figure 4C and D are graphs that were published as figure 2 in Alamer et al., 2020 [30] under CC BY 4.0 / Cropped from original.
- Figure 5: Created in Microsoft PowerPoint, icons from Microsoft PowerPoint
- Figure 6: Figure 6A and 6B were created in Microsoft Excel based on simulated data, and combined using Microsoft PowerPoint.
- Figure 7: Created in Microsoft PowerPoint. Icons included in Microsoft PowerPoint. Prognostic model icons by Iconjam and Becris Lineal.
- Figure 8: Created in Microsoft PowerPoint. Icons included in Microsoft PowerPoint.
- Figure 9: Created using the R package robvis, based on simulated data.
- Figure 10: Created using the R package metafor or metamisc. Based on fictional data.
- Figure 11: Created in Microsoft PowerPoint. Icons included in Microsoft PowerPoint.
- Figure 12: Visual abstract created in Microsoft PowerPoint. Icons for chemotherapy and radiotherapy by Freepik. Prognosis outcome by mia elysia. Prognostic model by Iconjam. Population icons by Freepik, Puckung and Iconsmind. PET-scan was adapted from figure 3b in Picardi et al., 2021 (66) [published under CCA 4.0 International License]. All other icons by Microsoft PowerPoint.
- Figure 13: Visual abstract created in Microsoft PowerPoint. Icons for chemotherapy, radiotherapy, and antibody by Freepik. Prognosis outcome by mia elysia. Sample icon by

Kiranshastry. PET-scan was adapted from figure 3b in Picardi et al., 2021 (66) [published under CCA 4.0 International License]. All other icons by Microsoft PowerPoint.

- Figure 14: Figure 22 from Kreuzberger et al. 2020 (93), included publication in this dissertation project.
- Figure 15: Figure 27 from Kreuzberger et al. 2020 (93), included publication in this dissertation project.
- Figure 16: Figure 28 from Kreuzberger 2020 (93), included publication in this dissertation project.
- Figure 17: Visual abstract created in Microsoft PowerPoint. Icon for prognostic model by Iconjam. Sample icons by Freepik, iconsmind, and Microsoft PowerPoint.
- Figure 18: Visual abstract created in Microsoft PowerPoint. Icons for chemotherapy, radiotherapy, and antibody by Freepik. Prognosis outcome by mia elysia. Sample icon by Kiranshastry. PET-scan was adapted from figure 3b in Picardi et al., 2021 (66) [published under CCA 4.0 International License]. All other icons by Microsoft PowerPoint.

Funding

Two out of three systematic reviews received direct funding for its conduct (publication 1 and publication 2) from the German Ministry of Education and Research.

- Publication 1 (Interim PET-results for prognosis in adults with Hodgkin lymphoma: a systematic review and meta-analysis of prognostic factor studies): Grant by the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, Grant no. 01KG1709), Germany
- Publication 2 (Prognostic models for newly-diagnosed chronic lymphocytic leukaemia in adults: a systematic review and meta-analysis): Grant by the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, Grant no. 01KG1711), Germany
- Publication 3 (Positron emission tomography-adapted therapy for first-line treatment in adults with Hodgkin lymphoma): This project did not receive any financial support.

Conflicts of Interest

The author of this doctoral thesis, Nina Kreuzberger, declares no financial conflicts of interest regarding the publications included in this manuscript. She is member of the editorial base of Cochrane Haematology. However, she was not involved in the editorial process of publications she led or contributed to. She is a member of the Cochrane Prognosis Methods group and GRADE Working Group since 2018 / 2019.

Appendices

Appendix I: Scientific Contribution of Authors

Publication 1: Interim PET-results for prognosis in adults with Hodgkin lymphoma: a systematic review and meta-analysis of prognostic factor studies

Authors: 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

Journal: Cochrane Library of Systematic Reviews

Scientific contribution by the doctoral candidate:

The doctoral candidate Nina Kreuzberger was a co-author on this project, being involved mainly with the assessment of risk of bias and participated in discussions around grading of the evidence with the GRADE prognosis group. The lead author of this systematic review was Angela Aldin, who developed her dissertation project based on this publication. All other co-authors contributed according to the table below (taken over from the original publication).

Author	Contribution
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.
Mariarena 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.

Publication 2: Prognostic models for newly-diagnosed chronic lymphocytic leukaemia in adults: a systematic review and meta-analysis

Authors: Kreuzberger N, Damen JAAG, Trivella M, Estcourt LJ, Aldin A, Umlauff L, Vazquez-Montes MDLA, Wolff R, Moons KG, Monsef I, Foroutan F, Kreuzer K, Skoetz N

Journal: Cochrane Library of Systematic Reviews

Scientific contribution by the doctoral candidate:

The doctoral candidate Nina Kreuzberger was the lead author of this publication, coordinating and conducting all steps of the systematic review. This includes the study selection for inclusion, the development of the data extraction form, data extraction and characterization of studies, risk of bias assessment, the meta-analysis using R, and the drafting of the manuscript. The co-authors contributed according to the table below (taken over from the original publication) either in conducting the second, independent screening of the literature, data extraction, and risk of bias assessment (MT, LJE, AA, LU, MVM, NS), contributed their clinical expertise (LE, KK), or advised on the specifics of prognostic model studies and reviews (JAAGD, MT, RW, KM, FF).

Author	Contribution
Nina Kreuzberger	Screening and selection of studies, development of data extraction form and data extraction, characteristics of studies, 'Risk of bias' assessment, statistical analysis, writing and drafting of the review, communication with and between authors
Johanna AAG Damen	Statistical analysis

Marialena Trivella	Statistical input, screening and selection of studies, data extraction, 'Risk of bias' assessment
Lise J Estcourt	Medical and content input, screening and selection of studies, data extraction, 'Risk of bias' assessment
Angela Aldin	Screening and selection of studies, data extraction, 'Risk of bias' assessment
Lisa Umlauff	Data extraction, 'Risk of bias' assessment, characteristics of included studies
Maria Vasquez Montes	Prototype of data extraction form, support in data extraction, 'Risk of bias' assessment
Robert Wolff	'Risk of bias' input
Karel Moons	Methodological input on reviews of prognosis studies
Ina Monsef	Search strategy development
Farid Foroutan	'Risk of bias' input
Karl-Anton Kreuzer	Medical and content input
Nicole Skoetz	Protocol development, screening and selection of studies, data extraction, 'Risk of bias' assessment, extensive proofread and comments on the review draft

Publication 3: Positron emission tomography-adapted therapy for first-line treatment in adults with Hodgkin lymphoma

Authors: Kreuzberger N, Goldkuhle M, von Tresckow B, Kobe C, Sickinger M, Monsef I, Skoetz N

Journal: Cochrane Library of Systematic Reviews

Scientific contribution by the doctoral candidate:

The doctoral candidate Nina Kreuzberger was the lead author of this publication, coordinating and conducting all steps of the systematic review. This includes the study selection for inclusion of studies, the development of the data extraction form, data extraction and characterization of studies, risk of bias assessment, the meta-analysis using RevMan Web, and the drafting of the manuscript. The co-authors contributed according to the table below (taken over from the original publication) either in conducting the second, independent screening of the literature, data extraction, and risk of bias assessment (MG, MTS, NS) or contributed their clinical expertise (BvT, CK, MTS, NS).

Author	Contribution
Nina Kreuzberger	Study screening, data extraction, risk of bias assessment, data analysis and interpretation, writing of the review
Marius Goldkuhle	Study screening, data extraction, risk of bias assessment, data analysis check
Bastian von Tresckow	Clinical expertise
Carsten Kobe	Clinical expertise
Marie-Therese Sickinger	Clinical expertise, screening, data extraction, risk of bias assessment, data analysis of studies included in the previous version
Ina Monsef	Search strategies
Nicole Skoetz	Study screening, data extraction, risk of bias assessment, data analysis and interpretation of the previous version of this review, proofreading of the protocol and review, statistical and methodological advice and supervision, content input

Appendix II: Publication 1



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](https://doi.org/10.1002/14651858.CD012643.pub3).

www.cochranelibrary.com

TABLE OF CONTENTS

HEADER	1
ABSTRACT	1
PLAIN LANGUAGE SUMMARY	3
SUMMARY OF FINDINGS	5
BACKGROUND	7
OBJECTIVES	8
METHODS	9
Figure 1.	11
RESULTS	14
Figure 2.	20
Figure 3.	22
Figure 4.	26
DISCUSSION	33
AUTHORS' CONCLUSIONS	37
ACKNOWLEDGEMENTS	37
REFERENCES	39
CHARACTERISTICS OF STUDIES	56
DATA AND ANALYSES	162
Analysis 1.1. Comparison 1: Univariable comparison of PET+ve vs. PET-ve, Outcome 1: Overall survival	162
Analysis 1.2. Comparison 1: Univariable comparison of PET+ve vs. PET-ve, Outcome 2: Progression-free survival	163
Analysis 2.1. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 1: OS by radiotherapy	165
Analysis 2.2. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 2: OS by study design	166
Analysis 2.3. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 3: OS by chemotherapy	167
Analysis 2.4. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 4: OS for PET/CT vs PET	167
Analysis 2.5. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 5: OS by disease stage	168
Analysis 2.6. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 6: Timing of interim PET ...	169
Analysis 2.7. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 7: OS by HR type of estimation	169
Analysis 3.1. Comparison 3: Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve, Outcome 1: PFS by study design	171
Analysis 3.2. Comparison 3: Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve, Outcome 2: PFS by chemotherapy	172
Analysis 3.3. Comparison 3: Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve, Outcome 3: PFS for PET/CT vs PET ..	173
Analysis 3.4. Comparison 3: Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve, Outcome 4: PFS by disease stage	174
Analysis 3.5. Comparison 3: Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve, Outcome 5: PFS by radiotherapy	175
Analysis 3.6. Comparison 3: Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve, Outcome 6: Timing of interim PET...	176
Analysis 3.7. Comparison 3: Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve, Outcome 7: PFS by HR type of estimation.....	177
APPENDICES	177
WHAT'S NEW	185
HISTORY	185
CONTRIBUTIONS OF AUTHORS	185
DECLARATIONS OF INTEREST	186
SOURCES OF SUPPORT	186
DIFFERENCES BETWEEN PROTOCOL AND REVIEW	187
INDEX TERMS	187

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

Angela Aldin¹, Lisa Umlauff¹, Lise J Estcourt², Gary Collins³, Karel GM Moons⁴, Andreas Engert⁵, Carsten Kobe⁶, Bastian von Tresckow⁵, Madhuri Haque¹, Farid Foroutan⁷, Nina Kreuzberger¹, Marialena Trivella^{3a}, Nicole Skoetz^{8b}

¹Cochrane 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. ²Haematology/Transfusion Medicine, NHS Blood and Transplant, Oxford, UK. ³Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK. ⁴Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands. ⁵Department I of Internal Medicine, Center for Integrated Oncology Aachen Bonn Cologne Duesseldorf, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. ⁶University of Cologne, Faculty of Medicine and University Hospital Cologne, Department for Nuclear Medicine, Cologne, Germany. ⁷Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada. ⁸Cochrane 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

^acontributed equally. ^bcontributed 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](https://doi.org/10.1002/14651858.CD012643.pub3).

Copyright © 2020 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

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](https://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, IR = 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, IR = 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

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.

Outcome	Anticipated absolute effects* (95% CI)		Relative (95% CI)	A of participants	Certainty of the evidence	Comment
	Risk with Interim PET-negative	Risk with Interim PET-positive				
Overall survival	Low		HR 5.09 (2.64 to 9.81)	1802 (9 studies)	⊕⊕⊕⊕	
Follow up: 3 years	900 per 1.000 ¹	585 per 1.000 ¹ (356 to 757)			MODERATE ^{2 3 4}	
	High					
	980 per 1.000 ¹	902 per 1.000 ¹ (820 to 948)				
Progression-free survival	Low		HR 4.90 (3.47 to 6.90)	2079 (14 studies)	⊕⊕⊕⊕	
Follow up: 3 years	850 per 1.000 ⁵	451 per 1.000 ⁵ (326 to 569)			VERY LOW ^{6 7 8}	
	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 heterogeneous 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](https://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.

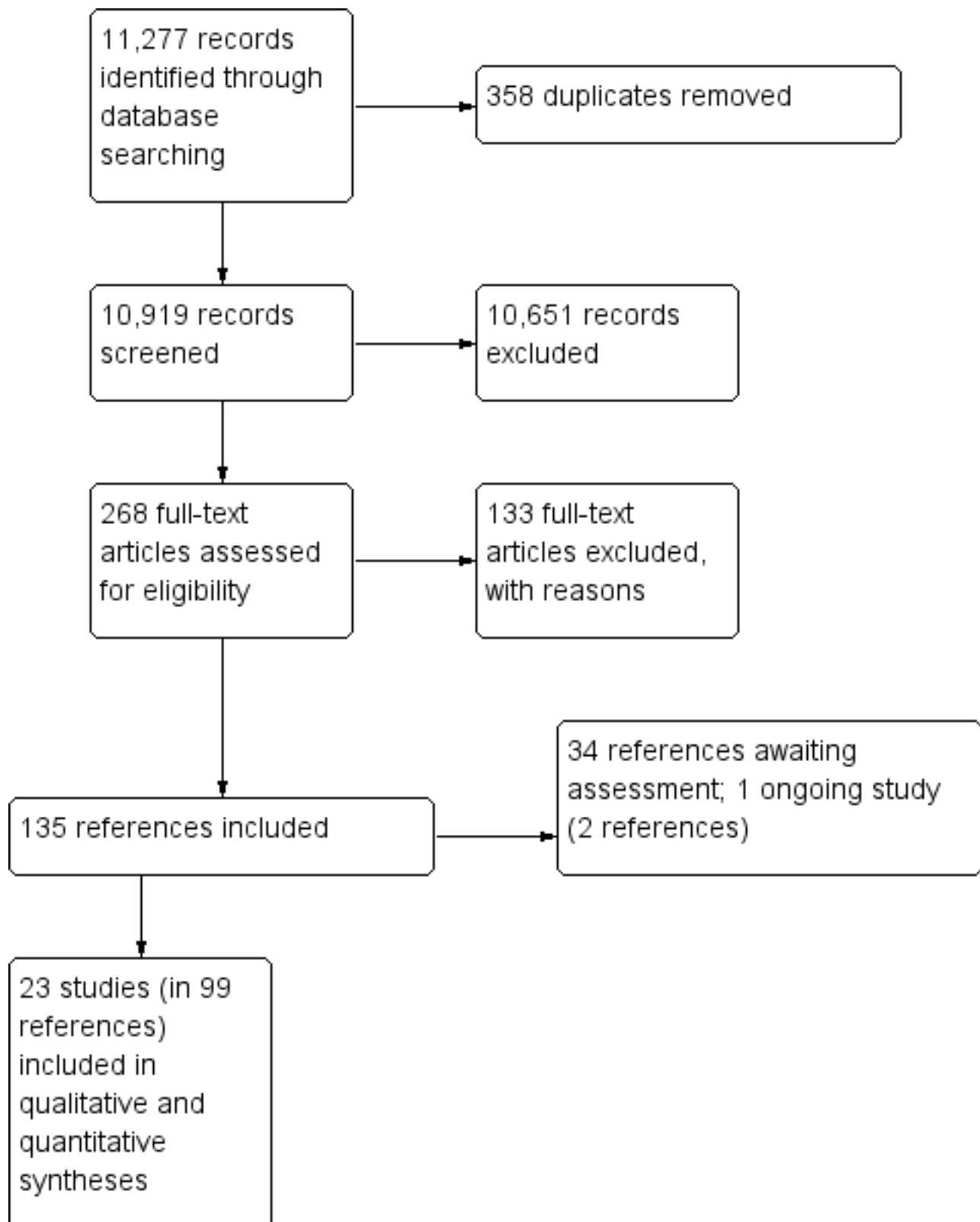
Data collection and analysis

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 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 Tau² 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

FiNeen 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 $\leq$ 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 recalculated 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 recalculated 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; Orlicchio 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; Yagur 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
Progression-free survival	Zaucha 2017	Low	Low	Moderate	Low	High	Low
	Zinzani 2011	Low	Low	Low	Low	High	Low
	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 (Zauch 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 (Zauch 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; Zauch 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 (Zauch 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; Zauch 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

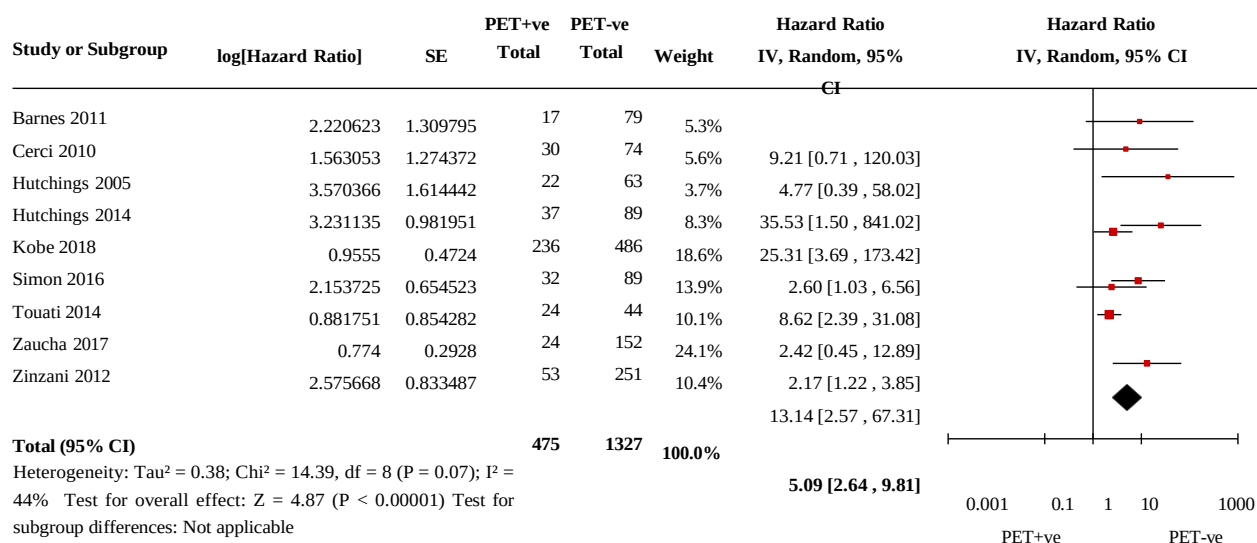
studies that were not included in meta-analyses, we reported the main results narratively in this review.

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](#); [Zauch a 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



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](#); [Zauch a 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 over-estimation 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	<u>PET2 results (N = 398)</u> Intention-to-treat analysis 5-year OS for entire arm = 95.2% (95% CI 91.1 to 97.4), 13 events Per-protocol 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	<u>PET2 results (N = 77)</u> PET-negative N = 61, no deaths PET-positive N = 16, 2 deaths Logrank test for difference between groups: P < .01 <u>PET4 results (N = 64)</u> 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	<u>Interim PET-positivity (DS 4)</u> 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	<u>Interim PET-positivity</u> 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									

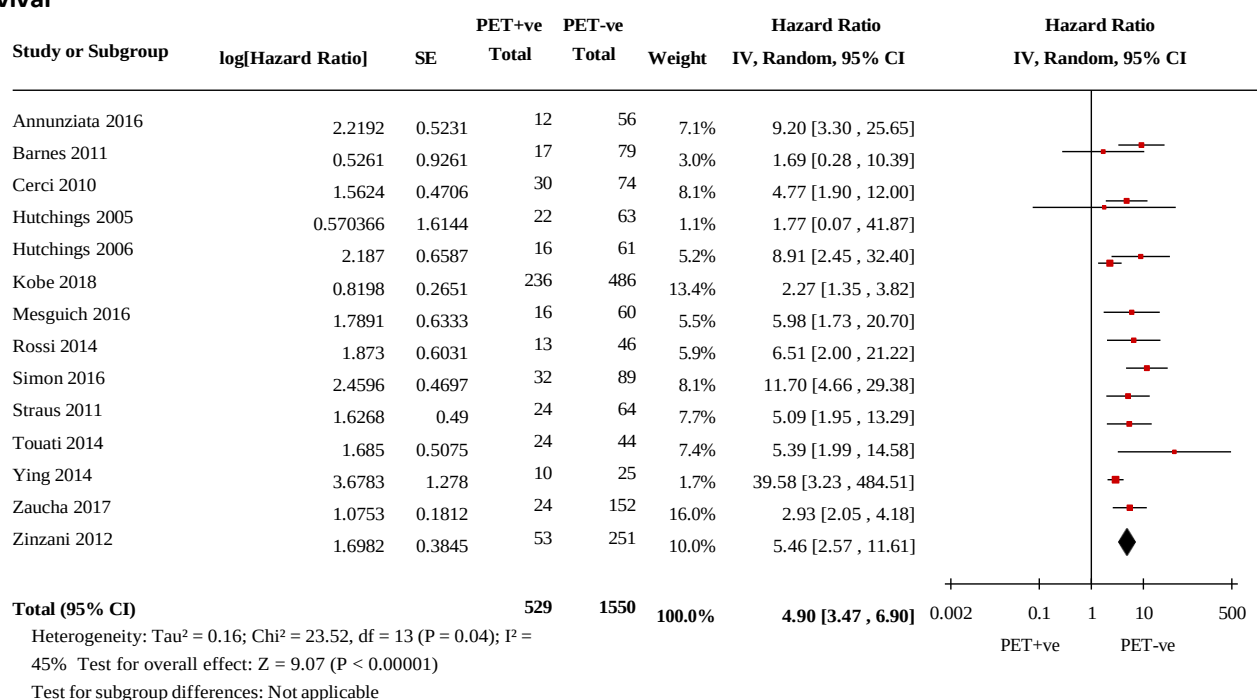
Progression-free survival (PFS)

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, IR = 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



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	<u>*PET-negative</u> Favourable group: N = 2 events (both relapses) in the ABVD + IN-RT 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. <u>*PET-positive</u> 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	<u>PET2 results (N = 398)</u> 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) <u>Results for entire standard arm</u> 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)	<u>PET1 results (N = 126)</u> 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 <u>PET1 vs. PET2 results (N = 89)</u> 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	<u>3-year PFS rates in PET2-negative versus PET-positive by disease subgroups</u> 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	<u>PET2</u> 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	<u>Early interim PET</u> 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	<i>Model 1 (interim PET2 + clinical stage + extranodal disease)</i> <u>PET2</u> HR = 36.281 (95% CI 7.179 to 183.4), P < .001 <i>Model 2 (interim PET2 + extranodal disease)</i> <u>PET2</u> 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	<u>Interim PET-positivity (DS 4)</u> 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	-	-
<i>Model 1 (interim PET + disease stage)</i> <u>Positive interim PET</u> HR = 3.73 (95% CI 1.35 to 10.35), P = 0.0112 <i>Model 2 (interim PET + bulky disease)</i> <u>Positive interim PET</u> 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
<u>SUVmax PET0-PET2</u> 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
<u>Interim PET-positivity</u> 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	<u>Analysis of imaging at different time points</u> <i>Baseline imaging</i> 77 participants had baseline PET-CT scans, 1 had only chest X-ray due to pregnancy at baseline <i>Imaging during and at the end of treatment</i> 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 <i>Follow-up imaging</i> - 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 <u>Need for surveillance imaging</u> 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).	<u>Interim PET results</u> - Negative interim PET2: 104 - Positive interim PET2: 28 <u>End PET results</u> <i>Negative interim PET2 group</i> - Negative final PET: 102/104 - Positive final PET: 2/104 <i>Positive interim PET2 group</i> - 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 IR 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 heterogeneous 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; Orlicchio

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, Orlicchio 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). Contrasting 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 improved, 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 (115)

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.

Cottreau AS, Versari A, LoN 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, Casanovas 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**(Suppl 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, Czereminski 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, Cottreau 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, LoN 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, Huttman 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.

Zauchna JM, Chauvie S, Malkowski B, Warszevska 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.

* Zauchna 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.

Zauchna 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 with 18F-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, Gregorian 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, Althoefer 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, Mashlach 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, Mashlach 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, Mashlach 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, Neuberger 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, Neuberger 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.

Haïoun 2005 {published data only}

Haïoun 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, LoN 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, Obrtlíkova 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, Pytlík 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, Althoefer 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, Klenner P. [Predictive features of positron emission tomography after two cycles of induction therapy in malignant lymphoma]. *Casopis Lékarů Českých [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, Attali A, Duggento A, Chiaravallotti 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, CliNon-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, Gregorian 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, Kachchi 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 CHIVPP/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):S72-3.

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, Bansa A, Giraldo P. The importance of PET/CT as method of evaluation of early response to treatment in HL. *Haematologica* 2012;**97**(Suppl. 1):S66.

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, LoN 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}

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, Gaftor-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 Borgh 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](https://doi.org/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, HooN 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](https://doi.org/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, Barıkta 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, Dedekova 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, CroN 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](https://doi.org/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, HooN 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, CroN 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](https://doi.org/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](https://doi.org/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, Gherzi 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](https://doi.org/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, Cottureau 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)

	<u>Median follow-up time (range)</u> • 55 months
Participants	<u>Number of included participants</u> • Total: 1925 • Randomised to standard treatment without change in protocol because of interim PET: 954 <u>Inclusion criteria</u> • Previously untreated • Classic supradiaphragmic stage I and II HL • Age 15 to 70 years <u>Exclusion criteria</u> • 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 2007¹) <u>Consent</u> • Yes; written informed consent <u>Recruitment period</u> • November 2006 to June 2011 <u>Age (range, in years)</u> • Favourable, standard treatment group median: 31 (15-49) • Unfavourable, standard treatment group median: 32 (15-70) <u>Ethnic group(s)</u> • Not reported <u>Stages of disease</u> • Early stages (I and II) <u>Comorbidities</u> • Not reported, except for the exclusion criteria <u>Therapy regimen</u> • ABVD and radiotherapy depending on treatment arm, favourable/unfavourable disease, and early PET (ePET) positivity
Prognostic factor(s)	<u>Prognostic factor(s)</u> • Early PET (ePET) <u>Definition of prognostic factor(s)</u> • 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 in univariate analysis 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, accommodation, expenses received from Genentech, Takeda, Gilead Sciences
- Brice P: research funding received from Merck Sharp & Dohme Oncology, Takeda; travel, accommodation, expenses received from Takeda
- Specht L: consulting or advisory role at Takeda; research funding received from Varian Medical Systems; travel, accommodation, 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, accommodation, 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, accommodation, 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), Associazione 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	<u>Secondary citation(s)</u> - NA <u>Language of publication</u> - English <u>Study design</u> - Retrospective, single-centre study <u>Study centre(s)</u> - Not reported <u>Country</u> - Italy <u>Median follow-up time (range)</u> - Not reported
Participants	<u>Number of included participants</u> 68 <u>Inclusion criteria</u> - HL diagnosis <u>Exclusion criteria</u> - Not reported <u>Consent</u> - Not reported <u>Recruitment period</u> - January 2007 to December 2014 <u>Age (range, in years)</u> 39 (16-72) <u>Ethnic group(s)</u> - 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 2014¹)

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)

	<u>Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)?</u> Not reported
Missing data	<u>Participants with any missing value?</u> - No <u>If yes, how were missing data handled?</u> - Not applicable
Analysis	Univariable analysis: Yes <u>Total number of participants included in univariable analysis for each outcome</u> - PFS: all <u>Statistical method</u> - Receiver operating characteristic (ROC) approach - Kaplan-Meier (survival analysis) - Log-rank (differences between groups) - Cox proportional hazards model <u>How was the prognostic factor treated?</u> - Binary Multivariable analysis: No
Risk of bias (QUIPS)	<u>Study participation</u> - Unclear - Clear description of participants and study characteristics. No inclusion and exclusion criteria provided. <u>Study attrition</u> - Unclear risk - No loss to follow-up reported. No length of follow-up reported. <u>Prognostic factor measurement</u> - Low risk - Adequate measurement and description. Prognostic factor measured the same way for all participants. Outcome: Overall survival <i>Not reported</i> Outcome: Progression-free survival <u>Outcome measurement</u> - Low risk - Clear definition. Outcome measured the same way for all participants. <u>'Other prognostic factors (covariates)'</u> - 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)	<u>Primary outcome(s) and definition(s)</u> • Progression-free survival (PFS), defined as the time from randomisation to first progression, relapse, or death from any cause or last follow-up <u>Secondary outcome(s) and definition(s)</u> • 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 <u>Timing of outcome measurement</u> • PFS: at 5 years <u>Was the same definition and method for measurement used in all participants?</u> • Yes <u>Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)?</u> • Not reported
Missing data	<u>Participants with any missing value?</u> • Yes; N = 11 stopped treatment before PET2, and further N = 14 stopped treatment before PET4 <u>If yes, how were missing data handled?</u> • All 413 participants included in ITT analysis, N = 412 included in safety analysis, N = 372 included in per-protocol analysis
Analysis	Univariable analysis: Yes <u>Total number of participants included in univariable analysis for each outcome</u> • PFS, OS: N = 413 in ITT analysis, N = 372 in per-protocol analysis <u>Statistical method</u> • 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, Cellegne, 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 Cellegne, Gilead, Janssen, and Roche/Genentech. RD has received personal fees from Bristol-Myers Squibb, Cellegne, 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)

	<u>Study centre(s)</u> - São Paulo University Clinics Hospital, Brazil <u>Country</u> - Brazil <u>Median follow-up time (range)</u> 36 months (32-40)
Participants	<u>Number of included participants</u> 104 <u>Inclusion criteria</u> - Newly diagnosed, biopsy-proven, classic HL <u>Exclusion criteria</u> - Pregnancy <u>Consent</u> - Yes; written <u>Recruitment period</u> - August 2005 to December 2007 <u>Age (range, in years)</u> 28 (13-82) <u>Ethnic group(s)</u> - Not reported <u>Stages of disease</u> - All stages <u>Comorbidities</u> - Not reported <u>Therapy regimen</u> - 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)	<u>Prognostic factor(s)</u> - Interim PET <u>Definition of prognostic factor(s)</u> - Whole-body PET scan <u>Timing of prognostic factor measurement</u> - After 2 cycles of ABVD, as late as possible within the week before start of cycle 3 <u>Method for measurement (use of specific scale and cut-off)</u>

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	<u>Secondary citation(s)</u> - Agostinelli 2016, Biggi 2013, Gallamini 2006, Gallamini 2007 <u>Language of publication</u> - English <u>Study design</u> - Retrospective, international, multi-centre study (17 centres) <u>Study centre(s)</u> 17 academic institutions worldwide <u>Countries</u> - Various <u>Median follow-up time (range)</u> 37 months (2-110)
Participants	<u>Number of included participants</u> 260 <u>Inclusion criteria</u> - 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 <u>Exclusion criteria</u> - Missing CT data, baseline PET, interim PET, CT or PET slices; poor quality PET images; miscellaneous reasons (n=9) <u>Consent</u> - No; due to retrospective study design <u>Recruitment period</u> - January 2002 to December 2009 <u>Age (range, in years)</u> 37.3 (14-82) <u>Ethnic group(s)</u> - Not reported <u>Stages of disease</u> - Early stage unfavourable (IIA HL with adverse prognostic factors) - Advanced stages (IIB – IVB)

Gallamini 2014 (Continued)

	<u>Comorbidities</u> - Not reported <u>Therapy regimen</u> 4-8 cycles ABVD with or without involved-field radiotherapy or consolidation radiotherapy
Prognostic factor(s)	<u>Prognostic factor(s)</u> - Interim PET <u>Definition of prognostic factor(s)</u> - Not reported <u>Timing of prognostic factor measurement</u> - A median of 12.3 days (range, 7-22) after cycle 2 of ABVD <u>Method for measurement (use of specific scale and cut-off)</u> - 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 <u>Was the same definition and method for measurement used in all participants?</u> - Yes <u>Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)?</u> - Yes
Outcome(s)	<u>Primary outcome(s) and definition(s)</u> - 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 <u>Secondary outcome(s) and definition(s)</u> - Inter-observer agreement using the 5-PS for PET2 interpretation <u>Timing of outcome measurement</u> - At 3 years <u>Was the same definition and method for measurement used in all participants?</u> - Yes <u>Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)?</u> - Yes
Missing data	<u>Participants with any missing value?</u> No <u>If yes, how were missing data handled?</u>

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	<u>Conflict of interest</u> • Not reported <u>Funding</u> • 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	<u>Secondary citation(s)</u> • NA <u>Language of publication</u> • English <u>Study design</u> • Retrospective study

Gandikota 2015 (Continued)

	<u>Study centre(s)</u> - Not reported <u>Country/Countries</u> - Not reported <u>Median follow-up time (range)</u> 46 months (24-126)
Participants	<u>Number of included participants</u> 78 <u>Inclusion criteria</u> - 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 <u>Exclusion criteria</u> - None <u>Consent</u> - No <u>Recruitment period</u> - January 2000 to December 2012 <u>Age (range, in years)</u> 43 (median; 22-86) <u>Ethnic group(s)</u> - Not reported <u>Stages of disease</u> - Early stages (IA to IIB) <u>Comorbidities</u> - Not reported <u>Therapy regimen</u> - ABVD (number of cycles based on risk factors and institutional guidelines) followed by involved-field or extended-field radiotherapy
Prognostic factor(s)	<u>Prognostic factor(s)</u> - Interim PET <u>Definition of prognostic factor(s)</u> - PET-CT scan (from base of the skull to upper thigh) <u>Timing of prognostic factor measurement</u>

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)	<u>Primary outcome(s) and definition(s)</u> • Outcomes relevant to this review were not explored in the study <u>Secondary outcome(s) and definition(s)</u> • Not applicable <u>Timing of outcome measurement</u> • Not applicable <u>Was the same definition and method for measurement used in all participants?</u> • Not applicable <u>Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)?</u> • Not applicable
Missing data	<u>Participants with any missing value?</u> • Yes: one patient without baseline PET due to pregnancy; one patient without detectable disease on the baseline scan (excision of single site disease) <u>If yes, how were missing data handled?</u> • 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)	<i>No risk of bias assessment, since outcomes relevant to this review were not explored in the study.</i>
Notes	<u>Conflict of interest</u> • The authors made no disclosure. <u>Funding</u> • No specific funding was disclosed.

Hutchings 2005

Study characteristics

Methods	<u>Secondary citation(s)</u> • NA <u>Language of publication</u> • English <u>Study design</u> • Not reported <u>Study centre(s)</u> • Guy's and St. Thomas' Hospital, London <u>Country</u> • UK <u>Median follow-up time (range)</u> • 40.2 months (6-125)
Participants	<u>Number of included participants</u> • 85 <u>Inclusion criteria</u> • Histologically-confirmed HL • Early interim FDG-PET scans <u>Exclusion criteria</u> • None <u>Consent</u> • Not reported <u>Recruitment period</u> • May 1993 to January 2004 <u>Age (range, in years)</u> • 36.7 (15-73) <u>Ethnic group(s)</u> • Not reported <u>Stages of disease</u> • All stages <u>Comorbidities</u> • Not reported <u>Therapy regimen</u>

Hutchings 2005 (Continued)

- According to departmental protocols: mainly ABVD, number of cycles not reported; additional radio-therapy depending on stage and site of HL
- Alternative therapy for participants without satisfactory remission during initial chemotherapy

Prognostic factor(s)	<u>Prognostic factor(s)</u> • Interim PET <u>Definition of prognostic factor(s)</u> • Half-body PET scan (mid-brain to upper thigh) <u>Timing of prognostic factor measurement</u> • 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 <u>Method for measurement (use of specific scale and cut-off)</u> • 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 <u>Was the same definition and method for measurement used in all participants?</u> • Yes <u>Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)?</u> • Not reported
Outcome(s)	<u>Primary outcome(s) and definition(s)</u> • 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 <u>Secondary outcome(s) and definition(s)</u> • None <u>Timing of outcome measurement</u> • At 2 and 5 years <u>Was the same definition and method for measurement used in all participants?</u> • Yes <u>Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)?</u> • Not reported
Missing data	<u>Participants with any missing value?</u> • No <u>If yes, how were missing data handled?</u> • 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	<u>Secondary citation(s)</u> • Borchmann 2017 <u>Language of publication</u> • English <u>Study design</u> • Open-label, international, randomised phase 3 trial <u>Study centre(s)</u> • 301 hospitals and private practices in five European countries <u>Countries</u> • Germany, Switzerland, Austria, the Netherlands, Czech Republic <u>Median follow-up time (range)</u> • Not reported for entire study population
Participants	<u>Number of included participants</u> • Total: 2101 • Qualified for randomisation: 1945 <u>Inclusion criteria</u> • 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 <u>Exclusion criteria</u> • 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)	<u>Primary outcome(s) and definition(s)</u> • Progression-free survival (PFS), defined as the time from completion of staging until progression, relapse, or death from any cause <u>Secondary outcome(s) and definition(s)</u> • Overall survival (OS), defined as time from completion of staging until death from any cause <u>Timing of outcome measurement</u> • At 3 years <u>Was the same definition and method for measurement used in all participants?</u> • Yes <u>Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)?</u> • Not reported
Missing data	<u>Participants with any missing value?</u> • 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 <u>If yes, how were missing data handled?</u> • Participants with missing data were excluded from analysis
Analysis	Univariable analysis: Yes <u>Total number of participants included in univariable analysis for each outcome</u> • OS: 722 • PFS: 722 <u>Statistical method</u> • Kaplan-Meier (survival analysis) • Cox regression analysis (hazard ratios) <u>How was the prognostic factor treated?</u> • Binary Multivariable analysis: Yes <u>Total number of participants included in multivariable analysis for each outcome</u> • OS: 722 • PFS: 722 <u>Statistical method</u> • 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	<u>Conflict of interest</u> • We declare no competing interests. <u>Funding</u> • 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	<u>Secondary citation(s)</u> • Markova 2009 <u>Language of publication</u> • English <u>Study design</u> • Retrospective, single-centre study <u>Study centre(s)</u> • Prague, institution not reported <u>Country</u> • Czech Republic <u>Median follow-up time (range)</u> • 52 months
Participants	<u>Number of included participants</u> • 69 <u>Inclusion criteria</u> • Newly diagnosed, histologically proven HL • Clinical stage IIB with large mediastinal mass and/or extranodal disease, stage III or IV • Age 18-60 years <u>Exclusion criteria</u> • 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 ($\pm$ 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 $\geq$ 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)	<u>Primary outcome(s) and definition(s)</u> - Progression-free survival (PFS), defined as the time from diagnosis to the first evidence of progression or relapse, or death from any cause <u>Secondary outcome(s) and definition(s)</u> - None <u>Timing of outcome measurement</u> - After cycle 4, 6/8 and 3 months after completion of chemotherapy <u>Was the same definition and method for measurement used in all participants?</u> - Yes <u>Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)?</u> - Not reported
Missing data	<u>Participants with any missing value?</u> - No <u>If yes, how were missing data handled?</u> - Not applicable
Analysis	Univariable analysis: Yes <u>Total number of participants included in univariable analysis for each outcome</u> - PFS: all <u>Statistical method</u> - Kaplan-Meier (survival analysis) - Log-rank test (comparison between groups) <u>How was the prognostic factor treated?</u> - Binary Multivariable analysis: No
Risk of bias (QUIPS)	<u>Study participation</u> - Low risk - All eligible participants included. Clear description of participants and study characteristics. Consecutive sampling. Inclusion and exclusion criteria provided. <u>Study attrition</u> - Low risk - No loss to follow-up. <u>Prognostic factor measurement</u> - 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)	<u>Primary outcome(s) and definition(s)</u> • Progression-free survival (PFS), defined as the time from diagnosis to either failure of first-line treatment, relapse or death <u>Secondary outcome(s) and definition(s)</u> • None <u>Timing of outcome measurement</u> • At 5 years <u>Was the same definition and method for measurement used in all participants?</u> • Yes <u>Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)?</u> • Not reported
Missing data	<u>Participants with any missing value?</u> • No <u>If yes, how were missing data handled?</u> • NA
Analysis	Univariable analysis: Yes <u>Total number of participants included in univariate analysis for each outcome</u> • PFS: all <u>Statistical method</u> • Kaplan Meier analysis curve • Log-rank test <u>How was the prognostic factor treated?</u> • Binary Multivariable analysis: Yes <u>Total number of participants included in multivariable analysis for each outcome</u> • PFS: all <u>Statistical method</u> • Cox proportional hazard models <u>How was the prognostic factor treated?</u> • Binary <u>Number of candidate covariates</u> • 3 <u>List of all candidate covariates</u>

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)

	<u>Therapy regimen</u> ABVD with or without radiotherapy
Prognostic factor(s)	<u>Prognostic factor(s)</u> - Interim PET <u>Definition of prognostic factor(s)</u> - Not reported <u>Timing of prognostic factor measurement</u> - After 2 or 3 cycles of ABVD <u>Method for measurement (use of specific scale and cut-off)</u> - Deauville 5-point scoring system - Scores of 1-3 considered negative, scores of 4-5 considered positive - Independent assessment by 3 nuclear medicine physicians <u>Was the same definition and method for measurement used in all participants?</u> - No, 10 participants had only PET without CT scan <u>Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)?</u> - Yes
Outcome(s)	<u>Primary outcome(s) and definition(s)</u> - Progression-free survival (PFS), defined as the time from diagnosis to disease progression, relapse or death from any cause <u>Secondary outcome(s) and definition(s)</u> - None <u>Timing of outcome measurement</u> - At 3 years <u>Was the same definition and method for measurement used in all participants?</u> - Yes <u>Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)?</u> - No
Missing data	<u>Participants with any missing value?</u> - No <u>If yes, how were missing data handled?</u> - Not applicable
Analysis	Univariable analysis: Yes <u>Total number of participants included in univariable analysis for each outcome</u> PFS: all <u>Statistical method</u>

Okii 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	<u>Secondary citation(s)</u> • NA <u>Language of publication</u> • English <u>Study design</u> • Retrospective, multi-centre study (6 centres) <u>Study centre(s)</u> • 6 centres in London, UK <u>Country</u> • UK <u>Median follow-up time (range)</u> • 27 months
Participants	<u>Number of included participants</u> • 23 <u>Inclusion criteria</u> • Newly diagnosed, histologically confirmed classic HL • Advanced stage • HIV positivity <u>Exclusion criteria</u> • None <u>Consent</u> • Not reported <u>Recruitment period</u> • June 2007 to August 2010 <u>Age (range, in years)</u> • 42 (median, 32-60) <u>Ethnic group(s)</u> • Not reported <u>Stages of disease</u> • Advanced stages: stage III –IV or stage IIB with at least one adverse prognostic factor <u>Comorbidities</u> • HIV positive participants only <u>Therapy regimen</u>

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 jirovecii*

Prognostic factor(s)	<u>Prognostic factor(s)</u> • Interim PET <u>Definition of prognostic factor(s)</u> • Half-body PET-CT scan <u>Timing of prognostic factor measurement</u> • After 2-3 cycles of ABVD, within the week before start of the next cycle <u>Method for measurement (use of specific scale and cut-off)</u> • 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 <u>Was the same definition and method for measurement used in all participants?</u> • Yes <u>Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)?</u> • Not reported
Outcome(s)	<u>Primary outcome(s) and definition(s)</u> • Progression-free survival (PFS), defined as the time from diagnosis to disease progression or relapse or last follow-up <u>Secondary outcome(s) and definition(s)</u> • 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 <u>Timing of outcome measurement</u> • At 2 years <u>Was the same definition and method for measurement used in all participants?</u> • Yes <u>Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)?</u> • Not reported
Missing data	<u>Participants with any missing value?</u> • No <u>If yes, how were missing data handled?</u> • Not applicable
Analysis	Univariable analysis: Yes <u>Total number of participants included in univariate analysis for each outcome</u>

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	<u>Secondary citation(s)</u> • NA <u>Language of publication</u> • English <u>Study design</u> • Retrospective, single-centre study <u>Study centre(s)</u> • Policlinico Universitario TorVergata, Rome, Italy <u>Country</u> • Italy <u>Median follow-up time (range)</u> • Not reported
Participants	<u>Number of included participants</u> • 132 <u>Inclusion criteria</u> • HL diagnosis based on biochemical tests and bone marrow biopsy • PET-MDCT staging examination, interim PET-MDCT and end of treatment PET-MDCT performed <u>Exclusion criteria</u> • None <u>Consent</u> • Not reported <u>Recruitment period</u> • January 2005 to June 2010 <u>Age (range, in years)</u> • 34 (mean, 16-74) <u>Ethnic group(s)</u> • Not reported <u>Stages of disease</u> • All stages <u>Comorbidities</u> • Not reported <u>Therapy regimen</u>

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)	<u>Prognostic factor(s)</u> • Interim PET <u>Definition of prognostic factor(s)</u> • PET scan from pelvis to head <u>Timing of prognostic factor measurement</u> • At the end of the second ABVD cycle <u>Method for measurement (use of specific scale and cut-off)</u> • International Harmonization Project guidelines • Rated by a radiologist and nuclear medicine specialist, confirmation by semi-quantitative analysis <u>Was the same definition and method for measurement used in all participants?</u> • Yes <u>Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)?</u> • Not reported
Outcome(s)	<u>Primary outcome(s) and definition(s)</u> • 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 <u>Secondary outcome(s) and definition(s)</u> • None <u>Timing of outcome measurement</u> • At the end of treatment <u>Was the same definition and method for measurement used in all participants?</u> • Yes <u>Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)?</u> • Not reported
Missing data	<u>Participants with any missing value?</u> • No <u>If yes, how were missing data handled?</u> • 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)	<i>No risk of bias assessment, since outcomes relevant to this review were not explored in this study.</i>
Notes	<u>Conflict of interest</u> • None <u>Funding</u> • Not reported

Rossi 2014

Study characteristics

Methods	<u>Secondary citation(s)</u> • NA <u>Language of publication</u> • English <u>Study design</u> • Retrospective, single-centre study <u>Study centre(s)</u> • Hospital of Dijon, France <u>Country</u> • France <u>Median follow-up time (range)</u> • 50 months (22-71)
Participants	<u>Number of included participants</u> • 59 <u>Inclusion criteria</u> • First diagnosis of classic HL <u>Exclusion criteria</u> • 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
- oSUVmax (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

- oSUVmax (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)	<u>Primary outcome(s) and definition(s)</u> • Overall survival (OS), not defined • Progression-free survival (PFS), not defined <u>Secondary outcome(s) and definition(s)</u> • None <u>Timing of outcome measurement</u> • At 5 years after diagnosis <u>Was the same definition and method for measurement used in all participants?</u> • Yes <u>Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)?</u> • Not reported
Missing data	<u>Participants with any missing value?</u> • No <u>If yes, how were missing data handled?</u> • NA
Analysis	Univariable analysis: Yes <u>Total number of participants included in univariable analysis for each outcome</u> • OS: all • PFS: all <u>Statistical method</u> • Kaplan-Meier (survival analysis) • Log-rank test (comparison between groups) • Cox proportional hazard model (effect of variants on survival) <u>How was the prognostic factor treated?</u> • Binary Multivariable analysis: Yes <u>Total number of participants included in multivariable analysis for each outcome</u> • 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)	<u>Primary outcome(s) and definition(s)</u> - 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 <u>Secondary outcome(s) and definition(s)</u> - None <u>Timing of outcome measurement</u> - At 5 years <u>Was the same definition and method for measurement used in all participants?</u> - Yes <u>Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)?</u> - Not reported
Missing data	<u>Participants with any missing value?</u> - No <u>If yes, how were missing data handled?</u> - Not applicable
Analysis	Univariable analysis: Yes <u>Total number of participants included in univariable analysis for each outcome</u> - PFS: 68 - OS: 68 <u>Statistical method</u> - Kaplan-Meier (survival analysis) - Chi-squared test or t-test (differences between groups) - ANOVA (comparison of means) <u>How was the prognostic factor treated?</u> - Binary Multivariable analysis: No
Risk of bias (QUIPS)	<u>Study participation</u> - 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). <u>Study attrition</u> - Low risk - All participants with available interim PET included. <u>Prognostic factor measurement</u>

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)

	<u>Timing of prognostic factor measurement</u> - After 2 to 4 cycles of treatment <u>Method for measurement (use of specific scale and cut-off)</u> - Interpretation of scans by 2 experienced PET-CT physicians - Scale and cut-off not reported <u>Was the same definition and method for measurement used in all participants?</u> - Not reported <u>Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)?</u> - Not reported
Outcome(s)	<u>Primary outcome(s) and definition(s)</u> - Progression-free survival (PFS), defined as the interval from diagnosis to first signs of tumour progression, patient death, or end of follow-up <u>Secondary outcome(s) and definition(s)</u> - None <u>Timing of outcome measurement</u> - At 3 years <u>Was the same definition and method for measurement used in all participants?</u> - Unclear, follow-up was conducted via telephone and/or outpatient visits <u>Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)?</u> - Not reported
Missing data	<u>Participants with any missing value?</u> - Only 35/50 participants underwent interim PET <u>If yes, how were missing data handled?</u> - Not reported
Analysis	Univariable analysis: Yes <u>Total number of participants included in univariable analysis for each outcome</u> - PFS: 35 <u>Statistical method</u> - Kaplan-Meier curves and life tables (survival analysis) - Log-rank tests (comparison between groups) <u>How was the prognostic factor treated?</u> - Binary Multivariable analysis: No
Risk of bias (QUIPS)	<u>Study participation</u> 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)	<u>Primary outcome(s) and definition(s)</u> • Progression-free survival (PFS), not defined <u>Secondary outcome(s) and definition(s)</u> • Kinetics of response <u>Timing of outcome measurement</u> • At 3 years <u>Was the same definition and method for measurement used in all participants?</u> • Yes <u>Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)?</u> • Yes
Missing data	<u>Participants with any missing value?</u> • Yes; only 198 participants had PET2 scans <u>If yes, how were missing data handled?</u> • Not reported
Analysis	Univariable analysis: Yes <u>Total number of participants included in univariable analysis for each outcome</u> • 286 (PET1) / 198 (PET2) <u>Statistical method</u> • Kaplan-Meier (survival analysis) • Cox proportional hazard regression analysis (HR between treatment groups) <u>How was the prognostic factor treated?</u> • Binary Multivariable analysis: No
Risk of bias (QUIPS)	<u>Study participation</u>

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.
Bucerus 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.
Haoun 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
Huelten Schmidt 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)
	<u>Aim</u>
	To establish the prognostic value of interim PET scans in limited stage patients with non-bulky disease
	<u>Study design</u>
	Retrospective
	<u>Country/treatment centre(s)</u>

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 $\geq 1/3$ 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) <u>Aim</u> • 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 <u>Study design</u> • Not reported <u>Country/treatment centre(s)</u> • International study <u>Number of included participants</u> • 262 <u>Inclusion criteria</u> • Individuals diagnosed with stage IIB-IV HL <u>Exclusion criteria</u> • None <u>Treatment</u> • ABVD <u>Primary outcome measure(s)</u> • 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. <u>Aim</u> • To evaluate the significance of iPET in elderly individuals with HL <u>Study design</u> • Retrospective study (1998 to 2016) <u>Country/treatment centre(s)</u> • Unclear, multicentre study (5 centres) <u>Number of included participants</u> • 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 ChIVPP/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 ChIVPP/ABVVP regimen.

(Conference abstract)

Aim

- To evaluate the prognostic value of interim 18 FDG-PET in advanced HL patients treated with intensified ChIVPP/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 ChIVPP/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) <u>Aim</u> - To assess how FDG-PET is currently used in individuals with HL <u>Study design</u> - Unclear, 2006 to 2008 <u>Country/treatment centre(s)</u> - Italy (Cancer Registries in Modena, Ferrara, Parma and Reggio Emilia) <u>Number of included participants</u> 136 <u>Inclusion criteria</u> - Diagnosed HL, adults (18 to 75 years), HIV negative <u>Exclusion criteria</u> - None <u>Treatment</u> - Not reported <u>Primary outcome measure(s)</u> - Overall survival - Relapse-free survival - Failure-free survival

Luminari 2011

Notes	Title: 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 Authors need to be contacted to clarify whether the treatment has been adapted based on the interim PET results. <u>Aim</u> - 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. <u>Study design</u> - Retrospective <u>Country/treatment centre(s)</u> - 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)
- Histologically confirmed 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.

(Conference 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	<u>Study design</u> - Randomised clinical trial (phase III) including 1150 participants with HL <u>Country/treatment centre</u> 1st Department of Medicine, Cologne University Hospital, Cologne, Germany <u>Number of included participants</u> - Total: 1150 <u>Inclusion criteria</u> - 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 <u>Exclusion criteria</u> - Leukocytes $< 3000/\mu\text{l}$ - Platelets $< 100000/\mu\text{l}$ - Hodgkin lymphoma as composite lymphoma - Activity index (WHO) > 2 <u>Arms and interventions</u> - Active comparator (A): two cycles ABVD followed by 20 Gy IF-RT, irrespective of FDG-PET results after chemotherapy - Experimental (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 <u>Primary outcome measure(s)</u> - Progression-free survival - Time frame: five years <u>Secondary outcome measure(s)</u> - Overall survival - Acute toxicity - Late toxicity - Complete response rate - Time frame: five years <u>Estimated study completion date</u> - 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

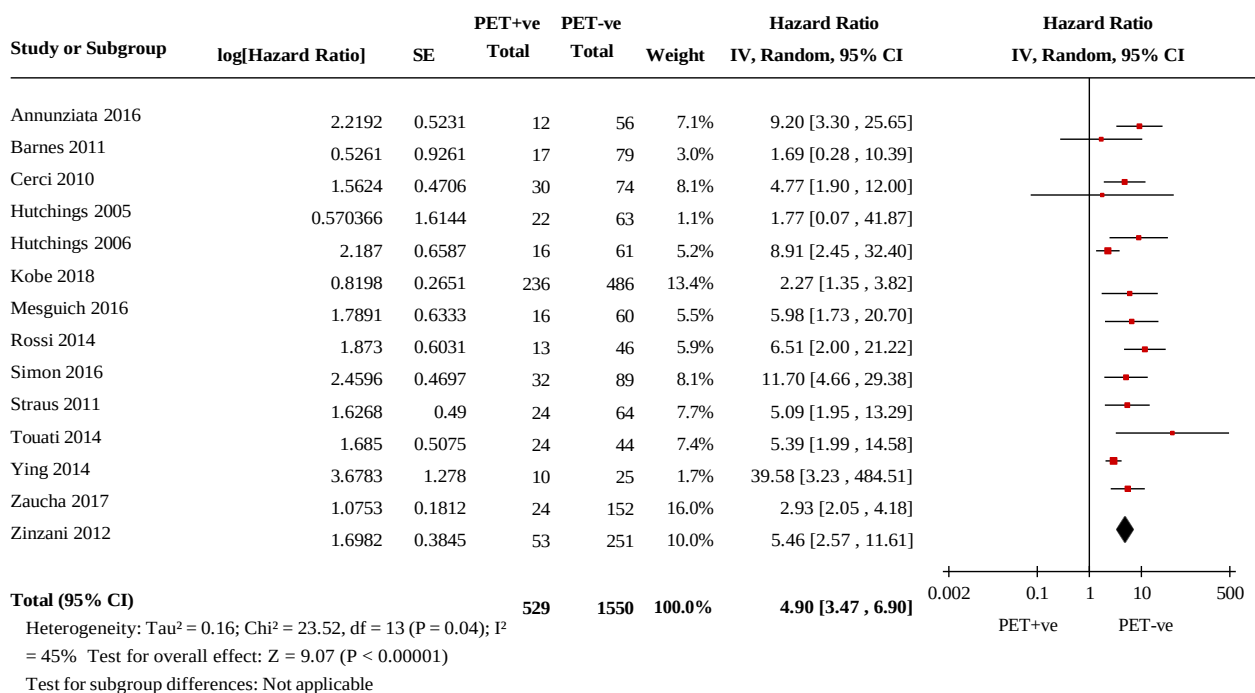
Study or Subgroup	log[Hazard Ratio]	SE	PET+ve Total	PET-ve Total	Weight	Hazard Ratio IV, Random, 95% CI	Hazard Ratio IV, Random, 95% CI
Barnes 2011	2.220623	1.309795	17	79	5.3%		
Cerci 2010	1.563053	1.274372	30	74	5.6%	9.21 [0.71 , 120.03]	
Hutchings 2005	3.570366	1.614442	22	63	3.7%	4.77 [0.39 , 58.02]	
Hutchings 2014	3.231135	0.981951	37	89	8.3%	35.53 [1.50 , 841.02]	
Kobe 2018	0.9555	0.4724	236	486	18.6%	25.31 [3.69 , 173.42]	
Simon 2016	2.153725	0.654523	32	89	13.9%	2.60 [1.03 , 6.56]	
Touati 2014	0.881751	0.854282	24	44	10.1%	8.62 [2.39 , 31.08]	
Zaucha 2017	0.774	0.2928	24	152	24.1%	2.42 [0.45 , 12.89]	
Zinzani 2012	2.575668	0.833487	53	251	10.4%	2.17 [1.22 , 3.85]	
Total (95% CI)			475	1327	100.0%	13.14 [2.57 , 67.31]	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

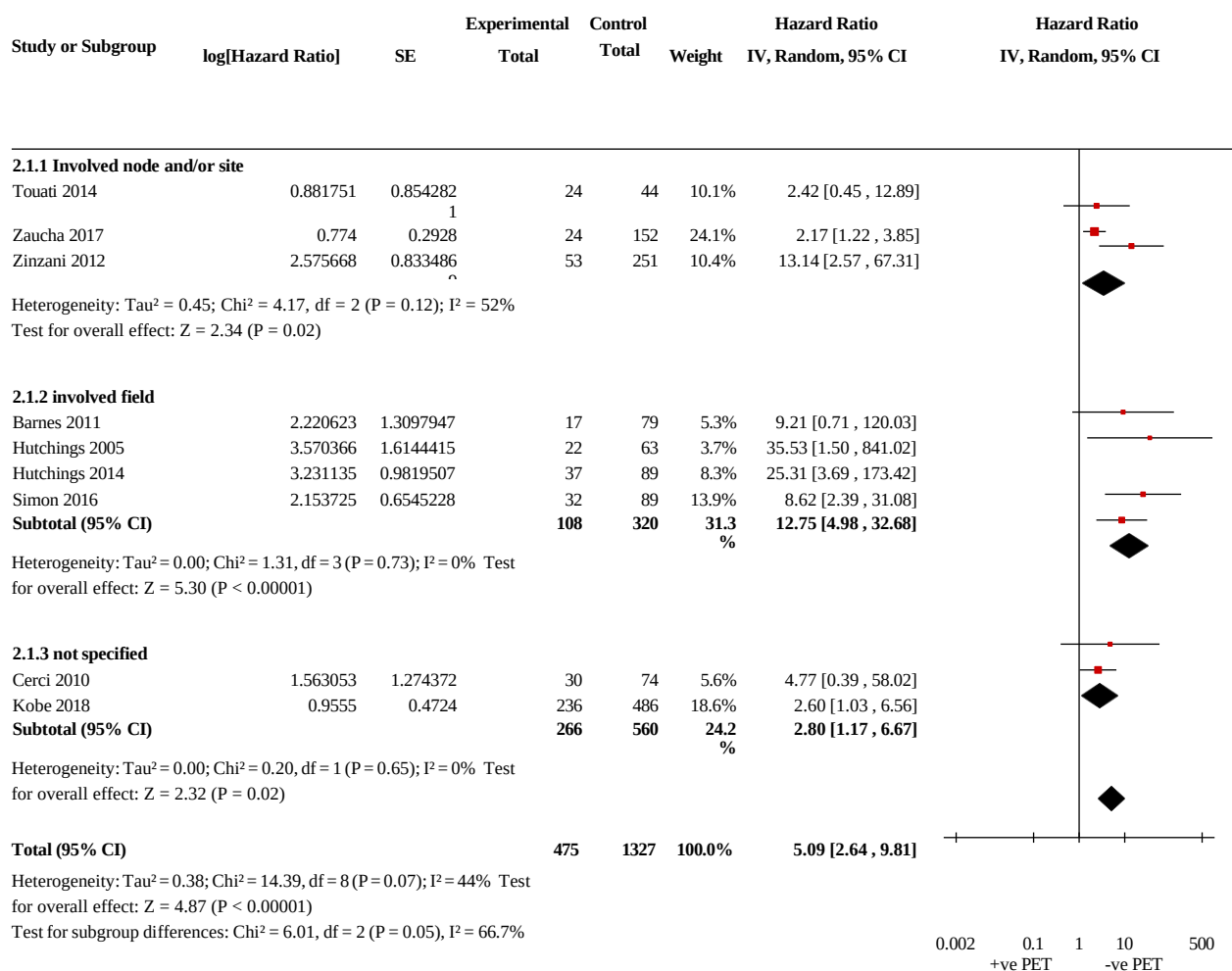
0.001 0.1 1 10 1000
PET+ve PET-ve

Analysis 1.2. Comparison 1: Univariable comparison of PET+ve vs. PET-ve, Outcome 2: Progression-free survival**Comparison 2. Subgroups in univariable comparison of OS: PET+ve vs. PET-ve**

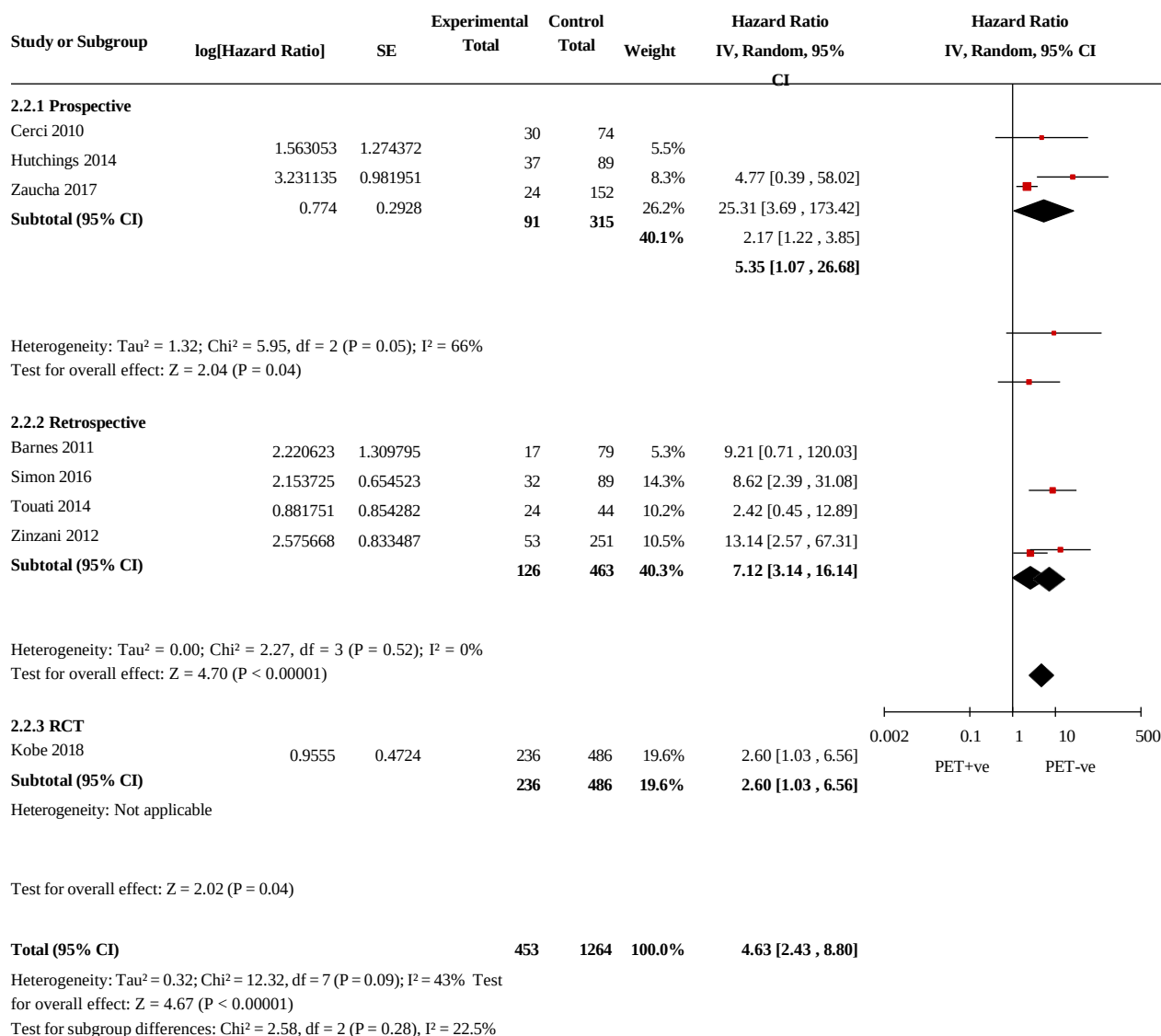
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
2.1 OS by radiotherapy	9	1802	Hazard Ratio (IV, Random, 95% CI)	5.09 [2.64, 9.81]
2.1.1 Involved node and/or site	3	548	Hazard Ratio (IV, Random, 95% CI)	3.45 [1.22, 9.72]
2.1.2 involved field	4	428	Hazard Ratio (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]

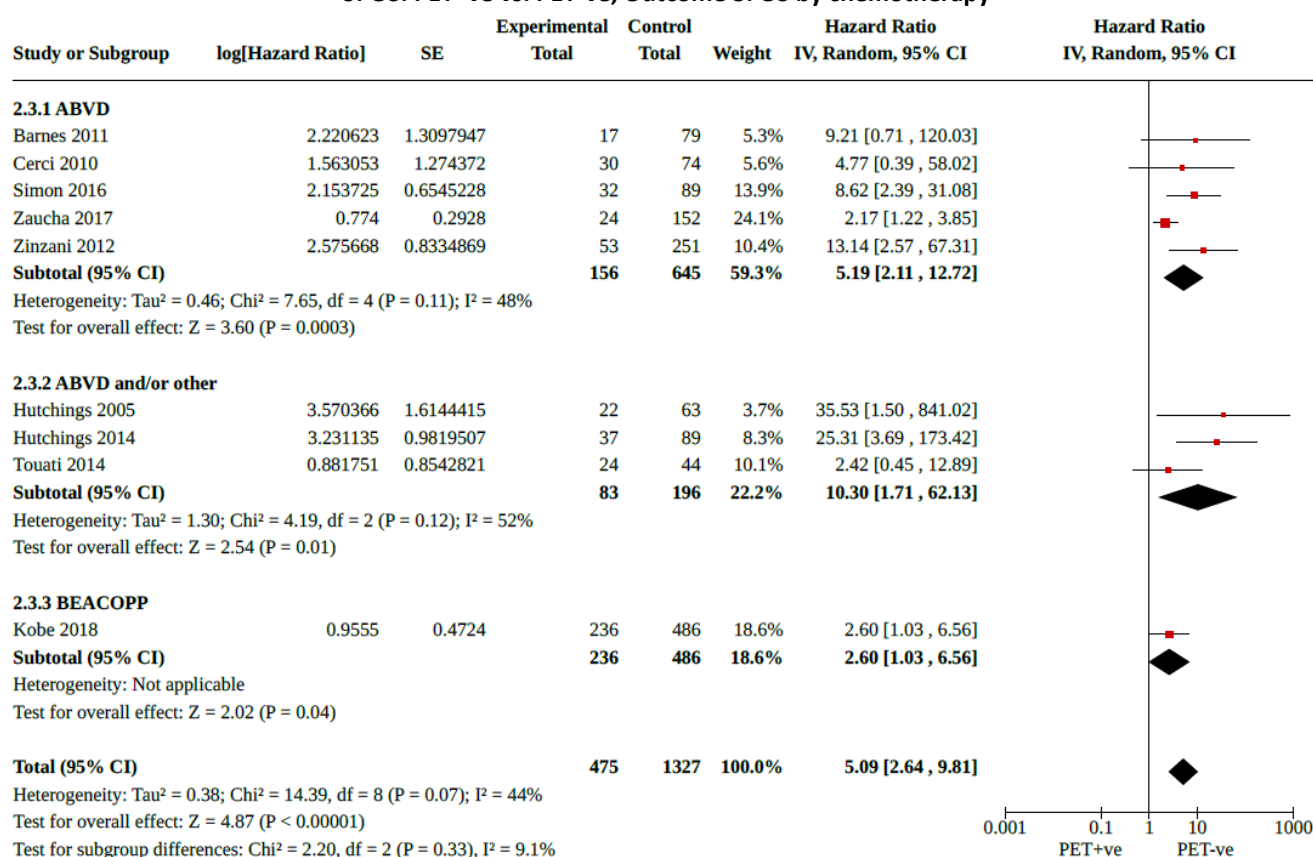
Analysis 2.1. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 1: OS by radiotherapy



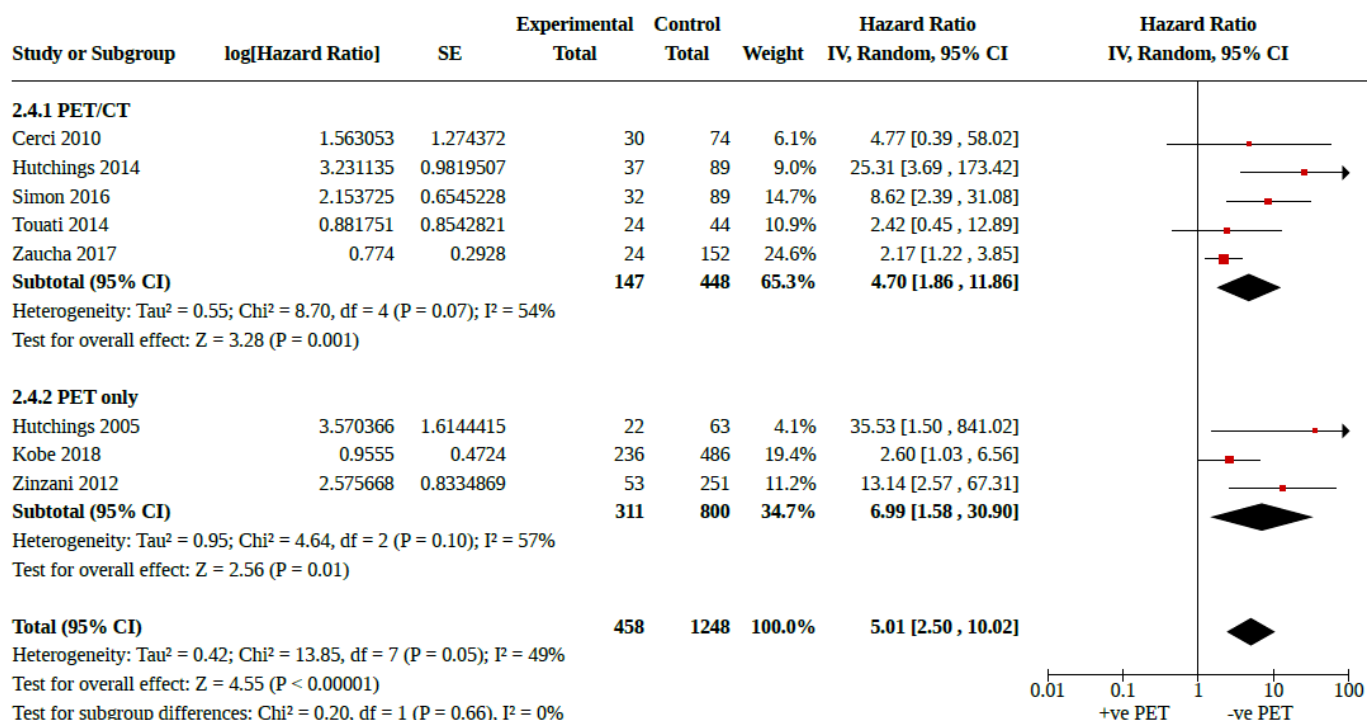
Analysis 2.2. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 2: OS by study design



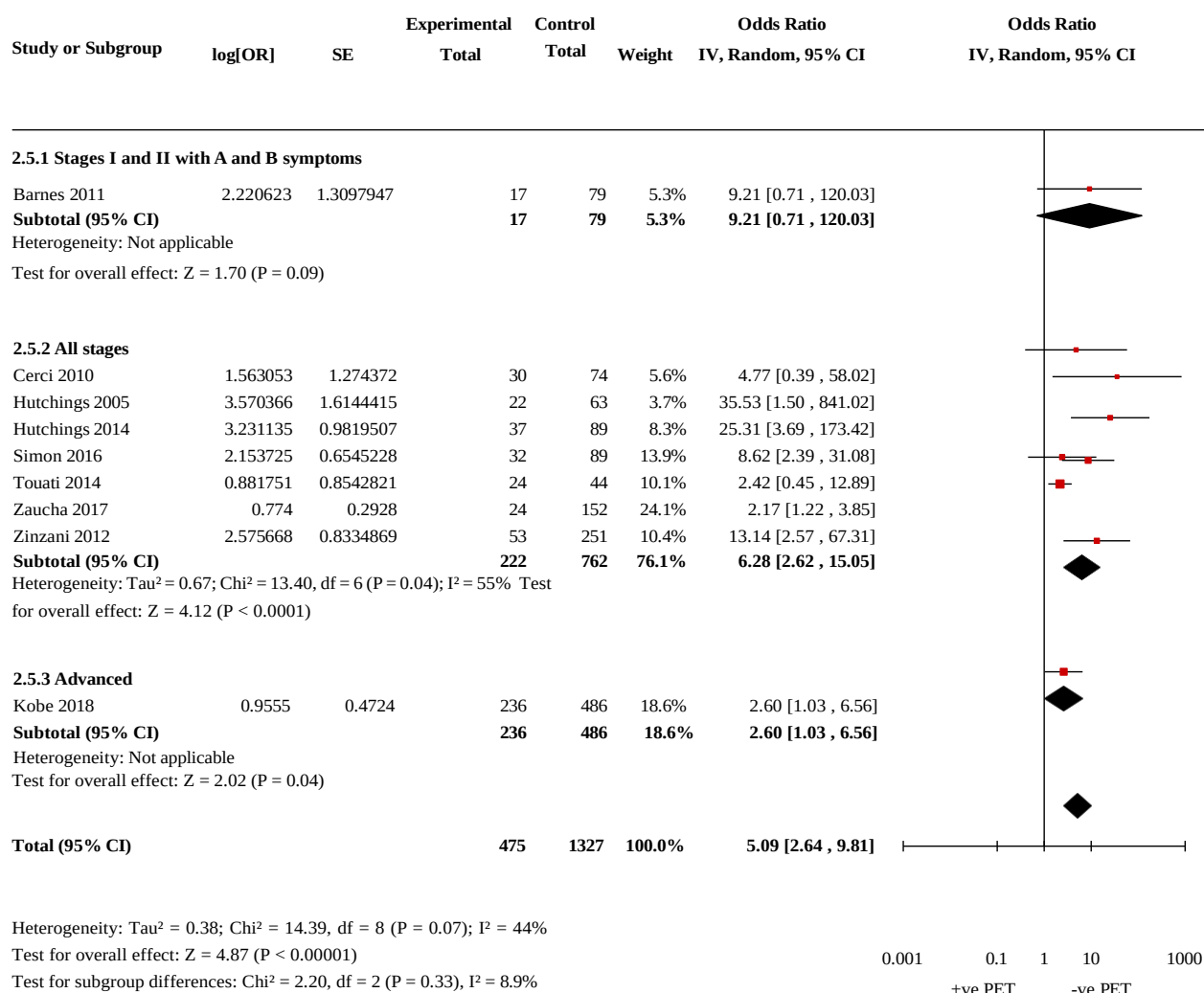
Analysis 2.3. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 3: OS by chemotherapy



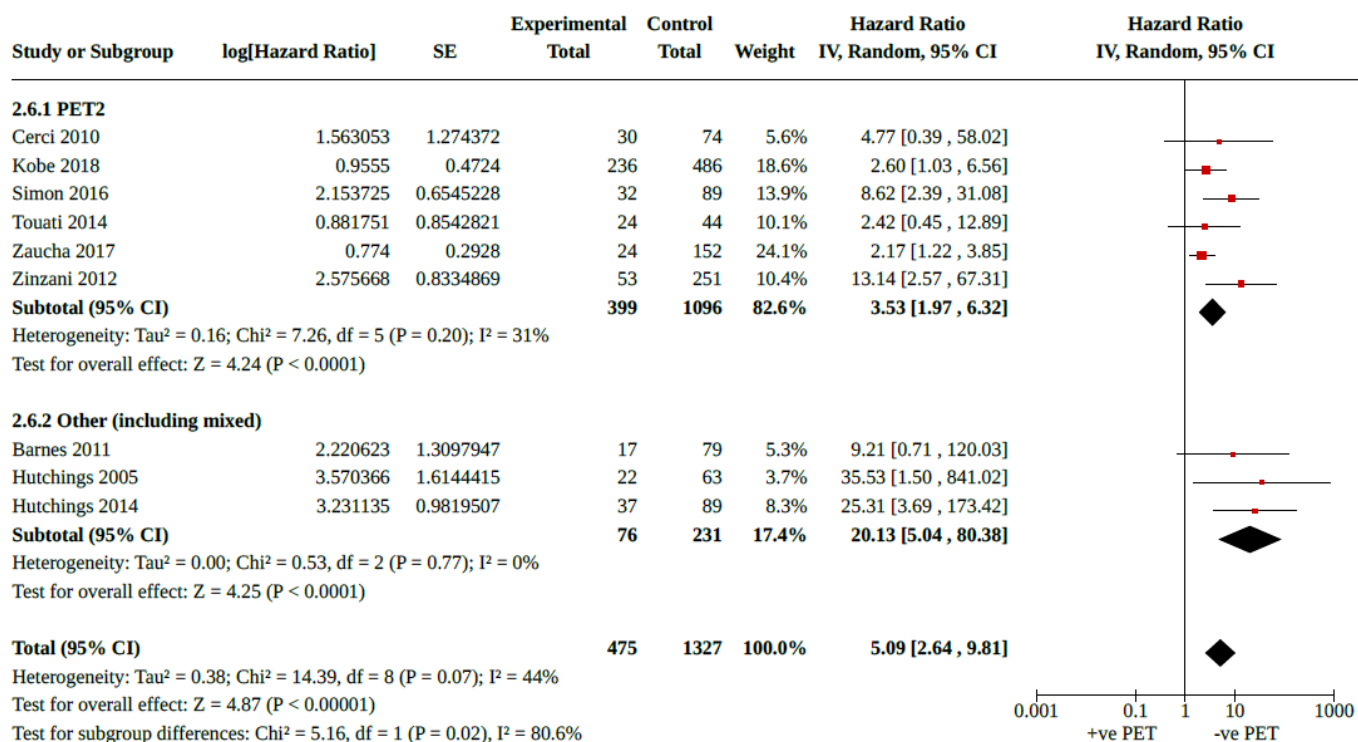
Analysis 2.4. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 4: OS for PET/CT vs PET



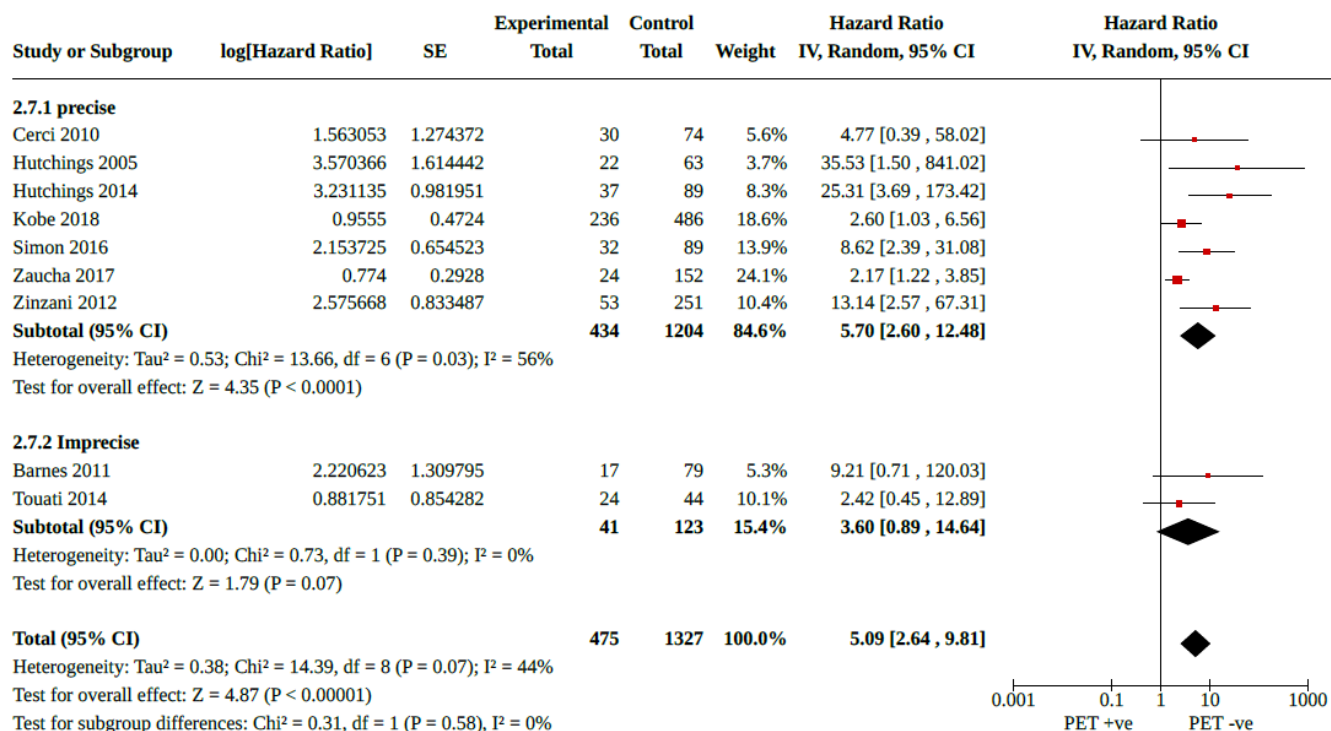
Analysis 2.5. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 5: OS by disease stage



Analysis 2.6. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 6: Timing of interim PET



Analysis 2.7. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 7: OS by HR type of estimation

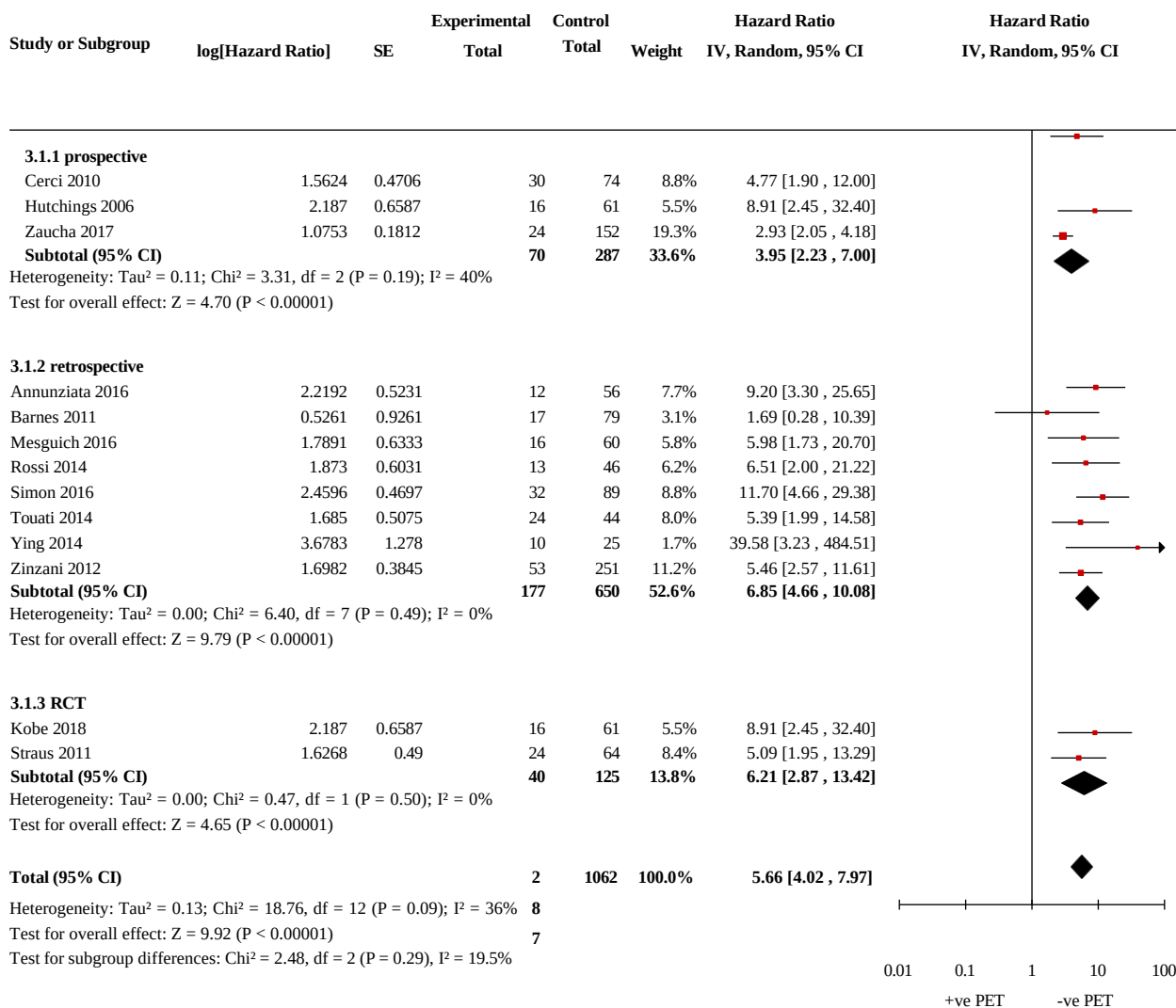


Comparison 3. Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve

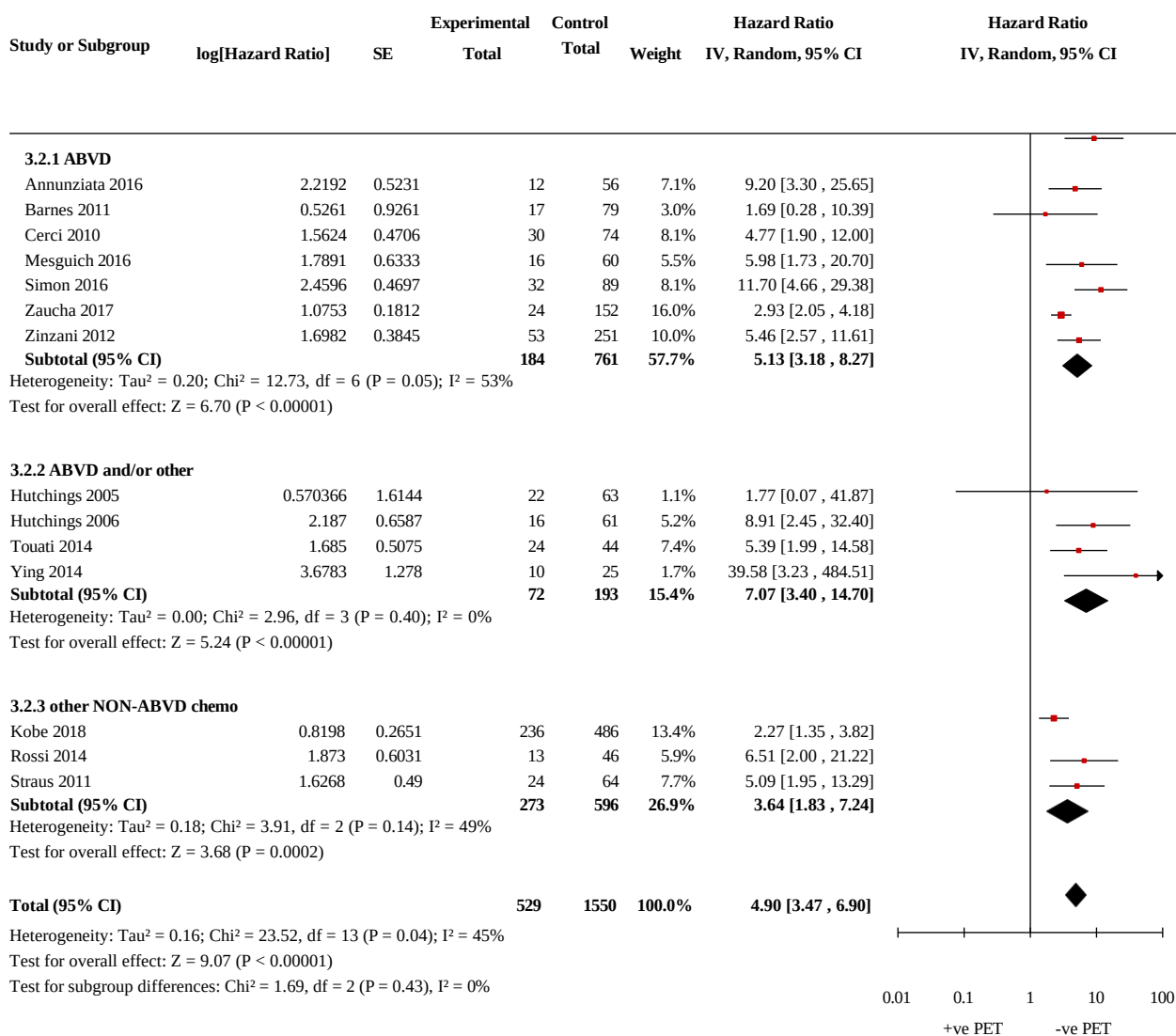
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
3.1 PFS by study design	13	1349	Hazard Ratio (IV, Random, 95% CI)	5.66 [4.02, 7.97]
3.1.1 prospective	3	357	Hazard Ratio (IV, Random, 95% CI)	3.95 [2.23, 7.00]
3.1.2 retrospective	8	827	Hazard Ratio (IV, Random, 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 participants	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]

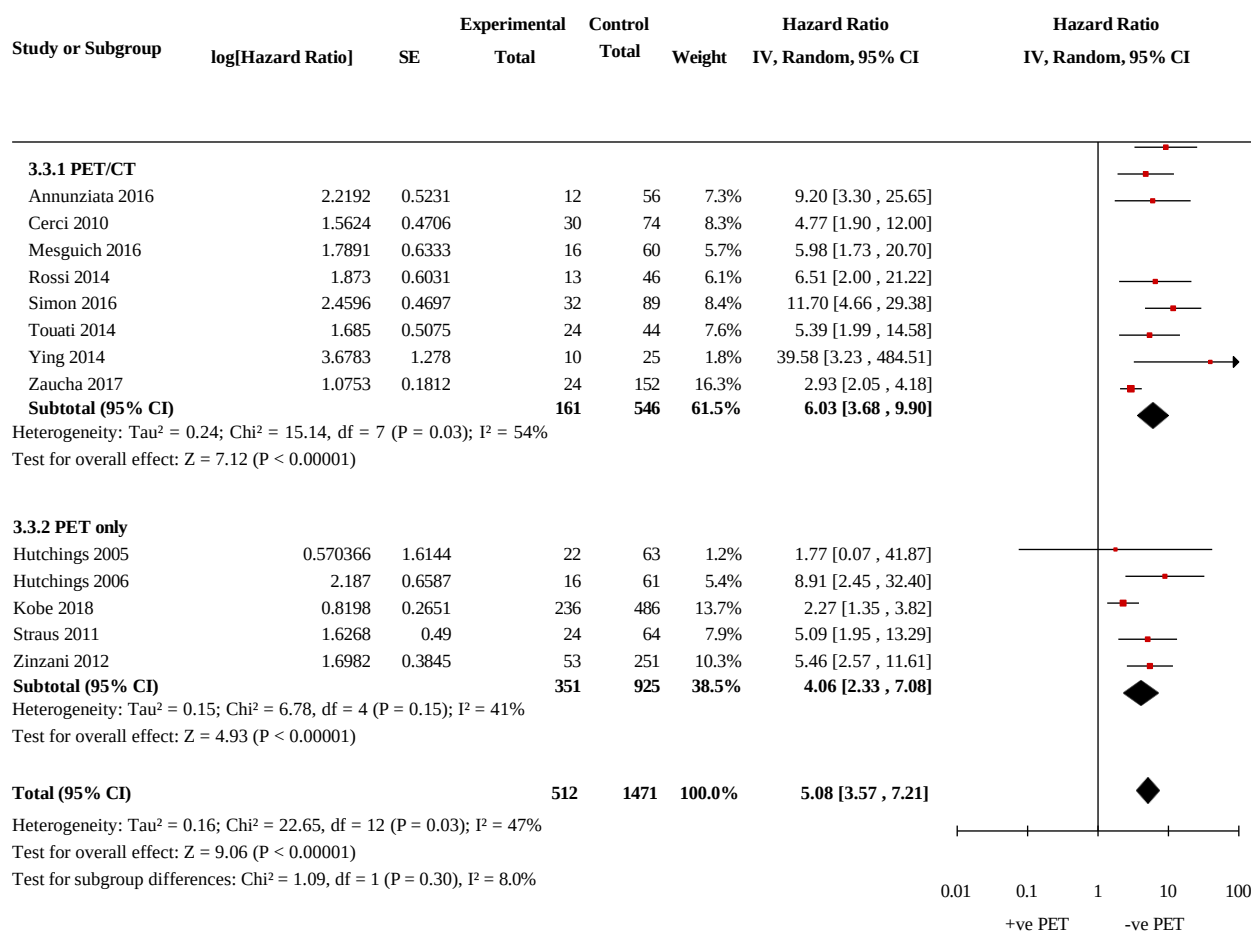
Analysis 3.1. Comparison 3: Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve, Outcome 1: PFS by study design



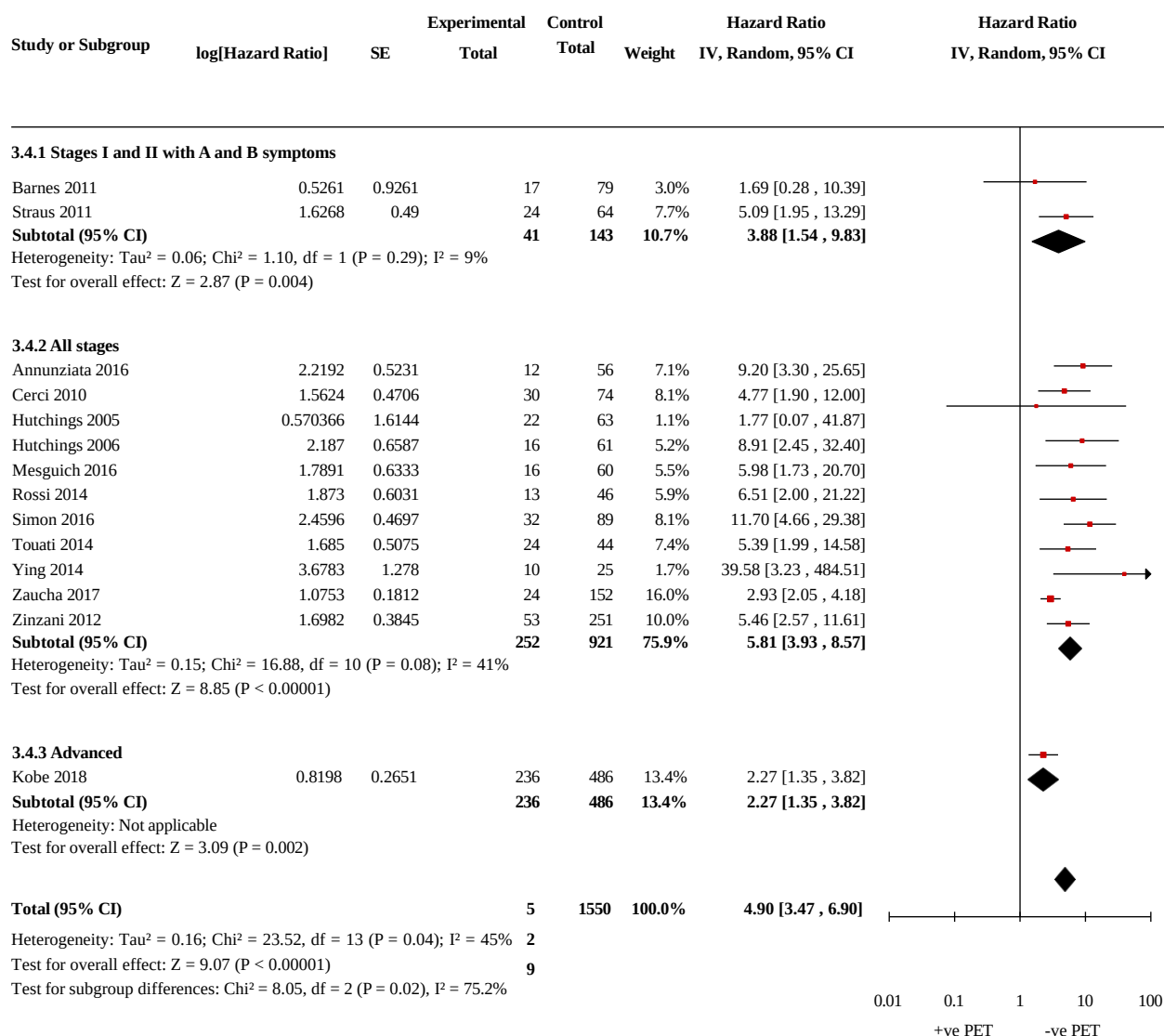
**Analysis 3.2. Comparison 3: Subgroups in univariable comparison of
PFS: PET+ve vs. PET-ve, Outcome 2: PFS by chemotherapy**



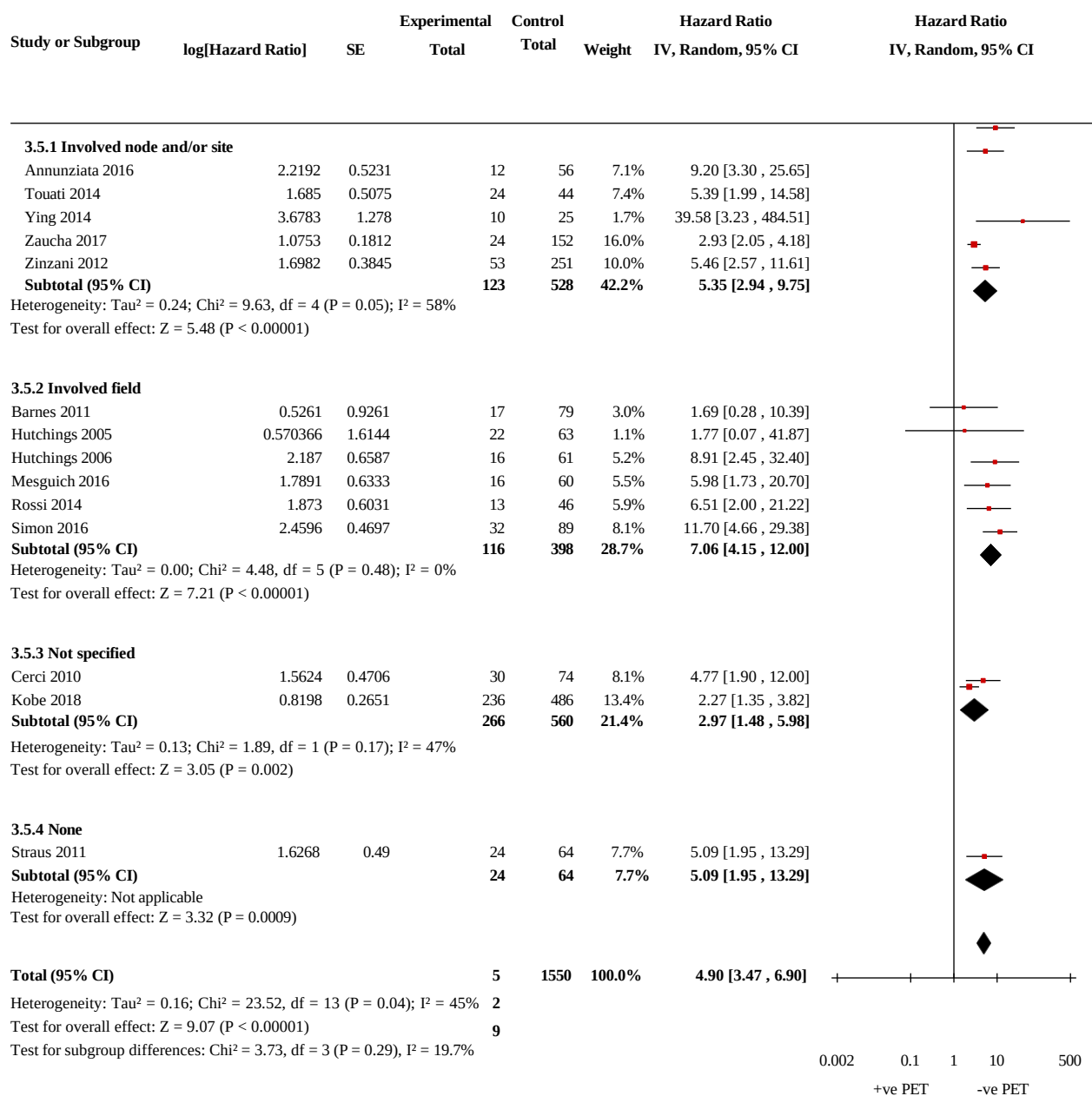
**Analysis 3.3. Comparison 3: Subgroups in univariable comparison of
PFS: PET+ve vs. PET-ve, Outcome 3: PFS for PET/CT vs PET**



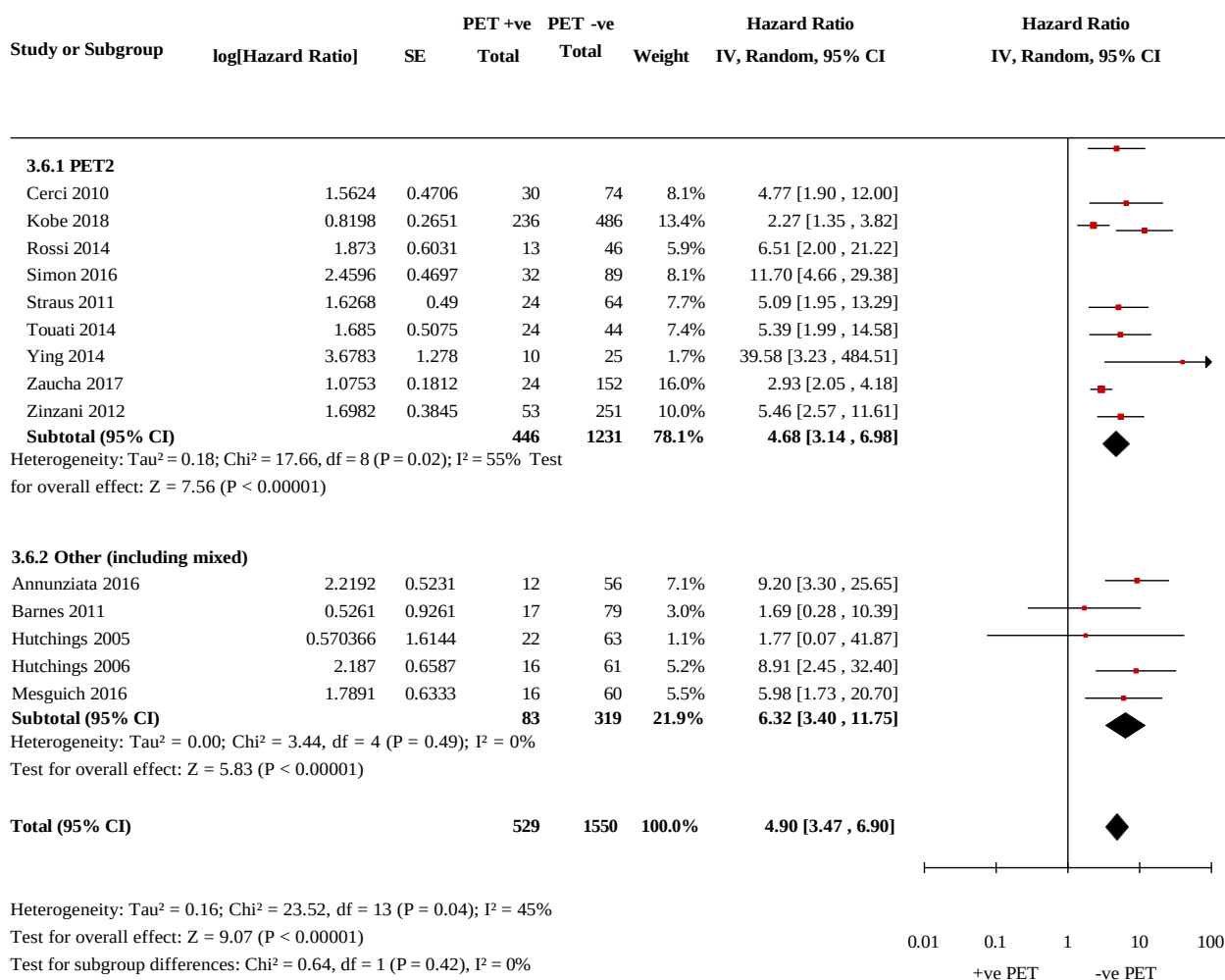
Analysis 3.4. Comparison 3: Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve, Outcome 4: PFS by disease stage



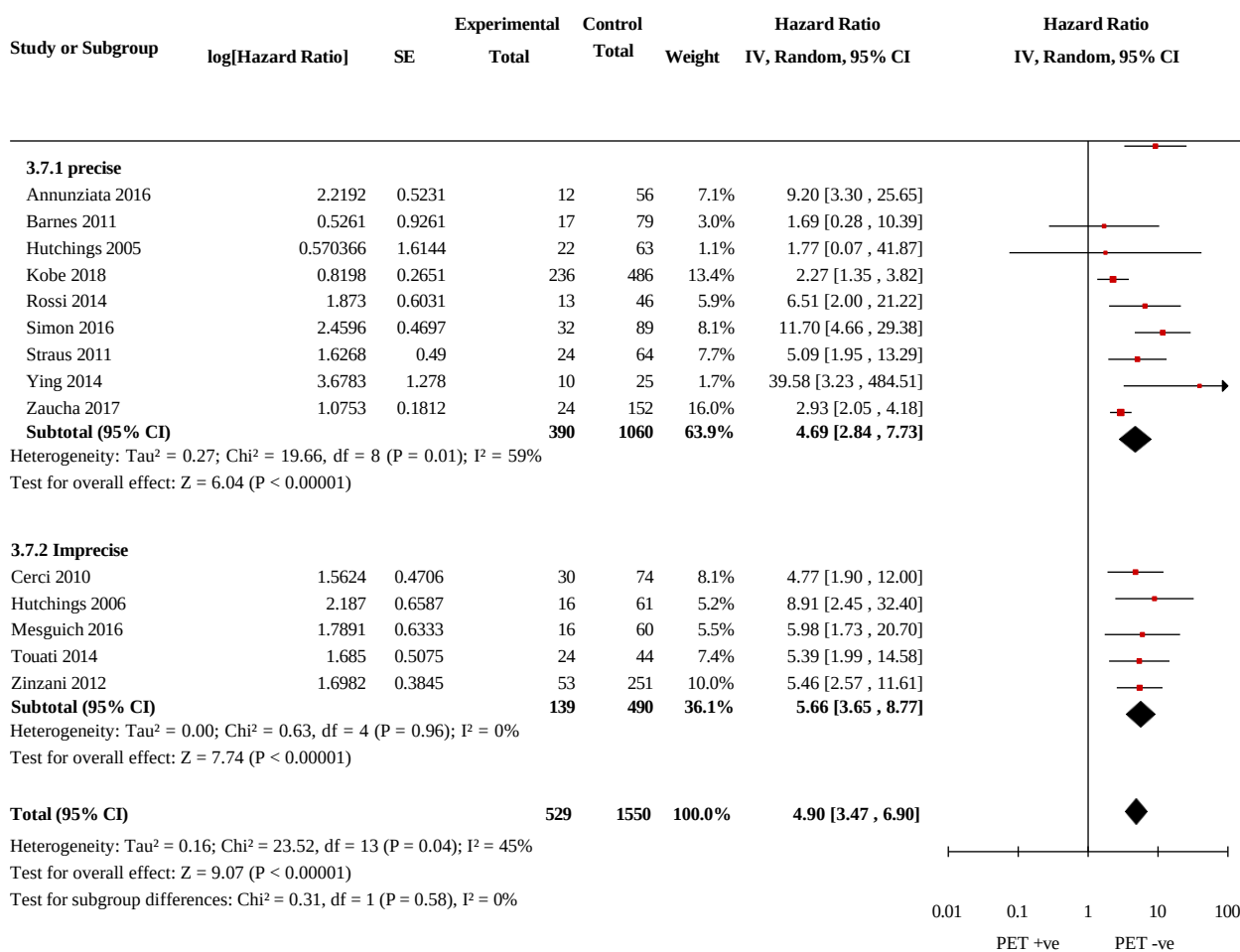
Analysis 3.5. Comparison 3: Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve, Outcome 5: PFS by radiotherapy



**Analysis 3.6. Comparison 3: Subgroups in univariable comparison of
PFS: PET+ve vs. PET-ve, Outcome 6: Timing of interim PET**



Analysis 3.7. Comparison 3: Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve, Outcome 7: PFS by HR type of estimation



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 18fluorodeoxyglucose* 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 18fluorodeoxyglucose* 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 18fluorodeoxyglucose* 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 18fluorodeoxyglucose* 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 fluoro-desoxyglucose\$ or fludeoxyglucose\$ or fluordeoxyglucose\$ or fluordesoxyglucose\$ or 18fluoro-desoxyglucose\$ 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 fluo- rodesoxyglucose\$ or fludeoxyglucose\$ or fluorodeoxyglucose\$ or fluordesoxyglucose\$ or 18flu- rodesoxyglucose\$ or 18fluordesoxyglucose\$ 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 fluo- rodesoxyglucose\$ or fludeoxyglucose\$ or fluordeoxyglucose\$ or fluordesoxyglucose\$ or 18fluo- rodesoxyglucose\$ or 18fluordesoxyglucose\$ 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 fluo- rodesoxyglucose\$ or fludeoxyglucose\$ or fluordeoxyglucose\$ or fluordesoxyglucose\$ or 18fluo- rodesoxyglucose\$ or 18fluordesoxyglucose\$ 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 fluo- rodesoxyglucose* or fludeoxyglucose* or fluordeoxyglucose* or fluordesoxyglucose* or 18flu- orodeoxyglucose* 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 draNing 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 draN.

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.

Marielena 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.

Marielena 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: Publication 2



Cochrane Database of Systematic Reviews

Prognostic models for newly-diagnosed chronic lymphocytic leukaemia in adults: a systematic review and meta-analysis (Review)

Kreuzberger N, Damen JAAG, Trivella M, Estcourt LJ, Aldin A, Umlau L, Vazquez-Montes MDLA, Wolff R, Moons KGM, Monsef I, Foroutan F, Kreuzer KA, Skoetz N

Kreuzberger N, Damen JAAG, Trivella M, Estcourt LJ, Aldin A, Umlau L, Vazquez-Montes MDLA, Wolff R, Moons KG, Monsef I, Foroutan F, Kreuzer K-A, Skoetz N.

Prognostic models for newly-diagnosed chronic lymphocytic leukaemia in adults: a systematic review and meta-analysis.

Cochrane Database of Systematic Reviews 2020, Issue 7. Art. No.:

CD012022. DOI: [10.1002/14651858.CD012022.pub2](https://doi.org/10.1002/14651858.CD012022.pub2).

www.cochranelibrary.com

TABLE OF CONTENTS

ABSTRACT	1
PLAIN LANGUAGE SUMMARY	2
SUMMARY OF FINDINGS	4
Figure 19.	5
Figure 20.	7
Figure 21.	9
BACKGROUND	10
OBJECTIVES	11
METHODS	11
RESULTS	17
Figure 1.	17
Figure 2.	19
Figure 3.	20
Figure 4.	21
Figure 5.	22
Figure 6.	22
Figure 7.	23
Figure 8.	23
Figure 9.	23
Figure 10.	24
Figure 11.	24
Figure 12.	24
Figure 13.	25
Figure 14.	26
Figure 15.	27
Figure 16.	28
Figure 17.	28
Figure 18.	29
Figure 22.	32
Figure 23.	33
Figure 24.	33
Figure 25.	34
Figure 26.	34
Figure 27.	36
Figure 28.	38
Figure 29.	42
DISCUSSION	42
AUTHORS' CONCLUSIONS	45
ACKNOWLEDGEMENTS	46
REFERENCES	47
CHARACTERISTICS OF STUDIES	67
APPENDICES	216
HISTORY	229
CONTRIBUTIONS OF AUTHORS	229
DECLARATIONS OF INTEREST	229
SOURCES OF SUPPORT	230
DIFFERENCES BETWEEN PROTOCOL AND REVIEW	230
NOTES	231
INDEX TERMS	231

[Prognosis Review]

Prognostic models for newly-diagnosed chronic lymphocytic leukaemia in adults: a systematic review and meta-analysis

Nina Kreuzberger¹, Johanna AAG Damen², Marialena Trivella³, Lise J Estcourt⁴, Angela Aldin¹, Lisa Umlauff¹, Maria DLA Vazquez-Montes⁵, Robert Wolff⁶, Karel GM Moons⁷, Ina Monsef¹, Farid Foroutan⁸, Karl-Anton Kreuzer⁹, Nicole Skoetz¹⁰

¹Cochrane 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. ²Cochrane Netherlands, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands. ³Centre for Statistics in Medicine, University of Oxford, Oxford, UK. ⁴Haematology/Transfusion Medicine, NHS Blood and Transplant, Oxford, UK. ⁵Nu&iield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK. ⁶Kleijnen Systematic Reviews Ltd, York, UK. ⁷Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands. ⁸Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada. ⁹Center of Integrated Oncology Cologne-Bonn, Department I of Internal Medicine, University of Cologne, Faculty of Medicine and University Hospital Cologne, Cologne, Germany. ¹⁰Cochrane 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

Contact: Nicole Skoetz, nicole.skoetz@uk-koeln.de.

Editorial group: Cochrane Haematology Group.

Publication status and date: New, published in Issue 7, 2020.

Citation: Kreuzberger N, Damen JAAG, Trivella M, Estcourt LJ, Aldin A, Umlauff L, Vazquez-Montes MDLA, Wolff R, Moons KG, Monsef I, Foroutan F, Kreuzer K-A, Skoetz N. Prognostic models for newly-diagnosed chronic lymphocytic leukaemia in adults: a systematic review and meta-analysis. *Cochrane Database of Systematic Reviews* 2020, Issue 7. Art. No.: CD012022. DOI: [10.1002/14651858.CD012022.pub2](https://doi.org/10.1002/14651858.CD012022.pub2).

Copyright © 2020 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

ABSTRACT

Background

Chronic lymphocytic leukaemia (CLL) is the most common cancer of the lymphatic system in Western countries. Several clinical and biological factors for CLL have been identified. However, it remains unclear which of the available prognostic models combining those factors can be used in clinical practice to predict long-term outcome in people newly-diagnosed with CLL.

Objectives

To identify, describe and appraise all prognostic models developed to predict overall survival (OS), progression-free survival (PFS) or treatment-free survival (TFS) in newly-diagnosed (previously untreated) adults with CLL, and meta-analyse their predictive performances.

Search methods

We searched MEDLINE (from January 1950 to June 2019 via Ovid), Embase (from 1974 to June 2019) and registries of ongoing trials (to 5 March 2020) for development and validation studies of prognostic models for untreated adults with CLL. In addition, we screened the reference lists and citation indices of included studies.

Selection criteria

We included all prognostic models developed for CLL which predict OS, PFS, or TFS, provided they combined prognostic factors known before treatment initiation, and any studies that tested the performance of these models in individuals other than the ones included in model development (i.e. 'external model validation studies'). We included studies of adults with confirmed B-cell CLL who had not received treatment prior to the start of the study. We did not restrict the search based on study design.

Data collection and analysis

We developed a data extraction form to collect information based on the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS). Independent pairs of review authors screened references, extracted data and assessed risk of bias according to the Prediction model Risk Of Bias ASsessment Tool (PROBAST). For models that were externally validated at least three times, we aimed to perform a quantitative meta-analysis of their predictive performance, notably their calibration (proportion of people predicted to experience the outcome who do so) and discrimination (ability to differentiate between people with and without the event) using a random-effects model. When a model categorised individuals into risk categories, we pooled outcome frequencies per risk group (low, intermediate, high and very high). We did not apply GRADE as guidance is not yet available for reviews of prognostic models.

Main results

From 52 eligible studies, we identified 12 externally validated models: six were developed for OS, one for PFS and five for TFS. In general, reporting of the studies was poor, especially predictive performance measures for calibration and discrimination; but also basic information, such as eligibility criteria and the recruitment period of participants was often missing. We rated almost all studies at high or unclear risk of bias according to PROBAST. Overall, the applicability of the models and their validation studies was low or unclear; the most common reasons were inappropriate handling of missing data and serious reporting deficiencies concerning eligibility criteria, recruitment period, observation time and prediction performance measures.

We report the results for three models predicting OS, which had available data from more than three external validation studies:

CLL International Prognostic Index (CLL-IPI)

This score includes five prognostic factors: age, clinical stage, IgHV mutational status, B2-microglobulin and TP53 status. *Calibration:* for the low-, intermediate- and high-risk groups, the pooled five-year survival per risk group from validation studies corresponded to the frequencies observed in the model development study. In the very high-risk group, predicted survival from CLL-IPI was lower than observed from external validation studies. *Discrimination:* the pooled c-statistic of seven external validation studies (3307 participants, 917 events) was 0.72 (95% confidence interval (CI) 0.67 to 0.77). The 95% prediction interval (PI) of this model for the c-statistic, which describes the expected interval for the model's discriminative ability in a new external validation study, ranged from 0.59 to 0.83.

Barcelona-Brno score

Aimed at simplifying the CLL-IPI, this score includes three prognostic factors: IgHV mutational status, del(17p) and del(11q). *Calibration:* for the low- and intermediate-risk group, the pooled survival per risk group corresponded to the frequencies observed in the model development study, although the score seems to overestimate survival for the high-risk group. *Discrimination:* the pooled c-statistic of four external validation studies (1755 participants, 416 events) was 0.64 (95% CI 0.60 to 0.67); 95% PI 0.59 to 0.68.

MDACC 2007 index score

The authors presented two versions of this model including six prognostic factors to predict OS: age, B2-microglobulin, absolute lymphocyte count, gender, clinical stage and number of nodal groups. Only one validation study was available for the more comprehensive version of the model, a formula with a nomogram, while seven studies (5127 participants, 994 events) validated the simplified version of the model, the index score. *Calibration:* for the low- and intermediate-risk groups, the pooled survival per risk group corresponded to the frequencies observed in the model development study, although the score seems to overestimate survival for the high-risk group. *Discrimination:* the pooled c-statistic of the seven external validation studies for the index score was 0.65 (95% CI 0.60 to 0.70); 95% PI 0.51 to 0.77.

Authors' conclusions

Despite the large number of published studies of prognostic models for OS, PFS or TFS for newly-diagnosed, untreated adults with CLL, only a minority of these (N = 12) have been externally validated for their respective primary outcome. Three models have undergone sufficient external validation to enable meta-analysis of the model's ability to predict survival outcomes. Lack of reporting prevented us from summarising calibration as recommended. Of the three models, the CLL-IPI shows the best discrimination, despite overestimation. However, performance of the models may change for individuals with CLL who receive improved treatment options, as the models included in this review were tested mostly on retrospective cohorts receiving a traditional treatment regimen. In conclusion, this review shows a clear need to improve the conducting and reporting of both prognostic model development and external validation studies. For prognostic models to be used as tools in clinical practice, the development of the models (and their subsequent validation studies) should adapt to include the latest therapy options to accurately predict performance. Adaptations should be timely.

PLAIN LANGUAGE SUMMARY

**How well do tools predict what happens with adults with newly-diagnosed chronic lymphocytic leukaemia (CLL) over time?
What was the aim of this review?**

There are many types of blood cancers called leukaemia. Chronic lymphocytic leukaemia (CLL) is the most common type. Twenty-five per cent of people who have leukaemia have CLL. It is natural for people with newly-diagnosed CLL and their families to want to know what will happen with their health in the future. They may be wondering if or when they will need treatment, if or when their disease will get worse or how long people live with CLL.

Researchers identified several characteristics that are associated with these outcomes. From these characteristics, they have tried to design tools that help predict what may happen to groups of people with newly-diagnosed CLL.

The aim of this Cochrane Review is to evaluate and summarise those tools and studies that test the tools with other patient data.

What are the key messages from this review?

Reviewers found that there is no reliable way to predict what might happen over time to people who have (untreated) CLL. One reason is because the prediction tools have not been tested enough times with enough different people to know how well they really work.

Another reason is because researchers continue to develop more effective CLL treatment options that have better results, and the prediction tools have not kept up with advances in treatment.

What are the main results of the review?

We identified 52 tools that were designed to predict what may happen to people newly-diagnosed with CLL. To find the best tools, we had to select the studies carefully. To apply these tools in clinical practice:

- a tool has to be tested by different researchers to predict what may happen with individuals with CLL in different geographic locations using different groups of people (i.e. age, gender, stage) with CLL. In other words, we would not include a tool if it was only tested on the people who provided their data to create it;
- the results of the tool should be consistent to prove that it works;
- the tests of the tool have to provide enough information to show how well the tool works. For example, the tests have to include large groups of people and enough information about the type of CLL they have.

We found three tools that met these requirements: the CLL International Prognostic Index (CLL-IPI), the Barcelona-Brno score, and the MDACC 2007 index score.

The CLL-IPI did the best job at identifying people who would survive longer with CLL and people who would survive less long. However, we rated the quality of the CLL-IPI studies as low because they did not provide all the information necessary to know how accurate the tool was. The Barcelona-Brno score and the MDACC 2007 index score, tested on a smaller overall number of patients, showed lower discrimination between persons with a good as compared to a worse prognosis, and showed a similarly low quality of the studies.

Conclusion

More and better research is needed to develop and test the tools to help predict how CLL will behave for different groups of people over time. The tools must also adapt to accurately predict the performance of new treatments.

SUMMARY OF FINDINGS

Summary of findings 1. CLL International Prognostic Index

Prognostic models for chronic lymphocytic leukaemia in adult patients

Population: untreated individuals with CLL

Index model: CLL international prognostic index

Timing: moment of prediction at diagnosis of CLL; moment of outcome occurrence not prespecified (any moment after diagnosis was included)

Setting: inpatient and outpatient care

Outcomes	Measure	N° of participants	Summary measure	Pooled result (95% CI)				Comments
				risk group	n (deaths)	pooled result (95% CI)	95%-PI	
Overall survival (OS) ^a	Discrimination	7 studies	c-statistic of external validation studies			0.72 (95% CI 0.67 to 0.77)	0.59 to 0.83	GRADE ^c
		3307 patients						
		917 deaths						
	Calibration ^b	8 studies	survival per risk group	low	2497 (249)	92.5% (89.2% to 94.8%)	82.4% to 97.0%	Survival at 5 years
		4891 patients		intermediate	1428 (233)	85.0% (79.7% to 89.1%)	68.8% to 93.6%	GRADE ^c
		875 deaths		high	765 (280)	64.9% (56.4% to 72.6%)	41.9% to 82.6%	
				very high	201 (113)	40.4% (29.3% to 52.6%)	13.2% to 72.4%	

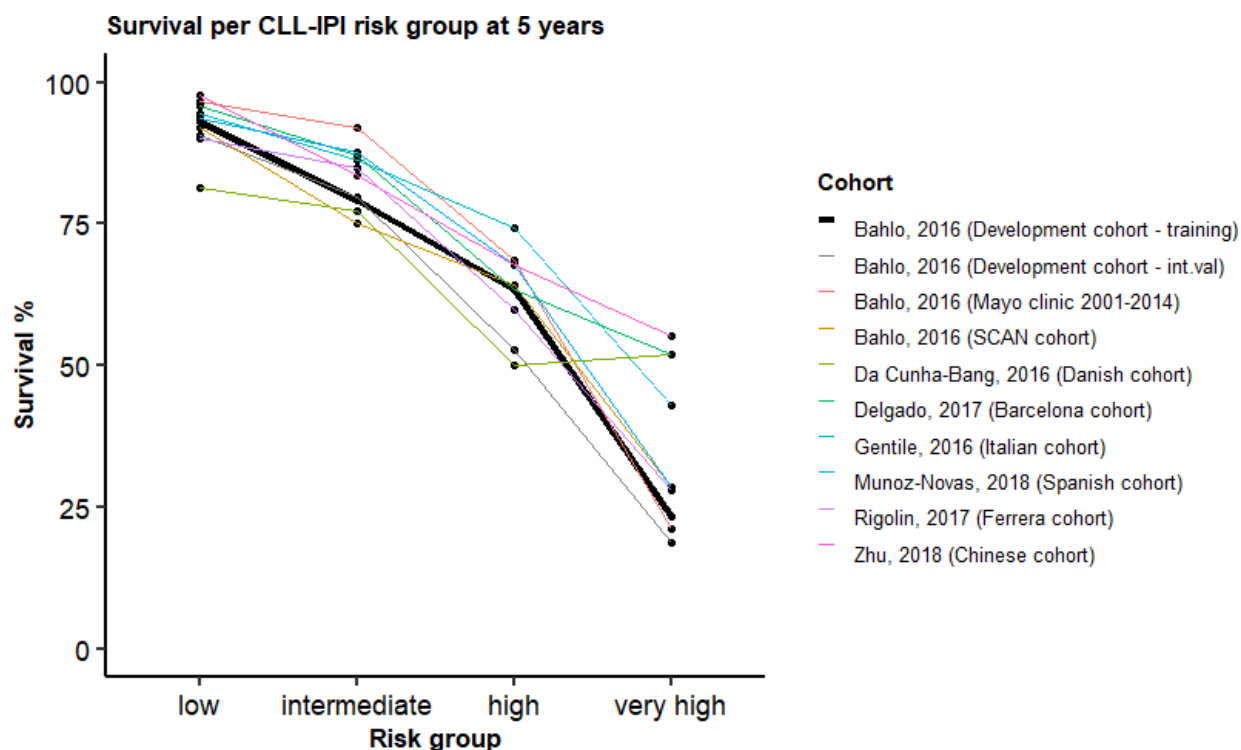
CI: confidence interval; **PI:** prediction interval

^aThe CLL-IPI was developed to predict overall survival. Although we identified external validation studies for other outcomes, we limited our analysis to the primary outcome of the development study.

^bNo calibration measures were reported, and so we used data on survival frequencies per group to compare expected versus observed survival – see [Figure 19](#).

^cGRADE was not conducted, as currently, no GRADE guidance for prognostic models exists.

Figure 19. Representation of survival per risk group per development and external validation study of the CLL-IPI (Bahlo 2016)



Summary of findings 2. Barcelona-Brno model

Prognostic models for chronic lymphocytic leukaemia in adult patients

Population: untreated individuals with CLL

Index model: Barcelona-Brno model

Timing: moment of prediction at diagnosis of CLL; moment of outcome occurrence not prespecified (any moment after diagnosis was included)

Setting: inpatient and outpatient care

Outcomes	Measure	N° of participants	Summary measure	Pooled result (95%)				Comments
				risk group	n (deaths)	pooled result (95% CI)	95% PI	



Overall survival (OS) ^a	Discrimination	4 studies	c-statistic of external validation studies		0.64 (95% CI 0.60 to 0.67)	0.59 to 0.68	GRADE ^c	
		1755 patients						
		416 deaths						
	Calibration ^b	3 studies	survival per risk group	low	1042 (88)	90.5% (85.1% to 94.0%)	80.4% to 95.7%	Survival at 5 years
		1974 patients						
		317 deaths		intermediate	673 (131)	79.7% (70.7% to 86.5%)	63.1% to 90.0%	GRADE ^c
				high	259 (98)	62.5% (49.3% to 74.1%)	41.3% to 79.7%	

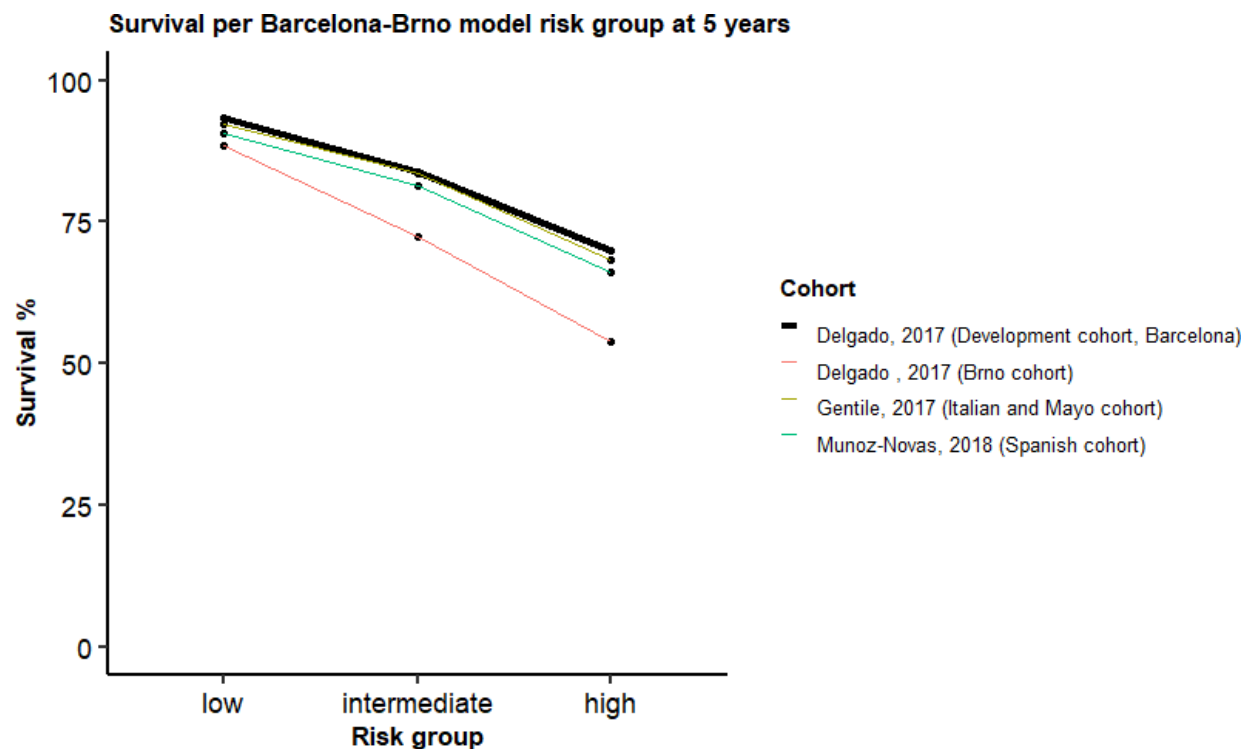
CI: confidence interval; **PI:** prediction interval

^aThe Barcelona-Brno model was developed to predict overall survival. Although we identified external validation studies for other outcomes, we limited our analysis to the primary outcome of the development study.

^bNo calibration measures were reported, and so we used data on survival frequencies per group to compare expected versus observed survival – see [Figure 20](#).

^cGRADE was not conducted, as currently, no GRADE guidance for prognostic models exists.

Figure 20. Representation of survival per risk group per development and external validation study of the Barcelona-Brno model (Delgado 2017)



Summary of findings 3. MDACC 2007 index score

Prognostic models for chronic lymphocytic leukaemia in adult patients

Population: untreated individuals with CLL

Index model: MDACC 2007 index score

Timing: moment of prediction at diagnosis of CLL; moment of outcome occurrence not prespecified (any moment after diagnosis was included)

Setting: inpatient and outpatient care

Outcomes	Measure	N° of participants	Summary measure	Pooled result (95%)			Comments
				Risk group	n (deaths)	pooled result (95% CI)	95% PI



Overall survival (OS) ^a	Discrimination	7 studies 5127 patients 994 deaths	c-statistic of external validation studies		0.65 (95% CI 0.60 to 0.70)	0.51 to 0.77	GRADE ^c
	Calibration ^b	5 studies 3786 patients 511 deaths	survival per risk group	low	1202 (35)	97.0% (94.3% to 98.4%)	90.9% to 99.0% Survival at 5 years
				intermediate	2425 (393)	82.3% (74.6% to 88.0%)	61.5% to 93.1% GRADE ^c
				high	159 (83)	45.6% (31.3% to 60.5%)	21.2% to 72.3%

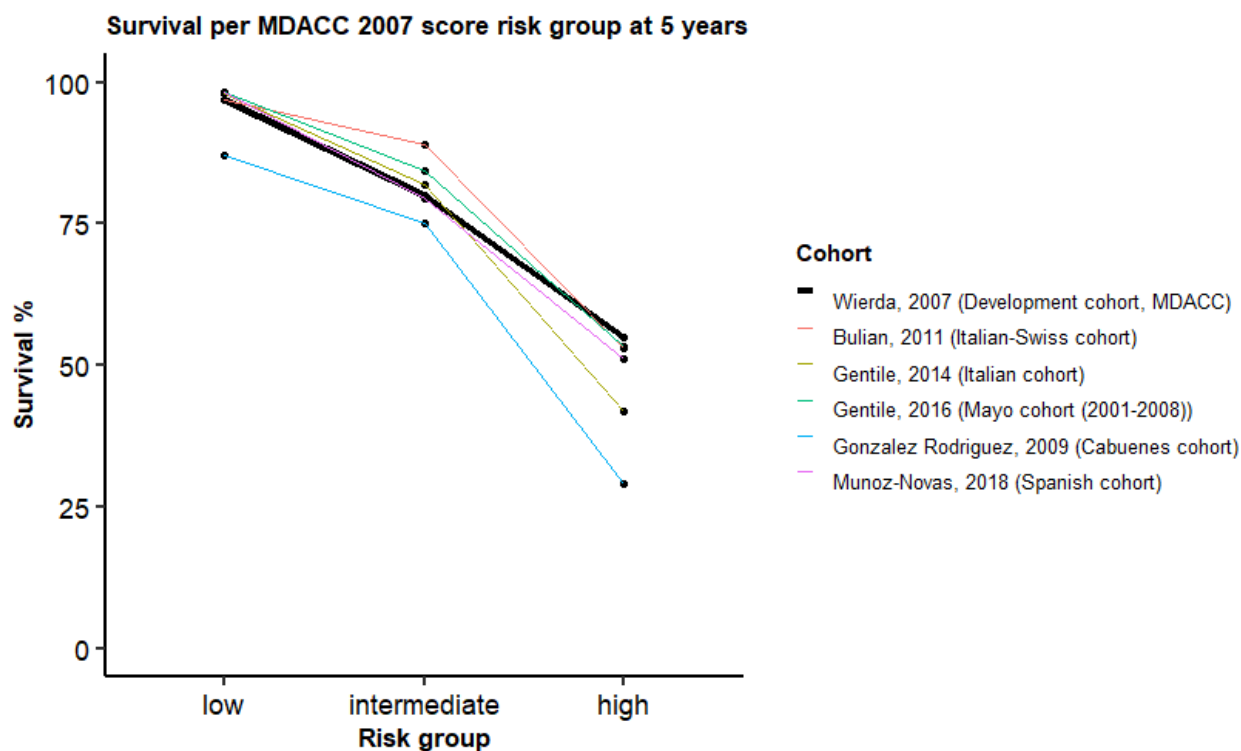
CI: confidence interval; **PI:** prediction interval

^aThe MDACC 2007 index score was developed to predict overall survival. Although we identified external validation studies for other outcomes, we limited our analysis to the primary outcome of the development study.

^bNo calibration measures were reported, and so we used data on survival frequencies per group to compare expected versus observed survival – see [Figure 21](#).

^cGRADE was not conducted, as currently, no GRADE guidance for prognostic models exists.

Figure 21. Representation of survival per risk group per development and external validation study of the MDACC 2007 model (Wierda 2007)



BACKGROUND

Description of the condition

Chronic lymphocytic leukaemia (CLL) is the most common form of malignant neoplasm (cancer) of the lymphatic system in Western countries. It is responsible for 25% of all leukaemias and occurs mainly in the elderly population ([Chiorazzi 2005](#)). The reported age-adjusted incidence rate of CLL in the USA between 2012 and 2016 was 4.9 per 100,000 persons with an estimated 20,720 new cases and an age-adjusted death rate of 1.2 per 100,000 persons per year ([Howlader 2019](#)). In the European Union, an estimated 46,000 individuals were living with CLL five years post-diagnosis in 2006 ([Watson 2008](#)).

In CLL, mature B cells accumulate in the blood, bone marrow, lymph nodes and spleen ([Herishanu 2013](#)). The diagnosis of CLL is generally established based on blood counts, differential counts, blood smears and immunophenotyping ([Hallek 2017](#)). The requirement for a diagnosis of CLL has been modified from a chronic absolute lymphocytosis with more than 5.0×10^9 cells per L to an absolute count of more than 5.0×10^9 monoclonal B cells with CLL immunophenotype per L peripheral blood for the duration of at least three months ([Hallek 2008](#); [Hallek 2017](#)). In the case of a monoclonal B-cell count lower than 5.0×10^9 per L and the absence of disease-related symptoms, cytopenias or tissue involvement other than bone marrow, the condition is defined as monoclonal B-lymphocytosis ([Hallek 2008](#)). A diagnosis of small lymphocytic lymphoma (SLL) is made when lymphadenopathy (enlarged lymph nodes) or splenomegaly (enlarged spleen) are caused by infiltrating CLL cells, and B lymphocytes in the peripheral blood do not exceed 5.0×10^9 per L ([Hallek 2008](#)).

CLL is characterised by a highly variable clinical course and prognosis. Some individuals with CLL experience no or only few symptoms over many years, do not require treatment, and have a life expectancy comparable to that of a healthy individual. Other individuals already experience symptoms at diagnosis or shortly thereafter, and die within a few years despite treatment. The heterogeneity in clinical presentation makes it difficult for the physician to predict accurately whether a patient may benefit from an early, aggressive treatment strategy, and to provide the patient with relevant prognostic information.

The most commonly used early attempts to group individuals with CLL according to their risk are the two staging systems by Binet 1981 ([Binet 1981](#)) and Rai 1975 ([Rai 1975](#)), which distinguish between early (Rai 0; Binet A), intermediate (Rai I, II; Binet B) and advanced stages (Rai III, IV; Binet C). The disease stage is determined by the number of lymphocytes in the peripheral blood, presence of enlarged lymph nodes, presence of anaemia or thrombocytopenia (low platelet count), and presence of an enlarged liver or spleen. The prognostic value of the two staging systems is limited as survival times vary significantly within these stages. According to the guidelines by the European Society for Medical Oncology, B-symptoms at diagnosis (fever, night sweats or weight loss) categorise people with early stage disease according to Binet or Rai, and are considered to have aggressive disease, requiring treatment ([Eichhorst 2015](#)).

Prognostic models

A prognostic model is a mathematical function that considers at least two prognostic factors, also called predictors, simultaneously with the aim to provide an estimate of an individual patient's probability to experience a certain health event within a defined time frame in the future ([Alba 2017](#); [Moons 2009](#); [Riley 2019](#); [Steyerberg 2013](#)). Prognostic factors may be characteristics of the individual or the disease (e.g. age, gender, disease stage, biological or genetic information), which are likely to predict a patient-relevant outcome, such as overall survival or disease progression ([Riley 2013](#)). By statistical modelling methods, these factors are often combined to form a weighted model able to accurately predict the likelihood of this outcome. Such a model aims at assisting the clinician to estimate the patient's prognosis and enhance shared decision-making. Establishing a prognosis for the individual patient may also lead to risk-stratified treatment recommendations ([Alba 2017](#); [Debray 2017](#); [Moons 2009](#); [Riley 2019](#); [Steyerberg 2013](#)).

To develop a prognostic model for a specific outcome, such as death or disease progression, various data sources such as cohort, nested case-control or case-cohort studies are considered appropriate, especially when data are prospectively collected. Data originating from randomised clinical trials, as a special form of prospectively collected data, can be used, but may limit generalisability due to more restrictive eligibility criteria, selective participation of specialised centres, trial effects and unrealistically precise predictor assessments ([Collins 2015](#); [Moons 2014](#); [Moons 2019](#); [Pajouheshnia 2019](#)). Before a prognostic model is used in clinical practice, its predictive performance should be quantified. Apparent performance of the model is the model performance estimated from the same data as used for model development and usually provides an overly optimistic estimate due to overfitting of the model to this specific dataset. Internal validation tests the model performance in the development dataset by using techniques such as bootstrapping or cross-validation. In order to test the generalisability of a prognostic model, predictive performance should be ideally assessed in several independent sets of data of individuals that were not used in the development and internal validation of the model, preferably by independent investigators to reduce bias. This process is called external validation ([Moons 2015](#)).

In the past few years, a number of clinical and biological CLL prognostic factors have been identified, including genomic aberrations, gene abnormalities (p53, ATM), mutation status of the variable segments of the immunoglobulin heavy chain genes (IgHV), or surrogate markers for these factors, such as CD38 and ZAP-70 expression ([Döhner 2000](#); [El Roubi 1993](#); [Kay 2007](#)). Assessing recent molecular markers at time of diagnosis is expected to provide more reliable information regarding optimal time for treatment initiation, type of therapy and individual prognosis ([Kay 2007](#); [Shanafelt 2004](#); [Wierda 2011](#); [Zenz 2011](#)). Thus, progressive and smouldering forms of the disease can now be separated more accurately than by using Rai or Binet staging systems alone. Moreover, the early recognition of aggressive stage A and indolent stage B and C disease would allow rational application of risk-adapted treatment strategies. Factors influencing the choice of treatment include age, fitness to tolerate chemotherapy or immunotherapy or both, TP53 status, previous or current immune

cytopenias, and evidence of lymphomatous transformation (Goede 2012).

The recent increase in availability of biological markers for CLL presents a challenge, as well as an opportunity to develop more precise prognostic models or algorithms that integrate a combination of markers and may guide counselling and treatment decisions for the individual patient. In order to identify the best-performing tool to estimate prognosis for untreated individuals with CLL, a comprehensive evaluation of all currently available prognostic models and meta-analysis of their predictive performance in external validation studies is urgently needed.

Health outcomes

The highly variable course of CLL and the possibility of having a normal life expectancy without progression or need for treatment entails that overall survival (OS) is one of the most important outcomes to be predicted by a prognostic model. In the USA, the median age at diagnosis is 70 years and the five-year survival rate with CLL is 85.1% (Howlader 2019). Therefore, it is important to observe patients as long as possible to obtain a prognostic model that is meaningful not only for high-risk individuals, but also for people with a less aggressive disease and longer survival.

As individuals with CLL are usually older, which implies an increased prevalence of comorbidities and decreased physical fitness, treatment may lead to serious adverse events and interactions with other medications. Hence, alternative meaningful outcomes to be predicted by a prognostic model include progression-free survival (PFS) or treatment-free survival (TFS,

also sometimes referred to as time-to-first-treatment). Treatment options for CLL have improved over time, thus, affecting survival rates but not the rates of treatment indication.

Why it is important to do this review

Although several prognostic factors have been identified during the last decade (Pflug 2014; Stilgenbauer 2014; Zaja 2013), they are controversial and there is no single prognostic factor available to determine treatment options in CLL patients. These factors have also been combined into numerous prognostic models. To date, no systematic review has been conducted to evaluate and assess the predictive performance of prognostic models in CLL, which would inform us which models have the greatest validity and therefore, would be preferred to guide clinical decision-making. To shed light on this important research question, we conducted a systematic review and, where possible, a meta-analysis of existing prognostic models for CLL and their corresponding validation studies.

OBJECTIVES

To identify, describe and appraise all prognostic models developed to predict overall survival (OS), progression-free survival (PFS) or treatment-free survival (TFS) in newly-diagnosed (previously untreated) adults with CLL, and meta-analyse their predictive performances.

METHODS

Criteria for considering studies for this review

Table 1. PICOTS system for prognostic models

P	Population	Untreated individuals with chronic lymphocytic leukaemia (CLL) at time of prediction
I	Index model(s)	All developed prognostic models for CLL and their corresponding external validation studies
C	Comparator	No predefined comparator
O	Outcome(s)	Overall survival (OS), progression-free survival (PFS), treatment-free survival (TFS)
T	Timing	Moment of prediction at diagnosis of CLL; moment of outcome occurrence not prespecified (any moment after diagnosis was included)
S	Setting	Not specified

Types of studies

According to the checklist for critical appraisal and data extraction for systematic reviews of prediction modelling studies (CHARMS) (Moons 2014), we included:

- prognostic model development studies without external validation in independent data;
- prognostic model development studies with external validation in independent data reported in the same study; and

- external model validation studies of previously reported models.

A prognostic model was defined as some form of mathematical function including at least two independent factors to predict OS, PFS or TFS. We included only models that combined prognostic factors known and assessed before treatment initiation. We excluded prognostic factor finding studies (i.e. studies aiming to establish one or several variables as independent prognostic factor(s) associated with an outcome, but not aiming to develop

a model to be used for individualised predictions (i.e. to provide absolute prognostic outcome probabilities) in new patients, and model impact studies (i.e. studies aiming to investigate the impact of the use of a model in practice) (Bouwmeester 2012). We excluded studies that were published only as conference proceedings. We did not exclude studies with treatment during follow-up time.

Participants

We included studies on individuals with previously untreated, confirmed B-cell CLL. We included both male and female adults (age ≥ 18 years).

Types of outcomes to be predicted

Primary outcome

- Models that predict OS as an outcome

We chose this outcome because it has the greatest clinical relevance and is most important for patients. Furthermore, death due to any cause is an objective endpoint not susceptible to bias of the outcome assessor. We did not require studies to have a minimum follow-up time for inclusion in this review.

Secondary outcomes

- Models that predict either PFS or TFS

We chose these outcomes as patients with similar survival may nevertheless have differing lengths of time without symptoms or need for treatment, depending both on initial treatment and disease characteristics. In case of immediate start of treatment, identification of patients with a lower probability to obtain a good response will help in making decisions regarding treatment, for example, deciding which patients might receive new or more aggressive therapy regimens. In case of a watch-and-wait strategy, differences in estimated prognosis can influence patient management regarding surveillance and treatment.

Search methods for identification of studies

Electronic searches

Reporting and therefore retrieval of prognostic model studies is very poor, as guidelines on reporting of prediction models have only recently been published (Collins 2015). For our first search, we did not use a specific search filter (Appendix 1). As this search strategy was not very specific, it yielded many results, which had to be screened in detail by two review authors. For the updated search, we integrated the search filter by Geersing 2012 to allow for a more specific strategy (Appendix 2).

We searched the following databases without applying any language restrictions in order to reduce language bias.

- MEDLINE via Ovid (searched 24 June 2019; Appendix 1; Appendix 2).
- Embase (searched 24 June 2019; Appendix 3).

We searched the following databases for ongoing trials.

- ClinicalTrials.gov (clinicaltrials.gov; searched 5 March 2020; Appendix 4).
- World Health Organization International Clinical Trials Registry Platform (WHO ICTRP) (searched 5 March 2020; Appendix 5).

Initially, we planned to include only models published after 1990. However, prognostic model studies are often conducted based on retrospective data, which in some cases included the analysis of blood samples frozen at diagnosis and analysed many years later. Thus, we decided not to limit our search strategy to date of publication. We did not screen conference proceedings, according to the section 'Types of prognostic models', where we stated that we would exclude conference abstracts based on limited information that would not allow us to assess risk of bias. In contrast to our planning (Skoetz 2016), we did not search the database of prognostic studies by the Prognostic Methods Group as it has not been developed. We did not search conference proceedings because of their limited information that would have complicated the inclusion and exclusion of studies as well as the risk of bias rating.

Searching other resources

As prespecified in the protocol (Skoetz 2016), we searched the following sources.

- Handsearching of references
 - References of all identified trials, relevant review articles and current treatment guidelines for further literature.
- Personal contacts
 - We contacted authors of relevant studies, study groups, experts and investigators from transplantation centres worldwide who are known to be active in the field for unpublished material or further information on ongoing studies.

In addition, and not reported in the protocol of this review, we tracked the citations of all included studies in October 2018 to identify possible additional validation studies of a developed model (Web of Science citation indexing).

Data collection and analysis

Selection of studies

During independent screening of titles and abstracts by two review authors (AA, LE, MT, NK, NS), we organised regular group discussions for congruency due to the novelty of the topic. The main discussion points can be found below. After title and abstract screening, full texts for all eligible studies were obtained and independently selected by two review authors (AA, LE, MT, NK, NS). Disagreements were solved by involving one or more additional review authors (Lefebvre 2019).

As recommended in the PRISMA statement (Moher 2009), we documented the total number of retrieved references and numbers of included and excluded studies in a flow chart.

Solving disagreements in study inclusion

The poor reporting and the fine line between the development of a prognostic model and prognostic factor (identification) studies made the screening of titles and abstracts and the final decision regarding the inclusion of studies challenging. During several group discussions, the following issues emerged repeatedly.

- As we changed our pre-planned limitation of the search to include studies from inception of the database, we obtained many studies labelled as "staging system." A staging system intends to describe the severity of a disease and, thus, indicate

some prognostic and therapeutic information. Two staging systems for CLL, Rai and Binet, are commonly used in clinical practice. However, the predicted survival of individuals with the same disease stage can be highly variable. We decided to exclude staging systems, as they are not used by clinicians any longer (except for Rai and Binet). Instead, we focused on new algorithms with a high chance to include more recently identified prognostic factors.

- We decided to exclude studies that build genetic signatures aiming to distinguish between good and poor prognosis of an individual, due to several reasons. First, the development of a genetic signature is often preceded by the identification of genetic markers on the same cohort from several hundred candidate markers, which introduces a high potential for overfitting of the algorithm to the development cohort (e.g. [Houldsworth 2014](#)). Second, genetic signatures can consist of a tremendous number of genes compared to the number of individuals included in the cohort (e.g. [Ferreira 2014](#)). Third, it is currently difficult to apply the algorithm of a genetic signature to an external cohort due to the complexity of the model.
- The fine line between a multivariable model, which includes several prognostic factors to prove their independence from each other, and a prognostic model was not always evident from the abstract. When an abstract mentioned a multivariable model and the formation of risk groups, we considered the full text of the paper. We then excluded papers that did not proceed to build a prognostic model or score explicitly.
- References were classified as an external validation study if the term "validation" was explicitly stated or the application of the prognostic model was clearly the main focus of the paper. We also included some publications where validation was not explicitly mentioned and performance measures were not reported, but where the authors put their focus on the application of one of the previously developed prognostic models (e.g. [CLL-IPI V - Reda 2017 \(Milano cohort\)](#); [CLL-IPI V - Rigolin 2017 \(Ferrera cohort\)](#)).

The decisions detailed above may seem subjective and were made in the case of this specific review. We would advise future authors who aim to evaluate prognostic models for a certain disease to clearly define the eligibility of different types of prognostic models at the protocol stage. This will help with objectivity throughout the screening process.

Data extraction and data management

Teams of two review authors (AA, LE, MT, NK, NS) independently extracted the data to enable assessment of applicability of the model for the review and the 'Risk of bias' assessment (see below). We contacted authors of individual studies for additional information, where required. We developed a standardised data extraction form containing the following items based on the CHARMS checklist ([Moons 2014](#)).

- General information
 - Author, title, publication date, country, language, duplicate publications
- Source of data
 - E.g. cohort, case-control, randomised trial 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
- Details of treatments received
- Study dates
- Outcomes to be predicted
 - Definition and method for measurement of outcome
 - Was the same outcome definition (and method for measurement) used in all patients?
 - Was the outcome assessed without knowledge of the candidate predictors (i.e. blinded)?
 - Time of outcome occurrence or summary of duration of follow-up
- Candidate predictors
 - Number and type of predictors (e.g. demographics, patient history, physical examination, additional testing, disease characteristics, tumour markers)
 - Definition and method for measurement of candidate predictors
 - Timing of predictor measurement (e.g. at patient presentation, at diagnosis, at treatment initiation)
 - Were predictors assessed blinded for outcome, and for each other (if relevant)?
 - Handling of predictors in the modelling (e.g. continuous, linear, non-linear transformations or categorised)
- Sample size
 - Number of participants and number of outcomes/events
 - Number of outcomes/events in relation to the number of candidate predictors (events per variable)
- Missing data
 - Number of participants with any missing value (include predictors and outcomes)
 - Number of participants with missing data for each predictor
 - Handling of missing data (e.g. complete-case analysis, imputation, or other methods)
- Model development
 - Modelling method (e.g. logistic, survival, neural networks, or machine learning techniques)
 - Modelling assumptions satisfied
 - Method for selection of predictors for inclusion in multivariable modelling (e.g. all candidate predictors, pre-selection based on unadjusted association with the outcome)
 - Method for selection of predictors during multivariable modelling (e.g. full model approach, backward or forward selection) and criteria used (e.g. P value, Akaike Information Criterion)
 - Shrinkage of predictor weights or regression coefficients (e.g. no shrinkage, uniform shrinkage, penalised estimation)
- Model performance
 - Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) and discrimination (C-statistic, D-statistic, log-rank) measures with confidence intervals
 - Classification measures (e.g. sensitivity, specificity, predictive values, net reclassification improvement) and whether a priori cut-points were used
- Model evaluation

- o Method used for testing model performance: development dataset only (random split of data, re-sampling methods, e.g. bootstrap or cross-validation, none) or separate external validation (e.g. temporal, geographical, different setting, different investigators)
- o In case of poor validation, whether model was adjusted or updated (e.g. intercept re-calibrated, predictor effects adjusted, or new predictors added)
- Results
 - o Final and other multivariable models (e.g. basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard errors or confidence intervals)
 - o Any alternative presentation of the final prediction models, e.g. sum score, nomogram, score chart, predictions for specific risk subgroups with performance
 - o Comparison of the distribution of predictors (including missing data) for development and validation datasets
- Interpretation and discussion
 - o Interpretation of presented models (confirmatory, i.e. model useful for practice versus exploratory, i.e. more research needed)
 - o Comparison with other studies, discussion of generalisability, strengths and limitations

References of model development studies, for which no external validation studies have yet been published, are summarised in [Studies awaiting classification](#). These studies fit the formal inclusion criteria of this review. However, they have not yet been tested in any additional cohort, which renders summary of performance in external cohorts impossible. They cannot yet be evaluated regarding their clinical usefulness, and are therefore not yet described in detail.

Risk of bias and applicability assessment

Since the publication of the review protocol, a 'Risk of bias' tool specifically designed for prognostic model studies (Prediction model Risk Of Bias ASsessment Tool, PROBAST) was published ([Moons 2019](#); [Wolff 2019](#)). Instead of the CHARMS checklist for critical appraisal of prognostic modelling studies as reported in our protocol, we used this new tool recommended by the Cochrane Prognosis Methods Group to assess the risk of bias of the individual prediction models investigated in the included primary studies. Teams of two review authors (AA, LE, MT, NK, NS) independently assessed the risk of bias and applicability for each study. We assessed only validation studies with the outcome that a model was developed for, in line with meta-analyses, hence PROBAST assessments reflect only the outcome of interest and not any additional outcomes reported in the same study.

After classifying each study into one of the three categories (model development with or without external validation in the same publication and external validation study of a previously developed model only), we assessed risk of bias according to the following four PROBAST domains (once per developed or validated model and per outcome).

- Participants
- Predictors
- Outcome

Analysis

We answered signalling questions within each domain with one out of five options ('yes', 'probably yes', 'probably no', 'no', 'no information'); 'yes' always implied the absence of a potential bias-generating aspect. We rated domain-level 'Risk of bias' assessments using one of the following three options.

- Low risk of bias: if the criterion is adequately fulfilled in the study, i.e. the study is at a low risk of bias for the given domain.
- High risk of bias: if the criterion is not fulfilled in the study, i.e. the study is at high risk of bias for the given domain.
- Unclear risk of bias: if the study report does not provide sufficient information to allow for a clear judgement or if the risk of bias is unknown for one of the domains listed above.

PROBAST additionally requires the judgement of the applicability of the model to the research question. The assessment occurs per domain (for the first three domains only) with the following response options: 'low concern regarding applicability', 'high concern regarding applicability' and 'unclear concern regarding applicability' (equivalent to the categories for risk of bias).

Based on these domain-level judgements, we established an overall judgement per study based on PROBAST guidance for risk of bias:

- Low risk of bias: if a prediction model evaluation is judged as low on all domains relating to bias or applicability.
- High risk of bias: if an evaluation is judged as high for at least one domain.
- Unclear risk of bias: if the prediction model evaluation is unclear in one or more domains and was rated as low in the remaining domains.

When information for a complete judgement was missing, we contacted the corresponding authors via email to request additional information to be able to make sound judgements on the 'Risk of bias' assessment. According to PROBAST, models that were developed, but not (yet) externally validated can be classified as 'high risk of bias', with the exception of extremely large samples ([Wolff 2019](#)). Due to the novelty of PROBAST and the subjectivity of ratings, we devoted a small section of the results to present our experience and agreements. Two authors of PROBAST (RW, KGMM) are also authors in our review and RW was involved in group discussions addressing disagreements between ratings.

Solving disagreements regarding PROBAST ratings

Due to the novelty of the 'Risk of bias' tool (PROBAST), we arranged frequent team meetings to discuss disagreements and discrepancies regarding ratings between the six authors who extracted data and assessed risk of bias and applicability. For reasons of transparency, we report the following challenges and our agreements.

- The relevant observation time varies between diseases. For an indolent neoplasm like CLL, the clinicians in our team considered a median follow-up of five years as appropriate for OS in a cohort with a normal case mix (i.e. no limitation to patients with late stage or high-risk disease such as patients with del(17p) only).

- The signalling question for the participant domain asks whether all inclusions and exclusions of participants were appropriate (item 1.2). Some publications listed missing values as part of the exclusion criteria. Missing values are also treated in the analysis domain (items 4.3 and 4.4). To avoid duplicate rating of the same issue, we decided to rate the risk of bias as high for the participant domain when complete data were required for eligibility, and rated the risk of bias as high for the analysis domain when the number and characteristics of the background sample were described but persons were excluded from analysis due to missing values.
- Some publications did not report eligibility criteria and/or recruitment period. We rated these studies as unclear not only for risk of bias, but also for concerns of applicability, because we could not be sure if the included individuals matched our review question and a current application of the model.
- When a publication lacked information necessary to answer a signalling question, but other available information could be used as an indication, we made assumptions regarding the answer (probably yes, probably no) to avoid the option 'no information'.
- For the CLL-IPI, we frequently encountered the problem that one predictor was replaced by a proxy predictor in the analysis of an external validation study. We assumed that this may happen in clinical practice, and included the external validation study to validate the CLL-IPI, although strictly speaking, they did not apply the original model. We rated the concern for applicability as unclear in the domain predictor.
- Some models developed to predict TFS include predictors also used for treatment indication, which means that the predictors are not excluded from the outcome definition (PROBAST item 3.3; e.g. O-CLL1 model: Rai staging is included in the prognostic score, and is, in combination with other characteristics, a factor for treatment indication). This may lead to an overestimation of the association between predictor and outcome, which is called incorporation bias (Moons 2019). We answered the corresponding PROBAST item 3.3 with 'no' and considered the risk of bias to be high for the outcome domain. We decided that all corresponding external validations, which validate a model that include a predictor also used for treatment indication, are also considered to contain a high risk of bias.
- The PROBAST guidance for item 4.1, which states that a study should include at least 20 events per candidate predictor for development studies and 100 events for validation studies is recommended for regression analysis with a binary outcome. The models included in our systematic review are mainly based on Cox proportional hazard models (Moons 2019). We therefore considered a cohort with more than 100 events to be an appropriate sample size for development and external validation samples (Collins 2016).
- The testing of model assumptions (PROBAST item 4.6) was very rarely reported. We did not let this item influence our judgement regarding the analysis domain because, by default, all except one study would have been considered to be of unclear risk of bias. We assumed that authors tested the pre-conditions before using their analysis type. However, we would advise authors of primary studies to improve reporting of the entire model development process.

Measures of prediction model performance

In contrast to 'classical' meta-analysis focusing on treatment efficacy as the parameter of interest, there is no single recommended methodology to meta-analyse the predictive performance of prognostic models (Debray 2012; Debray 2014). Hence, we extracted the reported predictive performance measures for each studied model from the development and validation studies, notably the model's discrimination and calibration in the development cohort (i.e. apparent performance, the performance measures after internal validation) and the performance found in a model's external validation.

Calibration refers to the accuracy of the predicted risk probabilities, which means the agreement between estimated and observed number of events in a cohort. A model might predict more events than the number of observed events, i.e. there is overestimation of the number of events. In contrast, a model can predict fewer events than observed meaning that the number of events is underestimated. In both situations, the model is miscalibrated. The use of a miscalibrated model in clinical practice can misguide clinical decisions based on such models (Van Calster 2019). Calibration can be presented as a calibration plot (expected probabilities plotted against observed outcome frequencies), as ratios between observed and expected number of events or outcome frequencies (O:E ratios) or calibration table (Debray 2017).

Discrimination refers to the ability of a prediction model to differentiate between those who do or do not experience the outcome event. A model has perfect discrimination if the predicted risks for all individuals who develop the outcome are higher than those for all individuals who do not experience the outcome. Discrimination is commonly estimated by the so-called concordance index (c-statistic, also sometimes called c-index). The c-statistic reflects the probability that for any randomly selected pair of individuals, one with and one without the outcome, the model assigns a higher probability to the individual with the outcome. The c-index can range from 0 to 1, with 1 indicating perfect discriminative ability and 0.5 meaning the model's predictions equal chance. The c-index is identical to the area under the receiver-operating characteristic (ROC) curve for models with binary endpoints, and can be generalised for time-to-event (survival) models accounting for censoring (Debray 2017).

Both measures are needed to assess the performance of a prognostic model.

Unit of analysis issues

Ideally, a prognostic model is developed and tested on specially designed, prospectively followed cohort studies consisting of a representative case mix. However, more often we identified models or validations derived from retrospective data or data from randomised controlled trials (RCTs). Data from these populations are used for several purposes and models. Where a model was developed and externally validated in the same publication, we compared the institution and year of inclusion of individuals to assure independence of the cohorts. We also compared the institution and year of inclusion of all external validation cohorts belonging to one developed model to avoid overlaps in participants. When the same cohort was used to validate two or more different prognostic models, we considered them as separate studies. When RCT data were used, the randomisation of

individuals was not considered by the authors of prognostic model studies.

Dealing with missing data

When data were missing, we requested additional information from the original investigators. When confidence intervals for measures of discrimination were still missing after we contacted the author, we calculated these according to a method proposed by Newcombe (Debray 2017; Debray 2018a; Newcombe 2006).

Investigation/description of heterogeneity

For the c-statistic, we used the between-study standard deviation (τ) to quantify possible heterogeneity. As we did not summarise O:E ratios to pool calibration, we merely visually inspected differences in survival per subgroup between cohorts. We planned to investigate and discuss clinical and statistical heterogeneity and design aspects of included studies mentioned in the section 'Data extraction and data management' based on subgroup analysis. We were unable to perform prespecified subgroup analysis based on diagnostic criteria due to the low number of studies with clear distinction between criteria.

Discussing reporting deficiencies

It is widely recommended that all prognostic models assess and report calibration and discrimination (Collins 2015; Moons 2015). However, it is known from numerous systematic reviews on methodological conduct and reporting of prognostic models in various disciplines that calibration is rarely reported and when it is reported (Heus 2018), it is done quite poorly. Hence, we also evaluated 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 are some indications that prognosis research is cluttered with false-positive studies which would not have been published if the results were negative. Moreover, studies evaluating development studies of prognostic models are not prospectively registered, and usually no protocol is published (Peat 2014). Therefore, it is difficult to assess publication bias. We used sensitive search strategies to increase retrieval (Geersing 2012).

Data synthesis

Data synthesis and meta-analysis approaches

In the protocol, we explained that we would pool the performance measures for calibration and discrimination. During the review process however, new methodological developments became available and we chose to adopt them, as follows. Although we assessed both model development and external validations for risk of bias, we performed meta-analysis only for validation studies, as there was only one development study per model.

- When a particular model has been validated at least three times for the primary outcome, we applied a random-effects model for pooling the logit transformation of the discrimination measure (the c-statistic) using the meta-analysis packages 'metamisc' and 'metafor' in the R statistical language (Debray 2014; Debray 2018b; Viechtbauer 2010).
- We pooled only validation studies with the outcome that a model was developed for (e.g. we did not meta-analyse

validation studies that tested a model to predict TFS if the model was developed for OS).

- We used random-effects meta-analysis, because validation studies typically differ regarding several parameters, including patient characteristics and design.
- Where the c-statistic was not reported, we did not estimate it. When the c-statistic was provided without measures of uncertainties, we calculated the standard error and confidence intervals based on the P value or the combination of sample size and number of events, if available, according to Newcombe and colleagues (Debray 2017; Debray 2018a; Newcombe 2006).

We planned to summarise the measures of calibration. Unfortunately, calibration measures and the expected survival, both in development and external validation cohorts, were rarely reported. Instead, many studies reported the observed outcome frequency per subgroup at a specific time point. Though it is possible to estimate approximate O:E ratios from the reported observed survival per risk category and probabilities reported in the model development study (for scores), in collaboration with the Cochrane Prognosis Methods Group, we decided against this estimation because it is merely an approximate indication for the calibration of a model. For time-to-event outcomes, we would need to account for censoring (Debray 2018a), which was not possible based on the limited reporting of the studies. Instead, we illustrated the survival per risk category for each model with sufficient data graphically. We pooled survival per risk category for all external validations of one model and reported a prediction interval. The 95% prediction interval (95% PI) is an estimate of the range in which a future average survival frequency in a new validation study of the prognostic model will fall with a 95% probability.

Subgroup analysis

We planned to investigate whether the change of definition of CLL over time has affected the performance of a model (i.e. Cheson 1996; Hallek 2008). We did not perform this analysis because the number of studies that reported a clear distinction between the definitions was too low.

Sensitivity analysis

During the review process, and before data extraction and analysis, we decided to conduct sensitivity analysis based on the 'Risk of bias' rating, if a sufficient number of validation studies were available per model.

We decided post hoc to conduct the following model-specific meta-analyses.

- Test the effect of studies that reported the area under the curve (AUC) instead of the c-statistic (which we meta-analysed together). These two measures for discrimination are the same for binary outcomes, but may differ for time-to-event data depending on the time point chosen to calculate the AUC.
- Explore the effect of the estimation of the 95% CI for the c-statistic. Estimating confidence intervals can introduce imprecision, therefore we aimed to explore the extend of the changes between reported and estimated 95% CIs.
- For the CLL-IPI, several external validation studies used a proxy prognostic factor for TP53 deletion or mutation, the cytogenetic aberration based on FISH, del(17p). Although we expected the effect to be small as the concordance between the two

predictors is larger than 90% (Dicker 2009), using a proxy can reduce predictive performance.

Rating the certainty of evidence and summary of findings

Originally, we planned to use the most up-to-date GRADE guidance for this systematic review. However, no GRADE guidance for grading the certainty of results from meta-analysis of prognostic models yet exists. Hence, for this review, we decided to refrain from applying GRADE.

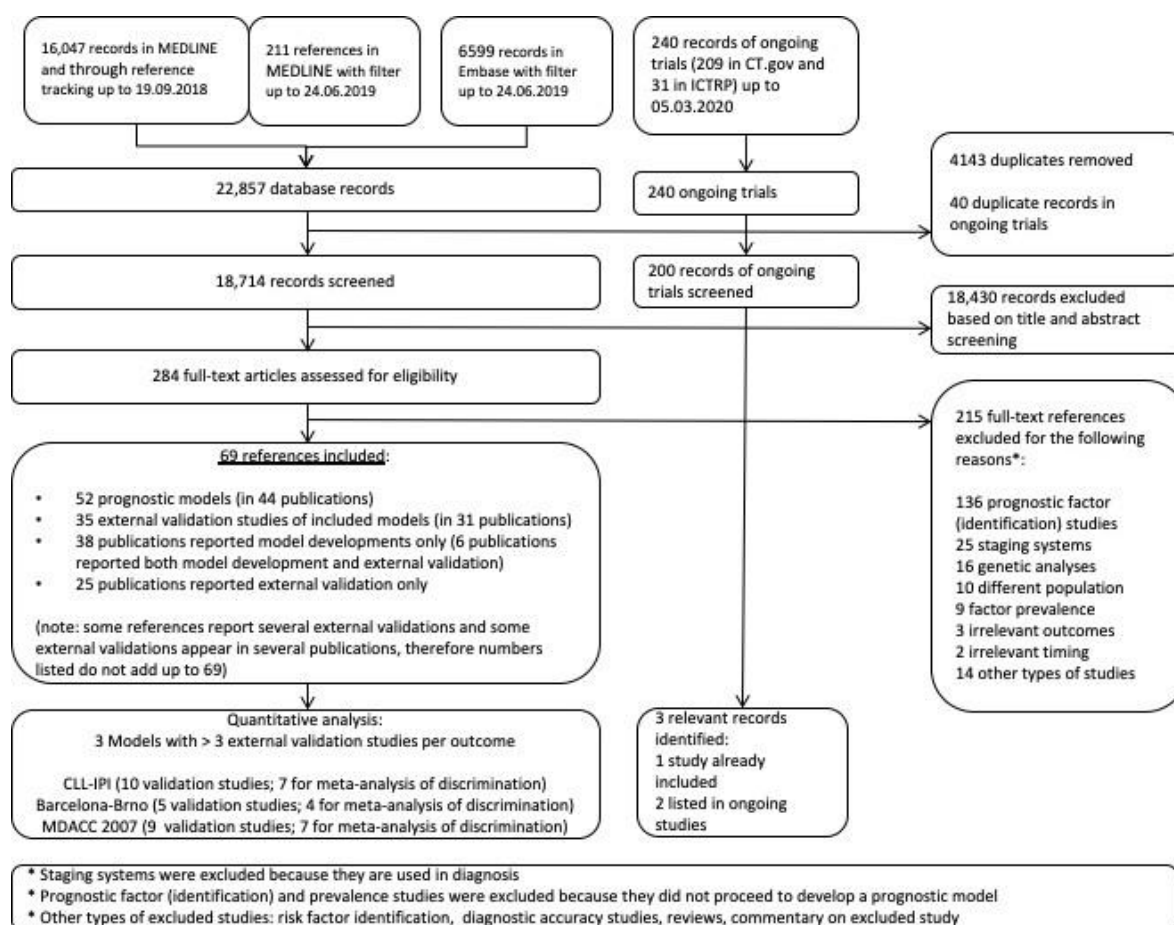
RESULTS

Description of studies

Results of the search

Our literature search (24 June 2019; MEDLINE and Embase) resulted in 22,857 potentially relevant references related to prognostic models for newly-diagnosed untreated individuals with chronic lymphocytic leukaemia (CLL). Of these, we removed 4144 duplicates, leaving 18,713 references for title and abstract screening. We identified 284 full-text references, which might fulfil our predefined inclusion criteria, including two references identified by reference and citation tracking. Of these, we excluded 214 references. We documented the overall number of screened, included and excluded references in a PRISMA flow diagram (Figure 1).

Figure 1. Study flow diagram



This led to a total of 70 included references. Of these, 38 reported model developments only, six publications reported both model development and external validation(s), and 25 publications reported external validation only. In total, we identified 52 prognostic models (in 44 publications), and 35 external validation studies of included models that predicted the primary outcome of one of the included developed models (in 31 publications). Several of these publications offered secondary analyses with another

outcome. These are only listed in this review if the primary outcome was not reported.

Out of the 52 developed models, 12 models were externally validated at least once either in the primary publication or as described in an additional publication (for an overview, see Appendix 6). Forty-one models did not undergo any external validation.

The search of databases for ongoing trials resulted in 240 records (ClinicalTrials.gov: 209 and ICTRP: 31; 5 March 2020). Of these, 40 records were duplicates. We identified three records that referred to the potential use of developing a prognostic model. One cohort had already been included from the database search (O-CLL-GISL, development of a model and several validations, e.g. [O-CLL-1 D - Gentile 2016 \(O-CLL-1-GISL\)](#)). One study is currently recruiting ([NCT03436524](#)), and one mentioned the aim to develop a staging system, however, it does not refer to any relevant publication yet ([NCT00275054](#)).

It is important to note the difference between 'reference' and 'study': one reference of a model development study may in addition to the model development also report one or several external validation studies of their model in independent cohorts. Furthermore, some references of external validations may report and compare the performances of several prognostic models in their cohort. Several publications used the same dataset to validate the same model(s).

References of model development studies without any external validation studies are summarised in [Studies awaiting classification](#) and [Appendix 7](#).

Included studies

We described the characteristics of included studies per model and per predicted outcome. The numbers of external validations belonging to each developed model and outcome may not correspond to the number of included references. We defined a study as a validation of one model in one independent cohort, which implies that the validation of two different models in the same cohort are counted as two separate studies for the description in this review. In some publications, one model was externally validated in several cohorts, while other publications validated several models in one cohort. We presented the number of participants included in analysis instead of the number of participants in the study sample or complete registry as, often, missing values were reported to be an exclusion reason. We only included external validation studies that predicted the same outcome that the prognostic model was developed for.

Models with meta-analysis (> 3 external validations)

- CLL-IPI (OS, [CLL-IPI D - Bahlo 2016 \(development cohort\)](#))
- Barcelona-Brno model (OS, [Barcelona-Brno D - Delgado 2017 \(Barcelona cohort\)](#))
- MDACC 2007 (OS, [MDACC 2007 D - Wierda 2007 \(MDACC\)](#))

Models without meta-analysis (1-3 external validations)

- GCLLSG model (OS, [GCLLSG D - Pflug 2014 \(GCLLSG\)](#))
- Rossi model (OS, [Rossi D - Rossi 2013 \(Italian cohort\)](#))

- Stephens (OS, [Stephens OS D - Stephens 2015 \(Ohio cohort\)](#))
- Baliakas model (TFS, [Baliakas D - Baliakas 2019 \(multicentre\)](#))
- GIMEMA model (PFS, [GIMEMA D - Molica 2005 \(GIMEMA cohort\)](#))
- MDACC 2011 (TFS, [MDACC 2007 D - Wierda 2007 \(MDACC\)](#))
- Morabito model (TFS, [Morabito D - Morabito 2009 \(Italian cohort\)](#))
- O-CLL1 model (TFS, [O-CLL-1 D - Gentile 2016 \(O-CLL-1-GISL\)](#))
- Stephens model (TFS, [Stephens TFS D - Stephens 2015 \(Ohio cohort\)](#))

Models with more than three external validation studies per outcome

Prognostic models for the prediction of OS (with meta-analysis, > 3 external validations)

CLL-IPI (CLL-IPI D - Bahlo 2016 (development cohort))

The CLL-IPI model predicts OS and was derived from combined data from eight phase three RCTs conducted in multiple countries (France, Germany, Poland, UK, USA), with a total of 1799 individuals included for the model development. The recruitment period for the trials lasted from 1997 to 2009 ([CLL-IPI D - Bahlo 2016 \(development cohort\)](#)). The median follow-up time was 79.9 months (interquartile range (IQR) 79.9 to 101.4 months). All disease stages were included; most individuals were classified as Rai I/ II or Binet B stage (50% and 41% respectively). The final model includes five predictors: TP53 status, IgHV mutational status, serum B2-microglobulin, clinical stage (Rai or Binet), and age. The original weighting of predictors was simplified to present as a risk score for easier use in clinical practice. We identified 11 external validation cohorts ([CLL-IPI V - Bahlo 2016 \(Mayo clinic 2001-2014\)](#); [CLL-IPI V - Bahlo 2016 \(SCAN cohort\)](#); [CLL-IPI V - Da Cunha-Bang 2016 \(Danish cohort\)](#); [CLL-IPI V - Delgado 2017 \(Barcelona cohort\)](#); [CLL-IPI V - Gentile 2016 \(Italian cohort\)](#); [CLL-IPI V - Molica 2016 \(O-CLL1-GISL\)](#); [CLL-IPI V - Muñoz-Novas 2018 \(Spanish cohort\)](#); [CLL-IPI V - Rani 2018 \(Indian cohort\)](#); [CLL-IPI V - Reda 2017 \(Milano cohort\)](#); [CLL-IPI V - Rigolin 2017 \(Ferrera cohort\)](#); [CLL-IPI V - Zhu 2018 \(Chinese cohort\)](#)), with 10 validations for OS and eight for TFS. The validation studies for OS included a total of 5485 individuals with available data presenting between 1983 and 2016. Two relevant validation studies lacked information regarding the recruitment period ([CLL-IPI V - Delgado 2017 \(Barcelona cohort\)](#); [CLL-IPI V - Rani 2018 \(Indian cohort\)](#)). Due to lack of reporting, only 3307 individuals with 917 deaths from seven external validation studies were included in the meta-analysis of the discrimination ([CLL-IPI V - Bahlo 2016 \(Mayo clinic 2001-2014\)](#); [CLL-IPI V - Bahlo 2016 \(SCAN cohort\)](#); [CLL-IPI V - Delgado 2017 \(Barcelona cohort\)](#); [CLL-IPI V - Gentile 2016 \(Italian cohort\)](#); [CLL-IPI V - Muñoz-Novas 2018 \(Spanish cohort\)](#); [CLL-IPI V - Rani 2018 \(Indian cohort\)](#); [CLL-IPI V - Zhu 2018 \(Chinese cohort\)](#)). We provide a summary of the main characteristics of studies regarding the CLL-IPI in [Figure 2](#).

Figure 2. CLL international prognostic index - summary of characteristics of included studies

Characteristics of included studies for the CLL-IPI externally validated model											
Model		Study	Characteristics of included studies								
			recruitment	country	number of events / number persons	age	gender	stage	treatment	design	follow-up
CLL-IPI model (overall survival)	Development	Bahlo 2016 (Development cohort)	1997 - 2009	France, Germany, Poland, UK, USA, etc.	462 events 1214 persons (train) 243 events 585 persons (test)	median 61 years (range: 27 - 86 years)	70% male	Rai 0: 14% Rai I/II: 50% Rai III/IV: 36%	690 F 1017 FC 113 CC 475 FCR 81 FCCAM 479 CHR	combined RCT data	median 79.9 months, IQR 79.9 - 101.4
		Bahlo 2016 (Mayo clinic 2001-2014)	2001 - 2014	USA	144 events 838 persons	≤ 65 years: 533 persons > 65 years: 305 persons	67.8% male	Rai 0: 57.4% Rai I-II: 38.5% Rai III-IV: 4.1%	168 purine analogs w/o MoAbs 37 Alkylator-based Chemoimmunotherapy 11 Alkylator w/o MoAbs 97 antibody only 7 RTK inhibitors 6 other	prospective cohort	median 63.2 months, IQR 30.2 - 91.8
	Meta-analysed c-statistic	Bahlo 2016 (SCAN cohort)	1999 - 2002	Denmark, Sweden	215 events 416 persons	≤ 65 years: 242 persons > 65 years: 174 persons	62.7% male	Binet A: 78.1% Binet B: 16.6% Binet C: 5.3%	NR	unclear	median 151 months, IQR 124.8 - 163.0
		Delgado 2017 (Barcelona cohort)	NR	Spain	212 events 524 persons	median 62 years (range: 22 - 93 years)	60% male	Rai 0: 62%; Rai I-IV: 38%	83 FCR or similar 82 purine analogs w/o MoAbs 86 received alkylating agents 23 received others	retrospective cohort	median 99.6 months (range 1 - 456 months)
		Gentile 2016 (Italian cohort)	1985 - 2015	Italy	174 events 858 persons	median 65.5 years (range: not reported)	56.2% male	Rai 0: 58.4% Rai I: 19.8% Rai II: 15.2% Rai III: 1.5% Rai IV: 5.1%	130 received chemotherapy 174 received chemo-immunotherapy	retrospective cohort	median 69.6 months, range not reported
		Muñoz-Novaz 2018 (Spanish cohort)	1989 - 2013	Spain	47 events 258 persons	median 65.7 years (IQR: 55.2 - 73.5)	62.8% male	Rai 0: 58.9%; Rai I-IV: 41.1%	NR	retrospective cohort	median 68 months (range 3 - 277 months)
		Rani 2018 (Indian cohort)	NR	India	86 events 198 persons	median 60 years (range: not reported)	77% male	Rai 0: 14.6%; Rai I: 21.2%; Rai II: 34.9%; Rai III: 14.1%; Rai IV: 15.2%	62 CHL 56 rituximab-based therapy 20 other therapies	NR	median 40.5 months (range 1 - 215 months)
		Zhu 2018 (Chinese cohort)	2002 - 2017	China	46 events 215 persons	median 60 years (range 16 - 85 years)	66.5% male	Rai 0: 16.7% Rai I-II: 54.0% Rai III-IV: 29.3%	147 treated with benzodiazepine FCR bendamoxetone R R+FFP+HDMP	retrospective cohort	median 48 months, (range: 1 - 192 months)
		Narrative	Da Cunha-Bang 2016 (Danish cohort)	2008 - 2015	Denmark	249 events 1514 persons	median 69 years (range: not reported)	60% male	Binet A: 80% Binet B/C: 20%	NR	prospective cohort
	Molica 2016 (O-CLL1-GISL)		2007 - NR	Italy	91 events (TFS) 337 persons	median 61 years (range: 33 - 70 years)	57.2% male	Rai 0: 77.8% Rai ≥ I: 22.2%	NR	prospective cohort	median 42 months, 2038 person-years
	Reda 2017 (Milan cohort)		1983 - 2016	Italy	NR events 329 persons	median 65 years (range: 32 - 91 years)	57% male	Rai 0-I: 74%; Rai II-IV: NR	190 CHL 108 FCR 12 O-Benda 44 BR 65 alemtuzumab 29 Ibrutinib or Idelalisib	retrospective cohort	median 144 months (range 0 - 360 months)
	Rigolin 2017 (Ferrera cohort)		2006 - 2016	Italy	NR events 335 persons	median 68.7 years (range: 33 - 96 years)	57.9% male	Binet A: 77.9% Binet B: 14.3% Binet C: 7.8%	114 chemotherapy	retrospective cohort	NR

Abbreviations: NR: not reported; CHL: chlorambucil; F: fluradabine; CC: cyclophosphamide + cladribine; FC: fluradabine + cyclophosphamide; FCR: fluradabine + cyclophosphamide + rituximab; FCCAM: FC + alemtuzumab; MoAbs: monoclonal antibody; R: rituximab; EEP: fresh frozen plasma; HDMP: high-dose methotrexate/cisplatin

Abbreviations: NR: not reported; CHL: chlorambucil; F: fludarabine; CC: cyclophosphamide + cladribine; FC: fludarabine + cyclophosphamide; FCR: fludarabine + cyclophosphamide + rituximab; FCCAM: FC + alemtuzumab; MoAbs: monoclonal antibody; R: rituximab; FFP: fresh frozen plasma; HDMP: high-dose methylprednisolone

Barcelona-Brno model (Barcelona-Brno D - Delgado 2017 (Barcelona cohort))

The Barcelona-Brno model predicts OS and was derived from data from a retrospective, single-centre cohort study conducted in Spain (Barcelona-Brno D - Delgado 2017 (Barcelona cohort)). TFS was reported as a secondary outcome. Model development included 524 individuals. The recruitment period was not reported. The median follow-up time was 99.6 months, ranging from 1 to 456 months. All disease stages were included; most individuals were Rai 0 or Binet A stage (62% and 83% respectively). The model aimed to simplify the CLL-IPI described below, therefore only combinations of the factors included in the CLL-IPI were tested. The final model included three predictors: IgHV mutational status, and the genomic aberrations del(17p), and del(11q). We identified six external validation cohorts (Barcelona-Brno V - Delgado 2017 (Brno cohort); Barcelona-Brno V - Gentile 2017 (Italian & Mayo);

Barcelona-Brno V - Molica 2017 (O-CLL1-GISL); Barcelona-Brno V - Muñoz-Novas 2018 (Spanish coh.); Barcelona-Brno V - Rani 2018 (Indian cohort); Barcelona-Brno V - Reda 2017 (Milan cohort)), with five validations for OS and six for TFS. The validation studies for OS included a total of 2501 individuals presenting between 1983 and 2016. Two validation studies lacked information regarding the recruitment period (Barcelona-Brno V - Delgado 2017 (Brno cohort); Barcelona-Brno V - Rani 2018 (Indian cohort)). Due to lack of reporting, only 1755 individuals with 416 deaths from three studies were included in meta-analysis of discrimination (Barcelona-Brno V - Delgado 2017 (Brno cohort); Barcelona-Brno V - Gentile 2017 (Italian & Mayo); Barcelona-Brno V - Muñoz-Novas 2018 (Spanish coh.); Barcelona-Brno V - Rani 2018 (Indian cohort)). We provide a summary of the main characteristics of studies regarding the Barcelona-Brno model in Figure 3.

Figure 3. Barcelona-Brno score - summary of characteristics of included studies

Characteristics of included studies for the Barcelona-Brno externally validated model											
Model		Study	Characteristics of included studies								
			recruitment	country	number of events / number persons	age	gender	stage	treatment	design	follow-up
Barcelona-Brno model (overall survival)	Development	Delgado 2017 (Barcelona cohort)	NR	Spain	212 events 524 persons	median 62 years (range: 22 - 93 years)	60% male	Rai I: 62%; Rai I-IV: 38%	83 FCR or similar 82 purine analogs w/o MoAbs 86 received alkylating agents 23 received others	retrospective cohort	median 99.6 months (range 1 - 456 months)
		Delgado 2017 (Brno cohort, CZ)	NR	Czech Republic	158 events 417 persons	median 62 years (range: 32 - 85 years)	66% male	Rai I: 41%; Rai I-II: 43%; Rai III-IV: 16%	125 FCR or similar 41 purine analogs w/o MoAbs 53 alkylating agents 57 received others	NR	median 59 months (range 3 - 330 months)
	Meta-analysed c-statistic	Gentile 2017 (Italian multicentre and Mayo cohort)	1985 - 2015	Italy & USA	283 events 1299 persons	median 63 years (range: 27 - 92 years)	61.3% male	Rai I: 57.9%; Rai I: 28.6% Rai II: 8.6%; Rai III: 1.5%; Rai IV: 3.5%	194 chemotherapy 316 chemoimmunotherapy	retrospective cohort	median 82 months (range 3 - 330 months)
		Muñoz-Novas 2018 (Spanish cohort)	1989 - 2013	Spain	47 events 258 persons	median 65.7 years (IQR: 55.2 - 73.5)	62.8% male	Rai I: 58.9%; Rai I-IV: 41.1%	137 received therapy, unclear which therapy	retrospective cohort	median 68 months (range 3 - 277 months)
		Rani 2018 (Indian cohort)	NR	India	86 events 198 persons	median 60 years (range: not reported)	77% male	Rai I: 14.6%; Rai I: 21.2%; Rai II: 34.9%; Rai III: 14.1%; Rai IV: 15.2%	62 CHL 56 rituximab-based therapy 20 other therapies	NR	median 40.5 months (range 1 - 215 months)
		Narrative	Molica 2017 (O-CLL1-GISL)	2007 - NR	Italy	91 events 337 persons	median 61 years (range: 33 - 70 years)	57.2% male	Rai I: 77.8%; Rai I-II: 22.2%	NR	prospective cohort
	Reda 2017 (Milan cohort)		1983 - 2016	Italy	NR events 329 persons	median 65 years (range: 32 - 91 years)	57% male	Rai I-II: 74%; Rai II-IV: NR	190 CHL 108 FCR 12 O-Benda 44 BR 65 alemtuzumab 29 ibrutinib or idelalisib	retrospective cohort	median 144 months (range 0 - 360 months)
Abbreviations: NR: not reported; CHL: chlorambucil; O-Benda: ofatumumab/bendamustine; BR: bendamustine-rituximab; MoAbs: monoclonal antibodies											

Abbreviations: NR: not reported; CHL: chlorambucil; O-Benda: ofatumumab/bendamustine; BR: bendamustine-rituximab; MoAbs: monoclonal antibodies

MDACC 2007 (Wierda 2007) (MDACC 2007 D - Wierda 2007 (MDACC))

The MDACC 2007 model predicts OS and was derived from data from a retrospective, single-centre cohort study conducted in the USA (MDACC 2007 D - Wierda 2007 (MDACC)). A total of 1674 individuals were included for the model development. The recruitment period for the study lasted from 1981 to 2004. The median follow-up time was 58.8 months (95% confidence interval (CI) 55.2 to 61.2 months). All disease stages were included; most individuals were classified as Rai I or Binet A stage (739/1674 and 1019/1674 respectively). The final model included six predictors: age, absolute lymphocyte count, serum B2-microglobulin, nodal groups, Rai stage, and gender. The publication reports both the formula to be used in combination with a nomogram, and a simplified index score. We identified 10 external cohorts (MDACC 2007 V - Trajkova 2013 (Macedonia); MDACC 2007 V - Bulian 2011 (Italian-Swiss); MDACC 2007 V - Gentile 2014 (Italian cohort); MDACC 2007 V - Gentile 2016 (Mayo cohort); MDACC 2007 V - González Rodríguez (Cabueñes); MDACC 2007 V - Molica 2010 (GIMEMA cohort); MDACC 2007 V - Molica 2015 (O-CLL1-GISL); MDACC 2007 V - Muñoz-Novas 2018

(Spanish cohort); MDACC 2007 V - Pflug 2014 (3 RCTs); MDACC 2007 V - Rani 2018 (Indian cohort)), with eight validations for OS and six for TFS. Seven studies used the index score and three additionally calculated the formula to predict OS. For one study, it was unclear if the index score or the formula were used. The eight cohorts for which OS was predicted included 6928 individuals presenting between 1983 and 2013. One validation study lacked information regarding the recruitment period (MDACC 2007 V - Rani 2018 (Indian cohort)). Only the simplified index score was sufficiently externally validated to be summarised in meta-analysis. In total, 5127 individuals with 994 deaths from seven studies remained for inclusion in the meta-analysis of discrimination (MDACC 2007 V - Bulian 2011 (Italian-Swiss); MDACC 2007 V - Gentile 2014 (Italian cohort); MDACC 2007 V - Gentile 2016 (Mayo cohort); MDACC 2007 V - González Rodríguez (Cabueñes); MDACC 2007 V - Muñoz-Novas 2018 (Spanish cohort); MDACC 2007 V - Pflug 2014 (3 RCTs); MDACC 2007 V - Rani 2018 (Indian cohort)). We provide a summary of the main characteristics of studies regarding the MDACC 2007 model in Figure 4.

Figure 4. MDACC 2007 index score - summary of characteristics of included studies

Characteristics of included studies for the MDACC 2007 externally validated model											
Model		Study	Characteristics of included studies								
			recruitment	country	number of events / number persons	age	gender	stage	treatment	design	follow-up
MDACC 2007 model (overall survival)	Development	Wierda 2007 (Development cohort)	1981 - 2004	USA	437 events 1617 persons	median 58 years (range: 0 - 90 years)	1029 male	Rai O: 469 Rai I: 739 Rai II: 235 Rai III: 93 Rai IV: 127	NR	retrospective cohort	median 58.8 months, 95% CI: 55.2 - 61.2
		Bulian (Italian-Swiss cohort)	1996 - 2011	Italy, Switzerland	151 events 1037 persons	median 65 years (range: 21 - 94 years)	58.2% male	Binet A: 74.6% Binet B: 18.4% Binet C: 6.9%	306 chemotherapy 129 chemoimmunotherapy 42 FCR 49 FR 38 R-other 40 with missing data	retrospective cohort	median 63.6 months, range not reported
	Meta-analysed c-statistic for the index score	Gentile 2014 (Italian multicentre cohort)	1983 - 2013	Italy	277 events 1502 persons	median 67 years (range: not reported)	55.7% male	Rai O: 56.5% Rai I: 21.4% Rai II: 15% Rai III: 2.6% Rai IV: 4.5%	337 chemotherapy 142 chemo-immunotherapy	retrospective cohort	median 68.4 months, range not reported
		Gentile 2016 (Mayo cohort 2001-2008)	2001 - 2008	USA	114 events 506 persons	median 62.5 years (range: 36 - 89 years)	68.6% male	Rai O: 62.5% Rai I: 28.9% Rai II: 5.1% Rai III: 1.2% Rai IV: 2.4%	NR	prospective cohort	median 86.4 months, range not reported
		González Rodríguez 2009 (Cabueñes hospital cohort)	1997 - 2007	Spain	76 events 257 persons	median 71.7 years (range: 42 - 94 years)	NR	Binet A: 76.8% Binet B: 14.8% Binet C: 8.4%	NR	retrospective cohort	NR
		Muñoz-Novaz 2018 (Spain multicentre cohort)	1989 - 2013	Spain	92 events 483 persons	median 67 years (range: 25 - 90 years)	64.2 % male	Rai O-II: 93.8% Rai III-IV: 6.2%	NR	retrospective cohort	median 46 months, range not reported
		Pflug 2014 (German CLL study group cohort)	1997 - 2006	Australia, Austria, Belgium, Czech Republic, Denmark, France, Germany, Israel, Italy, New Zealand, Spain	180 events 1144 persons	median 60 years (range: 30 - 81 years)	68.4% male	Rai O: 20.8% Rai I: 18.4% Rai II: 37.9% Rai III: 8.6% Rai IV: 14.3%	early F FC FCR	combined RCT data	median 63.4 months, range not reported
		Rani 2018 (Indian cohort)	NR	India	86 events 198 persons	median 60 years (range: not reported)	77% male	Rai O: 14.6%; Rai I: 21.2%; Rai II: 34.9%; Rai III: 14.1%; Rai IV: 15.2%	62 CHL 56 rituximab-based therapy 20 other therapies	NR	median 40.5 months (range 1 - 215 months)
	Narrative	Molica 2010 (GIMEMA cohort 1991-2000)	1991 - 2000	Italy	NR events 310 persons	median 64 years (range: not reported)	56.1% male	Rai O: 67% Rai I-III: 32.9% Rai III/IV: 0%	NR	retrospective cohort	median 41.5 months, range not reported
		Molica 2015 (O-CLL1 GISEL)	2007 - NR	Italy	91 events (TFS) 337 persons	median 61 years (range: 33 - 70 years)	57.1% male	Rai O: 77.5% Rai I-II: 22.5%	NR	prospective cohort	median 42 months, range not reported
		Trajkova 2013 (Macedonian cohort)	2011 - 2013	Macedonia		median 64.8 years (range: 47 - 78 years)	63% male	Rai O: 54% Rai I: 5% Rai II: 32% Rai III: 5% Rai IV: 4%	NR	unclear	NR

Abbreviations: NR: not reported; F: fludarabine; FCR: fludarabine + cyclophosphamide + rituximab; FR: fludarabine + rituximab; R-other: rituximab with other antineoplas or rituximab monotherapy; FC: fludarabine + cyclophosphamide; CHL: chlorambucil

Abbreviations: NR: not reported; F: fludarabine; FCR: fludarabine + cyclophosphamide + rituximab; FR: fludarabine + rituximab; R-other: rituximab with other regimens or rituximab monotherapy; FC: fludarabine + cyclophosphamide; CHL: chlorambucil

Prognostic models for the prediction of PFS or TFS (with meta-analysis, > 3 external validations)

We did not identify any prognostic model development studies that were externally validated more than three times for the outcomes PFS or TFS.

Models with one to three external validation studies per outcome

Prognostic models for the prediction of OS (without meta-analysis, 1 to 3 external validations)

GCLLSG model (Pflug 2014) (GCLLSG D - Pflug 2014 (GCLLSG))

The GCLLSG model predicts OS and was derived from combined data from three RCTs conducted in multiple countries (Australia, Austria, Belgium, Czech Republic, Denmark, France, Germany, Israel, Italy, New Zealand and Spain), with a total of 2007 individuals

included for the model development (GCLLSG D - Pflug 2014 (GCLLSG)). The recruitment period for the trials lasted from 1997 to 2006. The median follow-up time was 63.4 months; ranges were not reported. All disease stages were included; most individuals were Rai II or Binet A stage (37.9% and 42.6%, respectively). The final model included eight predictors: gender, age, Eastern Co-operative Oncology Group Performance Status (ECOG PS), del17p, del11q, IgHV mutational status, serum thymidine kinase, and serum B2-microglobulin. We identified three external validation cohorts, with one validation for OS (676 individuals, GCLLSG V - Pflug 2014 (Mayo cohort)), and two for TFS (GCLLSG V - Molica 2015 (O-CLL1-GISL); GCLLSG V - Rani 2018 (Indian cohort)). All validation studies lacked information regarding the recruitment period. We provide a summary of the main characteristics of studies regarding the GCLLSG model in Figure 5.

Figure 5. GCLLSG model - summary of characteristics of included studies

Characteristics of included studies for the GCLLSG externally validated model											
Model		Study	Characteristics of included studies								
			recruitment	country	number of events / number	age	gender	stage	treatment	design	follow-up
GCLLSG model (overall survival)	Development	Pflug 2014 (Development cohort)	1997 - 2006	Australia, Austria, Belgium, Czech Republic, Denmark, France, Germany, Israel, Italy, New Zealand, Spain	180 events 1144 persons	median 60 years (range: 30 - 81 years)	68.4% male	Rai 0: 20.8% Rai I: 18.4% Rai II: 37.9% Rai III: 8.6% Rai IV: 14.3%	early F FC FCR	combined RCT data	median 63.4 months, range not reported
		Molica 2015 (O-CLL-GISL)	2007 - NR	Italy	91 events 337 persons	median 61 years (range: 33 - 70 years)	57.1% male	Rai 0: 77.8%; Rai I-IV: NR	NR	prospective cohort	median 42 months (range 1 - 82 months; 2038 person-years)
	Narrative	Pflug 2014 (Mayo cohort)	NR	USA	190 events 676 persons	median 61.5 years (range: 32 - 89 years)	67% male	Rai 0: 57.1% Rai I: 34% Rai II: 5.9% Rai III: 1% Rai VI: 1.9%	NR	prospective cohort	median 57 months (range not reported)
		Rani 2018 (Indian cohort)	NR	India	86 events 198 persons	median 60 years (range: not reported)	77% male	Rai 0: 14.6%; Rai I: 21.2%; Rai II: 34.9%; Rai III: 14.1%; Rai IV: 15.2%	62 CHL 56 rituximab-based therapy 20 other therapies	NR	median 40.5 months (range 1 - 215 months)
Abbreviations: NR: not reported; F: fludarabine; FC: fludarabine + cyclophosphamide; FCR: fludarabine + cyclophosphamide-rituximab; CHL: chlorambucil											

Abbreviations: NR: not reported; F: fludarabine; FC: fludarabine + cyclophosphamide; FCR: fludarabine + cyclophosphamide + rituximab; CHL: chlorambucil

Rossi (Rossi 2013) (Rossi D - Rossi 2013 (Italian cohort))

The Rossi model predicts OS and was derived from data from a retrospective, multicentre cohort study conducted in Italy (Rossi D - Rossi 2013 (Italian cohort)). A total of 637 individuals were included for the model development. The recruitment period for the study lasted from 1996 to 2011. The median follow-up time was 67.2 months; ranges were not reported. All disease stages were included; most individuals were classified as Rai 0 stage (74.9%). The final model included five predictors: TP53, BIRC3 DIS, SF3B1 M, NOTCH1

M, and del11q22-q23. We identified two external validations (Rossi V - Jeromin 2014 (Munich cohort); Rossi V - Rossi 2013 (unclear)). One validation study included a total of 1160 individuals presenting between 2005 and 2010 (Rossi V - Jeromin 2014 (Munich cohort)). The second validation study was conducted between 1996 and 2011; the authors failed to report the number of participants included in the validation cohort (Rossi V - Rossi 2013 (unclear)). We provide a summary of the main characteristics of studies regarding the Rossi model in Figure 6.

Figure 6. Rossi model - summary of characteristics of included studies

Characteristics of included studies for the Rossi externally validated model											
Model		Study	Characteristics of included studies								
			recruitment	country	number of events / number	age	gender	stage	treatment	design	follow-up
Rossi model (overall survival)	Development	Rossi 2013 (Development cohort)	1996 - 2011	Italy	178 events 583 persons	51.2% >= 65 years	58.4% male	Rai 0: 74.9% Rai I-II: 11.5%	122 FCR, FR, or pentostatin + CR 64 F or FC 80 alkylator-based (e.g. CHL)	retrospective cohort	median 67.2 months (range not reported)
	Narrative	Jeromin 2014 (Munich cohort)	2005 - 2010	Unclear	NR events 930 persons	median 67 years (range: 29.6 - 90.5 years)	64.6% male	NR	NR	prospective cohort	median 55.2 months, (range not reported)
		Rossi 2013 (External validation cohort)	1996 - 2011	Italy	62 events 370 persons	NR	55.9% male	Rai 0-I: 75.4% Rai II: 17.3% Rai III-IV: 7.3%	NR	retrospective cohort	median 70.8 months, (range not reported)
Abbreviations: NR: not reported; F: fludarabine; FC: fludarabine + cyclophosphamide; FCR: fludarabine + cyclophosphamide + rituximab; FR: fludarabine/rituximab; CR: cyclophosphamide + rituximab; CHL: chlorambucil											

Abbreviations: NR: not reported; F: fludarabine; FC: fludarabine + cyclophosphamide; FCR: fludarabine + cyclophosphamide + rituximab; FR: fludarabine + rituximab; CR: cyclophosphamide + rituximab; CHL: chlorambucil

Stephens model (Stephens 2015) (Stephens OS D - Stephens 2015 (Ohio cohort))

Stephens and colleagues developed models for two outcomes, OS and TFS (Stephens OS D - Stephens 2015 (Ohio cohort); Stephens TFS D - Stephens 2015 (Ohio cohort)). They were derived from data from a retrospective, single-centre cohort study conducted in the USA. A total of 114 individuals were included for the model development. The recruitment period for the study lasted from 2002 to 2012. The follow-up time was not reported. All disease stages were included; most individuals were classified as Rai I/II

stage (46%). The final models for both outcomes were simplified for use in clinical practice. The simplified risk score for OS includes three predictors: ECOG PS, age, and lactate dehydrogenase. The simplified risk score for TFS included five predictors: ECOG PS, Rai stage, age, white blood cell count, and del11q22.3. We identified one external validation study (Stephens OS V - Stephens 2015 (MDACC)), which included 129 individuals. The recruitment period of the validation study was not reported. We provide a summary of the main characteristics of studies regarding the Stephens model in Figure 7.

Figure 7. Stephens model for OS - summary of characteristics of included studies

Characteristics of included studies for the Stephens externally validated model											
Model		Study	Characteristics of included studies								
			recruitment	country	number of events / number	age	gender	stage	treatment	design	follow-up
Stephens model (overall survival)	Development	Stephens 2015 (Development cohort)	2002 - 2012	USA	NR events 114 persons	median 62 years (range: 40 - 92 years)	NR	Rai 0: 33%; Rai I-II: 46%; Rai III-IV: 21%	35 purine analogs 18 treatment on a clinical trial 5 other treatment	retrospective cohort	NR
	Narrative	Stephens 2015 (MDACC cohort)	NR	USA	NR events 129 persons	median 63 years (range: 40 - 85 years)	NR	Rai 0: 29%; Rai I-II: 52%; Rai III-IV: 19%	NR	retrospective cohort	NR
Abbreviations: NR: not reported											

Abbreviations: NR: not reported

Prognostic models for the prediction of PFS or TFS (without meta-analysis, 1 to 3 external validations)

Baliakas model (Baliakas 2019) (Baliakas D - Baliakas 2019 (multicentre))

The Baliakas model predicts TFS and was derived from data from a retrospective, multicentre cohort study conducted in 10 European institutions (Baliakas D - Baliakas 2019 (multicentre)). A total of 1900 individuals were included for the development of two separate models, dividing the study population into mutated (M-CLL) and unmutated CLL (U-CLL) patients. The recruitment period for the study was not reported. The median follow-up time

was 7.1 years, ranging from 0.1 to 33.1 years. Only individuals with disease stage Binet A were included. The final model for M-CLL included three predictors: TP53 abnormality, trisomy 12, and stereotyped subset #2 defined as IGHV3-21/IGLV3-21 BcR IG. The final model for U-CLL included three predictors: TP53 abnormality, del11q, and gender. We identified one external validation study (Baliakas V - Baliakas 2019 (MLL + Scan.)), which included a total of 649 persons from two separate studies, one conducted at the Munich Leukemia Laboratory (508 persons) and one Scandinavian population-based study (141 persons). We provide a summary of the main characteristics of studies regarding the Baliakas model in Figure 8.

Figure 8. Baliakas model - summary of characteristics of included studies

Characteristics of included studies for the Baliakas externally validated model											
Model		Study	Characteristics of included studies								
			recruitment	country	number of events / number	age	gender	stage	treatment	design	follow-up
Baliakas model	Development	Baliakas 2019 (Development cohort)	NR	multiple European countries	NR events 1900 persons	median 64.3 years (range: 22 - 92 years)	61% male	Binet A: 80% Binet B: 12% Binet C: 8%	NR	retrospective cohort	median 7.1 years (range 0.1 - 33.1 years)
(time-to-first-treatment)	Narrative	Baliakas 2019 (MLL + Scan. cohort)	NR	Germany, Scandinavia	NR events 649 persons	median 63.6 years (range: 29 - 89 years)	62% male	Binet A: 100%	NR	retrospective cohort	NR
Abbreviations: NR, not reported											

Abbreviations: NR: not reported

GIMEMA model (Molica 2005) (GIMEMA D - Molica 2005 (GIMEMA cohort))

The GIMEMA model predicts PFS and was derived from data from a retrospective, multicentre cohort study conducted in Italy (GIMEMA D - Molica 2005 (GIMEMA cohort)). A total of 1138 individuals were included for the model development. The recruitment period for the study lasted from 1991 to 2000. The median follow-up time was 54 months, ranging from 4 to 309 months. Only individuals with

disease stage Binet A were included. The final model included four predictors: lymphocyte doubling time, absolute peripheral blood lymphocytosis, Rai stage, and gender. We identified one external validation study (GIMEMA V - González Rodríguez 2009 (Cabueñes coh.)), which included 265 persons presenting between 1997 and 2007. We provide a summary of the main characteristics of studies regarding the GIMEMA model in Figure 9.

Figure 9. GIMEMA model - summary of characteristics of included studies

Characteristics of included studies for the GIMEMA externally validated model											
Model		Study	Characteristics of included studies								
			recruitment	country	number of events / number	age	gender	stage	treatment	design	follow-up
GIMEMA model (progression-free survival)	Development	Molica 2005 (Development cohort)	1991 - 2000	Italy	235 events 593 persons	median 65 years (range: 27 - 100 years)	54% male	Rai 0: 77.2 % Rai I-III: 22.7 %	133 CHL 20 F or FC 7 CHOP-like regimen 40 other (CVP, C +/-prednisone)	retrospective cohort	median 54 months, range 4 - 309 months
	Narrative	González Rodríguez 2009 (Cabueñes hospital cohort)	1997 - 2007	Spain	76 events 257 persons	median 71.7 years (range: 42 - 94 years)	NR	Binet A: 76.8 % Binet B: 14.8 % Binet C: 8.4 %	NR	retrospective cohort	NR
Abbreviations: NR: not reported; CHL: chloambucil; F: fludarabine; FC: fludarabine + cyclophosphamide; CHOP: doxorubicin + cyclophosphamide + vincristine + prednisone; CVP: cyclophosphamide + vincristine sulfate + prednisone											

Abbreviations: NR: not reported; CHL: chlorambucil; F: fludarabine; FC: fludarabine + cyclophosphamide; CHOP: doxorubicin + cyclophosphamide + vincristine + prednisone; CVP: cyclophosphamide + vincristine sulfate + prednisone

MDACC 2011 (Wierda 2011) (MDACC 2011 D - Wierda 2011 (MDACC))

The MDACC 2011 model predicts TFS and was derived from data from a retrospective, single-centre cohort study conducted in the USA (MDACC 2011 D - Wierda 2011 (MDACC)). A total of 930 individuals were included for the model development. The recruitment period for the study lasted from 2004 to 2009. The median follow-up time was 26 months, ranging from three to

73 months. All disease stages were included; most individuals were classified as Rai I stage (51%). The final model included six predictors: IgHV mutational status, diameter of largest palpated lymph node, FISH category (del11q or del17p versus none), number of involved lymph node sites, lactate dehydrogenase (LDH), and the IgHV mutational status and LDH interaction term. We identified one external validation study (MDACC 2011 V - Molica 2016 (O-CLL1-

GISL)), which included 328 persons presenting between 2006 and

2010. We provide a summary of the main characteristics of studies regarding the MDACC 2011 model in [Figure 10](#).

Figure 10. MDACC 2011 model - summary of characteristics of included studies

Characteristics of included studies for the MDACC 2011 externally validated model											
Model		Study	Characteristics of included studies								
			recruitment	country	number of events / number	age	gender	stage	treatment	design	follow-up
MDACC 2011 model (time-to-first-treatment)	Development	Wierda 2011 (Development cohort)	2004 - 2009	USA	193 events 687 persons	median 59 years (range: 30 - 89 years)	61% male	Rai 0: 36% Rai I: 51% Rai II: 8% Rai III-IV: 5%	NR	retrospective cohort	median 26 months (range 3 - 73 months)
	Narrative	Molica 2016 (O-CLL1-GISL)	2006 - 2010	Italy	91 events 337 persons	median 61 years (range: 33 - 70 years)	58.2% male	Rai 0: 76.5% Rai I-II: 23.5%	NR	prospective cohort	median 30 months (range 1 - 65 months)
Abbreviations: NR: not reported											

Abbreviations: NR: not reported

Morabito model (Morabito 2009) (Morabito D - Morabito 2009 (Italian cohort))

The Morabito model predicts TFS and was derived from data from a multicentre study conducted in Italy ([Morabito D - Morabito 2009 \(Italian cohort\)](#)). The study design was not reported. A total of 262 individuals were included for the model development. The recruitment period was not reported. The median follow-up time was 36 months, ranging from 12 to 180 months. Only individuals

with disease stage Binet A were included. The final model included three predictors: IgHV mutational status, CD38, and ZAP-70. We identified one external validation study ([Morabito V - Gentile 2014 \(O-CLL1-GISL\)](#)), which included 480 persons presenting from 2007. There was no information regarding the end of the recruitment period. We provide a summary of the main characteristics of studies regarding the Morabito model in [Figure 11](#).

Figure 11. Morabito model - summary of characteristics of included studies

Characteristics of included studies for the Morabito externally validated model											
Model		Study	Characteristics of included studies								
			recruitment	country	number of events / number	age	gender	stage	treatment	design	follow-up
Morabito model (time-to-first-treatment)	Development	Morabito 2009 (Development cohort)	NR	Italy	51 events 252 persons	median 65 years (range not reported)	61.5% male	Binet A: 100%	NR	NR	median 36 months (range 12 - 180 months)
	Narrative	Gentile 2014 (O-CLL1-GISL)	2007 - NR	Italy	121 events 468 persons	47.4 % < 60 years	58.5% male	Rai 0: 77.3% Rai I-II: 22.7%	NR	prospective cohort	median 38.5 months (range 6 - 82 months)
Abbreviations: NR: not reported											

Abbreviations: NR: not reported

O-CLL1 model (Gentile 2016) (O-CLL-1 D - Gentile 2016 (O-CLL-1-GISL))

The O-CLL1 model predicts TFS and was derived from data from a prospective, multicentre cohort study conducted in Italy ([O-CLL-1 D - Gentile 2016 \(O-CLL-1-GISL\)](#)). A total of 480 individuals were included for the model development. Recruitment for the study started in 2007; the end date was not reported. The median follow-up time was 42 months, ranging from 6 to 82 months. Only individuals with disease stage Binet A were included. The final model included four predictors: Rai stage, absolute lymphocyte

count (ALC), serum B2-microglobulin, and IgHV mutational status. We identified two external validation cohorts for TFS ([O-CLL1 V - Gentile 2016 \(Mayo cohort\)](#); [O-CLL-1 V - Rani 2018 \(Indian cohort\)](#)), which included a total of 626 persons. One validation study recruited participants between 2001 and 2008 ([O-CLL1 V - Gentile 2016 \(Mayo cohort\)](#)); the other study lacked information regarding the recruitment period ([O-CLL-1 V - Rani 2018 \(Indian cohort\)](#)). We provide a summary of the main characteristics of studies regarding the O-CLL1 model in [Figure 12](#).

Figure 12. O-CLL-1 model - summary of characteristics of included studies

Characteristics of included studies for the O-CLL1 externally validated model											
Model		Study	Characteristics of included studies								
			recruitment	country	number of events / number	age	gender	stage	treatment	design	follow-up
O-CLL1 model (time-to-first-treatment)	Development	Gentile 2016 (O-CLL1-GISL)	2007 - NR	Italy	84 events 337 persons	46.7% <= 60 years	57.3% male	Rai 0: 77.7% Rai I-II: 22.3%	NR	prospective cohort	median 42 months (range 6 - 82 months)
	Narrative	Rani 2018 (Indian cohort)	NR	India	86 events 198 persons	median 60 years (range: not reported)	77% male	Rai 0: 14.6%; Rai I: 21.2%; Rai II: 34.9%; Rai III: 14.1%; Rai IV: 15.2%	62 CHL 56 rituximab-based therapy 20 other therapies	NR	median 40.5 months (range 1 - 215 months)
		Gentile 2016 (Mayo cohort 2001-2008)	2001 - 2008	USA	130 events 428 persons	40.2% <= 60 years	66.4% male	Rai 0: 67.5% Rai I-II: 32.5%	NR	NR	median 97 months (range not reported)
Abbreviations: NR: not reported; CHL: chlorambucil											

Abbreviations: NR: not reported; CHL: chlorambucil

Stephens model (Stephens 2015) (Stephens TFS D - Stephens 2015 (Ohio cohort))

For a description of the model for TFS in Stephens and colleagues ([Stephens TFS D - Stephens 2015 \(Ohio cohort\)](#)), please see

description in section OS (above), as the authors used the same development and validation cohorts for both outcomes. We provide a summary of the main characteristics of studies regarding the Stephens model in [Figure 13](#).

Figure 13. Stephens model for TFS - summary of characteristics of included studies

Characteristics of included studies for the Stephens externally validated model											
Model		Study	Characteristics of included studies								
			recruitment	country	number of events / number	age	gender	stage	treatment	design	follow-up
Stephens model (treatment-free survival)	Development	Stephens 2015 (Development cohort)	2002 - 2012	USA	58 events 114 persons	median 62 years (range: 40 - 92 years)	NR	Rai 0: 33%; Rai I-II: 46%; Rai III-IV: 21%	35 purine analogs 18 treatment on a clinical trial 5 other treatment	retrospective cohort	NR
	Narrative	Stephens 2015 (MDACC cohort)	NR	USA	NR events 129 persons	median 63 years (range: 40 - 85 years)	NR	Rai 0: 29%; Rai I-II: 52%; Rai III-IV: 19%	NR	retrospective cohort	NR
Abbreviations: NR: not reported											

Abbreviations: NR: not reported

Models without external validation studies

The 34 references with 41 model or score developments can be found in [Studies awaiting classification](#), as it is widely recommended that developed prediction models should not be used in daily practice before they were validated at least once ([Moons 2009](#); [Moons 2015](#); [Moons 2019](#); [Steyerberg 2013](#)). We have provided an overview of the main characteristics of these developed (but never validated) models in [Appendix 7](#). In total, 19 models were developed to predict OS, 19 to predict TFS and three to predict PFS. These models were published between 1982 and 2019 and contained an average number of 3.8 predictors (range between 2 and 6). Nineteen models were derived from retrospective cohorts, three were on prospective cohorts and 19 did not report the study design clearly. The most commonly included predictors were IgHV status (mutated versus unmutated), B2-microglobulin, age, clinical stage, genomic aberrations as defined by [Döhner 2000](#), ZAP-70 expression, CD38 expression and gender.

Excluded studies

During abstract screening, we excluded 18,713 references that clearly did not match our inclusion criteria. Of the remaining 283 full-text references, we excluded 213 for the following reasons.

- References of prognostic factor studies or prognostic factor identification studies (130 references) ([Berke 2019](#); [Bo 2014](#); [Brejcha 2010](#); [Brugiatelli 2007](#); [Bulian 2014](#); [Byrd 2006](#); [Cailliod 2005](#); [Callea 1999](#); [Catovsky 1989](#); [Cesano 2013](#); [Chang 2003](#); [Chauzeix 2018](#); [Chen 1997](#); [Chena 2008](#); [Chevallier 2002](#); [Chiaretti 2014](#); [Christiansen 1994](#); [Cicone 2012](#); [Claus 2012](#); [Claus 2014](#); [Cmunt 2002](#); [Cocco 2005](#); [Corcoran 2005](#); [Cordone 1998](#); [Cortese 2014](#); [Coscia 2012](#); [Crespo 2003](#); [Cro 2009](#); [D'Arena 2001](#); [D'Arena 2007](#); [Damle 1999](#); [DeAndres-Galiana 2016](#); [Degan 2004](#); [Delgado 2009](#); [Delgado 2014](#); [Del Giudice 2011](#); [Del Poeta 2010](#); [Del Principe 2004](#); [Del Principe 2006](#); [Di Raimondo 2001](#); [Dong 2011](#); [Dong 2014](#); [Durak 2009](#); [El-Kinawy 2012](#); [Gattei 2008](#); [Gdynia 2018](#); [Gentile 2016](#); [Giudice 2018](#); [Gogia 2014](#); [Grabowski 2005](#); [Han 1984](#); [Hock 2010](#); [Hus 2006](#); [Jaksic 1981](#); [Josefsson 2007](#); [Juliusson 1986](#); [Juliusson 1990](#); [Kahraman 2014](#); [Kardum-Skelin 2008](#); [Karmiris 1994](#); [Khalifa 2002](#); [Kim 2004](#); [Kimby 1988](#); [Knospe 1977](#); [Koberda 1989](#); [Korycka-Wolowiec 2011](#); [Krober 2002](#); [Kryachok 2011](#); [Kurec 1992](#); [Lai 2002](#); [Lech-Maranda 2012](#); [Lech-Maranda 2013](#); [Lecouvet 1997](#); [Li 2008](#); [Li 2017a](#); [Lin 2002](#); [Lin 2014](#); [Lozano-Santos 2014](#); [Lucas 2015](#); [Maffei 2007](#); [Maffei 2010](#); [Mansouri 2013](#); [Marasca 2005](#); [Marasca 2013](#); [Martinelli 2008](#); [Masic 1998](#); [Mateva 2001](#); [Matthews 2006](#); [Matthews 2007](#); [Matutes 2013](#); [Miao 2018](#); [Molica 1986](#); [Molica 1988](#); [Molica 1991](#); [Molica 1994](#); [Molica 1998](#); [Molica 1999a](#); [Molica 1999b](#); [Molica 2008](#); [Montserrat 1991](#); [Morabito 2001](#); [Morabito 2015a](#); [Morabito 2018c](#); [Morilla 2008](#); [Nenova 2000](#); [Nipp 2014](#); [Nowakowski 2009](#); [Nuckel 2006](#); [Nuckel 2009](#); [Ocana](#)

2007; [Oliveira 2011](#); [Oscier 1990](#); [Paolino 1984](#); [Prokocimer 1985](#); [Qin 2017](#); [Resegotti 1989](#); [Rissiek 2014](#); [Ronchetti 2016](#); [Sarmiento 2002](#); [Shanafelt 2010](#); [Spacek 2009](#); [Stamatopoulos 2017](#); [Stre&ord 2015](#); [Szymczyk 2018](#); [Vojdeman 2017](#); [Vural 2014](#); [Weiss 2011](#); [Wierda 2003](#); [Winkler 2010](#); [Zenz 2009](#)), and five references aimed at identifying prognostic factor thresholds ([Dasgupta 2015](#); [Davis 2016](#); [Rossi 2010a](#); [Tobin 2005a](#); [Tobin 2005b](#)).

- The main focus was on the development of a diagnostic CLL staging system (21 references) ([Apelgren 2006](#); [Baccarani 1982](#); [Binet 1981](#); [Binet 1977](#); [Chelazzi 1979](#); [Ciocoiu 1988](#); [De Faria 2000](#); [De Rossi 1989](#); [Ferrara 1981](#); [Jaksic 1992](#); [Molica 1984](#); [Moreno 2019](#); [Rai 1990](#); [Rai 1975](#); [Rossi 1986](#); [Rozman 1979](#); [Santoro 1979](#); [Scolozzi 1981](#); [Velardi 1980](#); [Wu 2010](#); [Zengin 1997](#)).
- Involved genetic analysis only (16 references), e.g. genetic subgrouping, genetic signature(s) or genetic clustering ([Baliakas 2015](#); [Bomben 2009](#); [Bou Samra 2014](#); [Chuang 2012](#); [Ferreira 2014](#); [Friedman 2009](#); [Herold 2011](#); [Houldsworth 2014](#); [Morabito 2015b](#); [Orgueira 2019](#); [Queiros 2015](#); [Raponi 2018](#); [Rodriguez 2007](#); [Vallat 2013](#); [Van Damme 2012](#); [Zucchetto 2006](#)).
- Included a population which did not match our PICOTS question (10 references), e.g. previously treated individuals with CLL ([Giles 2003](#); [Kardum-Skelin 2009](#); [Keating 2000](#); [Krober 2006](#); [Melo 1987](#); [Nola 2004](#); [O'Brien 1993](#); [Rossi 2010b](#); [Rossi 2011](#); [Weinberg 2007](#)).
- Nine references focused on the characterisation of CLL and the prevalence of prognostic factors ([Criel 1997](#); [Cro 2010](#); [Cuneo 2004](#); [D'Arena 2012](#); [Geisler 1997](#); [Gonzalez 2013](#); [Gonzalez-Gascon 2015](#); [Gonzalez-Rodriguez 2010](#); [Hallek 1999](#)).
- Nine references of other types of studies, e.g. risk factor study, diagnostic accuracy study ([Cuneo 2018](#); [Deslandes 2007](#); [Dimier 2018](#); [Fang 2019](#); [Kay 2018](#); [Kleinstern 2018](#); [Plesingerova 2017](#); [Salomon-Nguyen 1995](#); [Savopoulos 2016](#)).
- Four references developed or used an outdated staging system, which is no longer used in today's clinical practice ([Bettini 1986](#); [Chastang 1985](#); [French Cooperative Group on CLL 1988](#); [Mandelli 1987](#)).
- Five other types of references, e.g. review or comment to an excluded study ([Bomben 2005](#); [Grever 2006](#); [Jaksic 2014](#); [Matutes 2017](#); [Nedeva 2018](#)).
- Three references evaluated outcomes not relevant to this review ([Shanafelt 2017](#); [Stamatopoulos 2009](#); [Tallarico 2018](#)).
- Two references focused on scores or models which included predictors available at treatment initiation only (incorrect timing) ([Gentile 2018](#); [Nabhan 2017](#)).

Reporting deficiencies

Appendix 8 describes the reporting deficiencies of the included studies per model for the primary outcome each model was developed for. The number of studies includes the development study. Data obtained by contacting authors are included as available.

For the Barcelona-Brno model, the number of events for two studies and calibration for one study was obtained by contacting the corresponding authors of the primary studies. For the CLL-IPI, information on total sample size and number of events was provided by one author. Information on calibration was not reported in publications at all, but was obtained by contacting the corresponding authors. For the GCLLSG model, the confidence intervals of the c-statistic and calibration for the development study publication with one external validation was obtained from the authors. For the MDACC 2007 model, the c-statistic and 95% CI for one external validation study was provided by the corresponding author. We did not obtain any additional information for the other models. We received primary data of one study, but were unable to reconstruct the analysis.

Of the 41 models without external validations, all studies reported the total number of individuals included for analysis. Eighteen studies lacked information on the recruitment period, 19 did not clearly report their study design, 24 did not report the calibration of their model and 31 did not report any measure of discrimination for their model (Appendix 7). Since we included studies with a wide range of publication years, the large amount of missing information may be a result of the methodological evolution of prognostic model research. Moreover, with our decision to also include scores where authors decided seemingly ad hoc to assign points and create a prognostic score, we have included studies that show more resemblance with prognostic factor studies (i.e. testing of the independence of factors from each other) than with a prognostic model development.

Summarised over all categories of included studies (development studies without external validations (N = 40), development studies with external validations (N = 12), external validations of primary

outcome of an included model (N = 35)), information was especially lacking on calibration and discrimination. Calibration was reported in 17 and discrimination in 45 out of 87 studies.

Risk of bias and applicability assessment of included studies

Models with more than three external validation studies per outcome

Prognostic models for the prediction of OS (with meta-analysis, > 3 external validations)

CLL-IPI (CLL-IPI D - Bahlo 2016 (development cohort))

We rated the risk of bias of the development study as low for the domains: participants, predictors and outcome (CLL-IPI D - Bahlo 2016 (development cohort)), and high for the domain 'analysis' due to several reasons (univariable selection and dichotomisation of factors, missing data handling). We rated the risk of bias for the validation studies as low for the domain 'participants' for two studies (CLL-IPI V - Gentile 2016 (Italian cohort); CLL-IPI V - Muñoz-Novas 2018 (Spanish cohort)), unclear in the case of six studies due to unclear eligibility criteria, and high for two studies due to inappropriate inclusion criteria (CLL-IPI V - Bahlo 2016 (Mayo clinic 2001-2014); CLL-IPI V - Rani 2018 (Indian cohort)). We rated the risk of bias for the domains 'predictors' and 'outcome' as low for most studies, except one study with an unclear rating for the domain 'predictors' because predictor assessment has probably changed (CLL-IPI V - Gentile 2016 (Italian cohort)), and three studies with an unclear rating for the domain 'outcome' due to short observation time (CLL-IPI V - Da Cunha-Bang 2016 (Danish cohort); CLL-IPI V - Rani 2018 (Indian cohort); CLL-IPI V - Rigolin 2017 (Ferrera cohort)). Concerning the domain 'analysis', we considered seven studies to have a high risk of bias due to inappropriate handling of missing data, low number of events and lack of reporting of performance measures. Three studies had a low rating (CLL-IPI V - Bahlo 2016 (Mayo clinic 2001-2014); CLL-IPI V - Bahlo 2016 (SCAN cohort); CLL-IPI V - Delgado 2017 (Barcelona cohort)), and one had an unclear rating due to lack of performance measures (CLL-IPI V - Da Cunha-Bang 2016 (Danish cohort)). One validation study did not examine the outcome of interest (CLL-IPI V - Molica 2016 (O-CLL1-GISL)) (Figure 14).

Figure 14. Risk of bias (PROBAST) assessment of the CLL-IPI model (Bahlo 2016)

Model	Study	Prediction model Risk Of Bias Assessment Tool (PROBAST)				
		Domain 1: Participants	Domain 2: Predictors	Domain 3: Outcome	Domain 4: Analysis	Overall judgement
CLL-IPI model (overall survival)	Development	Bahlo 2016 (Development cohort)	low	low	low	high
		Bahlo 2016 (Mayo clinic 2001-2014)	high	low	low	high
	Meta-analysed c-statistic	Bahlo 2016 (SCAN cohort)	unclear	low	low	unclear
		Delgado 2017 (Barcelona cohort)	unclear	low	low	high
		Gentile 2016 (Italian cohort)	low	unclear	low	high
		Muñoz-Novas 2018 (Spanish cohort)	low	low	high	high
		Rani 2018 (Indian cohort)	high	low	unclear	high
		Zhu 2018 (Chinese cohort)	unclear	low	low	high
		Da Cunha-Bang 2016 (Danish cohort)	unclear	low	unclear	unclear
	Narrative	Molica 2016 (O-CLL1-GISL)	NA	NA	NA	NA
		Reda 2017 (Milan cohort)	unclear	low	low	high
		Rigolin 2017 (Ferrera cohort)	unclear	low	unclear	high

We rated the concern for applicability of the development study as high for the domain 'participants' due to the high proportion of individuals with treatment indication (CLL-IPI D - Bahlo 2016 (development cohort)). We rated this low for the domains 'predictors' and 'outcome'. We rated concern for applicability as unclear for the domain 'participants' in all validation studies but two, which received a low rating (CLL-IPI V - Da Cunha-Bang 2016

(Danish cohort); CLL-IPI V - Muñoz-Novas 2018 (Spanish cohort)), due to unclear eligibility criteria. We rated concern for applicability as unclear for the domain 'predictors' in six validation studies because one predictor (TP53 mutation) was replaced by a proxy predictor (del(17p)). The remaining four were rated as low (CLL-IPI V - Bahlo 2016 (SCAN cohort); CLL-IPI V - Delgado 2017 (Barcelona cohort); CLL-IPI V - Rigolin 2017 (Ferrera cohort); CLL-IPI V - Zhu

2018 (Chinese cohort)). We considered the concern for applicability to be low for the domain 'outcome' in all validation studies. One

validation study (CLL-IPI V - Molica 2016 (O-CLL1-GISL))) did not examine the outcome of interest (Figure 15).

Figure 15. Applicability assessment for all developed models with external validations

Model		Study	PROBAST - Applicability		
			Domain 1: Participants	Domain 2: Predictors	Domain 3: Outcome
Baliakas model (time-to-first-treatment)	Development	Baliakas 2019 (Development cohort)	high	low	low
	Narrative	Baliakas 2019 (MLL + Scan. cohort)	unclear	low	low
Barcelona-Brno model (overall survival)	Development	Delgado 2017 (Development cohort)	unclear	low	low
	Meta-analysed c-statistic	Delgado 2017 (Brno cohort, CZ)	unclear	low	low
		Gentile 2017 (Italian multicentre and Mayo cohort)	unclear	low	low
		Muñoz-Novaz 2018 (Spain multicentre cohort)	low	low	low
		Rani 2018 (Indian cohort)	unclear	unclear	low
	Narrative	Molica 2017 (O-CLL1-GISL)	NA	NA	high
CLL-IPI model (overall survival)	Development	Reda 2017 (Milan cohort)	unclear	unclear	low
	Meta-analysed c-statistic	Bahlo 2016 (Development cohort)	high	low	low
		Bahlo 2016 (Mayo clinic 2001-2014)	unclear	unclear	low
		Bahlo 2016 (SCAN cohort)	unclear	low	low
		Delgado 2017 (Barcelona hospital cohort)	unclear	low	low
		Gentile 2016 (Italian multicentre cohort)	unclear	unclear	low
		Muñoz-Novaz 2018 (Spain multicentre cohort)	low	unclear	low
		Rani 2018 (Indian cohort)	unclear	unclear	low
		Zhu 2018 (Chinese cohort)	unclear	low	low
	Narrative	Da Cunha-Bang 2016 (Danish cohort)	low	unclear	low
		Molica 2016 (O-CLL1-GISL)	NA	NA	high
		Reda 2017 (Milan cohort)	unclear	unclear	low
MDACC 2007 model (overall survival)	Development	Rigolin 2017 (Ferrera cohort)	unclear	low	low
	Meta-analysed c-statistic for the index score	Wierda 2007 (Development cohort)	low	low	low
		Bulian (Italian-Swiss cohort)	low	low	low
		Gentile 2014 (Italian multicentre cohort)	unclear	low	low
		Gentile 2016 (Mayo cohort 2001-2008)	unclear	low	low
		González Rodríguez 2009 (Cabueñes hospital cohort)	low	low	low
		Muñoz-Novaz 2018 (Spain multicentre cohort)	low	low	low
		Rani 2018 (Indian cohort)	unclear	unclear	low
	Narrative	Molica 2010 (GIMEMA cohort 1991-2000)	low	high	high
		Molica 2015 (O-CLL1-GISL)	NA	NA	high
		Pflug 2014 (German CLL study group cohort)	unclear	low	low
GCLLSG model (overall survival)	Development	Trajkova 2013 (Macedonian cohort)	low	low	low
		Pflug 2014 (Development cohort)	unclear	low	low
	Narrative	Molica 2015 (O-CLL1-GISL)	NA	NA	high
		Pflug 2014 (Mayo cohort)	unclear	low	low
GIMEMA model (progression-free survival)	Development	Rani 2018 (Indian cohort)	NA	NA	high
	Narrative				
MDACC 2011 model (time-to-first-treatment)	Development	Molica 2005 (Development cohort)	low	low	low
	Narrative	González Rodríguez 2009 (Cabueñes hospital cohort)	low	low	low
Morabito model (time-to-first-treatment)	Development	Wierda 2011 (Development cohort)	unclear	low	low
	Narrative	Molica 2013 (O-CLL1-GISL)	low	low	low
O-CLL1 model (time-to-first-treatment)	Development	Molica 2009 (Development cohort)	unclear	high	low
	Narrative	Gentile 2014 (O-CLL1-GISL)	low	low	low
Rossi model (overall survival)	Development	Gentile 2016 (O-CLL1-GISL)	low	low	low
	Narrative	Rani 2018 (Indian cohort)	unclear	unclear	low
		Gentile 2016 (Mayo cohort 2001-2008)	unclear	low	low
Stephens model (overall survival)	Development	Rossi 2013 (Development cohort)	low	low	low
	Narrative	Jeromin 2014 (Munich cohort)	low	low	low
		Rossi 2013 (External validation cohort)	low	low	low
Stephens model (treatment-free survival)	Development	Stephens 2015 (Development cohort)	high	low	low
	Narrative	Stephens 2015 (MDACC cohort)	high	low	low
Stephens model (treatment-free survival)	Development	Stephens 2015 (Development cohort)	high	low	low
	Narrative	Stephens 2015 (MDACC cohort)	high	low	low

Barcelona-Brno model (Barcelona-Brno D - Delgado 2017 (Barcelona cohort))

We rated the risk of bias of the development study as unclear for the domain 'participants' due to missing information on eligibility criteria (Barcelona-Brno D - Delgado 2017 (Barcelona cohort)), low for the domains 'predictors' and 'outcome', and high for the domain

'analysis' due to the model-building process (factor selection and weighting). We rated the risk of bias between the validation studies as unclear for the domain 'participants' in three studies (Barcelona-Brno V - Delgado 2017 (Brno cohort); Barcelona-Brno V - Rani 2018 (Indian cohort); Barcelona-Brno V - Reda 2017 (Milan cohort)) due to missing eligibility criteria and recruitment period, and low

in two studies ([Barcelona-Brno V - Gentile 2017 \(Italian & Mayo\)](#); [Barcelona-Brno V - Muñoz-Novas 2018 \(Spanish coh.\)](#)). We rated the risk of bias for the domains 'predictors' and 'outcome' as low for all validation studies except one ([Barcelona-Brno V - Rani 2018 \(Indian cohort\)](#)), which we rated as unclear for the domain 'outcome' due to the short observation time. Concerning the domain 'analysis',

we considered all studies to be at high risk of bias (especially due to handling missing data, low number of events and reporting of performance measures) except one ([Barcelona-Brno V - Delgado 2017 \(Brno cohort\)](#)), which we rated as unclear. One validation study did not examine the outcome of interest ([Barcelona-Brno V - Molica 2017 \(O-CLL1-GISL\)](#)) ([Figure 16](#)).

Figure 16. Risk of bias (PROBAST) assessment of the Barcelona-Brno model (Delgado 2017)

Model	Study	Prediction model Risk Of Bias ASessment Tool (PROBAST)				
		Domain 1: Participants	Domain 2: Predictors	Domain 3: Outcome	Domain 4: Analysis	Overall judgement
Barcelona-Brno model (overall survival)	Development	Delgado 2017 (Development cohort)	unclear	low	low	high
	Meta-analysed c-statistic	Delgado 2017 (Brno cohort, CZ)	unclear	low	low	unclear
		Gentile 2017 (Italian multicentre and Mayo cohort)	low	low	low	high
		Muñoz-Novas 2018 (Spanish cohort)	low	low	low	high
		Rani 2018 (Indian cohort)	unclear	low	unclear	high
	Narrative	Molica 2017 (O-CLL1-GISL)	NA	NA	NA	NA
		Reda 2017 (Milan cohort)	unclear	low	low	high

We rated the concern for applicability of the development study as unclear for the domain 'participants' and low for the domains 'predictors' and 'outcome' ([Barcelona-Brno D - Delgado 2017 \(Barcelona cohort\)](#)). Similarly, we rated concern for applicability as unclear for the domain 'participants', and low for the domains 'predictors' and 'outcome' in three validation studies ([Barcelona-Brno V - Delgado 2017 \(Brno cohort\)](#); [Barcelona-Brno V - Gentile 2017 \(Italian & Mayo\)](#); [Barcelona-Brno V - Muñoz-Novas 2018 \(Spanish coh.\)](#)) due to undefined eligibility criteria. We assigned two validation studies ([Barcelona-Brno V - Rani 2018 \(Indian cohort\)](#); [Barcelona-Brno V - Reda 2017 \(Milan cohort\)](#)) the rating unclear for the domain 'predictors' due to lack of information on the timing of predictor assessment. One validation study did not examine the outcome of interest ([Barcelona-Brno V - Molica 2017 \(O-CLL1-GISL\)](#)) ([Figure 15](#)).

MDACC 2007 (Wierda 2007) (MDACC 2007 D - Wierda 2007 (MDACC))

We rated the risk of bias for the development study as low across all domains ([MDACC 2007 D - Wierda 2007 \(MDACC\)](#)). We rated the

risk of bias across the validation studies as low for the domain 'participants' in six validation studies and high in three validation studies ([MDACC 2007 V - Gentile 2016 \(Mayo cohort\)](#); [MDACC 2007 V - Rani 2018 \(Indian cohort\)](#); [MDACC 2007 V - Trajkova 2013 \(Macedonia\)](#)) due to unclear eligibility criteria. We considered all validation studies to have a low risk of bias for the domains 'predictors'. We rated three studies as unclear for the domain 'outcome' ([MDACC 2007 V - Molica 2010 \(GIMEMA cohort\)](#); [MDACC 2007 V - Rani 2018 \(Indian cohort\)](#); [MDACC 2007 V - Trajkova 2013 \(Macedonia\)](#)) because of the short observation time. Concerning the domain 'analysis', we considered all validation studies to have a high or unclear risk of bias due to the low number of events and inappropriate handling of missing data. One validation study did not examine the outcome of interest ([MDACC 2007 V - Molica 2015 \(O-CLL1-GISL\)](#)) ([Figure 17](#)).

Figure 17. Risk of bias (PROBAST) assessment of the MDACC 2007 model (Wierda 2007)

Model	Study	Prediction model Risk Of Bias ASessment Tool (PROBAST)				
		Domain 1: Participants	Domain 2: Predictors	Domain 3: Outcome	Domain 4: Analysis	Overall judgement
MDACC 2007 model (overall survival)	Development	Wierda 2007 (Development cohort)	low	low	low	low
	Meta-analysed c-statistic for the index score	Bulian (Italian-Swiss cohort)	low	low	low	high
		Gentile 2014 (Italian multicentre cohort)	low	low	low	high
		Gentile 2016 (Mayo cohort 2001-2008)	high	low	low	high
		González Rodríguez 2009 (Cabeñes hospital cohort)	low	low	low	high
		Muñoz-Novas 2018 (Spain multicentre cohort)	low	low	low	high
		Pflug 2014 (German CLL study group cohort)	low	low	low	high
		Rani 2018 (Indian cohort)	high	low	unclear	high
		Molica 2010 (GIMEMA cohort 1991-2000)	low	low	unclear	high
		Molica 2015 (O-CLL1-GISL)	NA	NA	NA	NA
	Narrative	Trajkova 2013 (Macedonian cohort)	high	low	unclear	high

We rated the concern for applicability of the development study as low across all domains ([MDACC 2007 D - Wierda 2007 \(MDACC\)](#)). We rated concern for applicability as low for the domain 'participants' in five validation studies and unclear in four validation studies due to unclear eligibility criteria ([MDACC 2007 V - Gentile 2014 \(Italian cohort\)](#); [MDACC 2007 V - Gentile 2016 \(Mayo cohort\)](#); [MDACC 2007 V - Pflug 2014 \(3 RCTs\)](#); [MDACC 2007 V - Rani 2018 \(Indian cohort\)](#)). Concerning the domain 'predictors', we considered the concern for applicability to be low in seven validation studies, high in one study ([MDACC 2007 V - Molica 2010 \(GIMEMA cohort\)](#)), and unclear in one study ([MDACC 2007 V - Rani 2018 \(Indian cohort\)](#)). Concern

for applicability was rated as low for the domain 'outcome' in all validation studies except one, for which we assigned a high rating ([MDACC 2007 V - Molica 2010 \(GIMEMA cohort\)](#)). One validation study did not examine the outcome of interest ([MDACC 2007 V - Molica 2015 \(O-CLL1-GISL\)](#)) ([Figure 15](#)).

Prognostic models for the prediction of PFS or TFS (with meta-analysis, > 3 external validations)

We did not identify any prognostic model development studies that were externally validated more than three times for the outcomes PFS, TFS or TFS.

Models with one to three external validation studies per outcome

Prognostic models for the prediction of OS (without meta-analysis, 1 to 3 external validations)

GCLLSG model (Pflug 2014) (GCLLSG D - Pflug 2014 (GCLLSG))

We rated the risk of bias for the development study as low for all domains except 'analysis' (GCLLSG D - Pflug 2014 (GCLLSG)). We rated 'analysis' at high risk of bias due to dichotomisation and

univariable selection of factors, inappropriate handling of missing data and simplification of the model (loss of original weighting). We considered the only validation study with a matching outcome to have a high risk of bias for the domains 'participants' and 'analysis' due to inclusion of participants with available data and low number of events (GCLLSG V - Pflug 2014 (Mayo cohort)). Risk of bias was low for the domains 'predictors' and 'outcome'. The other two validation studies did not examine the outcome of interest (GCLLSG V - Molica 2015 (O-CLL1-GISL); GCLLSG V - Rani 2018 (Indian cohort)) (Figure 18).

Figure 18. Risk of bias (PROBAST) assessment of other models that were externally validated

Model	Study	Prediction model Risk Of Bias Assessment Tool (PROBAST)				
		Domain 1: Participants	Domain 2: Predictors	Domain 3: Outcome	Domain 4: Analysis	Overall judgement
Baliakas model (time-to-first-treatment)	Development	Baliakas 2019 (Development cohort)	high	low	low	high
	Narrative	Baliakas 2019 (MLL + Scan. cohort)	unclear	unclear	unclear	high
GCLLSG model (overall survival)	Development	Pflug 2014 (Development cohort)	low	low	low	high
	Narrative	Molica 2015 (O-CLL1-GISL)	NA	NA	NA	NA
		Pflug 2014 (Mayo cohort)	high	low	low	high
		Rani 2018 (Indian cohort)	NA	NA	NA	NA
GIMEMA model (progression-free survival)	Development	Molica 2005 (Development cohort)	low	low	high	high
	Narrative	González Rodríguez 2009 (Cabueñes hospital cohort)	low	low	high	high
MDACC 2011 model (time-to-first-treatment)	Development	Wierda 2011 (Development cohort)	high	low	high	high
	Narrative	Molica 2013 (O-CLL1-GISL)	low	low	high	high
Morabito model (time-to-first-treatment)	Development	Morabito 2009 (Development cohort)	unclear	low	low	high
	Narrative	Gentile 2014 (O-CLL1-GISL)	low	low	low	unclear
O-CLL1 model (time-to-first-treatment)	Development	Gentile 2016 (O-CLL1-GISL)	low	low	high	high
	Narrative	Rani 2018 (Indian cohort)	unclear	low	high	high
		Gentile 2016 (Mayo cohort 2001-2008)	high	low	high	unclear
Rossi model (overall survival)	Development	Rossi 2013 (Development cohort)	low	low	low	unclear
	Narrative	Jeromin 2014 (Munich cohort)	unclear	low	low	high
		Rossi 2013 (External validation cohort)	unclear	low	low	high
Stephens model (overall survival)	Development	Stephens 2015 (Development cohort)	high	low	low	high
	Narrative	Stephens 2015 (MDACC cohort)	high	unclear	low	high
Stephens model (treatment-free survival)	Development	Stephens 2015 (Development cohort)	high	low	high	high
	Narrative	Stephens 2015 (MDACC cohort)	high	unclear	high	high

We rated the concern for applicability of the development study as unclear for the domain 'participants' because the sample was based on RCT data that may not be representative for all individuals with CLL (GCLLSG D - Pflug 2014 (GCLLSG)). Concern was low for the domains 'predictors' and 'outcome'. Similarly, we rated concern for applicability of the validation study as unclear for the domain 'participants' due to eligibility criteria and low for the domains 'predictors' and 'outcome' (GCLLSG V - Pflug 2014 (Mayo cohort)). The other two validation studies did not examine the outcome of interest (GCLLSG V - Molica 2015 (O-CLL1-GISL); GCLLSG V - Rani 2018 (Indian cohort)) (Figure 15).

Rossi (Rossi 2013) (Rossi D - Rossi 2013 (Italian cohort))

We rated the risk of bias of the development study as low across all domains except 'analysis' (Rossi D - Rossi 2013 (Italian cohort)), which we considered to be of unclear risk because handling of missing data handling was not reported. We rated the risk of bias for the two validation studies as unclear due to lack of information on study design for the domain 'participants', low for the domains 'predictors' and 'outcome', and high for the domain 'analysis' due to missing data handling and low number of events (Rossi V - Jeromin 2014 (Munich cohort); Rossi V - Rossi 2013 (unclear)) (Figure 18).

We rated the concern for applicability of the development study as low across all domains (Rossi D - Rossi 2013 (Italian cohort)). We rated concern for applicability of the two validation studies as low for the domains 'participants', 'predictors' and 'outcome' (Rossi V - Jeromin 2014 (Munich cohort); Rossi V - Rossi 2013 (unclear)) (Figure 15).

Stephens model (Stephens 2015) (Stephens OS D - Stephens 2015 (Ohio cohort))

We rated the risk of bias for the development study as high for the domains 'participants' (Stephens OS D - Stephens 2015 (Ohio cohort)), because the study included only individuals with del(17p), which implies that only persons with available FISH assessment could have been included. We considered the domains 'predictors' and 'outcomes' to be of low risk. We rated the domain 'analysis' as high due to small sample size, inappropriate missing data handling and inconsistent reporting of performance measures. In a similar manner, we rated the risk of bias for the validation study as high for the domains 'participants' and 'analysis' due to the same reasons, unclear for the domain 'predictors' and low for the domain 'outcome' (Stephens OS V - Stephens 2015 (MDACC)) (Figure 18).

We rated the concern for applicability of the development study (Stephens OS D - Stephens 2015 (Ohio cohort)) and validation study (Stephens OS V - Stephens 2015 (MDACC)) as high for the domain 'participants' due to the selective sample, and low for the domains 'predictors' and 'outcome' (Figure 15).

Prognostic models for the prediction of PFS or TFS (without meta-analysis, 1 to 3 external validations)

Baliakas model (Baliakas 2019) (Baliakas D - Baliakas 2019 (multicentre))

We rated the risk of bias of the development study as low for the domains 'predictors' and 'outcome', and high for the domains 'participants' since participants with missing data were excluded (Baliakas D - Baliakas 2019 (multicentre)). We rated the

domain 'analysis' as high due to inappropriate handling of missing data, and univariable selection of predictors. Due to reporting deficiencies concerning inclusion criteria, predictor assessment and outcome definition of the validation study (Baliakas V - Baliakas 2019 (MLL + Scan.)), we rated the risk of bias as unclear across all domains except 'analysis', which we considered to be of high risk due to lack of reporting performance measures (Figure 18).

We rated the concern for applicability of the development study as high for the domain 'participants' due to the sample not being representative of all CLL patients and low for the domains 'predictors' and 'outcome' (Baliakas D - Baliakas 2019 (multicentre)). We rated as unclear concern for applicability of the validation study for the domain 'participants' due to lack of information about eligibility criteria and recruitment period (Baliakas V - Baliakas 2019 (MLL + Scan.)), and low for the domains 'predictors' and 'outcome' (Figure 15).

GIMEMA model (Molica 2005) (GIMEMA D - Molica 2005 (GIMEMA cohort))

We rated the risk of bias for the development study as low for the domains 'participants' and 'predictors' (GIMEMA D - Molica 2005 (GIMEMA cohort)), and high for the domains 'outcome' because a predictor was not excluded from the outcome definition and 'analysis' due to inappropriate handling of missing data, univariable selection of predictors, and simplification of the model. Likewise, we considered the validation study to have a low risk of bias for the domains 'participants' and 'predictors' (GIMEMA V - González Rodríguez 2009 (Cabueñes coh.)), while we rated the domains 'outcome' and 'analysis' as high risk of bias due to lack of information on performance measures (Figure 18).

We rated the concern for applicability of the development study (GIMEMA D - Molica 2005 (GIMEMA cohort)) and validation study (GIMEMA V - González Rodríguez 2009 (Cabueñes coh.)) as low across all domains (Figure 15).

MDACC 2011 (Wierda 2011) (MDACC 2011 D - Wierda 2011 (MDACC))

We rated the risk of bias for the development study as high for the domain 'participants' due to inappropriate inclusion criteria (MDACC 2011 D - Wierda 2011 (MDACC)), for the 'outcome' because outcome definition did not exclude one of the predictors and for 'analysis' because of univariable selection of predictors, unclear modelling procedure and lack of performance measures. We only gave a low rating to the domain 'predictors'. Similarly, we considered the risk of bias of the validation study to be high for the domains 'outcome' for the same reason and 'analysis' due to unclear handling of missing data and low number of events (MDACC 2011 V - Molica 2016 (O-CLL1-GISL)), while we rated the domains 'predictors' and 'participants' as low risk (Figure 18).

We rated the concern for applicability of the development study as unclear for the domain 'participants' due to inappropriate inclusion criteria and low for the domains 'predictors' and 'outcome' (MDACC 2011 D - Wierda 2011 (MDACC)). We rated concern for applicability of the validation study as low across all domains (MDACC 2011 V - Molica 2016 (O-CLL1-GISL)) (Figure 15).

Morabito model (Morabito 2009) (Morabito D - Morabito 2009 (Italian cohort))

We rated the risk of bias of the development study as unclear for the domain 'participants' because it was not clear if individuals

with missing data were considered (Morabito D - Morabito 2009 (Italian cohort)), low for the domains 'predictors' and 'outcome', and high for the domain 'analysis' due to low number of events, dichotomisation and univariable selection of factors and missing performance measures. In the case of the validation study (Morabito V - Gentile 2014 (O-CLL1-GISL)), we rated the risk of bias as low across all domains except 'analysis', which we considered to be of unclear risk due to lack of reporting performance measures (Figure 18).

We rated the concern for applicability of the development study as unclear for the domain 'participants' due to unclear eligibility criteria, high for the domain 'predictors' and low for the domain 'outcome' (Morabito D - Morabito 2009 (Italian cohort)). We rated as low, concern for applicability of the validation study across all domains (Morabito V - Gentile 2014 (O-CLL1-GISL)) (Figure 15).

O-CLL1 model (Gentile 2016) (O-CLL-1 D - Gentile 2016 (O-CLL-1-GISL))

We rated the risk of bias of the development study as low across the domains 'participants' and 'predictors' (O-CLL-1 D - Gentile 2016 (O-CLL-1-GISL)). We rated the domain 'outcome' as high risk of bias because one predictor was also included in the outcome definition. We rated the domain 'analysis' as high risk of bias due to unclear handling of missing data, categorisation and univariable selection of factors.

Concerning the domain 'participants', we rated one validation study as unclear due to missing eligibility criteria (O-CLL-1 V - Rani 2018 (Indian cohort)), and one as high risk due to inappropriate inclusion criteria (O-CLL1 V - Gentile 2016 (Mayo cohort)). We rated both validation studies as low risk of bias for the domain 'predictors' and high risk for the domain 'outcome' because one predictor was also included in the outcome definition. We considered the risk of bias for the domain 'analysis' to be high for one validation study due to the low number of events and no report for handling of missing data (O-CLL-1 V - Rani 2018 (Indian cohort)), while we rated the other study as unclear due to unclear handling of missing data (O-CLL1 V - Gentile 2016 (Mayo cohort)) (Figure 18).

We rated the concern for applicability of the development study as low across all domains (O-CLL-1 D - Gentile 2016 (O-CLL-1-GISL)). We rated concern for applicability of the two validation studies (O-CLL1 V - Gentile 2016 (Mayo cohort); O-CLL-1 V - Rani 2018 (Indian cohort)) as unclear for the domains 'participants' due to undefined eligibility criteria, the 'predictors' as low for the first and unclear for the second due to undefined timing of prognostic factor measurement, and low for both studies for the domain 'outcome' (Figure 15).

Stephens model (Stephens 2015) (Stephens TFS D - Stephens 2015 (Ohio cohort))

For risk of bias of the model for TFS in Stephens and colleagues (Stephens TFS D - Stephens 2015 (Ohio cohort)), please see description in the section for OS (above), as the authors used the same development and validation cohorts for both outcomes.

Deficiencies of prognostic model development studies (models with PROBAST rating)

Among the 12 prognostic model development studies with at least one external validation study (including the three meta-analysed models), we rated nine at high or unclear risk of bias for the domain

'analysis'. We identified the following most common issues in these studies regarding model development.

- *Missing data handling*: only one model development study imputed missing values; all other studies based their analysis on complete-case analyses.
- *Predictor selection based on univariable analysis*: eight studies derived their multivariable model on prior univariable selection using P values as a criterion.
- *Correction of estimates*: several development studies used split-sample or bootstrapping techniques. However, none of the studies reported a correction of their regression coefficients to reduce overfitting to the development cohort.
- *Categorisation of factors*: at some stage during the development process, all studies reported categorisation or dichotomisation of continuous prognostic factors.
- *Model weights*: 10 development studies presented a simplification of the original model formula in the form of a point score (weighting, grouping, or counting of disadvantageous prognostic factors), which results in loss of information.

Findings

Reporting of calibration was rare in the studies that we identified. If reported or provided afterwards via email, the format varied so that we were unable to summarise the 'traditional' calibration measures (calibration plots, calibration tables, O:E ratios). Therefore, we

decided to use a different way of presenting calibration graphically (Figure 19; Figure 20; Figure 21), based on something that most studies reported (survival frequencies per group). The black lines represent the development study and each coloured line one external validation cohort. The survival per risk group of the development cohort can be interpreted as the expected survival frequencies, and the survival frequencies in the external validation cohorts the observed survival frequencies. The pooled result gives the pooled observed survival frequencies.

Models with more than three external validation studies per outcome and meta-analysis

Prognostic models for the prediction of OS (with meta-analysis, > 3 external validations)

CLL-IPI (CLL-IPI D - Bahlo 2016 (development cohort))

The CLL-IPI is a prognostic score derived from univariable selection of predictors which were then entered into a forward-stepwise proportional-hazards Cox regression model with a hierarchy based on completeness of the predictor information (CLL-IPI D - Bahlo 2016 (development cohort)). The dataset was split into a training and external validation dataset after univariable selection of factors (66% and 33% of the data, respectively). Continuous factors were dichotomised based on published thresholds and quartiles. The authors integrated prognostic factors that were independently associated with the outcome in the final multivariable model in a weighted manner to construct their prognostic index (Table 2).

Table 2. Scoring of the CLL-IPI (Bahlo 2016)

Prognostic factor	Point distribution
Age (≥ 65 years)	1
Clinical stage (Rai I-IV or Binet B-C)	1
IgHV mutational status (unmutated)	2
B2-microglobulin (> 3.5 mg/L)	2
TP53 status (deleted or mutated)	4

Notes: IgHV: immunoglobulin heavy chain variable region genes

Calibration

None of the included studies regarding the CLL-IPI reported a measure of calibration in their publication.

We included eight external validation studies of this model, with a total of 4891 individuals to calculate the pooled survival frequencies per risk group at five years (Figure 19). The pooled observed survival frequencies of all validation studies were 92.5% (89.2% to 94.8%) for the low-risk group (score 0-1), 85.0% (79.7% to 89.1%) for the intermediate-risk group (score 2-3), 64.9% (56.4% to 72.6%) for the high-risk group (score 4-6) and 40.4% (29.3%

to 52.6%) for the very high-risk group (score 7-10). For the low, intermediate and high-risk group, the pooled result of the external validation studies approximated the survival frequencies of the model development study. In the very high-risk group, survival according to the development study would have been lower than overall in the external validation studies. For the low and intermediate-risk groups, the 95% PIs were relatively small. The other risk groups showed wide 95% PIs, indicating uncertainty (Table 3), probably due to the low number of individuals in these subgroups (N = 765 and N = 201, respectively). We judged heterogeneity (visual inspection) between external validation cohorts to be low for the calibration of the CLL-IPI.

Table 3. Survival per CLL-IPI risk group

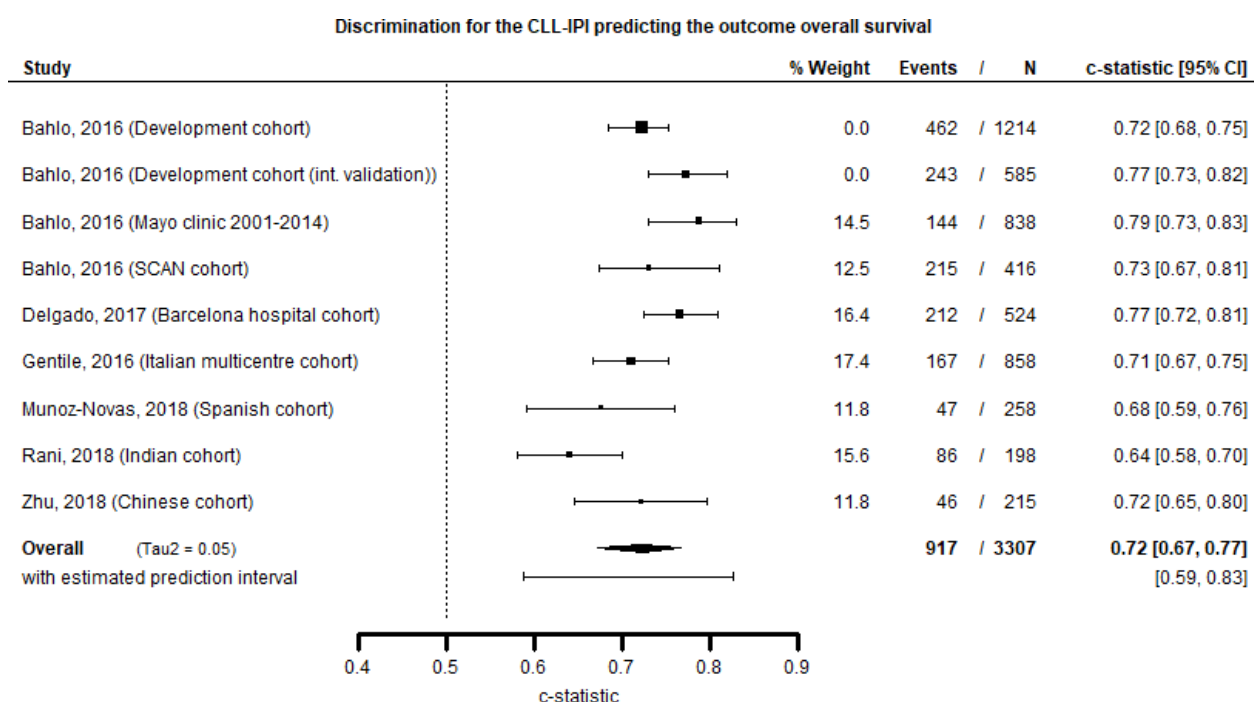
Risk group	Risk score	Development study (training dataset): percentage of persons surviving at 5 years (95% CI)	Pooled percentage of persons surviving at 5 years (95% CI)	95% prediction interval (PI)
Low risk	0-1	93.2% (90.5% to 96.0%)	92.5% (89.2% to 94.8%)	82.5% to 97.0%
Intermediate risk	2-3	79.3% (75.5% to 83.2%)	85.0% (79.7% to 89.1%)	68.8% to 93.6%
High risk	4-6	63.3% (57.9% to 68.8%)	64.9% (56.4% to 72.6%)	41.9% to 82.6%
Very high risk	7-10	23.3% (12.5% to 34.1%)	40.4% (29.3% to 52.6%)	13.2% to 72.4%

Discrimination

We included seven external validation studies with a total of 3307 individuals and 917 deaths observed during the overall observation time. Two studies that presented a measure for the c-statistic did not report a measure of uncertainty for this measure, which we estimated using the Newcombe method (CLL-IPI V - Gentile 2016

(Italian cohort); CLL-IPI V - Muñoz-Novas 2018 (Spanish cohort)). For one study (CLL-IPI V - Zhu 2018 (Chinese cohort)), we included the AUC as a measure of discrimination instead of the c-statistic. The pooled c-statistic of the seven studies was 0.72 (95% CI 0.67 to 0.77), with some degree of heterogeneity (between-study standard deviation, $\tau = 0.21$). The 95% PI ranged from 0.59 to 0.83 (Figure 22).

Figure 22. Meta-analysis of the c-statistic for the CLL-IPI model (Bahlo 2016)



The c-statistic of the model development study was 0.72 (95% CI 0.68 to 0.75). Based on sensitivity analysis omitting the study that used the AUC instead of the c-statistic (Figure 23), and comparing estimated and reported 95% CI of the c-statistic (Figure 24; Figure

25), we conclude that this had no effect on the result. Four out of seven external validation studies replaced the prognostic factor TP53 with a proxy, del(17p); sensitivity analysis showed no substantial differences in discriminative performance (Figure 26).

Figure 23. Sensitivity analysis of the c-statistic for the CLL-IPI excluding the study reporting AUC

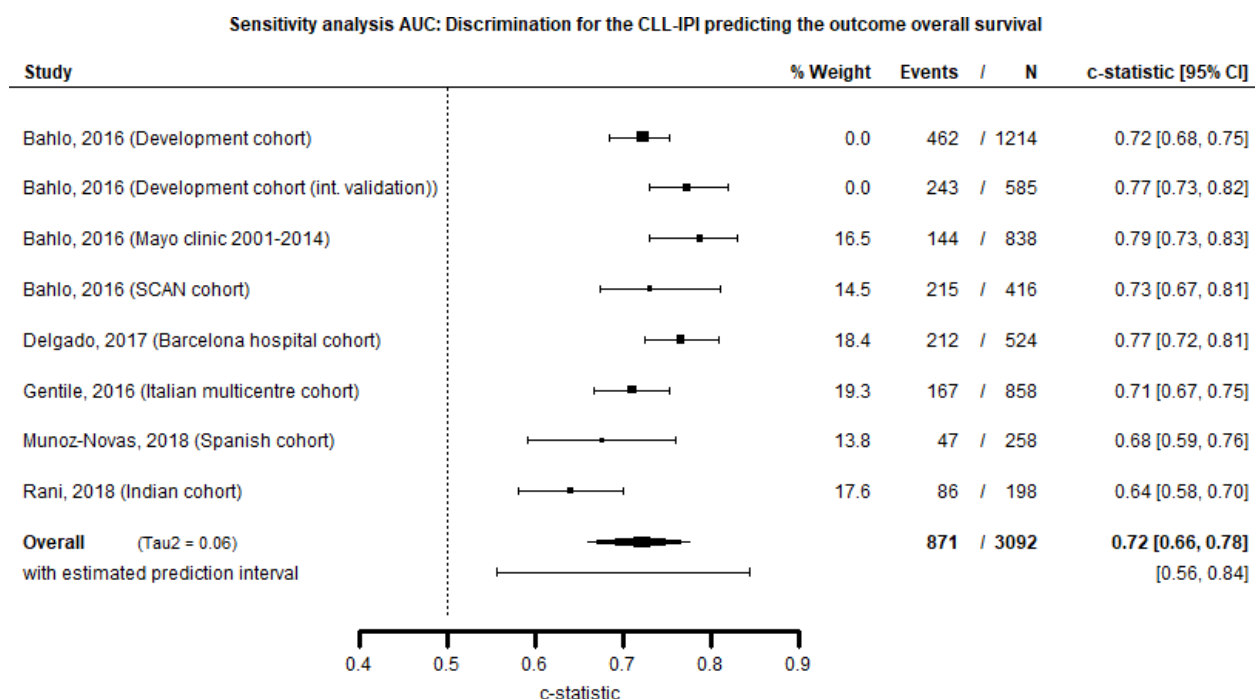


Figure 24. Sensitivity analysis of the c-statistic for the CLL-IPI excluding the study reporting no 95% CI

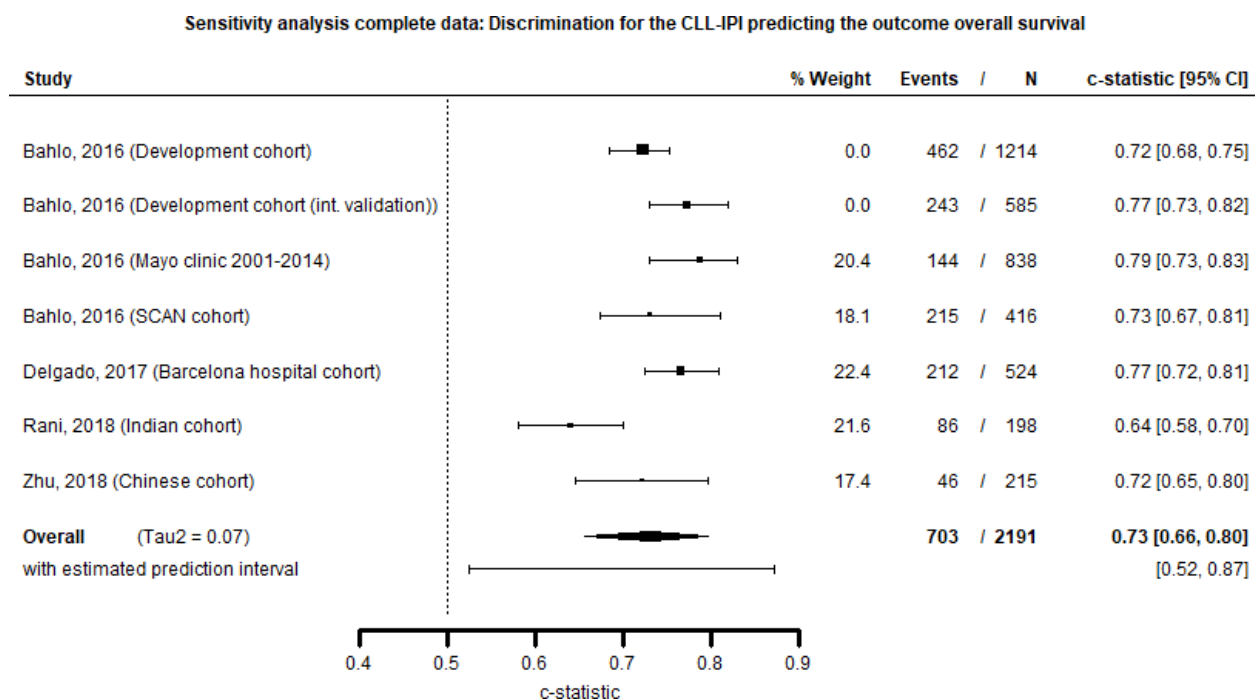


Figure 25. Sensitivity analysis of the c-statistic for the CLL-IPI excluding the study reporting no 95% CI. All 95% CIs were replaced by estimates using the Newcombe method for a comparison with Figure 27

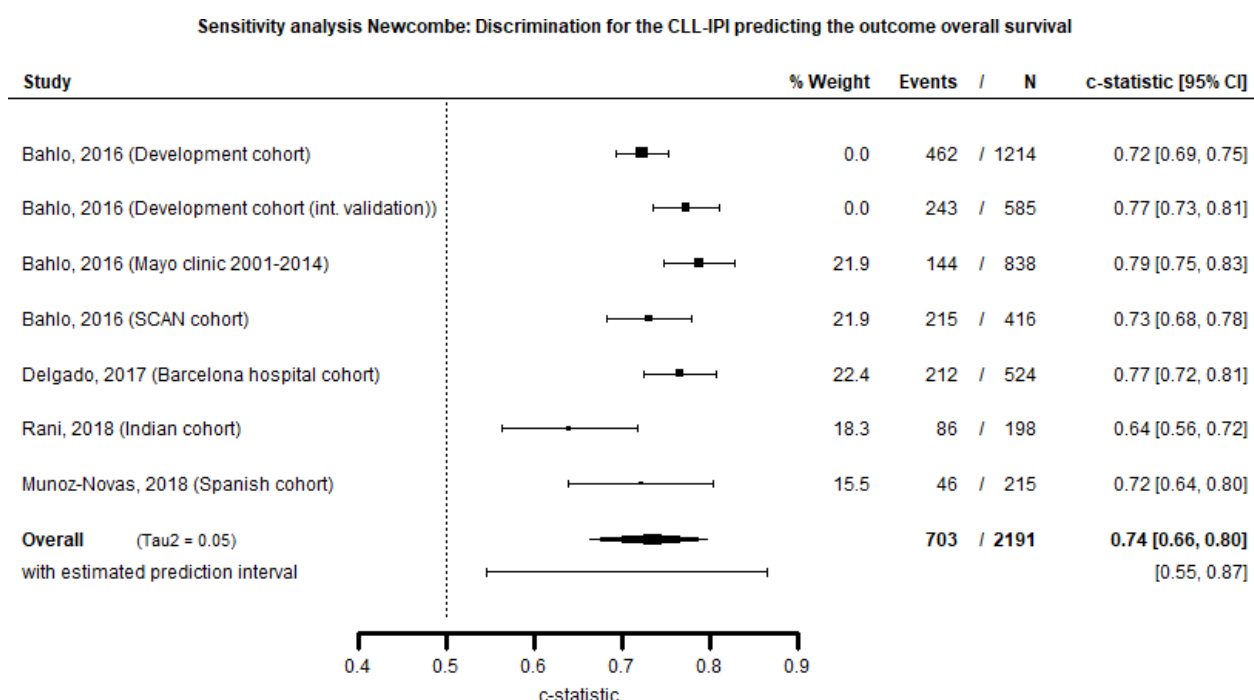
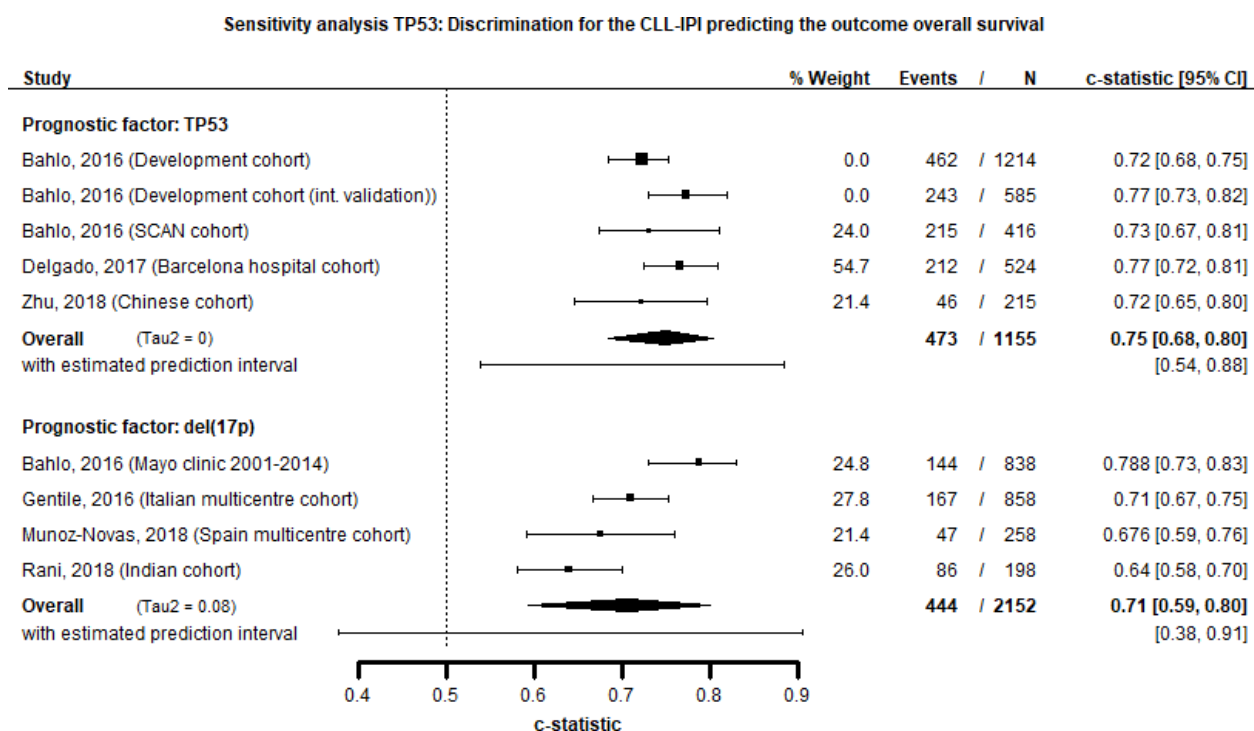


Figure 26. Sensitivity analysis of the c-statistic for the CLL-IPI regarding the availability of the predictor TP53 and its proxy del(17p)



Barcelona-Brno model (Barcelona-Brno D - Delgado 2017 (Barcelona cohort))

The Barcelona-Brno score is a prognostic score constructed by comparing the best combinations of the five factors included in another model for CLL ([Barcelona-Brno D - Delgado 2017 \(Barcelona cohort\)](#)), the CLL-IPI by Bahlo 2016 [CLL-IPI D - Bahlo 2016 \(development cohort\)](#)). Incomplete data were most likely part of the exclusion criteria, and therefore persons with missing data were excluded from the analysis in the primary publication. The score classified individuals into three risk groups: persons without del(11q) or del(17p) and mutated IgHV status were classified as low risk; persons with del(11q) or del(17p) and unmutated IgHV status were classified as high risk; all other persons were classified as intermediate risk.

Calibration

None of the included studies in reference to the Barcelona-Brno model reported a measure of calibration in their publication.

We included three external validation studies of this score with a total of 1974 individuals to calculate the pooled survival frequencies per risk group at five years ([Figure 20](#)). The pooled observed survival frequencies of all validation studies were 90.5% (95% CI 85.1% to 94.0%) for the low-risk group, 79.7% (95% CI 70.7% to 86.5%) for the intermediate-risk group and 62.5% (95% CI 49.3% to 74.1%) for the high-risk group. For the low and intermediate- risk group, the pooled result of the external validation studies approximated the survival frequencies of the model development study. In the high-risk group, survival according to the development study would have been higher than overall in the external validation studies. For the low-risk group, the 95% PI was relatively small. However, the intermediate and high-risk groups showed a wide 95% PI, indicating uncertainty ([Table 4](#)). We judged heterogeneity (visual inspection) between external validation cohorts to be low for the calibration of the Barcelona-Brno score.

Table 4. Survival per Barcelona-Brno model risk group

Risk group	Risk group	Development study: percentage of persons surviving at 5 years (95% CI)	Pooled percentage of persons surviving at 5 years (95% CI)	95% prediction interval (PI)
Low risk	No del(11q) or del(17p) and mutated IgHV status	93.4% (90.4% to 96.5%)	90.5% (85.1% to 94.0%)	80.4% to 95.7%
Intermediate risk	All other	83.8% (78.4% to 89.6%)	79.7% (70.7% to 86.5%)	63.1% to 90.0%
High risk	Del(11q) or del(17p) and unmutated IgHV status	70.0% (59.3% to 82.7%)	62.5% (49.3% to 74.1%)	41.3% to 79.7%

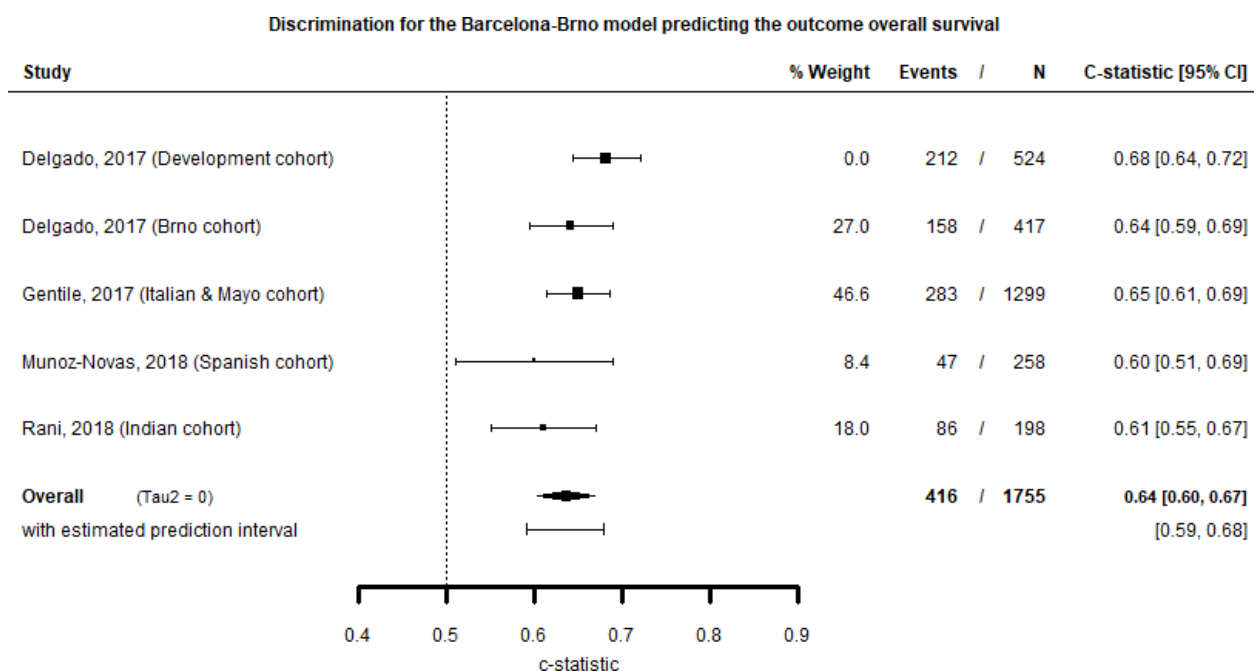
Notes: IgHV: immunoglobulin heavy chain variable region genes

Discrimination

For our meta-analysis of the c-statistic, we included four external validation studies with a total of 1755 individuals and 416 deaths observed during the overall observation time. Two studies that presented a measure for the c-statistic did not report a measure

of uncertainty, which we estimated using the Newcombe method ([Barcelona-Brno V - Gentile 2017 \(Italian & Mayo\)](#); [Barcelona-Brno V - Muñoz-Novas 2018 \(Spanish coh.\)](#)). The pooled c-statistic of the four studies was 0.64 (95% CI 0.60 to 0.67), with no heterogeneity between studies ($\tau = 0.00$). The 95% PI ranged from 0.59 to 0.68 ([Figure 27](#)).

Figure 27. Meta-analysis of the c-statistic for the Barcelona-Brno model (Delgado 2017)



The c-statistic of the model development study was 0.68 (95% CI 0.64 to 0.72); it was unclear whether bootstrapping techniques were used to obtain this estimate.

MDACC 2007 (Wierda 2007) (MDACC 2007 D - Wierda 2007 (MDACC))

The MDACC 2007 model is a prognostic model based on univariable selection of predictors that were, when significant, entered into

a proportional-hazards Cox regression model (MDACC 2007 D - Wierda 2007 (MDACC)). Bootstrapping was used to account for optimism in the c-statistic. The authors provided the precise formula to calculate a point score which could be translated graphically by a nomogram into individual survival probabilities at five or 10 years. In addition, for simplified use in clinical practice, the authors created an index score in a weighted manner (Table 5).

Table 5. Scoring of the MDACC 2007 model

	Point distribution			
Prognostic factor	0	1	2	3
Age (years)		< 50	50-65	> 65
β-2 microglobulin (mg/L)	< ULN	1-2 × ULN	> 2 × ULN	
ALC (× 10 ⁹ /L)	< 20	20-50	> 50	
Gender	Female	Male		
Rai stage	0-II	III-IV		
No. of involved nodal groups	≤ 2	3		
Notes: ULN: upper limit of normal; ALC: acute lymphocyte count				

Calibration

The development study provided a calibration plot based on bootstrap sampling. The data points representing the observed versus expected survival probabilities are close to the ideal line. At the higher frequencies, the nomogram slightly underestimates survival. Measures of calibration were not reported in any of the external validation studies.

We included five external validation studies of this model with a total of 3786 individuals for calculating the pooled survival frequencies per risk group at 5 years (graphically represented in [Figure 21](#)). The pooled observed survival frequencies of all

validation studies was 97.0% (95% CI 94.3% to 98.4%) for the low-risk group (score 1-3), 82.3% (74.6% to 88.0%) for the intermediate-risk group (score 4-7) and 45.6% (31.3% to 60.5%) for the high-risk group (score ≥ 8). For the low and intermediate-risk group, the pooled result of the external validation studies approximated the survival frequencies of the model development study. In the high-risk group, survival according to the development study would have been higher than overall in the external validation studies. For the low-risk group, the 95% PI was relatively small. The intermediate and high-risk group showed wide 95% PIs, indicating uncertainty ([Table 6](#)). We judged heterogeneity (visual inspection) of calibration between external validation cohorts to be low for the MDACC 2007 index score.

Table 6. Survival per MDACC 2007 model risk group

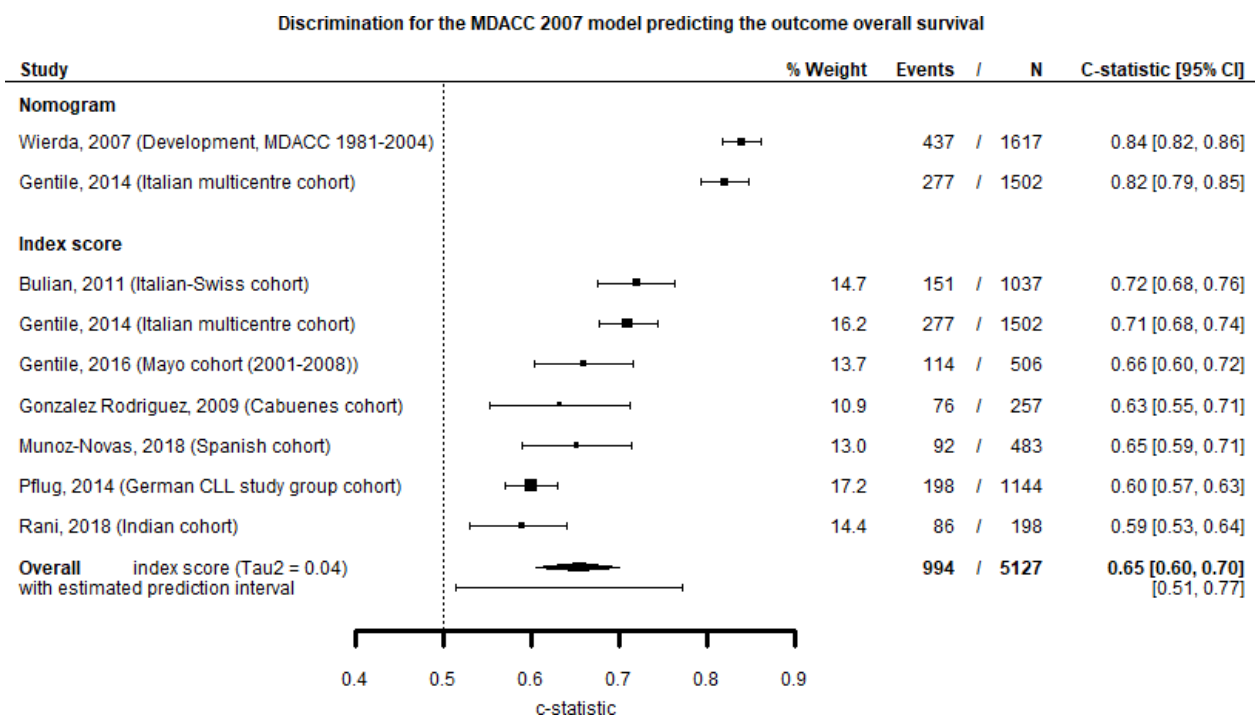
Risk group	Index score	Development study: Percentage of persons surviving at 5 years (95% CI)	Pooled percentage of persons surviving at 5 years (95% CI)	95% prediction interval (PI)
Low risk	1-3	97.0% (95.0% to 99.0%)	97.0% (94.3% to 98.4%)	90.9% to 99.0%
Intermediate risk	4-7	80.0% (78.0% to 82.0%)	82.3% (74.6% to 88.0%)	61.5% to 93.1%
High risk	≥ 8	55.0% (47.2% to 62.8%)	45.6% (31.3% to 60.5%)	21.2% to 72.3%

Discrimination

We included seven external validation studies with a total of 5127 individuals and 994 deaths observed during follow-up for meta-analysis of the c-statistic in the case of the index score. Only one external validation study provided sufficient data on performance for the nomogram, which is presented together with the development study discrimination on the upper part of the forest plot ([Figure 28](#)). Four studies that reported a measure for the c-statistic did not report a measure of uncertainty, which we estimated using the Newcombe method ([MDACC 2007 V - Bulian](#)

[2011 \(Italian-Swiss\)](#); [MDACC 2007 V - Gentile 2014 \(Italian cohort\)](#); [MDACC 2007 V - Gentile 2016 \(Mayo cohort\)](#); [MDACC 2007 V - Muñoz-Novas 2018 \(Spanish cohort\)](#)). For one study ([MDACC 2007 V - González Rodríguez \(Cabueñes\)](#)), we included the AUC as measure of discrimination instead of the c-statistic. The pooled estimate of the seven studies was 0.65 (95% CI 0.60 to 0.70), with some degree of heterogeneity (between-study standard deviation, $\tau = 0.21$). The prediction interval, which describes a range for the predicted model discrimination in a new validation study of the model, ranged from 0.51 to 0.77.

Figure 28. Meta-analysis of the c-statistic for the MDACC 2007 model (Wierda 2007)



The bootstrap-corrected c-statistic of the model development study was 0.84 (95% CI 0.82 to 0.86). The validation study for the nomogram showed a similar discrimination of 0.82 (95% CI 0.79 to 0.85). Other studies that used the nomogram provided the point score of the formula. However, due to the poor graphical quality of the presented nomogram, authors did not attempt to estimate the expected individual survival probabilities per individual ([MDACC 2007 V - Bulian 2011 \(Italian-Swiss\)](#); [MDACC 2007 V - Gentile 2014 \(Italian cohort\)](#)).

Prognostic models for the prediction of PFS or TFS (with meta-analysis, > 3 external validations)

We did not identify any prognostic model development studies that were externally validated more than three times for the outcomes PFS, TFS or TFS.

Models with one to three external validation studies

Prognostic models for the prediction of OS (without meta-analysis, 1 to 3 external validations)

GCLLSG model (Pflug 2014) (GCLLSG D - Pflug 2014 (GCLLSG))

The GCLLSG model was derived from univariable selection of predictors which were entered into a forward and backward stepwise proportional-hazards Cox regression model ([GCLLSG D - Pflug 2014 \(GCLLSG\)](#)). The authors controlled for several factors specific to the RCT data that were used for model-building (e.g. study, treatment etc). Bootstrapping techniques were used to test the robustness of the Cox regression model. Persons with missing data were excluded from analysis. Factors were categorised based on published thresholds and quartiles. The authors assigned risk scores to factors, which proved to be independent in the final multivariable model, in a weighted manner to construct the prognostic index presented in Table 7.

Table 7. Scoring of the GCLLSG model

Prognostic factor	Point distribution
FISH category del(17p)	6
Serum thymidine kinase (> 10.0 U/L)	2
B2-microglobulin (> 3.5 mg/L)	2
B2-microglobulin (> 1.7 and ≤ 3.5 mg/L)	1

IgHV mutational status (unmutated)	1
ECOG PS (> 0)	1
FISH category del(11q)	1
Gender (male)	1
Age (> 60 years)	1
Notes: FISH: fluorescence in situ hybridisation; IgHV = immunoglobulin heavy chain variable region genes; ECOG PS: Eastern Cooperative Oncology Group Performance Status	

Four risk groups were determined: low risk (0-2 points), intermediate risk (3-5 points), high risk (6-10 points) and very high risk (> 10 points). Five- and six-year OS was reported for each point score.

Calibration

Calibration was not reported for the development and external validation study ([GCLLSG V - Pflug 2014 \(Mayo cohort\)](#)).

In the development study, survival was 95.2%, 86.9%, 67.6% and 18.7% at five years for the low, intermediate, high and very high-risk group, respectively. In the validation study, survival was 95.2%, 91.1%, 71.7% and 13.6% at five years for the low, intermediate, high and very high-risk group, respectively.

Discrimination

The c-statistic of the prognostic score in the model development study was 0.75 (95% CI 0.70 to 0.78), discrimination in the external validation cohort was 0.77 (95% CI 0.70 to 0.83).

Rossi (Rossi 2013) (Rossi D - Rossi 2013 (Italian cohort))

The prognostic score by Rossi and colleagues was developed in several steps ([Rossi D - Rossi 2013 \(Italian cohort\)](#)). Firstly, the authors identified factors that were independently associated with OS by univariable and multivariable Cox regression analysis. Bootstrapping techniques were used to test the robustness of the model. In the next step, the factors were entered into a decision tree algorithm to divide individuals in subgroups. To test the stability of the decision tree, the random survival forest method with amalgamation algorithm was used. It was unclear how missing data were handled.

Calibration

No study reported a measure of calibration for this model.

In the development study, OS at five years was 86.9% in the very low-risk group (del13q14 only), 77.6% in the low-risk group (normal/+12), 65.9% in the intermediate-risk group (NOTCH1 and/or SF3B1 mutations and/or del11q22-q23 in the absence of TP53 and BIRC3 abnormalities) and 50.9% in the high-risk group (TP53 disruption and/or BIRC3 disruption independent of co-occurring lesions). One external validation study reported median survival per risk group (not reached for the very low-risk group, 13.8 years for the low-risk group, 11.2 years for the intermediate-risk

group and 7.7 years for the high-risk group) ([Rossi V - Rossi 2013 \(unclear\)](#)). In the other validation study, OS at five years was 91% in the very low-risk group, 90% in the low-risk group, 75.2% in the intermediate-risk group and 62.1% in the high-risk group ([Rossi V - Jeromin 2014 \(Munich cohort\)](#)).

Discrimination

The c-statistic of the prognostic score in the development cohort was 0.64. Only one of the two validation studies reported a measure of discrimination; the c-statistic was 0.66 (95% CI not reported).

Stephens model (Stephens 2015) (Stephens OS D - Stephens 2015 (Ohio cohort))

To develop prognostic models for OS, the authors identified significant prognostic factors in multivariable proportional hazards models using backwards selection. Based on the final model, they defined a simplified risk score based on the strength of the association. Missing data were accounted for by multiple imputation techniques. The model and score for OS included the factors ECOG performance status, age and lactate dehydrogenase (LDH).

Calibration

No study reported a measure of calibration for the prognostic models.

In the development cohort, the percentage of alive persons at two years was 89% (95% CI 74% to 96%) for score 0, 66% (95% CI 41% to 82%) for score 2 and 0% for score 4. In the validation cohort, the percentage of alive persons at two years was 95% (95% CI 83% to 99%) for score 0, 80% (95% CI 55% to 92%) for score 2 and 20% (95% CI 1% to 58%) for score 3.

Discrimination

The discrimination of the exact model for OS was 0.76 ($P < 0.03$) and 0.73 ($P < 0.0001$) for the development cohort. The validation cohort showed a c-statistic of 0.68 for the simplified score for OS.

Prognostic models for the prediction of PFS or TFS (without meta-analysis, 1 to 3 external validations)

Baliakas model (Baliakas 2019) (Baliakas D - Baliakas 2019 (multicentre))

The model presented by Baliakas and colleagues was derived from univariable selection of predictors which were entered into a proportional-hazards Cox regression model (Baliakas D - Baliakas 2019 (multicentre)), that was internally validated by bootstrapping and further confirmed by recursive partitioning based on conditional inference trees and merging of terminal nodes by an amalgamation algorithm. First, the authors split individuals in two groups, mutated and unmutated IgHV. The risk groups for mutated cases were as follows: (1) low risk: non-TP53 abnormality/+12/subset #2 membership and Binet stage A; (2) intermediate risk: Binet A with one of the following: TP53 abnormality and/or +12 and/or subset #2 membership, (3) high risk: Binet B, (4) very high risk: Binet C. The risk groups for unmutated cases were as follows: (1) very low risk: non-TP53abn/SF3B1mut/del(11q) female Binet A; (2) low risk: non-TP53abn/SF3B1mut/del(11q) male Binet A; (3) intermediate risk: Binet A with one of the following: TP53abn and/or SF3B1mut and/or del(11q); (4) high risk: Binet B; (5) very high risk: Binet C.

Calibration

No measures of calibration were reported for this model.

In the development study, the five-year treatment probability for mutated cases was 12% for group 1, 40% for group 2, 64% for group 3 and 92% for group 4. For unmutated cases, the five-year treatment probability was 45% for group 1, 65% for group 2, 78% for group 3, 90% for group 4 and 100% for group 5.

In the external validation cohort, the five-year treatment probability was only reported for the intermediate-risk groups (for mutated cases 43%, group 2; and for unmutated cases 74%, group 3).

Discrimination

In the development cohort, the discrimination was 0.745 (SE = 0.013) and 0.753 (SE = 0.013) for the cases with mutated IgHV and unmutated IgHV status, respectively. Discrimination was not reported for the external validation cohort.

GIMEMA model (Molica 2005) (GIMEMA D - Molica 2005 (GIMEMA cohort))

The GIMEMA model was derived from univariable selection of predictors which were entered into a proportional-hazards Cox regression model (GIMEMA D - Molica 2005 (GIMEMA cohort)). No internal validation was reported. Negative risk factors that were independent in multivariable analysis (lymphocyte doubling time < 12 months, lymphocyte count $>30 \times 10^9$ per L and Rai stage I to II) were assigned one point each (not weighted for HR) to create a score. Gender was added afterwards as relevant for individuals with score 0, but not for individuals with scores 1 to 3. Three risk groups were created: (1) females with score 0; (2) males with score 0; and (3) individuals with score 1 to 3 of either gender.

Calibration

No measure of calibration was reported for this model.

In the development study, the 10-year PFS was 76.2% for group 1, 61.4% for group 2 and 37.8% for group 3, respectively. In the validation study, the 10-year treatment-free survival was 89% for group 1, 54% for group 2 and 0% for group 3 (GIMEMA V - González Rodríguez 2009 (Cabueñes coh.)).

Discrimination

The development study did not report a measure of discrimination. The AUC for the validation study was 0.58 (95% CI 0.49 to 0.66).

MDACC 2011 (Wierda 2011) (MDACC 2011 D - Wierda 2011 (MDACC))

The MDACC 2011 model (MDACC 2011 D - Wierda 2011 (MDACC)) is a prognostic model derived from univariable selection of predictors. After univariable selection, the dataset was split into a training and test set. Significant predictors entered into a proportional-hazards Cox regression model (forward selection procedure with $P < 0.1$, and removed if not $P < 0.5$ in final model) were developed on the test set and applied to the test set. The factors of the final model were entered into Cox regression models for 1000 bootstrap samples to show their robustness. The final model formula can be used to derive a point score that can be translated graphically by a nomogram into individual survival probabilities at five or 10 years.

The formula is as follows: $[I(\text{No. of lymph node sites involved} = 3) \times 7.370 + I(\text{FISH11q del}) \times 9.312 + I(\text{FISH 17p del}) \times 11.285 + (\text{diameter of largest cervical lymph node in cm}) \times 4.172 + (\text{LDH}/100) \times I(\text{IgHV gene mutated}) \times 5.000 + (\text{LDH}/100) \times I(\text{IgHV gene} = \text{unmutated}) \times 1.065] + 35.467$.

Calibration

No study reported a measure of calibration for this model.

Discrimination

The development study did not report a measure of discrimination. The validation study included a total of 337 participants (91 treated) and reported a c-statistic of 0.71 (95% CI 0.60 to 0.82) (MDACC 2011 V - Molica 2016 (O-CLL1-GISL)).

Morabito model (Morabito 2009) (Morabito D - Morabito 2009 (Italian cohort))

The prognostic model developed by Morabito and colleagues (Morabito D - Morabito 2009 (Italian cohort)) was derived from univariable analysis, including dichotomisation of factors with cut-off determination by ROC curves. Significant ($P > 0.05$) prognostic factors were entered into proportional-hazards Cox regression models. Each unfavourable marker that remained significant was assigned one point, and the sum of these points formed the final score (CD38 positive = 1, ZAP-70 positive = 1, IgHV unmutated status = 1). Individuals with scores of two or three points were categorised together to form the high-risk group.

Calibration

No study reported a measure of calibration for this model.

The five-year PFS of the validation study for this model was 91.7% for the low-risk group (score 0), 82.9% for the intermediate-risk group (score 1) and 57.4% for the high-risk group (score 2-3) (Morabito V - Gentile 2014 (O-CLL1-GISL)).

Discrimination

Both the development and validation study did not report a measure of discrimination.

O-CLL1 model (Gentile 2016) (O-CLL-1 D - Gentile 2016 (O-CLL-1-GISL))

The prognostic score developed by Gentile and colleagues was derived from univariable selection of predictors (O-CLL-1 D - Gentile 2016 (O-CLL-1-GISL)). Before entering the factors into proportional-hazards Cox regression models ($P < 0.05$), continuous factors

were dichotomised using established thresholds or ROC curves. Bootstrapping techniques were used to test the robustness of the Cox regression model. There were no persons with missing values in the sample. The authors assigned risk scores to each prognostic factor that was significant in multivariable analysis based on their hazard ratio (Table 8). Individuals were divided into three different risk categories for TFS, low (score 0 to 2), intermediate (score 3 to 5) and high risk (score 6 to 7).

Table 8. Scoring of the O-CLL1 model

Prognostic factor	Point distribution		
	0	1	2
Rai staging system	0	I-II	-
B2-microglobulin	Normal	-	Elevated
ALC ($10^9/L$)	< 10	-	≥ 10
IgHV mutational status	Mutated	-	Unmutated

Notes: ALC: acute lymphocyte count; IgHV: immunoglobulin heavy chain variable region genes

Calibration

No study reported a measure of calibration for this model.

In the development study, treatment-free survival at three years was 95.3% in the low risk, 74.5% in the intermediate-risk and 28.6% in the high-risk group. Outcome frequencies were also provided per risk score. One external validation study showed similar percentages of treatment-free survival at three years with 95.5%, 78.9% and 40.6% for the low, intermediate and high-risk group, respectively (O-CLL1 V - Gentile 2016 (Mayo cohort)). The other external validation study presented results in the form of a Kaplan-Meier curve (O-CLL-1 V - Rani 2018 (Indian cohort)).

Discrimination

The c-statistic of the prognostic score in the model development study was 0.75 ($P < 0.001$). The c-statistic in the external validation cohorts were 0.72 ($P < 0.001$) (O-CLL1 V - Gentile 2016 (Mayo cohort)), and 0.55 (95% CI 0.49 to 0.60) (O-CLL-1 V - Rani 2018 (Indian cohort)).

Stephens model (Stephens 2015) (Stephens TFS D - Stephens 2015 (Ohio cohort))

To develop prognostic models for TFS, the authors identified significant prognostic factors in multivariable proportional hazards models using backwards selection. Based on the final model, they defined a simplified risk score based on the strength of the association. Missing data were accounted for by multiple imputation techniques. The model and score for TFS include the factors ECOG performance status, Rai stage, age, white blood cell count (WBC) and FISH category del(11q22.3). Risk groups were summarised as 0 to 1 points, 2 to 3 points and ≥ 4 points.

Calibration

No study reported a measure of calibration for the prognostic model for TFS.

In the development cohort, the percentage of persons who were treatment-free at two years was 85% (95% CI 0.60 to 0.95) for the low-risk group, 51% (95% CI 32 to 67) for the intermediate-risk group, and 0% for the high-risk group. In the validation cohort, the percentage of persons who were treatment-free at three years was 63% (95% CI 39 to 79) for the low-risk group, 26% (95% CI 15 to 39) for the intermediate-risk group and 16% (95% CI 6 to 29) for the high-risk group.

Discrimination

The discrimination of the exact model for TFS was 0.84 ($P < 0.017$) for the development cohort. The validation cohort showed a c-statistic of 0.66 for the simplified score for TFS.

Models without any external validation studies

The main characteristics of prognostic models for which no external validation studies could be found are described in Appendix 7.

Subgroup differences

We did not explore subgroup differences as the number of external validation studies per model with available data were too small. Three of the four external validation studies for the Barcelona-Bрно model defined CLL by the International Workshop on CLL Guideline (Hallek 2008), and one did not report diagnostic criteria. Among the seven external validations of the CLL-IPI, three studies used the International Workshop on CLL Guideline (Hallek 2008), and four did not clearly refer to the diagnostic criteria or used a combination

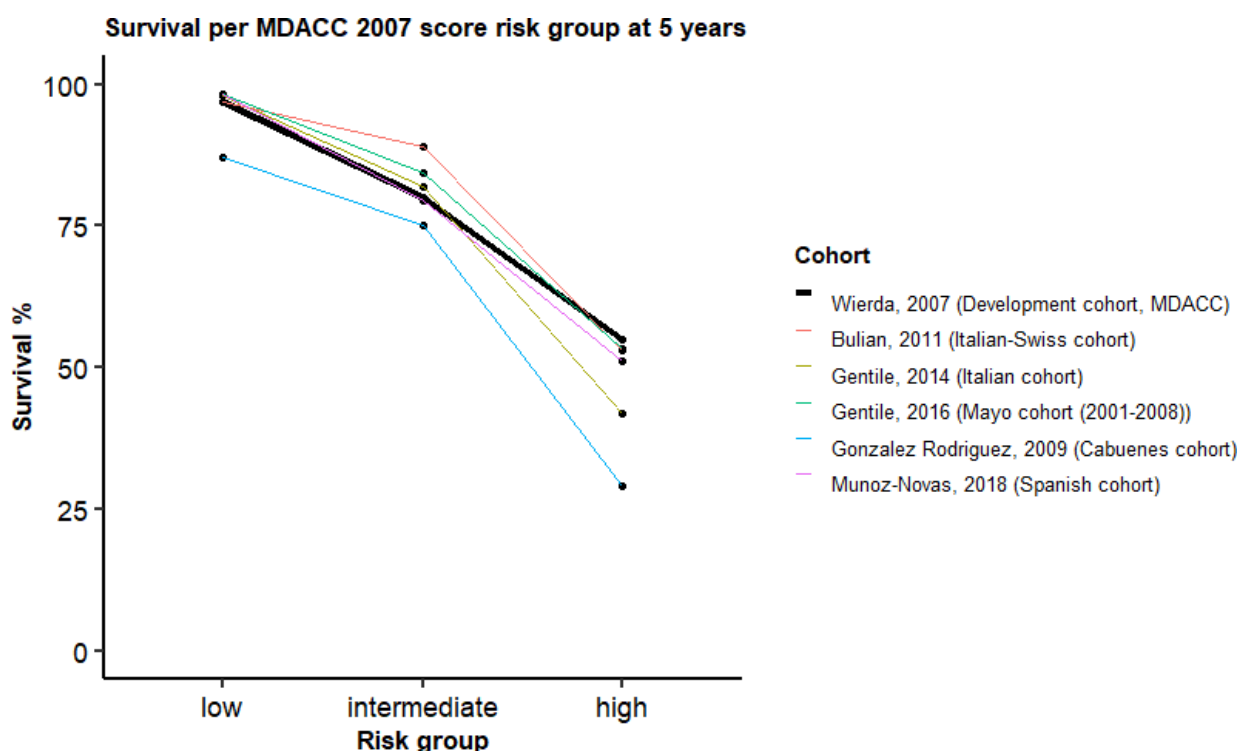
of criteria. One external validation study of the MDACC 2007 model used the International Workshop on CLL Guideline (Hallek 2008), two studies' cohorts were based on the NCI working group criteria (Cheson 1996), and four studies did not clearly report the criteria or used a combination of criteria.

Sensitivity analysis

We planned to explore the effect of risk of bias on the performance measures for the external validation studies per model. We were not able to conduct this analysis as we rated nearly all studies as high or unclear risk of bias, leaving no studies with a low risk of bias for inclusion in the sensitivity analysis.

One external validation study for the CLL-IPI reported the AUC instead of the c-statistic (CLL-IPI V - Zhu 2018 (Chinese cohort)), which we included in our meta-analysis. The pooled performance estimate for discrimination did not change (0.72, 95% CI 0.67 to 0.77 including the study; 0.72, 95% CI 0.66 to 0.78 excluding the study; see Figure 23). For the MDACC 2007 model, we also included one AUC instead of the c-statistic in the analysis. By removing this study (MDACC 2007 V - González Rodríguez (Cabueñes)), the pooled performance estimate for discrimination barely changed (from 0.65 to 0.66, 95% CI 0.60 to 0.71; for illustration, see Figure 29).

Figure 29. Sensitivity analysis of the c-statistic for the MDACC 2007 model excluding the study reporting AUC



To examine the effect of estimating the 95% CIs according to the Newcombe method, we conducted a sensitivity analysis exemplary for one included prognostic model, the CLL-IPI. In Figure 24, we limited the meta-analysis of the c-statistic to all studies that reported the 95% CI, which did not change the overall pooled estimate and 95% CIs substantially (from 0.72, 95% CI 0.67 to 0.77 for the main analysis to 0.73, 95% CI 0.66 to 0.80 for the analysis limited to studies with 95% CI reported). We calculated the 95% CIs based on Newcombe for all studies and the pooled results to compare this extreme scenario with the scenario that all 95% CIs were provided. Figure 25 shows that the estimation did not change the 95% CIs. The 95% PI changed negligibly.

For the CLL-IPI, one of the predictors was commonly missing (TP53 mutation) and replaced by a proxy prognostic factor (del(17p)). To see if the replacement limited discriminative performance, we conducted separate meta-analyses (Figure 26). The 95% CI of the pooled c-statistic overlapped, however, the number of studies in each analysis was very small (N = 3 and N = 4).

DISCUSSION

Summary of main results

We aimed to identify and describe all prognostic models for untreated CLL and their corresponding external validation studies. We identified 12 prognostic models with at least one external validation (Appendix 6). Of these, three models were validated externally more than three times, with appropriate reporting to allow us to conduct a meta-analysis of the c-statistic, which is a measure of discriminative performance of a model. Additionally, we identified 40 models or scores without any external validation study, which are further described in Appendix 7 and Characteristics of studies awaiting classification.

Results for models with more than three external validation studies per outcome for meta-analysis

Prognostic models for the prediction of OS (with meta-analysis, > 3 external validations)

The pooled c-statistic of the CLL-IPI, a point score calculated from five factors, was based on seven external validation studies with a total of 3307 individuals. The pooled estimate of its discriminative performance in seven external validation studies was 0.72 (95% CI 0.67 to 0.77). The 95% PI ranged from 0.59 to 0.83, indicating that from 100 external validation studies for this model, 95 will show a c-statistic within this range. Calibration was not reported in the publications, and thus the prognostic score was not updated based on calibration. Studies showed somewhat heterogeneous results regarding the pooled survival frequencies (observed survival frequencies) as compared to the development study survival frequencies (expected survival frequencies). Due to our diverging representation of calibration, to make assumptions regarding the calibration of the models, we assumed that the survival frequency of the development cohort can be interpreted as our expected frequency. The pooled survival per risk group was higher than observed in the development study for the very high-risk group, indicating that the model underpredicted survival in this group. This may be explained by the fact that the model was developed using data including a higher proportion of individuals with treatment indication at recruitment, who are participants with a worse prognosis than the general population with CLL. The curves showed a relatively homogeneous pattern regarding the differences between risk groups, although frequencies varied more widely for the very high-risk group, probably due to the low number of participants.

The pooled c-statistic of the Barcelona-Brno score, which contains two of the prognostic factors also included in the CLL-IPI, was derived from four external cohorts with a total of 1974 individuals. The pooled c-statistic was 0.64 (95% CI 0.60 to 0.67) and the 95% PI ranged from 0.59 and 0.68. Results between cohorts were somewhat heterogeneous, although the number of studies was relatively low to draw meaningful conclusions. Calibration was not reported in the publications, and thus the prognostic score was not updated based on calibration. The pooled survival per risk group was lower than observed in the development study, especially for the high-risk group, which means that across cohorts, the model overestimated survival as compared to the development cohort. This judgement is mainly driven by the Czech cohort. The pattern was consistent across studies.

The pooled c-statistic of the MDACC 2007 index score, which includes six prognostic factors, was derived from seven external validation studies with a total of 5127 individuals. The pooled c-statistic was 0.65 (95% CI 0.60 to 0.70; 95% PI: 0.51 to 0.77); discrimination was somewhat heterogeneous between studies. Calibration was not reported in any of the studies validating the index score, thus, no model updates were reported. The pooled survival per risk group matched the survival in the development study for the low and intermediate-risk groups. In the high-risk group, the pooled survival per risk group was lower than observed in the development study, thus the model overestimated the survival of the high-risk group. The pattern was consistent across studies.

From these three prognostic models for untreated individuals with CLL included in meta-analysis, the CLL-IPI performed best regarding the discrimination between persons with a 'as good as' compared to a worse prognosis. The score underpredicted survival for the very high-risk group, which was a small group of individuals defined by TP53 mutation or deletion.

The three models consist of unweighted ([Barcelona-Brno D - Delgado 2017 \(Barcelona cohort\)](#)) or weighted ([CLL-IPI D - Bahlo 2016 \(development cohort\)](#), [MDACC 2007 D - Wierda 2007 \(MDACC\)](#)) simplified point scores, which implies that information concerning the correct weighting of the prognostic factors was lost during simplification. Instead of individual outcome frequencies, the models result in a score on a limited scale, which reduced the possible outcome range compared to the original formula. Consequently, the simplification may have resulted in a reduction of the discriminative performance.

Prognostic models for the prediction of PFS or TFS (with meta-analysis, > 3 external validations)

We did not identify any prognostic model development studies that were externally validated more than three times for the outcomes PFS, TFS or TFS.

Models with one to three external validation studies

Prognostic models for the prediction of OS (without meta-analysis, 1 to 3 external validations)

In total, three models were developed to predict OS ([GCLLSG D - Pflug 2014 \(GCLLSG\)](#); [Rossi D - Rossi 2013 \(Italian cohort\)](#); [Stephens OS D - Stephens 2015 \(Ohio cohort\)](#)), with one, two and one external validation study, respectively. Due to the limited information on the predictive performance outside the model development setting, we would not yet consider these models ready for use in clinical practice.

Prognostic models for the prediction of PFS or TFS (without meta-analysis, 1 to 3 external validations)

In total, one model was developed to predict PFS ([GIMEMA D - Molica 2005 \(GIMEMA cohort\)](#)) and five were developed to predict TFS ([Baliakas D - Baliakas 2019 \(multicentre\)](#); [MDACC 2011 D - Wierda 2011 \(MDACC\)](#); [Morabito D - Morabito 2009 \(Italian cohort\)](#); [Stephens TFS D - Stephens 2015 \(Ohio cohort\)](#); [O-CLL-1 D - Gentile 2016 \(O-CLL-1-GISL\)](#)), with one external validation study per model for the first five and two for the last mentioned. Due to the limited information on the predictive performance outside the model development setting, we would not yet consider these models ready for use in clinical practice.

Models without any external validation studies

An abundance of prognostic models and scores without any application in external cohorts has been published, with a varying degree of reporting quality. Without any further knowledge on their performance in different settings than the development cohort, we would not recommend their use for patients.

Overall completeness, certainty of the evidence and study limitations of externally validated models

Overall completeness of the data

As expected from results of previous research ([Heus 2018](#)), reporting of the included models and their validation studies was poor. We have summarised the missing information per model for the development study and external validation studies for the primary model outcome in [Appendix 8](#).

For the CLL-IPI, none of the studies reported calibration (although we obtained information for five studies from corresponding authors). Reporting of discrimination was higher (8 out of 11, 6 times with 95% CI).

For the Barcelona-Brno score, none of the studies reported calibration (although we obtained information for one study from the corresponding authors). Reporting of discrimination was slightly higher (5 times out of 6, 3 times with 95% CI).

For the MDACC 2007 index score, two of the studies reported calibration. Reporting of discrimination was slightly higher (8 times out of 10, 7 times with 95% CI).

Of the other nine models with 17 validation studies that were not meta-analysed, none reported calibration (we obtained information on calibration for one study from the corresponding authors). Reporting of discrimination was higher (12 times, seven times with 95% CI).

Overall, across all categories of included studies (development studies with and without external validations, external validations), information on calibration and discrimination was lacking. Calibration was reported in six studies and discrimination in 37 out of 85 studies.

Many studies either excluded persons with missing prognostic factor data at enrolment, or excluded persons with missing data from the analysis. In most cases, we do not have information on the number of persons in the complete database or sample, and a comparison between included and excluded persons is rarely provided. Instead of removing data of individuals with missing data completely, multiple imputation would be preferable, but should be interpreted carefully ([Moons 2019](#); [Steyerberg 2014](#)).

None of the studies explicitly mentioned reasons for censoring, a common phenomenon in time-to-event data, in their cohort which may be particularly relevant for studies based on RCT data.

Certainty of the evidence

At the date of submission of this review, no official GRADE guidance for grading the summarised results of meta-analysis of prognostic models was available. Hence, we refrained from rating the certainty of the evidence.

Study limitations of prognostic model development studies

The majority of PROBAST ratings for model development studies was high. The main reason for this was handling missing data, predictor selection based on univariable selection of predictors, failure to correct estimates for optimism, categorisation of prognostic factors and simplification of models.

Potential bias in the review process

To prevent bias in the review process, we performed all relevant steps in duplicate and solved discrepancies in group discussions. We developed a sensitive search strategy and tracked the references and citations of all included references. However, we did not search conference abstracts or trial registries, since conference abstracts do not provide sufficient information for 'Risk of bias' assessment and prognostic studies are rarely prospectively registered, but rather built on existing retrospective databases. Thus, we would not expect to find any relevant studies by searching those additional databases.

As this is a new review type, our methods evolved during the process. We decided to adapt the newest methods regardless of the protocol, which may have introduced bias.

- *Prespecification of prognostic factors*: we did not limit our inclusion to models with e.g. a minimum set of clinically relevant factors or a specific kind of factor (such as non-invasive or genetic factors only). Therefore, we made no distinction regarding which predictors were included in the retrieved models – we included all developed and validated models that were the subject of our review, regardless of which type of predictors they included. We did not aim to rate the individual factors, examine the relationship between any one factor with our outcomes or the strength of any one factor above other factors, as this would be the subject of a prognostic factor review.
- *Search*: we have slightly adapted the search to make it more inclusive, and extended it to include studies published since the inception of MEDLINE instead of using the cut-off year 1990. Based on the recommendation of the Cochrane fast-track service, we have added a search strategy for Embase. We believe that screening more references has not introduced bias.
- *Screening*: the definition of prognostic models and the reporting standards and, with that, the amount of information expected in a publication has changed substantially over time. To avoid bias, we have included all studies that explicitly stated the aim of developing a score for prognostication, although from today's point of view, we would not consider these scores a well-developed prediction model, because they did not follow the recommended steps in the model-building process. Most of these studies can be found in [Appendix 7](#), as models without external validation studies.
- *Risk of bias*: since publication of the protocol, PROBAST was published. In our protocol, we did not specify the data analysis process. To avoid bias and use the most recently developed methods, we worked in close collaboration with the Cochrane Prognosis Methods group, and some authors of PROBAST are co-authors of this review. For transparency of our group decisions, we added the text under 'Solving disagreements'.
- *Analysis*: we planned to summarise the performance measures for calibration and discrimination. In the review we specified, according to most recent developments, that we would summarise the external validation studies per model. For some studies, we estimated the 95% CI for the c-statistic due to lack of reporting. Based on our sensitivity analysis, we assume that this estimation has not introduced bias.

We did not assess the likelihood of publication bias, i.e. a preferential publication of studies which show better model

performance, because currently, there is no established standard for this.

We would like to mention that external validation studies are ideally carried out by independent researchers (using independent samples). For the studies included in this review, this was not always the case (e.g. [CLL-IPI V - Bahlo 2016 \(Mayo clinic 2001-2014\)](#); [CLL-IPI V - Bahlo 2016 \(SCAN cohort\)](#)), and may have led to additional bias to the one assessed by PROBAST.

Applicability of findings to clinical practice and policy

We identified 52 prognostic models for various outcomes. Due to the lack of external validation, we disregarded 40 of these as not (yet) applicable. Nine more prognostic models were validated only one to three times, which we would not consider to be sufficient information to make a judgement of applicability to different settings. The three models included in meta-analysis of the c-statistic offer more information concerning their performance in different populations. The CLL-IPI seems to perform best among the three models in both the development and external validation cohorts. However, due to the relatively small number of external validation studies ($N < 10$) and serious limitations in reporting, we cannot draw final conclusions regarding the generalisability and usefulness of this model in clinical practice.

The three models that were included in meta-analyses all predict OS. Until 2015, treatment options for individuals with CLL had not changed drastically over time. However, since the introduction of new agents, such as ibrutinib and idelalisib, survival has improved for all risk groups. For individuals with a poor prognosis, there is hope that they will benefit from new and better tailored treatments. Improvements in survival for all risk groups results in systematic underprediction of survival of prognostic models that were developed using data of patients receiving traditional treatment regimens, while the improvement of prognosis especially for high-risk individuals will lead to lowered discriminative ability of a prognostic model. The prognostic models discussed in detail in this review may thus need to be updated, taking this improved baseline risk into account. In contrast, other patient-relevant outcomes, such as time-to-treatment or progression would not be influenced by improved treatment options as long as a watch-and-wait strategy is adopted, and may be an interesting alternative outcome for prediction models. Some of the models were already externally validated for both of these outcomes, and may be included in a future update of this review.

A further limitation of prognostic models in general is that they usually consider only one outcome per model, thereby not considering a balance between different outcomes, such as survival, quality of life, and other outcomes of interest.

CLL has experienced several changes in diagnostic criteria over time (in particular, in 2008 the criterion of having an absolute lymphocyte count (ALC) more than or equal to $5.0 \times 10^9/L^2$ to a B-cell count more than or equal to $5.0 \times 10^9/L$), changing the categorisation of individuals with clinical monoclonal lymphocytosis (cMBL) or CLL ([Cheson 1996](#); [Hallek 2008](#)). We aimed to explore this change in subgroup analysis but did not, due to the low number of included studies with clear division. However, this change is not expected to have an extensive effect on the performance of a model, as prognosis and patient characteristics are comparable in individuals with cMBL and Rai stage I.

Agreements and disagreements with other studies or reviews

To our knowledge, the current review is the first systematic review that aimed to identify all prognostic models developed for untreated individuals with CLL and their corresponding external validation studies.

We are aware of two other systematic reviews that cover one of the prognostic models that we identified, the CLL-IPI ([Molica 2018a](#); [Molica 2018b](#)). These systematic reviews aimed at identifying all published studies that have used CLL-IPI to predict the clinical outcome of CLL; one in patients who received chemoimmunotherapy or targeted therapies ([Molica 2018a](#)), the other without restriction of therapy ([Molica 2018b](#)). There are some differences between these reviews and our systematic review. In [Molica 2018a](#), studies were limited to individuals who received chemoimmunotherapy or targeted therapies and also included studies of relapsed/refractory CLL from conference abstracts. The authors summarised the survival frequencies per CLL-IPI risk group at two years for OS, whereas we pooled OS at five years. [Molica 2018b](#) included patients undergoing any treatment, and summarised both OS and TFS at five years. Both these reviews included data from conference proceedings, whereas we did not. Moreover, they did not report measures of discrimination such as the c-statistic (which is pivotal when assessing model performance) and did not assess risk of bias or applicability.

The included studies in [Molica 2018b](#) mostly overlapped with our identified studies for OS. The exception was two external validation studies published as abstracts ($N = 2$, CLL11 study population; [Goede 2016](#) and [Ferrer Lores 2016](#)). In contrast, we identified three additional cohorts which had not yet been published at the time of publication of [Molica 2018a](#) and [Molica 2018b](#) (CLL-IPI V - [Muñoz-Novas 2018 \(Spanish cohort\)](#); CLL-IPI V - [Rani 2018 \(Indian cohort\)](#); CLL-IPI V - [Zhu 2018 \(Chinese cohort\)](#)).

AUTHORS' CONCLUSIONS

Implications for practice

We identified three prognostic models that were validated more than three times for their primary outcome OS: the CLL-IPI, the Barcelona-Brno score, and the MDACC 2007 index score. Of the three models, the CLL-IPI seems to perform best regarding the discrimination between individuals with a good prognosis as compared to individuals with a worse prognosis. However, the number of external validation studies was relatively low to draw definite conclusions. Especially for the very high-risk group, the index underpredicts survival, which may be a result of the selective sample that was used to develop the index.

This review has been developed in a time of rapidly expanding treatment options that may drastically change the prognostic implications for the therapy and longevity of these patients. The external validation cohorts may not be representative anymore of the currently available, improved treatment options. Therefore, the models can be used to provide an approximate classification, but may need testing and updating before being applied to new patients.

Implications for research

This systematic review shows that for CLL, an abundance of studies of prognostic models and scores can be identified. Based on our very inclusive definition of a prognostic score, we found that these studies (including the newer studies) were not performed according to the current standards of prognostic model development. For the future, we recommend authors of model development studies incorporate the most recent checklists and tools, such as CHARMS and TRIPOD, as orientation to avoid common issues. We would also like to emphasise that reporting in more recent papers can benefit from improvements. The minimum reported information should include participant information (recruitment, study design, eligibility criteria, diagnostic criteria used, etc.), predictor assessment, outcome definition, and the relevant performance measures.

Our review also shows that the proportion of external model validation studies as compared to model development studies is very low. Before investing in the development of a new model, we recommend the validation and direct comparison of the existing models in different settings and populations. Most suitable for this aim are well-designed prospective studies that also take into account novel agents and emerging molecular markers. To improve predictive performance, a model can be updated with these markers or tailored to a specific population.

With time, newly identified prognostic factors may be added.

ACKNOWLEDGEMENTS

The research was supported by National Health Service (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.

We would like to thank Sarah Hodgkinson (Associate Editor, Cochrane Editorial and Methods Department), Helen Wakeford and

Clare Dooley (Managing Editors, Cochrane Editorial and Methods Department), and Anne Lethaby (Copy Editor, Copy Edit Support) for their assistance regarding the review. We would also like to thank Robin Featherstone (Information Specialist, Cochrane Editorial and Methods Department) for suggestions to improve our search strategy.

We also thank Alexandra McAleenan (Methods), Lucinda Archer (Methods), Sherry Reisner (Health Writer), Dr Stefano Molica (Department of Hematology-Oncology, Azienda Ospedaliera Pugliese-Ciaccio, Catanzaro, Italy), Dr Mary Ann Anderson (Clinical Haematology Department of The Royal Melbourne Hospital and Peter MacCallum Cancer Centre; Blood Cells and Blood Cancer Division of the Walter and Eliza Hall Institute, Melbourne, Australia) and Dr SG Agrawal (Senior Lecturer and Honorary Consultant, Division of Haemato-Oncology, St Bartholomew's Hospital, Barts Health NHS Trust; and Centre for Immunobiology, Blizard Institute, Queen Mary, University of London) for their valuable peer review.

We would like to thank Thomas Debray from the Cochrane Prognosis Methods group for statistical advice and Natali Pflug for clinical advice.

We would like to thank Yuan Chi (Centre for Evidence-Based Chinese Medicine, Beijing University of Chinese Medicine) for the translation of a Chinese publication ([CLL-IPI V - Zhu 2018 \(Chinese cohort\)](#)) and Leonardo Perales-Guerrero (medical student, Universidad de Guadalajara, Mexico) for the translation of a Spanish publication ([GIMEMA V - González Rodríguez 2009 \(Cabueñas coh.\)](#)).

We thank all principal investigators who have replied to our inquiries and/or provided us with further information or data concerning their publications. Their help enabled us to include as many studies as possible in analysis. We want to thank Massimo Gentile, Caspar da Cunha-Bang, Carolina Muñoz-Novas, Lata Rani, Gian Matteo Rigolin, Huayuan Zhu, Julio Delgado, Sanja Trajkova, Natali Pflug, Jasmin Bahlo, Pietro Bulian, and Neil E Kay.

REFERENCES

References to studies included in this review

Baliakas D - Baliakas 2019 (multicentre) {published data only}

Baliakas P, Moysiadis T, Hadzidimitriou A, Xochelli A, Jeromin S, Agathangelidis A, et al. Tailored approaches grounded on immunogenetic features for refined prognostication in chronic lymphocytic leukemia. *Haematologica* 2019;104(2):360-9.

Baliakas V - Baliakas 2019 (MLL + Scan.) {published data only}

Baliakas P, Moysiadis T, Hadzidimitriou A, Xochelli A, Jeromin S, Agathangelidis A, et al. Tailored approaches grounded on immunogenetic features for refined prognostication in chronic lymphocytic leukemia. *Haematologica* 2019;104(2):360-9.

Barcelona-Brno D - Delgado 2017 (Barcelona cohort) {published data only}

Delgado J, Doubek M, Baumann T, Kotaskova J, Molica S, Mozas P, et al. Chronic lymphocytic leukemia: a prognostic model comprising only two biomarkers (IGHV mutational status and FISH cytogenetics) separates patients with different outcome and simplifies the CLL-IPI. *American Journal of Hematology* 2017;92(4):375-80.

Barcelona-Brno V - Delgado 2017 (Brno cohort) {published data only}

Delgado J, Doubek M, Baumann T, Kotaskova J, Molica S, Mozas P, et al. Chronic lymphocytic leukemia: a prognostic model comprising only two biomarkers (IGHV mutational status and FISH cytogenetics) separates patients with different outcome and simplifies the CLL-IPI. *American Journal of Hematology* 2017;92(4):375-80.

Barcelona-Brno V - Gentile 2017 (Italian & Mayo) {published data only}

Gentile M, Shanafelt TD, Mauro FR, Laurenti L, Rossi D, Molica S, et al. Comparison between the CLL-IPI and the Barcelona-Brno prognostic model: analysis of 1299 newly diagnosed cases. *American Journal of Hematology* 2017;93(2):E35-7.

Barcelona-Brno V - Molica 2017 (O-CLL1-GISL) {published data only}

Delgado J, Doubek M, Baumann T, Kotaskova J, Molica S, Mozas P, et al. Chronic lymphocytic leukemia: a prognostic model comprising only two biomarkers (IGHV mutational status and FISH cytogenetics) separates patients with different outcome and simplifies the CLL-IPI. *American Journal of Hematology* 2017;92(4):375-80.*

Molica S, Giannarelli D, Mirabelli R, Levato L, Gentile M, Morabito F, et al. Reliability of six prognostic models to predict time-to-first-treatment in patients with chronic lymphocytic leukaemia in early phase. *American Journal of Hematology* 2017;92(6):E91-3.

Barcelona-Brno V - Muñoz-Novas 2018 (Spanish coh.) {published data only}

Munoz-Novas C, Poza-Santaella M, Gonzalez-Gascon YMI, Hernandez-Sanchez M, Rodriguez-Vicente AE, Infante MS, et al.

The international prognostic index for patients with chronic lymphocytic leukemia has the higher value in predicting overall outcome compared with the Barcelona-Brno biomarkers only prognostic model and the MD Anderson Cancer Center Prognostic Index. *Biomed Research International* 2018;9506979. [DOI: [10.1155/2018/9506979](https://doi.org/10.1155/2018/9506979)]

Barcelona-Brno V - Rani 2018 (Indian cohort) {published data only}

Rani L, Gogia A, Singh V, Kumar L, Sharma A, Kaur G, et al. Comparative assessment of prognostic models in chronic lymphocytic leukemia: evaluation in Indian cohort. *Annals of Hematology* 2018;98(2):437-43.

Barcelona-Brno V - Reda 2017 (Milan cohort) {published data only}

Reda G, Cassin R, Fattizzo B, Giannarelli D, Mattiello V, Barcellini W, et al. Chronic lymphocytic leukemia and prognostic models: a bridge between clinical and biological markers. *American Journal of Hematology* 2017;92(7):E135-7.

CLL-IPI D - Bahlo 2016 (development cohort) {published data only}

International CLL-IPI working group. An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. *Lancet Oncology* 2016;17(6):779-90.

CLL-IPI V - Bahlo 2016 (Mayo clinic 2001-2014) {published data only}

Gentile M, Shanafelt TD, Rossi D, Laurenti L, Mauro FR, Molica S, et al. Validation of the CLL-IPI and comparison with the MDACC prognostic index in newly diagnosed patients. *Blood* 2016;128(16):2093-5.

*

ZZZ <label> ZZZ*

International CLL-IPI Working Group. An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. *Lancet Oncology* 2016;17(6):779-90.

Molica S, Shanafelt TD, Giannarelli D, Gentile M, Mirabelli R, Cutrona G, et al. The chronic lymphocytic leukemia international prognostic index predicts time to first treatment in early CLL: independent validation in a prospective cohort of early stage patients. *American Journal of Hematology* 2016;91:1090-5.

CLL-IPI V - Bahlo 2016 (SCAN cohort) {published data only}

International CLL-IPI working group. An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. *Lancet Oncology* 2016;17(6):779-90.

CLL-IPI V - Da Cunha-Bang 2016 (Danish cohort) {published data only}

Da Cunha-Bang C, Christiansen I, Niemann CU. The CLL-IPI applied in a population-based cohort. *Blood* 2016;128(17):2181-3.

CLL-IPI V - Delgado 2017 (Barcelona cohort) {published data only}

Delgado J, Doubek M, Baumann T, Kotaskova J, Molica S, Mozas P, et al. Chronic lymphocytic leukemia: a prognostic model comprising only two biomarkers (IGHV mutational status and FISH cytogenetics) separates patients with different outcome and simplifies the CLL-IPI. *American Journal of Hematology* 2017;**92**(4):375-80.

CLL-IPI V - Gentile 2016 (Italian cohort) {published data only} Gentile M, Shanafelt TD, Mauro FR, Laurenti L, Rossi D, Molica S, et al. Comparison between the CLL-IPI and the Barcelona-Bruno prognostic model: analysis of 1299 newly diagnosed cases.

American Journal of Hematology 2017;**93**(2):E35-7.

*

ZZZ <label> ZZZ*

Gentile M, Shanafelt TD, Rossi D, Laurenti L, Mauro FR, Molica S, et al. Validation of the CLL-IPI and comparison with the MDACC prognostic index in newly diagnosed patients. *Blood* 2016;**128**(16):2093-5.

CLL-IPI V - Molica 2016 (O-CLL1-GISL) {published data only} Molica S, Giannarelli D, Levato L, Mirabelli R, Gentile M, Morabito F. Assessing time to first treatment in early chronic lymphocytic leukemia (CLL): a comparative performance

analysis of five prognostic models with inclusion of CLL-international prognostic index (CLL-IPI). *Leukemia & Lymphoma* 2017;**58**(7):1736-9.

Molica S, Giannarelli D, Mirabelli R, Levato L, Gentile M, Morabito F, et al. Reliability of six prognostic models to predict time-to-first-treatment in patients with chronic lymphocytic leukaemia in early phase. *American Journal of Hematology* 2017;**92**(6):E91-3.

*

ZZZ <label> ZZZ*

Molica S, Shanafelt TD, Giannarelli D, Gentile M, Mirabelli R, Cutrona G, et al. The chronic lymphocytic leukemia international prognostic index predicts time to first treatment in early CLL: independent validation in a prospective cohort of early stage patients. *American Journal of Hematology* 2016;**91**(11):1090-5.

CLL-IPI V - Muñoz-Novas 2018 (Spanish cohort) {published data only}

Munoz-Novas C, Poza-Santaella M, Gonzalez-Gascon YMI, Hernandez-Sanchez M, Rodriguez-Vicente AE, Infante MS, et al. The international prognostic index for patients with chronic lymphocytic leukemia has the higher value in predicting overall outcome compared with the Barcelona-Bruno biomarkers only prognostic model and the MD Anderson Cancer Center Prognostic Index. *Biomed Research International* 2018:9506979. [DOI: [10.1155/2018/9506979](https://doi.org/10.1155/2018/9506979)]

CLL-IPI V - Rani 2018 (Indian cohort) {published data only}

Rani L, Gogia A, Singh V, Kumar L, Sharma A, Kaur G, et al. Comparative assessment of prognostic models in chronic lymphocytic leukemia: evaluation in Indian cohort. *Annals of Hematology* 2018;**98**(2):437-43.

CLL-IPI V - Reda 2017 (Milano cohort) {published data only} Reda G, Cassin R, Fattizzo B, Giannarelli D, Mattiello V, Barcellini W, et al. Chronic lymphocytic leukemia and prognostic models: a bridge between clinical and biological markers. *American Journal of Hematology* 2017;**92**(7):E135-7.

CLL-IPI V - Rigolin 2017 (Ferrera cohort) {published data only}

Rigolin GM, Cavallari M, Quaglia FM, Formigaro L, Lista E, Urso A, et al. In CLL, comorbidities and the complex karyotype are associated with an inferior outcome independently of CLL-IPI. *Blood* 2017;**129**(26):3495-8.

CLL-IPI V - Zhu 2018 (Chinese cohort) {published data only}

Zhu HY, Wang L, Qiao J, Zou YX, Xia Y, Wu W, et al. Prognostic significance of CLL-IPI for Chinese patients with chronic lymphocytic leukemia. *Chung Hua Hsueh Yeh Hsueh Tsa Chi* 2018;**39**:392-7.

GCLLSG D - Pflug 2014 (GCLLSG) {published data only}
ZZZ <label> ZZZ*

Pflug N, Bahlo J, Shanafelt TD, Eichhorst BF, Bergmann MA, Elter T, et al. Development of a comprehensive prognostic index for patients with chronic lymphocytic leukemia. *Blood* 2014;**124**(1):49-62.

Tam CS, Seymour JF. A new prognostic score for CLL. *Blood* 2014;**124**(1):1-2.

GCLLSG V - Molica 2015 (O-CLL1-GISL) {published data only}

Molica S, Giannarelli D, Gentile M, Cutrona G, Di Renzo N, Di Raimondo F, et al. The utility of two prognostic models for predicting time to first treatment in early chronic lymphocytic leukemia patients: results of a comparative analysis. *Leukemia Research* 2013;**37**(8):943-7.

Molica S, Giannarelli D, Levato L, Mirabelli R, Gentile M, Morabito F. Assessing time to first treatment in early chronic lymphocytic leukemia (CLL): a comparative performance analysis of five prognostic models with inclusion of CLL-international prognostic index (CLL-IPI). *Leukemia & Lymphoma* 2017;**58**(7):1736-9.

Molica S, Giannarelli D, Mirabelli R, Levato L, Gentile M, Morabito F, et al. Reliability of six prognostic models to predict time-to-first-treatment in patients with chronic lymphocytic leukaemia in early phase. *American Journal of Hematology* 2017;**92**(6):E91-3.

*

ZZZ <label> ZZZ*

Molica S, Giannarelli D, Mirabelli R, Levato L, Russo A, Linardi M, et al. Unavailability of thymidine kinase does not preclude the use of German comprehensive prognostic index: results of an external validation analysis in early chronic lymphocytic leukemia and comparison with MD Anderson Cancer Center model. *European Journal of Haematology* 2015;**96**(1):72-7.

GCLLSG V - Pflug 2014 (Mayo cohort) {published data only}

Pflug N, Bahlo J, Shanafelt TD, Eichhorst BF, Bergmann MA, Elter T, et al. Development of a comprehensive prognostic

index for patients with chronic lymphocytic leukemia. *Blood* 2014;**124**(1):49-62.

GCLSG V - Rani 2018 (Indian cohort) {published data only} Rani L, Gogia A, Singh V, Kumar L, Sharma A, Kaur G, et al. Comparative assessment of prognostic models in chronic lymphocytic leukemia: evaluation in Indian cohort. *Annals of Hematology* 2018;**98**(2):437-43.

GIMEMA D - Molica 2005 (GIMEMA cohort) {published data only} Molica S, Mauro FR, Callea V, Gentile M, Giannarelli D, Lopez M, et al. A gender-based score system predicts the clinical outcome of patients with early B-cell chronic lymphocytic leukemia. *Leukemia & Lymphoma* 2005;**46**(4):553-60.

GIMEMA V - González Rodríguez 2009 (Cabueñes coh.) {published data only}

Gonzalez Rodriguez AP, Gonzalez Garcia E, Fernandez Alvarez C, Gonzalez Huerta AJ, Gonzalez Rodriguez S. B-chronic lymphocytic leukemia: epidemiological study and comparison of MDACC and GIMENA prognostic indexes [Estudio epidemiológico y comparación de los índices pronósticos del MD Anderson Cancer Center y el índice del Gruppo Italiano Malattie Ematologiche Maligne dell' Adulto en pacientes con leucemia linfática crónica de células B]. *Medicina Clinica* 2009;**133**(5):161-6.

MDACC 2007 D - Wierda 2007 (MDACC) {published data only} Wierda WG, O'Brien S, Wang X, Faderl S, Ferrajoli A, Do KA, et al. Prognostic nomogram and index for overall survival in previously untreated patients with chronic lymphocytic leukemia. *Blood* 2007;**109**(11):4679-85.

MDACC 2007 V - Bulian 2011 (Italian-Swiss) {published data only}

Bulian P, Tarnani M, Rossi D, Forconi F, Del Poeta G, Bertoni F, et al. Multicentre validation of a prognostic index for overall survival in chronic lymphocytic leukaemia. *Hematological Oncology* 2011;**29**(2):91-9.

MDACC 2007 V - Gentile 2014 (Italian cohort) {published data only}

*

ZZZ <label> ZZZ*

Gentile M, Mauro FR, Rossi D, Vincelli I, Tripepi G, Recchia AG, et al. Italian external and multicentric validation of the MD Anderson Cancer Center nomogram and prognostic index for chronic lymphocytic leukaemia patients: analysis of 1502 cases. *British Journal of Haematology* 2014;**167**:224-32.

Gentile M, Shanafelt TD, Rossi D, Laurenti L, Mauro FR, Molica S, et al. Validation of the CLL-IPI and comparison with the MDACC prognostic index in newly diagnosed patients. *Blood* 2016;**128**(16):2093-5.

MDACC 2007 V - Gentile 2016 (Mayo cohort) {published data only}

Gentile M, Shanafelt TD, Cutrona G, Molica S, Tripepi G, Alvarez I, et al. A progression-risk score to predict treatment-free survival for early stage chronic lymphocytic leukemia patients. *Leukemia* 2016;**30**(6):1440-3.

*

ZZZ <label> ZZZ*

Gentile M, Shanafelt TD, Rossi D, Laurenti L, Mauro FR, Molica S, et al. Validation of the CLL-IPI and comparison with the MDACC prognostic index in newly diagnosed patients. *Blood* 2016;**128**(16):2093-5.

Shanafelt TD, Jenkins G, Call TG, Zent CS, Slager S, Bowen DA, et al. Validation of a new prognostic index for patients with chronic lymphocytic leukemia. *Cancer* 2009;**115**:363-72.

MDACC 2007 V - González Rodríguez (Cabueñes) {published data only}

Gonzalez Rodriguez AP, Gonzalez Garcia E, Fernandez Alvarez C, Gonzalez Huerta AJ, Gonzalez Rodriguez S. B-chronic lymphocytic leukemia: epidemiological study and comparison of MDACC and GIMENA prognostic indexes [Estudio epidemiológico y comparación de los índices pronósticos del MD Anderson Cancer Center y el índice del Gruppo Italiano Malattie Ematologiche Maligne dell' Adulto en pacientes con leucemia linfática crónica de células B]. *Medicina Clinica* 2009;**133**(5):161-6.

MDACC 2007 V - Molica 2010 (GIMEMA cohort) {published data only}

Molica S, Mauro FR, Callea V, Giannarelli D, Lauria F, Rotoli B, et al. The utility of a prognostic index for predicting time to first treatment in early chronic lymphocytic leukemia: the GIMEMA experience. *Haematologica* 2010;**95**(3):464-9.

MDACC 2007 V - Molica 2015 (O-CLL1-GISL) {published data only}

Gentile M, Shanafelt TD, Cutrona G, Molica S, Tripepi G, Alvarez I, et al. A progression-risk score to predict treatment-free survival for early stage chronic lymphocytic leukemia patients. *Leukemia* 2016;**30**(6):1440-3.

Molica S, Di Raimondo F, Cutrona G, Fabris S, Mauro F, Brugiattelli M, et al. Clinical categories identified by a new prognostic index reflect biological characteristics of patients in early chronic lymphocytic leukemia: the Gruppo Italiano Studio Linfomi (GISL) experience. *Leukemia Research* 2010;**34**(8):e217-8.

Molica S, Giannarelli D, Levato L, Mirabelli R, Gentile M, Morabito F. Assessing time to first treatment in early chronic lymphocytic leukemia (CLL): a comparative performance analysis of five prognostic models with inclusion of CLL-international prognostic index (CLL-IPI). *Leukemia & Lymphoma* 2017;**58**(7):1736-9.

Molica S, Giannarelli D, Mirabelli R, Levato L, Gentile M, Morabito F, et al. Reliability of six prognostic models to predict time-to-first-treatment in patients with chronic lymphocytic leukaemia in early phase. *American Journal of Hematology* 2017;**92**(6):E91-3.

*

ZZZ <label> ZZZ*

Molica S, Giannarelli D, Mirabelli R, Levato L, Russo A, Linardi M, et al. Unavailability of thymidine kinase does not preclude the use of German comprehensive prognostic index: results of an external validation analysis in early chronic lymphocytic leukemia and comparison with MD Anderson Cancer Center model. *European Journal of Haematology* 2015;**96**(1):72-7.

MDACC 2007 V - Muñoz-Novas 2018 (Spanish cohort) {published data only}

Munoz-Novas C, Poza-Santaella M, Gonzalez-Gascon YMI, Hernandez-Sanchez M, Rodriguez-Vicente AE, Infante MS, et al. The international prognostic index for patients with chronic lymphocytic leukemia has the higher value in predicting overall outcome compared with the Barcelona-Brno biomarkers only prognostic model and the MD Anderson Cancer Center Prognostic Index. *Biomed Research International* 2018;9506979. [DOI: [10.1155/2018/9506979](https://doi.org/10.1155/2018/9506979)]

MDACC 2007 V - Pflug 2014 (3 RCTs) {published data only} Pflug N, Bahlo J, Shanafelt TD, Eichhorst BF, Bergmann MA, Elter T, et al. Development of a comprehensive prognostic index for patients with chronic lymphocytic leukemia. *Blood* 2014;124(1):49-62.

MDACC 2007 V - Rani 2018 (Indian cohort) {published data only}

Rani L, Gogia A, Singh V, Kumar L, Sharma A, Kaur G, et al. Comparative assessment of prognostic models in chronic lymphocytic leukemia: evaluation in Indian cohort. *Annals of Hematology* 2018;98(2):437-43.

MDACC 2007 V - Trajkova 2013 (Macedonia) {published data only}

Trajkova S, Cevreska L, Pivkova-Veljanovska A, Ivanovski M, Dukovski D, Popova-Simjanovska M, et al. Multivariable model consisting of clinical and biological markers for time to first treatment in CLL patients: preliminary results from single centre experience. *Makedonska Akademija na Naukite i Umetnostite Oddelenie Za Bioloski i Meditsinski Nauki Prilozi* 2013;34(3):39-48.

MDACC 2011 D - Wierda 2011 (MDACC) {published data only}

Wierda WG, O'Brien S, Wang X, Faderl S, Ferrajoli A, Do KA, et al. Multivariable model for time to first treatment in patients with chronic lymphocytic leukemia. *Journal of Clinical Oncology* 2011;29(31):4088-95.

MDACC 2011 V - Molica 2016 (O-CLL1-GISL) {published data only}

Molica S, Giannarelli D, Gentile M, Cutrona G, Di Renzo N, Di Raimondo F, et al. External validation on a prospective basis of a nomogram for predicting the time to first treatment in patients with chronic lymphocytic leukemia. *Cancer* 2013;119(6):1177-85.

Molica S, Giannarelli D, Gentile M, Cutrona G, Di Renzo N, Di Raimondo F, et al. The utility of two prognostic models for predicting time to first treatment in early chronic lymphocytic leukemia patients: results of a comparative analysis. *Leukemia Research* 2013;37(8):943-7.

Molica S, Giannarelli D, Levato L, Mirabelli R, Gentile M, Lentini M, et al. A prognostic algorithm including a modified version of MD Anderson Cancer Center (MDACC) score predicts time to first treatment of patients with clinical monoclonal lymphocytosis (cMBL)/Rai stage 0 chronic lymphocytic leukemia (CLL). *International Journal of Hematology* 2014;100(3):290-5.

Molica S, Giannarelli D, Levato L, Mirabelli R, Gentile M, Morabito F. Assessing time to first treatment in early chronic lymphocytic leukemia (CLL): a comparative performance

analysis of five prognostic models with inclusion of CLL-international prognostic index (CLL-IPI). *Leukemia & Lymphoma* 2017;58(7):1736-9.

*

ZZZ <label> ZZZ*

Molica S, Giannarelli D, Mirabelli R, Levato L, Gentile M, Morabito F, et al. Reliability of six prognostic models to predict time-to-first-treatment in patients with chronic lymphocytic leukaemia in early phase. *American Journal of Hematology* 2017;92(6):E91-3.

Molica S, Shanafelt TD, Giannarelli D, Gentile M, Mirabelli R, Cutrona G, et al. The chronic lymphocytic leukemia international prognostic index predicts time to first treatment in early CLL: independent validation in a prospective cohort of early stage patients. *American Journal of Hematology* 2016;91(11):1090-5.

Morabito D - Morabito 2009 (Italian cohort) {published data only}

Morabito F, Cutrona G, Gentile M, Matis S, Todoerti K, Colombo M, et al. Definition of progression risk based on combinations of cellular and molecular markers in patients with Binet stage A chronic lymphocytic leukaemia. *British Journal of Haematology* 2009;146(1):44-53.

Morabito V - Gentile 2014 (O-CLL1-GISL) {published data only}

Gentile M, Cutrona G, Mosca L, Matis S, Fabris S, Lionetti M, et al. Prospective validation of a risk score based on biological markers for predicting progression free survival in Binet stage A chronic lymphocytic leukemia patients: results of the multicenter O-CLL1-GISL study. *American Journal of Hematology* 2014;89(7):743-50.

O-CLL-1 D - Gentile 2016 (O-CLL-1-GISL) {published data only}

*

ZZZ <label> ZZZ*

Gentile M, Shanafelt TD, Cutrona G, Molica S, Tripepi G, Alvarez I, et al. A progression-risk score to predict treatment-free survival for early stage chronic lymphocytic leukemia patients. *Leukemia* 2016;30(6):1440-3.

Molica S, Giannarelli D, Levato L, Mirabelli R, Gentile M, Morabito F. Assessing time to first treatment in early chronic lymphocytic leukemia (CLL): a comparative performance analysis of five prognostic models with inclusion of CLL-international prognostic index (CLL-IPI). *Leukemia & Lymphoma* 2017;58(7):1736-9.

Molica S, Giannarelli D, Mirabelli R, Levato L, Gentile M, Morabito F, et al. Reliability of six prognostic models to predict time-to-first-treatment in patients with chronic lymphocytic leukaemia in early phase. *American Journal of Hematology* 2017;92(6):E91-3.

Molica S, Shanafelt TD, Giannarelli D, Gentile M, Mirabelli R, Cutrona G, et al. The chronic lymphocytic leukemia international prognostic index predicts time to first treatment in early CLL: independent validation in a prospective cohort of early stage patients. *American Journal of Hematology* 2016;91(11):1090-5.

O-CLL1 V - Gentile 2016 (Mayo cohort) {published data only} Gentile M, Shanafelt TD, Cutrona G, Molica S, Tripepi G, Alvarez I, et al. A progression-risk score to predict treatment-free survival for early stage chronic lymphocytic leukemia patients. *Leukemia* 2016;30(6):1440-3.

O-CLL-1 V - Rani 2018 (Indian cohort) {published data only} Rani L, Gogia A, Singh V, Kumar L, Sharma A, Kaur G, et al. Comparative assessment of prognostic models in chronic lymphocytic leukemia: evaluation in Indian cohort. *Annals of Hematology* 2018;98(2):437-43.

Rossi D - Rossi 2013 (Italian cohort) {published data only}

Rossi D, Rasi S, Spina V, Bruscaggin A, Monti S, Ciardullo C, et al. Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia. *Blood* 2013;121(8):1403-12.

Rossi V - Jeromin 2014 (Munich cohort) {published data only} Jeromin S, Weissmann S, Haferlach C, Dicker F, Bayer K, Grossmann V, et al. SF3B1 mutations correlated to cytogenetics and mutations in NOTCH1, FBXW7, MYD88, XPO1 and TP53 in 1160 untreated CLL patients. *Leukemia* 2014;28(1):108-17.

Rossi V - Rossi 2013 (unclear) {published data only}

Rossi D, Rasi S, Spina V, Bruscaggin A, Monti S, Ciardullo C, et al. Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia. *Blood* 2013;121(8):1403-12.

Stephens OSD - Stephens 2015 (Ohio cohort) {published data only}

Stephens DM, Ruppert AS, Weirda WG, Jones JA, Woyach JA, Maddocks K, et al. Externally validated predictive clinical model for untreated del(17p13.1) chronic lymphocytic leukemia patients. *American Journal of Hematology* 2015;90(11):967-9.

Stephens OS V - Stephens 2015 (MDACC) {published data only} Stephens DM, Ruppert AS, Weirda WG, Jones JA, Woyach JA, Maddocks K, et al. Externally validated predictive clinical model for untreated del(17p13.1) chronic lymphocytic leukemia patients. *American Journal of Hematology* 2015;90(11):967-9.

Stephens TFS D - Stephens 2015 (Ohio cohort) {published data only}

Stephens DM, Ruppert AS, Weirda WG, Jones JA, Woyach JA, Maddocks K, et al. Externally validated predictive clinical model for untreated del(17p13.1) chronic lymphocytic leukemia patients. *American Journal of Hematology* 2015;90(11):967-9.

Stephens TFS V - Stephens 2015 (MDACC) {published data only} Stephens DM, Ruppert AS, Weirda WG, Jones JA, Woyach JA, Maddocks K, et al. Externally validated predictive clinical model for untreated del(17p13.1) chronic lymphocytic leukemia patients. *American Journal of Hematology* 2015;90(11):967-9.

References to studies excluded from this review

Apelgren 2006 {published data only}

Apelgren P, Hasselblom S, Werlenius O, Nilsson-Ehle H, Andersson PO. Evaluation of clinical staging in chronic lymphocytic leukemia - population-based study. *Leukemia & Lymphoma* 2006;47(12):2505-16.

Baccarani 1982 {published data only}

Baccarani M, Cavo M, Gobbi M, Lauria F, Tura S. Staging of chronic lymphocytic leukemia. *Blood* 1982;59(6):1191-6.

Baliakas 2015 {published data only}

Baliakas P, Agathangelidis A, Hadzidimitriou A, Sutton LA, Minga E, Tsanousa A, et al. Not all IGHV3-21 chronic lymphocytic leukemias are equal: prognostic considerations. *Blood* 2015;125(5):856-9.

Berke 2019 {published data only}

Berke Mentese I, Yegin ZA, Gokcen S, Ozkurt ZN, Yagci M. Prognostic significance of serum BAFF, APRIL, TACI and BCMA levels in chronic lymphocytic leukemia. *Indian Journal of Hematology and Blood Transfusion* 2019;35(2):265-71.

Bettini 1986 {published data only}

Bettini R, Rapazzini P, Ferrari V, Steidl L. Clinical staging of chronic lymphatic leukemia. *Recenti Progressi in Medicina* 1986;77(2):72-6.

Binet 1977 {published data only}

Binet JL, Leporrier M, Dighiero G, Charron D, D'Athis P, Vaugier G, et al. A clinical staging system for chronic lymphocytic leukemia: prognostic significance. *Cancer* 1977;40(2):855-64.

Binet 1981 {published data only}

Binet JL, Auquier A, Dighiero G, Chastang C, Piguat H, Goasguen J, et al. A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. *Cancer* 1981;48(1):198-206.

Bo 2014 {published data only}

Bo MD, Del Principe MI, Pozzo F, Ragusa D, Bulian P, Rossi D, et al. NOTCH1 mutations identify a chronic lymphocytic leukemia patient subset with worse prognosis in the setting of a rituximab-based induction and consolidation treatment. *Annals of Hematology* 2014;93(10):1765-74.

Bomben 2005 {published data only}

Bomben R, Dal Bo M, Zucchetto A, Zaina E, Nanni P, Sonogo P, et al. Mutational status of IgV(H) genes in B-cell chronic lymphocytic leukemia and prognosis: percent mutations or antigen-driven selection? *Leukemia* 2005;19(8):1490-2.

Bomben 2009 {published data only}

Bomben R, Dal Bo M, Capello D, Forconi F, Ma&ei R, Laurenti L, et al. Molecular and clinical features of chronic lymphocytic leukaemia with stereotyped B cell receptors: results from an Italian multicentre study. *British Journal of Haematology* 2009;144(4):492-506.

Bou Samra 2014 {published data only}

Bou Samra E, Klein B, Commes T, Moreaux J. Identification of a 20-gene expression-based risk score as a predictor of clinical outcome in chronic lymphocytic leukemia patients. *Biomed Research International* 2014;**2014**:423174.

Brejcha 2010 {published data only}

Brejcha M, Doubek M, Cmunt E, Schwarz J, Pospisil Z, Kozmon P, et al. New prognostic markers of chronic lymphocytic leukemia (CLL) in the everyday hematological practice. An analysis of data from four hematology departments. [Czech]. *Transfuze a Hematologie Dnes* 2010;**16**(Suppl 1):62-6.

Brugiatelli 2007 {published data only}

Brugiatelli M, Loteta B, Nocilli L, Mannina D. B-cell chronic lymphocytic leukemia: clinical impact of biological prognostic factors and updated treatment strategies. *Lijechnicki Vjesnik* 2007;**129** Suppl 3:26-8.

Bulian 2014 {published data only}

Bulian P, Shanafelt TD, Fegan C, Zucchetto A, Cro L, Nuckel H, et al. CD49d is the strongest flow cytometry-based predictor of overall survival in chronic lymphocytic leukemia. *Journal of Clinical Oncology* 2014;**32**(9):897-904.

Byrd 2006 {published data only}

Byrd JC, Gribben JG, Peterson BL, Grever MR, Lozanski G, Lucas DM, et al. Select high-risk genetic features predict earlier progression following chemoimmunotherapy with fludarabine and rituximab in chronic lymphocytic leukemia: justification for risk-adapted therapy. *Journal of Clinical Oncology* 2006;**24**(3):437-43.

Cailliod 2005 {published data only}

Cailliod R, Quantin C, Carli PM, Jooste V, Le Teu & G, Binquet C, et al. A population-based assessment of the prognostic value of the CD19 positive lymphocyte count in B-cell chronic lymphocytic leukemia using Cox and Markov models. *European Journal of Epidemiology* 2005;**20**(12):993-1001.

Callea 1999 {published data only}

Callea V, Morabito F, Oliva BM, Stelitano C, Levato D, Dattilo A, et al. Surface CD14 positivity in B-cell chronic lymphocytic leukaemia is related to clinical outcome. *British Journal of Haematology* 1999;**107**(2):347-52.

Catovsky 1989 {published data only}

Catovsky D, Fooks J, Richards S, MRC Working Party on Leukaemia in Adults. Prognostic factors in chronic lymphocytic leukaemia: the importance of age, sex and response to treatment in survival. A report from the MRC CLL 1 trial. *British Journal of Haematology* 1989;**72**(2):141-9.

Cesano 2013 {published data only}

Cesano A, Perbellini O, Evensen E, Chu CC, Cio&i F, Ptacek J, et al. Association between B-cell receptor responsiveness and disease progression in B-cell chronic lymphocytic leukemia: results from single cell network profiling studies. *Haematologica* 2013;**98**(4):626-34.

Chang 2003 {published data only}

Chang CC, Liu CZ, Cleveland RP. Relative importance of CD38 expression over myeloid-associated markers expression in predicting the clinical course of B-CLL patients. *Leukemia and Lymphoma* 2003;**44**(6):977-82.

Chastang 1985 {published data only}

Chastang C, Travade P, Auquier A. Critical discussion of the assessment of a three-stage prognostic classification for chronic lymphocytic leukemia. *Statistics in Medicine* 1985;**4**(3):287-93.

Chauzeix 2018 {published data only}

Chauzeix J, Laforet MP, Deveza M, Crowther L, Marcellaud E, Derouault P, et al. Normal serum protein electrophoresis and mutated IGHV genes detect very slowly evolving chronic lymphocytic leukemia patients. *Cancer Medicine* 2018;**7**(6):2621-8.

Chelazzi 1979 {published data only}

Chelazzi G, Bettini R, Michetti A, Ricci G. Prognostic value of the clinical staging of chronic lymphocytic leukemia (author's transl). *Haematologica* 1979;**64**(4):463-71.

Chen 1997 {published data only}

Chen PM, Lin SH, Fan SF, Chiou TJ, Hsieh RK, Yu IT, et al. Genotypic characterization and multivariate survival analysis of chronic lymphocytic leukemia in Taiwan. *Acta Haematologica* 1997;**97**(4):196-204.

Chena 2008 {published data only}

Chena C, Avalos JS, Bezares RF, Arrossagaray G, Turdo K, Bistmans A, et al. Biallelic deletion 13q14.3 in patients with chronic lymphocytic leukemia: cytogenetic, FISH and clinical studies. *European Journal of Haematology* 2008;**81**(2):94-9.

Chevallier 2002 {published data only}

Chevallier P, Penther D, Avet-Loiseau H, Robillard N, Ifrah N, Mahe B, et al. CD38 expression and secondary 17p deletion are important prognostic factors in chronic lymphocytic leukaemia. *British Journal of Haematology* 2002;**116**(1):142-50.

Chiaretti 2014 {published data only}

Chiaretti S, Marinelli M, Del Giudice I, Bonina S, Piciocchi A, Messina M, et al. NOTCH1, SF3B1, BIRC3 and TP53 mutations in patients with chronic lymphocytic leukemia undergoing first-line treatment: correlation with biological parameters and response to treatment. *Leukemia & Lymphoma* 2014;**55**(12):2785-92.

Christiansen 1994 {published data only}

Christiansen I, Gidlof C, Wallgren AC, Simonsson B, Totterman TH. Serum levels of soluble intercellular adhesion molecule 1 are increased in chronic B-lymphocytic leukemia and correlate with clinical stage and prognostic markers. *Blood* 1994;**84**(9):3010-6.

Chuang 2012 {published data only}

Chuang HY, Rassenti L, Salcedo M, Licon K, Kohlmann A, Haferlach T, et al. Subnetwork-based analysis of chronic lymphocytic leukemia identifies pathways that associate with disease progression. *Blood* 2012;**120**(13):2639-49.

Ciccone 2012 {published data only}

Ciccone M, Agostinelli C, Rigolin GM, Piccaluga PP, Cavazzini F, Righi S, et al. Proliferation centers in chronic lymphocytic leukemia: correlation with cytogenetic and clinicobiological features in consecutive patients analyzed on tissue microarrays. *Leukemia* 2012;**26**(3):499-508.

Ciocioiu 1988 {published data only}

Ciocioiu A, Motoiu I, Niculescu R, Berceanu S. The staging of chronic lymphatic leukemia. Its prognostic value and therapeutic implications. *Revista de Medicină Internă, Neurologie, Psihiatrie, Neurochirurgie, Dermato-venerologie* 1988;**40**(1):41-5.

Claus 2012 {published data only}

Claus R, Lucas DM, Stilgenbauer S, Ruppert AS, Yu L, Zucknick M, et al. Quantitative DNA methylation analysis identifies a single CpG dinucleotide important for ZAP-70 expression and predictive of prognosis in chronic lymphocytic leukemia. *Journal of Clinical Oncology* 2012;**30**(20):2483-91.

Claus 2014 {published data only}

Claus R, Lucas DM, Ruppert AS, Williams KE, Weng D, Patterson K, et al. Validation of ZAP-70 methylation and its relative significance in predicting outcome in chronic lymphocytic leukemia. *Blood* 2014;**124**(1):42-8.

Cmunt 2002 {published data only}

Cmunt E, Michalova K, Sindelarova L, Karban J, Zemanova Z, Kurkova S, et al. Importance of prognostic factors in patients with chronic B-lymphocytic leukemia at the time of diagnosis. *Sbornik Lekarsky* 2002;**103**(3):359-70.

Cocco 2005 {published data only}

Cocco AE, Osei ES, Thut DM, Edinger AK, Powers JJ, Fu P, et al. Bimodal cell populations are common in chronic lymphocytic leukemia but do not impact overall survival. *American Journal of Clinical Pathology* 2005;**123**(6):818-25.

Corcoran 2005 {published data only}

Corcoran M, Parker A, Orchard J, Davis Z, Wirtz M, Schmitz OJ, et al. ZAP-70 methylation status is associated with ZAP-70 expression status in chronic lymphocytic leukemia. *Haematologica* 2005;**90**(8):1078-88.

Cordone 1998 {published data only}

Cordone I, Masi S, Mauro FR, Soddu S, Morsilli O, Valentini T, et al. p53 expression in B-cell chronic lymphocytic leukemia: a marker of disease progression and poor prognosis. *Blood* 1998;**91**(11):4342-9.

Cortese 2014 {published data only}

Cortese D, Sutton LA, Cahill N, Smedby KE, Geisler C, Gunnarsson R, et al. On the way towards a 'CLL prognostic index': focus on TP53, BIRC3, SF3B1, NOTCH1 and MYD88 in a population-based cohort. *Leukemia* 2014;**28**(3):710-3.

Coscia 2012 {published data only}

Coscia M, Vitale C, Peola S, Foglietta M, Rigoni M, Griggio V, et al. Dysfunctional V9V2 T cells are negative prognosticators

and markers of dysregulated mevalonate pathway activity in chronic lymphocytic leukemia cells. *Blood* 2012;**120**(16):3271-9.

Crespo 2003 {published data only}

Crespo M, Bosch F, Villamor N, Bellosillo B, Colomer D, Rozman M, et al. ZAP-70 expression as a surrogate for immunoglobulin-variable-region mutations in chronic lymphocytic leukemia. *New England Journal of Medicine* 2003;**348**(18):1764-75.

Criel 1997 {published data only}

Criel A, Verhoef G, Vlietinck R, Mecucci C, Billiet J, Michaux L, et al. Further characterization of morphologically defined typical and atypical CLL: a clinical, immunophenotypic, cytogenetic and prognostic study on 390 cases. *British Journal of Haematology* 1997;**97**(2):383-91.

Cro 2009 {published data only}

Cro L, Morabito F, Zucal N, Fabris S, Lionetti M, Cutrona G, et al. CD26 expression in mature B-cell neoplasia: its possible role as a new prognostic marker in B-CLL. *Hematological Oncology* 2009;**27**(3):140-7.

Cro 2010 {published data only}

Cro L, Ferrario A, Lionetti M, Bertoni F, Zucal NN, Nobili L, et al. The clinical and biological features of a series of immunophenotypic variant of B-CLL. *European Journal of Haematology* 2010;**85**(2):120-9.

Cuneo 2004 {published data only}

Cuneo A, Rigolin GM, Bigoni R, De Angeli C, Veronese A, Cavazzini F, et al. Chronic lymphocytic leukemia with 6q- shows distinct hematological features and intermediate prognosis. *Leukemia* 2004;**18**(3):476-83.

Cuneo 2018 {published data only}

Cuneo A, Follows G, Rigolin GM, Picicocchi A, Tedeschi A, Trentin L, et al. Efficacy of bendamustine and rituximab as first salvage treatment in chronic lymphocytic leukemia and indirect comparison with ibrutinib: a GIMEMA, ERIC and UK CLL FORUM study. *Haematologica* 2018;**103**(7):1209-17.

D'Arena 2001 {published data only}

D'Arena G, Musto P, Cascavilla N, Dell'Olio M, Di Renzo N, Perla G, et al. CD38 expression correlates with adverse biological features and predicts poor clinical outcome in B-cell chronic lymphocytic leukemia. *Leukemia & Lymphoma* 2001;**42**(1-2):109-14.

D'Arena 2007 {published data only}

D'Arena G, Tarnani M, Rumi C, Vaisitti T, Aydin S, De Filippi R, et al. Prognostic significance of combined analysis of ZAP-70 and CD38 in chronic lymphocytic leukemia. *American Journal of Hematology* 2007;**82**(9):787-91.

D'Arena 2012 {published data only}

D'Arena G, D'Auria F, Simeon V, Laurenti L, Deaglio S, Mansueto G, et al. A shorter time to the first treatment may be predicted by the absolute number of regulatory T-cells in patients with Rai stage 0 chronic lymphocytic leukemia. *American Journal of Hematology* 2012;**87**(6):628-31.

Damle 1999 {published data only}

Damle RN, Wasil T, Fais F, Ghiotto F, Valetto A, Allen SL, et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood* 1999;**94**(6):1840-7.

Dasgupta 2015 {published data only}

Dasgupta A, Mahapatra M, Saxena R. A study for proposal of use of regulatory T cells as a prognostic marker and establishing an optimal threshold level for their expression in chronic lymphocytic leukemia. *Leukemia & Lymphoma* 2015;**56**(6):1831-8.

Davis 2016 {published data only}

Davis Z, Forconi F, Parker A, Gardiner A, Thomas P, Catovsky D, et al. The outcome of Chronic lymphocytic leukaemia patients with 97% IGHV gene identity to germline is distinct from cases with < 97% identity and similar to those with 98% identity. *British Journal of Haematology* 2016;**173**(1):127-36.

DeAndres-Galiana 2016 {published data only}

DeAndres-Galiana EJ, Fernandez-Martinez JL, Luaces O, Del Coz JJ, Huergo-Zapico L, Acebes-Huerta A, et al. Analysis of clinical prognostic variables for chronic lymphocytic leukemia decision-making problems. *Journal of Biomedical Informatics* 2016;**60**:342-51.

De Faria 2000 {published data only}

De Faria JR, De Oliveira JS, Delbone de Faria RM, Silva MR, Gohman S, Yamamoto M, et al. Prognosis related to staging systems for chronic lymphocytic leukemia. *Revista Paulista de Medicina [Sao Paulo Medical Journal]* 2000;**118**(4):83-8.

Degan 2004 {published data only}

Degan M, Bomben R, Bo MD, Zucchetto A, Nanni P, Rupolo M, et al. Analysis of IgV gene mutations in B cell chronic lymphocytic leukaemia according to antigen-driven selection identifies subgroups with different prognosis and usage of the canonical somatic hypermutation machinery. *British Journal of Haematology* 2004;**126**(1):29-42.

Delgado 2009 {published data only}

Delgado J, Pratt G, Phillips N, Briones J, Fegan C, Nomdedeu J, et al. Beta2-microglobulin is a better predictor of treatment-free survival in patients with chronic lymphocytic leukaemia if adjusted according to glomerular filtration rate. *British Journal of Haematology* 2009;**145**(6):801-5.

Delgado 2014 {published data only}

Delgado J, Salaverria I, Baumann T, Martinez-Trillos A, Lee E, Jimenez L, et al. Genomic complexity and IGHV mutational status are key predictors of outcome of chronic lymphocytic leukemia patients with TP53 disruption. *Haematologica* 2014;**99**(11):e231-4.

Del Giudice 2011 {published data only}

Del Giudice I, Mauro FR, De Propriis MS, Santangelo S, Marinelli M, Peragine N, et al. White blood cell count at diagnosis and immunoglobulin variable region gene mutations are independent predictors of treatment-free survival in

young patients with stage A chronic lymphocytic leukemia. *Haematologica* 2011;**96**(4):626-30.

Del Poeta 2010 {published data only}

Del Poeta G, Del Principe MI, Maurillo L, Rossi FM, Buccisano F, Ammatuna E, et al. Spontaneous apoptosis and proliferation detected by BCL-2 and CD71 proteins are important progression indicators within ZAP-70 negative chronic lymphocytic leukemia. *Leukemia & Lymphoma* 2010;**51**(1):95-106.

Del Principe 2004 {published data only}

Del Principe MI, Del Poeta G, Venditti A, Buccisano F, Maurillo L, Marini R, et al. Clinical significance of soluble p53 protein in B-cell chronic lymphocytic leukemia. *Haematologica* 2004;**89**(12):1468-75.

Del Principe 2006 {published data only}

Del Principe MI, Del Poeta G, Buccisano F, Maurillo L, Venditti A, Zucchetto A, et al. Clinical significance of ZAP-70 protein expression in B-cell chronic lymphocytic leukemia. *Blood* 2006;**108**(3):853-61.

De Rossi 1989 {published data only}

De Rossi G. Prognosis in chronic lymphocytic leukemia (CLL): experience of the Italian cooperative group (GIMEMA). *Bone Marrow Transplantation* 1989;**4** Suppl 1:162-4.

Deslandes 2007 {published data only}

Deslandes E, Chevret S. Assessing surrogacy from the joint modelling of multivariate longitudinal data and survival: application to clinical trial data on chronic lymphocytic leukaemia. *Statistics in Medicine* 2007;**26**(30):5411-21.

Dimier 2018 {published data only}

Dimier N, Delmar P, Ward C, Morariu-Zamfir R, Fingerle-Rowson G, Bahlo J, et al. A model for predicting effect of treatment on progression-free survival using MRD as a surrogate end point in CLL. *Blood* 2018;**131**(9):955-62.

Di Raimondo 2001 {published data only}

Di Raimondo F, Giustolisi R, Lerner S, Cacciola E, O'Brien S, Kantarjian H, et al. Retrospective study of the prognostic role of serum thymidine kinase level in CLL patients with active disease treated with fludarabine. *Annals of Oncology* 2001;**12**(5):621-5.

Dong 2011 {published data only}

Dong HJ, Zhou LT, Zhu DX, Wang DM, Fang C, Zhu HY, et al. The prognostic significance of TP53 mutations in Chinese patients with chronic lymphocytic leukemia is independent of del(17p13). *Annals of Hematology* 2011;**90**(6):709-17.

Dong 2014 {published data only}

Dong HJ, Fang C, Wang L, Fan L, Xu J, Wu JZ, et al. TP53 Pro72 allele potentially increases the poor prognostic significance of TP53 mutation in chronic lymphocytic leukemia. *Medical Oncology* 2014;**31**(4):908.

Durak 2009 {published data only}

Durak B, Akay OM, Aslan V, Ozdemir M, Sahin F, Artan S, et al. Prognostic impact of chromosome alterations detected by FISH

- in Turkish patients with B-cell chronic lymphocytic leukemia. *Cancer Genetics and Cytogenetics* 2009;**188**(2):65-9.
- El-Kinawy 2012** {published data only}
- El-Kinawy NS, Sharaf HM, El-Hamid MA. Prognostic significance of del 17p, ZAP-70 and CD38 as independent indicators for B-CLL: correlation to response to treatment and disease outcome. *Egyptian Journal of Medical Human Genetics* 2012;**13**(2):173-81.
- Fang 2019** {published data only}
- Fang H, Reichard KK, Rabe KG, Hanson CA, Call TG, Ding W, et al. IGH translocations in chronic lymphocytic leukemia: clinicopathologic features and clinical outcomes. *American Journal of Hematology* 2019;**94**(3):338-45.
- Ferrara 1981** {published data only}
- Ferrara F, Borrelli G, Fasanaro A, Mettievier V, Rametta V, Cimino R. Prognostic evaluation of chronic lymphocytic leukemia in relation to the Rai classification. Critical analysis of 145 personal cases. *Minerva Medica* 1981;**72**(12):741-4.
- Ferreira 2014** {published data only}
- Ferreira PG, Jares P, Rico D, Gomez-Lopez G, Martinez-Trillos A, Villamor N, et al. Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia. *Genome Research* 2014;**24**(2):212-26.
- French Cooperative Group on CLL 1988** {published data only} French Cooperative Group on Chronic Lymphocytic Leukemia. Comparison of the (A, B, C) staging and the Rai's staging from a large prospective series (935 patients). *Nouvelle Revue Française d'Hématologie* 1988;**30**:363-7.
- Friedman 2009** {published data only}
- Friedman DR, Weinberg JB, Barry WT, Goodman BK, Volkheimer AD, Bond KM, et al. A genomic approach to improve prognosis and predict therapeutic response in chronic lymphocytic leukemia. *Clinical Cancer Research* 2009;**15**(22):6947-55.
- Gattei 2008** {published data only}
- Gattei V, Bulian P, Del Principe MI, Zucchetto A, Maurillo L, Buccisano F, et al. Relevance of CD49d protein expression as overall survival and progressive disease prognosticator in chronic lymphocytic leukemia. *Blood* 2008;**111**(2):865-73.
- Gdynia 2018** {published data only}
- Gdynia G, Robak T, Kopitz J, Heller A, Grekova S, Duglova K, et al. Distinct activities of glycolytic enzymes identify chronic lymphocytic leukemia patients with a more aggressive course and resistance to chemo-immunotherapy. *EBioMedicine* 2018;**32**:125-33.
- Geisler 1997** {published data only}
- Geisler CH, Philip P, Christensen BE, Hou-Jensen K, Pedersen NT, Jensen OM, et al. In B-cell chronic lymphocytic leukaemia chromosome 17 abnormalities and not trisomy 12 are the single most important cytogenetic abnormalities for the prognosis: a cytogenetic and immunophenotypic study of 480 unselected newly diagnosed patients. *Leukemia Research* 1997;**21**(11-12):1011-23.
- Gentile 2016** {published data only}
- Gentile M, Cutrona G, Molica S, Ilariucci F, Mauro FR, Di Renzo N, et al. Prospective validation of predictive value of abdominal computed tomography scan on time to first treatment in Rai 0 chronic lymphocytic leukemia patients: results of the multicenter O-CLL1-GISL study. *European Journal of Haematology* 2016;**96**(1):36-45.
- Gentile 2018** {published data only}
- Gentile M, Shanafelt TD, Mauro FR, Reda G, Rossi D, Laurenti L, et al. Predictive value of the CLL-IPI in CLL patients receiving chemo-immunotherapy as first-line treatment. *European Journal of Haematology* 2018;**101**(5):703-6.
- Giles 2003** {published data only}
- Giles FJ, Bekele BN, O'Brien S, Cortes JE, Verstovsek S, Balerdi M, et al. A prognostic model for survival in chronic lymphocytic leukaemia based on p53 expression. *British Journal of Haematology* 2003;**121**(4):578-85.
- Giudice 2018** {published data only}
- Giudice ID, Rigolin GM, Raponi S, Ca&orio L, Ilari C, Wang J, et al. Refined karyotype-based prognostic stratification of chronic lymphocytic leukemia with a low- and very-low-risk genetic profile. *Leukemia* 2018;**32**(2):543-6.
- Gogia 2014** {published data only}
- Gogia A, Raina V, Gupta R, Gajendra S, Kumar L, Sharma A, et al. Prognostic and predictive significance of smudge cell percentage on routine blood smear in chronic lymphocytic leukemia. *Clinical Lymphoma, Myeloma & Leukemia* 2014;**14**(6):514-7.
- Gonzalez 2013** {published data only}
- Gonzalez D, Else M, Wren D, Usai M, Buhl AM, Parker A, et al. CLLU1 expression has prognostic value in chronic lymphocytic leukemia after first-line therapy in younger patients and in those with mutated IGHV genes. *Haematologica* 2013;**98**(2):274-8.
- Gonzalez-Gascon 2015** {published data only}
- Gonzalez-Gascon YMI, Hernandez-Sanchez M, Rodriguez-Vicente AE, Sanzo C, Aventin A, Puiggros A, et al. A high proportion of cells carrying trisomy 12 is associated with a worse outcome in patients with chronic lymphocytic leukemia. *Hematological Oncology* 2015;**17**(10):84-92.
- Gonzalez-Rodriguez 2010** {published data only}
- Gonzalez-Rodriguez AP, Contesti J, Huergo-Zapico L, Lopez-Soto A, Fernandez-Guizan A, Acebes-Huerta A, et al. Prognostic significance of CD8 and CD4 T cells in chronic lymphocytic leukemia. *Leukemia & Lymphoma* 2010;**51**(10):1829-36.
- Grabowski 2005** {published data only}
- Grabowski P, Hultdin M, Karlsson K, Tobin G, Aleskog A, Thunberg U, et al. Telomere length as a prognostic parameter in chronic lymphocytic leukemia with special reference to VH gene mutation status. *Blood* 2005;**105**(12):4807-12.

Grever 2006 {published data only}

Grever MR, Dewald GW, Neuberg DS, Reed JC, Kitada S, Flinn IW, et al. Select high risk genetic features predict earlier progression following chemotherapy in chronic lymphocytic leukemia: prospective randomized trial (Intergroup E2997) to evaluate justification for risk-adapted therapy [abstract]. *Journal of Clinical Oncology: ASCO Annual Meeting Proceedings* 2006;24:342.

Hallek 1999 {published data only}

Hallek M, Langenmayer I, Nerl C, Knauf W, Dietzfelbinger H, Adorf D, et al. Elevated serum thymidine kinase levels identify a subgroup at high risk of disease progression in early, nonmolding chronic lymphocytic leukemia. *Blood* 1999;93(5):1732-7.

Han 1984 {published data only}

Han T, Barcos M, Emrich L, Ozer H, Gajera R, Gomez GA, et al. Bone marrow infiltration patterns and their prognostic significance in chronic lymphocytic leukemia: correlations with clinical, immunologic, phenotypic, and cytogenetic data. *Journal of Clinical Oncology* 1984;2(6):562-70.

Herold 2011 {published data only}

Herold T, Jurinovic V, Metzeler KH, Boulesteix AL, Bergmann M, Seiler T, et al. An eight-gene expression signature for the prediction of survival and time to treatment in chronic lymphocytic leukemia. *Leukemia* 2011;25(10):1639-45.

Hock 2010 {published data only}

Hock BD, McKenzie JL, McArthur L, Tansley S, Taylor KG, Fernyhough LJ. CD38 as a prognostic marker in chronic lymphocytic leukaemia at a single New Zealand centre: patient survival in comparison to age- and sex-matched population data. *Internal Medicine Journal* 2010;40(12):842-9.

Houldsworth 2014 {published data only}

Houldsworth J, Guttapalli A, Thodima V, Yan XJ, Mendiratta G, Zielonka T, et al. Genomic imbalance defines three prognostic groups for risk stratification of patients with chronic lymphocytic leukemia. *Leukemia & Lymphoma* 2014;55(4):920-8.

Hus 2006 {published data only}

Hus I, Podhorecka M, Bojarska-Junak A, Rolinski J, Schmitt M, Sieklucka M, et al. The clinical significance of ZAP-70 and CD38 expression in B-cell chronic lymphocytic leukaemia. *Annals of Oncology* 2006;17(4):683-90.

Jaksic 1981 {published data only}

Jaksic B, Vitale B. Total tumour mass score (TTM): a new parameter in chronic lymphocyte leukaemia. *British Journal of Haematology* 1981;49(3):405-13.

Jaksic 1992 {published data only}

Jaksic B, Kusec R. More on early chronic lymphocytic leukemia (CLL): clinical staging systems and indications for treatment. *Hematologic Pathology* 1992;6(4):219-21.

Jaksic 2014 {published data only}

Jaksic O, Vitale B, Jaksic B. An old and simple solution for a new problem - more on clinical staging and evaluation of response

in B-cell chronic lymphocytic leukaemia in the era of new therapies. *British Journal of Haematology* 2014;165(5):737-40.

Josefsson 2007 {published data only}

Josefsson P, Geisler CH, Le&ers H, Petersen JH, Andersen MK, Jurlander J, et al. CLLU1 expression analysis adds prognostic information to risk prediction in chronic lymphocytic leukemia. *Blood* 2007;109(11):4973-9.

Juliusson 1986 {published data only}

Juliusson G. Immunologic and cytogenetic studies improve prognosis prediction in chronic B-lymphocytic leukemia. A multivariate analysis of 24 variables. *Cancer* 1986;58(3):688-93.

Juliusson 1990 {published data only}

Juliusson G, Oscier DG, Fitchett M, Ross FM, Stockdill G, Mackie MJ, et al. Prognostic subgroups in B-cell chronic lymphocytic leukemia defined by specific chromosomal abnormalities. *New England Journal of Medicine* 1990;323(11):720-4.

Kahraman 2014 {published data only}

Kahraman Cetin N, Kacar Doger F, Kurtgoz S, Ruscuklu D, Yavasoglu I. Prognostic significance of the IgVH mutation status and immunohistochemical analysis of ZAP70 and CD38 in bone marrow biopsies in chronic lymphocytic leukemia. *Turkiye Klinikleri Journal of Medical Sciences* 2014;34(3):334-44.

Kardum-Skelin 2008 {published data only}

Kardum-Skelin I, Planinc-Peraica A, Ostojic Kolonic S, Radic-Kristo D, Milas M, Vrhovac R, et al. Clinical and laboratory prognostic parameters for leukemic types of chronic lymphoproliferative diseases. *Acta Medica Croatica* 2008;62(4):351-64.

Kardum-Skelin 2009 {published data only}

Kardum-Skelin I, Jaksic O, Kolonic SO, Vrhovac R, Fabijanic I, Jelic-Puskaric B, et al. New parameters of diploid histogram of image DNA cytometry and newly characterized types of nucleolar organizer region structures in defining the proliferative-kinetic index in chronic leukemic lymphoproliferative disorders. *Analytical & Quantitative Cytology & Histology* 2009;31(5):313-23.

Karmiris 1994 {published data only}

Karmiris T, Rohatiner AZ, Love S, Carter M, Ganjoo RK, Amess J, et al. The management of chronic lymphocytic leukemia at a single centre over a 24-year period: prognostic factors for survival. *Hematological Oncology* 1994;12(1):29-39.

Kay 2018 {published data only}

Kay NE, LaPlant BR, Pettinger AM, Call TG, Leis JF, Ding W, et al. Cumulative experience and long term follow-up of pentostatin-based chemoimmunotherapy trials for patients with chronic lymphocytic leukemia. *Expert Review of Haematology* 2018;11(4):337-49.

Keating 2000 {published data only}

Keating MJ, Smith TL, Lerner S, O'Brien S, Robertson LE, Kantarjian H, et al. Prediction of prognosis following

- fludarabine used as secondary therapy for chronic lymphocytic leukemia. *Leukemia and Lymphoma* 2000;**37**(1-2):71-85.
- Khalifa 2002** {published data only}
- Khalifa M, Chehata S, Laatiri MA, Grira C, Gharbi O, Kortas M, et al. Epidemiologic, clinical and therapeutic aspects of chronic lymphoid leukemia: apropos of 120 cases. *Tunisie Medicale* 2002;**80**(10):584-9.
- Kim 2004** {published data only}
- Kim SZ, Chow KU, Kukoc-Zivojnov N, Boehrer S, Brieger A, Steimle-Grauer SA, et al. Expression of ZAP-70 protein correlates with disease stage in chronic lymphocytic leukemia and is associated with, but not generally restricted to, non-mutated Ig VH status. *Leukemia and Lymphoma* 2004;**45**(10):2037-45.
- Kimby 1988** {published data only}
- Kimby E, Mellstedt H, Nilsson B, Bjorkholm M, Holm G, Lindemalm C, et al. Blood lymphocyte characteristics as predictors of prognosis in chronic lymphocytic leukemia of B-cell type. *Hematological Oncology* 1988;**6**(1):47-55.
- Kleinstern 2018** {published data only}
- Kleinstern G, Camp NJ, Goldin LR, Vachon CM, Vajdic CM, De Sanjose S, et al. Association of polygenic risk score with the risk of chronic lymphocytic leukemia and monoclonal B-cell lymphocytosis. *Blood* 2018;**131**:2541-51.
- Knospe 1977** {published data only}
- Knospe WH, Gregory SA, Trobaugh FE Jr, Stedronsky JA, Schrek R. Chronic lymphocytic leukemia: correlation of clinical course and therapeutic response with in vitro testing and morphology of lymphocytes. *American Journal of Hematology* 1977;**2**(1):73-101.
- Koberda 1989** {published data only}
- Koberda J, Czyz J, Hellmann A. Leukocyte doubling time as a prognostic factor in chronic lymphocytic leukemia. *Acta Haematologica Polonica* 1989;**20**(2):195-200.
- Korycka-Wolowiec 2011** {published data only}
- Korycka-Wolowiec A, Kotkowska A, Wawrzyniak E, Blonski JZ, Robak T. Do chromosomal aberrations detected in chronic lymphocytic leukemia patients by conventional cytogenetics with DSP30 have the clinical significance? Preliminary results. *Acta Haematologica Polonica* 2011;**42**(4):669-80.
- Krober 2002** {published data only}
- Krober A, Seiler T, Benner A, Bullinger L, Bruckle E, Lichter P, et al. V(H) mutation status, CD38 expression level, genomic aberrations, and survival in chronic lymphocytic leukemia. *Blood* 2002;**100**(4):1410-6.
- Krober 2006** {published data only}
- Krober A, Bloehdorn J, Hafner S, Buhler A, Seiler T, Kienle D, et al. Additional genetic high-risk features such as 11q deletion, 17p deletion, and V3-21 usage characterize discordance of ZAP-70 and VH mutation status in chronic lymphocytic leukemia. *Journal of Clinical Oncology* 2006;**24**(6):969-75.
- Kryachok 2011** {published data only}
- Kryachok I, Abramenko I, Bilous N, Chumak A, Martina Z, Filonenko I. IGHV gene rearrangements as outcome predictors for CLL patients: experience of Ukrainian group. *Medical Oncology* 2011;**29**(2):1093-101.
- Kurec 1992** {published data only}
- Kurec AS, Threatte GA, Gottlieb AJ, Smith JR, Anderson J, Davey FR. Immunophenotypic subclassification of chronic lymphocytic leukaemia (CLL). *British Journal of Haematology* 1992;**81**(1):45-51.
- Lai 2002** {published data only}
- Lai R, O'Brien S, Maushouri T, Rogers A, Kantarjian H, Keating M, et al. Prognostic value of plasma interleukin-6 levels in patients with chronic lymphocytic leukemia. *Cancer* 2002;**95**(5):1071-5.
- Lech-Maranda 2012** {published data only}
- Lech-Maranda E, Grzybowska-Izydorczyk O, Wyka K, Mlynarski W, Borowiec M, Antosik K, et al. Serum tumor necrosis factor-alpha and interleukin-10 levels as markers to predict outcome of patients with chronic lymphocytic leukemia in different risk groups defined by the IGHV mutation status. *Archivum Immunologiae et Therapiae Experimentalis* 2012;**60**(6):477-86.
- Lech-Maranda 2013** {published data only}
- Lech-Maranda E, Mlynarski W, Grzybowska-Izydorczyk O, Borowiec M, Pastorczak A, Cebula-Obrzut B, et al. Polymorphisms of TNF and IL-10 genes and clinical outcome of patients with chronic lymphocytic leukemia. *Genes, Chromosomes & Cancer* 2013;**52**(3):287-96.
- Lecouvet 1997** {published data only}
- Lecouvet FE, Vande Berg BC, Michaux L, Schmitz PJ, Malghem J, Jamart J, et al. Early chronic lymphocytic leukemia: prognostic value of quantitative bone marrow MR imaging findings and correlation with hematologic variables. *Radiology* 1997;**204**(3):813-8.
- Li 2008** {published data only}
- Li FJ, Ding S, Pan J, Shakhmatov MA, Kashentseva E, Wu J, et al. FCRL2 expression predicts IGHV mutation status and clinical progression in chronic lymphocytic leukemia. *Blood* 2008;**112**(1):179-87.
- Li 2017a** {published data only}
- Li H, Xiong W, Liu H, Yi S, Yu Z, Liu W, et al. Serum LDH level may predict outcome of chronic lymphocytic leukemia patients with a 17p deletion: a retrospective analysis of prognostic factors in China. *Chinese Journal of Cancer Research* 2017;**29**(2):156-65.
- Lin 2002** {published data only}
- Lin K, Sherrington PD, Dennis M, Matrai Z, Cawley JC, Pettitt AR. Relationship between p53 dysfunction, CD38 expression, and IgV(H) mutation in chronic lymphocytic leukemia. *Blood* 2002;**100**(4):1404-9.
- Lin 2014** {published data only}
- Lin TT, Norris K, Heppel NH, Pratt G, Allan JM, Allsup DJ, et al. Telomere dysfunction accurately predicts clinical

outcome in chronic lymphocytic leukaemia, even in patients with early stage disease. *British Journal of Haematology* 2014;**167**(2):214-23.

Lozano-Santos 2014 {published data only}

Lozano-Santos C, Martinez-Velasquez J, Fernandez-Cuevas B, Polo N, Navarro B, Millan I, et al. Vascular endothelial growth factor A (VEGFA) gene polymorphisms have an impact on survival in a subgroup of indolent patients with chronic lymphocytic leukemia. *PLOS One* 2014;**9**(6):e101063.

Lucas 2015 {published data only}

Lucas DM, Ruppert AS, Lozanski G, Dewald GW, Lozanski A, Claus R, et al. Cytogenetic prioritization with inclusion of molecular markers predicts outcome in previously untreated patients with chronic lymphocytic leukemia treated with fludarabine or fludarabine plus cyclophosphamide: a long-term follow-up study of the US intergroup phase III trial E2997. *Leukemia and Lymphoma* 2015;**56**(11):3031-7.

Maffei 2007 {published data only}

Maffei R, Marasca R, Martinelli S, Castelli I, Santachiara R, Morandi E, et al. Angiopoietin-2 expression in B-cell chronic lymphocytic leukemia: association with clinical outcome and immunoglobulin heavy-chain mutational status. *Leukemia* 2007;**21**(6):1312-5.

Maffei 2010 {published data only}

Maffei R, Martinelli S, Santachiara R, Rossi D, Guarnotta C, Sozzi E, et al. Angiopoietin-2 plasma dosage predicts time to first treatment and overall survival in chronic lymphocytic leukemia. *Blood* 2010;**116**(4):584-92.

Mandelli 1987 {published data only}

Mandelli F, De Rossi G, Mancini P, Alberti A, Cajozzo A, Grignani F, et al. Prognosis in chronic lymphocytic leukemia: a retrospective multicentric study from the GIMEMA group. *Journal of Clinical Oncology* 1987;**5**(3):398-406.

Mansouri 2013 {published data only}

Mansouri L, Grabowski P, Degerman S, Svenson U, Gunnarsson R, Cahill N, et al. Short telomere length is associated with NOTCH1/SF3B1/TP53 aberrations and poor outcome in newly diagnosed chronic lymphocytic leukemia patients. *American Journal of Hematology* 2013;**88**(8):647-51.

Marasca 2005 {published data only}

Marasca R, Maffei R, Morselli M, Zucchini P, Castelli I, Martinelli S, et al. Immunoglobulin mutational status detected through single-round amplification of partial V(H) region represents a good prognostic marker for clinical outcome in chronic lymphocytic leukemia. *Journal of Molecular Diagnostics* 2005;**7**(5):566-74.

Marasca 2013 {published data only}

Marasca R, Maffei R, Martinelli S, Fiorcari S, Bulgarelli J, Debbia G, et al. Clinical heterogeneity of de novo 11q deletion chronic lymphocytic leukaemia: prognostic relevance of extent of 11q deleted nuclei inside leukemic clone. *Hematological Oncology* 2013;**31**(2):348-55.

Martinelli 2008 {published data only}

Martinelli S, Ma&ei R, Castelli I, Santachiara R, Zucchini P, Fontana M, et al. Increased expression of angiopoietin-2 characterizes early B-cell chronic lymphocytic leukemia with poor prognosis. *Leukemia Research* 2008;**32**(4):593-7.

Masic 1998 {published data only}

Masic N, Gagro A, Rabatic S, Sabioncello A, Dasic G, Jaksic B, et al. Decision-tree approach to the immunophenotype-based prognosis of the B-cell chronic lymphocytic leukemia. *American Journal of Hematology* 1998;**59**(2):143-8.

Mateva 2001 {published data only}

Mateva N, Nenova I, Dimitrov I, Karnolski I, Ananoshtev N. A predictive model for survival of patients with chronic lymphocytic leukemia - results of a multivariate survival analysis. *Journal of the Balkan Union of Oncology* 2001;**6**(4):429-33.

Matthews 2006 {published data only}

Matthews C, Catherwood MA, Morris TC, Kettle PJ, Drake MB, Gilmore WS, et al. Serum TK levels in CLL identify Binet stage A patients within biologically defined prognostic subgroups most likely to undergo disease progression. *European Journal of Haematology* 2006;**77**(4):309-17.

Matthews 2007 {published data only}

Matthews C, Catherwood MA, Morris TC, Alexander HD. V(H)3-48 and V(H)3-53, as well as V(H)3-21, gene rearrangements define unique subgroups in CLL and are associated with biased lambda light chain restriction, homologous LCDR3 sequences and poor prognosis. *Leukemia Research* 2007;**31**(2):231-4.

Matutes 2013 {published data only}

Matutes E, Bosanquet AG, Wade R, Richards SM, Else M, Catovsky D. The use of individualized tumor response testing in treatment selection: second randomization results from the LRF CLL4 trial and the predictive value of the test at trial entry. *Leukemia* 2013;**27**(2):507-10.

Matutes 2017 {published data only}

Matutes E, Polliack A. Predicting time to first treatment in early stage CLL: scores, values and comparative prognostic models. *Leukemia & Lymphoma* 2017;**58**(7):1528-9.

Melo 1987 {published data only}

Melo JV, Catovsky D, Gregory WM, Galton DA. The relationship between chronic lymphocytic leukaemia and prolymphocytic leukaemia. IV. Analysis of survival and prognostic features. *British Journal of Haematology* 1987;**65**(1):23-9.

Miao 2018 {published data only}

Miao Y, Miao Y, Shi K, Sun Q, Zhao SS, Xia Y, et al. A higher percentage of cells with 13q deletion predicts worse outcome in Chinese patients with chronic lymphocytic leukemia carrying isolated 13q deletion. *Annals of Hematology* 2018;**97**:1663-9.

Molica 1984 {published data only}

Molica S, Puzzon P, Iannaccaro P, Pane C, Alberti A. Prognosis of chronic lymphocytic leukemia: analysis of one hundred cases. *Tumori* 1984;**70**(5):399-402.

Molica 1986 {published data only}

Molica S, Alberti A. Prognostic value of "total tumor mass score" (TTM): a retrospective analysis of 130 patients with chronic lymphocytic leukemia. *Tumori* 1986;**72**(6):559-64.

Molica 1988 {published data only}

Molica S, Alberti A. Investigation of nuclear clefts as a prognostic parameter in chronic lymphocytic leukemia. *European Journal of Haematology* 1988;**41**(1):62-5.

Molica 1991 {published data only}

Molica S. Progression and survival studies in early chronic lymphocytic leukemia. *Blood* 1991;**78**(4):895-9.

Molica 1994 {published data only}

Molica S, Brugiatielli M, Callea V, Morabito F, Levato D, Nobile F, et al. Comparison of younger versus older B-cell chronic lymphocytic leukemia patients for clinical presentation and prognosis. A retrospective study of 53 cases. *European Journal of Haematology* 1994;**52**(4):216-21.

Molica 1998 {published data only}

Molica S, Levato D, Dattilo A, Mannella A. Clinico-prognostic relevance of quantitative immunophenotyping in B-cell chronic lymphocytic leukemia with emphasis on the expression of CD20 antigen and surface immunoglobulins. *European Journal of Haematology* 1998;**60**(1):47-52.

Molica 1999a {published data only}

Molica S, Levato D, Cascavilla N, Levato L, Musto P. Clinico-prognostic implications of simultaneous increased serum levels of soluble CD23 and beta2-microglobulin in B-cell chronic lymphocytic leukemia. *European Journal of Haematology* 1999;**62**(2):117-22.

Molica 1999b {published data only}

Molica S, Levato D, Dattilo A. Natural history of early chronic lymphocytic leukemia. A single institution study with emphasis on the impact of disease-progression on overall survival. *Haematologica* 1999;**84**(12):1094-9.

Molica 2008 {published data only}

Molica S, Vitelli G, Cutrona G, Todoerti K, Mirabelli R, Digiesi G, et al. Serum thrombopoietin compared with ZAP-70 and immunoglobulin heavy-chain gene mutation status as a predictor of time to first treatment in early chronic lymphocytic leukemia. *Leukemia and Lymphoma* 2008;**49**(1):62-7.

Montserrat 1991 {published data only}

Montserrat E, Gomis F, Vallespi T, Rios A, Romero A, Soler J, et al. Presenting features and prognosis of chronic lymphocytic leukemia in younger adults. *Blood* 1991;**78**(6):1545-51.

Morabito 2001 {published data only}

Morabito F, Mangiola M, Oliva B, Stelitano C, Callea V, Deaglio S, et al. Peripheral blood CD38 expression predicts survival in B-cell chronic lymphocytic leukemia. *Leukemia Research* 2001;**25**(11):927-32.

Morabito 2015a {published data only}

Morabito F, Cutrona G, Gentile M, Fabris S, Matis S, Vigna E, et al. Is ZAP70 still a key prognostic factor in early stage chronic lymphocytic leukaemia? Results of the analysis from a prospective multicentre observational study. *British Journal of Haematology* 2015;**168**(3):455-9.

Morabito 2015b {published data only}

Morabito F, Cutrona G, Mosca L, D'Anca M, Matis S, Gentile M, et al. Surrogate molecular markers for IGHV mutational status in chronic lymphocytic leukemia for predicting time to first treatment. *Leukemia Research* 2015;**39**(8):840-5.

Morabito 2018c {published data only}

Morabito F, Shanafelt TD, Gentile M, Reda G, Mauro FR, Rossi D, et al. Immunoglobulin heavy chain variable region gene and prediction of time to first treatment in patients with chronic lymphocytic leukemia: mutational load or mutational status? Analysis of 1003 cases. *American Journal of Hematology* 2018;**93**(9):E216-9.

Moreno 2019 {published data only}

Moreno C, Delgado J, Byrd JC, Zvagelsky WL, Suzuki S, Hsu E, et al. Changes in clinical stage identify patients with CLL and different outcome within iwCLL partial response: RESONATE study. *British Journal of Haematology* 2019;**185**(1):148-50. [DOI: [10.1111/bjh.15397](https://doi.org/10.1111/bjh.15397)]

Morilla 2008 {published data only}

Morilla A, Gonzalez de Castro D, Del Giudice I, Osuji N, Else M, Morilla R, et al. Combinations of ZAP-70, CD38 and IGHV mutational status as predictors of time to first treatment in CLL. *Leukemia & Lymphoma* 2008;**49**(11):2108-15.

Nabhan 2017 {published data only}

Nabhan C, Mato A, Flowers CR, Grinblatt DL, Lamanna N, Weiss MA, et al. Characterizing and prognosticating chronic lymphocytic leukemia in the elderly: prospective evaluation on 455 patients treated in the United States. *BMC Cancer* 2017;**17**(1):198.

Nedeva 2018 {published data only}

Nedeva A, Naseva E, Kindekov I, Petkova N, Nikolov I, Raynov J. The prognostic significance of interphase cytogenetic abnormalities in chronic lymphocytic leukemia. *Archives of Hellenic Medicine* 2018;**35**(4):520-6.

Nenova 2000 {published data only}

Nenova I, Karnolski I, Mateva N, Ananoshtev N. A study of prognostic factors in patients with chronic lymphocytic leukemia. *Journal of Balkan Union of Oncology* 2000;**5**(1):49-54.

Nipp 2014 {published data only}

Nipp RD, Volkheimer AD, Davis ED, Chen Y, Weinberg JB, Friedman DR. CD38 variation as a prognostic factor in chronic lymphocytic leukemia. *Leukemia & Lymphoma* 2014;**55**(1):191-4.

Nola 2004 {published data only}

Nola M, Pavletic SZ, Weisenburger DD, Smith LM, Bast MA, Vose JM, et al. Prognostic factors influencing survival in patients

with B-cell small lymphocytic lymphoma. *American Journal of Hematology* 2004;**77**(1):31-5.

Nowakowski 2009 {published data only}

Nowakowski GS, Hoyer JD, Shanafelt TD, Zent CS, Call TG, Bone ND, et al. Percentage of smudge cells on routine blood smear predicts survival in chronic lymphocytic leukemia. *Journal of Clinical Oncology* 2009;**27**(11):1844-9.

Nuckel 2006 {published data only}

Nuckel H, Huttman A, Klein-Hitpass L, Schroers R, Fuhrer A, Sellmann L, et al. Lipoprotein lipase expression is a novel prognostic factor in B-cell chronic lymphocytic leukemia. *Leukemia & Lymphoma* 2006;**47**(6):1053-61.

Nuckel 2009 {published data only}

Nuckel H, Collins CH, Frey UH, Sellmann L, Durig J, Si&ert W, et al. FCRL2 mRNA expression is inversely associated with clinical progression in chronic lymphocytic leukemia. *European Journal of Haematology* 2009;**83**(6):541-9.

O'Brien 1993 {published data only}

O'Brien S, Kantarjian H, Beran M, Smith T, Koller C, Estey E, et al. Results of fludarabine and prednisone therapy in 264 patients with chronic lymphocytic leukemia with multivariate analysis-derived prognostic model for response to treatment. *Blood* 1993;**82**(6):1695-700.

Ocana 2007 {published data only}

Ocana E, Delgado-Perez L, Campos-Caro A, Munoz J, Paz A, Franco R, et al. The prognostic role of CXCR3 expression by chronic lymphocytic leukemia B cells. *Haematologica* 2007;**92**(3):349-56.

Oliveira 2011 {published data only}

Oliveira AC, De La Banda E, Domingo-Domenech E, Encuentra M, Mercadal S, Domingo A, et al. Prospective study of clinical and biological prognostic factors at diagnosis in patients with early stage B-cell chronic lymphocytic leukemia. *Leukemia and Lymphoma* 2011;**52**(3):429-35.

Orgueira 2019 {published data only}

Orgueira AM, Rodriguez BA, Vence NA, Lopez AB, Arias JA, Varela ND, et al. Time to treatment prediction in chronic lymphocytic leukemia based on new transcriptional patterns. *Frontiers in Oncology* 2019;**9**:79.

Oscier 1990 {published data only}

Oscier DG, Stevens J, Hamblin TJ, Pickering RM, Lambert R, Fitchett M. Correlation of chromosome abnormalities with laboratory features and clinical course in B-cell chronic lymphocytic leukaemia. *British Journal of Haematology* 1990;**76**(3):352-8.

Paolino 1984 {published data only}

Paolino W, Infelise V, Levis A, Marmont F, Vitolo U, Paolino F, et al. Adenosplenomegaly and prognosis in uncomplicated and complicated chronic lymphocytic leukemia. A study of 362 cases. *Cancer* 1984;**54**(2):339-46.

Plesingerova 2017 {published data only}

Plesingerova H, Librova Z, Plevova K, Libra A, Tichy B, Skuhrova Francova H, et al. COBL1, LPL and ZAP70 expression defines prognostic subgroups of chronic lymphocytic leukemia patients with high accuracy and correlates with IGHV mutational status. *Leukemia & Lymphoma* 2017;**58**(1):70-9.

Prokocimer 1985 {published data only}

Prokocimer M, Modan M, Lusky A, Hershko C. Multivariate analysis of prognostic factors in chronic lymphocytic leukemia. *Israel Journal of Medical Sciences* 1985;**21**(6):490-8.

Qin 2017 {published data only}

Qin SC, Xia Y, Miao Y, Zhu HY, Wu JZ, Fan L, et al. MYD88 mutations predict unfavorable prognosis in chronic lymphocytic leukemia patients with mutated IGHV gene. *Blood Cancer Journal* 2017;**7**(12):651.

Queiros 2015 {published data only}

Queiros AC, Villamor N, Clot G, Martinez-Trillos A, Kulis M, Navarro A, et al. A B-cell epigenetic signature defines three biologic subgroups of chronic lymphocytic leukemia with clinical impact. *Leukemia* 2015;**29**(3):598-605.

Rai 1975 {published data only}

Rai KR, Sawitsky A, Cronkite EP, Chanana AD, Levy RN, Pasternack BS. Clinical staging of chronic lymphocytic leukemia. *Blood* 1975;**46**(2):219-34.

Rai 1990 {published data only}

Rai KR, Han T. Prognostic factors and clinical staging in chronic lymphocytic leukemia. *Hematology/Oncology Clinics of North America* 1990;**4**(2):447-56.

Raponi 2018 {published data only}

Raponi S, Del Giudice I, Marinelli M, Wang J, Cafforio L, Ilari C, et al. Genetic landscape of ultra-stable chronic lymphocytic leukemia patients. *Annals of Oncology* 2018;**29**(4):966-72.

Resegotti 1989 {published data only}

Resegotti L, Levis A, Decrescenzo A, Ficara F, Pini M. Significance and independence of the immunologic, histological, and clinical features as prognostic factors in chronic lymphatic leukemia. *Haematologica* 1989;**74**(5 Suppl):368-79.

Rissiek 2014 {published data only}

Rissiek A, Schulze C, Bacher U, Schieferdecker A, Thiele B, Jacholkowski A, et al. Multidimensional scaling analysis identifies pathological and prognostically relevant profiles of circulating T-cells in chronic lymphocytic leukemia. *International Journal of Cancer* 2014;**135**(10):2370-9.

Rodriguez 2007 {published data only}

Rodriguez A, Villuendas R, Yanez L, Gomez ME, Diaz R, Pollan M, et al. Molecular heterogeneity in chronic lymphocytic leukemia is dependent on BCR signaling: clinical correlation. *Leukemia* 2007;**21**(9):1984-91.

Ronchetti 2016 {published data only}

Ronchetti D, Manzoni M, Agnelli L, Vinci C, Fabris S, Cutrona G, et al. lncRNA profiling in early-stage chronic lymphocytic leukemia

identifies transcriptional fingerprints with relevance in clinical outcome. *Blood Cancer Journal* 2016;**6**(9):e468.

Rossi 1986 {published data only}

Rossi PL, Alfano G, D'Onofrio G, Menichella G, Mango G. Depletion lymphocytopheresis in chronic lymphocytic leukemias: criteria for predicting which patients will respond to treatment. *International Journal of Artificial Organs* 1986;**9**(1):59-62.

Rossi 2010a {published data only}

Rossi FM, Del Principe MI, Rossi D, Irno Consalvo M, Luciano F, Zucchetto A, et al. Prognostic impact of ZAP-70 expression in chronic lymphocytic leukemia: mean fluorescence intensity T/B ratio versus percentage of positive cells. *Journal of Translational Medicine* 2010;**8**:23.

Rossi 2010b {published data only}

Rossi S, Shimizu M, Barbarotto E, Nicoloso MS, Dimitri F, Sampath D, et al. MicroRNA fingerprinting of CLL patients with chromosome 17p deletion identify a miR-21 score that stratifies early survival. *Blood* 2010;**116**(6):945-52.

Rossi 2011 {published data only}

Rossi D, Spina V, Deambrogi C, Rasi S, Laurenti L, Stamatopoulos K, et al. The genetics of Richter syndrome reveals disease heterogeneity and predicts survival after transformation. *Blood* 2011;**117**(12):3391-401.

Rozman 1979 {published data only}

Rozman C, Montserrat Costa E, Morey M, Aranalde JM, Feliu E, Granena A, et al. Chronic lymphoid leukemia. Survival in relation to clinical stages. Statistical analysis of 95 cases. *Medicina Clinica* 1979;**72**(7):265-71.

Salomon-Nguyen 1995 {published data only}

Salomon-Nguyen F, Valensi F, Merle-Beral H, Flandrin G. A scoring system for the classification of CD5-B CLL versus CD5+ B CLL and B PLL. *Leukemia & Lymphoma* 1995;**16**(5-6):445-50.

Santoro 1979 {published data only}

Santoro A, Musumeci R, Rilke F, Franchi F, Valagussa P, Bajetta E, et al. Clinical classification and survival in chronic lymphocyte leukemia. *Tumori* 1979;**65**(1):39-49.

Sarmiento 2002 {published data only}

Sarmiento MA, Palacios MF, Scolnik MP, Ramirez FR, Stanganelli C, Cabrera J, et al. Evolution of chronic lymphocytic leukemia: predictive value of immunophenotype, soluble CD23 and morphology. *Medicina* 2002;**62**(4):305-12.

Savvopoulos 2016 {published data only}

Savvopoulos S, Misener R, Panoskatsis N, Pistikopoulos EN, Mantalaris A. A personalized framework for dynamic modeling of disease trajectories in chronic lymphocytic leukemia. *IEEE Transactions on Biomedical Engineering* 2016;**63**(11):2396-404.

Scolozzi 1981 {published data only}

Scolozzi R, Boccafogli A, Romanini D, Ramazzina E, Careri C, Guidoboni CA. Prognosis of chronic lymphatic leukemia.

Retrospective study of the survival period 47 patients. *Minerva Medica* 1981;**72**(4-5):205-11.

Shanafelt 2010 {published data only}

Shanafelt TD, Rabe KG, Kay NE, Zent CS, Jelinek DF, Reinalda MS, et al. Age at diagnosis and the utility of prognostic testing in patients with chronic lymphocytic leukemia. *Cancer* 2010;**116**(20):4777-87.

Shanafelt 2017 {published data only}

Shanafelt TD, Parikh SA, Noseworthy PA, Goede V, Chaffee KG, Bahlo J, et al. Atrial fibrillation in patients with chronic lymphocytic leukemia (CLL). *Leukemia & Lymphoma* 2017;**58**(7):1630-9.

Spacek 2009 {published data only}

Spacek M, Pekova S, Bezdicikova L, Kozak T. Risk factors in chronic lymphocytic leukemia, validation of flow cytometric detection using RT-PCR and immunohistochemistry. *Transfuzie a Hematologie Dnes* 2009;**15**(2):91-6.

Stamatopoulos 2009 {published data only}

Stamatopoulos B, Meuleman N, Haibe-Kains B, Saussoy P, Van Den Neste E, Michaux L, et al. MicroRNA-29c and microRNA-223 down-regulation has in vivo significance in chronic lymphocytic leukemia and improves disease risk stratification. *Blood* 2009;**113**(21):5237-45.

Stamatopoulos 2017 {published data only}

Stamatopoulos B, Timbs A, Bruce D, Smith T, Clifford R, Robbe P, et al. Targeted deep sequencing reveals clinically relevant subclonal IgHV rearrangements in chronic lymphocytic leukemia. *Leukemia* 2017;**31**(4):837-45.

Strefford 2015 {published data only}

Strefford JC, Kadalayil L, Forster J, Rose-Zerilli MJ, Parker A, Lin TT, et al. Telomere length predicts progression and overall survival in chronic lymphocytic leukemia: data from the UK LRF CLL4 trial. *Leukemia* 2015;**10**(10):217.

Szymczyk 2018 {published data only}

Szymczyk A, Chocholska S, Macheta A, Szczepanek D, Hus M, Podhorecka M. Assessment of micro RNAs expression in leukemic cells as prognostic markers in chronic lymphocytic leukemia: micro RNAs can predict survival in a course of the disease. *Oncotarget* 2018;**9**(27):19136-46.

Tallarico 2018 {published data only}

Tallarico M, Foster JC, Seisler D, Lafky JM, Hurria A, Jatoi A, et al. Frequency and impact of grade three or four toxicities of novel agents on outcomes of older patients with chronic lymphocytic leukemia and non-Hodgkin lymphoma (alliance A151611). *Journal of Geriatric Oncology* 2018;**9**:321-8.

Tobin 2005a {published data only}

Tobin G. The immunoglobulin genes and chronic lymphocytic leukemia (CLL). *Upsala Journal of Medical Sciences* 2005;**110**(2):97-113.

Tobin 2005b {published data only}

Tobin G, Thunberg U, Laurell A, Karlsson K, Aleskog A, Willander K, et al. Patients with chronic lymphocytic leukemia with mutated VH genes presenting with Binet stage B or C form a subgroup with a poor outcome. *Haematologica* 2005;**90**(4):465-9.

Vallat 2013 {published data only}

Vallat L, Kemper CA, Jung N, Maumy-Bertrand M, Bertrand F, Meyer N, et al. Reverse-engineering the genetic circuitry of a cancer cell with predicted intervention in chronic lymphocytic leukemia. *Proceedings of the National Academy of Sciences of the United States of America* 2013;**110**(2):459-64.

Van Damme 2012 {published data only}

Van Damme M, Crompot E, Meuleman N, Mineur P, Bron D, Lagneaux L, et al. HDAC isoenzyme expression is deregulated in chronic lymphocytic leukemia B-cells and has a complex prognostic significance. *Epigenetics* 2012;**7**(12):1403-12.

Velardi 1980 {published data only}

Velardi A, Spinozzi F, Siracusa A, Colozza M, Aversa F, Rambotti P, et al. Clinical staging of chronic lymphocytic leukaemia and its relationship to survival. *Haematologica* 1980;**65**(1):82-95.

Vojdeman 2017 {published data only}

Vojdeman FJ, Herman SE, Kirkby N, Wiestner A, Van T' Veer MB, Tjonnfjord GE, et al. Soluble CD52 is an indicator of disease activity in chronic lymphocytic leukemia. *Leukemia & Lymphoma* 2017;**58**(10):2356-62.

Vural 2014 {published data only}

Vural F, Karaca E, Soyer N, Gunduz C, Sahin F, Kosova B, et al. Comparison of CD38, ZAP70 and hTERT expression with known prognostic markers in patients with chronic lymphocytic leukemia during five-year follow-up period. *Uluslararası Hematoloji-Onkoloji Dergisi (UHOD)* 2014;**24**(3):179-84.

Weinberg 2007 {published data only}

Weinberg JB, Volkheimer AD, Chen Y, Beasley BE, Jiang N, Lanasa MC, et al. Clinical and molecular predictors of disease severity and survival in chronic lymphocytic leukemia. *American Journal of Hematology* 2007;**82**(12):1063-70 [Erratum appears in American Journal of Hematology 2008;83(1):90].

Weiss 2011 {published data only}

Weiss L, Melchardt T, Egle A, Grabmer C, Greil R, Tinhofer I. Regulatory T cells predict the time to initial treatment in early stage chronic lymphocytic leukemia. *Cancer* 2011;**117**(10):2163-9.

Wierda 2003 {published data only}

Wierda WG, Johnson MM, Do KA, Manshouri T, Dey A, O'Brien S, et al. Plasma interleukin 8 level predicts for survival in chronic lymphocytic leukaemia. *British Journal of Haematology* 2003;**120**(3):452-6.

Winkler 2010 {published data only}

Winkler D, Schneider C, Zucknick M, Bogelein D, Schulze K, Zenz T, et al. Protein expression analysis of chronic lymphocytic

leukemia defines the effect of genetic aberrations and uncovers a correlation of CDK4, P27 and P53 with hierarchical risk. *Haematologica* 2010;**95**(11):1880-8.

Wu 2010 {published data only}

Wu T, Li ZJ, Wang YF, Qiu LG. Applicability of different clinical staging systems in chronic lymphocytic leukemia for Chinese patients. *Journal of Leukemia and Lymphoma* 2010;**19**(3):136-9.

Zengin 1997 {published data only}

Zengin N, Kars A, Kansu E, Ozdemir O, Barista I, Gullu I, et al. Comparison of Rai and Binet classifications in chronic lymphocytic leukemia. *Hematology* 1997;**2**(2):125-9.

Zenz 2009 {published data only}

Zenz T, Benner A, Dührsen U, Durig J, Dohner H, Siffert W, et al. BCL2-938C>A polymorphism and disease progression in chronic lymphocytic leukemia. *Leukemia & Lymphoma* 2009;**50**(11):1837-42.

Zucchetto 2006 {published data only}

Zucchetto A, Bomben R, Dal Bo M, Sonogo P, Nanni P, Rupolo M, et al. A scoring system based on the expression of six surface molecules allows the identification of three prognostic risk groups in B-cell chronic lymphocytic leukemia. *Journal of Cellular Physiology* 2006;**207**(2):354-63.

References to studies awaiting assessment

Antic 2011 {published data only}

Antic D, Mihaljevic B, Cokic V, Fekete MD, Djurasevic TK, Pavlovic S, et al. Patients with early stage chronic lymphocytic leukemia: new risk stratification based on molecular profiling. *Leukemia & Lymphoma* 2011;**52**(7):1394-7.

Baumann 2014 {published data only}

Baumann T, Delgado J, Santacruz R, Martínez-Trillos A, Royo C, Navarro A, et al. Chronic lymphocytic leukemia in the elderly: clinico-biological features, outcomes, and proposal of a prognostic model. *Haematologica* 2014;**99**(10):1599-604.

Bulian 2011 {published data only}

Bulian P, Tarnani M, Rossi D, Forconi F, Del Poeta G, Bertoni F, et al. Multicentre validation of a prognostic index for overall survival in chronic lymphocytic leukaemia. *Hematological Oncology* 2011;**29**(2):91-9.

Bulian 2012 {published data only}

Bulian P, Rossi D, Forconi F, Del Poeta G, Bertoni F, Zucca E, et al. IGHV gene mutational status and 17p deletion are independent molecular predictors in a comprehensive clinical-biological prognostic model for overall survival prediction in chronic lymphocytic leukemia. *Journal of Translational Medicine* 2012;**10**:18.

Cavallini 2017 {published data only}

Cavallini C, Visco C, Putta S, Rossi D, Mimiola E, Purvis N, et al. Integration of B-cell receptor-induced ERK1/2 phosphorylation and mutations of SF3B1 gene refines prognosis in treatment-naïve chronic lymphocytic leukemia. *Haematologica* 2017;**102**(4):e144-7.

Del Giudice 2005 {published data only}

Del Giudice I, Morilla A, Osuji N, Matutes E, Morilla R, Burford A, et al. Zeta-chain associated protein 70 and CD38 combined predict the time to first treatment in patients with chronic lymphocytic leukemia. *Cancer* 2005;**104**(10):2124-32.

Friedrichs 2011 {published data only}

Friedrichs B, Siegel S, Reimer R, Barsoum A, Coggin J Jr, Kabelitz D, et al. High expression of the immature laminin receptor protein correlates with mutated IGVH status and predicts a favorable prognosis in chronic lymphocytic leukemia. *Leukemia Research* 2011;**35**(6):721-9.

Furundarena 1994 {published data only}

Furundarena Salsamendi JR, Martinez D, Navarro E, Aguinaco R, Vidal MJ, Lasa R, et al. Chronic lymphocytic leukemia: a retrospective analysis of diagnostic characteristics and survival predictive factors in 150 patients. *Biología y Clínica Hematológica* 1994;**16**(2):64-70.

Gentile 2009 {published data only}

Gentile M, Cutrona G, Neri A, Molica S, Ferrarini M, Morabito F. Predictive value of beta2-microglobulin (beta2-m) levels in chronic lymphocytic leukemia since Binet A stages. *Haematologica* 2009;**94**(6):887-8.

Haferlach 2010 {published data only}

Haferlach C, Dicker F, Weiss T, Schnittger S, Beck C, Grote-Metke A, et al. Toward a comprehensive prognostic scoring system in chronic lymphocytic leukemia based on a combination of genetic parameters. *Genes, Chromosomes & Cancer* 2010;**49**(9):851-9.

Jarque 1991 {published data only}

Jarque I, Sanz G, Gomis F, Martinez J, Martin G, De la Rubia J, et al. Chronic lymphatic leukemia. II. Analysis of prognostic factors and development of survival predicting models. Study of 187 patients. *Sangre* 1991;**36**(4):285-94.

Lee 1987 {published data only}

Lee JS, Dixon DO, Kantarjian HM, Keating MJ, Talpaz M. Prognosis of chronic lymphocytic leukemia: a multivariate regression analysis of 325 untreated patients. *Blood* 1987;**69**(3):929-36.

Leotard 2000 {published data only}

Leotard S, Chastang C, Travade P, Jaudon MC, Tournilhac O, Baudet S, et al. Prognostic relevance of a scoring system based on clinical and biological parameters in early chronic lymphocytic leukemia. *Hematology Journal* 2000;**1**(5):301-6.

Letestu 2010 {published data only}

Letestu R, Levy V, Eclache V, Baran-Marszak F, Vaur D, Naguib D, et al. Prognosis of Binet stage A chronic lymphocytic leukemia patients: the strength of routine parameters. *Blood* 2010;**116**(22):4588-90.

Li 2017b {published data only}

Li H, Yi SH, Xiong WJ, Liu HM, Lyu R, Wang TY, et al. Chronic lymphocytic leukemia prognostic index: a new integrated scoring system to predict the Time to First Treatment in Chinese

patients with chronic lymphocytic leukemia. *Chinese Medical Journal* 2017;**130**(2):135-42.

Liang 2018 {published data only}

Liang JH, Gao R, Dai JC, Gale RP, Li W, Fan L, et al. The prognostic role of HBV infection in chronic lymphocytic leukemia. *Journal of Cancer Research and Clinical Oncology* 2018;**144**:1309-15.

Metze 2000 {published data only}

Metze K, Lobo AM, Lorand-Metze I. Nucleolus organizer regions (AgNORs) and total tumor mass are independent prognostic parameters for treatment-free period in chronic lymphocytic leukemia. *International Journal of Cancer* 2000;**89**(5):440-3.

Miao 2019 {published data only}

Miao Y, Zou YX, Gu DL, Zhu HC, Zhu HY, Wang L, et al. SF3B1 mutation predicts unfavorable treatment-free survival in Chinese chronic lymphocytic leukemia patients. *Annals of Translational Medicine* 2019;**7**(8):176.

Molica 1990 {published data only}

Molica S, Reverter JC, Alberti A, Montserrat E. Timing of diagnosis and lymphocyte accumulation patterns in chronic lymphocytic leukemia: analysis of their clinical significance. *European Journal of Haematology* 1990;**44**(5):277-81.

Molica 2010 {published data only}

Molica S, Digiesi G, Battaglia C, Cutrona G, Antenucci A, Molica M, et al. Ba& serum level predicts time to first treatment in early chronic lymphocytic leukemia. *European Journal of Haematology* 2010;**85**(4):314-20.

Molica 2015 {published data only}

Molica S, Giannarelli D, Levato L, Gentile M, Mirabelli R, Morabito F. Do biologic parameters affect the time to first treatment of clinical monoclonal B-cell lymphocytosis and chronic lymphocytic leukemia Rai stage 0? Results of a prospective analysis. *Clinical Lymphoma, Myeloma and Leukemia* 2015;**15**(3):e55-60.

Molica 2019 {published data only}

Molica S, Giannarelli D, Levato L, Mirabelli R, Levato D, Lentini M, et al. A simple score based on geriatric assessment predicts survival in elderly newly diagnosed chronic lymphocytic leukemia patients. *Leukemia & Lymphoma* 2019;**60**(3):845-7.

Morabito 2011 {published data only}

Morabito F, De Filippi R, Laurenti L, Zirlik K, Recchia AG, Gentile M, et al. The cumulative amount of serum-free light chain is a strong prognosticator in chronic lymphocytic leukemia. *Blood* 2011;**118**(24):6353-61.

Pepper 2012 {published data only}

Pepper C, Majid A, Lin TT, Hewamana S, Pratt G, Walewska R, et al. Defining the prognosis of early stage chronic lymphocytic leukaemia patients. *British Journal of Haematology* 2012;**156**(4):499-507.

Qin 2018 {published data only}

Qin S, Fan L, Liang J, Gale R, Miao Y, Wu Y, et al. Definition of disease-progression risk stratification in untreated chronic lymphocytic leukemia using combined clinical, molecular and virological variables. *Hematological Oncology* 2018;**36**(4):656-62.

Rozman 1982 {published data only}

Rozman C, Montserrat E, Feliu E, Granena A, Marin P, Nomdedeu B, et al. Prognosis of chronic lymphocytic leukemia: a multivariate survival analysis of 150 cases. *Blood* 1982;**59**(5):1001-5.

Rozman 1984 {published data only}

Rozman C, Montserrat E, Rodriguez-Fernandez JM, Ayats R, Vallespi T, Parody R, et al. Bone marrow histologic pattern - the best single prognostic parameter in chronic lymphocytic leukemia: a multivariate survival analysis of 329 cases. *Blood* 1984;**64**(3):642-8.

Schweighofer 2011 {published data only}

Schweighofer CD, Coombes KR, Barron LL, Diao L, Newman RJ, Ferrajoli A, et al. A two-gene signature, SKI and SLAMF1, predicts time-to-treatment in previously untreated patients with chronic lymphocytic leukemia. *PLOS One* 2011;**6**(12):e28277.

Stamatopoulos 2010 {published data only}

Stamatopoulos B, Meuleman N, De Bruyn C, Pieters K, Anthoine G, Mineur P, et al. A molecular score by quantitative PCR as a new prognostic tool at diagnosis for chronic lymphocytic leukemia patients. *PLOS One* 2010;**5**(9):e12780.

Tsimberidou 2007 {published data only}

Tsimberidou AM, Wen S, O'Brien S, McLaughlin P, Wierda WG, Ferrajoli A, et al. Assessment of chronic lymphocytic leukemia and small lymphocytic lymphoma by absolute lymphocyte counts in 2,126 patients: 20 years of experience at the University of Texas M.D. Anderson Cancer Center. *Journal of Clinical Oncology* 2007;**25**(29):4648-56.

Vetro 2018 {published data only}

Vetro C, Haferlach T, Jeromin S, Stengel A, Zenger M, Nadarajah N, et al. Identification of prognostic parameters in CLL with no abnormalities detected by chromosome banding and FISH analyses. *British Journal of Haematology* 2018;**183**(1):47-59.

Visentin 2015 {published data only}

Visentin A, Facco M, Frezzato F, Castelli M, Trimarco V, Martini V, et al. Integrated CLL scoring system, a new and simple index to predict time to treatment and overall survival in patients with chronic lymphocytic leukemia. *Clinical Lymphoma, Myeloma & Leukemia* 2015;**15**(10):612-20.e5.

Wierda 2009 {published data only}

Wierda WG, O'Brien S, Wang X, Faderl S, Ferrajoli A, Do K-A, et al. Characteristics associated with important clinical end points in patients with chronic lymphocytic leukemia at initial treatment. *Journal of Clinical Oncology* 2009;**27**(10):1637-43.

References to ongoing studies

NCT00275054 {published data only}

NCT00275054. Rituximab, fludarabine, and cyclophosphamide or observation alone in treating patients with stage 0, stage I, or stage II chronic lymphocytic leukemia. clinicaltrials.gov/ct2/show/NCT00275054 (first received 11 January 2006).

NCT03436524 {published data only}

NCT03436524. A prognostic tool for early stage CLL. clinicaltrials.gov/ct2/show/NCT03436524 (first received 19 February 2018).

Additional references

Alba 2017

Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, et al. Discrimination and calibration of clinical prediction models: users guides to the medical literature. *JAMA* 2017;**318**(14):1377-84.

Bouwmeester 2012

Bouwmeester W, Zuihtho 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.

Cheson 1996

Cheson BD, Bennett JM, Grever M, Kay N, Keating MJ, O'Brien S, et al. National Cancer Institute-sponsored Working Group guidelines for chronic lymphocytic leukemia: revised guidelines for diagnosis and treatment. *Blood* 1996;**87**(12):4990-7.

Chiorazzi 2005

Chiorazzi N, Rai KR, Ferrarini M. Chronic lymphocytic leukemia. *New England Journal of Medicine* 2005;**352**(8):804-15. [PMID: 15728813]

Collins 2015

Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD). *Annals of Internal Medicine* 2015;**162**(1):735-6.

Collins 2016

Collins GS, Ma J, Gerry S, Ohuma E, Odondi L, Trivella M, et al. Risk prediction models in perioperative medicine: methodological considerations. *Current Anesthesiology Reports* 2016;**6**(3):267-75.

Debray 2012

Debray TP, Ko&ijberg H, Vergouwe Y, Moons KG, Steyerberg EW. Aggregating published prediction models with individual participant data: a comparison of different approaches. *Statistics in Medicine* 2012;**31**(23):2697-712.

Debray 2014

Debray TP, Ko&ijberg H, Nieboer D, Vergouwe Y, Steyerberg EW, Moons KG. Meta-analysis and aggregation of multiple published prediction models. *Statistics in Medicine* 2014;**33**(14):2341-62.

Debray 2017

Debray TP, Damen JA, Snell KI, Ensor J, HooL L, Reitsma JB, et al. A guide to systematic review and meta-analysis of prediction model performance. *BMJ* 2017;**356**:i6460.

Debray 2018a

Debray TP, Damen JA, Riley RD, Snell K, Reitsma JB, HooL L, et al. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. *Statistical Methods in Medical Research* 2018;**28**(9):2768-76.

Debray 2018b [Computer program]

metamisc: Diagnostic and prognostic meta-Analysis. R package version 0.1.9. Debray T, de Jong V. Vienna, Austria: R Foundation for Statistical Computing, 2018. Available at www.R-project.org.

Dicker 2009

Dicker F, Herholz H, Schnittger S, Nakao A, Patten N, Wu L, et al. The detection of TP53 mutations in chronic lymphocytic leukemia independently predicts rapid disease progression and is highly correlated with a complex aberrant karyotype. *Leukemia* 2009;**23**(1):117-24.

Döhner 2000

Döhner H, Stilgenbauer S, Benner A, Leupolt E, Krober A, Bullinger L, et al. Genomic aberrations and survival in chronic lymphocytic leukemia. *New England Journal of Medicine* 2000;**343**(26):1910-6. [PMID: 11136261]

Eichhorst 2015

Eichhorst B, Robak T, Montserrat E, Ghia P, Hillmen P, Hallek M, et al. ESMO Guidelines Committee. Chronic lymphocytic leukaemia: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology* 2015;**26**(Suppl 5):v78-84.

El Rouby 1993

El Rouby S, Thomas A, Costin D, Rosenberg CR, Potmesil M, Silber R, et al. p53 gene mutation in B-cell chronic lymphocytic leukemia is associated with drug resistance and is independent of MDR1/MDR3 gene expression. *Blood* 1993;**82**(11):3452-9. [PMID: 8241511]

Ferrer Lores 2016

Ferrer Lores B, Navarro Cubells B, Serrano Alcala A, Castillo Martin I, Teruel Casaus A, Ballester Martinez S, et al. Prognostic impact of the new CLL-IPi Index in a single center CLL Spanish cohort. *Blood* 2016;**128**(22):2019.

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**:e32844.

Goede 2012

Goede V, Hallek M. Optimal pharmacotherapeutic management of chronic lymphocytic leukaemia: considerations in the elderly. *Drugs and Aging* 2012;**28**(3):163-76.

Goede 2016

Goede V, Bahlo J, Kutsch N, Fischer K, Fink AM, Fingerle-Rowson G, et al. Evaluation of the International Prognostic Index for Chronic Lymphocytic Leukemia (CLL-IPi) in elderly patients with comorbidities: analysis of the CLL11 study population. *Blood* 2016;**128**(22):4401.

Hallek 2008

Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Döhner H, et al. International Workshop on Chronic Lymphocytic Leukemia. Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. *Blood* 2008;**111**(12):5446-56.

Hallek 2017

Hallek M. Chronic lymphocytic leukemia: 2017 update on diagnosis, risk stratification, and treatment. *American Journal of Hematology* 2017;**92**(9):946-65.

Herishanu 2013

Herishanu Y, Katz BZ, Lipsky A, Wiestner A. Biology of chronic lymphocytic leukemia in different microenvironments: clinical and therapeutic implications. *Hematology/Oncology Clinics of North America* 2013;**27**(2):173-206.

Heus 2018

Heus P, Damen JA, Pajouheshnia R, Scholten RJ, Reitsma JB, Collins GS, et al. Poor reporting of multivariable prediction model studies: towards a targeted implementation strategy of the TRIPOD statement. *BMC Medicine* 2018;**16**(1):120.

Howlader 2019

Howlader N, Noone AM, Krapcho M, Miller D, Brest A, Yu M, et al. National Cancer Institute. SEER Cancer Statistics Review, 1975-2016. www.seer.cancer.gov/archive/csr/1975_2016 (based on November 2018 SEER data submission, posted to the SEER web site, April 2019) (accessed May 2019).

Kay 2007

Kay NE, O'Brien SM, Pettitt AR, Stilgenbauer S. The role of prognostic factors in assessing 'high-risk' subgroups of patients with chronic lymphocytic leukemia. *Leukemia* 2007;**21**(9):1885-91. [PMID: 17568813]

Lefebvre 2019

Lefebvre C, Glanville J, Briscoe S, Littlewood A, Marshall C, Metzendorf M-I, 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. Available from www.training.cochrane.org/handbook.

Moher 2009

Moher D, Liberati A, Tetzla J, Altman DG, the PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Journal of Clinical Epidemiology* 2009;**62**(10):1006-12.

Molica 2018a

Molica S, Giannarelli D, Mirabelli R, Levato L, Shanafelt TD. Chronic lymphocytic leukemia international prognostic index (CLL-IPI) in patients receiving chemoimmuno or targeted therapy: a systematic review and meta-analysis. *Annals of Hematology* 2018;**97**(10):2005-8.

Molica 2018b

Molica S, Giannarelli D, Mirabelli R, Levato L, Kay NE, Shanafelt TD. Chronic lymphocytic leukemia international prognostic index: a systematic review and meta-analysis. *Blood* 2018;**131**(3):365-8.

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.

Moons 2015

Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): explanation and elaboration. *Annals of Internal Medicine* 2015;**162**(1):W1-W73.

Moons 2019

Moons KG, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Annals of Internal Medicine* 2019;**170**(1):W1-W33.

Newcombe 2006

Newcombe RG. Confidence intervals for an effect size measure based on the Mann-Whitney statistic. Part 2: asymptotic methods and evaluation. *Statistics in Medicine* 2006;**25**(4):559-73.

Pajouheshnia 2019

Pajouheshnia R, Groenwold RH, Peelen LM, Reitsma JB, Moons KG. When and how to use data from randomised trials to develop or validate prognostic models. *BMJ* 2019;**365**:l2154.

Peat 2014

Peat G, Riley RD, CroL 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**:e1001671.

Pflug 2014

Pflug N, Bahlo J, Shanafelt TD, Eichhorst BF, Bergmann MA, Elter T, et al. Development of a comprehensive prognostic index for patients with chronic lymphocytic leukemia. *Blood* 2014;**124**(1):49-62. [PMID: 24797299]

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.

Riley 2019

Riley RD, Van der Windt DA, CroL P, Moons KGM. Prognosis Research in Healthcare. Concepts, Methods, and Impact. Oxford: Oxford University Press, 2019.

Shanafelt 2004

Shanafelt TD, Geyer SM, Kay NE. Prognosis at diagnosis: integrating molecular biologic insights into clinical practice for patients with CLL. *Blood* 2004;**103**(4):1202-10. [PMID: 14576043]

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.

Steyerberg 2014

Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *European Heart Journal* 2014;**35**(29):1925-31.

Stilgenbauer 2014

Stilgenbauer S, Schnaiter A, Paschka P, Zenz T, Rossi M, Döhner K, et al. Gene mutations and treatment outcome in chronic lymphocytic leukemia: results from the CLL8 trial. *Blood* 2014;**123**(21):3247-54.

Van Calster 2019

Van Calster B, McLernon DJ, Van Smeden M, Wynants L, Steyerberg EW, Bossuyt P, et al. Calibration: the Achilles heel of predictive analytics. *BMC Medicine* 2019;**17**(1):230.

Viechtbauer 2010

Viechtbauer W. Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software* 2010;**36**(3):1-48.

Watson 2008

Watson L, Wyld P, Catovsky D. Disease burden of chronic lymphocytic leukaemia within the European Union. *European Journal of Haematology* 2008;**81**(4):253-8.

Wierda 2011

Wierda WG, O'Brien S, Wang X, Faderl S, Ferrajoli A, Do KA, et al. Multivariable model for time to first treatment in patients with chronic lymphocytic leukemia. *Journal of Clinical Oncology* 2011;**29**(31):4088-95. [PMID: 21969505]

Wolff 2019

Wolff RF, Moons KG, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine* 2019;**170**(1):51-8.

Zaja 2013

Zaja F, Mian M, Volpetti S, Visco C, Sissa C, Nichele I, et al. Bendamustine in chronic lymphocytic leukemia: outcome according to different clinical and biological prognostic

factors in the everyday clinical practice. *American Journal of Hematology* 2013;**88**(11):955-60.

Zenz 2011

Zenz T, Mertens D, Dohner H, Stilgenbauer S. Importance of genetics in chronic lymphocytic leukemia. *Blood Reviews* 2011;**25**(3):131-7. [PMID: 21435757]

References to other published versions of this review

Skoetz 2016

Skoetz N, Trivella M, Kreuzer KA, Collins G, Köhler N, Wol M, et al. Prognostic models for chronic lymphocytic leukaemia: an exemplar systematic review and meta-analysis. *Cochrane Database of Systematic Reviews* 2016, Issue 1. Art. No: CD012022. [DOI: [10.1002/14651858.CD012022](https://doi.org/10.1002/14651858.CD012022)]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies [ordered by study ID]

Baliakas D - Baliakas 2019 (multicentre)

Study characteristics

General information	<u>Model and type of study</u>
	- Baliakas 2019 (development cohort) - Development study
	<u>Secondary citations</u>
	not applicable
	<u>Language of publication</u>
	English
	<u>Study design</u>
	multicentre, retrospective cohort
	<u>Follow-up time</u>
	median 7.1 years (range 0.1 to 33.1 years)
Participants	<u>Number of included persons in the cohort</u>
	2366 persons* *analysis included only 1900 persons with Binet A, divided into M-CLL (n = 1224) and U-CLL (n = 676)
	<u>Setting</u>
	10 European institutions, multicentre study
	<u>Recruitment period</u>
	not reported
	<u>Age (in years)</u>
	median 64.3 years (range: 22-92 years)
	<u>Sex</u>
	61% male
	<u>Stages of disease</u>
	Binet A: 80%; Binet B: 12%; Binet C: 8%

Baliakas D - Baliakas 2019 (multicentre) *(Continued)*

Treatment

- not reported

Inclusion criteria

- diagnosed CLL, available immunogenetic data

Exclusion criteria

- not reported

Predictors

Number of candidate predictors

- univariable analysis: 15
- multivariable analysis: 15

List of predictors in final model (including cut-points for dichotomised factors)

- M-CLL: TP53 abnormality, trisomy12, subset #2 (IGHV3-21/IGLV3-21 BcR IG)
- U-CLL: TP53 abnormality, del11q, gender

Timing of predictor measurement

- at diagnosis

Outcome(s)

Primary outcome in study

- TTFT: evaluated from the diagnostic date until the date of initial treatment; untreated cases were censored at the time of last follow-up.

Additional outcome(s)

- NA

Missing data

Participants with any missing data?

- yes
- M-CLL: 306, U-CLL: 288

If yes, how was missing data handled?

- excluded from multivariable analysis if data were missing for one of the factors included in the model

Analysis

Number of participants and number of events (specific time points where reported)

- TTFT: 1900 persons, number of events not reported

Predictor selection method

- selection of prognostic factors based on univariable analysis

Statistical method

- decision tree based on binary recursive partitioning and subsequent application of an amalgamation algorithm

Simplification of model?

- yes

Performance measures reported?

- Calibration: not reported
- Discrimination: reported (c-statistic)

Baliakas D - Baliakas 2019 (multicentre) (Continued)

Creation of risk groups?

- yes
- M-CLL: 4 risk groups; i) very high risk: Binet C with identical 5- and 10-year TP of 92%; ii) high risk: Binet B, 5y-TP and 10y-TP: 64% and 84%, respectively; iii) intermediate risk: Binet A with one of the following: TP53abn and/or +12 and/or subset #2 membership, 5y-TP and 10y-TP: 40% and 55%, respectively (of note, among 18 non-censored cases with no treatment indication for more than 10 years after diagnosis, 5 (30%) carried TP53abn); and iv) low risk: non-TP53abn/+12/subset #2 Binet A, 5y-TP and 10y-TP: 12% and 25%, respectively
- U-CLL: 5 risk groups; i) very high risk: Binet C with 5- and 10-year TP of 100%; ii) high risk: Binet B, 5y-TP and 10y-TP: 90% and 100%, respectively; iii) intermediate risk: Binet A with one of the following: TP53abn and/or SF3B1mut and/or del(11q), 5y-TP and 10y-TP: 78% and 98%, respectively; iv) low risk: nonTP53abn/SF3B1mut/del(11q) male Binet A, 5y-TP and 10yTP: 65% and 85%, respectively; and v) very low risk: non-TP53abn/SF3B1mut/del(11q) female Binet A, 5y-TP and 10y-TP: 45% and 65%, respectively

PROBAST: Applicability

Domain 1: Participant selection

- high - sample not representative of entire CLL population as only participants with disease stage Binet A were analysed

Domain 2: Predictors

- low

Domain 3: Outcome

- low

Notes

Funding and conflict of interest

- "... the Swedish Cancer Society, the Swedish Research Council, the Knut and Alice Wallenberg Foundation, Karolinska Institutet, Stockholm, the Lion's Cancer Research Foundation, Uppsala, the Marcus Borgström Foundation and Selander's Foundation, Uppsala; H2020 "AEGLE, An analytics framework for integrated and personalized healthcare services in Europe" by the EU; H2020 "MEDGENET, Medical Genomics and Epigenomics Network" (No.692298) by the EU; H2020 "CLLassify, Innovative risk assessment for individualizing treatment in chronic lymphocytic leukemia" (No.702714) by the EU; Associazione Italiana per la Ricerca sul Cancro AIRC Investigator grants #20246, and Special Program Molecular Clinical Oncology AIRC 5 per mille #9965; Progetti di Rilevante Interesse Nazionale (PRIN) #2015ZMRFEA, MIUR, Rome, Italy; TRANSCAN-179 NOVEL JTC 2016; project CEITEC 2020 (LQ1601) by MEYS-CZ, project AZV-MH-CZ 15-30015A-4/2015; JCS was funded by Bloodwise (11052, 12036), the Kay Kendall Leukaemia Fund (873), Cancer Research UK (C34999/A18087, ECMC C24563/A15581), Wessex Medical Research and the Bournemouth Leukaemia Fund; Special Program Molecular Clinical Oncology 5 x 1000 No. 10007, Associazione Italiana per la Ricerca sul Cancro Foundation Milan, Italy; Progetto Ricerca Finaliz"
- "KS and PG received research support from Janssen Pharmaceuticals and Gilead Sciences. KS, PG and AH received research support from Novartis SA, Abbvie and Roche Hellas. The other authors declare no conflict of interests in relation to the present study."

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	No	Missing data used as a reason for exclusion: only participants: (quote) "for whom immunogenetic data was available were included in this multicentre retrospective study".
Domain 2: Predictors	Yes	Detailed description of predictor assessment in the appendix
Domain 3: Outcome	Yes	Predefined outcome definition

Baliakas D - Baliakas 2019 (multicentre) (Continued)

Domain 4: Analysis	No	Participants with missing data were excluded from multivariable analyses: (quote) "we considered only those cases with available data for all the factors included in the model (n = 918 for M-CLL and n = 384 for U-CLL)"; univariable selection of predictors; assumptions of Cox proportional hazards model were checked.
Overall judgement	No	

Baliakas V - Baliakas 2019 (MLL + Scan.)

Study characteristics

General information	<u>Model and type of study</u> - Baliakas 2019 (MLL + Scan cohort) - Validation study <u>Secondary citations</u> - not applicable <u>Language of publication</u> - English <u>Study design</u> - retrospective cohort <u>Follow-up time</u> - not reported
Participants	<u>Number of included persons in the cohort</u> 649 persons <u>Setting</u> - Munich Leukemia Laboratory (n = 508) - Scandinavian population-based study (n = 141) <u>Recruitment period</u> - not reported <u>Age (in years)</u> - median 63.6 years (range: 29-89 years) <u>Sex</u> 62% male <u>Stages of disease</u> - Binet A: 100% <u>Treatment</u> - not reported

Baliakas V - Baliakas 2019 (MLL + Scan.) (Continued)

	<u>Inclusion criteria</u> • CLL, disease stage Binet A <u>Exclusion criteria</u> • not reported
Predictors	<u>List of predictors used for validation (and changes between original predictors and predictors in validation study)</u> • M-CLL: TP53 abnormality, trisomy12, subset #2 • U-CLL: TP53 abnormality and/or SF3B1 mutation, del(11q), gender <u>Timing of predictor measurement</u> • at diagnosis
Outcome(s)	<u>Primary outcome in study</u> • TTFT was evaluated from the diagnostic date until the date of initial treatment, untreated cases were censored at the time of last follow-up. <u>Additional outcome(s)</u> • NA <u>Outcome in model development</u> • TTFT
Missing data	<u>Participants with any missing data?</u> • no <u>If yes, how was missing data handled?</u> • not applicable
Analysis	<u>Number of participants and number of events (specific time points where reported)</u> • time-to-first-treatment: 649 persons, number of events not reported <u>Which model was used?</u> • classification into risk groups <u>Was the model updated?</u> • no <u>Performance measures reported?</u> • Calibration: not reported • Discrimination: not reported <u>Creation of risk groups?</u> • no
PROBAST: Applicability	<u>Domain 1: Participant selection</u> • unclear - neither eligibility criteria nor recruitment period reported <u>Domain 2: Predictors</u>

Baliakas V - Baliakas 2019 (MLL + Scan.) (Continued)

- low

Domain 3: Outcome

- low

Notes

Funding and conflict of interest

- "... the Swedish Cancer Society, the Swedish Research Council, the Knut and Alice Wallenberg Foundation, Karolinska Institutet, Stockholm, the Lion's Cancer Research Foundation, Uppsala, the Marcus Borgström Foundation and Selander's Foundation, Uppsala; H2020 "AEGLE, An analytics framework for integrated and personalized healthcare services in Europe" by the EU; H2020 "MEDGENET, Medical Genomics and Epigenomics Network" (No.692298) by the EU; H2020 "CLLassify, Innovative risk assessment for individualizing treatment in chronic lymphocytic leukemia" (No.702714) by the EU; Associazione Italiana per la Ricerca sul Cancro AIRC Investigator grants #20246, and Special Program Molecular Clinical Oncology AIRC 5 per mille #9965; Progetti di Rilevante Interesse Nazionale (PRIN) #2015ZMRFEA, MIUR, Rome, Italy; TRANSCAN-179 NOVEL JTC 2016; project CEITEC 2020 (LQ1601) by MEYS-CZ, project AZV-MH-CZ 15-30015A-4/2015; JCS was funded by Bloodwise (11052, 12036), the Kay Kendall Leukaemia Fund (873), Cancer Research UK (C34999/A18087, ECMC C24563/A15581), Wessex Medical Research and the Bournemouth Leukaemia Fund; Special Program Molecular Clinical Oncology 5 x 1000 No. 10007, Associazione Italiana per la Ricerca sul Cancro Foundation Milan, Italy; Progetto Ricerca Finaliz"
- "KS and PG received research support from Janssen Pharmaceuticals and Gilead Sciences. KS, PG and AH received research support from Novartis SA, Abbvie and Roche Hellas. The other authors declare no conflict of interests in relation to the present study."

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	Eligibility criteria not reported; not stated whether participants were excluded based on missing values; recruitment period not reported
Domain 2: Predictors	Unclear	Lack of information about predictor assessment, especially since the validation cohort consisted of two separate cohorts which were merged for analysis
Domain 3: Outcome	Unclear	Outcome definition not reported
Domain 4: Analysis	No	Performance measures (calibration or discrimination) not reported
Overall judgement	No	

Barcelona-Brno D - Delgado 2017 (Barcelona cohort)

Study characteristics

General information

Model and type of study

- Delgado 2017 (development cohort (Barcelona))
- Development study

Secondary citations

- not applicable

Language of publication

- English

Barcelona-Brno D - Delgado 2017 (Barcelona cohort) (Continued)

Study design

- retrospective cohort

Follow-up time

- median 99.6 months (range 1-456 months)

Participants

Number of included persons in the cohort

- 524 persons

Setting

- Spain, single-centre study

Recruitment period

- not reported

Age (in years)

- median 62 years (range: 22-93 years)

Sex

- 60% male

Stages of disease

- Rai 0: 62%; Rai I-IV: 38%
- Binet A: 83%; Binet B/C: 17%

Treatment

- 83 received FCR or similar
- 82 received purine analogs without MoAbs
- 86 received alkylating agents
- 23 received others

Inclusion criteria

- Information at diagnosis included age, clinical stage (Rai and Binet), IgHV mutational status, beta2-microglobulin (B2M), and FISH-detected cytogenetic abnormalities

Exclusion criteria

- not reported

Predictors

Number of candidate predictors

- univariable analysis: 5
- multivariable analysis: 5

List of predictors in final model (including cut-points for dichotomised factors)

- IgHV mutational status (continuous: categorical/binary transformation; binary, cut-o& not given but usually at 98%); FISH-detected genetic aberrations (del17p, del11q) (binary)

Timing of predictor measurement

- at diagnosis

Outcome(s)

Primary outcome in study

Barcelona-Brno D - Delgado 2017 (Barcelona cohort) (Continued)

- OS: calculated from the date of diagnosis to the date of death or last follow-up

Additional outcome(s)

- TTFT: calculated from the date of diagnosis to the date of first treatment or last follow-up, considering disease-unrelated deaths as competing events

Missing data	<u>Participants with any missing data?</u> • no <u>If yes, how was missing data handled?</u> • not applicable
Analysis	<u>Number of participants and number of events (specific time points where reported)</u> • overall survival: 524 persons, 212 events • TTFT: 524 persons, 292 events <u>Predictor selection method</u> • aim was the simplification of the CLL-IPI. Therefore, the 5 factors of this model were combined and the most discriminatory combination was chosen. The exact procedure was unclear. <u>Statistical method</u> • Identification of combination of biomarkers with significant discriminatory value (based on CLL-IPI model) <u>Simplification of model?</u> • no <u>Performance measures reported?</u> • Calibration: reported upon request • Discrimination: reported (c-statistic) <u>Creation of risk groups?</u> • yes, 3 risk groups (no 11q or 17p and mutated IgHV); (all others); (11q or 17p and unmutated IgHV) • at 10 years • percentage of group without event at time point (low risk: 82%, intermediate risk: 52%, high risk: 27%)
PROBAST: Applicability	<u>Domain 1: Participant selection</u> • unclear - no clear eligibility criteria and recruitment period reported <u>Domain 2: Predictors</u> • low <u>Domain 3: Outcome</u> • low
Notes	<u>Funding and conflict of interest</u> • "Red Tematica de Investigacion Cooperativa en Cancer RT, Grant No.: 06/0020/002051 and RD12/0036/0023; Instituto de Salud Carlos III (ISCIII), Grant No.: FISS PI080304; ICGC-CLL Genome Project, Generalitat de Catalunya, Grant No.: 2009SGR1008; "Emili Letang" (T.B.)" • "The authors declare no conflict of interest."

Barcelona-Brno D - Delgado 2017 (Barcelona cohort) (Continued)

Other comments

- It seemed that the model was built for OS, but also applied to TTFT, no clear description of this process provided.
- The authors have provided additional information (calibration and discrimination, number of events) upon request via email.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	Retrospective retrieval of individuals with available information for their model - individuals without the necessary data may have been left out. Unclear recruitment period
Domain 2: Predictors	Yes	No explicit statement on predictor measurement, however, there are indications that in this single-centre study, lab procedures remained similar. This was a retrospective cohort, possibility to look forward
Domain 3: Outcome	Yes	No clear description of outcome assessment. We assumed that assessment of the objective outcome OS was similar within this single-centre study and therefore did not rate as high risk (no information on e.g. registry, frequency of follow-up)
Domain 4: Analysis	No	Dichotomisation of predictors. Although the choice of predictors was based on a previous model (CLL-IPI), no formal factor selection procedure or comparison of tested factor combinations were reported. One point assigned per factor, no formally established factor weights
Overall judgement	No	

Barcelona-Brno V - Delgado 2017 (Brno cohort)

Study characteristics

General information	<u>Model and type of study</u> • Delgado 2017 (Brno cohort, Czech Republic) • Validation study <u>Secondary citations</u> • not applicable <u>Language of publication</u> • English <u>Study design</u> • not reported <u>Follow-up time</u> • median 59 months (range 3-330 months)
Participants	<u>Number of included persons in the cohort</u>

Barcelona-Brno V - Delgado 2017 (Brno cohort) (Continued)

- 417 persons

Setting

- Czech Republic (Brno), single-centre study

Recruitment period

- not reported

Age (in years)

- median 62 years (range: 32-85 years)

Sex

- 66% male

Stages of disease

- Rai 0: 41%; Rai I-II: 43%; Rai III-IV: 16%
- Binet A: 72%; Binet B: 12%; Binet C: 16%

Treatment

- 125 received FCR or similar
- 41 received purine analogs without MoAbs
- 53 received alkylating agents
- 57 received others

Inclusion criteria

- not reported

Exclusion criteria

- not reported

Predictors

List of predictors used for validation (and changes between original predictors and predictors in validation study)

- IgHV mutational status (continuous: categorical/binary transformation; binary, cut-off not given but usually at 98%); FISH-detected genetic aberrations (del17p, del11q) (binary)
- no differences between development and validation predictors

Timing of predictor measurement

- at diagnosis

Outcome(s)

Primary outcome in study

- overall survival: OS was calculated from the date of diagnosis to the date of death or last follow-up.

Additional outcome(s)

- TTFT: calculated from the date of diagnosis to the date of first treatment or last follow-up, considering disease-unrelated deaths as competing events

Outcome in model development

- overall survival

Missing data

Participants with any missing data?

- no

Barcelona-Brno V - Delgado 2017 (Brno cohort) (Continued)

If yes, how was missing data handled?

- not applicable

Analysis

Number of participants and number of events (specific time points where reported)

- OS: 417 persons, 158 events
- TTFT: 417 persons, 276 events

Which model was used?

- original model

Was the model updated?

- no

Performance measures reported?

- Calibration: provided upon request (calibration plot)
- Discrimination: provided upon request (c-statistic)

Creation of risk groups?

- yes, 3 risk groups (no 11q or 17p and mutated IgHV); (all others); (11q or 17p and unmutated IgHV)
- at 5 years
- percentage of group without event at time point (low risk: 88.4% (83.0 to 94.1%), intermediate risk: 72.5% (65.4 to 80.4%), high risk: 53.9% (44.2 to 65.8%))

PROBAST: Applicability

Domain 1: Participant selection

- unclear - recruitment period and eligibility criteria unclear

Domain 2: Predictors

- low

Domain 3: Outcome

- low

Notes

Funding and conflict of interest

- "Ministry of Health of the Czech Republic, Grant No.: AZV 15-31834A/ 2015 and AZV 15-30015A/2015; Ministry of Education, Youth and Sports of the Czech Republic project NPU1 - CEITEC 2020, Grant No.: LQ1601."
- "The authors declare no conflict of interest."

Other comments

- The authors have provided additional information (calibration and discrimination, number of events) upon request via email.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	No recruitment period and study design reported No clear inclusion and exclusion criteria (we do not know if missing values have been part of the exclusion criteria)

Barcelona-Brno V - Delgado 2017 (Brno cohort) (Continued)

Domain 2: Predictors	Yes	Not explicitly stated, but predictors probably assessed in a similar way
Domain 3: Outcome	Yes	No clear description of outcome assessment. We assumed that assessment of the objective outcome OS was similar within this single-centre study and therefore did not rate as high risk (no information on e.g. registry, frequency of follow-up)
Domain 4: Analysis	Unclear	No information on missing values; information on calibration and discrimination was provided upon request

Barcelona-Brno V - Gentile 2017 (Italian & Mayo)

Study characteristics

General information	<u>Model and type of study</u> - Gentile 2017 (Italian multicentre and Mayo clinic cohort) - Validation study <u>Secondary citations</u> - not applicable <u>Language of publication</u> - English <u>Study design</u> - retrospective cohort <u>Follow-up time</u> - median 82 months (range 3-330 months)
Participants	<u>Number of included persons in the cohort</u> 1299 persons <u>Setting</u> - Italy, USA, multicentre study <u>Recruitment period</u> 1985 - 2015 <u>Age (in years)</u> - median 63 years (range: 27-92 years) <u>Sex</u> 61.3% male <u>Stages of disease</u> - Rai 0: 57.9%; Rai I: 28.6%; Rai II: 8.6%; Rai III: 1.5%; Rai IV: 3.5% - Binet A/B/C: not reported <u>Treatment</u>

Barcelona-Brno V - Gentile 2017 (Italian & Mayo) (Continued)

- 194 received chemotherapy
- 316 received chemoimmunotherapy

Inclusion criteria

- newly-diagnosed CLL persons

Exclusion criteria

- not reported

Predictors

List of predictors used for validation (and changes between original predictors and predictors in validation study)

- IgHV mutational status (continuous: categorical/binary transformation; binary, cut-o& not given but usually at 98%); FISH-detected genetic aberrations (del17p, del11q) (binary)
- no differences between development and validation predictors

Timing of predictor measurement

- at diagnosis

Outcome(s)

Primary outcome in study

- overall survival: for OS, the time interval was measured from the day of CLL diagnosis until death from all causes or last follow-up.

Additional outcome(s)

- TTFT: measured as the day of CLL diagnosis until the start of therapy or last follow-up

Outcome in model development

- overall survival

Missing data

Participants with any missing data?

- no

If yes, how was missing data handled?

- not applicable

Analysis

Number of participants and number of events (specific time points where reported)

- OS: 1299 persons, 283 events
- TTFT: 1299 persons, 510 events

Which model was used?

- original model

Was the model updated?

- no

Performance measures reported?

- Calibration: reported (explained variation)
- Discrimination: reported (c-statistic)

Creation of risk groups?

- yes, 3 risk groups (no 11q or 17p and mutated IgHV); (all others); (11q or 17p and unmutated IgHV)

Barcelona-Brno V - Gentile 2017 (Italian & Mayo) (Continued)

- at 5 years
- percentage of group without event at time point (low risk: 92.2%, intermediate risk: 83.6%, high risk: 68.2%)

PROBAST: Applicability	<u>Domain 1: Participant selection</u> • unclear - recruitment period spanned a broad time period, wherein CLL definitions and treatment options have changed <u>Domain 2: Predictors</u> • low <u>Domain 3: Outcome</u> • low
Notes	<u>Funding and conflict of interest</u> • Funding not reported • "The authors have no conflict of interest to disclose." <u>Other comments</u> • This publication referred to the same cohort as in Gentile 2016 • The authors provided no additional information on calibration upon request.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Appropriate data sources used. Many missing values due to retrospective design, rated in domain 4
Domain 2: Predictors	Yes	Quote: "IgHV mutation analysis and FISH were performed at the reference laboratory of each participating center. The IgHV mutation status was tested on tumour DNA collected at diagnosis, and was assessed according to the ERIC guidelines."
Domain 3: Outcome	Yes	Objective standard outcome: no clear description of outcome assessment. We assumed that assessment of the objective outcome OS was similar and therefore did not rate as high risk.
Domain 4: Analysis	No	> 70% of individuals were excluded based on missing values (quote: "because of the absence of the required laboratory data in the cohort initially considered. However, the individuals included were representative of the whole cohort and were similar for age, sex, and Rai stage distribution").
Overall judgement	No	

Barcelona-Brno V - Molica 2017 (O-CLL1-GISL)

Study characteristics

General information	<u>Model and type of study</u> • Molica 2017 (O-CLL1-GISL) • Validation study

Barcelona-Brno V - Molica 2017 (O-CLL1-GISL) (Continued)

Secondary citations

- Delgado 2017

Language of publication

- English

Study design

- prospective cohort

Follow-up time

- median 42 months (range 1 - 82 months)

Participants

Number of included persons in the cohort

- 337 persons

Setting

- Italy, multicentre study

Recruitment period

- 2007 - not reported

Age (in years)

- median 61 years (range: 33 - 70 years)

Sex

- 57.2% male

Stages of disease

- Rai 0: 77.8%; Rai I-II: 22.2%
- Binet A: 100%; Binet B/C: not applicable

Treatment

- not reported

Inclusion criteria

- Newly-diagnosed individuals with CLL from several Italian Institutions who were seen within 12 months of diagnosis (confirmed by the biological review committee according to flow cytometry analysis (positive clusters of differentiation antigen 5 (CD5), 19 (CD19), 23 (CD23)); established diagnosis of B-CLL by NCI criteria, performed by local haematologist; age above 18 years and below 70 years; Binet stage A; NCI watch-and-wait policy

Exclusion criteria

- diagnosis more than 12 months ago, aged above 70 years, patients with leukaemic phase of lymphoproliferative disorders of B origin CD5- and/or CD23- according to flow cytometry analysis, Binet B or C

Predictors

List of predictors used for validation (and changes between original predictors and predictors in validation study)

- IgHV mutational status (continuous: categorical/binary transformation; binary, cut-off not given but usually at 98%); FISH-detected genetic aberrations (del17p, del11q) (binary)
- TP53 and del17p were used as composite predictor in place of del17p only.

Timing of predictor measurement

Barcelona-Brno V - Molica 2017 (O-CLL1-GISL) (Continued)

- at study entry, within 12 months of diagnosis

Outcome(s)	<u>Primary outcome in study</u> • time-to-first-treatment: the primary endpoint, TTFT, was defined as the interval between the date of registration and the date of initiation of first CLL treatment. (Molica 2016) <u>Additional outcome(s)</u> • not applicable <u>Outcome in model development</u> • overall survival
Missing data	<u>Participants with any missing data?</u> • no <u>If yes, how was missing data handled?</u> • not applicable
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> • time-to-first-treatment: 337 persons, 91 events <u>Which model was used?</u> • original model <u>Was the model updated?</u> • no <u>Performance measures reported?</u> • Calibration: not reported • Discrimination: reported (AUC) <u>Creation of risk groups?</u> • no
PROBAST: Applicability	Not applicable; outcome did not match primary outcome of the model
Notes	<u>Funding & conflict of interest</u> • "This work was supported by funding from Associazione Italiana per la Ricerca sul Cancro (AIRC 5xmille grant 9980, IG10492 to MF and FM and IG10136 to AN). We thank AIL Cosenza-Fondazione Amelia Scorza' onlus, Cosenza, Italy, and Brigida Gulino for precious secretarial assistance." (Gentile 2016) • "The authors have no conflict of interest to disclose." <u>Other comments</u> • This publication validated the Barcelona-Brno score in the same population as the validation cohort in the development study. • The outcome did not correspond to the primary outcome of the model, and was therefore not included for analysis and PROBAST rating.

Barcelona-Brno V - Muñoz-Novas 2018 (Spanish coh.)

Study characteristics

General information

Model and type of study

- Muñoz-Novas 2018 (Spanish cohort)
- Validation study

Secondary citations

- not applicable

Language of publication

- English

Study design

- retrospective cohort

Follow-up time

- median 68 months (range 3 - 277 months)

Participants

Number of included persons in the cohort

- 696 persons

Setting

- Spain, multicentre study

Recruitment period

- 1989 - 2013

Age (in years)

- median 65.7 years (IQR: 55.2 - 73.5)

Sex

- 62.8% male

Stages of disease

- Rai 0: 58.9%; Rai I-IV: 41.1%
- Binet A/B/C: not reported

Treatment

- 137 received treatment
- unclear which treatment

Inclusion criteria

- all persons diagnosed with CLL

Exclusion criteria

- not reported

Predictors

List of predictors used for validation (and changes between original predictors and predictors in validation study)

Barcelona-Brno V - Muñoz-Novas 2018 (Spanish coh.) (Continued)

- IgHV mutational status (continuous: categorical/binary transformation; binary, cut-o& not given but usually at 98%); FISH-detected genetic aberrations (del17p, del11q) (binary)
- no differences between development and validation predictors

Timing of predictor measurement

- at diagnosis

Outcome(s)

Primary outcome in study

- overall survival: Overall survival (OS) was calculated from the time of diagnosis to death or last follow-up.

Additional outcome(s)

- time-to-first-treatment: Time-to-first therapy (TTFT) was defined from the date of diagnosis to first treatment or last follow-up.

Outcome in model development

- overall survival

Missing data

Participants with any missing data?

- yes (438 persons)

If yes, how was missing data handled?

- complete-case analysis

Analysis

Number of participants & number of events (specific time points where reported)

- overall survival: 258 persons, 47 events
- time-to-first-treatment: 258 persons, 113 events

Which model was used?

- original model

Was the model updated?

- no

Performance measures reported?

- Calibration: not reported
- Discrimination: reported (c-statistic)

Creation of risk groups?

- yes, 3 risk groups (no 11q or 17p and mutated IgHV); (all others); (11q or 17p and unmutated IgHV)
- at 5 years
- percentage of group without event at time point (low risk: 90.7%, intermediate risk: 81.4%, high risk: 66.2%)

PROBAST: Applicability

Domain 1: Participant selection

- low

Domain 2: Predictors

- low

Domain 3: Outcome

Barcelona-Brno V - Muñoz-Novas 2018 (Spanish coh.) (Continued)

- low

Notes

Funding & conflict of interest

- Funding not reported
- "The authors declare that there are no conflicts of interest regarding the publication of this article."

Other comments

- none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Although individuals with incomplete data were excluded, there was a comparison and the sample seemed representative of the cohort.
Domain 2: Predictors	Yes	The type of biomarkers used seem to be standard, and although there was no explicit information we think that the methods were consistent across studies and models.
Domain 3: Outcome	Yes	Objective standard outcome - no clear description of outcome assessment - we assumed that assessment of the objective outcome OS was similar and therefore did not rate as high risk.
Domain 4: Analysis	No	Consecutive sampling of individuals in hospital routine, therefore around 2/3 of individuals with missing data for this model validation were excluded; no calibration reported.
Overall judgement	No	

Barcelona-Brno V - Rani 2018 (Indian cohort)

Study characteristics

General information

Model and type of study

- Rani 2018 (Indian cohort)
- Validation study

Secondary citations

- not applicable

Language of publication

- English

Study design

- unclear

Follow-up time

- median 40.5 months (range 1 - 215 months)

Participants

Number of included persons in the cohort

Barcelona-Brno V - Rani 2018 (Indian cohort) (Continued)

- 198 persons

Setting

- India, unclear if single or multicentre study

Recruitment period

- not reported

Age (in years)

- median 60 years (range: not reported)

Sex

- 77% male

Stages of disease

- Rai 0: 14.6%; Rai I: 21.2%; Rai II: 34.9%; Rai III: 14.1%; Rai IV: 15.2%
- Binet A/B/C: not reported

Treatment

- 62 received chlorambucil-based therapy
- 56 received rituximab-based therapy
- 20 received other therapies

Inclusion criteria

- treatment-naïve CLL persons according to Hallek 2008

Exclusion criteria

- not reported

Predictors

List of predictors used for validation (and changes between original predictors and predictors in validation study)

- IgHV mutational status (continuous: categorical/binary transformation; binary, cut-off not given but usually at 98%); FISH-detected genetic aberrations (del17p, del11q) (binary)
- no differences between development and validation predictors

Timing of predictor measurement

- not reported

Outcome(s)

Primary outcome in study

- overall survival: OS was defined as the time from the date of diagnosis to date of death or date of last follow-up.

Additional outcome(s)

- time-to-first-treatment: TTFT was defined as the time from the date of diagnosis to date of commencement of first therapy.

Outcome in model development

- overall survival

Missing data

Participants with any missing data?

- no

Barcelona-Brno V - Rani 2018 (Indian cohort) (Continued)

If yes, how was missing data handled?

- not applicable

Analysis

Number of participants & number of events (specific time points where reported)

- overall survival: 198 persons, 86 events
- time-to-first-treatment: 140 persons, 89 events

Which model was used?

- original model

Was the model updated?

- no

Performance measures reported?

- Calibration: not reported
- Discrimination: reported (c-statistic)

Creation of risk groups?

- yes, 3 risk groups (no 11q or 17p and mutated IgHV); (all others); (11q or 17p and unmutated IgHV)
- median survival (low risk: 138 years, intermediate risk: 63 years, high risk: 56 years)

PROBAST: Applicability

Domain 1: Participant selection

- unclear - recruitment period and eligibility criteria unclear

Domain 2: Predictors

- unclear - no information on timing of predictor because the study design is unclear

Domain 3: Outcome

- low

Notes

Funding & conflict of interest

- "... Department of Biotechnology (BT/PR11106/GBD/27/145/2008, BT/PR15438/MED/30/606/2011 and T/PR8680/AGR/36/754/2013), Ministry of Science and Technology, GOI, and All India Institute of Medical Sciences, New Delhi (8-60/A060/ 2011/RS) to RG to carry out this work."
- "The authors declare that they have no conflict of interest."

Other comments

- none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	Inclusion and exclusion criteria unclear; it seemed to us that only individuals with complete information on all predictors were included. Recruitment period unclear
Domain 2: Predictors	Yes	Not explicitly stated, but predictors probably assessed in a similar way (single-centre)

Barcelona-Brno V - Rani 2018 (Indian cohort) (Continued)

Domain 3: Outcome	Unclear	Observation time to observe survival was short.
Domain 4: Analysis	No	Low number of events, no information on handling of missing values; calibration was not reported.
Overall judgement	No	

Barcelona-Brno V - Reda 2017 (Milan cohort)

Study characteristics

General information	<u>Model and type of study</u> - Reda 2017 (Milan cohort) - Validation study <u>Secondary citations</u> - not applicable <u>Language of publication</u> - English <u>Study design</u> - retrospective cohort <u>Follow-up time</u> - median 144 months (range 0 - 360 months)
Participants	<u>Number of included persons in the cohort</u> 698 persons <u>Setting</u> - Italy, single-centre study <u>Recruitment period</u> 1983 - 2016 <u>Age (in years)</u> - median 65 years (range: 32 - 91 years) <u>Sex</u> 57% male <u>Stages of disease</u> - Rai 0-I: 74%; Rai II-IV: not reported - Binet A/B/C: not reported <u>Treatment</u> 190 received Chlorambucil (CHL) 108 received fludarabinecytosphamide-rituximab (FCR)

Barcelona-Brno V - Reda 2017 (Milan cohort) (Continued)

- 12 received ofatumumabbendamustine (O-Benda)
- 44 received bendamustine-rituximab (BR)
- 65 received alemtuzumab
- 29 received Ibrutinib or Idelalisib

Inclusion criteria

- all patients diagnosed with CLL, reclassified by 2008 criteria

Exclusion criteria

- not reported

Predictors	<u>List of predictors used for validation (and changes between original predictors and predictors in validation study)</u> • IgHV mutational status (continuous: categorical/binary transformation; binary, cut-o& not given but usually at 98%); FISH-detected genetic aberrations (del17p, del11q) (binary) • no differences between development and validation predictors <u>Timing of predictor measurement</u> • at diagnosis
Outcome(s)	<u>Primary outcome in study</u> • overall survival: overall survival (OS) was calculated from CLL diagnosis to death. <u>Additional outcome(s)</u> • time-to-first-treatment: Time-To-First-Treatment (TTFT) was evaluated from the time of diagnosis to firstline therapy start. <u>Outcome in model development</u> • overall survival
Missing data	<u>Participants with any missing data?</u> • yes (369 participants) <u>If yes, how was missing data handled?</u> • unclear
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> • overall survival: 329 persons, number of events not reported • time-to-first-treatment: 329 persons, number of events not reported <u>Which model was used?</u> • original model <u>Was the model updated?</u> • no <u>Performance measures reported?</u> • Calibration: not reported • Discrimination: not reported <u>Creation of risk groups?</u>

Barcelona-Brno V - Reda 2017 (Milan cohort) (Continued)

- yes, 3 risk groups (no 11q or 17p and mutated IgHV); (all others); (11q or 17p and unmutated IgHV)
- at 15 years
- percentage of group without event at time point (low risk: 80%, intermediate risk: 50%, high risk: 30%)

PROBAST: Applicability

Domain 1: Participant selection

- unclear - no clear information on eligibility criteria and study design

Domain 2: Predictors

- unclear - no information on the timing of the predictor

Domain 3: Outcome

- low

Notes

Funding & conflict of interest

- Funding not reported
- "Authors have no affiliations that they consider to be relevant and important with any organization that to any author's knowledge has a direct interest, particularly a financial interest, in the subject matter discussed."

Other comments

- The authors did not reply to our request via email for additional information.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	No eligibility criteria, unclear study design
Domain 2: Predictors	Yes	Not explicitly stated, but predictors probably assessed in a similar way
Domain 3: Outcome	Yes	No clear description of outcome assessment - we assumed that assessment of the objective outcome OS was similar and therefore did not rate as high risk.
Domain 4: Analysis	No	No information on number of events; no performance measures reported; patients with missing data were leL out from analysis.
Overall judgement	No	

CLL-IPI D - Bahlo 2016 (development cohort)

Study characteristics

General information

Model and type of study

- Bahlo 2016 (development cohort)
- Development study

Secondary citations

- not applicable

Language of publication

CLL-IPI D - Bahlo 2016 (development cohort) (Continued)

- English

Study design

- combined data (e.g. IPD or combination of cohorts); data from 8 phase-three RCTs

Follow-up time

- median 79.9 months, IQR: 79.9 - 101.4

Participants

Number of included persons in the cohort

- 3472 persons

Setting

- multicentre study
- France, Germany, Poland, UK, USA, and possibly more

Recruitment period

- 1997 - 2009

Age (in years)

- median 61 years (range: 27 - 86 years)

Sex

- 70% male

Stages of disease

- Rai 0: 14%; Rai I/II: 50%; Rai III/IV: 36%
- Binet A: 32%; Binet B: 41%; Binet C: 27%

Treatment

- 690 received fluradabine (F)
- 1017 received F + cyclophosphamide (C)
- 113 received C + cladribine
- 475 received FC + rituximab
- 81 received FC + alemtuzumab
- 479 received chlorambucil

Inclusion criteria

- untreated CLL - possibly more exclusion criteria in the individual RCTs; FR: between the ages of 18 and 65 years; absence of del17p, life expectancy longer than 6 months; DE - CLL4: 18 and 65 years; life expectancy more than 6 months; Binet stage C and stage B if they had rapid disease progression or symptoms as evidenced by enlarged lymph nodes and organs or if they had severe B symptoms, stage A with severe B-symptoms; DE - CLL5: 65 and 80 years; life expectancy more than 6 months; Binet stage C and in stage B or A if they had rapid disease progression (lymphocyte doubling time 3 months) or symptoms from enlarged lymph nodes and organs, or if they had severe B-symptoms; PL: more than 18 years, progressive disease; UK: stages B, C, and A-progressive who needed treatment; USA: 18 years or older with a diagnosis of progressive CLL

Exclusion criteria

- FR: comorbidities (CIRS more than 5, ECOG more than 2; autoimmune cytopenia, HIV, HBV, HCV, active second malignancy, transformation to aggressive B-cell malignancy, abnormal renal function, bilirubin, transaminase more than 2 times the upper limit; DE - CLL 4: ECOG more than 3; organ dysfunction, concomitant or previous neoplasms, autoimmune haemolytic anaemia or thrombocytopenia; DE - CLL5: ECOG more than 3; organ dysfunction, concomitant or previous neoplasms, autoimmune

CLL-IPI D - Bahlo 2016 (development cohort) (Continued)

haemolytic anaemia, thrombocytopenia; DE - CLL8: ECOG more than 2; CIRS more than 5; autoimmune cytopenia, active second disease; PL: poor performance status (WHO grade 4), autoimmune haemolytic anaemia (AIHA), autoimmune thrombocytopenia, active infections, abnormal liver or renal function, Richter's syndrome, or concomitant neoplasm; USA: second malignancy, serum creatinine more than 2 mg/dL; performance status more than 2; bilirubin more than 1 mg/dL; autoimmune haemolytic anaemia, thrombocytopenia; history of steroid treatment; pregnancy and lactating women; active infection

Predictors	<u>Number of candidate predictors</u> • univariable analysis: 27 • multivariable analysis: 25 <u>List of predictors in final model (including cut-points for dichotomised factors)</u> • TP53 status (binary; deleted or mutated); IgHV mutational status (continuous: categorical/binary transformation; cut-off not given but usually 98%); β2-microglobulin concentration (continuous: categorical/binary transformation; binary, cut-off 3.5 mg/L based on literature); clinical stage (Rai or Binet) (categorical); age (continuous: categorical/binary transformation; cut-off: 65 years) <u>Timing of predictor measurement</u> • before the start of first-line treatment
Outcome(s)	<u>Primary outcome in study</u> • overall survival: overall survival was calculated from study entry to date of death. <u>Additional outcome(s)</u> • not applicable
Missing data	<u>Participants with any missing data?</u> • yes (1673 participants) <u>If yes, how was missing data handled?</u> • complete-case analysis
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> • overall survival: 1214 persons, 462 events for training set; 585 persons, 243 events for internal validation set <u>Predictor selection method</u> • univariable analysis, during multivariable modelling: a stepwise modelling based on categories of missing values: step 1: N = 1908, age, gender, Binet, LDH, haemoglobin [ex: WBC, HGB, LYM]; step 2: N = 1563 (585 events), age, gender, clinical stage (Rai OR Binet), LDH, haemoglobin, ECOG, hierarchical type (del17p, del11q); step 3: N = 1192 (452 events); age, clinical stage (Rai or Binet), IgHV, B2M, hierarchical type (del17p); step 4: more MV analyses (table S11 P value or CI) <u>Statistical method</u> • Cox proportional hazard model <u>Simplification of model?</u> • yes <u>Performance measures reported?</u> • Calibration: not reported

CLL-IPI D - Bahlo 2016 (development cohort) (Continued)

- Discrimination: reported (c-statistic)

Creation of risk groups?

- yes, 4 risk groups, (cut-points at (0-1);(2-3);(4-6); (7-10))
- at 5 years
- percentage of group without event at time point (low risk: 93.2%, intermediate risk: 79.3%, high risk: 63.3%, very high risk: 23.3%)

PROBAST: Applicability

Domain 1: Participant selection

- high - most individuals had a previous treatment indication, in total only 17.8% of individuals were on watch-and-wait strategy at enrolment, therefore the population was rather selective.

Domain 2: Predictors

- low

Domain 3: Outcome

- low

Notes

Funding & conflict of interest

- "José Carreras Leukaemia Foundation"
- KB: BMBF grant. BFE: honorarium for advisory boards, honorarium and/or scientific grants from Janssen, Gilead, Mundipharma, GlaxoSmithKline (GSK) and Roche. CG: personal fees from Roche, Janssen, Gilead, Celgene, Novartis, and AbbVie, outside the submitted work. MG: National Cancer Institute during the conduct of the study; grants, personal fees, and non-financial support from Pharmacyclics and Acerta, outside the submitted work. TDS: grants from Genentech, Pharmacyclics Janssen, GSK, Celgene, Cephalon, Hospira, and Polyphenon E International, outside the submitted work. SS: grants, personal fees, and other from AbbVie, Amgen, Boehringer Ingelheim, Celgene, Genentech, Genzyme, Gilead, GSK, Janssen, Mundipharma, Novartis, Pharmacyclics, Hoffmann La-Roche, and Sanofi, during the conduct of the study. "JBa, NK, MAB, JBy, KGC, HD, ME, SL, DN, DO, TR, RR, and MH declare no competing interests." "The funder of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report."

Other comments

- The authors provided additional information upon request (cohort information, measures of calibration).

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Appropriate data sources used, missing values were excluded from analysis. This was rated in domain 4.
Domain 2: Predictors	Yes	Well-established predictors used - not explicitly stated, but predictors probably assessed in a similar way, according to RCT protocols.
Domain 3: Outcome	Yes	Objective standard outcome, follow-up in the context of each individual RCT Median observation time seemed on the lower limit (around 5 years), but reasonable for patients with treatment indication.
Domain 4: Analysis	No	It was unclear for which model the performance was assessed (score or formula). Predictors were categorised and dichotomised. Patients with missing values were dropped from the analysis - exclusion of approximately half of all patients. Univariable analysis was used for predictor selection.

CLL-IPI D - Bahlo 2016 (development cohort) (Continued)

Upon request, the authors provided information on the calibration of their model.

Overall judgement	No

CLL-IPI V - Bahlo 2016 (Mayo clinic 2001-2014)

Study characteristics

General information	<u>Model and type of study</u>

- Bahlo 2016 (Mayo clinic 2001 - 2014)
- Validation study

Secondary citations

- Gentile 2016; Molica 2016

Language of publication

- English

Study design

- prospective cohort

Follow-up time

- median 63.2 months, IQR 30.2 to 91.8

Participants	<u>Number of included persons in the cohort</u>

- 838 persons

Setting

- USA, single-centre study

Recruitment period

- 2001 - 2014

Age (in years)

- ≤ 65 years: 533 persons
- > 65 years: 305 persons

Sex

- 67.8% male

Stages of disease

- Rai 0: 57.4%; Rai I-II: 38.5%; Rai III-IV: 4.1%
- Binet A/B/C: not reported

Treatment

- 163 received purine analogue chemotherapy
- 5 received purine analogue monotherapy or combination without monoclonal antibody
- 37 received alkylator-based chemoimmunotherapy

CLL-IPI V - Bahlo 2016 (Mayo clinic 2001-2014) (Continued)

- 11 received alkylator monotherapy or combination without monoclonal antibody
- 97 received antibody only
- 7 received RTK inhibitors
- 6 received other treatments

Inclusion criteria

- complete baseline data

Exclusion criteria

- not reported

Predictors

List of predictors used for validation (and changes between original predictors and predictors in validation study)

- TP53 status (binary; deleted or mutated); IgHV mutational status (continuous: categorical/binary transformation; cut-o& not given but usually 98%); β 2-microglobulin concentration (continuous: categorical/binary transformation; binary, cut-o& 3.5 mg/L (based on literature)); clinical stage (Rai or Binet) (categorical); Age (continuous: categorical/binary transformation; cut-o& 65 years)
- missing data for TP53, therefore replaced by del(17p)

Timing of predictor measurement

- not reported, presumably at diagnosis

Outcome(s)

Primary outcome in study

- overall survival: OS calculated from diagnosis to death.

Additional outcome(s)

- not applicable

Outcome in model development

- overall survival

Missing data

Participants with any missing data?

- no

If yes, how was missing data handled?

- not applicable

Analysis

Number of participants & number of events (specific time points where reported)

- overall survival: 838 persons, 144 events

Which model was used?

- simplified model

Was the model updated?

- no

Performance measures reported?

- Calibration: not reported
- Discrimination: reported (c-statistic)

Creation of risk groups?

CLL-IPI V - Bahlo 2016 (Mayo clinic 2001-2014) (Continued)

- yes, 4 risk groups, (cut-points at (0-1);(2-3);(4-6); (7-10))
- at 5 years
- percentage of group without event at time point (low risk: 96.6%, intermediate risk: 92%, high risk: 68.5%, very high risk: 21.2%)

PROBAST: Applicability	<u>Domain 1: Participant selection</u> • unclear - insufficient information on CLL diagnostic criteria and inclusion criteria (i.e. availability of baseline data). <u>Domain 2: Predictors</u> • unclear - predictor TP53 not available and replaced by a proxy, del(17p) <u>Domain 3: Outcome</u> • low
Notes	<u>Funding & conflict of interest</u> • "José Carreras Leukaemia Foundation" • KB: BMBF grant. BFE: honorarium for advisory boards, honorarium and/or scientific grants from Janssen, Gilead, Mundipharma, GlaxoSmithKline (GSK) and Roche. CG: personal fees from Roche, Janssen, Gilead, Celgene, Novartis, and AbbVie, outside the submitted work. MG: National Cancer Institute during the conduct of the study; grants, personal fees, and non-financial support from Pharmacyclics and Acerta, outside the submitted work. TDS: grants from Genentech, Pharmacyclics Janssen, GSK, Celgene, Cephalon, Hospira, and Polyphenon E International, outside the submitted work. SS: grants, personal fees, and other from AbbVie, Amgen, Boehringer Ingelheim, Celgene, Genentech, Genzyme, Gilead, GSK, Janssen, Mundipharma, Novartis, Pharmacyclics, Hoffmann La-Roche, and Sanofi, during the conduct of the study. "JBa, NK, MAB, JBy, KGC, HD, ME, SL, DN, DO, TR, RR, and MH declare no competing interests." "The funder of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report." <u>Other comments</u> • none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	No	No information on inclusion criteria, definition of CLL and if/how many patients were excluded due to missing baseline data
Domain 2: Predictors	Yes	Well-established predictors used - not explicitly stated, but predictors probably assessed in a similar way (single-centre)
Domain 3: Outcome	Yes	Objective standard outcome: no clear description of outcome assessment - we assumed that assessment of the objective outcome OS was similar and therefore did not rate as high risk (no info on e.g. registry, frequency of follow-up etc.)
Domain 4: Analysis	Yes	Number of events sufficient Patients with missing values possibly dropped at enrolment, not clearly stated: 'There were no missing values for the MAYO cohort.'. This was rated in domain 1. Upon request, the authors provided information on the calibration of their model.

CLL-IPI V - Bahlo 2016 (Mayo clinic 2001-2014) (Continued)

Overall judgement No

CLL-IPI V - Bahlo 2016 (SCAN cohort)

Study characteristics

General information	<u>Model and type of study</u> - Bahlo 2016 (SCAN cohort) - Validation study <u>Secondary citations</u> - not applicable <u>Language of publication</u> - English <u>Study design</u> - other: cohort of CLL patients within a so-called case-control study, which was actually a register; uncertain whether prospective or retrospective study design <u>Follow-up time</u> - median 151 months, IQR 124.8 to 163.0
Participants	<u>Number of included persons in the cohort</u> 416 persons <u>Setting</u> - Denmark and Sweden, multicentre study <u>Recruitment period</u> 1999 - 2002 <u>Age (in years)</u> ≤ 65 years: 242 persons > 65 years: 174 persons <u>Sex</u> 62.7% male <u>Stages of disease</u> - Rai: not reported - Binet A: 78.1%; Binet B: 16.6%; Binet C: 5.3% <u>Treatment</u> - not reported <u>Inclusion criteria</u> - complete baseline data

CLL-IPI V - Bahlo 2016 (SCAN cohort) (Continued)

Exclusion criteria

- The source population for the SCALE study was restricted to subjects with sufficient knowledge of the Danish or Swedish language to answer questions in a telephone interview and without a history of organ transplantation, human immunodeficiency virus infection, or other hematopoietic malignancy (Smedby 2005).

Predictors

List of predictors used for validation (and changes between original predictors and predictors in validation study)

- TP53 status (binary; deleted or mutated); IgHV mutational status (continuous: categorical/binary transformation; cut-o& not given but usually 98%); β 2-microglobulin concentration (continuous: categorical/binary transformation; binary, cut-o& 3.5 mg/L (based on literature)); clinical stage (Rai or Binet) (categorical); age (continuous: categorical/binary transformation; cut-o& 65 years)
- 4% of β 2-microglobulin imputed (multiple imputation)

Timing of predictor measurement

- not reported, presumably at diagnosis

Outcome(s)

Primary outcome in study

- overall survival: OS was calculated from diagnosis to death.

Additional outcome(s)

- not applicable

Outcome in model development

- overall survival

Missing data

Participants with any missing data?

- yes (17 participants)

If yes, how was missing data handled?

- multiple imputation

Analysis

Number of participants & number of events (specific time points where reported)

- overall survival: 416 persons, 215 events

Which model was used?

- simplified model

Was the model updated?

- no

Performance measures reported?

- Calibration: not reported
- Discrimination: reported (c-statistic)

Creation of risk groups?

- yes, 4 risk groups, (cut-points at (0-1);(2-3);(4-6); (7-10))
- at 5 years
- percentage of group without event at time point (low risk: 92.1%, intermediate risk: 75%, high risk: 64.3%, very high risk: 28.6%)

CLL-IPI V - Bahlo 2016 (SCAN cohort) (Continued)

PROBAST: Applicability

Domain 1: Participant selection

- unclear - the SCALE study database encompassed the entire population in Denmark and Sweden in a specified time frame. For this model validation, only the CLL cohort was relevant. We did not know exactly how many participants were excluded due to missing information. No eligibility criteria were defined. We were unsure if the population can reflect the general CLL population.

Domain 2: Predictors

- low

Domain 3: Outcome

- low

Notes

Funding & conflict of interest

- "José Carreras Leukaemia Foundation"
- KB: BMBF grant. BFE: honorarium for advisory boards, honorarium and/or scientific grants from Janssen, Gilead, Mundipharma, GlaxoSmithKline (GSK) and Roche. CG: personal fees from Roche, Janssen, Gilead, Celgene, Novartis, and AbbVie, outside the submitted work. MG: National Cancer Institute during the conduct of the study; grants, personal fees, and non-financial support from Pharmacia and Acerta, outside the submitted work. TDS: grants from Genentech, Pharmacia Janssen, GSK, Celgene, Cephalon, Hospira, and Polyphenon E International, outside the submitted work. SS: grants, personal fees, and other from AbbVie, Amgen, Boehringer Ingelheim, Celgene, Genentech, Genzyme, Gilead, GSK, Janssen, Mundipharma, Novartis, Pharmacia, Hoffmann La-Roche, and Sanofi, during the conduct of the study. "JBa, NK, MAB, JBy, KGC, HD, ME, SL, DN, DO, TR, RR, and MH declare no competing interests." "The funder of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report."

Other comments

- Upon request, the authors provided information on the calibration of their model.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	The SCALE study database encompassed the entire population in Denmark and Sweden in a specified time frame. For this model validation, only the CLL cohort was relevant. We did not know exactly how many participants were excluded. No eligibility criteria were defined.
Domain 2: Predictors	Yes	Patient data collection between 1999 and 2002; we assumed that within this time frame, predictor assessment remained relatively homogenous.
Domain 3: Outcome	Yes	Objective standard outcome: no clear description of outcome assessment in this publication. However, this was a national cohort study, therefore we assumed standardised outcome assessment.
Domain 4: Analysis	Yes	Sufficient number of events Missing data was handled appropriately: 'we analysed missing values of the SCAN cohort using Little's MCAR test and imputed these using linear regression.' Upon request, the authors provided information on the calibration of their model.

CLL-IPI V - Da Cunha-Bang 2016 (Danish cohort)

Study characteristics

General information	<u>Model and type of study</u> - Da Cunha-Bang 2016 (Danish National CLL Registry) - Validation study <u>Secondary citations</u> - not applicable <u>Language of publication</u> - English <u>Study design</u> - prospective cohort <u>Follow-up time</u> - median 38.4 months, range not reported
Participants	<u>Number of included persons in the cohort</u> 3023 persons <u>Setting</u> - Denmark, multicentre study <u>Recruitment period</u> 2008 - 2015 <u>Age (in years)</u> - median 69 years (range: not reported) <u>Sex</u> 60% male <u>Stages of disease</u> - Rai: not reported - Binet A: 80%; Binet B/C: 20% <u>Treatment</u> 295 received treatment - unclear which treatment <u>Inclusion criteria</u> - all patients from national registry <u>Exclusion criteria</u> - missing data
Predictors	<u>List of predictors used for validation (and changes between original predictors and predictors in validation study)</u>

CLL-IPI V - Da Cunha-Bang 2016 (Danish cohort) (Continued)

- TP53 status (binary; deleted or mutated); IgHV mutational status (continuous: categorical/binary transformation; cut-o& not given but usually 98%); β 2-microglobulin concentration (continuous: categorical/binary transformation; binary, cut-o& 3.5 mg/L (based on literature)); clinical stage (Rai or Binet) (categorical); age (continuous: categorical/binary transformation; cut-o& 65 years)
- no differences between development and validation predictors reported

Timing of predictor measurement

- at diagnosis

Outcome(s)	<u>Primary outcome in study</u> • progression-free survival: TTE (time-to-event), event defined as treatment or death. <u>Additional outcome(s)</u> • overall survival: OS defined as time to death <u>Outcome in model development</u> • overall survival
Missing data	<u>Participants with any missing data?</u> • no <u>If yes, how was missing data handled?</u> • not applicable
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> • progression-free survival: 1514 persons, 544 events • overall survival: 1514 persons, 249 events <u>Which model was used?</u> • simplified model <u>Was the model updated?</u> • no <u>Performance measures reported?</u> • Calibration: not reported • Discrimination: not reported <u>Creation of risk groups?</u> • yes, 4 risk groups, (cut-points at (0-1);(2-3);(4-6); (7-10)) • no time point and numerical results reported, see figure 1 instead
PROBAST: Applicability	<u>Domain 1: Participant selection</u> • low <u>Domain 2: Predictors</u> • unclear - predictors not described, therefore it was unclear if TP53 or a proxy was used <u>Domain 3: Outcome</u> • low

CLL-IPI V - Da Cunha-Bang 2016 (Danish cohort) (Continued)

Notes

Funding & conflict of interest

- Funding not reported
- "During the study, C.U.N. received grants from the Danish Cancer Society, consultancy fees (from Janssen, Roche, Abbvie, and Gilead), and grants (from Novartis and Roche) outside the submitted work and is the principal investigator for clinical trials sponsored by Roche. The remaining authors declare no competing financial interests."

Other comments

- none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	Individuals with missing data previously excluded; comparison of included and excluded individuals showed similar baseline characteristics.
Domain 2: Predictors	Yes	Well-established predictors used - not explicitly stated, but predictors probably assessed in a similar way. Relatively recent and prospectively followed cohort
Domain 3: Outcome	Unclear	For mortality, the observation time was too short (our clinician recommended observation time to exceed 5 years, median observation time here: 3.2 years)
Domain 4: Analysis	Unclear	Patients with missing data previously excluded, although authors stated that they were comparable. Rated in domain 1. No information on model performance measures, however, a validation in our sense was not planned.

CLL-IPI V - Delgado 2017 (Barcelona cohort)

Study characteristics

General information

Model and type of study

- Delgado 2017 (Barcelona hospital cohort)
- Validation study

Secondary citations

- not applicable

Language of publication

- English

Study design

- retrospective cohort

Follow-up time

- median 99.6 months, range not reported

Participants

Number of included persons in the cohort

CLL-IPI V - Delgado 2017 (Barcelona cohort) (Continued)

- 524 persons

Setting

- Spain (Barcelona), single-centre study

Recruitment period

- not reported

Age (in years)

- median 62 years (range: 22 - 93 years)

Sex

- 60% male

Stages of disease

- Rai 0: 62%; Rai I-IV: 38%
- Binet A: 83%; Binet or C: 17%

Treatment

- 83 received FCR or similar
- 82 received purine analogs w/o MoAbs
- 86 received alkylating agents
- 23 received others

Inclusion criteria

- Availability of information on age, clinical stage, IgHV status, β 2m, FISH

Exclusion criteria

- not reported

Predictors

List of predictors used for validation (and changes between original predictors and predictors in validation study)

- TP53 status (binary; deleted or mutated); IgHV mutational status (continuous: categorical/binary transformation; cut-o& not given but usually 98%); β 2-microglobulin concentration (continuous: categorical/binary transformation; binary, cut-o& 3.5 mg/L (based on literature)); clinical stage (Rai or Binet) (categorical); age (continuous: categorical/binary transformation; cut-o& 65 years)
- no changes

Timing of predictor measurement

- at diagnosis

Outcome(s)

Primary outcome in study

- overall survival: OS was calculated from the date of diagnosis to the date of death or last follow-up.

Additional outcome(s)

- not applicable

Outcome in model development

- overall survival

Missing data

Participants with any missing data?

CLL-IPI V - Delgado 2017 (Barcelona cohort) (Continued)

- no

If yes, how was missing data handled?

- not applicable

Analysis

Number of participants & number of events (specific time points where reported)

- overall survival: 524 persons, number of events not reported

Which model was used?

- simplified model

Was the model updated?

- no

Performance measures reported?

- Calibration: not reported
- Discrimination: not reported

Creation of risk groups?

- yes, 4 risk groups, (cut-points at (0-1);(2-3);(4-6); (7-10), using Rai or Binet)
- at 5 years
- percentage of group without event at time point (low risk: 95.7%, intermediate risk: 87.1%, high risk: 63.5%, very high risk: 51.9%)

PROBAST: Applicability

Domain 1: Participant selection

- unclear - no clear eligibility criteria and recruitment period reported

Domain 2: Predictors

- low

Domain 3: Outcome

- low

Notes

Funding & conflict of interest

- "Red Tematica de Investigacion Cooperativa en Cancer RT, Grant No.: 06/0020/002051 and RD12/0036/0023; Instituto de Salud Carlos III (ISCIII), Grant No.: FISS PI080304; ICGC-CLL Genome Project, Generalitat de Catalunya, Grant No.: 2009SGR1008; "Emili Letang" (T.B.)."
- "The authors declare no conflict of interest."

Other comments

- Upon request, the authors provided additional information on the discrimination and calibration of the CLL-IPI model in their cohort.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	Retrospective retrieval of persons with available information for their model - individuals without the necessary data may have been left out. Unclear recruitment period

CLL-IPI V - Delgado 2017 (Barcelona cohort) (Continued)

Domain 2: Predictors	Yes	Well-established predictors used - not explicitly stated, but predictors probably assessed in a similar way (single-centre study)
Domain 3: Outcome	Yes	Objective standard outcome: no clear description of outcome assessment in this publication. We assumed standardised outcome assessment (single-centre study).
Domain 4: Analysis	Yes	Sufficient number of events. No missing values, however, possibly an exclusion reason. This was rated in domain 1. Upon request, the authors provided information on the calibration of the CLL-IPI in their cohort.
Overall judgement	No	

CLL-IPI V - Gentile 2016 (Italian cohort)

Study characteristics

General information	<u>Model and type of study</u> - Gentile 2016 (Italian multicentre cohort) - Validation study <u>Secondary citations</u> - Gentile 2017 <u>Language of publication</u> - English <u>Study design</u> - retrospective cohort <u>Follow-up time</u> - median 69.6 months, range not reported
Participants	<u>Number of included persons in the cohort</u> 858 persons <u>Setting</u> - Italy, multicentre study <u>Recruitment period</u> 1985 - 2015 <u>Age (in years)</u> - median 65.5 years (range: not reported) <u>Sex</u> 56.2% male

CLL-IPI V - Gentile 2016 (Italian cohort) (Continued)

Stages of disease

- Rai 0: 58.4%; Rai I: 19.8%; Rai II: 15.2%; Rai III: 1.5%; Rai IV: 5.1%
- Binet A: 79.7%; Binet B: 18.8%; Binet C: 1.5%

Treatment

- 130 received chemotherapy
- 174 received chemoimmunotherapy

Inclusion criteria

- All patients with CLL in one of the five centres that were evaluable for CLL-IPI (i.e. fully available data on age, staging, IgHV, β 2m and del17p/TP53)

Exclusion criteria

- not reported

Predictors	<u>List of predictors used for validation (and changes between original predictors and predictors in validation study)</u> • TP53 status (binary; deleted or mutated); IgHV mutational status (continuous: categorical/binary transformation; cut-o& not given but usually 98%); β2-microglobulin concentration (continuous: categorical/binary transformation; binary, cut-o& 3.5 mg/L (based on literature)); clinical stage (Rai or Binet) (categorical); age (continuous: categorical/binary transformation; cut-o& 65 years) • TP53 was replaced by the proxy del(17p). <u>Timing of predictor measurement</u> • at diagnosis
Outcome(s)	<u>Primary outcome in study</u> • overall survival: for OS, the time interval was measured from the day of CLL diagnosis until death from all causes or last follow-up. <u>Additional outcome(s)</u> • time-to-first-treatment: For TTFT from the day of CLL diagnosis until the start of therapy or last follow-up <u>Outcome in model development</u> • overall survival
Missing data	<u>Participants with any missing data?</u> • no <u>If yes, how was missing data handled?</u> • not applicable
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> • overall survival: 858 persons, 174 events • time-to-first-treatment: 858 persons, 304 events <u>Which model was used?</u> • simplified model <u>Was the model updated?</u>

CLL-IPI V - Gentile 2016 (Italian cohort) (Continued)

- no

Performance measures reported?

- Calibration: not reported, survival per risk group and explained variation on outcome available
- Discrimination: reported (c-statistic)

Creation of risk groups?

- yes, 4 risk groups, (cut-points at (0-1);(2-3);(4-6); (7-10))
- at 5 years
- percentage of group without event at time point (low risk: 94.5%, intermediate risk: 86.4%, high risk: 74.4%, very high risk: 42.9%)

PROBAST: Applicability	<u>Domain 1: Participant selection</u> • unclear - data was retrospectively collected for a large time period (1985 - 2015). During this time, the definition of CLL and treatment options have changed. Therefore, the cohort may be very heterogeneous and discrimination for OS may not reflect performance of the model for individuals diagnosed nowadays. <u>Domain 2: Predictors</u> • unclear - predictor TP53 not available and replaced by a proxy, del(17p) <u>Domain 3: Outcome</u> • low
Notes	<u>Funding & conflict of interest</u> • Funding source not reported • "S.A.P. received funding from Pharmacyclics. All other authors have no conflict of interest to disclose." <u>Other comments</u> • identical centres as presented in Gentile 2017

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Appropriate data sources used: consecutive sampling, but exclusion of individuals without baseline characteristics and FISH analysis from analysis (in total, only 22.5% of individuals were evaluable). This is rated in domain 4.
Domain 2: Predictors	Unclear	Due to the broad time range and multiple participating centres, we doubt that predictor assessments were homogeneous.
Domain 3: Outcome	Yes	Objective standard outcome: '... death, which were abstracted from clinical records at the time of inclusion and updated on an ongoing basis.'
Domain 4: Analysis	No	Consecutive sampling, but exclusion of individuals without baseline characteristics and FISH analysis from analysis (in total, only 22.5% of patients were evaluable). Missing values emerged due to the long time span. Some predictors were not assessed by default at diagnosis until recently. Comparison of baseline sample and included sample similar.
		Calibration not reported, however, survival per risk group available
Overall judgement	No	

CLL-IPI V - Molica 2016 (O-CLL1-GISL)

Study characteristics

General information	<u>Model and type of study</u>
	• Molica 2016 (O-CLL1-GISL) • Validation study
	<u>Secondary citations</u>
	• Molica 2017; Molica 2017; Molica 2016
	<u>Language of publication</u>
	• English
	<u>Study design</u>
	• prospective cohort
	<u>Follow-up time</u>
	• median 42 months, 2038 person-years
Participants	<u>Number of included persons in the cohort</u>
	• 337 persons
	<u>Setting</u>
	• Italy, multicentre study
	<u>Recruitment period</u>
	• 2007 - not reported
	<u>Age (in years)</u>
	• median 61 years (range: 33 - 70 years)
	<u>Sex</u>
	• 57.2% male
	<u>Stages of disease</u>
	• Rai 0: 77.8%; Rai ≥I: 22.2%
	• Binet A: 100%; Binet B/C: 0%
	<u>Treatment</u>
	• 91 received therapy • unclear which therapy
	<u>Inclusion criteria</u>
	• Newly-diagnosed CLL patients from several Italian Institutions who were seen within 12 months of diagnosis were prospectively enrolled into the OCLL1- GISL protocol.
	<u>Exclusion criteria</u>
	• not reported

CLL-IPI V - Molica 2016 (O-CLL1-GISL) (Continued)

Predictors	<u>List of predictors used for validation (and changes between original predictors and predictors in validation study)</u> - TP53 status (binary; deleted or mutated); IgHV mutational status (continuous: categorical/binary transformation; cut-o& not given but usually 98%); β2-microglobulin concentration (continuous: categorical/binary transformation; binary, cut-o& 3.5 mg/L (based on literature)); clinical stage (Rai or Binet) (categorical); age (continuous: categorical/binary transformation; cut-o& 65 years) - TP53 and del17p as composite <u>Timing of predictor measurement</u> - at study entry, within 12 months of diagnosis
Outcome(s)	<u>Primary outcome in study</u> - time-to-first-treatment: the primary endpoint, TTFT, was defined as the interval between the date of registration and the date of initiation of first CLL treatment. Patients underwent sequential monitoring, and the frequency of follow-up visits was individualised according to patient risk; this ranged from 3 to 6 months (median 6 months). All physicians who registered patients in this observational database stated that they had used the NCI-WG guidelines as a reference criterion for starting therapy. In particular, the absolute lymphocyte count was not used as the sole indicator for treatment. Active disease, requiring therapy, was defined when at least one of the criteria set out in the NCI sponsored Working Group guidelines was satisfied. <u>Additional outcome(s)</u> - not applicable <u>Outcome in model development</u> - overall survival
Missing data	<u>Participants with any missing data?</u> - no <u>If yes, how was missing data handled?</u> - not applicable
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> - time-to-first-treatment: 337 persons, 91 events <u>Which model was used?</u> - simplified model <u>Was the model updated?</u> - no - new cut-o& score for risk groups tested <u>Performance measures reported?</u> - Calibration: not reported - Discrimination: reported (c-statistic) <u>Creation of risk groups?</u> - yes, 4 risk groups, (cut-points at (0-1);(2-3);(4-6); (7-10)) - at 5 years

CLL-IPI V - Molica 2016 (O-CLL1-GISL) (Continued)

- percentage of group without event at time point (low risk: 76%, intermediate risk: 45%, high risk: 41%, very high risk: not available)

PROBAST: Applicability	Not applicable; outcome did not match primary outcome of the model.
Notes	<u>Funding & conflict of interest</u> "Contract grant sponsor: NIH; Contract grant no.: 1R01CA197120-01."; "Contract grant sponsors: AIRC (the Italian Association for Cancer Research), a non-profit organization to FM. "Special Program Molecular Clinical Oncology - 5 per mille" n. 9980, 2010/15 and AIRC "Innovative immunotherapeutic treatments of human cancer" n.16695, 2015/18." "Conflict of interest: Nothing to report" <u>Other comments</u> - The same model was validated on the same cohort in other publications.

CLL-IPI V - Muñoz-Novas 2018 (Spanish cohort)

Study characteristics

General information	<u>Model and type of study</u> - Munoz-Novas, 2018 (Spanish cohort) - Validation study <u>Secondary citations</u> - not applicable <u>Language of publication</u> - English <u>Study design</u> - retrospective cohort <u>Follow-up time</u> - median 68 months, range not reported
Participants	<u>Number of included persons in the cohort</u> 696 persons <u>Setting</u> - Spain, multicentre study <u>Recruitment period</u> 1989 - 2013 <u>Age (in years)</u> - median 65.7 years (range: IQR 55.2 - 73.5) <u>Sex</u> 62.8% male

CLL-IPI V - Muñoz-Novas 2018 (Spanish cohort) (Continued)

Stages of disease

- Rai 0: 58.9%; Rai I-IV: 41.1%
- Binet A: not reported

Treatment

- 137 received treatment
- unclear which treatment

Inclusion criteria

- all patients diagnosed with CLL

Exclusion criteria

- not reported

Predictors

List of predictors used for validation (and changes between original predictors and predictors in validation study)

- TP53 status (binary; deleted or mutated); IgHV mutational status (continuous: categorical/binary transformation; cut-o& not given but usually 98%); β 2-microglobulin concentration (continuous: categorical/binary transformation; binary, cut-o& 3.5 mg/L (based on literature)); clinical stage (Rai or Binet) (categorical); Age (continuous: categorical/binary transformation; cut-o& 65 years)
- TP53 not assessed, only del(17p) available

Timing of predictor measurement

- not reported, presumably at diagnosis

Outcome(s)

Primary outcome in study

- overall survival: Overall survival (OS) was calculated from the time of diagnosis to death or last follow-up.

Additional outcome(s)

- time-to-first-treatment: Time to first therapy (TTFT) from the date of diagnosis to first treatment or last follow-up.

Outcome in model development

- overall survival

Missing data

Participants with any missing data?

- yes (438 participants)

If yes, how was missing data handled?

- complete-case analysis

Analysis

Number of participants & number of events (specific time points where reported)

- overall survival: 258 persons, 47 events
- time-to-first-treatment: 258 persons, 113 events

Which model was used?

- simplified model

Was the model updated?

CLL-IPI V - Muñoz-Novas 2018 (Spanish cohort) (Continued)

- no

Performance measures reported?

- Calibration: not reported
- Discrimination: reported (c-statistic) and ROC curve (S2)

Creation of risk groups?

- yes, 4 risk groups, (cut points at (0-1);(2-3);(4-6); (7-10))
- at 5 years
- percentage of group without event at time point (low risk: 93.6%, intermediate risk: 87.6%, high risk: 67.8%, very high risk: 28.6%)

PROBAST: Applicability	<u>Domain 1: Participant selection</u> • low <u>Domain 2: Predictors</u> • unclear - Predictor TP53 not available and replaced by a proxy, del(17p) <u>Domain 3: Outcome</u> • low
Notes	<u>Funding & conflict of interest</u> • Funding not reported • "The authors declare that there are no conflicts of interest regarding the publication of this article." <u>Other comments</u> • none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Appropriate data source used: 'A total of 696 unselected CLL patients newly diagnosed and previously untreated from different institutions of the central region of Spain were included in this study.'
Domain 2: Predictors	Yes	Well-established predictors used - not explicitly stated, but predictors probably assessed in a similar way
Domain 3: Outcome	Yes	Objective standard outcome: no clear description of outcome assessment in this publication. We assume standardised outcome assessment.
Domain 4: Analysis	No	Insufficient number of events Participants with missing values not included in model - consecutive sampling of routine patients, therefore around 2/3 of patients with missing data Calibration not reported
Overall judgement	No	

CLL-IPI V - Rani 2018 (Indian cohort)

Study characteristics

General information	<u>Model and type of study</u> - Rani 2018 (Indian cohort) - Validation study <u>Secondary citations</u> - not applicable <u>Language of publication</u> - English <u>Study design</u> - unclear <u>Follow-up time</u> - median 40.5 months (range: 1 - 215 months)
Participants	<u>Number of included persons in the cohort</u> 198 persons <u>Setting</u> - India, unclear if single or multicentre study <u>Recruitment period</u> - not reported <u>Age (in years)</u> - median 60 years (range: not reported) <u>Sex</u> 77% male <u>Stages of disease</u> - Rai 0: 14.6%; Rai I: 21.2%; Rai II: 34.8%; Rai III: 14.1%; Rai IV: 15.2% - Binet A: not reported <u>Treatment</u> 62 received chlorambucil-based therapy 56 received rituximab-based therapy 20 received other therapies <u>Inclusion criteria</u> - treatment-naïve CLL persons according to Hallek 2008 <u>Exclusion criteria</u> - not reported
Predictors	<u>List of predictors used for validation (and changes between original predictors and predictors in validation study)</u>

CLL-IPI V - Rani 2018 (Indian cohort) (Continued)

- TP53 status (binary; deleted or mutated); IgHV mutational status (continuous: categorical/binary transformation; cut-o& not given but usually 98%); β 2-microglobulin concentration (continuous: categorical/binary transformation; binary, cut-o& 3.5 mg/L (based on literature)); clinical stage (Rai or Binet) (categorical); age (continuous: categorical/binary transformation; cut-o& 65 years)
- only TP53 deletions were used to document TP53 aberrations because of unavailability of TP53 mutation status for all the patients.

Timing of predictor measurement

- not reported

Outcome(s)	<u>Primary outcome in study</u> • overall survival: OS was defined as the time from the date of diagnosis to date of death or date of last follow-up. <u>Additional outcome(s)</u> • time-to-first-treatment: TTFT was defined as the time from the date of diagnosis to date of commencement of first therapy. <u>Outcome in model development</u> • overall survival
Missing data	<u>Participants with any missing data?</u> • no <u>If yes, how was missing data handled?</u> • not applicable
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> • overall survival: 198 persons, 86 events • time-to-first-treatment: 140 persons, 89 events <u>Which model was used?</u> • simplified model <u>Was the model updated?</u> • no <u>Performance measures reported?</u> • Calibration: not reported • Discrimination: reported (c-statistic) <u>Creation of risk groups?</u> • yes, 4 risk groups (cut-points at (0-1);(2-3);(4-6); (7-10)) • median survival (low risk: not reached, intermediate risk: 190 years, high risk: 62 years, very high risk: 28 years)
PROBAST: Applicability	<u>Domain 1: Participant selection</u> • unclear - eligibility criteria were not reported, possibly patients with missing values excluded. We are unsure if the sample is representative for CLL patients. <u>Domain 2: Predictors</u>

CLL-IPI V - Rani 2018 (Indian cohort) (Continued)

- unclear - TP53 was used as deletion only instead of deletion or mutation.

Domain 3: Outcome

- low

Notes

Funding & conflict of interest

- "The financial support was provided by the Department of Biotechnology (BT/PR11106/GBD/27/145/2008, BT/PR15438/MED/30/606/2011 and T/PR8680/AGR/36/754/2013), Ministry of Science and Technology, GOI, and All India Institute of Medical Sciences, New Delhi (8-60/A060/ 2011/RS) to RG for carrying out this work."
- "The authors declare that they have no conflict of interest."

Other comments

- none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	No	Inclusion and exclusion criteria unclear, it seemed like only patients with complete datasets were included. The recruitment period unclear
Domain 2: Predictors	Yes	Well-established predictors used - not explicitly stated, but predictors probably assessed in a similar way (single-centre study).
Domain 3: Outcome	Unclear	Observation time was short.
Domain 4: Analysis	No	Low number of events; no information on handling of missing values; calibration was not reported.
Overall judgement	No	

CLL-IPI V - Reda 2017 (Milano cohort)

Study characteristics

General information

Model and type of study

- Reda 2017 (Milano cohort)
- Validation study

Secondary citations

- not applicable

Language of publication

- English

Study design

- retrospective cohort

Follow-up time

CLL-IPI V - Reda 2017 (Milano cohort) (Continued)

- median 144 months, range not reported

Participants	<u>Number of included persons in the cohort</u> • 698 persons <u>Setting</u> • Italy, single-centre study <u>Recruitment period</u> • 1983 - 2016 <u>Age (in years)</u> • median 65 years (range: 32 - 91 years) <u>Sex</u> • 57% male <u>Stages of disease</u> • Rai 0-I: 74%; Rai II-IV: not reported • Binet A: 82%; Binet B/C: not reported <u>Treatment</u> • 190 received Chlorambucil (CHL) • 108 received fludarabinecytosphamide-rituximab (FCR) • 12 received ofatumumab bendamustine (O-Benda) • 44 received bendamustine-rituximab (BR) • 65 received alemtuzumab • 29 received Ibrutinib or Idelalisib <u>Inclusion criteria</u> • all patients diagnosed with CLL reclassified by 2008 criteria <u>Exclusion criteria</u> • not reported
Predictors	<u>List of predictors used for validation (and changes between original predictors and predictors in validation study)</u> • TP53 status (binary; deleted or mutated); IgHV mutational status (continuous: categorical/binary transformation; cut-o& not given but usually 98%); β2-microglobulin concentration (continuous: categorical/binary transformation; binary, cut-o& 3.5 mg/L (based on literature)); clinical stage (Rai or Binet) (categorical); age (continuous: categorical/binary transformation; cut-o& 65 years) • It was not clear if TP53 was used, or its proxy del(17p) instead. <u>Timing of predictor measurement</u> • not reported, probably at diagnosis (medical records)
Outcome(s)	<u>Primary outcome in study</u> • overall survival: OS was defined as time from CLL diagnosis to death. <u>Additional outcome(s)</u> • time-to-first-treatment: TTFT was evaluated from the time of diagnosis to first-line therapy start.

CLL-IPI V - Reda 2017 (Milano cohort) (Continued)

	<u>Outcome in model development</u> • overall survival
Missing data	<u>Participants with any missing data?</u> • yes (369 participants) <u>If yes, how was missing data handled?</u> • unclear
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> • overall survival: 329 persons, number of events not reported • time-to-first-treatment: 329 persons, number of events not reported <u>Which model was used?</u> • simplified model <u>Was the model updated?</u> • no <u>Performance measures reported?</u> • Calibration: not reported • Discrimination: not reported <u>Creation of risk groups?</u> • yes, 4 risk groups, (cut-points at (0-1);(2-3);(4-6); (7-10)) • no time point and numerical results reported, see figure 1
PROBAST: Applicability	<u>Domain 1: Participant selection</u> • unclear - long observation period (1983 - 2016), no clear eligibility criteria <u>Domain 2: Predictors</u> • unclear - not much information on predictors reported, mentioning of FISH data but unclear if TP53 or del(17p) was used as predictor <u>Domain 3: Outcome</u> • low
Notes	<u>Funding & conflict of interest</u> • Funding not reported • "Authors have no affiliations that they consider to be relevant and important with any organization that to any author's knowledge has a direct interest, particularly a financial interest, in the subject matter discussed." <u>Other comments</u> • Hematology Unit, IRCCS Ca' Granda Ospedale Maggiore Policlinico of Milan, Milano, Italy
Item	Authors' judgement Support for judgement

CLL-IPI V - Reda 2017 (Milano cohort) (Continued)

Domain 1: Participant selection	Unclear	No eligibility criteria reported
Domain 2: Predictors	Yes	Well-established predictors used - not explicitly stated, but predictors probably assessed in a similar way (single-centre study)
Domain 3: Outcome	Yes	Objective standard outcome: no clear description of outcome assessment in this publication. We assumed standardised outcome assessment
Domain 4: Analysis	No	Unclear how many events and how many patients etc. were included for validation Patients with missing values were excluded from analysis. No performance measures and survival per group reported (figure 1 not sharp)
Overall judgement	No	

CLL-IPI V - Rigolin 2017 (Ferrera cohort)

Study characteristics

General information	<u>Model and type of study</u> - Rigolin 2017 (Ferrera cohort) - Validation study <u>Secondary citations</u> - not applicable <u>Language of publication</u> - English <u>Study design</u> - retrospective cohort <u>Follow-up time</u> - not reported
Participants	<u>Number of included persons in the cohort</u> 335 persons <u>Setting</u> - Italy, single-centre study <u>Recruitment period</u> 2006 - 2016 <u>Age (in years)rrr</u> - median 68.7 years (range: 33 - 96 years) <u>Sex</u>

CLL-IPI V - Rigolin 2017 (Ferrera cohort) (Continued)

- 57.9% male

Stages of disease

- Rai: not reported
- Binet A: 77.9%; Binet B: 14.3%; Binet C: 7.8%

Treatment

- 114 received chemotherapy

Inclusion criteria

- all patients diagnosed with CLL at 'our center'

Exclusion criteria

- not reported

Predictors	<u>List of predictors used for validation (and changes between original predictors and predictors in validation study)</u> • TP53 status (binary; deleted or mutated); IgHV mutational status (continuous: categorical/binary transformation; cut-o& not given but usually 98%); β2-microglobulin concentration (continuous: categorical/binary transformation; binary, cut-o& 3.5 mg/L (based on literature)); clinical stage (Rai or Binet) (categorical); age (continuous: categorical/binary transformation; cut-o& 65 years) • no changes <u>Timing of predictor measurement</u> • not reported
Outcome(s)	<u>Primary outcome in study</u> • overall survival: OS was calculated from the date of diagnosis until death due to any cause or until the last patient follow-up. <u>Additional outcome(s)</u> • time-to-first-treatment: TTFT was calculated as the interval between diagnosis and the start of first-line treatment. <u>Outcome in model development</u> • overall survival
Missing data	<u>Participants with any missing data?</u> • unclear or not reported <u>If yes, how was missing data handled?</u> • not applicable
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> • 335 participants, number of events not reported <u>Which model was used?</u> • simplified model <u>Was the model updated?</u> • no

CLL-IPI V - Rigolin 2017 (Ferrera cohort) (Continued)

Performance measures reported?

- Calibration: not reported
- Discrimination: not reported

Creation of risk groups?

- yes, 4 risk groups, (cut-points at (0-1);(2-3);(4-6); (7-10))
- no time point and numerical results reported, see figure 1

PROBAST: Applicability	<u>Domain 1: Participant selection</u> • unclear - no eligibility criteria <u>Domain 2: Predictors</u> • low <u>Domain 3: Outcome</u> • low
Notes	<u>Funding & conflict of interest</u> • "... Fondo di Ateneo per la Ricerca 2013, 2014, 2016 of the University of Ferrara (G.M.R., A.C., M.N.), Programma Ricerca Regione Università 2007-2009 University of Ferrara (G.M.R., A.C.), Program- mi di Ricerca di Rilevante Interesse Nazionale (PRIN) 2008 (A.C.), Ricerca Finalizzata (A.C.; project RF-2011-02349712), Ministero dell'Istruzione, dell'Università e della Ricerca PRIN 2015 (A.C.; project 2015ZMRFEA), and AIL Ferrara. E.V. and E.S. are supported by AIL Ferrara." • "The authors declare no competing financial interests." <u>Other comments</u> • The authors provided their primary data and, in addition, a screen capture of the Stata output of their analysis.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	No eligibility criteria
Domain 2: Predictors	Yes	Well-established predictors used - not explicitly stated, but predictors probably assessed in a similar way (single-centre study)
Domain 3: Outcome	Unclear	Objective standard outcome, however, observation time not reported
Domain 4: Analysis	No	No information on number of events, missing data and missing data handling; no performance measures reported
Overall judgement	No	

CLL-IPI V - Zhu 2018 (Chinese cohort)

Study characteristics

General information	<u>Model and type of study</u>

CLL-IPI V - Zhu 2018 (Chinese cohort) (Continued)

- Zhu 2018 (Chinese cohort)
- Validation study

Secondary citations

- not applicable

Language of publication

- Chinese

Study design

- retrospective cohort

Follow-up time

- median 48 months, (range: 1 - 192 months)

Participants

Number of included persons in the cohort

- 215 persons

Setting

- China, single-centre study

Recruitment period

- 2002 - 2017

Age (in years)

- median 60 years (range 16 - 85 years)

Sex

- 66.5% male

Stages of disease

- Rai 0: 16.7%; Rai I-II: 54.0%; Rai III-IV: 29.3%
- Binet A: 44.2%; Binet B: 27.9%; Binet C: 27.9%

Treatment

- 147 received any of the following:
- benzodiazepine
- FCR (fludarabine, cyclophosphamide, rituximab),
- bendamoxetone
- rituximab
- R+FFP+HDMP (rituximab + fresh frozen plasma + high-dose methylprednisolone)

Inclusion criteria

- newly-diagnosed CLL patients treated in our hospital between 2002 and 2017

Exclusion criteria

- not reported

Predictors

List of predictors used for validation (and changes between original predictors and predictors in validation study)

CLL-IPI V - Zhu 2018 (Chinese cohort) (Continued)

- TP53 status (binary; deleted or mutated); IgHV mutational status (continuous: categorical/binary transformation; cut-o& not given but usually 98%); β 2-microglobulin concentration (continuous: categorical/binary transformation; binary, cut-o& 3.5 mg/L (based on literature)); clinical stage (Rai or Binet) (categorical); age (continuous: categorical/binary transformation; cut-o& 65 years)
- no changes

Timing of predictor measurement

- at diagnosis

Outcome(s)

Primary outcome in study

- overall survival: OS time was defined as the interval from a definitive diagnosis to the end point of death or follow-up for any cause.

Additional outcome(s)

- time-to-first-treatment: TTFT was defined as the interval from the definitive diagnosis to the initiation of treatment.

Outcome in model development

- overall survival

Missing data

Participants with any missing data?

- no

If yes, how was missing data handled?

- not applicable

Analysis

Number of participants & number of events (specific time points where reported)

- overall survival: 215 persons, 46 events
- time-to-first-treatment: 215 persons, 147 events

Which model was used?

- simplified model

Was the model updated?

- no

Performance measures reported?

- Calibration: was provided upon request (observed-expected ratio per risk group, calibration plot, Hosmer and Lemeshow test)
- Discrimination: reported (AUC)

Creation of risk groups?

- yes, 4 risk groups, (cut-points at (0-1);(2-3);(4-6); (7-10))
- at 5 years
- percentage of group without event at time point (low risk: 97.6%, intermediate risk: 83.7%, high risk: 67.8%, very high risk: 55.2%)

PROBAST: Applicability

Domain 1: Participant selection

- unclear - no eligibility criteria reported.

Domain 2: Predictors

CLL-IPI V - Zhu 2018 (Chinese cohort) (Continued)

- low

Domain 3: Outcome

- low

Notes

Funding & conflict of interest

- not reported

Other comments

- original study translated by Yuan Chi (Centre for Evidence-Based Chinese Medicine, Beijing University of Chinese Medicine, China)
- the authors provided additional data upon request (calibration plot, O:E ratios per risk group, AUC).

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	No eligibility criteria reported, unclear if patients with missing values were excluded
Domain 2: Predictors	Yes	Well-established predictors used - not explicitly stated, but predictors probably assessed in a similar way
Domain 3: Outcome	Yes	Objective standard outcome
Domain 4: Analysis	No	Number of events low (46) No information on missing values The authors provided the calibration and discrimination upon request.
Overall judgement	No	

GCLLSG D - Pflug 2014 (GCLLSG)

Study characteristics

General information

Model and type of study

- Pflug 2014 (development cohort)
- Development study

Secondary citations

- Tam 2014

Language of publication

- English

Study design

- combined data (e.g. IPD or combination of cohorts); data from 3 RCTs

Follow-up time

GCLLSG D - Pflug 2014 (GCLLSG) (Continued)

- median 63.4 months (range not reported)

Participants	<u>Number of included persons in the cohort</u> • 2007 persons <u>Setting</u> • multicentre study • Australia, Austria, Belgium, Czech Republic, Denmark, France, Germany, Israel, Italy, New Zealand, Spain <u>Recruitment period</u> • 1997 - 2006 <u>Age (in years)</u> • median 60 years (range: 30 - 81 years) <u>Sex</u> • 68.4% male <u>Stages of disease</u> • Rai 0: 20.8%; Rai I: 18.4%; Rai II: 37.9%; Rai III: 8.6%; Rai IV: 14.3% • Binet A: 42.6%; Binet B: 37.7%; Binet C: 19.8% <u>Treatment</u> • participants were treated with: • early fludarabine (F) • F plus cyclophosphamide (FC) • FC plus rituximab (FCR) <u>Inclusion criteria</u> • different inclusion criteria per clinical trial (CLL1, CLL4 and CLL8 trials); common criteria: confirmed diagnosis of B-cell chronic lymphocytic leukaemia, untreated patient, life expectancy above 6 months <u>Exclusion criteria</u> • different exclusion criteria per clinical trial (CLL1, CLL4 and CLL8 trials)
Predictors	<u>Number of candidate predictors</u> • univariable analysis: 20 • multivariable analysis: 18 <u>List of predictors in final model (including cut-points for dichotomised factors)</u> • sex (binary); age (continuous, dichotomisation after univariable analysis); ECOG PS (categorical, dichotomisation after univariable analysis); del(17p) (continuous: categorical/binary transformation); del(11q) (continuous: categorical/binary transformation; literature-based); IgHV mutational status (continuous: categorical/binary transformation; literature-based); s-TK (continuous, dichotomisation after univariable analysis); s-beta2m (continuous, dichotomisation after univariable analysis) <u>Timing of predictor measurement</u> • not reported, presumably at study entry
Outcome(s)	<u>Primary outcome in study</u>

GCLLSG D - Pflug 2014 (GCLLSG) *(Continued)*

- overall survival (the main end point of statistical analyses was OS defined as the time between registration/randomisation and death)

Additional outcome(s)

- not applicable

Missing data

Participants with any missing data?

- yes (784 participants)

If yes, how was missing data handled?

- complete-case analysis

Analysis

Number of participants & number of events (specific time points where reported)

- overall survival: 1223 participants, number of events included for analysis unclear

Predictor selection method

- univariable analysis, during multivariable modelling: stepwise regression, (P value or CI)

Statistical method

- Cox proportional hazard model

Simplification of model?

- yes

Performance measures reported?

- Calibration: not reported
- Discrimination: reported (c-statistic)

Creation of risk groups?

- yes, 4 risk groups, (cut-points at (0-2); (3-5); (6-10); (11-14))
- at 5 years
- percentage of group without event at time point (low risk: 95.2 %, intermediate risk: 86.9 %, high risk: 67.6 %, very high risk: 18.7 %)

PROBAST: Applicability

Domain 1: Participant selection

- unclear - the study used RCT data with specific inclusion and exclusion criteria, patient population younger than average

Domain 2: Predictors

- low

Domain 3: Outcome

- low

Notes

Funding & conflict of interest

- "This manuscript was written on behalf of the German CLL Study Group. Studies CLL1, CLL4, and CLL8 were planned and conducted as investigator-initiated trials by the German CLL Study Group and were supported by research grants from German Cancer Aid, Medac Schering Onkologie, and F. Hoffmann-La Roche. T.D.S. is a clinical scholar of the Leukemia Lymphoma Society."
- "N.P. received Travel Grants from Roche. T.D.S. received research grants from Genentech, Celgene, Glaxo-Smith-Kline, Cephalon, Hospira, and Polyphenon E International. B.E. is a consultant and/or

GCLLSG D - Pflug 2014 (GCLLSG) (Continued)

holds an advisory role for Celgene and Pharmacyclics and has received honoraria and research funding from Roche and Mundipharma. S.S. is a consultant and/or holds an advisory role for Roche and Mundipharma, and received honoraria and research funding from both. H.D. received research grants from Roche. U.J. received honoraria and research funding from Roche. M.H. is a consultant and/or holds an advisory role and received research funding from Roche. J.B., M.A.B., T.E., K.B., G.M., K.G.R., M.J.E., G.H., R.B., A.-M.F., C.-M.W., K.F., and N.E.K. declare no competing financial interests."

Other comments

- none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Appropriate data sources used: combination of three RCT cohorts
Domain 2: Predictors	Yes	Not explicitly stated in this publication, but we assumed standard assessment in the context of each RCT
Domain 3: Outcome	Yes	Objective standard outcome, follow-up in the context of each individual RCT
Domain 4: Analysis	No	Univariable selection of predictors Bootstrapping, however only for multivariable modeling and not used for correction of optimism; complete case analysis, background sample similar but many missing values; calibration not reported; simplification of the model, thus the assigned weights did not correspond to the results of multivariable modeling.
Overall judgement	No	

GCLLSG V - Molica 2015 (O-CLL1-GISL)

Study characteristics

General information	<u>Model and type of study</u> • Molica 2015 (O-CLL1-GISL) • Validation study <u>Secondary citations</u> • Molica 2013; Molica 2017; Molica 2017 <u>Language of publication</u> • English <u>Study design</u> • prospective cohort <u>Follow-up time</u> • median 42 months (range: 1 - 82 months; 2038 person-years)
Participants	<u>Number of included participants in the cohort</u>

GCLLSG V - Molica 2015 (O-CLL1-GISL) (Continued)

- 337 participants

Setting

- Italy, multicentre study

Recruitment period

- 2007 - not reported

Age (in years)

- median 61 years (range: 33 - 70 years)

Sex

- 57.1% male

Stages of disease

- Rai 0: 77.8%; Rai I-IV: not reported
- Binet A: 100 %

Treatment

- 91 received treatment
- unclear which treatment

Inclusion criteria

- newly-diagnosed CLL patients from several Italian Institutions who were seen within 12 months of diagnosis were prospectively enrolled into the OCLL1-GISL protocol. The inclusion criteria for CLL diagnosis, employed at the time of study design, were those of the National Cancer Institute (NCI)-sponsored Working Group guidelines (NCI-WG). Patients enrolled did not require therapy according to NCI guidelines (i.e. asymptomatic Binet stage A).

Exclusion criteria

- not reported

Predictors

List of predictors used for validation (and changes between original predictors and predictors in validation study)

- sex (binary); age (continuous, dichotomisation after univariable analysis); ECOG PS (categorical, dichotomisation after univariable analysis); del(17p) (continuous: categorical/binary transformation); del(11q) (continuous: categorical/binary transformation; literature-based); IgHV mutational status (continuous: categorical/binary transformation; literature-based); s-TK (continuous, dichotomisation after univariable analysis); s-beta2m (continuous, dichotomisation after univariable analysis)
- TK was omitted.

Timing of predictor measurement

- at study entry, within 12 months of diagnosis

Outcome(s)

Primary outcome in study

- time-to-first-treatment (TTFT was defined as the interval between the date of database registration and the date of first CLL treatment. Patients underwent sequential monitoring, and the frequency of follow-up visits was individualised according to patient risk; this ranged from 3 to 6 months (median 6 months). All physicians who registered patients in this observational database stated that they had used the NCI-WG guidelines as a reference criterion for starting therapy. In particular, the absolute lymphocyte count was not used as the sole indicator for treatment. Active disease, requiring therapy, was defined when at least one of the criteria set out in the NCI-sponsored Working Group guidelines was satisfied).

GCLLSG V - Molica 2015 (O-CLL1-GISL) (Continued)

	<u>Additional outcome(s)</u> - not applicable <u>Outcome in model development</u> - overall survival
Missing data	<u>Participants with any missing data?</u> - no <u>If yes, how was missing data handled?</u> - not applicable
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> - time-to-first-treatment: 337 participants, 91 events <u>Which model was used?</u> - simplified model <u>Was the model updated?</u> - no <u>Performance measures reported?</u> - Calibration: not reported - Discrimination: reported (c-statistic) <u>Creation of risk groups?</u> - no
PROBAST: Applicability	Not applicable; outcome did not match primary outcome of the model.
Notes	<u>Funding & conflict of interest</u> "Contract grant sponsor: NIH; Contract grant number: 1R01CA197120-01."; "Contract grant sponsors: AIRC (the Italian Association for Cancer Research), a non-profit organization to FM. "Special Program Molecular Clinical Oncology - 5 per mille" n. 9980, 2010/15 and AIRC "Innovative immunotherapeutic treatments of human cancer" n.16695, 2015/18." (Molica 2016) "Conflict of interests: nothing to report." (Molica 2016) <u>Other comments</u> - The validation of this model in the O-CLL1-GISL cohort appeared in several publications. The c-statistic was reported differently in each of them. - The authors did not reply to our request for additional information.

GCLLSG V - Pflug 2014 (Mayo cohort)

Study characteristics

General information	<u>Model and type of study</u> - Pflug 2014 (Mayo clinic cohort, undefined time range) - Validation study

GCLLSG V - Pflug 2014 (Mayo cohort) (Continued)

Secondary citations

- not applicable

Language of publication

- English

Study design

- prospective cohort

Follow-up time

- median 57 months (range not reported)

Participants

Number of included participants in the cohort

- 676 participants

Setting

- USA, multicentre study

Recruitment period

- not reported

Age (in years)

- median 61.5 years (range: 32 - 89 years)

Sex

- 67% male

Stages of disease

- Rai 0: 57.1%; Rai I: 34%; Rai II: 5.9%; Rai III: 1%; Rai VI: 1.9%
- not reported

Treatment

- 190 received treatment
- unclear which treatment

Inclusion criteria

- consecutive series of 676 newly-diagnosed, prospectively-followed CLL patients cared for at Mayo Clinic who had baseline data on all considered variables except s-TK and/or s-B2M available and who had stored serum collected less than 36 months (median, 1 month) of diagnosis available for s-TK and s-B2M analysis.

Exclusion criteria

- not reported

Predictors

List of predictors used for validation (and changes between original predictors and predictors in validation study)

- sex (binary); age (continuous, dichotomisation after univariable analysis); ECOG PS (categorical, dichotomisation after univariable analysis); del(17p) (continuous: categorical/binary transformation); del(11q) (continuous: categorical/binary transformation; literature-based); IgHV mutational status (continuous: categorical/binary transformation; literature-based); s-TK (continuous, dichotomisation

GCLLSG V - Pflug 2014 (Mayo cohort) (Continued)

	after univariable analysis); serum β2-microglobulin (continuous, dichotomisation after univariable analysis) - TK from frozen blood samples within 36 months of diagnosis; different assessment methods for beta2m than in development study <u>Timing of predictor measurement</u> - at study entry
Outcome(s)	<u>Primary outcome in study</u> - overall survival: the main end point of statistical analyses was OS defined as the time between registration/randomisation and death. <u>Additional outcome(s)</u> - time-to-first-treatment: TFS was calculated from the date of diagnosis to the start of the first CLL treatment. <u>Outcome in model development</u> - overall survival
Missing data	<u>Participants with any missing data?</u> - yes (3 participants) <u>If yes, how was missing data handled?</u> - complete-case analysis
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> - overall survival: 673 participants, 85 events - time-to-first-treatment: 673 participants, 190 events <u>Which model was used?</u> - simplified model <u>Was the model updated?</u> - no <u>Performance measures reported?</u> - Calibration: not reported - Discrimination: reported (c-statistic) <u>Creation of risk groups?</u> - yes, 4 risk groups, (cut-points at (0-2); (3-5); (6-10); (11-14)) - at 5 years - percentage of group without event at time point (low risk: 95.2%, intermediate risk: 91.4%, high risk: 71.7%, very high risk: 13.6%)
PROBAST: Applicability	<u>Domain 1: Participant selection</u> - unclear - unclear recruitment period, data availability as inclusion criterion <u>Domain 2: Predictors</u> - low - differences in the measurement of one predictor, which is usually not available in the USA as compared to Germany (serum thymidine kinase): interassay calibration

GCLLSG V - Pflug 2014 (Mayo cohort) (Continued)

Domain 3: Outcome

- low

Notes

Funding & conflict of interest

- "This manuscript was written on behalf of the German CLL Study Group. Studies CLL1, CLL4, and CLL8 were planned and conducted as investigator-initiated trials by the German CLL Study Group and were supported by research grants from German Cancer Aid, Medac Schering Onkologie, and F. Hoffmann-La Roche. T.D.S. is a clinical scholar of the Leukemia Lymphoma Society."
- "N.P. received Travel Grants from Roche. T.D.S. received research grants from Genentech, Celgene, Glaxo-Smith-Kline, Cephalon, Hospira, and Polyphenon E International. B.E. is a consultant and/or holds an advisory role for Celgene and Pharmacyclics and has received honoraria and research funding from Roche and Mundipharma. S.S. is a consultant and/or holds an advisory role for Roche and Mundipharma, and received honoraria and research funding from both. H.D. received research grants from Roche. U.J. received honoraria and research funding from Roche. M.H. is a consultant and/or holds an advisory role and received research funding from Roche. J.B., M.A.B., T.E., K.B., G.M., K.G.R., M.J.E., G.H., R.B., A.-M.F., C.-M.W., K.F., and N.E.K. declare no competing financial interests."

Other comments

- none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	No	Only patients with available data included: "... who had baseline data on all considered variables except s-TK and/or s-b2m available and who had stored serum collected > 36 months ..."
Domain 2: Predictors	Yes	Standard laboratory measures used. Where s-TK and s-B2M were not available, blood samples were shipped to Germany and analysed.
Domain 3: Outcome	Yes	Objective standard outcome - no clear description of outcome assessment - we assumed that assessment of the objective outcome OS was similar and therefore did not rate as high risk (no info on e.g. registry, frequency of follow-up, etc.)
Domain 4: Analysis	No	Low number of events; patients with missing data were not enrolled in the study, rated in domain 1; calibration not reported
Overall judgement	No	

GCLLSG V - Rani 2018 (Indian cohort)

Study characteristics

General information

Model and type of study

- Rani 2018 (Indian cohort)
- Validation study

Secondary citations

- not applicable

Language of publication

GCLLSG V - Rani 2018 (Indian cohort) (Continued)

- English

Study design

- unclear

Follow-up time

- median 40.5 months (range: 1 - 215 months)

Participants	<u>Number of included participants in the cohort</u> 198 participants <u>Setting</u> - India, unclear if single or multicentre study <u>Recruitment period</u> - not reported <u>Age (in years)</u> - median 60 years (range not reported) <u>Sex</u> 77% male <u>Stages of disease</u> - Rai 0: 14.6%; Rai I: 21.2%; Rai II: 34.9%; Rai III: 14.1%; Rai IV: 15.2% - Binet A/B/C: not reported <u>Treatment</u> 62 received chlorambucil-based therapy 56 received rituximab-based therapy 20 received other therapies <u>Inclusion criteria</u> - treatment-naïve CLL patients according to Hallek 2008 <u>Exclusion criteria</u> - not reported
Predictors	<u>List of predictors used for validation (and changes between original predictors and predictors in validation study)</u> - sex (binary); age (continuous, dichotomisation after univariable analysis); ECOG PS (categorical, dichotomisation after univariable analysis); del(17p) (continuous: categorical/binary transformation); del(11q) (continuous: categorical/binary transformation; literature-based); IgHV mutational status (continuous: categorical/binary transformation; literature-based); s-TK (continuous, dichotomisation after univariable analysis); s-beta2m (continuous, dichotomisation after univariable analysis) <u>Timing of predictor measurement</u> - not reported
Outcome(s)	<u>Primary outcome in study</u>

GCLLSG V - Rani 2018 (Indian cohort) (Continued)

- time-to-first-treatment (TTFT was defined as the time from the date of diagnosis to date of commencement of first therapy)

Additional outcome(s)

- not applicable

Outcome in model development

- overall survival

Missing data	<u>Participants with any missing data?</u> - no <u>If yes, how was missing data handled?</u> - not applicable
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> - time-to-first-treatment: 140 participants, 89 events <u>Which model was used?</u> - simplified model <u>Was the model updated?</u> - no <u>Performance measures reported?</u> - Calibration: not reported - Discrimination: reported (c-statistic) <u>Creation of risk groups?</u> - yes, 4 risk groups, (cut-points at (0-2); (3-5); (6-10); (11-14) - median survival (low risk: not reached, intermediate risk: 36 years, high risk: 4 years, very high risk: 0.01 years)
PROBAST: Applicability	Not applicable, outcome did not match primary outcome of the model
Notes	<u>Funding & conflict of interest</u> "The financial support was provided by the Department of Biotechnology (BT/PR11106/GBD/27/145/2008, BT/PR15438/MED/30/606/2011 and T/PR8680/AGR/36/754/2013), Ministry of Science and Technology, GOI, and All India Institute of Medical Sciences, New Delhi (8-60/A060/ 2011/RS) to RG for carrying out this work." "The authors declare that they have no conflict of interest." <u>Other comments</u> - Graphs were provided to obtain survival at a specific time point, however, these graphs were printed very small.

GIMEMA D - Molica 2005 (GIMEMA cohort)

Study characteristics

GIMEMA D - Molica 2005 (GIMEMA cohort) (Continued)

General information	<u>Model and type of study</u> • Molica 2005 (GIMEMA cohort) • Development study <u>Secondary citations</u> • not applicable <u>Language of publication</u> • English <u>Study design</u> • retrospective cohort <u>Follow-up time</u> • median 54 months, range 4 - 309 months
Participants	<u>Number of included participants in the cohort</u> • 1138 participants <u>Setting</u> • Italy, multicentre study <u>Recruitment period</u> • 1991 - 2000 <u>Age (in years)</u> • median 65 years (range: 27 - 100 years) <u>Sex</u> • 54% male <u>Stages of disease</u> • Rai 0: 77.2 %; Rai I-III: 22.7% • Binet A: 100% <u>Treatment</u> • 133 received chlorambucil • 20 received fludarabine alone or in association with cyclophosphamide • 7 received CHOP-like regimens • 40 received other therapies (CVP, cyclophosphamide with or without prednisone) <u>Inclusion criteria</u> • All previously untreated B-cell CLL patients in Binet stage A whose diagnosis was immunologically confirmed (CD5 +/- Smlg weak) and who were observed at different GIMEMA centres during the period 1991 – 2000 were considered eligible for this study. <u>Exclusion criteria</u> • inadequate follow-up
Predictors	<u>Number of candidate predictors</u>

GIMEMA D - Molica 2005 (GIMEMA cohort) (Continued)

- univariable analysis: 10 or more
- multivariable analysis: 6

List of predictors in final model (including cut-points for dichotomised factors)

- Lymphocyte doubling time (LDT, continuous: categorical/binary transformation; dichotomised to above and below 12 months); absolute peripheral blood lymphocytosis (continuous: categorical/binary transformation; two categories, above and below $30 \times 10^9/l$); Rai stage (categorical; 0 vs. I-III); gender (binary)

Timing of predictor measurement

- at diagnosis

Outcome(s)	<u>Primary outcome in study</u> • progression-free survival (progressive disease was defined on the basis of a shIL to a more advanced clinical stage (i.e. from A to B or C) and/or LDT larger than 12 months. For those individuals who did not change Binet stage nor doubled their lymphocyte count, disease progression was considered when one of the following conditions was met: decrease in haemoglobin larger than 3 g/dL, splenomegaly larger than 6 cm below the leL costal margin, massive lymphadenopathy (i.e. larger than 8 cm) or progressive lymphadenopathy and/or peripheral blood lymphocytosis more than $100 \times 10^9 /l$) <u>Additional outcome(s)</u> • not applicable
Missing data	<u>Participants with any missing data?</u> • yes (545 participants) <u>If yes, how was missing data handled?</u> • complete-case analysis
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> • progression-free survival: 593 participants, 235 events in overall sample, unclear number of events in sample used for modeling <u>Predictor selection method</u> • univariable analysis, during multivariable modelling: unclear if stepwise regression was used (criteria were P value or CI) <u>Statistical method</u> • Cox proportional hazard model, stratification of the resulting score by gender <u>Simplification of model?</u> • yes <u>Performance measures reported?</u> • Calibration: not reported • Discrimination: not reported <u>Creation of risk groups?</u> • yes, 3 risk groups (risk groups were: no risk factors; one risk factor; more than one risk factor) • at 10 years

GIMEMA D - Molica 2005 (GIMEMA cohort) (Continued)

- percentage of group without event at time point (low risk: 67.8%, intermediate risk: 41%, high risk: 24.8%)

PROBAST: Applicability	<u>Domain 1: Participant selection</u> - low <u>Domain 2: Predictors</u> - low <u>Domain 3: Outcome</u> - low
Notes	<u>Funding & conflict of interest</u> - not reported <u>Other comments</u> - The model development characteristics were not well reported.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Appropriate data sources used (retrospective cohort); 1.7% of patients were excluded due to inadequate follow-up. We rated this as low proportion of exclusions.
Domain 2: Predictors	Yes	The predictors used seem to be routine demographic and laboratory information, and although there was no explicit information, we think that the methods were consistent.
Domain 3: Outcome	No	Knowledge of the predictor was used both in the definition of progression and the model (peripheral blood lymphocytes).
Domain 4: Analysis	No	Individuals with missing data were excluded: "The outcome of the 593 patients with complete data was different from the outcome of the 545 patients with incomplete data (10 year PFS 55.5% vs. 70.2%). The reasons for this difference rely on the higher number of patients in Rai stages I – III ($P = 0.01$) and increased serum levels of b2-m ($P = 0.003$) found in the subgroup with complete data"; univariable selection of predictors; no weighting of predictors: "These differences in the RR did not prevent us from constructing a risk score simply by adding the negative factors present in a single patient at the time of diagnosis."; no model performance measures reported
Overall judgement	No	

GIMEMA V - González Rodríguez 2009 (Cabueñes coh.)

Study characteristics

General information	<u>Model and type of study</u> - Gonzalez Rodriguez 2009 (Hospital de Cabuenes) - Validation study

GIMEMA V - González Rodríguez 2009 (Cabueñes coh.) (Continued)

Secondary citations

- not applicable

Language of publication

- Spanish

Study design

- retrospective cohort

Follow-up time

- not reported

Participants

Number of included participants in the cohort

- 265 participants

Setting

- Spain, single-centre study

Recruitment period

- 1997 - 2007

Age (in years)

- median 71.7 years (range: 42 - 94 years)

Sex

- 58.5% male

Stages of disease

- modified Rai low: 61.2%; Rai intermediate: 29.3%; Rai high: 9.5%
- Binet A: 76.8%; Binet B: 14.8%; Binet C: 8.4%

Treatment

- not reported

Inclusion criteria

- Only cases with an increase of more than $5 \times 10^9/l$ lymphocytes with morphology and immunophenotype of B-CLL and with Royal Marsden Scoring System greater than 3, and cases of CLL confirmed by lymph node biopsy have been included. The inclusion criteria were based on the initial data at the time of diagnosis.

Exclusion criteria

- not reported

Predictors

List of predictors used for validation (and changes between original predictors and predictors in validation study)

- Lymphocyte doubling time (LDT, continuous: categorical/binary transformation; dichotomised to above and below 12 months); absolute peripheral blood lymphocytosis (continuous: categorical/binary transformation; two categories, above and below $30 \times 10^9/l$); Rai stage (categorical; 0 vs. I-III); gender (binary)
- no changes

Timing of predictor measurement

GIMEMA V - González Rodríguez 2009 (Cabueñes coh.) (Continued)

	at diagnosis
Outcome(s)	<u>Primary outcome in study</u> - overall survival: to estimate the overall survival (OS), the date of diagnosis and the date of death for any cause or the final date of the study were considered. <u>Additional outcome(s)</u> - time-to-first-treatment: treatment-free survival was considered as the time elapsed from diagnosis until the start date of the treatment. <u>Outcome in model development</u> - progression-free survival
Missing data	<u>Participants with any missing data?</u> - yes (8 participants) <u>If yes, how was missing data handled?</u> - complete-case analysis
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> - overall survival: 257 participants, 76 events - time-to-first-treatment: 257 participants, 67 events <u>Which model was used?</u> - simplified model <u>Was the model updated?</u> - no <u>Performance measures reported?</u> - Calibration: not reported - Discrimination: reported (AUC) <u>Creation of risk groups?</u> - yes, 3 risk groups - at 5 years - percentage of group surviving at 5 years (low risk: 79%, intermediate risk: 67%, high risk: 79%)
PROBAST: Applicability	<u>Domain 1: Participant selection</u> - low <u>Domain 2: Predictors</u> - low <u>Domain 3: Outcome</u> - low
Notes	<u>Funding & conflict of interest</u> This work was partially funded through research project FIS06/ 0841 from the Health Research Fund.

GIMEMA V - González Rodríguez 2009 (Cabueñes coh.) (Continued)

Other comments

- The original study was translated by Leonardo Perales-Guerrero (medical student, Universidad de Guadalajara, Mexico).
- The authors could not be contacted since the email address of the contact author as stated in the publication was no longer valid.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Appropriate data sources used: retrospective cohort
Domain 2: Predictors	Yes	The predictors used seem to be routine demographic and laboratory information, and although there was no explicit information, we think that the methods are consistent.
Domain 3: Outcome	Unclear	No median observation time; standard outcome, but treatment indication not described
Domain 4: Analysis	No	Only small amount of missing outcome data (n = 8); discrimination reported in form of AUC, however, time point not clear (probably 5 years); calibration not reported
Overall judgement	No	

MDACC 2007 D - Wierda 2007 (MDACC)

Study characteristics

General information	<u>Model and type of study</u> • Wierda 2007 (development, MDACC 1981-2004) • Development study <u>Secondary citations</u> • not applicable <u>Language of publication</u> • English <u>Study design</u> • retrospective cohort <u>Follow-up time</u> • median 58.8 months, 95% CI: 55.2 - 61.2
Participants	<u>Number of included participants in the cohort</u> • 1674 participants <u>Setting</u> • US (Texas), single-centre study

MDACC 2007 D - Wierda 2007 (MDACC) (Continued)

Recruitment period

- 1981 - 2004

Age (in years)

- median 58 years (range: 0 - 90 years)

Sex

- 1029 male participants

Stages of disease

- Rai 0: 469 individuals; Rai I: 739 individuals; Rai II: 235 individuals; Rai III: 93 individuals; Rai IV: 127 individuals
- Binet A: 1019 individuals; Binet B: 470 individuals; Binet C: 168 individuals

Treatment

- 902 received treatment
- unclear which treatment

Inclusion criteria

- Previously untreated patients who presented for initial evaluation to MDACC from August 1981 through August 2004 were included in this analysis.

Exclusion criteria

- not reported

Predictors	<u>Number of candidate predictors</u> • univariable analysis: 18 • multivariable analysis: 16 <u>List of predictors in final model (including cut-points for dichotomised factors)</u> • Age (continuous: no transformation; in years); ALC (absolute lymphocyte count) (continuous: no transformation; $\times 10^9/L$); serum $\beta 2$-microglobulin (continuous: no transformation); nodal groups (binary); Rai stage (categorical; binary: stage III or IV vs. 0-II); sex (binary) <u>Timing of predictor measurement</u> • at study entry
Outcome(s)	<u>Primary outcome in study</u> • overall survival (the time interval was measured from the day of presentation to MDACC until death or last follow-up. Death from all causes was included). <u>Additional outcome(s)</u> • not applicable
Missing data	<u>Participants with any missing data?</u> • yes (57 participants) <u>If yes, how was missing data handled?</u> • complete-case analysis
Analysis	<u>Number of participants & number of events (specific time points where reported)</u>

MDACC 2007 D - Wierda 2007 (MDACC) (Continued)

- overall survival: 1617 participants, 437 events

Predictor selection method

- univariable analysis, during multivariable modelling: backward elimination (P value or CI)

Statistical method

- Cox proportional hazard model

Simplification of model?

- yes

Performance measures reported?

- Calibration: not reported
- Discrimination: reported (c-statistic)

Creation of risk groups?

- yes, 3 risk groups, (cut-points at (1-3);(4-7);(above 8))
- at 5 years
- percentage of group without event at time point (low risk: 0.97%, intermediate risk: 0.8%, high risk: 0.55%)

PROBAST: Applicability

Domain 1: Participant selection

- low

Domain 2: Predictors

- low

Domain 3: Outcome

- low

Notes

Funding & conflict of interest

- Funding not reported
- "The authors declare no competing financial interests."

Other comments

- none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Retrospective cohort, consecutive sampling
Domain 2: Predictors	Yes	Routine clinical data used: although the recruitment period was very broad, standard clinical data (gender, age, Rai stage, ALC and beta-2 microglobulin) have been included for modelling
Domain 3: Outcome	Yes	Objective standard outcome: no clear description of outcome assessment; we assumed that assessment of the objective outcome OS was similar and therefore did not rate as high risk.

MDACC 2007 D - Wierda 2007 (MDACC) (Continued)

Domain 4: Analysis	Yes	Number of missing values and therefore the number of patients excluded from analysis was low; continuous factors handled appropriately; performance measures of original model reported; bootstrap correction applied; original formula and a simplified score provided; univariable selection of factors, however, nearly all factors were included in multivariable modelling as well.
Overall judgement	Yes	

MDACC 2007 V - Bulian 2011 (Italian-Swiss)

Study characteristics

General information	<u>Model and type of study</u> - Bulian 2011 (Italian-Swiss cohort) - Validation study <u>Secondary citations</u> - not applicable <u>Language of publication</u> - English <u>Study design</u> - retrospective cohort <u>Follow-up time</u> - median 63.6 months, range not reported
Participants	<u>Number of included participants in the cohort</u> 1480 participants <u>Setting</u> - Italy, Switzerland, multicentre study <u>Recruitment period</u> 1996 - 2011 <u>Age (in years)</u> - median 65 years (range: 21 - 94 years) <u>Sex</u> 58.2% male <u>Stages of disease</u> - Rai 0-II: not reported; Rai III/IV: 7.2% - Binet A: 74.6%; Binet B: 18.4%; Binet C: 6.9% <u>Treatment</u> 306 received chemotherapy

MDACC 2007 V - Bulian 2011 (Italian-Swiss) (Continued)

- 129 received Chemoimmunotherapy
- 42 received FCR
- 49 received FR
- 38 received R-other
- 40 with missing data

Inclusion criteria

- 'The series included all cases observed at each centre over a given time period, although each centre allegedly failed to observe all the incident CLL cases.'

Exclusion criteria

- not reported

Predictors	<u>List of predictors used for validation (and changes between original predictors and predictors in validation study)</u> • Age (continuous: no transformation; in years); ALC (absolute lymphocyte count) (continuous: no transformation; $\times 10^9/L$); serum $\beta 2$-microglobulin (continuous: no transformation); nodal groups (binary); Rai stage (categorical; binary: stage III or IV vs. 0-II); gender (binary) • no changes <u>Timing of predictor measurement</u> • not explicitly reported, but probably at diagnosis, as this was a retrospective cohort using standard demographic and clinical information
Outcome(s)	<u>Primary outcome in study</u> • overall survival <u>Additional outcome(s)</u> • time-to-treatment ('Patient treatment was started according to NCI-WG indications.') <u>Outcome in model development</u> • overall survival
Missing data	<u>Participants with any missing data?</u> • no <u>If yes, how was missing data handled?</u> • not applicable
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> • overall survival: 1037 participants, 151 events <u>Which model was used?</u> • The authors tried to use the nomogram, but had problems due to the low quality of the graph in the development publication; additionally, they used the index score. <u>Was the model updated?</u> • yes • 'An additional model was developed by replacing the dichotomous variables Rai staging (0–II vs. III–IV) with the three-level Binet staging.' <u>Performance measures reported?</u>

MDACC 2007 V - Bulian 2011 (Italian-Swiss) (Continued)

- Calibration: not reported
- Discrimination: not reported

Creation of risk groups?

- yes, 3 risk groups, (cut-points at (1-3), (4-7), (>7))
- at 5 years
- percentage of group without event at time point (low risk: 97%, intermediate risk: 89%, high risk: 53%)

PROBAST: Applicability	<u>Domain 1: Participant selection</u> • low <u>Domain 2: Predictors</u> • low <u>Domain 3: Outcome</u> • low
Notes	<u>Funding & conflict of interest</u> • "Ministero della Salute (Ricerca Finalizzata I.R.C.C.S. and 'Alleanza Contro il Cancro'), Rome; Associazione Italiana contro le Leucemie, linfomi e mielomi (A.I.L.), Venezia Section, Pramaggiore Group; Ricerca Scientifica Applicata, Regione Friuli Venezia Giulia, Trieste ('Linfonet' project); Associazione Italiana per la Ricerca sul Cancro (Investigator Grant IG-8701), Milan, Italy; Helmut Horten Foundation, San Salvatore Foundation, Fondazione per la Ricerca e la Cura sui Linfomi (Lugano, Switzerland); Swiss Cancer League (Krebsliga Schweiz) grant KLS-01690-03-2005." • "The authors declare no competing financial interests." <u>Other comments</u> • none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	"Clinical and biological data of 1480 untreated CLL patients were retrospectively collected from 8 centres. The series included all cases observed at each centre over a given time period, ..."
Domain 2: Predictors	Yes	Measurements were taken in 8 different centres. However, the authors standardised the predictors so this should not be of influence.
Domain 3: Outcome	Yes	Objective standard outcome: no definition of outcome and assessment - we assumed that definition and assessment of the objective outcome OS was similar and therefore did not rate as high risk.
Domain 4: Analysis	No	Exclusion of individuals with missing values: "Some CLL cases with missing values were excluded from multivariate models. The death rate in this group was higher. The lack of inclusion of a number of individuals with possible worse prognosis could partly justify the observed bias in nomogram prediction."; expected OS estimated, but only point score: no discrimination calculated, because estimation based on nomogram not possible (quality of graph)
Overall judgement	No	

MDACC 2007 V - Gentile 2014 (Italian cohort)

Study characteristics

General information	<u>Model and type of study</u> - Gentile 2014 (Italian multicentre cohort) - Validation study <u>Secondary citations</u> - Gentile 2016 <u>Language of publication</u> - English <u>Study design</u> - retrospective cohort <u>Follow-up time</u> - median 68.4 months, range not reported
Participants	<u>Number of included participants in the cohort</u> 1502 individuals <u>Setting</u> - Italy, multicentre study <u>Recruitment period</u> 1983 - 2013 <u>Age (in years)</u> - median 67 years (range: not reported) <u>Sex</u> 55.7% male <u>Stages of disease</u> - Rai 0: 56.5%; Rai I: 21.4%; Rai II: 15%; Rai III: 2.6%; Rai IV: 4.5% - Binet A: 82.9%; Binet B: 11.6%; Binet C: 5.5% <u>Treatment</u> 337 received chemotherapy 142 received chemoimmunotherapy <u>Inclusion criteria</u> - The CLL databases of four Italian centres, which included all patients diagnosed with CLL since 1983, were utilised for research purposes. <u>Exclusion criteria</u> - Missing data, 12m
Predictors	<u>List of predictors used for validation (and changes between original predictors and predictors in validation study)</u>

MDACC 2007 V - Gentile 2014 (Italian cohort) (Continued)

- Age (continuous: no transformation; in years); ALC (absolute lymphocyte count) (continuous: no transformation; $\times 10^9/L$); serum $\beta 2$ -microglobulin (continuous: no transformation); nodal groups (binary); Rai stage (categorical; binary: stage III or IV vs. 0-II); gender (binary)

Timing of predictor measurement

- at study entry

Outcome(s)

Primary outcome in study

- overall survival: for OS, the time interval was measured from the day of CLL diagnosis until death from all causes or last follow-up.

Additional outcome(s)

- time-to-first-treatment: for TFS, it was from the day of CLL diagnosis until therapy start or last follow-up.

Outcome in model development

- overall survival

Missing data

Participants with any missing data?

- no

If yes, how was missing data handled?

- not applicable

Analysis

Number of participants & number of events (specific time points where reported)

- overall survival: 1502 participants, 277 events
- time-to-first-treatment: 1502 participants, 479 events

Which model was used?

- Authors used both nomogram and the index score, and also compared with previous validations.

Was the model updated?

- no

Performance measures reported?

- Calibration: not reported
- Discrimination: reported (c-statistic)

Creation of risk groups?

- yes, 3 risk groups, (cut-points at (1-3);(4-7);(above 7))
- at 5 years
- percentage of group without event at time point (low risk: 89%, intermediate risk: 82%, high risk: 42%)

PROBAST: Applicability

Domain 1: Participant selection

- unclear - very long period of retrospective enrolment: patients were probably treated differently. Inclusion of cMBL patients previously classified as CLL - sensitivity analysis showed no difference in discrimination.

Domain 2: Predictors

- low

MDACC 2007 V - Gentile 2014 (Italian cohort) (Continued)

Domain 3: Outcome

- low

Notes

Funding & conflict of interest

- Funding not reported
- "The authors indicated no potential conflicts of interest."

Other comments

- 4 of the 5 Italian centres overlapped with the cohort in the secondary citation. This paper was chosen as primary because more information was provided.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Retrospective cohort: "... which included all patients diagnosed with CLL since 1983, were utilized for research purposes."
Domain 2: Predictors	Yes	Routine clinical data used: although the recruitment period was very broad, standard clinical data (gender, age, Rai stage, ALC and beta-2 microglobulin) have been included for modelling.
Domain 3: Outcome	Yes	Objective standard outcome: no definition of outcome and assessment - we assumed that definition and assessment of the objective outcome OS was similar and therefore did not rate as high risk.
Domain 4: Analysis	No	Individuals with missing values were excluded. Although background characteristics were similar, we rated this as high risk because nearly half of all patients were excluded from analysis; measure for calibration not reported, however, reporting of survival per risk group
Overall judgement	No	

MDACC 2007 V - Gentile 2016 (Mayo cohort)

Study characteristics

General information

Model and type of study

- Gentile 2016 (Mayo clinic cohort)
- Validation study

Secondary citations

- Gentile 2016; Shanafelt 2009

Language of publication

- English

Study design

- prospective cohort

Follow-up time

MDACC 2007 V - Gentile 2016 (Mayo cohort) (Continued)

- median 86.4 months, range not reported

Participants	<u>Number of included participants in the cohort</u> • 506 participants <u>Setting</u> • USA, single-centre study <u>Recruitment period</u> • 2001 - 2008 (correspondence with authors) <u>Age (in years)</u> • median 62.5 years (range: 36 - 89 years) <u>Sex</u> • 68.6% male <u>Stages of disease</u> • Rai 0: 62.5%; Rai I: 28.9%; Rai II: 5.1%; Rai III: 1.2%; Rai IV: 2.4% • Binet A: not reported <u>Treatment</u> • 213 received treatment • unclear which treatment <u>Inclusion criteria</u> • not reported <u>Exclusion criteria</u> • not reported
Predictors	<u>List of predictors used for validation (and changes between original predictors and predictors in validation study)</u> • Age (continuous: no transformation; in years); ALC (absolute lymphocyte count) (continuous: no transformation; $\times 10^9/L$); serum $\beta 2$-microglobulin (continuous: no transformation); nodal groups (binary); Rai stage (categorical; binary: stage III or IV vs. 0-II); gender (binary) <u>Timing of predictor measurement</u> • at diagnosis
Outcome(s)	<u>Primary outcome in study</u> • overall survival: for OS, the time interval was measured from the day of CLL diagnosis until death from all causes or last follow-up. <u>Additional outcome(s)</u> • time-to-first-treatment: for TTFT from the day of CLL diagnosis until the start of therapy or last follow-up <u>Outcome in model development</u> • overall survival

MDACC 2007 V - Gentile 2016 (Mayo cohort) (Continued)

Missing data	<u>Participants with any missing data?</u> - no <u>If yes, how was missing data handled?</u> - not applicable
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> - overall survival: 506 participants, 114 events - time-to-first-treatment: 506 participants, 213 events <u>Which model was used?</u> - simplified model <u>Was the model updated?</u> - no <u>Performance measures reported?</u> - Calibration: reported - Discrimination: reported (c-statistic) <u>Creation of risk groups?</u> - yes, 3 risk groups, (cut-points at (1-3);(4-7);(above 7)) - at 5 years - percentage of group without event at time point (low risk: 98.3%, intermediate risk: 84.4%, high risk: 15.6%)
PROBAST: Applicability	<u>Domain 1: Participant selection</u> - unclear - no clear eligibility criteria <u>Domain 2: Predictors</u> - low <u>Domain 3: Outcome</u> - low
Notes	<u>Funding & conflict of interest</u> - Funding not reported "S.A.P. received funding from Pharmacyclics. All other authors have no conflict of interest to disclose." <u>Other comments</u> - The corresponding author provided additional information to describe the cohort (recruitment period, overlap of cohort with other publications)

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	No	Missing values as exclusion reason: 'The validation data set consisted of a consecutive series of 506 newly-diagnosed CLL patients, prospectively followed

MDACC 2007 V - Gentile 2016 (Mayo cohort) (Continued)

		at the Mayo Clinic, with baseline data available for all variables considered for both CLL-IPI and the MDACC score.'
Domain 2: Predictors	Yes	Routine clinical data used, prospective follow-up, relatively recent cohort (2001-2008)
Domain 3: Outcome	Yes	Objective standard outcome, clear definition and assessment: 'death, which were abstracted from clinical records at the time of inclusion and updated on an ongoing basis.'
Domain 4: Analysis	Unclear	Individuals with missing data were excluded from the study, rated in domain 1. Calibration not reported; survival per risk groups was available.
Overall judgement	No	

MDACC 2007 V - González Rodríguez (Cabueñes)

Study characteristics

General information	<u>Model and type of study</u> - Gonzalez Rodriguez 2009 (Cabuenes cohort) - Validation study <u>Secondary citations</u> - not applicable <u>Language of publication</u> - Spanish <u>Study design</u> - retrospective cohort <u>Follow-up time</u> - not reported
Participants	<u>Number of included participants in the cohort</u> 265 participants <u>Setting</u> - Spain, single-centre study <u>Recruitment period</u> 1997 - 2007 <u>Age (in years)</u> - median 71.7 years (range: 42 - 94 years) <u>Sex</u> - not reported

MDACC 2007 V - González Rodríguez (Cabueñes) (Continued)

Stages of disease

- reported as Rai modified stages; low risk: 61.2%; intermediate: 29.3%; high risk: 9.5%
- Binet A: 76.8%; Binet B: 14.8%; Binet C: 8.4%

Treatment

- 67 received treatment
- unclear which treatment

Inclusion criteria

- Only cases with an increase of more than $5 \times 10^9/l$ lymphocytes with morphology and immunophenotype of B-CLL and with Royal Marsden Scoring System greater than 3, and cases of CLL confirmed by lymph node biopsy have been included. The inclusion criteria were based on the initial data at the time of diagnosis.

Exclusion criteria

- not reported

Predictors

List of predictors used for validation (and changes between original predictors and predictors in validation study)

- Age (continuous: no transformation; in years); ALC (absolute lymphocyte count) (continuous: no transformation; $\times 10^9/L$); serum $\beta 2$ -microglobulin (continuous: no transformation); nodal groups (binary); Rai stage (categorical; binary: stage III or IV vs. 0-II); gender (binary)
- no changes

Timing of predictor measurement

- at diagnosis

Outcome(s)

Primary outcome in study

- overall survival: to estimate overall survival, the diagnosis date was considered and the date of death by any cause or the final date of the study

Additional outcome(s)

- time-to-first-treatment: treatment-free survival was considered as the time elapsed from diagnosis until the date of treatment initiation.

Outcome in model development

- overall survival

Missing data

Participants with any missing data?

- yes (8 participants)

If yes, how was missing data handled?

- complete-case analysis

Analysis

Number of participants & number of events (specific time points where reported)

- overall survival: 257 participants, 76 events
- time-to-first-treatment: 257 participants, 67 events

Which model was used?

- index score (simplified model)

MDACC 2007 V - González Rodríguez (Cabueñes) (Continued)

Was the model updated?

- no

Performance measures reported?

- Calibration: not reported
- Discrimination: reported (AUC)

Creation of risk groups?

- yes, 3 risk groups, (cut-points at (1-3);(4-7);(above 7))
- at 5 years
- percentage of group without event at time point (low risk: 87%, intermediate risk: 75%, high risk: 29%)

PROBAST: Applicability

Domain 1: Participant selection

- low

Domain 2: Predictors

- low

Domain 3: Outcome

- low

Notes

Funding & conflict of interest

- This work was partially funded through research project FIS06/0841 from the Health Research Fund.
- Conflict of interest not reported

Other comments

- original study translated by Leonardo Perales-Guerrero (Medical student, Universidad de Guadalajara, Mexico)
- data extraction from both the original and translated study

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Retrospective cohort, consecutive sampling: 'Un total de 265 pacientes se diagnosticaron de LLCB en el periodo de 10 años.'
Domain 2: Predictors	Yes	Routine clinical data assessed at diagnosis
Domain 3: Outcome	Yes	Objective standard outcome: no definition of outcome and assessment - we assumed that definition and assessment of the objective outcome OS was similar and therefore did not rate as high risk.
Domain 4: Analysis	No	Low number of events; individuals with missing values excluded from modelling, however this number was small; calibration not reported
Overall judgement	No	

MDACC 2007 V - Molica 2010 (GIMEMA cohort)

Study characteristics

General information

Model and type of study

- Molica 2010 (GIMEMA cohort)
- Validation study

Secondary citations

- not applicable

Language of publication

- English

Study design

- retrospective cohort

Follow-up time

- median 41.5 months, range not reported
-

Participants

Number of included participants in the cohort

- 1158 participants
- characteristics only reported for 310 participants with complete data

Setting

- Italy, multicentre study

Recruitment period

- 1991 - 2000

Age (in years)

- median 64 years (range: not reported)

Sex

- 56.1% male

Stages of disease

- Rai 0: 67%; Rai I-III: 32.9%; Rai III/IV: 0%
- Binet A: 100%

Treatment

- 51 received therapy
- unclear which treatment

Inclusion criteria

- previously untreated CLL participants in Binet stage A whose diagnosis was immunologically confirmed (CD5+/Smlg weak) and who were observed at different GIMEMA primary haematology centres during the period 1991–2000

Exclusion criteria

- not reported
-

MDACC 2007 V - Molica 2010 (GIMEMA cohort) (Continued)

Predictors	<u>List of predictors used for validation (and changes between original predictors and predictors in validation study)</u> - Age (continuous: no transformation; in years); ALC (absolute lymphocyte count) (continuous: no transformation; $\times 10^9/L$); serum $\beta 2$-microglobulin (continuous: no transformation); nodal groups (binary); Rai stage (categorical; binary: stage III or IV vs. 0-II); sex (binary) - cut-o& for Rai was adapted in the score <u>Timing of predictor measurement</u> - at diagnosis
Outcome(s)	<u>Primary outcome in study</u> - time-to-first-treatment: active, therapy-requiring disease was defined by very detailed criteria. <u>Additional outcome(s)</u> - overall survival: definition not reported <u>Outcome in model development</u> - overall survival
Missing data	<u>Participants with any missing data?</u> - no <u>If yes, how was missing data handled?</u> - not applicable
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> - overall survival: 310 participants, number of events not reported - time-to-first-treatment: 310 participants, 51 events <u>Which model was used?</u> - Authors used both nomogram and index score, and also compared with previous validations. <u>Was the model updated?</u> - no <u>Performance measures reported?</u> - Calibration: not reported - Discrimination: not reported <u>Creation of risk groups?</u> - yes, 3 risk groups, (cut-points at (1-3);(4-7);(above 7)) - at 5 years - percentage of group without event at time point (low risk: 99.3%, intermediate risk: 88.2%, high risk: no individuals in this category)
PROBAST: Applicability	<u>Domain 1: Participant selection</u> low <u>Domain 2: Predictors</u>

MDACC 2007 V - Molica 2010 (GIMEMA cohort) (Continued)

- high - the authors changed the cut-points for the predictor 'Rai stage'. Instead of a binary variable stage 0-II vs II-IV, Molica 2010 used Rai 0 vs. Rai I-II

Domain 3: Outcome

- High - outcome did not match primary outcome of the model. Survival per risk group reported

Notes

Funding & conflict of interest

- Funding not reported
- "The authors reported no potential conflicts of interest."

Other comments

- none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Retrospective cohort, inclusion of complete database: "The GIMEMA CLL database includes information on previously untreated CLL patients in Binet stage A whose diagnosis was immunologically confirmed".
Domain 2: Predictors	Yes	Although individuals came from 25 different centres, predictors were routine clinical data.
Domain 3: Outcome	Unclear	The median observation time was below 5 years (41.5 months); outcome time-to-first-treatment was not the primary outcome of the validated model.
Domain 4: Analysis	No	Individuals with missing values were excluded from the analysis ("The outcome of the 593 patients with complete data was different from the outcome of the 545 patients with incomplete data"); no performance measures reported
Overall judgement	No	

MDACC 2007 V - Molica 2015 (O-CLL1-GISL)

Study characteristics

General information

Model and type of study

- Molica 2015 (O-CLL1-GISL)
- Validation study

Secondary citations

- Gentile 2016; Molica 2010; Molica 2017 (leukemia & lymphoma); Molica 2017 (American Journal of Hematology)

Language of publication

- English

Study design

- prospective cohort

MDACC 2007 V - Molica 2015 (O-CLL1-GISL) (Continued)

Follow-up time

- median 42 months, range not reported

Participants

Number of included participants in the cohort

- 337 participants

Setting

- Italy, multicentre study

Recruitment period

- 2007 - not reported

Age (in years)

- median 61 years (range: 33 - 70 years)

Sex

- 57.1% male

Stages of disease

- Rai 0: 77.5%; Rai I-II: 22.5%
- Binet A: 100%; Binet B/C: not applicable

Treatment

- 91 received treatment
- unclear which treatment

Inclusion criteria

- Established diagnosis of B-CLL by NCI criteria, performed by local haematologist (diagnosis will be confirmed by the biological review committee according to flow cytometry analysis (positive clusters of differentiation antigen 5 (CD5), 19 (CD19), 23 (CD23)). Age above 18 years and below 70 years. Eastern Cooperative Oncology Group (ECOG) below 2. Binet stage A. Diagnosis performed within 12 months before inclusion in the study. Patients who do not necessitate therapy by NCI guidelines (watch-and-wait policy). Shipment of peripheral blood sample to centralised laboratory for biological assessment. Clinical data including baseline information on disease localisation and laboratory parameters at staging and assurance of follow-up updating for at least 3 years are requested. Written informed consent.

Exclusion criteria

- Patients with CLL whose diagnosis exceeded 12 months before registration. Patients with leukaemic phase of lymphoproliferative disorders of B origin CD5- and/or CD23- according to flow cytometry analysis. Clinical Binet stage B or C. Patients who necessitated therapy according to NCI guidelines (no watch-and-wait policy). Age above 70 years. Without a written informed consent.

Predictors

List of predictors used for validation (and changes between original predictors and predictors in validation study)

- Age (continuous: no transformation; in years); ALC (absolute lymphocyte count) (continuous: no transformation; $\times 10^9/L$); serum β_2 -microglobulin (continuous: no transformation); nodal groups (binary); Rai stage (categorical; binary: stage III or IV vs. 0-II); gender (binary)
- no changes

Timing of predictor measurement

- at study entry, within 12 months of diagnosis

MDACC 2007 V - Molica 2015 (O-CLL1-GISL) (Continued)

Outcome(s)	<u>Primary outcome in study</u> - time-to-first-treatment: the primary endpoint, TTFT, was defined as the interval between the date of database registration and the date of first CLL treatment (individuals underwent sequential monitoring, and the frequency of follow-up visits was individualised according to patient risk; this ranged from 3 to 6 months (median 6 months). All physicians who registered patients in this observational database stated that they had used the NCI-WG guidelines as a reference criterion for starting therapy. In particular, the absolute lymphocyte count was not used as the sole indicator for treatment. Active disease, requiring therapy, was defined when at least one of the criteria set out in the NCI-sponsored Working Group guidelines was satisfied). <u>Additional outcome(s)</u> - not applicable <u>Outcome in model development</u> - overall survival
Missing data	<u>Participants with any missing data?</u> - no <u>If yes, how was missing data handled?</u> - not applicable
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> - time-to-first-treatment: 337 participants, 91 events <u>Which model was used?</u> - Authors used nomogram/formula <u>Was the model updated?</u> - no <u>Performance measures reported?</u> - Calibration: not reported - Discrimination: reported (c-statistic) <u>Creation of risk groups?</u> - no
PROBAST: Applicability	Not applicable, outcome did not match primary outcome of the model.
Notes	<u>Funding & conflict of interest</u> "This work was supported by funding from Associazione Italiana per la Ricerca sul Cancro (AIRC 5xmille grant 9980, IG10492 to MF and FM and IG10136 to AN). We thank AIL Cosenza-Fondazione Amelia Scorza' onlus, Cosenza, Italy, and Brigida Gulino for precious secretarial assistance." (Gentile 2016) - Conflict of interest not reported <u>Other comments</u> - Cohort identical in several publications

MDACC 2007 V - Muñoz-Novas 2018 (Spanish cohort)

Study characteristics

General information	<u>Model and type of study</u> - Munoz-Novas 2018 (Spanish cohort) - Validation study <u>Secondary citations</u> - not applicable <u>Language of publication</u> - English <u>Study design</u> - retrospective cohort <u>Follow-up time</u> - median 46 months, range not reported
Participants	<u>Number of included participants in the cohort</u> 696 participants <u>Setting</u> - Spain, multicentre study <u>Recruitment period</u> 1989 - 2013 <u>Age (in years)</u> - median 67 years (range: 25 - 90 years) <u>Sex</u> 64.2 % male <u>Stages of disease</u> - Rai 0-II: 93.8%; Rai III-IV: 6.2% - Binet A/B/C: not reported <u>Treatment</u> 186 received treatment - unclear which treatment <u>Inclusion criteria</u> - all participants diagnosed with CLL <u>Exclusion criteria</u> - not reported
Predictors	<u>List of predictors used for validation (and changes between original predictors and predictors in validation study)</u>

MDACC 2007 V - Muñoz-Novas 2018 (Spanish cohort) (Continued)

- Age (continuous: no transformation; in years); ALC (absolute lymphocyte count) (continuous: no transformation; $\times 10^9/L$); serum $\beta 2$ -microglobulin (continuous: no transformation); nodal groups (binary); Rai stage (categorical; binary: stage III or IV vs. 0-II); gender (binary)

Timing of predictor measurement

- at diagnosis, data from medical records

Outcome(s)

Primary outcome in study

- overall survival: overall survival (OS) was calculated from the time of diagnosis to death or last follow-up.

Additional outcome(s)

- time-to-first-treatment: time-to-first-therapy (TTFT) from the date of diagnosis to first treatment or last follow-up

Outcome in model development

- overall survival

Missing data

Participants with any missing data?

- yes (213 participants)

If yes, how was missing data handled?

- complete-case analysis

Analysis

Number of participants & number of events (specific time points where reported)

- overall survival: 483 participants, 92 events
- time-to-first-treatment: 483 participants, 186 events

Which model was used?

- simplified model

Was the model updated?

- no

Performance measures reported?

- Calibration: not reported
- Discrimination: reported (c-statistic)

Creation of risk groups?

- yes, 3 risk groups, (cut-points at (1-3);(4-7);(above 7))
- at 5 years
- percentage of group without event at time point (low risk: 98.4%, intermediate risk: 79.6%, high risk: 51.1%)

PROBAST: Applicability

Domain 1: Participant selection

- low

Domain 2: Predictors

- low

Domain 3: Outcome

MDACC 2007 V - Muñoz-Novas 2018 (Spanish cohort) (Continued)

- low

Notes	<u>Funding & conflict of interest</u> • Funding not reported • "The authors declare that there are no conflicts of interest regarding the publication of this article." <u>Other comments</u> • none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Retrospective cohort: "A total of 696 unselected CLL patients newly diagnosed and previously untreated from different institutions of the central region of Spain were included in this study." Patients with missing values were later excluded at analysis stage, rated in domain 4.
Domain 2: Predictors	Yes	Routine clinical data assessed at diagnosis was used (from medical records).
Domain 3: Outcome	Yes	Objective standard outcome: no definition of outcome and assessment - we assumed that definition and assessment of the objective outcome OS was similar and therefore did not rate as high risk.
Domain 4: Analysis	No	Low number of events; exclusion of patients with missing values from analysis; calibration not reported
Overall judgement	No	

MDACC 2007 V - Pflug 2014 (3 RCTs)

Study characteristics

General information	<u>Model and type of study</u> • Pflug 2014 (German CLL study group cohort) • Validation study <u>Secondary citations</u> • not applicable <u>Language of publication</u> • English <u>Study design</u> • combined data (e.g. IPD or combination of cohorts) <u>Follow-up time</u> • median 63.4 months, range not reported
Participants	<u>Number of included participants in the cohort</u> • 2007 participants

MDACC 2007 V - Pflug 2014 (3 RCTs) (Continued)

Setting

- several, multicentre study (Australia, Austria, Belgium, Czech Republic, Denmark, France, Germany, Israel, Italy, New Zealand, Spain)

Recruitment period

- 1997 - 2006

Age (in years)

- median 60 years (range: 30 - 81 years)

Sex

- 68.4% male

Stages of disease

- Rai 0: 20.8%; Rai I: 18.4%; Rai II: 37.9%; Rai III: 8.6%; Rai IV: 14.3%
- Binet A: 42.6%; Binet B: 37.7%; Binet C: 19.8%

Treatment

- participants were treated with:
- early fludarabine (F)
- F plus cyclophosphamide (FC)
- FC plus rituximab (FCR)

Inclusion criteria

- CLL1 trial: 18-75 years, confirmed diagnosis of B-cell chronic lymphocytic leukaemia, Rai stage 0-II (Binet stage A), life expectancy above 6 months, no autoimmune haemolytic anaemia, no thrombocytopenia, no severe organ dysfunction, no other prior or concurrent malignancy, no prior or other concurrent chemotherapy
- CLL4 trial: < 65 years with untreated CLL in advanced stage (all Binet stage C patients; Binet stage B with symptoms, which require therapy; Binet stage A with severe B-symptoms)
- CLL8 trial: ≥ 18 years, diagnosed B-cell chronic lymphocytic leukaemia (CLL) defined by the National Cancer Institute (NCI) Working Group criteria, Binet stage C disease OR Binet stage B disease AND 1 of other predefined signs or symptoms, ECOG performance status 0-1, Cumulative Illness Rating Scale (CIRS) score above 6, life expectancy above 6 months, no active infection, no pregnancy or nursing, no concurrent or previous treatment of CLL by chemotherapy, radiotherapy, or immunotherapy, and more

Exclusion criteria

- CLL4 trial: age < 18 and > 65 years were excluded as well as patients with any previous treatment of CLL, life expectancy less than six months and an Eastern Cooperative Oncology Group performance status of more than two. Patients were also excluded if they had severe organ dysfunction, concomitant or previous other neoplasms or an autoimmune haemolytic anaemia or thrombocytopenia.

Predictors

List of predictors used for validation (and changes between original predictors and predictors in validation study)

- Age (continuous: no transformation; in years); ALC (absolute lymphocyte count) (continuous: no transformation; $\times 10^9/L$); serum $\beta 2$ -microglobulin (continuous: no transformation); nodal groups (binary); Rai stage (categorical; binary: stage III or IV vs. 0-II); gender (binary)
- no changes

Timing of predictor measurement

- at study entry (< 1 year: 63.3%, > 1 year: 36.7%)

MDACC 2007 V - Pflug 2014 (3 RCTs) (Continued)

Outcome(s)	<u>Primary outcome in study</u> - overall survival: OS was defined as the time between registration/randomisation and death. <u>Additional outcome(s)</u> - progression-free survival: PFS was calculated from registration/randomisation to start of the first CLL treatment or disease progression or death. <u>Outcome in model development</u> - overall survival
Missing data	<u>Participants with any missing data?</u> - yes (863 participants) <u>If yes, how was missing data handled?</u> - complete-case analysis
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> - overall survival: 1144 participants, number of events not reported in publication - 180 at 5 years <u>Which model was used?</u> - c-statistic reported for the nomogram <u>Was the model updated?</u> - no <u>Performance measures reported?</u> - Calibration: not reported - Discrimination: reported (c-statistic) <u>Creation of risk groups?</u> - no
PROBAST: Applicability	<u>Domain 1: Participant selection</u> - unclear - restricted inclusion and exclusion criteria based on RCTs <u>Domain 2: Predictors</u> - low <u>Domain 3: Outcome</u> - low
Notes	<u>Funding & conflict of interest</u> "This manuscript was written on behalf of the German CLL Study Group. Studies CLL1, CLL4, and CLL8 were planned and conducted as investigator-initiated trials by the German CLL Study Group and were supported by research grants from German Cancer Aid, Medac Schering Onkologie, and F. Hoffmann-La Roche. T.D.S. is a clinical scholar of the Leukemia Lymphoma Society." "N.P. received Travel Grants from Roche. T.D.S. received research grants from Genentech, Celgene, Glaxo-Smith-Kline, Cephalon, Hospira, and Polyphenon E International. B.E. is a consultant and/or holds an advisory role for Celgene and Pharmacyclics and has received honoraria and research funding from Roche and Mundipharma. S.S. is a consultant and/or holds an advisory role for Roche and

MDACC 2007 V - Pflug 2014 (3 RCTs) (Continued)

Mundipharma, and received honoraria and research funding from both. H.D. received research grants from Roche. U.J. received honoraria and research funding from Roche. M.H. is a consultant and/or holds an advisory role and received research funding from Roche. J.B., M.A.B., T.E., K.B., G.M., K.G.R., M.J.E., G.H., R.B., A.-M.F., C.-M.W., K.F., and N.E.K. declare no competing financial interests."

Other comments

- The authors provided additional data upon request (classification tables).

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Overlap of individuals between trials prevented; only concern could be comparability of the three trials.
Domain 2: Predictors	Yes	Prospective collection of predictors within clinical trials
Domain 3: Outcome	Yes	Objective standard outcome: follow-up in frame of each RCT
Domain 4: Analysis	No	Number of events sufficient; individuals with missing data were excluded from model development. It was not clear if this was the case for the validation of the MDACC2007 model; calibration not reported
Overall judgement	No	

MDACC 2007 V - Rani 2018 (Indian cohort)

Study characteristics

General information	<u>Model and type of study</u> • Rani 2018 (Indian cohort) • Validation study <u>Secondary citations</u> • not applicable <u>Language of publication</u> • English <u>Study design</u> • unclear <u>Follow-up time</u> • median 40.5 months (range not reported)
Participants	<u>Number of included participants in the cohort</u> • 198 participants <u>Setting</u> • India, unclear if single or multicentre study <u>Recruitment period</u>

MDACC 2007 V - Rani 2018 (Indian cohort) (Continued)

- not reported

Age (in years)

- median 60 years (range: not reported)

Sex

- 77% male

Stages of disease

- Rai 0: 14.6%; Rai I: 21.2%; Rai II: 34.9%; Rai III: 14.1%; Rai IV: 15.2%
- Binet A/B/C: not reported

Treatment

- 62 received chlorambucil-based therapy
- 56 received rituximab-based therapy
- 20 received other therapies

Inclusion criteria

- treatment-naïve CLL patients according to Hallek 2008

Exclusion criteria

- not reported

Predictors	<u>List of predictors used for validation (and changes between original predictors and predictors in validation study)</u> • Age (continuous: no transformation; in years); ALC (absolute lymphocyte count) (continuous: no transformation; $\times 10^9/L$); serum $\beta 2$-microglobulin (continuous: no transformation); nodal groups (binary); Rai stage (categorical; binary: stage III or IV vs. 0-II); gender (binary) <u>Timing of predictor measurement</u> • not reported
Outcome(s)	<u>Primary outcome in study</u> • overall survival: OS was defined as the time from the date of diagnosis to date of death or date of last follow-up. <u>Additional outcome(s)</u> • not applicable for this model (TTFT was assessed in this cohort for other prognostic models) <u>Outcome in model development</u> • overall survival
Missing data	<u>Participants with any missing data?</u> • no <u>If yes, how was missing data handled?</u> • not applicable
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> • overall survival: 198 participants, 86 events

MDACC 2007 V - Rani 2018 (Indian cohort) (Continued)

Which model was used?

- simplified model

Was the model updated?

- no

Performance measures reported?

- Calibration: not reported
- Discrimination: reported (c-statistic)

Creation of risk groups?

- yes, 3 risk groups
- survival probabilities not reported

PROBAST: Applicability

Domain 1: Participant selection

- unclear - recruitment period and eligibility criteria unclear

Domain 2: Predictors

- unclear - no information on timing of predictor because the study design was unclear

Domain 3: Outcome

- low

Notes

Funding & conflict of interest

- "The financial support was provided by the Department of Biotechnology (BT/PR11106/GBD/27/145/2008, BT/PR15438/MED/30/606/2011 and T/PR8680/AGR/36/754/2013), Ministry of Science and Technology, GOI, and All India Institute of Medical Sciences, New Delhi (8-60/A060/ 2011/RS) to RG for carrying out this work."
- "The authors declare that they have no conflict of interest."

Other comments

- none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	No	No clear eligibility criteria; possibly missing values as exclusion reason; no recruitment period
Domain 2: Predictors	Yes	Study design unclear, "... were obtained from medical records of the patients". Routine clinical predictors were used in this model, therefore we rated this study as low.
Domain 3: Outcome	Unclear	Short follow-up time
Domain 4: Analysis	No	Low number of events; calibration not reported; survival per risk groups graphically presented
Overall judgement	No	

MDACC 2007 V - Trajkova 2013 (Macedonia)

Study characteristics

General information	<u>Model and type of study</u> - Trajkova 2013 (Macedonian cohort) - Validation study <u>Secondary citations</u> - not applicable <u>Language of publication</u> - English <u>Study design</u> - unclear <u>Follow-up time</u> - not reported
Participants	<u>Number of included participants in the cohort</u> 100 participants <u>Setting</u> - Macedonia, single-centre study <u>Recruitment period</u> 2011 - 2013 <u>Age (in years)</u> - median 64.8 years (range: 47 - 78 years) <u>Sex</u> 63% male <u>Stages of disease</u> - Rai 0: 54%; Rai I: 5%; Rai II: 32%; Rai III: 5%; Rai IV: 4% - Binet A/B/C: not reported <u>Treatment</u> 41 received treatment - unclear which therapy <u>Inclusion criteria</u> - Our study incorporated 100 consecutive treatment-naïve CLL patients with IgHV mutational status. <u>Exclusion criteria</u> - not reported
Predictors	<u>List of predictors used for validation (and changes between original predictors and predictors in validation study)</u>

MDACC 2007 V - Trajkova 2013 (Macedonia) (Continued)

- Age (continuous: no transformation; in years); ALC (absolute lymphocyte count) (continuous: no transformation; $\times 10^9/L$); serum $\beta 2$ -microglobulin (continuous: no transformation); nodal groups (binary); Rai stage (categorical; binary: stage III or IV vs. 0-II); gender (binary)

Timing of predictor measurement

- at presentation at University Haematology Clinic

Outcome(s)	<u>Primary outcome in study</u> • time-to-first-treatment: time-to-event end point was defined as the time from first visit to the University Haematology Clinic to first CLL treatment. There was no restriction of time from diagnosis to presentation at the University Haematology Clinic. <u>Additional outcome(s)</u> • not applicable <u>Outcome in model development</u> • overall survival
Missing data	<u>Participants with any missing data?</u> • no <u>If yes, how was missing data handled?</u> • not applicable
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> • time-to-first-treatment: 100 participants, 41 events <u>Which model was used?</u> • unclear <u>Was the model updated?</u> • unclear <u>Performance measures reported?</u> • Calibration: not reported • Discrimination: not reported <u>Creation of risk groups?</u> • yes • No reporting of observed survival per subgroup
PROBAST: Applicability	<u>Domain 1: Participant selection</u> • low <u>Domain 2: Predictors</u> • low <u>Domain 3: Outcome</u> • low
Notes	<u>Funding & conflict of interest</u>

MDACC 2007 V - Trajkova 2013 (Macedonia) (Continued)

- not reported

Other comments

- "estimated and projected" survival without any comparison of observed survival provided; we could not use this information for our review and therefore, we did not continue with data extraction.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	No	"100 consecutive treatment naïve CLL patients"
Domain 2: Predictors	Yes	Routine clinical data
Domain 3: Outcome	Unclear	No observation time reported; outcome assessment not described
Domain 4: Analysis	No	Low number of participants and events; not really validating MDACC 2007 - reporting of projected survival only and additional univariable analysis
Overall judgement	No	

MDACC 2011 D - Wierda 2011 (MDACC)

Study characteristics

General information	<u>Model and type of study</u> • Wierda 2011 (development (MDACC 2004-2009)) • Development study <u>Secondary citations</u> • not applicable <u>Language of publication</u> • English <u>Study design</u> • retrospective cohort <u>Follow-up time</u> • median 26 months (range 3 - 73 months)
Participants	<u>Number of included participants in the cohort</u> • 930 participants <u>Setting</u> • US (Texas), single-centre study <u>Recruitment period</u> • 2004 - 2009

MDACC 2011 D - Wierda 2011 (MDACC) (Continued)

Age (in years)

- median 59 years (range: 30 - 89 years)

Sex

- 61% male

Stages of disease

- Rai 0: 36%; Rai I: 51%; Rai II: 8%; Rai III-IV: 5%

Treatment

- 232 received treatment
- unclear which treatment

Inclusion criteria

- no previous treatment; not recommended for first-line treatment at initial visit

Exclusion criteria

- not reported

Predictors

Number of candidate predictors

- univariable analysis: more than 25
- multivariable analysis: unclear

List of predictors in final model (including cut-points for dichotomised factors)

- IgHV status (continuous: categorical/binary transformation); diameter of largest palpated cervical LN, cm (continuous: no transformation); FISH category (categorical; del(11q) or del(17p) vs. none); number of involved lymph node sites (continuous: categorical/binary transformation); lactate dehydrogenase IU/L/100 (continuous: no transformation); IgHV status*LDH (interaction term)

Timing of predictor measurement

- at study entry (median time from diagnosis to presentation to MDACC was 3.4 months, range 0 to 428 months)

Outcome(s)

Primary outcome in study

- time-to-first-treatment: TTFT was defined as the time interval between the date of presentation to MDACC and date of first CLL treatment.

Additional outcome(s)

- not applicable

Missing data

Participants with any missing data?

- yes (243 participants)

If yes, how was missing data handled?

- complete-case analysis

Analysis

Number of participants & number of events (specific time points where reported)

- time-to-first-treatment: 687 participants, 193 events

Predictor selection method

MDACC 2011 D - Wierda 2011 (MDACC) (Continued)

- univariable analysis, during multivariable modelling: stepwise regression mix, (P value or CI)

Statistical method

- Cox proportional hazard model

Simplification of model?

- yes

Performance measures reported?

- Calibration: not reported
- Discrimination: not reported

Creation of risk groups?

- no

PROBAST: Applicability

Domain 1: Participant selection

- unclear - individuals with missing values were excluded. We were unsure if the sample was representative for all individuals with CLL

Domain 2: Predictors

- low

Domain 3: Outcome

- low

Notes

Funding & conflict of interest

- "Supported in part by the Leukemia and Lymphoma Society Clinical Scholar program (W.G.W.)"
- "The author(s) indicated no potential conflicts of interest."

Other comments

- none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	No	Individuals without evaluation of clinical factors and at least one newer factor were excluded: "We identified 930 previously untreated patients who presented to MD Anderson Cancer Center (MDACC) between January 2004 and December 2009, were not recommended for first-line treatment at initial visit, and were evaluated for traditional clinical and laboratory prognostic factors and one or more of the newer prognostic factors including IgHV mutation status, chromosome abnormalities by FISH analysis, and ZAP-70 expression by flow cytometry and/or immunohistochemistry (IHC)."
Domain 2: Predictors	Yes	Predictor assessment well described; comparison of assessment tools
Domain 3: Outcome	No	Some univariably tested predictors were not excluded from the outcome definition (progressive or symptomatic splenomegaly, massive nodes, progressive or symptomatic lymphadenopathy, Rai stage; Hallek 2008); short observation time
Domain 4: Analysis	No	Many predictors compared to events, 193 events (more than 25 predictors); univariable selection of predictors; unclear how many predictors entered in-

MDACC 2011 D - Wierda 2011 (MDACC)

(Continued)

to MV analysis; no performance measures reported; split sample, but after the model was already developed, which means that the data were internally validated on data that it was developed with. Bootstrapping was done, however not used to correct the model; assumption of proportionality was checked

Overall judgement	No

MDACC 2011 V - Molica 2016 (O-CLL1-GISL)

Study characteristics

General information

Model and type of study

- Molica 2013 (O-CLL1-GISL)
- Validation study

Secondary citations

- Molica 2013; Molica 2014; Molica 2016; Molica 2017; Molica 2017

Language of publication

- English

Study design

- prospective cohort

Follow-up time

- median 30 months (range 1 - 65 months)

Participants

Number of included participants in the cohort

- 328 participants

Setting

- Italy, multicentre study

Recruitment period

- 2006 - 2010

Age (in years)

- median 61 years (range: 33 - 70 years)

Sex

- 58.2% male

Stages of disease

- Rai 0: 76.5%; Rai I-II: 23.5%
- Binet A: 100%

Treatment

- 91 received treatment
- unclear which treatment

MDACC 2011 V - Molica 2016 (O-CLL1-GISL) (Continued)

Inclusion criteria

- The inclusion criteria for CLL diagnosis, which were used at the time of study design and initiation, followed the National Cancer Institute-sponsored Working Group guidelines, which require absolute lymphocytosis with a lower threshold of above 5000 mature-appearing lymphocytes/mL in the peripheral blood. Because the objective of this observational study was to evaluate the role of novel prognostic variables in younger patients with CLL, only those aged above 70 years were considered eligible.

Exclusion criteria

- not reported

Predictors

List of predictors used for validation (and changes between original predictors and predictors in validation study)

- IgHV status (continuous: categorical/binary transformation); diameter of largest palpated cervical LN, cm (continuous: no transformation); FISH category (categorical; del(11q) or del(17p) vs. None); number of involved lymph node sites (continuous: categorical/binary transformation); lactate dehydrogenase IU/L/100 (continuous: no transformation); IgHV status*LDH (interaction term)

Timing of predictor measurement

- at study entry, within 12 months of diagnosis

Outcome(s)

Primary outcome in study

- time-to-first-treatment: TTFT was defined as the interval between the date of registration and the date of initiation of first CLL treatment. Patients underwent sequential monitoring, and the frequency of follow-up visits was individualised according to patient risk; this ranged from 3 to 6 months (median 6 months). All physicians who registered patients in this observational database stated that they had used the NCI-WG guidelines as a reference criterion for starting therapy. In particular, the absolute lymphocyte count was not used as the sole indicator for treatment. Active disease, requiring therapy, was defined when at least one of the criteria set out in the NCI-sponsored Working Group guidelines was satisfied.

Additional outcome(s)

- not applicable

Outcome in model development

- time-to-first-treatment

Missing data

Participants with any missing data?

- no

If yes, how was missing data handled?

- not applicable

Analysis

Number of participants & number of events (specific time points where reported)

- time-to-first-treatment: 337 participants, 91 events
- in Molica 2013: time-to-first-treatment: 328 participants, 68 events

Which model was used?

- Cox proportional hazard model

Was the model updated?

- no

MDACC 2011 V - Molica 2016 (O-CLL1-GISL) (Continued)

Performance measures reported?

- Calibration: reported (calibration plot)
- Discrimination: reported (c-statistic)

Creation of risk groups?

- no

PROBAST: Applicability	<u>Domain 1: Participant selection</u> • low <u>Domain 2: Predictors</u> • low <u>Domain 3: Outcome</u> • low
Notes	<u>Funding & conflict of interest</u> • "This work was supported by grants from Associazione Italiana Ricerca sul Cancro (AIRC) (to Antonino Neri -IG4569, MFIG10492 and Fortunato Morabito -RG6432) and AIRC-Special Program Molecular Clinical Oncology-"5 per mille", grant 9980, 2010-15 to Antonino Neri, Manlio Ferrarini and Fortunato Morabito; Ricerca Finalizzata from Italian Ministry of Health 2006 (GC, FM and MF) and 2007 (GC)." • "The authors made no disclosures." <u>Other comments</u> • The O-CLL1-GISL cohort was used in various publications and to validate different CLL models.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Prospective cohort, detailed eligibility criteria in trial registry
Domain 2: Predictors	Yes	Predictors prospectively collected
Domain 3: Outcome	No	See development of the model: some predictors in the model were not excluded from the outcome definition (number and size of lymph nodes included in model, however these factors also formed Binet stage, which was a reason for treatment indication, Hallek 2008).
Domain 4: Analysis	No	Low number of events; no information on missing data reported
Overall judgement	No	

Morabito D - Morabito 2009 (Italian cohort)

Study characteristics

General information	<u>Model and type of study</u> • Morabito 2009 (Development cohort (Italian)) • Development study

Morabito D - Morabito 2009 (Italian cohort) (Continued)

Secondary citations

- not applicable

Language of publication

- English

Study design

- not reported

Follow-up time

- median 36 months (range 12 - 180 months)

Participants

Number of included participants in the cohort

- 262 participants

Setting

- Italy, multicentre study

Recruitment period

- not reported

Age (in years)

- median 65 years (range not reported)

Sex

- 61.5% male

Stages of disease

- Rai 0-IV: not reported
- Binet A: 100%

Treatment

- not reported

Inclusion criteria

- diagnosis of typical CLL based on the National Cancer Institute (NCI) Working Group criteria (Hallek 2008) and confirmed by flow cytometry

Exclusion criteria

- not reported

Predictors

Number of candidate predictors

- univariable analysis: 3
- multivariable analysis: not applicable

List of predictors in final model (including cut-points for dichotomised factors)

- IgHV mutational status (continuous: categorical/binary transformation; 2% threshold, as established in literature); CD38 (continuous: categorical/binary transformation; 30% threshold); ZAP-70 (continuous: categorical/binary transformation; threshold established by ROC curve)

Timing of predictor measurement

Morabito D - Morabito 2009 (Italian cohort) (Continued)

- markers were tested within two years in 70% of cases, within 4 years in 15% and from 4 to 5 years from diagnosis in the remaining 15%

Outcome(s)	<u>Primary outcome in study</u> • time-to-first-treatment: TTT was measured from diagnosis to first-line treatment or last follow-up. Treatment was decided uniformly in all participating centres based on documented progressive and symptomatic disease according to NCI working guidelines. <u>Additional outcome(s)</u> • not applicable
Missing data	<u>Participants with any missing data?</u> • no <u>If yes, how was missing data handled?</u> • complete-case analysis
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> • time-to-first-treatment: 252 participants, 51 events <u>Predictor selection method</u> • univariable analysis only (P value or CI); a scoring system was designed in which one point was assigned to each unfavourable prognostic marker and the final score was determined by the sum of unfavourable markers present. <u>Statistical method</u> • univariable analysis: Kaplan-Meier method <u>Simplification of model?</u> • not applicable <u>Performance measures reported?</u> • Calibration: not reported • Discrimination: not reported <u>Creation of risk groups?</u> • no
PROBAST: Applicability	<u>Domain 1: Participant selection</u> • unclear - not sufficient information (eligibility criteria, sample characteristics, inclusions and exclusions) on the sample to make a judgement <u>Domain 2: Predictors</u> • high - predictors not collected at study entry or at diagnosis: "Markers were tested within two years in 70% of cases, within 4 years in 15% and from 4 to 5 years from diagnosis in the remaining 15%." <u>Domain 3: Outcome</u> • low
Notes	<u>Funding & conflict of interest</u>

Morabito D - Morabito 2009 (Italian cohort) (Continued)

- "Supported in part from Associazione Italiana Ricerca sul Cancro (AIRC) (to FM, AN and MF), FIRB (Grant no RBIP06LCA9, to MF) MIUR, CIPE (2006–8, CBA project, to SZ), ISS (to SZ), Progetti Strategici – Ricerca Finalizzata Ministero Italiano della Salute 'RFPS_2006_3_33_99_60' (to GC) and 'RF-PS_2006_340196' (to FM and MF); progetto ordinario ricerca finalizzata Ministero Italiano della salute-2007 (to GC), progetto Compagnia San Paolo (to GC), the Fondazione Internazionale Ricerche Medicina Sperimentale (FIRMS) provided financial and administrative assistance, Fondazione 'Amelia Scorza' onlus, Cosenza, Italy; and Associazione Italiana contro le Leucemie -Milano. KT and SM are supported by fellowships from the Fondazione Italiana Ricerca sul Cancro (FIRC)".
- Conflicts of interests not reported

Other comments

- none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	Only one statement on inclusion criteria. No statement on baseline sample and possible exclusions. Very little baseline information on participants
Domain 2: Predictors	Yes	Prospective collection of predictor information, centralised laboratory
Domain 3: Outcome	Yes	Standard outcome definition used: "Treatment was decided uniformly in all participating centres based on documented progressive and symptomatic disease according to NCI working guidelines".
Domain 4: Analysis	No	Not sufficient events; continuous data not handled appropriately (ROC to find the 'best' cut-point); no information on missing data; no performance measure reported
Overall judgement	No	

Morabito V - Gentile 2014 (O-CLL1-GISL)

Study characteristics

General information	<u>Model and type of study</u> • Gentile 2014 (O-CLL1-GISL) • Validation study <u>Secondary citations</u> • not applicable <u>Language of publication</u> • English <u>Study design</u> • prospective cohort <u>Follow-up time</u> • median 38.5 months (range 6 - 82 months)
Participants	<u>Number of included participants in the cohort</u>

100

Morabito V - Gentile 2014 (O-CLL1-GISL) (Continued)

- 480 participants

Setting

- Italy, multicentre study

Recruitment period

- 2007 - not reported

Age (in years)

- 47.4 % of individuals were aged < 60 years

Sex

- 58.5% male

Stages of disease

- Rai 0: 77.3%; Rai I-II: 22.7%
- Binet A: 100%

Treatment

- not reported

Inclusion criteria

- the inclusion criteria for CLL diagnosis, employed at the time of study start, followed the NCI/WG guidelines established in 1996, required > 5.000 lymphocytes/mL in the peripheral blood.

Exclusion criteria

- exclusion criteria were the following: (i) CD5- and/or CD23- B-lymphoproliferative disorders; (ii) clinical Binet stage B or C, and (iii) need of therapy according to NCI/WG guidelines; (iv) age > 70 years.

Predictors	<u>List of predictors in final model (including cut-points for dichotomised factors)</u> • IgHV mutational status (continuous: categorical/binary transformation; 2% threshold, as established in literature); CD38 (continuous: categorical/binary transformation; 30% threshold); ZAP-70 (continuous: categorical/binary transformation; threshold established by ROC curve) • no difference between development and validation cohort predictors <u>Timing of predictor measurement</u> • at study entry, maximum 1 year after diagnosis
Outcome(s)	<u>Primary outcome in study</u> • progression-free survival: PFS has been defined as time to therapy requirement. <u>Additional outcome(s)</u> • not applicable <u>Outcome in model development</u> • time-to-first-treatment
Missing data	<u>Participants with any missing data?</u> • no <u>If yes, how was missing data handled?</u>

Morabito V - Gentile 2014 (O-CLL1-GISL) (Continued)

- not applicable

Analysis	<u>Number of participants & number of events (specific time points where reported)</u> • progression-free survival: 468 participants, 121 events <u>Which model was used?</u> • original model <u>Was the model updated?</u> • no <u>Performance measures reported?</u> • Calibration: not reported • Discrimination: not reported <u>Creation of risk groups?</u> • yes, 3 risk groups, (cut-points at 1; 2-3; 4) • at 3 years • percentage of group without event at time point (low risk: 91.7%, intermediate risk: 82.9%, high risk: 57.4%)
PROBAST: Applicability	<u>Domain 1: Participant selection</u> • low <u>Domain 2: Predictors</u> • low <u>Domain 3: Outcome</u> • low
Notes	<u>Funding & conflict of interest</u> • "Associazione Italiana per la Ricerca sul Cancro; Contract grant no.: 9980, IG10492, IG10136" • "Nothing to report" <u>Other comments</u> • The O-CLL1-GISL cohort was used in various publications and to validate different CLL models. • The corresponding author provided additional information regarding the cohort.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Prospectively registered study (O-CLL1-GISL) with defined inclusion and exclusion criteria
Domain 2: Predictors	Yes	Prospective collection of predictor information. Predictors assessed study entry at a maximum of 1 year after diagnosis.
Domain 3: Outcome	Yes	CLL progression criteria 2008 were used. Outcome labelled differently than model development study. However, progression-free survival (PFS) has been defined as time-to-therapy requirement, which is equivalent to time-to-treatment in Morabito 2009.

Morabito V - Gentile 2014 (O-CLL1-GISL) *(Continued)*

Domain 4: Analysis	Unclear	No relevant performance measures reported (progression per risk group was reported)

O-CLL-1 D - Gentile 2016 (O-CLL-1-GISL)

Study characteristics

General information	<u>Model and type of study</u> - Gentile 2016 (O-CLL1-GISL) - Development study <u>Secondary citations</u> - Molica 2016; Molica 2017; Molica 2017 <u>Language of publication</u> - English <u>Study design</u> - prospective cohort <u>Follow-up time</u> - median 42 months (range 6 - 82 months)
Participants	<u>Number of included participants in the cohort</u> 480 participants <u>Setting</u> - Italy, multicentre study <u>Recruitment period</u> 2007 - not reported <u>Age (in years)</u> 46.7% of individuals were younger than 60 years <u>Sex</u> 57.3% male <u>Stages of disease</u> - Rai 0: 77.7%; Rai I-II: 22.3% - Binet A: 100% <u>Treatment</u> 84 out of 337 received treatment - unclear which treatment <u>Inclusion criteria</u> - Established diagnosis of B-CLL by NCI criteria, performed by local haematologist (confirmed by the biological review committee according to flow cytometry analysis (positive clusters of differentiation

O-CLL-1 D - Gentile 2016 (O-CLL-1-GISL) (Continued)

antigen 5 (CD5), 19 (CD19), 23 (CD23)). Age more than 18 years and less than 70 years. Eastern Cooperative Oncology Group (ECOG) less than or equal to 2. Binet stage A. Diagnosis performed within 12 months before inclusion in the study. NCI watch-and-wait policy.

Exclusion criteria

- Patients with CLL whose diagnosis exceeded 12 months before registration. Patients with leukaemic phase of lymphoproliferative disorders of B origin CD5- and/or CD23- according to flow cytometry analysis. Clinical Binet stage B or C. Patients who necessitated therapy according to NCI guidelines. Age > 70 years. Without a written informed consent.

Predictors	<u>Number of candidate predictors</u> • univariable analysis: 10 • multivariable analysis: 7 <u>List of predictors in final model (including cut-points for dichotomised factors)</u> • Rai stage (categorical; 0 vs. I/II); ALC (continuous: categorical/binary transformation); beta2-microglobulin (continuous: categorical/binary transformation; normal/elevated (elevated above 2 mg/L)); IgHV status (continuous: categorical/binary transformation) <u>Timing of predictor measurement</u> • at study entry (< 12 months from diagnosis)
Outcome(s)	<u>Primary outcome in study</u> • time-to-first-treatment: TFS was calculated from diagnosis to the first CLL treatment (event) or to last follow-up (censoring) <u>Additional outcome(s)</u> • not applicable
Missing data	<u>Participants with any missing data?</u> • no <u>If yes, how was missing data handled?</u> • not applicable
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> • time-to-first-treatment: 337 participants, 84 events <u>Predictor selection method</u> • univariable analysis, unclear if stepwise approach was used in multivariable modelling (criterion: P value or CI) <u>Statistical method</u> • Cox proportional hazard model <u>Simplification of model?</u> • yes <u>Performance measures reported?</u> • Calibration: not reported • Discrimination: reported (c-statistic)

O-CLL-1 D - Gentile 2016 (O-CLL-1-GISL) (Continued)

Creation of risk groups?

- yes, 3 risk groups, (cut-points at (0-2); (3-5); (6-7))
- at 3 years
- percentage of group without event at time point (low risk: 95.3%, intermediate risk: 74.5%, high risk: 28.6%)

PROBAST: Applicability	<u>Domain 1: Participant selection</u> • low <u>Domain 2: Predictors</u> • low <u>Domain 3: Outcome</u> • low
Notes	<u>Funding & conflict of interest</u> • "Associazione Italiana per la Ricerca sul Cancro (AIRC 5xmille grant 9980, IG10492 to MF and FM and IG10136 to AN)." • "The authors declare no conflict of interest." <u>Other comments</u> • none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Prospective cohort, clear inclusion and exclusion criteria in trial registry
Domain 2: Predictors	Yes	Prospective and standardised collection of predictor information ("CLL cell phenotype, CD38 and ZAP-70 expression, and <i>IgHV</i> mutational status were performed in a central laboratory in Genova, while all FISH and genetic analyses were performed in Milan").
Domain 3: Outcome	No	Standard outcome definition, however, predictors were not excluded from outcome definition (e.g. Rai stage).
Domain 4: Analysis	No	Continuous data not handled appropriately ("Continuous variables of prognostic importance on TFS in univariate proportional hazards Cox regression were dichotomized using published thresholds and laboratory norms").
Overall judgement	No	

O-CLL-1 V - Rani 2018 (Indian cohort)

Study characteristics

General information	<u>Model and type of study</u> • Rani 2018 (Indian cohort) • Validation study

O-CLL-1 V - Rani 2018 (Indian cohort) *(Continued)*

Secondary citations

- not applicable

Language of publication

- English

Study design

- unclear

Follow-up time

- median 40.5 months (range: 1 - 215 months)

Participants

Number of included participants in the cohort

- 198 participants

Setting

- India, unclear if single or multicentre study

Recruitment period

- not reported

Age (in years)

- median 60 years (range not reported)

Sex

- 77% male

Stages of disease

- Rai 0: 14.6%; Rai I: 21.2%; Rai II: 34.9%; Rai III: 14.1%; Rai IV: 15.2%
- Binet A/B/C: not reported

Treatment

- 62 received chlorambucil-based therapy
- 56 received rituximab-based therapy
- 20 received other therapies

Inclusion criteria

- treatment-naïve CLL participants according to Hallek 2008

Exclusion criteria

- not reported

Predictors

List of predictors used for validation (and changes between original predictors and predictors in validation study)

- gender (binary); age (continuous, dichotomisation after univariable analysis); ECOG PS (categorical, dichotomisation after univariable analysis); del(17p) (continuous: categorical/binary transformation); del(11q) (continuous: categorical/binary transformation; literature-based); IgHV mutational status (continuous: categorical/binary transformation; literature-based); s-TK (continuous, i after univariable analysis); s-beta2m (continuous, dichotomisation after univariable analysis)

Timing of predictor measurement

O-CLL-1 V - Rani 2018 (Indian cohort) (Continued)

	not reported
Outcome(s)	<u>Primary outcome in study</u> - time-to-first-treatment: TTFT was defined as the time from the date of diagnosis to date of commencement of first therapy. <u>Additional outcome(s)</u> - not applicable <u>Outcome in model development</u> - time-to-first-treatment
Missing data	<u>Participants with any missing data?</u> - no <u>If yes, how was missing data handled?</u> - not applicable
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> - time-to-first-treatment: 140 participants, 89 events <u>Which model was used?</u> - simplified model <u>Was the model updated?</u> - no <u>Performance measures reported?</u> - Calibration: not reported - Discrimination: reported (c-statistic) <u>Creation of risk groups?</u> - yes, 3 risk groups, (cut-points at (0-2); (3-5); (6-7)) - median survival (low risk: 3 years, intermediate risk: 24 years, high risk: 79 years)
PROBAST: Applicability	<u>Domain 1: Participant selection</u> - unclear - recruitment period unclear <u>Domain 2: Predictors</u> - unclear - no information on timing of predictor because the study design was unclear <u>Domain 3: Outcome</u> - low
Notes	<u>Funding & conflict of interest</u> "The financial support was provided by the Department of Biotechnology (BT/PR11106/GBD/27/145/2008, BT/PR15438/MED/30/606/2011 and T/PR8680/AGR/36/754/2013), Ministry of Science and Technology, GOI, and All India Institute of Medical Sciences, New Delhi (8-60/A060/ 2011/RS) to RG for carrying out this work." "The authors declare that they have no conflict of interest."

O-CLL-1 V - Rani 2018 (Indian cohort) (Continued)

Other comments

- Graphs were provided to obtain the percentage of participants untreated at a specific time point, however, these graphs were printed very small.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	Inclusion and exclusion criteria unclear; it seemed to us that only individuals with complete information on all predictors were included; recruitment period unclear; study design unclear
Domain 2: Predictors	Yes	Not explicitly stated, but predictors probably assessed in a similar way (single-centre)
Domain 3: Outcome	No	See model development: Predictors were not excluded from outcome definition (e.g. Rai stage).
Domain 4: Analysis	No	Low number of events; no information on handling of missing values
Overall judgement	No	

O-CLL1 V - Gentile 2016 (Mayo cohort)

Study characteristics

General information	<u>Model and type of study</u> • Gentile 2016 (Mayo cohort 2001 - 2008) • Validation study <u>Secondary citations</u> • not applicable <u>Language of publication</u> • English <u>Study design</u> • not reported <u>Follow-up time</u> • median 97 months (range not reported)
Participants	<u>Number of included participants in the cohort</u> • 428 participants <u>Setting</u> • USA, single-centre study <u>Recruitment period</u> • 2001 - 2008

O-CLL1 V - Gentile 2016 (Mayo cohort) (Continued)

Age (in years)

- 40.2% ≤ 60 years

Sex

- 66.4% male

Stages of disease

- Rai 0: 67.5%; Rai I-II: 32.5%
- Binet A/B/C: not reported

Treatment

- 130 were treated
- unclear which treatment

Inclusion criteria

- not reported

Exclusion criteria

- not reported

Predictors

List of predictors used for validation (and changes between original predictors and predictors in validation study)

- Rai stage (categorical; 0 vs. I/II); ALC (continuous: categorical/binary transformation); beta2-microglobulin (continuous: categorical/binary transformation; normal/elevated (elevated above 2 mg/L)); IgHV status (continuous: categorical/binary transformation)

Timing of predictor measurement

- within 18 months of diagnosis

Outcome(s)

Primary outcome in study

- time-to-first-treatment: TFS was calculated from diagnosis to the first CLL treatment (event) or to last follow-up (censoring).

Additional outcome(s)

- not applicable

Outcome in model development

- time-to-first-treatment

Missing data

Participants with any missing data?

- not reported

If yes, how was missing data handled?

- not reported

Analysis

Number of participants & number of events (specific time points where reported)

- time-to-first-treatment: 428 participants, 130 events

Which model was used?

- simplified model

O-CLL1 V - Gentile 2016 (Mayo cohort) (Continued)

Was the model updated?

- no

Performance measures reported?

- Calibration: yes (Hosmer-Lemeshow test)
- Discrimination: reported (c-statistic)

Creation of risk groups?

- yes, 3 risk groups, (cut-points at (0-2); (3-5); (6-7))
- at 3 years
- percentage of group without event at time point (low risk: 95.5%, intermediate risk: 78.9%, high risk: 40.6%)

PROBAST: Applicability	<u>Domain 1: Participant selection</u> - unclear - recruitment period unclear <u>Domain 2: Predictors</u> - low <u>Domain 3: Outcome</u> - low
Notes	<u>Funding & conflict of interest</u> "Associazione Italiana per la Ricerca sul Cancro (AIRC 5xmille grant 9980, IG10492 to MF and FM and IG10136 to AN)." "The authors declare no conflict of interest." <u>Other comments</u> - none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	No	No eligibility criteria; individuals with missing values were excluded.
Domain 2: Predictors	Yes	Not explicitly stated, but predictors probably assessed in a similar way (single-centre).
Domain 3: Outcome	No	See model development: Predictors not excluded from outcome definition (e.g. Rai stage)
Domain 4: Analysis	Unclear	Calibration not reported; individuals with missing values were probably excluded from analysis, rated in domain 1.
Overall judgement	No	

Rossi D - Rossi 2013 (Italian cohort)

Study characteristics

General information	<u>Model and type of study</u> • Rossi 2013 • Development study <u>Secondary citations</u> • not applicable <u>Language of publication</u> • English <u>Study design</u> • retrospective cohort <u>Follow-up time</u> • median 67.2 months (range not reported)
Participants	<u>Number of included participants in the cohort</u> • 637 participants <u>Setting</u> • Italy, multicentre study <u>Recruitment period</u> • 1996 - 2011 <u>Age (in years)</u> • 51.2% of individuals were older than 65 years. <u>Sex</u> • 58.4% male <u>Stages of disease</u> • Rai 0: 74.9%; Rai I-II: 11.5% • Binet A: not reported <u>Treatment</u> • 122 patients received rituximab-based regimens (i.e. fludarabine-cyclophosphamide-rituximab, fludarabine-rituximab, or pentostatin-cyclophosphamide-rituximab) • 64 received fludarabine-based regimens (i.e. fludarabine or fludarabine-cyclophosphamide) • 80 received alkylator-based regimens (i.e. chlorambucil) <u>Inclusion criteria</u> • 'newly diagnosed and previously untreated CLL patients of whom 583 (91.5%) were provided with regular follow-up' <u>Exclusion criteria</u> • not reported
Predictors	<u>Number of candidate predictors</u>

Rossi D - Rossi 2013 (Italian cohort) (Continued)

- univariable analysis: no univariable analysis
- multivariable analysis: 18

List of predictors in final model (including cut-points for dichotomised factors)

- TP53 (binary); BIRC3 DIS (binary); SF3B1 M (binary); NOTCH1 M (binary); del11q22-q23 (binary)

Timing of predictor measurement

- at diagnosis

Outcome(s)	<u>Primary outcome in study</u> • overall survival: OS was measured from the date of initial presentation to the date of death from any cause (event) or last follow-up (censoring). <u>Additional outcome(s)</u> • Treatment-free survival: TFS was measured from date of initial presentation to date of progressive disease requiring treatment according to IWCLL-NCI guidelines (event), death or last follow-up (censoring).
Missing data	<u>Participants with any missing data?</u> • yes (54 participants) <u>If yes, how was missing data handled?</u> • not reported
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> • overall survival: 583 participants, 178 events <u>Predictor selection method</u> • full model, during multivariable modelling: hierarchical clustering <u>Statistical method</u> • recursive partitioning <u>Simplification of model?</u> • no <u>Performance measures reported?</u> • Calibration: not reported • Discrimination: reported (c-statistic) <u>Creation of risk groups?</u> • yes, 4 risk groups, (cut-points at very low risk: del13q14 sole lesion; low risk: +12 or all lesions; intermediate: NOTCH1 and/or SF3B1; high risk: TP53 and/or BIRC3) • at 5 years • percentage of group without event at time point (low risk: 86.9%, intermediate risk: 77.6%, high risk: 65.9%, very high risk: 50.9%)
PROBAST: Applicability	<u>Domain 1: Participant selection</u> • low <u>Domain 2: Predictors</u>

Rossi D - Rossi 2013 (Italian cohort) (Continued)

- low

Domain 3: Outcome

- low

Notes

Funding & conflict of interest

- "This study was supported by the Associazione Italiana per la Ricerca sul Cancro Foundation, Special Program Molecular Clinical Oncology, 5 1000, number 10007, Milan, Italy (to G.G. and to R.F.); Progetto Futuro in Ricerca 2008 (to D.R.); Programmi di Ricerca di Rilevante Interesse Nazionale (PRIN) 2008 (to G.G. and R.M.); PRIN 2009 (to D.R.); Progetto Futuro in Ricerca 2012 (to D.R.); Ministero dell'Istruzione, dell'Università e della Ricerca, Rome, Italy; Progetto Giovani Ricercatori 2008 (to D.R.); Progetto Giovani Ricercatori 2010 (to D.R.); Ricerca Sanitaria Finalizzata 2008 (to G.G.); Ministero della Salute, Rome, Italy; Novara-AIL Onlus Foundation, Novara, Italy (to G.G. and D.R.); Compagnia di San Paolo, Turin, Italy (to R.F.); Helmut Horten Foundation and San Salvatore Foundation, Lugano, Switzerland (to F.B.); Nelia et Amadeo Barletta Foundation, Lausanne (to F.B.); National Institutes of Health grant PO1-CA092625 (to R.D.-F.); and a Specialized Center of Research grant from the Leukemia and Lymphoma Society (to R.D.-F.). S.M. and S.C. are supported by fellowships from the Novara-AIL Onlus Foundation, Novara, Italy. L.P. is on leave from the University of Perugia Medical School."
- "The authors declare no competing financial interests."

Other comments

- none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Appropriate data sources used: "multicentric cohort of 637 newly diagnosed and previously untreated CLL patients who consecutively presented for initial evaluation at four institutions from June 1996 through June 2011"
Domain 2: Predictors	Yes	Detailed description of predictors and timing of measurements: "clinical information prospectively collected at clinically relevant time points (i.e. at diagnosis, progression, and last follow-up)"
Domain 3: Outcome	Yes	Objective standard outcome; regular clinical database update
Domain 4: Analysis	Unclear	Unclear if individuals with missing data were excluded; number of events sufficient; building of model by recursive partitioning (decision tree, tenfold cross-validation), testing for stability with random forest algorithm and Cox model with bootstrapping; correction of c-statistic; assumptions were checked

Rossi V - Jeromin 2014 (Munich cohort)

Study characteristics

General information

Model and type of study

- Jeromin 2014 (MLL Munich cohort)
- Validation study

Secondary citations

- not applicable

Language of publication

Rossi V - Jeromin 2014 (Munich cohort) (Continued)

- English

Study design

- prospective cohort

Follow-up time

- median 55.2 months (range not reported)

Participants

Number of included participants in the cohort

- 1160 participants

Setting

- probably Germany, unclear

Recruitment period

- 2005 - 2010

Age (in years)

- median 67 years (range: 29.6 - 90.5 years)

Sex

- 64.6% male

Stages of disease

- not reported

Treatment

- Not reported

Inclusion criteria

- not reported

Exclusion criteria

- prior treatment

Predictors

List of predictors used for validation (and changes between original predictors and predictors in validation study)

- TP53 (binary); BIRC3 DIS (binary); SF3B1 M (binary); NOTCH1 M (binary); del11q22-q23 (binary)
- BIRC3, which is part of the original model, was not analysed in this study.

Timing of predictor measurement

- at diagnosis for most participants (82.6%)

Outcome(s)

Primary outcome in study

- overall survival: OS was measured from the date of diagnosis until last follow-up or death.

Additional outcome(s)

- time to treatment: TTT was evaluated from the date of diagnosis until the date of initial treatment.

Outcome in model development

Rossi V - Jeromin 2014 (Munich cohort) (Continued)

	• overall survival	
Missing data	<u>Participants with any missing data?</u> • yes (160 participants) <u>If yes, how was missing data handled?</u> • complete-case analysis	
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> • overall survival: 930 participants, events not reported <u>Which model was used?</u> • original model as developed in Rossi 2013 <u>Was the model updated?</u> • no <u>Performance measures reported?</u> • Calibration: not reported • Discrimination: reported (c-statistic) <u>Creation of risk groups?</u> • yes, 4 risk groups (cut-points not reported) • at 5 years • percentage of group without event at time point (low risk: 91%, intermediate risk: 90%, high risk: 75.2%, very high risk: 62.1%)	
PROBAST: Applicability	<u>Domain 1: Participant selection</u> • low <u>Domain 2: Predictors</u> • low <u>Domain 3: Outcome</u> • low	
Notes	<u>Funding & conflict of interest</u> • "Next-generation deep sequencing studies were supported in part by the IRON-II study framework (Roche Diagnostics, Penzberg, Germany)." • "SS, WK, CH, and TH are part owners of the MLL Munich Leukemia Laboratory. SJ, SW, VG, KB, FD, TA, AR and AK are employed by the MLL Munich Leukemia Laboratory." <u>Other comments</u> • none	
Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	Study design unclear

Rossi V - Jeromin 2014 (Munich cohort) (Continued)

Domain 2: Predictors	Yes	Relatively short recruitment period (2005 - 2010); samples were sent to a laboratory which was not involved in treatment.
Domain 3: Outcome	Yes	Objective standard outcome; no information on outcome assessment. However, we assumed that for OS, risk of bias was low.
Domain 4: Analysis	No	Individuals with missing data excluded from analysis (complete data for OS in 935/1160 cases); no performance measures, just risk groups
Overall judgement	No	

Rossi V - Rossi 2013 (unclear)

Study characteristics

General information	<u>Model and type of study</u> - Rossi 2013 (unclear) - Validation study <u>Secondary citations</u> - not applicable <u>Language of publication</u> - English <u>Study design</u> - retrospective cohort <u>Follow-up time</u> - median 70.8 months, (range not reported)
Participants	<u>Number of included participants in the cohort</u> - not reported <u>Setting</u> - Italy, multicentre study <u>Recruitment period</u> 1996 - 2011 <u>Age (in years)</u> - not reported <u>Sex</u> 55.9% male <u>Stages of disease</u> - Rai 0-I: 75.4%; Rai II: 17.3%; Rai III-IV: 7.3% <u>Treatment</u>

Rossi V - Rossi 2013 (unclear) (Continued)

- not reported

Inclusion criteria

- not reported

Exclusion criteria

- not reported

Predictors	<u>List of predictors used for validation (and changes between original predictors and predictors in validation study)</u> • TP53 (binary); BIRC3 DIS (binary); SF3B1 M (binary); NOTCH1 M (binary); del11q22-q23 (binary) • same predictors <u>Timing of predictor measurement</u> • at diagnosis
Outcome(s)	<u>Primary outcome in study</u> • overall survival <u>Additional outcome(s)</u> • not applicable <u>Outcome in model development</u> • overall survival
Missing data	<u>Participants with any missing data?</u> • unclear <u>If yes, how was missing data handled?</u> • not applicable
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> • overall survival: 370 participants, 62 events <u>Which model was used?</u> • recursive partitioning <u>Was the model updated?</u> • yes • What has been updated? Excel sheet <u>Performance measures reported?</u> • Calibration: reported (observed-expected ratio) • Discrimination: reported (c-statistic) <u>Creation of risk groups?</u> • yes, 4 risk groups, (cut-points at very low risk: del13q14 sole lesion; low risk: +12 or all lesions; intermediate: NOTCH1 and/or SF3B1; high risk: TP53 and/or BIRC3) • at 5 years

Rossi V - Rossi 2013 (unclear) (Continued)

- percentage of group without event at time point (low risk: 86.9%, intermediate risk: 77.6%, high risk: 65.9%, very high risk: 50.9%)

PROBAST: Applicability	<u>Domain 1: Participant selection</u> - low <u>Domain 2: Predictors</u> - low <u>Domain 3: Outcome</u> - low
Notes	<u>Funding & conflict of interest</u> "This study was supported by the Associazione Italiana per la Ricerca sul Cancro Foundation, Special Program Molecular Clinical Oncology, 5 1000, number 10007, Milan, Italy (to G.G. and to R.F.); Progetto Futuro in Ricerca 2008 (to D.R.); Programmi di Ricerca di Rilevante Interesse Nazionale (PRIN) 2008 (to G.G. and R.M.); PRIN 2009 (to D.R.); Progetto Futuro in Ricerca 2012 (to D.R.); Ministero dell'Istruzione, dell'Università e della Ricerca, Rome, Italy; Progetto Giovani Ricercatori 2008 (to D.R.); Progetto Giovani Ricercatori 2010 (to D.R.); Ricerca Sanitaria Finalizzata 2008 (to G.G.); Ministero della Salute, Rome, Italy; Novara-AIL Onlus Foundation, Novara, Italy (to G.G. and D.R.); Compagnia di San Paolo, Turin, Italy (to R.F.); Helmut Horten Foundation and San Salvatore Foundation, Lugano, Switzerland (to F.B.); Nelia et Amadeo Barletta Foundation, Lausanne (to F.B.); National Institutes of Health grant PO1-CA092625 (to R.D.-F.); and a Specialized Center of Research grant from the Leukemia and Lymphoma Society (to R.D.-F.). S.M. and S.C. are supported by fellowships from the Novara-AIL Onlus Foundation, Novara, Italy. L.P. is on leave from the University of Perugia Medical School." "The authors declare no competing financial interests." <u>Other comments</u> - difference between external validation cohort and development cohort unclear

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	Study design not clear: "The validation series was represented by a cohort of 370 newly diagnosed and previously untreated CLL patients who consecutively presented for initial evaluation from June 1996 through June 2011 (Table S1) and provided with regular follow-up (at least three visits/year)."
Domain 2: Predictors	Yes	No detailed description of predictor assessment; we assumed that it was similar to the development cohort described in the same publication.
Domain 3: Outcome	Yes	Objective standard outcome, regular clinical database update
Domain 4: Analysis	No	Small number of events; no information on missing values; calibration not reported
Overall judgement	No	

Stephens OS D - Stephens 2015 (Ohio cohort)

Study characteristics

Stephens OS D - Stephens 2015 (Ohio cohort) (Continued)

General information	<u>Model and type of study</u>
	- Stephens 2015 - Development study
	<u>Secondary citations</u>
	not applicable
	<u>Language of publication</u>
	English
	<u>Study design</u>
	retrospective cohort
	<u>Follow-up time</u>
	not reported
Participants	<u>Number of included participants in the cohort</u>
	114 participants
	<u>Setting</u>
	USA, single-centre study (Ohio State University)
	<u>Recruitment period</u>
	2002 - 2012
	<u>Age (in years)</u>
	median 62 years (range: 40 - 92 years)
	<u>Sex</u>
	not reported
	<u>Stages of disease</u>
	Rai 0: 33%; Rai I-II: 46%; Rai III-IV: 21%
	<u>Treatment</u>
	35 received purine analogue-based therapy 18 received treatment on a clinical trial 5 received other treatment
	<u>Inclusion criteria</u>
	patients with CLL with del(17p13.1), previously untreated
	<u>Exclusion criteria</u>
	not reported
Predictors	<u>Number of candidate predictors</u>
	- univariable analysis: no univariable analysis - multivariable analysis: 18
	<u>List of predictors in final model (including cut-points for dichotomised factors)</u>

Stephens OS D - Stephens 2015 (Ohio cohort) (Continued)

- for OS: ECOG Performance Status (binary); age (binary); lactate dehydrogenase (LDH; binary)

Timing of predictor measurement

- at presentation, median 4.8 months after diagnosis (0 days to 19.7 years)

Outcome(s)	<u>Primary outcome in study</u> • overall survival: OS was calculated from date of first visit until date of death or last follow-up. <u>Additional outcome(s)</u> • treatment-free survival: TFS was calculated from date of first visit until date of first treatment or death, censoring patients alive and treatment-free at last follow-up.
Missing data	<u>Participants with any missing data?</u> • yes <u>If yes, how was missing data handled?</u> • multiple imputation technique
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> • OS: 114 participants, events not reported • TFS: 114 participants, 58 events <u>Predictor selection method</u> • univariable analysis, during multivariable modelling: backwards selection <u>Statistical method</u> • proportional hazards model <u>Simplification of model?</u> • yes • risk score (RS) based on the variables and regression coefficients of the multivariable model • simplified risk score (SRS) to be used in clinical practice based on the strength of associations with clinical outcome when all variables had been categorised <u>Performance measures reported?</u> • Calibration: not reported • Discrimination: reported for OS (c-statistic), but not reported for TFS (c-statistic for multivariable model only) <u>Creation of risk groups?</u> • for OS: yes, 3 risk groups (SRS = 0; 2; 4) • at 2 years
PROBAST: Applicability	<u>Domain 1: Participant selection</u> • high - selective sample of very high risk individuals with a del(17p) deletion <u>Domain 2: Predictors</u> • low <u>Domain 3: Outcome</u>

Stephens OS D - Stephens 2015 (Ohio cohort) (Continued)

- low

Notes	<u>Funding & conflict of interest</u> • "Contract grant sponsors: Four Winds Foundation, D. Warren Brown Foundation; Contract grant sponsors: Mr. and Mrs. Michael Thomas, Harry Mangurian Foundation; Contract grant no.: P50 CA140158; Contract grant sponsor: Leukemia and Lymphoma Society". • "The authors have no competing interests to disclose." <u>Other comments</u> • The authors reported a multivariable model, a risk score (RS) and a simplified risk score (SRS) for each outcome. Since the former two were not validated externally, only data for the SRS were extracted.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	No	Possibly selective inclusion of individuals, as only patients with del(17p13.1) were considered. Individuals without assessment could not have been included. Other eligibility criteria were unclear.
Domain 2: Predictors	Yes	Mostly laboratory data, probably assessed in a similar way: "Stimulated cytogenetic and fluorescent in situ hybridisation (FISH) analyses were performed on peripheral blood or bone marrow samples, as previously described"; single-centre study
Domain 3: Outcome	Yes	Objective standard outcome: although median observation time was not reported, median OS estimates were 5.2 years, therefore, we assumed a median observation time exceeding 5 years.
Domain 4: Analysis	No	Number of events only reported for TFS, not OS - the number of persons included was too small, total n = 114; missing data handling appropriate; inconsistent reporting of c-statistics; calibration was not reported
Overall judgement	No	

Stephens OS V - Stephens 2015 (MDACC)

Study characteristics

General information	<u>Model and type of study</u> • Stephens 2015 • Validation study <u>Secondary citations</u> • not applicable <u>Language of publication</u> • English <u>Study design</u> • retrospective cohort <u>Follow-up time</u>

Stephens OS V - Stephens 2015 (MDACC) (Continued)

	not reported
Participants	<u>Number of included participants in the cohort</u> 129 participants <u>Setting</u> - USA, single-centre study (MDACC) <u>Recruitment period</u> - not reported <u>Age (in years)</u> - median 63 years (range: 40 - 85 years) <u>Sex</u> - not reported <u>Stages of disease</u> - Rai 0: 29%; Rai I-II: 52%; Rai III-IV: 19% <u>Treatment</u> - not reported <u>Inclusion criteria</u> - patients with CLL with del(17p13.1), previously untreated <u>Exclusion criteria</u> - not reported
Predictors	<u>List of predictors used for validation (and changes between original predictors and predictors in validation study)</u> - for OS: ECOG Performance Status (binary); age (binary); lactate dehydrogenase (LDH; binary) - same predictors <u>Timing of predictor measurement</u> - not reported
Outcome(s)	<u>Primary outcome in study</u> - overall survival <u>Additional outcome(s)</u> - treatment-free survival <u>Outcome in model development</u> - overall survival - treatment-free survival
Missing data	<u>Participants with any missing data?</u> not reported <u>If yes, how was missing data handled?</u>

Stephens OS V - Stephens 2015 (MDACC) (Continued)

- not applicable

Analysis	<u>Number of participants & number of events (specific time points where reported)</u> • OS: 129 participants, events not reported • TFS: 129 participants, events not reported <u>Which model was used?</u> • simplified risk score (SRS) <u>Was the model updated?</u> • no <u>Performance measures reported?</u> • Calibration: not reported • Discrimination: reported for OS and TFS (c-statistic) <u>Creation of risk groups?</u> • for OS: yes, 3 risk groups (SRS = 0; 2; 3) • at 2 years
PROBAST: Applicability	<u>Domain 1: Participant selection</u> • high - selective sample of very high risk individuals with a del(17p) deletion <u>Domain 2: Predictors</u> • low <u>Domain 3: Outcome</u> • low
Notes	<u>Funding & conflict of interest</u> • "Contract grant sponsors: Four Winds Foundation, D. Warren Brown Foundation; Contract grant sponsors: Mr. and Mrs. Michael Thomas, Harry Mangurian Foundation; Contract grant no.: P50 CA140158; Contract grant sponsor: Leukemia and Lymphoma Society". • "The authors have no competing interests to disclose." <u>Other comments</u> • none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	No	Possibly selective inclusion of individuals, as only patients with del(17p13.1) were considered. Individuals without assessment could not have been included. Other eligibility criteria were unclear.
Domain 2: Predictors	Unclear	lack of information about similarity of predictor assessment to the development study; no recruitment period
Domain 3: Outcome	Yes	Objective standard outcome: although median observation time was not reported, median OS estimates were 6.4 years, therefore we assumed a median observation time exceeding 5 years.

Stephens OS V - Stephens 2015 (MDACC) *(Continued)*

Domain 4: Analysis	No	Number of included persons too small, total N = 129; calibration not reported
Overall judgement	No	

Stephens TFS D - Stephens 2015 (Ohio cohort)

Study characteristics

General information	<u>Model and type of study</u> - Stephens 2015 - Development study Different outcome but identical publication as Stephens OS D - Stephens 2015 (Ohio cohort) and Stephens OS V - Stephens 2015 (MDACC). Therefore, only relevant differences are reported in this table. For additional information, see tables above.
Participants	see Stephens OS D - Stephens 2015 (Ohio cohort)
Predictors	<u>List of predictors in final model (including cut-points for dichotomised factors)</u> for TFS: ECOG Performance Status (binary); Rai stage (binary); age (binary); white blood cell (WBC; binary); del(11q22.3) (binary)
Outcome(s)	<u>Primary outcome in study</u> - overall survival: OS was calculated from date of first visit until date of death or last follow-up. <u>Additional outcome(s)</u> - treatment-free survival: TFS was calculated from date of first visit until date of first treatment or death, censoring patients alive and treatment-free at last follow-up.
Missing data	
Analysis	<u>Number of participants & number of events (specific time points where reported)</u> - OS: 114 participants, events not reported - TFS: 114 participants, 58 events <u>Performance measures reported?</u> - Calibration: not reported - Discrimination: reported for OS (c-statistic), but not reported for TFS (c-statistic for multivariable model only) <u>Creation of risk groups?</u> - for TFS: yes, 3 risk groups (SRS = 0/1; 2/3; ≥ 4) - at 2 years
PROBAST: Applicability	
Notes	

Stephens TFS D - Stephens 2015 (Ohio cohort) (Continued)

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	No	Possibly selective inclusion of individuals, as only patients with del(17p13.1) were considered. Individuals without assessment could not have been included. Other eligibility criteria were unclear.
Domain 2: Predictors	Yes	Mostly laboratory data, probably assessed in a similar way: "Stimulated cytogenetic and fluorescent in situ hybridization (FISH) analyses were performed on peripheral blood or bone marrow samples, as previously described"; single-centre study
Domain 3: Outcome	No	Predictors were not excluded from the outcome definition (Rai stage; Hallek 2008)
Domain 4: Analysis	No	Number of events only reported for TFS, not OS - the number of persons included was too small, total N = 114; missing data handling appropriate; inconsistent reporting of c-statistics; calibration was not reported
Overall judgement	No	

Stephens TFS V - Stephens 2015 (MDACC)

Study characteristics

General information	<u>Model and type of study</u> - Stephens 2015 - Validation study Different outcome but identical publication as Stephens OS D - Stephens 2015 (Ohio cohort) and Stephens OS V - Stephens 2015 (MDACC). Therefore, only relevant differences are reported in this table. For additional information, see tables above.
Participants	
Predictors	<u>List of predictors used for validation (and changes between original predictors and predictors in validation study)</u> - for TFS: ECOG Performance Status (binary); Rai stage (binary); age (binary); white blood cell (WBC; binary); del(11q22.3) (binary) - same predictors
Outcome(s)	<u>Primary outcome in study</u> - overall survival <u>Additional outcome(s)</u> - treatment-free survival <u>Outcome in model development</u> - overall survival - treatment-free survival
Missing data	

Stephens TFS V - Stephens 2015 (MDACC) (Continued)

Analysis Number of participants & number of events (specific time points where reported)

- OS: 129 participants, events not reported
- TFS: 129 participants, events not reported

Performance measures reported?

- Calibration: not reported
- Discrimination: reported for OS and TFS (c-statistic)

Creation of risk groups?

- for TFS: yes, 3 risk groups (SRS = 0/1; 2/3; >= 4)
- at 2 years

PROBAST: Applicability

Notes

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	No	Possibly selective inclusion of individuals, as only patients with del(17p13.1) were considered. Individuals without assessment could not have been included. Other eligibility criteria were unclear.
Domain 2: Predictors	Unclear	Lack of information about similarity of predictor assessment to the development study, no recruitment period
Domain 3: Outcome	No	Predictors were not excluded from the outcome definition (Rai stage; Hallek 2008)
Domain 4: Analysis	No	Number of included persons too small, total N = 129; calibration not reported
Overall judgement	No	

AIHA: autoimmune haemolytic anaemia

ALC: absolute lymphocyte count

AUC: area under the curve

B2M: beta2-microglobulin

BcR IG: clonotypic B-cell receptor immunoglobulin

BIRC3: baculoviral IAP repeat containing 3

BR: bendamustine-rituximab

C: cyclophosphamide

CHL: chlorambucil

CHOP: cyclophosphamide, doxorubicin, vincristine, and prednisone

CI: confidence interval

CIRS: Cumulative Illness Rating Score

CLL: chronic lymphocytic leukaemia

CLL-IPI: Chronic Lymphocytic Leukaemia-International Prognostic Index

cMBL: clinical monoclonal B lymphocytosis

CVP: central venous pressure

DE: Germany

del11q: FISH-detected genetic aberration del11q

del17p: FISH-detected genetic aberration del17p

ECOG PS: Eastern Cooperative Oncology Group performance status

F: fludarabine

FC: fludarabine and cyclophosphamide

100

Prognostic models for newly-diagnosed chronic lymphocytic leukaemia in adults: a systematic review and meta-analysis (Review)

202

FCR: fludarabine, cyclophosphamide and rituximab
 FFP: fresh frozen plasma
 FR: France
 FISH: fluorescence in situ hybridisation
 HIV: human immunodeficiency virus
 HBV: hepatitis B virus
 HCV: hepatitis C virus
 HDMP: high-dose methylprednisolone
 IgHV: immunoglobulin heavy chain variable region genes
 IGLV: immunoglobulin lambda variable cluster
 IQR: Interquartile range
 IPD: individual patient data
 iwCLL: International workshop on Chronic Lymphocytic Leukemia
 LDH: lactate dehydrogenase
 LDT: lymphocyte doubling time
 LN: lymph node
 LYM: lymphocyte count
 MCAR: missing completely at random
 M-CLL: mutated chronic lymphocytic leukaemia
 MDACC: MD Anderson Cancer Center
 MLL: Munich Leukemia Laboratory
 MoAbs: monoclonal antibodies
 MV analyses: multivariable analysis
 NA: not available
 NCI: National Cancer Institute
 NOTCH1: notch receptor 1
 O-Benda: ofatumumab bendamustine
 O:E: observed-expected ratio
 OS: overall survival
 PFS: progression-free survival
 PL: Poland
 PROBAST: Prediction model Risk Of Bias ASsessment Tool
 R: rituximab
 RCT: randomised controlled trial
 ROC: receiver operating characteristics curve
 RR: risk ratio
 RS: risk score
 RTK: tyrosine kinase inhibitors
 s-B2M: serum beta-2 microglobulin
 SCALE: Scandinavian population-based study
 SCAN: Scandinavian population-based case-control study
 SF3B1: Splicing Factor 3b Subunit 1
 SRS: simplified risk score
 s-TK: serum thymidine kinase
 TFS: treatment-free survival
 TP53: tumour protein P53
 TP: treatment probability
 TTE: Time-to-Event
 TTFT: time-to-first-treatment
 U-CLL: unmutated chronic lymphocytic leukaemia
 WBC: white blood cell count
 WHO: World Health Organization

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Apelgren 2006	Study focus on staging system(s)
Baccarani 1982	Study focus on staging system(s)

Study	Reason for exclusion
Baliakas 2015	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Berke 2019	Prognostic factor study
Bettini 1986	Outdated staging system, which is no longer used in today's clinical practice
Binet 1977	Study focus on staging system(s)
Binet 1981	Study focus on staging system(s)
Bo 2014	Prognostic factor study
Bomben 2005	Comment or addition referring to an excluded study
Bomben 2009	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Bou Samra 2014	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Brejcha 2010	Prognostic factor study
Brugiatelli 2007	Prognostic factor study
Bulian 2014	Prognostic factor study
Byrd 2006	Prognostic factor study
Cailliod 2005	Prognostic factor study
Callea 1999	Prognostic factor study
Catovsky 1989	Prognostic factor study
Cesano 2013	Prognostic factor study
Chang 2003	Prognostic factor study
Chastang 1985	Outdated staging system, which is no longer used in today's clinical practice
Chauzeix 2018	Prognostic factor study
Chelazzi 1979	Study focus on staging system(s)
Chen 1997	Prognostic factor study
Chena 2008	Prognostic factor study
Chevallier 2002	Prognostic factor study
Chiaretti 2014	Prognostic factor study
Christiansen 1994	Prognostic factor study
Chuang 2012	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Ciccone 2012	Prognostic factor study

100

Study	Reason for exclusion
Ciocoiu 1988	Study focus on staging system(s)
Claus 2012	Prognostic factor study
Claus 2014	Prognostic factor study
Cmunt 2002	Prognostic factor study
Cocco 2005	Prognostic factor study
Corcoran 2005	Prognostic factor study
Cordone 1998	Prognostic factor study
Cortese 2014	Prognostic factor study
Coscia 2012	Prognostic factor study
Crespo 2003	Prognostic factor study
Criel 1997	Characterisation of CLL and prognostic factors
Cro 2009	Prognostic factor study
Cro 2010	Characterisation of CLL and prognostic factors
Cuneo 2004	Characterisation of CLL and prognostic factors
Cuneo 2018	Observational efficacy study
D'Arena 2001	Prognostic factor study
D'Arena 2007	Prognostic factor study
D'Arena 2012	Characterisation of CLL and prognostic factors
Damle 1999	Prognostic factor study
Dasgupta 2015	Study focus on identification of the threshold for a prognostic factor
Davis 2016	Study focus on identification of the threshold for a prognostic factor
De Faria 2000	Study focus on staging system(s)
De Rossi 1989	Study focus on staging system(s)
DeAndres-Galiana 2016	Study focus on identification of a prognostic factor
Degan 2004	Prognostic factor study
Del Guidice 2011	Prognostic factor study
Del Poeta 2010	Prognostic factor study
Del Principe 2004	Prognostic factor study

Study	Reason for exclusion
Del Principe 2006	Prognostic factor study
Delgado 2009	Prognostic factor study
Delgado 2014	Prognostic factor study
Deslandes 2007	Methodological study
Di Raimondo 2001	Prognostic factor study
Dimier 2018	Prediction model to evaluate the relationship between surrogacy outcome(s) and standard outcomes
Dong 2011	Prognostic factor study
Dong 2014	Prognostic factor study
Durak 2009	Prognostic factor study
El-Kinawy 2012	Prognostic factor study
Fang 2019	Study applies the CLL-IPI, but without the aim of validating it
Ferrara 1981	Study focus on staging system(s)
Ferreira 2014	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
French Cooperative Group on CLL 1988	Outdated staging system, which is no longer used in today's clinical practice
Friedman 2009	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Gattei 2008	Prognostic factor study
Gdynia 2018	Prognostic factor study
Geisler 1997	Characterisation of CLL and prognostic factors
Gentile 2016	Prognostic factor study
Gentile 2018	Score or model included predictors available only at treatment initiation
Giles 2003	Patient population did not match the review question; study included previously treated CLL patients
Giudice 2018	Prognostic factor study
Gogia 2014	Prognostic factor study
Gonzalez 2013	Characterisation of CLL and prognostic factors
Gonzalez-Gascon 2015	Characterisation of CLL and prognostic factors
Gonzalez-Rodriguez 2010	Characterisation of CLL and prognostic factors

Study	Reason for exclusion
Grabowski 2005	Prognostic factor study
Grever 2006	Abstract only
Hallek 1999	Characterisation of CLL and prognostic factors
Han 1984	Prognostic factor study
Herold 2011	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Hock 2010	Prognostic factor study
Houldsworth 2014	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Hus 2006	Prognostic factor study
Jaksic 1981	Prognostic factor study
Jaksic 1992	Study focus on staging system(s)
Jaksic 2014	Comment or addition referring to an excluded study
Josefsson 2007	Prognostic factor study
Juliussen 1986	Prognostic factor study
Juliussen 1990	Prognostic factor study
Kahraman 2014	Prognostic factor study
Kardum-Skelin 2008	Prognostic factor study
Kardum-Skelin 2009	Patient population did not match the review question; study included previously treated CLL patients
Karmiris 1994	Prognostic factor study
Kay 2018	Study applied the CLL-IPI, but without the aim of validating it
Keating 2000	Patient population did not match the review question; study included previously treated CLL patients
Khalifa 2002	Prognostic factor study
Kim 2004	Prognostic factor study
Kimby 1988	Prognostic factor study
Kleinstern 2018	Risk factor study
Knospe 1977	Prognostic factor study
Koberda 1989	Prognostic factor study
Korycka-Wolowiec 2011	Prognostic factor study

Study	Reason for exclusion
Krober 2002	Prognostic factor study
Krober 2006	Patient population did not match the review question; study included previously treated CLL patients
Kryachok 2011	Prognostic factor study
Kurec 1992	Prognostic factor study
Lai 2002	Prognostic factor study
Lech-Maranda 2012	Prognostic factor study
Lech-Maranda 2013	Prognostic factor study
Lecouvet 1997	Prognostic factor study
Li 2008	Prognostic factor study
Li 2017a	Prognostic factor study
Lin 2002	Prognostic factor study
Lin 2014	Prognostic factor study
Lozano-Santos 2014	Prognostic factor study
Lucas 2015	Prognostic factor study
Maffei 2007	Prognostic factor study
Maffei 2010	Prognostic factor study
Mandelli 1987	Development study of an outdated staging system, which is no longer used in today's clinical practice
Mansouri 2013	Prognostic factor study
Marasca 2005	Prognostic factor study
Marasca 2013	Prognostic factor study
Martinelli 2008	Prognostic factor study
Masic 1998	Prognostic factor identification study
Mateva 2001	Prognostic factor model, no full-text for further investigation
Matthews 2006	Prognostic factor study
Matthews 2007	Prognostic factor study
Matutes 2013	Prognostic factor study
Matutes 2017	Review

Study	Reason for exclusion
Melo 1987	Patient population did not match the review question; study was not restricted to previously untreated CLL patients
Miao 2018	Prognostic factor study
Molica 1984	Study focus on staging system(s)
Molica 1986	Prognostic factor study
Molica 1988	Prognostic factor study
Molica 1991	Prognostic factor study
Molica 1994	Prognostic factor study
Molica 1998	Prognostic factor study
Molica 1999a	Prognostic factor study
Molica 1999b	Prognostic factor study
Molica 2008	Prognostic factor study
Montserrat 1991	Prognostic factor study
Morabito 2001	Prognostic factor study
Morabito 2015a	Prognostic factor study
Morabito 2015b	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Morabito 2018c	Study aimed to identify cut-o& points for a specific prognostic factor
Moreno 2019	Study focus on staging system(s)
Morilla 2008	Prognostic factor study
Nabhan 2017	Score or model included predictors available only at treatment initiation
Nedeva 2018	Evaluation of an excluded model
Nenova 2000	Prognostic factor study
Nipp 2014	Prognostic factor study
Nola 2004	Patient population did not match the review question; study was not restricted to previously untreated CLL patients
Nowakowski 2009	Prognostic factor study
Nuckel 2006	Prognostic factor study
Nuckel 2009	Prognostic factor study

Study	Reason for exclusion
O'Brien 1993	Patient population did not match the review question; study was not restricted to previously untreated CLL patients
Ocana 2007	Prognostic factor study
Oliveira 2011	Prognostic factor study
Orgueira 2019	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Oscier 1990	Prognostic factor study
Paolino 1984	Prognostic factor study
Plesingerova 2017	Study focus on surrogacy outcome(s)
Prokocimer 1985	Prognostic factor study
Qin 2017	Prognostic factor study
Queiros 2015	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Rai 1975	Study focus on staging system(s)
Rai 1990	Study focus on staging system(s)
Raponi 2018	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Resegotti 1989	Prognostic factor study
Rissiek 2014	Prognostic factor study
Rodriguez 2007	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Ronchetti 2016	Prognostic factor study
Rossi 1986	Study focus on staging system(s)
Rossi 2010a	Study focus on identification of the threshold for a prognostic factor
Rossi 2010b	Patient population did not match the review question; study was not restricted to previously untreated CLL patients
Rossi 2011	Patient population did not match the review question; study was not restricted to previously untreated CLL patients
Rozman 1979	Study focus on staging system(s)
Salomon-Nguyen 1995	Diagnostic score
Santoro 1979	Study focus on staging system(s)
Sarmiento 2002	Prognostic factor study
Savvopoulos 2016	Methodological study

Study	Reason for exclusion
Scolozzi 1981	Study focus on staging system(s)
Shanafelt 2010	Prognostic factor study
Shanafelt 2017	Outcome not relevant for the review
Spacek 2009	Prognostic factor study
Stamatopoulos 2009	Prognostic score, however, this score was built with a proxy as outcome ('distinguished mutated and unmutated cases')
Stamatopoulos 2017	Prognostic factor study
Strefford 2015	Prognostic factor study
Szymczyk 2018	Prognostic factor study
Tallarico 2018	Outcome not relevant for the review
Tobin 2005a	Study focus on identification of the threshold for a prognostic factor
Tobin 2005b	Study focus on identification of the threshold for a prognostic factor
Vallat 2013	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Van Damme 2012	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Velardi 1980	Study focus on staging system(s)
Vojdeman 2017	Prognostic factor study
Vural 2014	Prognostic factor study
Weinberg 2007	Patient population did not match the review question; study was not restricted to previously untreated CLL patients
Weiss 2011	Prognostic factor study
Wierda 2003	Prognostic factor study
Winkler 2010	Prognostic factor study
Wu 2010	Study focus on staging system(s)
Zengin 1997	Study focus on staging system(s)
Zenz 2009	Prognostic factor study
Zucchetto 2006	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering

CLL: chronic lymphocytic leukaemia

CLL-IPI: chronic lymphocytic leukaemia International Prognostic Index

Characteristics of studies awaiting classification *[ordered by study ID]*

Antic 2011

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Baumann 2014

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Bulian 2011

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Bulian 2012

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Cavallini 2017

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Del Guidice 2005

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Friedrichs 2011

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Furundarena 1994

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Gentile 2009

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Haferlach 2010

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Jarque 1991

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Lee 1987

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Leotard 2000

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Letestu 2010

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Li 2017b

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Liang 2018

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Metze 2000

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Miao 2019

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Molica 1990

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Molica 2010

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Molica 2015

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Molica 2019

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Morabito 2011

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Pepper 2012

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Qin 2018

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Rozman 1982

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Rozman 1984

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Schweighofer 2011

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Stamatopoulos 2010

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Tsimberidou 2007

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Vetro 2018

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Visentin 2015

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7 .

Wierda 2009

Notes Details on study design, recruitment period, prognostic factors and more can be found in [Appendix 7](#).

Characteristics of ongoing studies [ordered by study ID]

NCT00275054

Study name	Rituximab, fludarabine, and cyclophosphamide or observation alone in treating patients with stage 0, stage I, or stage II chronic lymphocytic leukemia
Starting date	2005
Contact information	Michael Hallek, MD, Medizinische Universitätsklinik I at the University of Cologne
Notes	- inclusion reason: "Development of a new prognostic staging system" - no staging system or prognostic model published yet

NCT03436524

Study name	A prognostic tool for early stage CLL
Starting date	2018
Contact information	Davide Rossi, MD, PhD, Principal investigator, Oncology Institute of Southern Switzerland
Notes	- inclusion reason: "The study aims at developing a model for the prediction of time to first treatment in chronic lymphocytic leukemia patients presenting with asymptomatic early stage disease". - study is currently recruiting

APPENDICES

Appendix 1. MEDLINE (via Ovid) search strategy up to 19 September 2018

1	exp LEUKEMIA, LYMPHOCYTIC, CHRONIC, B-CELL/
2	((lymphocytic* or b-cell or b-lymphocytic or lymphoblastic or lymphatic) adj1 leuk?em\$ adj3 (chronic\$ or cronic\$ or chroniq\$ or well-differentia\$)).tw,kf,ot.
3	((lymphoplasmacytoid or lymphocytic) adj1 lymphom* adj3 (chronic\$ or cronic\$ or chroniq\$ or well-differentia\$)).tw,kf,ot.
4	((small cell\$ or small-cell\$) adj3 lymphom\$).tw,kf,ot.
5	(lymphom\$ adj2 lymphocyt\$).tw,kf,ot.
6	lymphoplasma?ytoid.tw,kf,ot.

100

(Continued)

7	(cII or b-cII or bcII).tw.
8	sII.tw.
9	or/1-8
10	(predict\$ or clinical\$ or outcome\$ or risk\$).mp.
11	validat\$.mp. or predict\$.ti. or rule\$.mp.
12	(predict\$ and (outcome\$ or risk\$ or models\$)).mp.
13	((history or variable\$ or criteria or scor\$ or characteristic\$ or finding\$ or factor\$) and (predict\$ or model\$ or decision\$ or identif\$ or prognos\$)).mp.
14	decision\$.mp. and ((model\$ or clinical\$).mp. or LOGISTIC MODELS/)
15	(prognostic and (history or variable\$ or criteria\$ or scor\$ or characteristic\$ or finding\$ or factor\$ or model\$)).mp.
16	or/10-15
17	validat\$.mp. or predict\$.ti. or DECISION SUPPORT TECHNIQUES/ or rule\$.mp. or PREDICTIVE VALUE OF TESTS/
18	(predict\$ and (clinical\$ or identif\$)).mp.
19	17 or 18
20	RISK ASSESSMENT/
21	(risk\$ adj scores\$).tw,kf,ot.
22	exp RISK FACTORS/
23	(risk\$ adj (score\$ and factor\$)).tw,kf,ot.
24	DECISION SUPPORT TECHNIQUES/
25	(decision\$ adj2 (techniqu\$ or model\$)).tw,kf,ot.
26	(decision\$ and support\$ and technique\$).tw,kf,ot.
27	(prediction\$ and rule\$ and clinical\$).tw,kf,ot.
28	(decision\$ adj2 (modeling\$ or aid\$ or analys\$ or technique\$)).tw,kf,ot.
29	or/20-28
30	ANIMALS/ not HUMANS/
31	9 and (16 or 19 or 29)
32	31 not 30

Appendix 2. MEDLINE (via Ovid) search strategy, 19 September 2018 to 24 June 2019, including study filter

1	exp LEUKEMIA, LYMPHOCYTIC, CHRONIC, B-CELL/
2	((lymphocytic* or b-cell or b-lymphocytic or lymphoblastic or lymphatic) adj1 leuk?em\$ adj3 (chronic\$ or cronic\$ or chroniq\$ or well-differentia\$)).tw,kf,ot.
3	((lymphoplasmacytoid or lymphocytic) adj1 lymphom* adj3 (chronic\$ or cronic\$ or chroniq\$ or well-differentia\$)).tw,kf,ot.
4	((small cell\$ or small-cell\$) adj3 lymphom\$).tw,kf,ot.
5	(lymphom\$ adj2 lymphocyt\$).tw,kf,ot.
6	lymphoplasma?ytoid.tw,kf,ot.
7	(cli or b-cli or bcli).tw.
8	sll.tw.
9	or/1-8
10	Validat\$.tw. or Predict\$.ti. or Rule\$.tw. or (Predict\$ and (Outcome\$ or Risk\$ or Model\$)).tw. or ((History or Variable\$ or Criteria or Scor\$ or Characteristic\$ or Finding\$ or Factor\$) and (Predict\$ or Model\$ or Decision\$ or Identif\$ or Prognos\$)).tw. or (Decision\$.tw. and ((Model\$ or Clinical\$).tw. or LOGISTIC MODELS/)) or (Prognostic and (History or Variable\$ or Criteria or Scor\$ or Characteristic\$ or Finding\$ or Factor\$ or Model\$)).tw. or ("Stratification" or "Discrimination" or "Discriminate" or c-statistic or "Area under the curve" or "AUC" or "Calibration" or "Indices" or "Algorithm" or "Multivariable").tw.
11	9 and 10
12	exp ANIMALS/ not HUMANS/
13	11 not 12
14	limit 13 to ed=20180919-20190624

key: exp # /: explode # MeSH subject heading, tw: text word, kf: keyword heading word, ot: original title, * or \$: truncation, ?: wildcard, adj#: adjacent within # number of words

search line #10: Geersing G-J, Bouwmeester W, Zuithoof P, Spijker R, Leeftink M, Moons KG, et al. (2012) Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. PLOS One 2012; 7(2): e32844, doi:10.1371/journal.pone.0032844

Appendix 3. Embase search strategy up to 24 June 2019

1	exp CHRONIC LYMPHATIC LEUKEMIA/
100	

(Continued)

2	((lymphocytic* or b-cell or b-lymphocytic or lymphoblastic or lymphatic) adj1 leuk?em* adj3 (chronic* or cronic* or chroniq* or well-differentia*)).tw,kw.
3	((lymphoplasmacytoid or lymphocytic) adj1 lymphom* adj3 (chronic* or cronic* or chroniq* or well-differentia*)).tw,kw.
4	((small cell* or small-cell*) adj3 lymphom*).tw,kw.
5	(lymphoma* adj2 lymphocyt*).tw,kw.
6	lymphoplasma?ytoid.tw,kw.
7	(cli or b-cli or bcli).tw.
8	sll.tw.
9	or/1-8
10	Validat*.tw. or Predict*.ti. or Rule*.tw. or (Predict* and (Outcome* or Risk* or Model*)).tw. or ((History or Variable* or Criteria or Scor* or Characteristic* or Finding* or Factor*) and (Predict* or Model* or Decision* or Identif* or Prognos*)).tw. or (Decision*.tw. and ((Model* or Clinical*).tw. or STATISTICAL MODEL/)) or (Prognostic and (History or Variable* or Criteria or Scor* or Characteristic* or Finding* or Factor* or Model*)).tw. or ("Stratification" or "Discrimination" or "Discriminate" or c-statistic or "Area under the curve" or "AUC" or "Calibration" or "Indices" or "Algorithm" or "Multivariable").tw.
11	9 and 10
12	exp ANIMAL/ not HUMAN/
13	11 not 12
14	limit 13 to embase
15	limit 14 to em=197401-201926

key: exp # / : explode # MeSH subject heading, tw: text word, kw: keyword, *: truncation, ?: wildcard, adj#: adjacent within # number of words
searchline #10: Geersing G-J, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons KG, et al. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. PLOS One 2012; 7(2): e32844, doi:10.1371/journal.pone.0032844 (translated in Embase)

Appendix 4. ClinicalTrials.gov search strategy up to 05 March 2020

Basic search

prognostic OR predictive OR model OR score | chronic lymphocytic leukemia OR CLL

Advanced search

Intervention/treatment: prognostic OR predictive OR Model OR score

Recruitment: all studies

Study type: all studies

Appendix 5. WHO ICTRP search strategy

Basic search

Chronic lymphocytic leukemia AND prognostic

Chronic lymphocytic leukemia AND predictive

Chronic lymphocytic leukemia AND model

Chronic lymphocytic leukemia AND score

CLL AND prognostic

CLL AND predictive

CLL AND model

CLL AND score

Advanced search

Condition: chronic lymphocytic leukemia OR CLL

Intervention: prognostic OR predictive Or model OR score

Recruitment status: ALL



Appendix 6. Prognostic models with external validation studies

Author, year	Model (= 1) or Score (= 2)	Outcome	Recruitment period	Number of patients	Study design	Number of predictors included in final model	Discrimination	Calibration
Delgado 2017	2	OS	NR	524	Retrospective cohort	IgHV; FISH abnormalities (2)	0.682	Calibration plot for survival at 5 years
International CLL IPI work-group	2	OS	1997 to 2007	3472	RCTs	IgHV; B2M; age; stage; TP53 (5)	0.723	NR
	2	OS	NR	262	Unclear	IgHV; ZAP-70; CD38 (3)	NR	NR
	2	OS	1997 to 2006	1223	RCTs	IgHV; B2M; age; gender; del17p; ECOG; sTK; del11q (8)	0.75	NR
Wierda 2007	1	OS	1981 to 2004	1674	Retrospective cohort	IgHV; B2M; age; stage; gender; ALC (6)	0.84	Calibration plot for survival at 5 years
Stephens 2015	2	OS	2002 to 2012	114	Retrospective cohort	age; LDH; ECOG (3)	0.73	NR
Molica 2005	2	PFS	1991 to 2000	1138	Retrospective cohort	stage; gender; lymphocytosis; LDT (4)	NR	NR
Baliakas 2019	2	TFS	NR	1900	Retrospective cohort	IgHV; TP53; FISH abnormalities; gender (4)	0.745	NR
Gentile 2016	2	TFS	2007 to NR	480	Prospective cohort	IgHV; B2M; stage; ALC (4)	0.75	Hosmer-May test

(Continued)

Rossi 2013	2	TFS	NR	673	Retrospective cohort	TP53; FISH abnormalities; SF3B1 (3)	0.642	Bias-corrected calibration slope, 0.965
Wierda 2011	1	TFS	2004 to 2009	930	Retrospective cohort	IgHV; LDH; number of involved lymph nodes; FISH abnormalities; lymph node size in neck; interaction term (6)	NR	NR
Stephens 2015	2	TFS	2002 to 2012	114	Retrospective cohort	age; stage; ECOG; del11q; WBC count (5)	0.84	NR

Abbreviations: OS: overall survival; PFS: progression-free survival; TFS: treatment-free survival; NR: not reported; RCT: randomised controlled trial; IgHV: immunoglobulin heavy chain variable region gene mutational status; B2M: beta 2 microglobulin; FISH: fluorescence in situ hybridisation genomic aberrations; Zap-70: zeta-chain-associated protein kinase 70; CD38: cluster of differentiation 38; LDH: lactate dehydrogenase; LPL: lipoprotein lipase; ECOG: Eastern Cooperative Oncology Group performance status

Appendix 7. Prognostic models without external validation studies

Author, year	Model (=1) or Score (=2)	Outcome	Recruit-ment peri-od	Number of patients	Study design	Predictors included in final model (number)	Discrimina-tion	Calibration
Baumann 2014	2	OS	1990 to 2012	949	Retrospective cohort	B2M; stage; Zap-70; Comorbirities (4)	NR	NR
Bulian 2012	1	OS	1996 to 2008	620	Retrospective cohort	IgHV; B2M; age stage; FISH; gender (6)	0.79	NR
Bulian 2011	1	OS	1996 to NR	1480	Retrospective cohort	B2M; age; stage; gender (4)	0.78	Calibration plot for survival at 5 years
Furundarena	1	OS	1973 to 1992	150	Retrospective cohort	age; stage; gender; splenomegaly (4)	NR	NR
Hafenlach 2010	2	OS	2005 to 2008	399	NR	IgHV; age; TP53; translocation IGH@ on 14q32; number of chromosome aberrations based on CBA; WBC count (6)	NR	NR
Jarque 1991, model A	1	OS	1969 to 1988	187	Retrospective cohort	Age; Spinal infiltration; BUN (3)	NR	NR
Jarque 1991, model B	1	OS	1969 to 1988	187	Retrospective cohort	Albumin; Spinal infiltration; BUN; cervical adenopathies (4)	NR	NR

(Continued)

Lee 1987	1, 2	OS	1970 to 1983	325	Retrospective cohort	Age; LDH; uric acid; alkaline phosphatase; external lymphadenopathy (5)	NR	NR
Liang 2018, model 3a	2	OS	2000 to 2014	501	Retrospective cohort	IgHV; B2M; stage; TP53; albumin; HBV infection (6)	time-dependent ROC curve	NR
Molica 1990	1	OS	NR	221	Unclear	age; stage; LAR (3)	NR	O/E ratios
Molica 2019	1, 2	OS	2000 to NR	108	Retrospective cohort	age; ADL; CIRS (3)	c = 0.70 (95% CI 0.53 to 0.87)	NR
Pepper 2012	1	OS	NR	1154	Unclear	IgHV; age; CD38; LDT (4)	NR	NR
Rozman 1982	1	OS	NR	150	Unclear	Splenomegaly; lymphocytosis; anaemia; thrombocytopenia (4)	NR	O/E probabilities
Rozman 1984	1	OS	NR	329	Unclear	lymphadenopathy; haemoglobin; bone marrow pattern; hepatomegaly (4)	NR	O/E ratios
Stamatopoulos 2010	2	OS	NR	170	Unclear	ZAP-70; LPL; miR-29 (3)	NR	NR
Tsimberidou 2007	2	OS	1985 to 2005	1893	Retrospective cohort	Age; B2M; Del17p; albumin; creatinine (5)	NR	NR
Visentin 2015	2	OS	1983 to 2013	608	Retrospective cohort	IgHV; CD38; FISH abnormalities (3)	0.88	NR

(Continued)

Wierda 2009	1, 2	OS	1985 to 2004	595	Unclear	Age; B2M; treatment (3)	NR	Concordance index for calibration curve (nomogram): 0.81
Friedrichs 2011	2	PFS	NR	134	NR	IgHV; CD38; iLR (3)	NR	NR
Leotard 2000	1	PFS	1985 to 1997	88	Prospective cohort	B2M; LDH; albumin; sCD23 (4)	NR	NR
Letestu 2010	1, 2	PFS	NR	339	NR	B2M; CD38; lymphocytosis; sTK (4)	NR	NR
Antic 2011	2	TFS	NR	33	Unclear	B2M; LPL; sVEGF (3)	NR	NR
Cavallini 2017	2	TFS	NR	125	NR	SF3B1; ERK1/2 phosphorylation (2)	NR	NR
Del Guidice 2005	2	TFS	2003 to 2004	201	Prospective cohort	ZAP-70; CD38 (2)	NR	NR
Gentile 2009	2	TFS	NR	222	NR	IgHV; B2M; CD38 (3)	NR	NR
Haferlach 2010	2	TTT	2005 to 2008	399	NR	IgHV; ATM deletion; translocation involving IGH@ locus on 14q32; number of chromosome abnormalities (4)	NR	NR
Li 2017b	2	TFS	2007 to 2015	406	NR	IgHV; stage; Del17p (3)	NR	NR

(Continued)

Liang 2018 model 3b	2	TFS	2000 to 2014	501	Retrospective cohort	IgHV; LD; lymphocytosis; platelets; HBV (5)	time-dependent ROC curve	NR
Metze 2000	2	TFS	1995 to 1998	57	Prospective cohort	total tumor mass score; percentage of cells with 1 AgNOR cluster (2)	NR	NR
Miao 2019, model A (CLL-IPI-S)	1, 2	TFS	2000 to 2017	399	Retrospective cohort	IgHV; B2M; Age; stage; TP53; SF3B1 (6)	AUC = 0.762	calibration plot
Miao 2019, model B (CLL-PI)	1, 2	TFS	2000 to 2017	399	Retrospective cohort	IgHV; B2M; stage; TP53; SF3B1 (5)	AUC = 0.773	calibration plot
Molica 2010	2	TFS	1998 to 2008	150	Retrospective cohort	IgHV; BAFF (2)	0.86	NR
Molica 2015	1	TFS	NR	322	Unclear	IgHV; B-cell count; interaction term (3)	NR	NR
Morabito 2011	2	TFS	NR	449	Retrospective cohort	stage; ZAP-70; FISH abnormalities; SFLC (kappa + delta) (4)	NR	NR
Pepper 2012	1	TFS	NR	1154	Unclear	IgHV; age; LDT; CD38 (4)	NR	NR
Qin 2018	2	TTT	2008 to 2016	334	Retrospective cohort	IgHV; EBV DNA positive; stage; TP53; ALC; HBsAg+; (6)	AUC = 0.768	NR
Schweighofer 2011	1	TFS	NR	131	Retrospective cohort	SKI (gene); SLAMF1 (gene) (2)	NR	NR

(Continued)

Stamatopoulos 2010	2	TFS	NR	170	Unclear	ZAP-70; LPL; miR-29c (3)	NR	NR
Vetro 2018	1	TTT	NR	171	Unclear	IgHV; SF3B1; percentage of B cells; genomic aberrations (4)	NR	NR
Wierda 2009	1	TFS	1985 to 2004	595	Unclear	Age; B2M; treatment; bone marrow lymphocytes (4)	NR	NR

Abbreviations: OS: overall survival; PFS: progression-free survival; TFS: treatment-free survival; TTT: time to treatment; NR: not reported; RCT: randomised controlled trial; IgHV: immunoglobulin heavy chain variable region gene mutational status; B2M: beta 2 microglobulin; FISH: fluorescence in situ hybridisation genomic aberrations; Zap-70: zeta-chain-associated protein kinase 70; CD38: cluster of differentiation 38; LDH: lactate dehydrogenase; LPL: lipoprotein lipase; ECOG: Eastern Cooperative Oncology Group performance status; LAR: lymphocyte accumulation time; sVEGF: soluble vascular endothelial growth factor; iLR: immature laminin receptor; CIRS: Cumulative Illness Rating Score; ADL: Katz Activity of Daily Living; EBV: Epstein-Barr virus; HBsAg: surface antigen of hepatitis B virus

Appendix 8. Reporting deficiencies in all included studies with at least one external validation study (development + validation studies, listed per model)

Prognostic model	Number of cohorts	Recruitment period reported in:	Study design reported in:	Observation time (range) reported in:	Total sample size reported in:	Number of events reported in:	Calibration reported in:	Discrimination (95% CI or SE) reported in:
Baliakas (Baliakas 2019)	2	0	2	1 (1)	2	0	0	1 (1)
Barcelona-Brno (Delgado 2017)	6	3	4	6 (6)	6	5	1	5 (3)
CLL-IPI (Bahlo 2016)	11	9	9	10 (6)	10	9	5	8 (6)
GCLLSG (Pflug 2014)	2	1	2	2 (0)	2	2	2	2 (2)
GIMEMA (Molica 2005)	2	2	2	1 (1)	2	1	0	1 (1)
MDACC 2007 (Wierda 2007)	10	7	8	8 (9)	9	9	2	8 (7)
MDACC 2011 (Wierda 2011)	2	1	2	2 (2)	2	2	0	1 (1)
Morabito model (Morabito, 2009)	2	2	1	2 (2)	2	2	0	0 (0)
O-CLL1 model (Gentile 2016)	3	1	1	3 (2)	3	3	0	3 (2)
Rossi (Rossi 2013)	3	3	3	3 (0)	3	2	0	2 (0)
Stephens (Stephens 2015)	2	1	1	0 (0)	2	0	0	2 (0)
Stephens (Stephens 2015)	2	1	1	0 (0)	2	0	0	2 (0)

HISTORY

Protocol first published: Issue 1, 2016

Review first published: Issue 7, 2020

CONTRIBUTIONS OF AUTHORS

Nina Kreuzberger: screening and selection of studies, development of data extraction form and data extraction, characteristics of studies, 'Risk of bias' assessment, statistical analysis, writing and drafting of the review, communication with and between authors

Johanna AAG Damen: statistical analysis

Marialena Trivella: statistical input, screening and selection of studies, data extraction, 'Risk of bias' assessment

Lise J Estcourt: medical and content input, screening and selection of studies, data extraction, 'Risk of bias' assessment

Angela Aldin: screening and selection of studies, data extraction, 'Risk of bias' assessment

Lisa Umlauff: data extraction, 'Risk of bias' assessment, characteristics of included studies

Maria Vazquez Montes: prototype of data extraction form, support in data extraction, 'Risk of bias' assessment

Robert Wolff: 'Risk of bias' input

Karel GM Moons: methodological input on reviews of prognosis studies

Ina Monsef: search strategy development

Farid Foroutan: 'Risk of bias' input

Karl-Anton Kreuzer: medical and content input

Nicole Skoetz: protocol development, screening and selection of studies, data extraction, 'Risk of bias' assessment, extensive proofread and comments on the review draft

DECLARATIONS OF INTEREST

Nina Kreuzberger: My institution received a grant from the Federal Ministry of Education and Research, Germany to conduct this review.

Johanna AAG Damen: none known

Marialena Trivella: I am working as a statistical editor in a number Cochrane groups, and I declare that my work as a statistical editor is independent to this published work where I participate. The University of Oxford received a small part of the grant from the Federal Ministry of Education and Research, Germany, as reimbursement for my time spent on the project.

Lise J Estcourt: My institution received a grant from the Federal Ministry of Education and Research, Germany to conduct this review.

Angela Aldin: My institution received a grant from the Federal Ministry of Education and Research, Germany to conduct this review.

Lisa Umlauff: My institution received a grant from the Federal Ministry of Education and Research, Germany to conduct this review.

Maria Vazquez Montes: I am sponsored by the BHF to contribute on a similar prognostic models review for heart failure which allowed me to effectively participate in the work under consideration.

Robert Wolff: As employee of Kleijnen Systematic Reviews, I was the lead author of a systematic review on prostate cancer, commissioned by Elekta, Nucletron.

Karel Moons: none known

Ina Monsef: none known

Farid Foroutan: none known

Karl-Anton Kreuzer: board member and consultant for AbbVie, Alexion, Amgen, Ariad, Baxter, Bayer Health Care, Biotest, Boehringer Ingelheim, Bristol Myers Squibb, Celgene, Chugai, Gilead, Glaxo-SmithKline, Grifols, Hexal, Janssen, Jazz Pharmaceuticals, Leo, Mundipharma, MSD, Novartis, Pfizer, Roche, Shire, Teva. Grants, fees, honoraria and travel grants from AbbVie, Alexion, Amgen, Ariad,

Baxter, Bayer Health Care, Biotest, Boehringer Ingelheim, Bristol Myers Squibb, Celgene, Chugai, Gilead, Glaxo-SmithKline, Grifols, Hexal, Janssen, Jazz Pharmaceuticals, Leo, Mundipharma, MSD, Novartis, Pfizer, Roche, Shire, Teva.

Nicole Skoetz: My institution received a grant from the Federal Ministry of Education and Research, Germany to conduct this review.

SOURCES OF SUPPORT

Internal sources

- University Hospital of Cologne, Department I of Internal Medicine, Germany
- NHS Blood and Transplant, UK

External sources

- Grant by the Federal Ministry of Education and Research (Grant no. 01KG1711), Germany

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

We changed the title of the review to clarify the included population (newly-diagnosed adults with CLL). In the protocol, the title was 'Prognostic models for chronic lymphocytic leukaemia: an exemplar systematic review and meta-analysis'.

Structure of the review

As prognosis reviews are a new review type within Cochrane, little guidance had been published at protocol stage. Recently, a review template has been developed and published by the Cochrane Prognosis Methods group, which we have adopted for this review (methods.cochrane.org/prognosis/our-publications). In addition, we have added two sections ('Selection of studies' and 'Risk of bias' assessment') to report our group decisions transparently.

Search methods

We did not search PubMed, because content from PubMed can be identified via MEDLINE.

We added a search for Embase based on editorial comments.

We added a search of the following databases for ongoing trials instead of the metaRegister of Controlled Trials:

- ClinicalTrials.gov
- World Health Organization International Clinical Trials Registry Platform (WHO ICTRP)

We did not search conference proceedings for abstracts, because the paucity of information would not allow us to apply the 'Risk of bias' tool. To ensure that we identified all validation studies corresponding to a developed model, we screened the citations of all included studies (Web of Science, October 2018).

Inclusion and exclusion of studies

Originally, we planned to include only models published after 1990. However, prognostic model studies are often conducted based on retrospective data, which sometimes include the analysis of blood samples frozen at diagnosis and analysed many years after. To give the same chance to studies with samples that were analysed immediately and after several years for inclusion in this review, we decided not to limit our search strategy based on date of publication. As a clarification, any studies that explicitly aimed at defining a new staging system was excluded during title and abstract screening.

We decided to include prognostic models with the outcomes 'time-to-first treatment' and 'treatment-free survival' as additional outcomes, because individuals with low-risk CLL can often live a long time without disease progression and the need for treatment. Usually, at disease progression, treatment is indicated, which means that these two outcomes are quite similar and both are meaningful for the patient.

After several rounds of discussion, we decided that genetic signatures are beyond the scope of this review because: a) we prespecified that we would exclude factor identification studies, which most of the genetic signature studies are to start with; b) they are not yet ready to be applied in clinical practice due to the unavailability of detailed genetic information that these signatures use and the complexity of the algorithms; and c) more often than not, it is not feasible to assess a genetic signature study in the same context as a prognostic model.

Analysis and 'Risk of bias' assessment

Instead of the CHARMS checklist, we used the recently published PROBAST-tool to assess the risk of bias of individual studies ([Wolff 2019](#)). The checklist does not originally aim at the assessment of risk of bias, but as a guide for critical appraisal and data extraction. However, at the protocol stage, no tool specific to 'Risk of bias' assessment for prognostic models had been published.

The analysis was not specified precisely in the protocol, as methods for systematic reviews of prognosis are just developing. As described in the protocol, we decided to follow the current recommendations of the Cochrane Prognosis Methods Group. We added a more detailed description in the section Data collection and analysis. In short, we meta-analysed the performance measures of the various external validation studies per model where data were available. We planned to summarise the measures of calibration. Unfortunately, calibration in the form of O:E ratios or calibration plots was rarely reported. Instead, many studies only presented the observed outcome frequency per subgroup at a specific time point. To gain an overview of this information among several external validations, during the review process and in collaboration with the Cochrane Prognosis Methods Group, we decided to represent the outcome frequency graphically together with the pooled outcome frequency in a table.

As we only meta-analysed the external validation studies of a prognostic model, we analysed only models that were externally validated several times. Although we have included all prognostic models and scores that we identified, we did not describe models without any external validation studies in detail, because it is not recommended to use prognostic scores or models without any testing in independent cohorts, especially when the development sample was small (Moons 2015; Steyerberg 2013). References to these studies can be found in the list of studies awaiting classification; further information can be found in Appendix 7.

In the protocol, we did not specify any sensitivity analyses. We decided post hoc to explore the effect of including area under the curve (AUC) as a performance measure for discrimination instead of the c-statistic (Figure 23), the effect of the Newcombe estimation method for 95% CIs of the c-statistic (Figure 24; Figure 25), and the difference between studies using the original predictor or a proxy (del(17p) instead of TP53 mutation (Figure 26).

GRADE

We did not apply GRADE as no GRADE guidance has yet been developed to assess certainty of evidence from meta-analysis of prognostic models.

NOTES

Parts of this review, especially the methods, are from the Cochrane Haematology standard template.

INDEX TERMS

Medical Subject Headings (MeSH)

Age Factors; Bias; Biomarkers, Tumor; Calibration; Confidence Intervals; Discriminant Analysis; Disease-Free Survival; Genes, p53 [genetics]; Immunoglobulin Heavy Chains [genetics]; Immunoglobulin Variable Region [genetics]; Leukemia, Lymphocytic, Chronic, B-Cell [*mortality] [pathology]; *Models, Theoretical; Neoplasm Staging; Prognosis; Progression-Free Survival; Receptors, Antigen, B-Cell [genetics]; Reproducibility of Results; Tumor Suppressor Protein p53 [genetics]

MeSH check words

Adult; Female; Humans; Male

Appendix IV: Publication 3



Positron emission tomography-adapted therapy for first-line treatment in adults with Hodgkin lymphoma (Review)

Kreuzberger N, Goldkuhle M, von Tresckow B, Kobe C, Sickinger MT, Monsef I, Skoetz N

Kreuzberger N, Goldkuhle M, von Tresckow B, Kobe C, Sickinger M-T, Monsef I, Skoetz N.
Positron emission tomography-adapted therapy for first-line treatment in adults with Hodgkin lymphoma.
Cochrane Database of Systematic Reviews 2025, Issue 3. Art. No.: CD010533.
DOI: [10.1002/14651858.CD010533.pub3](https://doi.org/10.1002/14651858.CD010533.pub3).

www.cochranelibrary.com

TABLE OF CONTENTS

ABSTRACT	1
PLAIN LANGUAGE SUMMARY	3
SUMMARY OF FINDINGS	5
BACKGROUND	14
OBJECTIVES	14
METHODS	15
Figure 1.	17
RESULTS	20
Figure 2.	22
DISCUSSION Figure 2.	27
AUTHORS' CONCLUSIONS	29
ACKNOWLEDGEMENTS	30
REFERENCES	31
CHARACTERISTICS OF STUDIES	46
DATA AND ANALYSES	82
Analysis 1.1. Comparison 1: PET-adapted treatment adaptation in PET-negative early stage HL (omission of radiotherapy, with or without more chemotherapy), Outcome 1: Overall survival (OS)	83
Analysis 1.2. Comparison 1: PET-adapted treatment adaptation in PET-negative early stage HL (omission of radiotherapy, with or without more chemotherapy), Outcome 2: Progression-free survival (PFS)	83
Analysis 1.3. Comparison 1: PET-adapted treatment adaptation in PET-negative early stage HL (omission of radiotherapy, with or without more chemotherapy), Outcome 3: Treatment-related mortality	84
Analysis 1.4. Comparison 1: PET-adapted treatment adaptation in PET-negative early stage HL (omission of radiotherapy, with or without more chemotherapy), Outcome 4: Incidence of second primary malignancy	84
Analysis 2.1. Comparison 2: PET-adapted de-escalation in PET-negative, mixed early-and-intermediate-stage PET-negative HL (omission of radiotherapy), Outcome 1: Overall survival (OS)	85
Analysis 2.2. Comparison 2: PET-adapted de-escalation in PET-negative, mixed early-and-intermediate-stage PET-negative HL (omission of radiotherapy), Outcome 2: Progression-free survival (PFS)	85
Analysis 3.1. Comparison 3: PET-adapted de-escalation in PET-negative intermediate-stage HL (ITT, except where noted), Outcome 1: Overall survival (OS)	86
Analysis 3.2. Comparison 3: PET-adapted de-escalation in PET-negative intermediate-stage HL (ITT, except where noted), Outcome 2: Progression-free survival (PFS)	87
Analysis 3.3. Comparison 3: PET-adapted de-escalation in PET-negative intermediate-stage HL (ITT, except where noted), Outcome 3: Treatment-related mortality	87
Analysis 3.4. Comparison 3: PET-adapted de-escalation in PET-negative intermediate-stage HL (ITT, except where noted), Outcome 4: Serious adverse events	88
Analysis 3.5. Comparison 3: PET-adapted de-escalation in PET-negative intermediate-stage HL (ITT, except where noted), Outcome 5: Incidence of second primary malignancy	88
Analysis 4.1. Comparison 4: PET-adapted de-escalation in PET-negative, advanced stage HL (ITT) - de-escalation of chemotherapy, Outcome 1: Overall survival	89
Analysis 4.2. Comparison 4: PET-adapted de-escalation in PET-negative, advanced stage HL (ITT) - de-escalation of chemotherapy, Outcome 2: Progression-free survival	90
Analysis 4.3. Comparison 4: PET-adapted de-escalation in PET-negative, advanced stage HL (ITT) - de-escalation of chemotherapy, Outcome 3: Treatment-related mortality	90
Analysis 4.4. Comparison 4: PET-adapted de-escalation in PET-negative, advanced stage HL (ITT) - de-escalation of chemotherapy, Outcome 4: Adverse events, grade 3-4	91
Analysis 4.5. Comparison 4: PET-adapted de-escalation in PET-negative, advanced stage HL (ITT) - de-escalation of chemotherapy, Outcome 5: Incidence of second primary malignancy	91
Analysis 5.1. Comparison 5: PET-adapted de-escalation in PET-negative, advanced stage HL (ITT) - RT vs. no further RT, Outcome 1: Overall survival	92
Analysis 5.2. Comparison 5: PET-adapted de-escalation in PET-negative, advanced stage HL (ITT) - RT vs. no further RT, Outcome 2: Progression-free survival	93
Analysis 5.3. Comparison 5: PET-adapted de-escalation in PET-negative, advanced stage HL (ITT) - RT vs. no further RT, Outcome 3: Incidence of second primary malignancy	93

Analysis 6.1. Comparison 6: PET-adapted escalation in PET-positive, mixed early-and-intermediate-stage HL (ITT, except where noted), Outcome 1: Overall survival	94
Analysis 6.2. Comparison 6: PET-adapted escalation in PET-positive, mixed early-and-intermediate-stage HL (ITT, except where noted), Outcome 2: Progression-free survival	94
Analysis 6.3. Comparison 6: PET-adapted escalation in PET-positive, mixed early-and-intermediate-stage HL (ITT, except where noted), Outcome 3: Incidence of second primary malignancy	95
Analysis 7.1. Comparison 7: PET-adapted escalation in PET-positive, advanced-stage HL (ITT, except where noted), Outcome 1: Overall survival	96
Analysis 7.2. Comparison 7: PET-adapted escalation in PET-positive, advanced-stage HL (ITT, except where noted), Outcome 2: Progression-free survival	96
Analysis 7.3. Comparison 7: PET-adapted escalation in PET-positive, advanced-stage HL (ITT, except where noted), Outcome 3: Treatment-related mortality	97
Analysis 7.4. Comparison 7: PET-adapted escalation in PET-positive, advanced-stage HL (ITT, except where noted), Outcome 4: Serious adverse events	97
Analysis 7.5. Comparison 7: PET-adapted escalation in PET-positive, advanced-stage HL (ITT, except where noted), Outcome 5: Adverse events, grade 3-4	98
Analysis 7.6. Comparison 7: PET-adapted escalation in PET-positive, advanced-stage HL (ITT, except where noted), Outcome 6: Any grade adverse event	98
Analysis 7.7. Comparison 7: PET-adapted escalation in PET-positive, advanced-stage HL (ITT, except where noted), Outcome 7: Incidence of secondary malignancy	99
Analysis 8.1. Comparison 8: Subgroup analyses: PET-adapted de-escalation in PET-negative early-stage HL (additional chemotherapy), Outcome 1: Overall survival	100
Analysis 8.2. Comparison 8: Subgroup analyses: PET-adapted de-escalation in PET-negative early-stage HL (additional chemotherapy), Outcome 2: Progression-free survival	100
Analysis 8.3. Comparison 8: Subgroup analyses: PET-adapted de-escalation in PET-negative early-stage HL (additional chemotherapy), Outcome 3: Incidence of second primary malignancy	101
Analysis 9.1. Comparison 9: Subgroup analysis: PET-adapted de-escalation in PET-negative early-stage HL (NLPHL vs. cHL), Outcome 1: Progression-free survival	102
Analysis 10.1. Comparison 10: Subgroup analyses: PET-adapted de-escalation in PET-negative intermediate stage HL (additional chemotherapy), Outcome 1: Overall survival	103
Analysis 10.2. Comparison 10: Subgroup analyses: PET-adapted de-escalation in PET-negative intermediate stage HL (additional chemotherapy), Outcome 2: Progression-free survival	104
Analysis 10.3. Comparison 10: Subgroup analyses: PET-adapted de-escalation in PET-negative intermediate stage HL (additional chemotherapy), Outcome 3: Treatment-related mortality	105
Analysis 10.4. Comparison 10: Subgroup analyses: PET-adapted de-escalation in PET-negative intermediate stage HL (additional chemotherapy), Outcome 4: Incidence of second primary malignancy	106
Analysis 11.1. Comparison 11: Subgroup analyses: PET-adapted de-escalation in PET-negative advanced stage HL (initial chemotherapy regimen; ITT), Outcome 1: Overall survival	107
Analysis 11.2. Comparison 11: Subgroup analyses: PET-adapted de-escalation in PET-negative advanced stage HL (initial chemotherapy regimen; ITT), Outcome 2: Progression-free survival	108
Analysis 11.3. Comparison 11: Subgroup analyses: PET-adapted de-escalation in PET-negative advanced stage HL (initial chemotherapy regimen; ITT), Outcome 3: Treatment-related mortality	109
Analysis 11.4. Comparison 11: Subgroup analyses: PET-adapted de-escalation in PET-negative advanced stage HL (initial chemotherapy regimen; ITT), Outcome 4: Adverse events, grade 3-4	110
Analysis 11.5. Comparison 11: Subgroup analyses: PET-adapted de-escalation in PET-negative advanced stage HL (initial chemotherapy regimen; ITT), Outcome 5: Incidence of second primary malignancy	111
Analysis 12.1. Comparison 12: Subgroup analyses: PET-adapted de-escalation in PET-negative advanced stage HL (NLPHL vs. cHL), Outcome 1: progression-free survival	112
ADDITIONAL TABLES	113
APPENDICES	131
WHAT'S NEW.....	136
HISTORY.....	138
CONTRIBUTIONS OF AUTHORS.....	138
DECLARATIONS OF INTEREST	138
SOURCES OF SUPPORT.....	139
DIFFERENCES BETWEEN PROTOCOL AND REVIEW	139

NOTES	140
INDEX TERMS.....	140

[Intervention Review]

Positron emission tomography-adapted therapy for first-line treatment in adults with Hodgkin lymphoma

Nina Kreuzberger^{1,2}, Marius Goldkuhle^{1,2}, Bastian von Tresckow^{3,4}, Carsten Kobe⁵, Marie-Therese Sickinger⁶, Ina Monsef^{1,2}, Nicole Skoetz^{1,2}

¹Cochrane Haematology, Institute of Public Health, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. ²Cochrane Evidence Synthesis Unit (CESU) Germany/UK, Institute of Public Health, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. ³Department I of Internal Medicine, Center for Integrated Oncology Aachen Bonn Cologne Duesseldorf, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. ⁴Department of Hematology and Stem Cell Transplantation, West German Cancer Center, University Hospital Essen, University of Duisburg-Essen, Essen, Germany. ⁵Department for Nuclear Medicine, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. ⁶Clinical Department of Pediatrics and Adolescent Medicine, University Hospital St. Pölten, Sankt Pölten, Austria

Contact: Nina Kreuzberger, nina.kreuzberger@uk-koeln.de.

Editorial group: Cochrane Central Editorial Service.

Publication status and date: New search for studies and content updated (conclusions changed), published in Issue 3, 2025.

Citation: Kreuzberger N, Goldkuhle M, von Tresckow B, Kobe C, Sickinger M-T, Monsef I, Skoetz N. Positron emission tomography-adapted therapy for first-line treatment in adults with Hodgkin lymphoma. *Cochrane Database of Systematic Reviews* 2025, Issue 3. Art. No.: CD010533. DOI: [10.1002/14651858.CD010533.pub3](https://doi.org/10.1002/14651858.CD010533.pub3).

Copyright © 2025 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

ABSTRACT

Background

Hodgkin lymphoma (HL) is one of the most curable cancers worldwide. Treatment options comprise more- or less-intensified regimens of chemotherapy plus radiotherapy depending on the disease stage. An interim-[¹⁸F]-fluorodeoxy-D-glucose (FDG)-positron emission tomography (PET), a procedure to illustrate a tumour's metabolic activity, stage and progression, could be used during treatment to distinguish between individuals who are good or poor early responders to therapy. Subsequent therapy could be de-escalated in PET-negative individuals (good responders) or escalated in those who are PET-positive (poor responders).

Objectives

To assess the effects of interim [¹⁸F]-FDG-PET-imaging treatment modification in previously untreated individuals with HL.

Search methods

For this review update, we searched MEDLINE, the Cochrane Central Register of Controlled Trials (CENTRAL), Embase, clinicaltrials.gov and the WHO ICTRP up to 17 November 2023.

Selection criteria

We included randomised controlled trials (RCTs) comparing interim-FDG-PET-adapted therapy with non-adapted standard treatment in adults with untreated HL of all stages.

Data collection and analysis

Two review authors independently screened results for inclusion, extracted data into a standardised data extraction sheet and assessed the risk of bias according to the Cochrane risk of bias tool. We collected (modified) intention-to-treat effect estimates for the predefined outcomes: overall survival (OS), progression-free survival (PFS), treatment-related mortality (TRM), adverse events (AE) including secondary malignancies and quality of life (QoL), where available, and used random-effects models for meta-analysis. We analysed early,

intermediate and advanced stage HL and PET-negative versus PET-positive participants separately. We used the GRADE approach to rate our certainty in the evidence.

Main results

We included 10 studies covering early (1 RCT, 667 participants), intermediate (1 RCT, 651 participants), early-to-intermediate (3 RCTs, 1639 participants) and advanced-stage HL (5 RCTs; 3629 participants). We did not identify eligible ongoing studies.

Generally, the risk of bias was low or, sometimes, unclear except for detection bias, which was rated as high for all studies for subjective outcomes such as PFS, TRM and AE due to the lack of blinding.

PET-based adaptation in early-stage PET-negative participants

The effect of treatment adaptation (omission of radiotherapy with or without additional chemotherapy) on OS and PFS is uncertain (HR 0.84, 95% CI 0.13 to 5.32; and HR 4.52, 95% CI 0.72 to 28.41, 1034 participants). Adaptation may have little to no effect on the incidence of secondary malignancies (RR 0.83, 95% CI 0.46 to 1.50; 984 participants; low-certainty). No studies reported on TRM, serious adverse events (SAE) or QoL.

PET-based adaptation in intermediate-stage PET-negative participants

Treatment adaptation by omission of radiotherapy (with or without additional chemotherapy) may have little effect on OS (HR 0.91, 95% CI 0.42 to 1.96; 1073 participants; low-certainty) and PFS (HR 1.59, 95% CI 0.95 to 2.67; 1073 participants; low-certainty) compared to standard therapy. The effect on TRM is very uncertain. De-escalation may have little effect on the incidence of SAE (RR 1.01, 95% CI 0.84 to 1.21; 1096 participants, low-certainty) and secondary malignancies (RR 1.01, 95% CI 0.57 to 1.82; 1515 participants; low-certainty). No studies reported on QoL.

PET-based de-escalation in advanced-stage PET-negative participants

Three RCTs examined interim-PET-based de-intensification of chemotherapy compared with standard in advanced-stage PET-negative participants; this probably increases OS (HR 0.65, 95% CI 0.40 to 1.07; 2633 participants, moderate-certainty), although the confidence interval included the possibility of no effect, while it has probably little effect on PFS (HR 0.98, 95% CI 0.78 to 1.25; 2633 participants, moderate-certainty). Treatment de-escalation may reduce TRM (RR 0.21, 95% CI 0.06 to 0.73; 2761 participants, low-certainty) and the incidence of secondary malignancy (RR 0.87, 95% CI 0.60 to 1.26; 2757 participants; low-certainty), although for this latter finding, the CI included the possibility of no effect. No studies reported SAE and QoL.

Two RCTs considered combined modality treatment as standard for advanced stages and de-escalated by omitting radiotherapy. De-escalation may increase OS (HR 0.63, 95% CI 0.11 to 3.69; 296 participants; low-certainty), PFS (HR 0.78, 95% CI 0.43 to 1.43; 412 participants; low-certainty), and may reduce the incidence of secondary malignancy (RR 0.41, 95% CI 0.08 to 2.09; 349 participants; low-certainty), although for all these findings, the CI included the possibility of no effect. No studies reported TRM, SAE and QoL.

PET-based escalation in mixed early-and-intermediate-stage, PET-positive participants

One study compared intensified chemotherapy (BEACOPP_{escalated}) and radiotherapy with standard chemotherapy (ABVD) and radiotherapy based on positive interim-PET after two cycles in early-to-intermediate-stage HL. Treatment escalation may increase OS (HR 0.92, 95% CI 0.43 to 1.97; 260 participants; low-certainty) and PFS (0.67, 95% CI 0.37 to 1.20; 260 participants; low-certainty), although the CI included the possibility of no effect. The effect on secondary malignancies is very uncertain (RR 1.23, 95% CI 0.43 to 3.55; 234 participants, very low-certainty). No studies reported TRM, SAE and QoL.

PET-based escalation in advanced-stage, PET-positive participants

Two studies examined interim-PET-based escalation of PET-positive participants with rituximab in addition to chemotherapy in advanced-stage HL, which likely does not increase OS (HR 1.39, 95% CI 0.74 to 2.63; 795 participants; moderate-certainty) or PFS (HR 1.03, 95% CI 0.68 to 1.54; 582 participants; moderate-certainty). It may increase TRM (RR 4.00, 95% CI 0.45 to 35.5; 434 participants; very low-certainty), although the CI included the possibility of no effect. Escalation probably increases the number of participants with SAE (RR 1.61, 95% CI 1.00 to 2.60; 148 participants, moderate-certainty), and may reduce the number of participants with secondary malignancy (RR 0.67, 95% CI 0.28 to 1.60; 582 participants; low-certainty), although for this latter finding, the CI included the possibility of no effect. No study reported QoL.

Authors' conclusions

In early-stage HL, the effect of interim-PET-based treatment adaptation by omission of radiotherapy is uncertain. No effect was seen on long-term adverse events, although the follow-up of around five years may be too short to see an effect. In intermediate-stage HL, omission of radiotherapy may have little effect on both overall and progression-free survival, serious adverse events and secondary malignancies.

In advanced-stage HL, reducing chemotherapy upon negative interim-PET has the potential to increase overall survival while not negatively affecting progression-free survival and long-term adverse events. If combined modality treatment is opted for, omitting radiotherapy may increase both overall and progression-free survival, while reducing the negative effect of radiotherapy on secondary malignancies.

Interim-PET-positive treatment intensification by providing more chemotherapy in early-to-intermediate stage HL may be beneficial, while adding rituximab to standard chemotherapy in advanced-stage HL does not result in the expected improvement, but increases adverse events.

PLAIN LANGUAGE SUMMARY

Imaging-adapted therapy for individuals with Hodgkin lymphoma

Key findings

- The effect of treatment adaptation depends on the interim PET result and the stage of an individual with Hodgkin lymphoma (HL).
- While the effect of treatment de-escalation based on a negative interim PET scan is uncertain in early-stage HL, it may have little effect on overall and progression-free survival in intermediate-stage HL.
- In advanced-stage HL, reduction of chemotherapy or radiotherapy after a negative interim PET scan is probably beneficial for overall survival.
- Treatment escalation with rituximab after a positive interim PET scan probably increases adverse events without improving survival.

What is Hodgkin lymphoma?

Hodgkin lymphoma (HL) is a rare cancer of the lymphatic system (part of the immune system which keeps body fluid levels in balance and defends the body against infections). Its symptoms are painless swelling of lymphatic tissue, and some people experience fever, night sweats, and unintended weight loss. HL is one of the most curable cancers worldwide and is usually treated with chemotherapy, sometimes in combination with radiotherapy, depending on the stage.

What is positron emission tomography?

Positron emission tomography (PET) is an imaging technique that can visualise the activity of cells within body tissues using a radioactive liquid injection, i.e. based on sugar. As cancer cells usually grow quickly, they need a lot of energy. They thus take up more of the radio-labelled sugar injection and are brighter on the scan. PET is commonly used in the determination of the extent (stage) of cancers or as a response assessment tool to therapy. Recently, it has also been used to look at whether it can be used during treatment to adapt according to the cell's activity level. Treatment could be reduced in individuals who show a good response to chemotherapy, thus those with less activity, which is called PET-negative, to decrease the risk of long-term unwanted effects (adverse events). Treatment may be increased in those showing a poor response, thus more activity, which is called PET-positive.

What did we want to find out?

We wanted to know whether treatment adaptation after a PET scan during treatment improves overall survival, time without disease progression (progression-free survival), treatment-related mortality, unwanted effects, and quality of life. This included less treatment after a negative PET scan, or more treatment after a positive PET scan.

What did we do?

We systematically searched for studies in relevant medical databases and found 10 randomised controlled studies published before 17 November 2023, which we evaluated and meta-analysed according to PET status (negative or positive) and stage (early-to-intermediate or advanced).

What did we find?

From the 10 studies we found, one included 667 individuals with early-stage HL, one included 651 individuals with intermediate-stage HL, and three included 1639 individuals with early-to-intermediate stage HL (with one separating those two groups). Five studies included 3629 individuals with advanced-stage HL. All studies explored whether leaving out radiotherapy after a negative PET has beneficial outcomes, but only three studies looked at outcomes after more intensive treatment in those with a positive PET scan.

The standard therapy in early-stage and intermediate-stage HL usually includes chemotherapy and radiotherapy. In early-stage HL, it is uncertain whether leaving out radiotherapy after a negative PET scan is beneficial or not regarding survival and progression-free survival. It may have little to no effect on serious unwanted effects or the development of another cancer.

In intermediate-stage HL, leaving out radiotherapy after a negative interim PET may not affect overall survival and progression-free survival negatively.

The standard therapy for advanced-stage HL nowadays is more intensive chemotherapy. Reducing the intensity of the chemotherapy based on a negative PET scan probably increases overall survival and has little effect on progression-free survival. It may also reduce the development of another cancer. Some countries use, similar to early stages, chemotherapy and radiotherapy. Leaving out radiotherapy based on having negative PET scans both during and after chemotherapy may increase overall survival and progression-free survival, and may reduce the development of another cancer.

Evidence from one study suggests that more intensive chemotherapy after a positive PET scan in early-to-intermediate stage HL may increase overall survival and progression-free survival, but could not determine whether people develop more or fewer other cancers.

Two studies tested whether rituximab, a monoclonal antibody, in addition to standard chemotherapy could improve outcomes. It does probably not increase overall or progression-free survival, but probably increases serious unwanted effects and treatment-related mortality.

What are the limitations of the evidence?

Many studies did not report unwanted effects in the format we need to use the data, although they were assessed. None of the studies has published their data on quality of life.

The certainty in the evidence varies between comparisons. It is mostly moderate for leaving out radiotherapy in early-to intermediate-stage PET-negative individuals, less intensive chemotherapy in PET-negative advanced-stage HL, and rituximab in PET-positive advanced-stage participants, but low-to-very-low for omitting radiotherapy after negative PET in advanced stage and intensifying chemotherapy after positive PET in early-to-intermediate stage HL.



SUMMARY OF FINDINGS

Summary of findings 1. PET-adapted treatment compared to standard of care in PET-negative, early-stage Hodgkin lymphoma

PET-adapted de-escalation compared to standard of care in PET-negative, early-stage HL

Patient or population: PET-negative, early-stage Hodgkin lymphoma

Setting: inpatient and outpatient

Intervention: PET-adapted treatment (chemotherapy, no further radiotherapy)

Comparison: standard of care (chemotherapy plus radiotherapy)

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	C of partici- pants (studies)	Certainty of the evidence (GRADE)	Comments
	Frequency with standard of care	Frequency with PET- adapted de-es- calation				
Overall survival (OS) at 5 years	981 per 1000 [§]	984 per 1000 (903 to 998)	HR 0.84 (0.13 to 5.32) [death]	1024 (2 RCTs)	⊕⊕⊕⊕ Very low ^{a,b}	It is uncertain whether PET-adapted treatment in- creases or decreases overall survival compared to standard therapy.
Progression-free survival (PFS) at 5 years	964 per 1000 [§]	847 per 1000 (353 to 974)	HR 4.52 (0.72 to 28.41) [progression, death]	1034 (2 RCTs)	⊕⊕⊕⊕ Very Low ^{a,c}	It is uncertain whether PET-adapted treatment in- creases or decreases progression-free survival com- pared to standard therapy.
Treatment-related mortality (TRM) <i>during treatment</i>	0/237	0/238	not estimable	465 (1 RCT)	⊕⊕⊕⊕ Very Low ^{a,b,c}	It is uncertain whether PET-adapted treatment in- creases or decreases treatment-related mortality compared to standard therapy.
Serious adverse events (SAE)	-	-	-	0 RCTs	-	There were no studies with published results on seri- ous adverse events for this comparison.
Incidence of sec- ondary malignancy at 10 years	48 per 1000	40 per 1000 (22 to 72)	RR 0.83 (0.46 to 1.49)	983 (2 RCTs)	⊕⊕⊕⊕ Low ^{c,d}	PET-adapted de-escalation may result in little to no difference in the incidence of secondary malignancy compared to standard therapy.
Quality of life (QoL)	-	-	-	0 RCTs	-	There were no studies with published results on quality of life for this comparison.

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

§ Baseline risk estimated based on the median 5-year overall survival and progression-free survival of the standard arm in two included studies for this comparison (H10F, HD16)

CI: confidence interval; **HR:** hazard ratio; **RCT:** randomised controlled trial; **RR:** risk 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.

^aDowngraded two levels for serious imprecision due to very wide confidence intervals.

^bDowngraded one level for serious risk of bias due to high risk of attrition bias.

^cDowngraded one level for serious risk of bias (all trials were open-label, which may have affected subjective outcome assessments such as PFS, TRM; high risk of attrition bias in one study).

^dDowngraded one level for serious imprecision small number of events, wide confidence intervals.

Summary of findings 2. PET-adapted treatment compared to standard of care in PET-negative intermediate-stage HL

PET-adapted treatment compared to standard of care in PET-negative, intermediate-stage HL

Patient or population: PET-negative intermediate-stage Hodgkin lymphoma

Setting: inpatient and outpatient

Intervention: PET-adapted treatment (chemotherapy, no further radiotherapy)

Comparison: standard of care (chemotherapy plus radiotherapy)

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	C of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Frequency with standard of care	Frequency with PET- adapted de-es- calation				
Overall survival (OS) at 5 years	955 per 1000 [§]	959 per 1000 (914 to 981)	HR 0.91 (0.42 to 1.96) [death]	1073 (2 RCTs)	⊕⊕⊕⊕ Low ^{a,b}	PET-adapted de-escalation may have little to no effect on overall survival compared to standard therapy.

Progression-free survival (PFS) at 5 years	947 per 1000 [§]	917 per 1000 (865 to 950)	HR 1.59 (0.95 to 2.67) [progression, death]	1073 (2 RCTs)	⊕⊕⊕⊖ Low ^{a,b}	PET-adapted de-escalation may have little effect on progression-free survival compared to standard therapy.
Treatment-related mortality (TRM) during treatment	2 per 1000	7 per 1000 (1 to 67)	RR 2.94 (0.31 to 28.20)	1690 (2 RCTs)	⊕⊕⊕⊖ Very low ^{b,c}	The evidence is very uncertain about the effect of PET-adapted de-escalation on treatment-related mortality.
Serious adverse events (SAE) until 30 days after the last protocol treatment, or longer if related	295 per 1000	298 per 1000 (248 to 357)	RR 1.01 (0.84 to 1.21)	1096 (1 RCT)	⊕⊕⊕⊖ Low ^{b,d}	PET-adapted de-escalation may result in little to no difference in serious adverse events compared to standard therapy.
Incidence of secondary malignancy at 10 years	49 per 1000	50 per 1000 (28 to 90)	RR 1.01 (0.57 to 1.82)	1515 (2 RCTs)	⊕⊕⊕⊖ Low ^{a,b}	PET-adapted de-escalation may result in little to no difference in the incidence of secondary malignancy compared to standard therapy.
Quality of life (QoL)	-		-	0 RCTs	-	There were no studies with published results on quality of life for this comparison.

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

[§] **Baseline risk** estimated based on the median 5-year overall survival and progression-free survival of the standard arm in two included studies for this comparison (H10U, HD17)

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.

a. Downgraded one level for serious imprecision due to very wide confidence intervals

b. Downgraded one level for serious risk of bias due to high risk for attrition bias and no blinding of trials which may have affected subjective outcomes such as PFS, TRM, QoL.

c. Downgraded two levels for very serious imprecision (only two studies with two events, therefore wide 95% confidence intervals).

d. Downgraded one level for potential publication bias (most studies do assess serious adverse events; however, only one study reported SAEs).

Summary of findings 3. PET-adapted de-escalation compared to standard of care in PET-negative, advanced-stage Hodgkin lymphoma - reduction of chemotherapy

PET-adapted de-escalation compared to standard of care in PET-negative, advanced-stage Hodgkin lymphoma

Patient or population: PET-negative, advanced-stage Hodgkin lymphoma

Setting: inpatient and outpatient

Intervention: PET-adapted de-escalation (reduction of chemotherapy)

Comparison: standard of care

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	C of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Frequency with standard of care	Frequency with PET- adapted de-es- calation				
Overall survival (OS) at 5 years	972 per 1000 [§]	982 per 1000 (970 to 989)	HR 0.65 (0.40 to 1.07) [death]	2633 (3 RCTs)	⊕⊕⊕⊖ Moderate ^a	PET-adapted de-escalation probably increases overall survival compared to standard of care, although the CI included the possibility of no effect.
Progression-free survival (PFS) at 5 years	899 per 1000 [§]	901 per 1000 (875 to 920)	HR 0.98 (0.78 to 1.25) [progression, death]	2633 (3 RCTs)	⊕⊕⊕⊖ Moderate ^b	PET-adapted de-escalation probably results in little to no difference in progression-free survival compared to standard of care.
Treatment-related mortality (TRM) <i>during treatment</i>	12 per 1000	2 per 1000 (1 to 8)	RR 0.21 (0.06 to 0.73)	2761 (3 RCTs)	⊕⊕⊕⊖ Low ^{a,b}	PET-adapted de-escalation may reduce treatment-related mortality compared to standard of care.
Serious adverse events (SAE)	-	-	-	-	-	There were no studies with published results on serious adverse events for this comparison, although almost all studies assessed SAEs.
Incidence of sec- ondary malignancy at 5 years	40 per 1000	35 per 1000 (24 to 51)	RR 0.87 (0.60 to 1.26)	2757 (3 RCTs)	⊕⊕⊕⊖ Low ^{a,b}	PET-adapted de-escalation may reduce the incidence of secondary malignancy compared to standard of care, although the CI included the possibility of no effect.

Quality of life -

There were no studies with published results on quality of life for this comparison.

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

§ Baseline risk estimated based on the median 5-year overall survival and progression-free survival of the standard arm in three included studies for this comparison (AHL2011, HD18, RATHL).

CI: confidence interval; HR: hazard ratio; RCT: randomised controlled trial; RR: risk 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.

^a Downgraded one level for serious imprecision (OS: 95% confidence interval includes the line of zero effect, optimal information size (N = 2181 per arm based on sample incidence) not reached; TRM: low number of events, 16 in standard and 2 in de-escalation; incidence of secondary malignancy: 95% confidence interval includes the line of zero effect, optimal information size (N = 12,055 per arm based on sample incidence) not reached)

^b Downgraded one level for serious risk of bias (all trials were open-label, which may have affected subjective outcome assessments such as PFS, TRM, quality of life)

Summary of findings 4. PET-adapted de-escalation in advanced stage Hodgkin lymphoma - omission of radiotherapy

PET-adapted de-escalation by omission of radiotherapy compared to chemotherapy and radiotherapy in advanced-stage Hodgkin lymphoma

Patient or population: advanced-stage Hodgkin lymphoma

Setting: inpatient and outpatient

Intervention: PET-adapted de-escalation by omission of radiotherapy

Comparison: chemotherapy and radiotherapy

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	C of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with chemotherapy and radiotherapy	Risk with PET-adapted de-escalation by omission of radiotherapy [§]				

Overall survival at 3 years	987 per 1000	991 per 1000 (951 to 999)	HR 0.63 (0.11 to 3.69) [death]	296 (1 RCT) [#]	⊕⊕○○ Low ^a	PET-adapted de-escalation by omission of radiotherapy may increase overall survival, although the CI included the possibility of no effect.
Progression-free survival at 3 years	920 per 1000	937 per 1000 (888 to 965)	HR 0.78 (0.43 to 1.43) [progression or death]	412 (2 RCTs)	⊕⊕○○ Low ^{b,c}	PET-adapted de-escalation by omission of radiotherapy may increase progression-free survival, although the CI included the possibility of no effect.
Treatment-related mortality	-	-	-	-	-	There were no studies with published results on treatment-related mortality.
Serious adverse events	-	-	-	-	-	There were no studies with published results on serious adverse events.
Incidence of secondary malignancy at 3 years	28 per 1000	12 per 1000 (2 to 59)	RR 0.41 (0.08 to 2.09)	349 (2 RCTs)	⊕⊕○○ Low ^{c,d}	PET-adapted de-escalation by omission of radiotherapy may reduce the incidence of secondary malignancy, although the CI included the possibility of no effect.
Quality of life	-	-	-	-	-	There were no studies with published results on quality of life.

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

§ Baseline risk estimated based on the 2 or 3-year overall survival and progression-free survival of the standard arm of HD0607 and HD0801.

Although two RCTs reported this outcome, one study (HD0801) could not estimate the effect as there was only one event.

CI: confidence interval; HR: hazard ratio; RR: risk 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.

^aDowngraded two levels for very serious imprecision due to low number of events, small sample size and wide confidence intervals.

^bDowngraded one level for serious risk of bias; all studies were unblinded, which may have affected subjective outcomes; in HD0801, the PP analysis showed an effect in opposition to ITT analysis, which may be caused by performance bias (those not receiving the assigned treatment being different than those who followed through the protocol).

^cDowngraded one level for serious imprecision due to small sample size and wide confidence intervals including the line of no effect.

^dDowngraded one level for serious risk of bias; all studies were unblinded, which may have affected subjective outcomes.

Summary of findings 5: PET-adapted escalation compared to standard of care in PET-positive, early to intermediate stage Hodgkin lymphoma

PET-adapted escalation compared to standard of care in PET-positive, early-to-intermediate stage Hodgkin lymphoma

Patient or population: PET-positive, early-to-intermediate-stage Hodgkin lymphoma

Intervention: PET-adapted escalation (intensified chemotherapy)

Setting: inpatient and outpatient

Comparison: standard of care

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	C of partici- pants (studies)	Certainty of the evidence (GRADE)	Comments
	Frequency with standard of care	Frequency with PET- adapted esca- lation				
Overall survival (OS) at 10 years	904 per 1000 [§]	911 per 1000 (820 to 958)	HR 0.92 (0.43 to 1.97) [death]	260 (1 RCT)	⊕⊕⊕⊕ Low ^a	PET-adapted treatment escalation with more intensive chemotherapy may increase overall survival, although the CI included the possibility of no effect.
Progression-free survival (PFS) at 10 years	792 per 1000 [§]	855 per 1000 (756 to 917)	HR 0.67 (0.29 to 1.55) [progression, death]	260 (1 RCT)	⊕⊕⊕⊕ Low ^{b,c}	PET-adapted escalation may increase progression-free survival compared to standard of care, although the CI included the possibility of no effect.
Treatment-related mortality (TRM)	-	-	-	-	-	There were no studies with published results on treatment-related mortality for this comparison.
Serious adverse events (SAE)	-	-	-	-	-	There were no studies with published results on serious adverse events for this comparison.
Secondary malignancy <i>Follow-up: 54 months</i>	51 per 1000	62 per 1000 (22 to 180)	RR 1.23 (0.43 to 3.55)	234 (1 RCT)	⊕⊕⊕⊕ Very low ^{a,b}	The evidence about the effect of PET-adapted treatment escalation on the incidence of secondary malignancy is very uncertain.
Quality of life	-	-	-	-	-	There were no studies with published results on quality of life for this comparison.

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

§ **Baseline risk** estimated based on the median 10-year overall survival and progression-free survival of the standard arm in one included study for this comparison (H10F & H10U)

CI: confidence interval; HR: hazard ratio; RCT: randomised controlled trial; RR: risk ratio; PET: position 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.

^a Downgraded two levels for very serious imprecision (small sample size and small number of events).

^b Downgraded one level for serious risk of bias (all trials were open label, which may have affected subjective outcome assessments such as PFS, TRM, quality of life).

^c Downgraded one level for serious imprecision (small sample size).

Summary of findings 6: PET-adapted escalation compared to standard of care in PET-positive, advanced-stage Hodgkin lymphoma

PET-adapted escalation compared to standard of care in PET-positive, advanced-stage Hodgkin lymphoma

Patient or population: PET-positive, advanced stage Hodgkin lymphoma

Setting: inpatient and outpatient

Intervention: PET-adapted escalation (standard treatment with additional rituximab)

Comparison: standard of care

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	C of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Frequency with standard of care	Frequency with PET- adapted esca- lation				
Overall survival (OS) at 5 years	933 per 1000 [§]	907 per 1000 (832 to 950)	HR 1.39 (0.74 to 2.63) [death]	795 (2 RCTs)	⊕⊕⊕⊖ Moderate ^a	PET-adapted treatment escalation with rituximab probably does not increase overall survival compared to standard of care.

Progression-free survival (PFS) at 5 years	735 per 1000 [§]	728 per 1000 (622 to 811)	HR 1.03 (0.68 to 1.54) [progression, death]	582 (2 RCTs)	⊕⊕⊕⊕ Moderate ^b	PET-adapted treatment escalation with rituximab probably does not increase progression-free survival compared with standard of care.
Treatment-related mortality (TRM) <i>during treatment</i>	5 per 1000	18 per 1000 (2 to 164)	RR 4.00 (0.45 to 35.50)	434 (1 RCT)	⊕⊕⊕⊕ Very low ^{b,c}	PET-adapted treatment escalation with rituximab may increase treatment-related mortality compared to standard of care, although the CI included the possibility of no effect.
Serious adverse events (SAE) <i>during treatment</i>	250 per 1000	403 per 1000 (250 to 650)	RR 1.61 (1.00 to 2.60)	148 (1 RCT)	⊕⊕⊕⊕ Moderate ^b	PET-adapted treatment escalation with rituximab probably increases the number of participants with serious adverse events.
Secondary malignancy at 5 years	41 per 1000	27 per 1000 (11 to 66)	RR 0.67 (0.28 to 1.60)	582 (2 RCTs)	⊕⊕⊕⊕ Low ^{a,b}	PET-adapted treatment escalation with rituximab may reduce the incidence of secondary malignancy, although the CI included the possibility of no effect.
Quality of life	-	-	-	-	-	There were no studies with published results on quality of life for this comparison.

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

[§] Baseline risk estimated based on the median 5-year overall survival and progression-free survival of the standard arms in two included study for this comparison (HD0607, HD18)

CI: confidence interval; HR: hazard ratio; RCT: randomised controlled trial; RR: risk 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.

^a Downgraded one level for serious imprecision (low number of events, small sample size).

^b Downgraded one level for serious risk of bias (all trials were open-label, which may have affected subjective outcome assessments such as PFS, TRM and quality of life).

^c Downgraded two levels for very serious imprecision (very low number of events, 1 and 4 in standard and escalation arm; small sample size).

BACKGROUND

Description of the condition

Hodgkin lymphoma (HL) is a cancer of the lymphatic system that usually involves the lymph nodes. In 20% of cases, the spleen is affected, and other organs such as the liver, lung, bone, or bone marrow can be involved (Swerdlow 2017). It has an incidence of 2 to 3 per 100,000 inhabitants in high-income countries (SEER 2022) and a lower incidence of around 1 per 100,000 in lower-income countries (Huang 2022), and is one of the most common malignancies in young adults (Metzger 2019; SEER 2022; Swerdlow 2017).

HL is a B-cell lymphoma characterised by the presence of a small proportion of tumour cells known as Hodgkin- and Reed-Sternberg cells (Fraga 2007). These malignant cells usually account for only about 1% of cells in the tumour tissue surrounded by a specific non-tumour micro-environment (Diaz 2011). It can be distinguished into classic HL, which accounts for around 90% of all HLs, and nodular lymphocyte-predominant HL (NLPHL), which constitutes the remaining 10%. The two types differ in morphology, phenotype and molecular features, and therefore in clinical behaviour as well as clinical presentation (Swerdlow 2017). Classic HL is further divided into four subtypes (nodular sclerosis, mixed cellularity, lymphocyte-rich classic HL and lymphocyte-depleted classic HL), which are treated similarly (Swerdlow 2017).

The disease usually develops in lymph nodes in the upper part of the body, mostly the cervical lymph nodes, and results in painless swelling of the lymphatic tissue involved. Peripheral extranodal involvement is rare. Up to 40% of individuals present with B-symptoms such as fever, drenching night sweats and a loss of more than 10% of body weight (Swerdlow 2017).

The Ann Arbor classification is used for staging to distinguish between four different tumour stages. Stages one to three indicate the degree of lymph node involvement, whereas stage four is indicative of disseminated organ involvement (Lister 1989). Risk factors for a poor prognosis include large mediastinal (bulky) mass, extranodal lesion, a high erythrocyte sedimentation rate and three or more involved lymph node areas (Swerdlow 2017), although definitions may vary slightly between different study groups. Additionally, the occurrence of bulky disease (largest tumour diameter greater than 10 cm), often referred to as the Cotswold modification (Lister 1989), is taken into consideration. HL is classified as early favourable, early unfavourable (intermediate) or advanced-stage disease (Engert 2007; Klimm 2005).

With cure rates of up to 90%, HL is one of the most curable cancers worldwide (Engert 2010; Engert 2012; Von Tresckow 2012). The currently accepted standard chemotherapy regimen is a combination of adriamycin, bleomycin, vinblastine and dacarbazine (ABVD) (Canellos 1992; Engert 2010). Individuals with early-stage disease usually receive a combination of ABVD and involved-field radiation (IF-RT) (Engert 2010; Von Tresckow 2012), whereas those with advanced-stage disease usually receive an intensified regimen such as BEACOPP (bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine and prednisone) (Borchmann 2011; Engert 2012; Skoetz 2017). In individuals who relapse or whose disease is refractory to initial treatment, high-dose chemotherapy followed by autologous

haematopoietic stem cell transplantation (ASCT) is the therapy of choice (Rancea 2013).

Description of the intervention

The intervention to be examined in this review is de-escalation or intensification of treatment of HL based on imaging with [18F]-fluorodesoxy-D-glucose (FDG)-positron emission tomography (FDG-PET, also called PET scanning) any time during treatment, which is called interim-PET. FDG-PET is an imaging tool based on a radio-labelled glucose analogue to detect metabolically active tissue such as cancers, which often show a higher glucose uptake. The 18F-FDG molecule comprises two parts, a tracer (2-deoxy-D-glucose) and 18F, a positron-emitting nuclide. When FDG is taken up by cells, it can be detected by scintigraphy thanks to the annihilation photons that are emitted. This process is used to give evidence about FDG-avid tumours, such as HL, and their state and degree of progression both for staging and response evaluation (Boellaard 2010).

Early interim FDG-PET is a good predictor of the prognosis of HL (Aldin 2020). The ability to differentiate between responding and non-responding individuals at an early point during therapy has led to randomised trials exploring response-adapted therapeutic strategies, such as de-escalation in responding individuals or escalation in non-responding individuals (Sickinger 2015).

How the intervention might work

As interim-FDG-PET, indicating the metabolic activity of the tumour, is a prognostic factor for long-term outcomes (Aldin 2020), treatment modification based on low as compared to high metabolic activity may align the two groups in terms of long-term outcomes. To improve the prognosis of those with a positive PET scan, treatment may be intensified (escalation) by i.e. increasing the number of cycles of chemotherapy, increasing or changing the agents, or additional immunotherapy. For those with a negative PET scan and thus low metabolic tumour activity after some cycles of chemotherapy, a reduction in treatment may still be efficacious in terms of overall survival (OS) and progression-free survival (PFS) while reducing the incidence of long-term adverse events such as secondary malignancies (Trotman 2021).

Why it is important to do this review

As PET-guided therapy adaptation is very important and pivotal for decision-making, the potential advantages and disadvantages of this approach have been evaluated in several randomised controlled trials (RCTs, i.e. H10; Picardi 2007; UK NCRI RAPID). This update of a systematic review on the effectiveness of interim PET-based treatment modification in individuals with HL extends the previous version by adding new studies regarding advanced stages, data on treatment de-escalation, comparisons and longer-term follow-up data from already included studies.

OBJECTIVES

To assess the effects of interim [18F]-FDG-PET imaging treatment modification in previously untreated individuals with HL.

METHODS

Criteria for considering studies for this review

Types of studies

We included randomised controlled trials (RCTs) with treatment adaptation based on [18F]-FDG-PET scan results. We included both full-text and abstract publications if sufficient information was available on the study design, characteristics of participants, interventions and outcomes. We excluded quasi-randomised trials (e.g. assignment to treatment in alternation or by date of birth) and cross-over trials.

Types of participants

We included trials enrolling individuals with a newly confirmed diagnosis of HL. For this review, we have applied no age, sex, gender or ethnicity restrictions. We considered individuals with all stages and subtypes of previously untreated HL. In trials involving mixed populations of individuals with haematological malignancies, we used only data from those with HL. We intended to exclude trials in which fewer than 80% of participants had HL unless subgroup data for these individuals were provided. As in the last version of this review ([Sickinger 2015](#)), we explicitly excluded trials that consisted of more than 80% of participants younger than 18 years of age.

In addition, we separated patients into three stages (early, intermediate and advanced) instead of two (early, advanced). We considered the terms "early unfavourable" and "intermediate" stage to be equivalent.

Types of interventions

The main experimental intervention was PET-adapted treatment modification, independent of the exact treatment, compared to standard therapy for HL (current at the time the study was conducted) without modification.

We considered the following comparisons:

- In individuals with early-stage or intermediate-stage PET-negative HL (subgroup analysis for early versus intermediate stage)
 - Experimental: treatment modification (e.g. no further treatment/no radiotherapy or further chemotherapy instead of radiotherapy)
 - Control: standard approach (e.g. further chemotherapy plus radiotherapy or additional radiotherapy only)
- In individuals with early-stage or intermediate-stage PET-positive HL (subgroup analysis for early versus intermediate stage)
 - Experimental: treatment modification (chemotherapy intensification, e.g. shift to BEACOPP, plus radiotherapy)
 - Control (e.g. further chemotherapy plus radiotherapy)
- In individuals with advanced-stage PET-negative HL
 - Experimental: treatment modification (chemotherapy reduction, e.g. number of cycles or chemotherapeutic agents)
 - Control: standard chemotherapy (e.g. BEACOPP or ABVD)
- In individuals with advanced-stage PET-positive HL
 - Experimental: treatment modification (chemotherapy intensification, e.g. number of cycles or number

of chemotherapeutic agents or plus additional immunotherapy)

- Control: standard chemotherapy (e.g. BEACOPP or ABVD)

Types of outcome measures

Primary outcomes

- Overall survival (OS), defined as the time from randomisation to death as a result of any cause.

Secondary outcomes

- Progression-free survival (PFS), defined as the time interval from random treatment assignment to the first confirmed progression, relapse or death from any cause, or to the last follow-up;
- Response rate, measured as overall response rate (ORR), complete response (CR) or partial response (PR)
 - We intended to use the definitions of ORR, CR and PR as provided in the publication;
- Treatment-related mortality (TRM);
- Adverse events (AEs);
 - Number of participants with any AE
 - Number of participants with grade 3 to 4 AE
 - Number of participants with serious adverse events (SAE)
 - Incidence of secondary malignancy
- Quality of life (QoL), measured using reliable and valid instruments.

We did not aim to evaluate the diagnostic or prognostic value of FDG-PET.

Timing of outcome measurement

Per study, we aimed to choose the effect estimate carrying the maximum information, i.e. more participants included, later follow-up time point provided. For practical reasons, we extracted point estimates for the most commonly reported follow-up time point.

For time-to-event outcomes, we chose the outcome measures incorporating the highest degree of information (hazard ratios where available, longest follow-up time).

Search methods for identification of studies

An experienced information specialist (IM) developed search strategies according to the *Cochrane Handbook for Systematic Reviews of Interventions* ([Lefebvre 2022](#)). We did not exclude any publications based on language to limit language bias, although we did not specifically search non-English databases completely.

Electronic searches

We searched the following databases and sources:

- MEDLINE (via Ovid, from the last review version in 2014 to 17 November 2023);
- Embase (via Ovid, from the last review version in 2014 to 17 November 2023);
- Cochrane Central Register of Controlled Trials (CENTRAL; via Cochrane Library, Issue 9, 2014 for the first version of this review;

Issue 10, 2023 for the current version; searched 17 November 2023);

- Sources of ongoing trials:
 - o ClinicalTrials.gov (up to 17 November 2023);
 - o WHO ICTRP (up to 17 November 2023).

All search strategies are included in [Appendix 1](#).

Handsearching

- We checked the reference lists of all identified trials, relevant review articles and current treatment guidelines for further literature.

Data collection and analysis

Selection of studies

Two review authors (NK, MG) independently screened the results of the search strategies for eligibility for the update of this review by reading the titles and abstracts. In case of disagreement, we obtained the full-text publication. If no consensus could be reached, we asked a third review author (NS) for the final decision, in accordance with standard Cochrane methods ([Lefebvre 2022](#)).

We documented the study selection process in a flow chart, as recommended in the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analysis) statement ([Moher 2009](#)), showing the total number of retrieved references and the number of included and excluded studies with exclusion reasons at full-text stage (see [Figure 1](#)).

Figure 1. PRISMA flow diagram

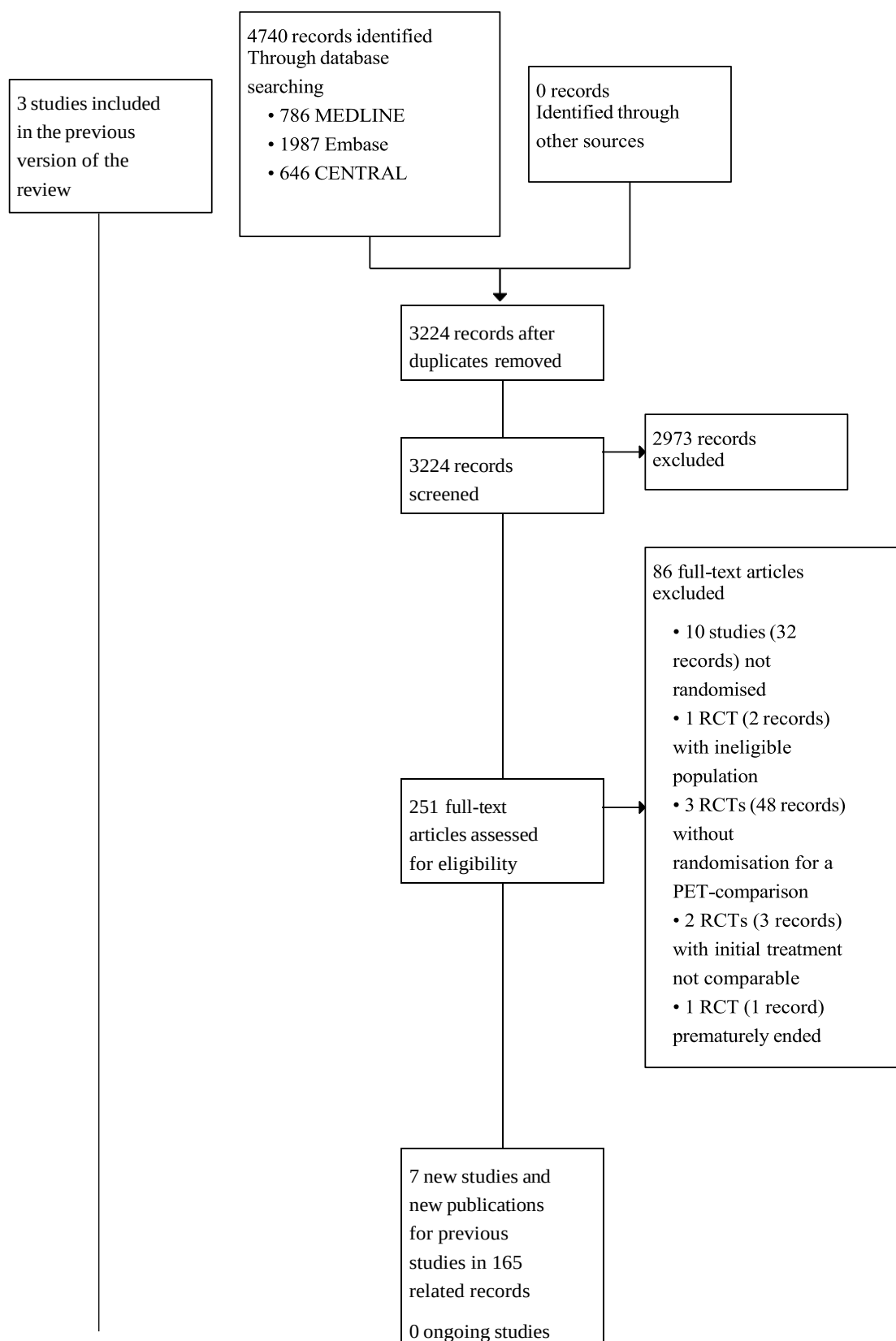
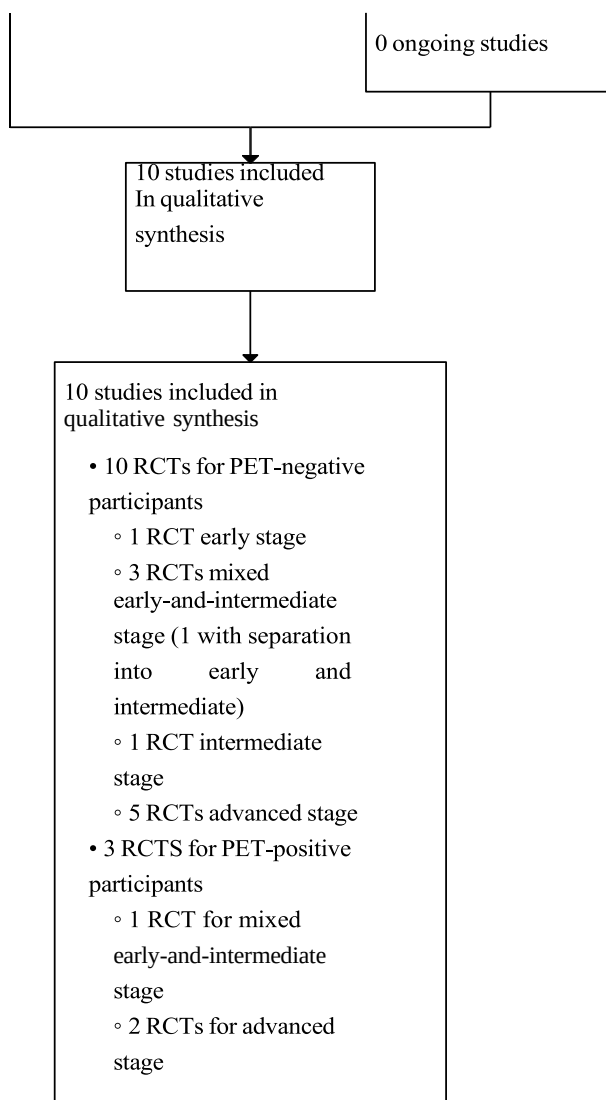


Figure 1. (Continued)



Data extraction and management

Two review authors (NK, MG, YSP) extracted data for the update of this review independently and in duplicate. If required, we contacted the authors of particular studies for supplementary information (Li 2023).

For the data extraction, we used a standardised form containing the following items:

- **General information:** author, title, source, publication date, country, language, duplicate publications;
- **Study characteristics:** trial design, aims, setting and dates, eligibility criteria and power calculations;
- **Participant characteristics:** age, sex, ethnicity, number of participants recruited/allocated/evaluated, participants lost to follow-up, additional diagnoses, stage of disease;
- **Interventions:** PET technique, the timing of PET assessment, the cutoff for PET-imaging, type of (multi-agent) chemotherapy

(intensity of regimen, number of cycles), field and dose of radiotherapy, duration of follow-up, treatment cross-overs, compliance with assigned treatment;

- **Outcomes:** outcome definition, statistical methods, events and sample size, effect estimates, 95% confidence intervals or other measure of variability for the outcomes listed under [Types of outcome measures](#).

Assessment of risk of bias in included studies

Two review authors (NK, MG) independently assessed the risk of bias in each study using the Cochrane risk of bias tool following the criteria outlined in the *Cochrane Handbook for Systematic Reviews of Interventions*. Discrepancies were solved by discussion (Higgins 2017).

- sequence generation;
- allocation concealment;
- blinding (participants, personnel, outcome assessors);

- incomplete outcome data;
- selective outcome reporting;
- other sources of bias.

For each criterion, we made a judgement using one of three categories:

- 'low risk': if the criterion was adequately fulfilled in the study.
- 'high risk': if the criterion was not fulfilled in the study.
- 'unclear risk': if the study report did not provide sufficient information to allow for a judgement of 'low risk' or 'high risk', or if the risk of bias was unknown for the rated criterion.

Dichotomous data

For binary outcomes, we calculated risk ratios (RRs) with 95% confidence intervals (CIs) per trial and outcome from the extracted number of participants randomised and the number of participants with events per study arm, where available.

Time-to-event data

For time-to-event outcomes, we extracted hazard ratios (HRs) from published data where available. When HR and corresponding 95% CI were not available, we recalculated these according to [Parmar 1998](#) and [Tierney 2007](#).

Continuous data

We intended to calculate continuous outcomes (e.g. QoL) as mean differences; however, no continuous outcomes were reported in the included studies.

Unit of analysis issues

We included studies that randomised individuals to treatment arms.

In case we encountered cluster-randomised or cross-over trials, we would have followed the methods according to [Higgins 2023b](#). Where events instead of number of participants were reported, we did not meta-analyse the outcome (i.e. for adverse events).

Dealing with missing data

Where possible, we attempted to recalculate effect estimates from published data. If outcome data were missing and could not be recalculated, we contacted the primary study investigators. Recalculations and author requests can be found in [Table 1](#).

Missing data for individuals within a study are assessed as part of the risk of bias ratings.

Assessment of heterogeneity

We assessed the heterogeneity of treatment effects between trials using a χ^2 test with a significance level at P value < 0.1 . We used the I^2 statistic to quantify possible heterogeneity ($I^2 > 30\%$ moderate heterogeneity, $I^2 > 75\%$ considerable heterogeneity) ([Deeks 2022](#)).

We intended to explore potential causes of heterogeneity through sensitivity and subgroup analyses; however, no analysis showed considerable heterogeneity. Sensitivity analyses are reported in [Appendix 2](#).

Assessment of reporting biases

In meta-analyses involving at least 10 trials, we intended to explore potential publication bias by generating a funnel plot and statistically testing this by conducting a linear regression test. We would have considered a P value of < 0.1 as significant for this test. However, as we included only a maximum of six trials per comparison in the review, this test was not conducted ([Page 2024](#)).

Data synthesis

We performed analyses according to the recommendations of the *Cochrane Handbook for Systematic Reviews of Interventions* ([Deeks 2022](#)), using Review Manager Web for the analyses. In contrast to the protocol, we used a random-effects model as the primary model for all analyses.

We based the estimation of the between-study variance of dichotomous outcomes on the Mantel-Haenszel method and used the inverse variance method for continuous outcomes and time-to-event outcomes where hazard ratios (HR) were available.

Subgroup analysis and investigation of heterogeneity

We intended to perform subgroup analyses of the following characteristics:

- Age;
- Type of HL (classic HL versus NLPHL);
- Intensity of chemotherapeutic regimen (e.g. aggressive therapy, such as BEACOPP, or less-aggressive therapy, such as ABVD);
- Duration of follow-up.

For age and duration of follow-up, we did not identify sufficient data. We examined stage and treatment intensity in subgroups or separate analyses. For instance, in early-to-intermediate-stage HL, we explored whether the addition of more chemotherapy in the control group would change the results. In advanced stage HL, we explored whether initiation with ABVD in contrast to BEACOPP would yield a subgroup difference. One study in early-stage HL ([HD16](#)) and one in advanced-stage HL ([HD18](#)) published results for NLPHL separately. From [HD16](#), only the data on PFS in the NLPHL group, not the cHL group, could be used. Although [HD17](#) included participants with NLPHL, the sample size was too small for analysis.

Sensitivity analysis

We intended to perform sensitivity analyses of the following characteristics if there were sufficient data:

- Measures of study quality (including full-text publications/abstracts);
- Preliminary results versus mature results.

All studies were published as full texts and reported final analyses. We performed only a sensitivity analysis of fixed-effect versus random-effects models for OS and PFS. Ad hoc, we also performed a sensitivity analysis for intention-to-treat versus per-protocol analysis for the main outcomes of OS and PFS.

Summary of findings and assessment of the certainty of the evidence

We used the GRADE (Grading of Recommendations Assessment, Development and Evaluation) framework to rate our certainty in

the evidence according to the five domains: study limitations, indirectness, inconsistency, imprecision and publication bias (Schünemann 2024). To create summary of findings tables, we used the software GRADEpro (<https://gradepro.org/>). For the expression of our results, we applied the wording suggestions by Santesso 2020.

We decided to present comparisons for early, intermediate and advanced stages separately. Mixed groups after a negative interim PET scan are presented in the Appendix. All treatment strategies may be important, as, e.g. the initial chemotherapy in advanced stage HL may differ between settings (Eichenauer 2018); therefore, we decided to present all treatment strategies as SoF tables.

We prioritised outcomes according to their relevance to patients. The most important outcome was OS, followed by PFS, TRM, AEs including second primary malignancies, and QoL.

RESULTS

Description of studies

Results of the search

The updated database search on 17 November 2023 resulted in 4740 database records (786 MEDLINE, 1987 Embase, 646 CENTRAL). After removal of duplicates, we screened 3224 titles and abstracts for inclusion. We excluded 2973 because they did not fulfil our eligibility criteria. We assessed 251 records in full-text format and 86 of these were excluded (see [Characteristics of excluded studies](#) and [Excluded studies](#)). Overall, in this update search, we identified seven new studies and updates on two already included studies in 165 study-related references (see [Characteristics of included studies](#)). Currently, we have not identified any ongoing studies fulfilling our eligibility criteria.

The search for the first version of this review led to the identification of 308 potentially relevant references. At the initial screening stage, 298 records were excluded due to a lack of conformity with the predefined inclusion criteria. Of the remaining 10 publications, three trials (H10; Picardi 2007; UK NCRI RAPID) reported in nine publications and involving four comparisons were included.

We documented the overall number of references screened, identified, selected, excluded and included in a PRISMA flow diagram (Figure 1).

Trial design

The included studies varied regarding multiple design aspects, i.e. whether they only adapted treatment for PET-negative participants, their time point of randomisation, and the definition of their analysis population (Table 2).

All included studies examined treatment adaptation in interim-PET-negative participants, while only three included a randomised treatment escalation strategy for those with a positive interim PET scan. Four out of 10 studies randomised participants at study enrolment (AHL2011; H10; HD16; HD17), compared to six studies randomised after interim or end-of-treatment PET scan. Two studies that randomised at study enrolment and did not include treatment escalation analysed PET-negative and PET-positive participants together as PET-adapted treatment arms, while all other studies separated PET-negative and PET-positive treatment arms.

All studies were open-label, and most studies had a full or partial central review of interim PET scans.

Setting and characteristics of participants

Except for HD0801, Picardi 2007 and UK NCRI RAPID, which were conducted in Italy and the UK, trials were multinational trials with the majority of countries being European (Table 3). Participants were recruited earliest in 2000 (Picardi 2007) and latest in 2017 (HD17). Sample sizes ranged from 116 to 1005 for treatment de-escalation comparisons and from 148 to 438 for treatment escalation comparisons.

One study included early-stage (HD16), three included early and intermediate stage (H10 analysed early and intermediate stage separately for PET-negative participants; Picardi 2007; UK NCRI RAPID), one included intermediate stage (HD17), and five covered advanced stage HL (AHL2011; HD0607; HD0801; HD18; RATHL). The median age in the included studies was between 30 and 39 years, including participants up to 79 years old. The percentage of female participants ranged from 37% to 57.1%. Only HD16, HD17 and HD18 included NPLHL (Table 3).

Characteristics of the intervention

FDG-PET scan

The majority of studies conducted the interim PET-scan for treatment adaptation after two cycles of initial chemotherapy, except for HD0607 and HD0801 (PET after cycles two and six must be negative), Picardi 2007 (after six cycles) and UK NCRI RAPID (after three cycles).

Cutoffs for scoring criteria differed between studies; H10 used the International Harmonization Criteria, HD0801 used a two-point visual score, Picardi 2007 did not report standardised criteria, while all other studies used the five-point score according to Lugano criteria (a five-point scale, previously Deauville score). HD16, HD17, HD18 and UK NCRI RAPID defined a score of 1 or 2 as PET-negative and 3 to 5 as positive, while AHL2011, HD0607, and RATHL defined a score of 1 to 3 as negative and a score of 4 or 5 as positive (Table 2). For this review, we decided to use the PET-negative and positive groups as provided in the primary studies without further subgroup analyses, as the criteria may contain even more variability due to technical advancements increasing sensitivity over time.

Treatment adaptation in PET-negative participants

All included trials randomised interim-PET negative participants to a treatment adaptation arm compared to standard of care. For studies at early and intermediate stages, this meant leaving out radiotherapy after initial chemotherapy with ABVD, ABVD plus BEACOPP^{escalated} or VEBEP, although some studies added one (H10, favourable; UK NCRI RAPID) or two (H10, unfavourable) cycles of additional chemotherapy compared with standard of care.

Comparisons in studies on advanced stages were more heterogeneous; after initial treatment with ABVD (HD0607; HD0801; RATHL) or BEACOPP (AHL2011; HD18), one study compared four cycles of BEACOPP with de-escalation to four cycles of ABVD (AHL2011), another de-escalated from four cycles of eBEACOPP to two cycles of eBEACOPP (HD18), one study de-escalated from four cycles of ABVD to AVD (RATHL), and two studies considered radiotherapy after six cycles of ABVD as standard of care and de-escalated to no further radiotherapy (HD0607; HD0801).

Treatment escalation in PET-positive participants

Of the three trials that randomised interim-PET positive participants to treatment escalation, one included early-to- intermediate stages and escalated participants after the initial two cycles of ABVD from the standard of care of two cycles of ABVD plus radiotherapy to two cycles of eBEACOPP plus radiotherapy (H10). The other two trials included advanced-stage participants and escalated from eBEACOPP (eight cycles or six cycles, HD0607 and HD18) to the same number of cycles of BEACOPP with additional rituximab.

Outcomes

All trials except for Picardi 2007 (number of events reported) and HD0801 (only one mortality event) reported overall survival as time-to-event analyses. All studies reported PFS as time-to-event except Picardi, who provided event-free survival as a time-to-event outcome (including secondary neoplasm events, which was only one; definitions are provided in Table 4).

Other outcomes were reported less consistently. Response was assessed in all studies but reported in only four (Table 5). Data on fertility was assessed in seven out of 10 studies but published for four studies only (AHL2011; HD16; HD18; RATHL), however, it was not always reported per randomised treatment group (Table 6). Quality of life was assessed in multiple studies, but no data have yet been published.

Treatment-related mortality was reported in six out of 10 studies. The rate of secondary malignancy was reported in eight of 10 studies. None of the studies reported the number of participants with any adverse event in a way we could analyse, although one primary study author provided data. Grade 3 to 4 adverse events were reported or provided by three studies. Two studies reported or provided data on serious adverse events (Table 7).

Analyses and assumptions of analysis for OS and PFS

Reporting and recalculations for included studies are presented in Table 1. Most studies provided the hazard ratio and Kaplan-Meier curves for their time-to-event outcomes; two study authors responded to an author's request. Reporting and presentation of censored events were variable. Only two studies (HD0607; HD17) checked the assumption of proportional hazards for the time-to-event analyses for OS and PFS; one study planned to do this in their protocol (UK NCRI RAPID).

All studies reported intention-to-treat analyses or analyses based on a predefined safety participant set for most outcomes. In case

only per-protocol analyses were reported, this is marked as a footnote in the respective forest plot.

Study funding and potential conflicts of interest

AHL2011 and HD16 were funded by governments or governments only, while HD0607, HD0801, HD17, and Picardi 2007 obtained grants from non-profit organisations such as independent research networks or foundations. UK NCRI RAPID and RATHL received funding by non-profit organisations and government agencies, while H10 and HD18 obtained a combination of governmental or non-profit and pharmaceutical industry support (for details, see table Table 8).

All studies except one (Picardi 2007) provided statements regarding potential conflicts of interest (for details see Table 8 and Characteristics of included studies).

Excluded studies

For the previous version of this review, 298 trials were excluded that did not match our inclusion criteria at the title and abstract stage. We re-included one study (HD0607) that was excluded previously due to the unavailability of data in the identified conference proceedings.

For the current version of the review, a total of 86 records covering 17 studies listed in the Characteristics of excluded studies were excluded because they did not match our initial inclusion criteria:

- Ten trials (32 records) provided non-randomised PET-adapted treatment (CALGB 50604; CALGB50801; EORTC-COBRA; GATLA LH-05; JCOG1305; NCT03226249; NCT03712202; NCT05922904; RAFTING; SWOG S0816);
- One trial (two records) included a majority of participants aged under 18 years old (AHOD0031);
- Three trials (48 records) were randomised trials that described different pathways for interim-PET-negative and PET-positive participants but did not randomise within PET classification (ECHELON-1; HD15; HD21);
- Two trials (three records) adapted treatment based on interim-PET, but started already with different treatment arms (FIL; RADAR);
- One trial (one record) was prematurely ended and was thus unable to provide data (H11).

Risk of bias in included studies

A summary of the risk of bias ratings can be found in Figure 2.

Figure 2. _ 'Risk of bias' summary: review authors' judgements about each 'Risk of bias' item for each included study.

	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding (performance bias): All outcomes	Blinding of outcome assessment (detection bias): Mortality outcomes	Blinding of outcome assessment (detection bias): Subjective outcomes (all except mortality)	Incomplete outcome data (attrition bias): Efficacy outcomes	Incomplete outcome data (attrition bias): Safety outcomes	Selective reporting (reporting bias)	Other bias
AHL2011	+	+	+	+	-	+	+	+	+
H10	+	+	+	+	-	-	-	?	?
HD0607	+	?	+	+	-	+	+	+	+
HD0801	+	+	?	+	-	?	+	+	+
HD16	+	+	?	+	-	+	-	?	+
HD17	+	+	-	+	-	+	+	?	?
HD18	+	+	+	+	-	+	+	?	+
Picardi 2007	?	?	+	+	-	+	+	?	+
RATHL	+	+	+	+	-	?	+	+	+
UK NCRI RAPID	+	+	?	+	-	+		?	+

Allocation

The risk of selection bias due to inadequate random sequence generation was judged low for all studies except [Picardi 2007](#), which did not provide further information on the randomisation. The allocation was described explicitly as concealed or randomisation was central for all studies except [HD0607](#) and [Picardi 2007](#), which did not provide sufficient information for a judgement.

Performance bias

All included studies were unblinded. We judged performance bias to be low for five studies ([AHL2011](#); [H10](#); [HD0607](#); [Picardi 2007](#); [RATHL](#)), as we assumed that there were no systematic differences in additional treatments, treatment discontinuation or follow-up of participants due to preplanned schedules and similar rates of participants discontinuing or dropping out. We rated the risk of performance bias as unclear for three studies due to changes between treatment arms ([HD16](#); [UK NCRI RAPID](#)) or interruption without further specification ([HD0801](#)), and high for one study due to unbalanced participant dropout ([HD17](#)).

Detection bias

For the objective outcome, overall survival, we considered the lack of blinding as irrelevant and judged all studies as being at low risk of bias. For subjective outcomes such as PFS, TRM, etc., we judged all studies as being at high risk of detection bias as no study reported blinding of the outcome assessors explicitly.

Incomplete outcome data

We rated the majority of studies as being at low risk of attrition bias, as the amount of missing data was low and censoring was balanced between treatment arms. However, two studies were at unclear risk of attrition bias ([HD0801](#); [RATHL](#)) for efficacy outcomes, because censored observations were not marked and the length of follow-up was not reported per treatment arm. One study (with two comparisons) was rated as being at high risk of bias due to lacking participants in the long-term follow-up without further elaboration ([H10](#)).

Selective reporting

The risk of bias due to selective reporting was judged as low for four studies. We rated six studies as being unclear due to either a missing protocol ([Picardi 2007](#)) or some outcomes such as long-term toxicities that require long-term follow-up not yet being published ([HD16](#); [HD17](#); [HD18](#); [UK NCRI RAPID](#)), and one study ([H10](#)) because it changed the definitions of one of the main outcomes (EFS). Quality of life data were not available for any study, although they were measured in the majority.

Other potential sources of bias

We rated all studies except two as not having other risks of bias. [H10](#) was rated as being unclear because it was halted early based on a preplanned interim analysis. In [HD17](#), the standard arm used IF-RT for both PET-positive and PET-negative participants, while the experimental arm had no further RT for PET-negative and IN-RT for PET-positive patients, which may limit comparability.

Effects of interventions

See: [Summary of findings 1](#) Summary of findings: PET-adapted treatment compared to standard of care in PET-negative, early-

stage Hodgkin lymphoma; [Summary of findings 2](#) Summary of findings: PET-adapted treatment compared to standard of care in PET-negative intermediate-stage HL; [Summary of findings](#)

[3](#) Summary of findings: PET-adapted de-escalation compared to standard of care in PET-negative, advanced-stage Hodgkin lymphoma - reduction of chemotherapy; [Summary of findings](#)

[4](#) Summary of findings: PET-adapted de-escalation in advanced stage Hodgkin lymphoma - omission of radiotherapy; [Summary of findings 5](#) Summary of findings 6: PET-adapted escalation compared to standard of care in PET-positive, early to intermediate stage Hodgkin lymphoma; [Summary of findings 6](#) Summary of findings 7: PET-adapted escalation compared to standard of care in PET-positive, advanced-stage Hodgkin lymphoma

Interim PET-negative participants

Early-stage HL

The comparison of treatment adaptation in interim-PET negative participants with early-stage HL included two studies ([H10](#), favourable HL participants; [HD16](#)). Treatment adaptation constituted the omission of radiotherapy in both studies, replaced by more chemotherapy in [H10](#). Median study follow-up ranged between 66 months and 9.5 years. From [H10](#), we included the data only up to the safety amendment in August 2010 that closed the chemotherapy-only arm, as enrolment after this time constituted treating both arms with ABVD and radiotherapy. The results for this comparison are presented in [Summary of findings 1](#).

Overall survival

It is uncertain whether PET-based treatment adaptation has an effect on overall survival compared to standard of care with combined chemotherapy and radiotherapy (HR 0.84, 95% CI 0.13 to 5.32; 2 RCTs, 1024 participants; very low-certainty evidence; [Analysis 1.1](#)). We downgraded our certainty in the evidence by two levels due to serious imprecision due to extremely large confidence intervals, and one level due to serious risk of attrition bias in [H10](#).

Progression-free survival

It is uncertain whether PET-based treatment adaptation has an effect on progression-free survival compared to standard of care with combined chemotherapy and radiotherapy (HR 4.52, 95% CI 0.72 to 28.41; 2 RCTs, 1034 participants; very low-certainty evidence; [Analysis 1.2](#)). We downgraded our certainty in the evidence by two levels due to serious risk of bias because all trials were open-label, which may have affected subjective outcomes, and two levels for very serious imprecision due to very wide confidence intervals.

Treatment-related mortality

It is uncertain whether PET-based treatment adaptation increases or decreases treatment-related mortality compared to standard of care with combined chemotherapy and radiotherapy, as the effect was not estimable due to a lack of events in [H10](#) and no data reported in [HD16](#) ([Analysis 1.3](#)).

Adverse events

No study on PET-based treatment adaptation in PET-negative, early-stage HL reported adverse events or serious adverse events as predefined for this review. Further data on adverse events were reported in a format that could not be analysed in this review ([Table 7](#)), i.e. for both arms together, per grade, or individual AEs.

PET-based treatment adaptation may result in little to no difference in the incidence of second primary malignancy compared to standard of care (RR 0.83, 95% CI 0.46 to 1.49; 2 RCTs including 993 participants; low-certainty evidence; [Analysis 1.4](#)).

Quality of life

Data on quality of life have not yet been published for this comparison.

Other outcomes

Fertility outcomes were assessed in one of the two studies ([Table 6](#)). Response to treatment was assessed by all studies, but reported heterogeneously or not reported at all ([Table 5](#)). Discontinuation of treatment due to toxicity was reported in one study for this comparison ([Table 9](#)). The [H10](#) update also reports on long-term pulmonary adverse events (adaptation: 8.2%, standard: 7.5%) and cardiovascular adverse events (de-escalation: 6%, standard: 4.6%).

Subgroup analyses

Subgroup analyses separating studies that adapted by omitting radiotherapy as compared to replacing radiotherapy with additional chemotherapy showed no difference between the adaptation strategies, although the number of studies is too low to draw valid conclusions ([Analysis 8.1](#); [Analysis 8.2](#); [Analysis 8.3](#)).

Only [HD16](#) included participants with NPLHL and published a separate subgroup analysis. Although the sample of NPLHL participants was small, the effect in this population followed the same trend as the main analysis regarding progression-free survival (RR 7.62, 95% CI 0.43 to 133.73; 1 RCT, 46 participants; [Analysis 9.1](#)). No further outcomes were reported.

Mixed early- and intermediate-stage HL

The comparison of treatment adaptation in interim-PET negative participants with early-to-intermediate stage HL included two studies ([Picardi 2007](#); [UK NCRI RAPID](#); [Appendix 3](#)). Both studies omitted radiotherapy without replacement by more chemotherapy, hence this comparison will further be called de-escalation.

Overall survival

PET-adapted treatment de-escalation may increase overall survival compared to standard of care with combined chemotherapy and radiotherapy (HR 0.51, 95% CI 0.15 to 1.71; 1 RCT, 420 participants; low-certainty evidence; [Analysis 2.1](#)). We downgraded our certainty in the evidence by two levels for serious imprecision due to wide confidence intervals and small sample size.

Progression-free survival

PET-adapted treatment de-escalation may decrease progression-free survival compared to standard of care with combined chemotherapy and radiotherapy (HR 2.00, 95% CI 1.00 to 3.97; 2 RCTs, 580 participants; low-certainty evidence; [Analysis 2.2](#)). We downgraded our certainty in the evidence by one level for serious risk of bias because all trials were open-label, which may have affected subjective outcomes, and by one level for serious imprecision due to wide confidence intervals and small sample size.

Treatment-related mortality

No study reported on treatment-related mortality for this comparison.

Adverse events

Data on adverse events were reported in a format that could not be analysed in this review ([Table 7](#)), i.e. for both arms together, per grade, or individual AEs.

No study reported on the incidence of secondary primary malignancy for this comparison.

Quality of life

Data on quality of life have not yet been published for this comparison.

Other outcomes

Fertility outcomes and treatment discontinuation due to toxicity were not assessed. Response to treatment was assessed by all studies, but reported heterogeneously or not reported at all ([Table 5](#)).

Subgroup analyses

For this comparison, no subgroup analyses were conducted.

Intermediate-stage HL

The comparison of treatment adaptation in interim-PET negative participants with intermediate-stage HL included two studies ([H10](#), unfavourable HL participants, [HD17](#); [Summary of findings 2](#)). PET-based treatment adaptation consisted of the omission of radiotherapy in both studies and being replaced by more chemotherapy in [H10](#). Median study follow-up ranged between 46 and 114 months. We included the updated 10-year data from [H10](#) from participants enrolled up to the safety amendment in August 2010 that closed the chemotherapy-only arm.

Although a recently published update on recurrence patterns in [H10](#) and [UK NCRI RAPID](#) reported separate data for early and intermediate stage HL for both studies, the data did not fit the definition of PFS (death not included). We therefore analysed [UK NCRI RAPID](#) separately in the comparison of early-to-intermediate HL.

Overall survival

PET-based treatment adaptation may have little to no effect on overall survival compared to standard of care with combined chemotherapy and radiotherapy (HR 0.91, 95% CI 0.42 to 1.96; 2 RCTs, 1073 participants; low-certainty evidence; [Analysis 3.1](#)). We downgraded our certainty in the evidence by one level due to serious imprecision due to large confidence intervals, and one level for serious risk of bias due to potential attrition bias in [H10](#).

Progression-free survival

PET-based treatment de-escalation may have little effect on progression-free survival compared to standard of care with combined chemotherapy and radiotherapy (HR 1.59, 95% CI 0.95 to 2.67; 2 RCTs, 1073 participants; low-certainty evidence; [Analysis 3.2](#)). We downgraded our certainty in the evidence by one level due to serious risk of bias because all trials were open-label, which may have affected subjective outcomes, and one level due to serious imprecision.

Treatment-related mortality

It is uncertain whether PET-based treatment adaptation increases or decreases treatment-related mortality compared with standard of care with combined chemotherapy and radiotherapy (RR 2.94, 95% CI 0.31 to 28.20; 2 RCTs, 1690 participants; very low-certainty evidence; [Analysis 3.3](#)), as the confidence intervals were extremely wide (very serious imprecision) and the lack of masking may have affected the judgement on relatedness with the treatment (serious risk of bias).

Adverse events

The effect of PET-based treatment adaptation on serious adverse events was reported by only one study; there may be little to no difference compared to standard of care (RR 1.01, 95% CI 0.84 to 1.21; 1 RCT, 1096 participants; low-certainty evidence; [Analysis 3.4](#)). We downgraded by one level for serious risk of bias due to lack of masking and one level for potential publication bias, as serious adverse events were assessed in more trials according to their trial protocol.

Further data on adverse events were reported in a format that could not be analysed in this review ([Table 7](#)), i.e. for both arms together, per grade, or individual AEs.

PET-based treatment adaptation may result in little to no difference in the incidence of second primary malignancy compared to standard of care (RR 1.01, 95% CI 0.57 to 1.82; 2 RCTs, 1515 participants; low-certainty evidence; [Analysis 3.5](#)). We downgraded our certainty in the evidence by one level due to serious risk of bias because all trials were open-label, which may have affected subjective outcomes, and one level due to serious imprecision.

Quality of life

Data on quality of life have not yet been published for this comparison.

Other outcomes

Fertility outcomes were assessed in both studies, but have not yet been published ([Table 6](#)). Response to treatment was assessed by all studies, but reported heterogeneously or not reported at all ([Table 5](#)). Discontinuation of treatment due to toxicity was reported in [HD17](#) ([Table 9](#)).

Subgroup analyses

Subgroup analyses separating studies that de-escalated by omitting radiotherapy ([HD17](#)) as compared to replacing radiotherapy with additional chemotherapy ([H10](#)) showed no difference between the adaptation strategies, although the number of studies is too low to draw valid conclusions ([Analysis 10.1](#); [Analysis 10.2](#); [Analysis 10.3](#); [Analysis 10.4](#)).

Only [HD17](#) included participants with NLPHL; however, the number of participants was too low for analysis of the treatment adaptation effect.

The [H10](#) long-term follow-up also reported on long-term pulmonary adverse events (adaptation: 13.1%, standard: 10.7%) and cardiovascular adverse events (de-escalation: 13.1%, standard: 12.7%).

Advanced-stage HL

Due to clinical heterogeneity regarding the interventions, we decided to meta-analyse studies that examined treatment de-escalation reducing the number or intensity of chemotherapy cycles ([AHL2011](#); [HD18](#); [RATHL](#), summarised in [Summary of findings 3](#)) separately from those that compared six cycles of chemotherapy plus radiotherapy with six cycles of chemotherapy only ([HD0607](#); [HD0801](#); presented in [Summary of findings 4](#)).

Overall survival

In advanced-stage HL, PET-based reduction of chemotherapy likely increases overall survival compared to standard of care (HR 0.65, 95% CI 0.40 to 1.07; 3 RCTs, 2633 participants; moderate-certainty evidence; [Analysis 4.1](#)), although the CI included the possibility of no effect. We downgraded by one level for serious imprecision because the optimal information size of 2181 participants per treatment arm was not reached.

Treatment de-escalation by omission of radiotherapy compared with combination therapy may increase overall survival (HR 0.63, 95% CI 0.11 to 3.69; 1 RCT, 296 participants; low-certainty evidence, although the CI included the possibility of no effect; one study could not estimate the effect due to only one event, [Analysis 5.1](#)). We downgraded by two levels due to very serious imprecision.

Progression-free survival

PET-based reduction of chemotherapy probably results in little to no difference regarding progression-free survival compared with standard of care (HR 0.98, 95% CI 0.78 to 1.25; 3 RCTs, 2633 participants; moderate-certainty evidence; [Analysis 4.2](#)). We downgraded by one level for serious risk of bias because all trials were open-label, which may have affected subjective outcomes.

Treatment de-escalation by omission of radiotherapy compared with combination therapy may increase progression-free survival (HR 0.78, 95% CI 0.43 to 1.43; 2 RCTs, 412 participants; low-certainty evidence, although the CI included the possibility of no effect; [Analysis 5.2](#)). We downgraded by one level due to serious risk of bias (open-label trial) and one level due to serious imprecision (the 95% CI crossed the line of zero effect).

Treatment-related mortality

PET-based reduction of chemotherapy may reduce treatment-related mortality compared to standard of care (HR 0.21, 95% CI 0.06 to 0.73; 3 RCTs, 2761 participants; low-certainty evidence; [Analysis 4.3](#)). We downgraded by one level for serious risk of bias because all trials were open-label, which may have affected subjective outcomes, and one level for serious imprecision due to the low number of events.

Treatment-related mortality was not reported for the comparison of treatment de-escalation by omission of radiotherapy compared with combination therapy.

Adverse events

Data on serious adverse events were not reported for studies reducing chemotherapy based on interim PET in advanced-stage HL.

Treatment de-escalation probably slightly decreased the number of participants with at least one grade 3 or 4 adverse events

compared with standard of care (RR 0.94, 95% CI 0.92 to 0.97; 2 RCTs, 1930 participants; moderate-certainty evidence; [Analysis 4.4](#)). We downgraded for serious risk of bias as all trials were open-label.

PET-based reduction of chemotherapy may reduce the incidence of secondary primary malignancy compared to standard of care (HR 0.87, 95% CI 0.60 to 1.26; 3 RCTs, 2757 participants; low-certainty evidence, although the CI included the possibility of no effect; [Analysis 4.5](#)). We downgraded by one level for serious risk of bias because all trials were open-label, and one level for serious imprecision because the optimal information size was not reached for this outcome.

Treatment de-escalation by omission of radiotherapy may reduce the incidence of secondary primary malignancy compared with combination therapy, although the CI included the possibility of no effect (RR 0.41, 95% CI 0.08 to 2.09; 2 RCTs, 349 participants; low-certainty evidence; [Analysis 5.3](#)). We downgraded by one level for serious risk of bias due to lack of blinding and one level for serious imprecision due to the low number of events resulting in wide confidence intervals.

Quality of life

Data on quality of life have not yet been published for treatment de-escalation in advanced-stage HL.

Other outcomes

Fertility outcomes were assessed and reported in three out of five studies ([AHL2011](#); [HD18](#); [RATHL](#), [Table 6](#)), although [RATHL](#) did not report the outcome per randomised treatment arm. Response to treatment was assessed in all studies but reported in only three studies with advanced-stage HL ([Table 5](#)). Discontinuation of treatment due to toxicity was reported in three out of five studies in advanced-stage HL ([Table 9](#)).

Subgroup analyses

Subgroup analyses based on initial chemotherapy (ABVD versus BEACOPP_{escalated}) showed no difference based on the test for subgroup differences ([Analysis 11.1](#); [Analysis 11.2](#); [Analysis 11.3](#); [Analysis 11.4](#); [Analysis 11.5](#)).

[HD18](#) published additional data on NLPHL. The data for progression-free survival agreed with the overall study result of [HD18](#) (RR 0.24, 95% CI 0.03 to 2.01; 1 RCT, 53 participants; [Analysis 12.1](#)); this indicates that de-escalation in case of PET-negativity can be considered in NLPHL as well.

Interim PET-positive participants

Early and intermediate stage

Only one RCT, [H10](#), examined treatment escalation in interim-PET positive participants by intensification of chemotherapy from four cycles of ABVD to two cycles of ABVD and two cycles of BEACOPP_{escalated}, including radiotherapy in both treatment arms. For this comparison, the primary study authors analysed favourable and unfavourable participants together. The author's conclusions changed between five-year and 10-year follow-up regarding the interpretation of improvement of PFS with treatment escalation. This comparison is presented in [Summary of findings 5](#).

Overall survival

Interim PET-based intensified chemotherapy and radiotherapy may increase overall survival compared to standard chemotherapy and radiotherapy, although the CI included the possibility of no effect (HR 0.92, 95% CI 0.43 to 1.97; 1 RCT, 260 participants; low-certainty evidence; 10-year follow-up). We downgraded by two levels for very serious imprecision due to the low sample size and the small number of events. The 10-year effect estimate was reduced compared with results reported at five years (HR 0.45, 95% CI 0.19 to 1.07; 361 participants; five-year follow-up).

Progression-free survival

Interim PET-based intensified chemotherapy and radiotherapy may increase progression-free survival compared to standard chemotherapy and radiotherapy, although the CI included the possibility of no effect (HR 0.67, 95% CI 0.29 to 1.55; 1 RCT, 260 participants; low-certainty evidence; 10-year follow-up). We downgraded for serious risk of bias, as all studies were open-label, and for serious imprecision due to small sample size. Similarly, the effect at 10 years diminished compared to the five-year results (HR 0.42, 95% CI 0.23 to 0.75; 361 participants; five-year follow-up).

Treatment-related mortality

No studies reported treatment-related mortality for this comparison.

Adverse events

No studies reported adverse events in the way that was preplanned to be analysed for this systematic review.

We are uncertain whether interim PET-based intensified chemotherapy and radiotherapy increases or decreases the incidence of secondary primary malignancy compared to standard chemotherapy with radiotherapy (1 RCT including 234 participants, very low-certainty evidence; 10-year estimate). We downgraded by two levels for very serious imprecision due to low sample size, very low number of events and wide confidence intervals, and one level for serious risk of bias due to lack of blinding.

Quality of life

No studies reported data on quality of life for this comparison.

Other outcomes

Fertility outcomes were assessed, but have not yet been published. Response to treatment was assessed but not reported. Discontinuation of treatment due to toxicity was not reported.

The long-term follow-up reported on long-term pulmonary (escalation: 15.2%, standard: 13.5%) and cardiovascular (escalation: 8.3%, standard: 6.5%) adverse events.

Advanced stage

Two studies examined interim-PET-based escalation of the chemotherapy regimen with additional rituximab compared to chemotherapy only ([HD0607](#); [HD18](#)). In [HD0607](#), participants received two cycles of ABVD before interim PET and received four cycles of BEACOPP_{escalated} and four cycles of BEACOPP with or without rituximab. In [HD18](#), participants received two cycles of BEACOPP_{escalated} before interim-PET2 and six further cycles with or without rituximab before safety amendment. After the

safety amendment, the rituximab arm was closed and participants received only four additional cycles of BEACOPP_{escalated}. This comparison is presented in [Summary of findings 6](#).

Overall survival

Interim-PET treatment escalation with rituximab probably does not increase overall survival compared with standard therapy (HR 1.39, 95% CI 0.74 to 2.63; 2 RCTs, 795 participants; moderate-certainty evidence). We downgraded by one level for serious imprecision due to the low number of events and the small sample size.

Progression-free survival

Interim-PET treatment escalation with rituximab probably does not increase progression-free survival compared with standard therapy (HR 1.03, 95% CI 0.68 to 1.54; 2 RCTs, 582 participants; moderate-certainty evidence). We downgraded by one level for serious risk of bias, as all studies were open-label.

Treatment-related mortality

Interim-PET treatment escalation with rituximab may increase treatment-related mortality compared with standard therapy, although the CI included the possibility of no effect (RR 4.00, 95% CI 0.45 to 35.50; 1 RCT, 434 participants; very low-certainty evidence). We downgraded by one level for serious risk of bias, as all studies were open-label, and two levels for very serious imprecision due to a very low event rate.

Adverse events

PET-adapted treatment escalation with rituximab probably increases the number of participants with serious adverse events (RR 1.61, 95% CI 1.00 to 2.60; 1 RCT, 148 participants; moderate-certainty evidence). We downgraded by one level for serious risk of bias, as all studies were open-label.

PET-adapted treatment escalation with rituximab may have little to no effect on the number of people with grade 3 to 4 adverse events compared with standard of care (RR 1.06, 95% CI 0.77 to 1.45; 2 studies, 586 participants; low-certainty evidence). We downgraded by one level for serious risk of bias, as all studies were open-label, and one level for serious imprecision due to wide confidence intervals.

PET-adapted treatment escalation with rituximab may increase the number of participants with any grade adverse events slightly compared with standard of care, although the CI included the possibility of no effect (RR 1.16, 95% CI 0.89 to 1.51; 1 RCT, 148 participants; low-certainty evidence). We downgraded by one level for serious risk of bias, as all studies were open-label, and one level for serious imprecision due to wide confidence intervals.

PET-adapted treatment escalation with rituximab may decrease the number of participants with secondary primary malignancy compared with standard of care, although the CI included the possibility of no effect (RR 0.67, 95% CI 0.28 to 1.60; 2 RCTs, 582 participants; low-certainty evidence). We downgraded by one level for serious risk of bias, as all studies were open-label, and one level for serious imprecision due to small sample size and low number of events.

Quality of life

No studies reported data on quality of life for this comparison.

Other outcomes

Fertility outcomes were assessed, but have not yet been published for one study and were not assessed in another. Response to treatment was assessed and reported in both studies for this comparison ([Table 5](#)). Discontinuation of treatment due to toxicity was reported in one out of two studies ([Table 9](#)).

DISCUSSION

Summary of main results

The current review identified 10 randomised controlled trials that examined FDG-PET-based treatment adaptation. Two comparisons included early-stage, two intermediate-stage and two early-to-intermediate-stage individuals for treatment adaptation after negative interim-PET. Five studies included advanced-stage participants for treatment de-escalation in PET-negative participants, and three studies explored treatment escalation in PET-positive participants.

Early-stage HL

In early-stage participants, it is uncertain whether PET-based treatment adaptation consisting of omission of radiotherapy with or without more chemotherapy in PET-negative participants affects overall survival or progression-free survival, as the certainty of the evidence is very low due to very imprecise estimates and risk of bias. There may be little to no effect on the incidence of secondary malignancies. Data on quality of life have not yet been reported. Based on this uncertainty, there is a lack of evidence for treatment adaptation in early-stage HL ([Summary of findings 1](#)).

Mixed early- and intermediate-stage HL

In studies that included both early- and intermediate-stage participants, PET-based treatment de-escalation in PET-negative participants, consisting of omission of radiotherapy for both studies in this comparison, may increase overall survival. The intention-to-treat (ITT) data show a possible decrease in PFS. For the [UK NCRI RAPID](#) trial, this seemingly contradictory result may be explained by the choice of the ITT analysis, as five of eight participants who died in the radiotherapy arm did not receive radiotherapy. None died from HL. In [Picardi 2007](#), no deaths were observed; it thus did not contribute to the analysis. No other outcome data could be analysed for this comparison ([Appendix 3](#)).

Only one study explored treatment escalation by providing more intensive chemotherapy in mixed early- and intermediate-stage, PET-positive participants. PET-based treatment intensification may increase overall survival and progression-free survival. Data on the incidence of secondary malignancies were very uncertain, and no data were reported on treatment-related mortality, (serious) adverse events or quality of life ([Summary of findings 5](#)).

Intermediate-stage HL

In intermediate-stage participants, PET-based treatment de-escalation in PET-negative participants, consisting of omission of radiotherapy for both studies in this comparison, may have little to no effect on overall survival and progression-free survival. The certainty in the evidence regarding treatment-related mortality is too low to draw conclusions; however, there may be little to no effect on serious adverse events and the incidence of secondary

malignancies. Data on quality of life have not yet been reported ([Summary of findings 2](#)).

Advanced-stage HL

In advanced-stage HL, three studies examined PET-based reduction of chemotherapy while two compared the omission of radiotherapy after chemotherapy.

PET-based reduction of chemotherapy probably increases overall survival without a change in progression-free survival. The reduction may decrease the incidence of treatment-related mortality and secondary malignancies. No data was available regarding (serious) adverse events or quality of life ([Summary of findings 3](#)). The comparison with radiotherapy as standard of care showed that overall survival and progression-free survival may be prolonged, while the incidence of secondary malignancy may be decreased. Data on (serious) adverse events and quality of life were not available ([Summary of findings 4](#)).

Treatment escalation with rituximab added to standard of care in advanced PET-positive participants probably does not increase overall survival or progression-free survival but may increase treatment-related mortality and probably increases the number of participants with serious adverse events. Escalation may reduce the incidence of secondary malignancies ([Summary of findings 5](#)).

Treatment adaptation across stages

From our data, we cannot conclude from all comparisons that treatment de-escalation after a negative interim PET can safely be implemented. In early-stage HL, it seems that the combination therapy of two cycles ABVD plus RT controls the tumour better than omitting RT, while in intermediate-stage HL, the more intensive initial treatment, six cycles ABVD or a 2+2 regimen of chemotherapy, may suffice for tumour control. Similarly, de-escalation in advanced stages, which have more intensive upfront chemotherapy treatment, probably results in sufficient tumour control, reducing or not impacting mortality.

The addition of the monoclonal antibody rituximab, which binds to the protein CD20 on B cells and triggers cell death, did not result in additional clinical benefits, although it increased the frequency of serious adverse events. This is in line with another phase 2 study that could not identify a beneficial treatment effect of rituximab added to ABVD in advanced stage HL ([Strati 2019](#)).

Overall completeness and applicability of evidence

Compared to the previous version of this review, new data could fill the gaps regarding overall survival, advanced stage HL and PET-positive participants. Currently, we have not identified any additional ongoing or unpublished studies; therefore, we are confident that we have covered the current state of the study landscape. However, multiple studies have still not published data on all their planned or assessed outcomes; there may thus be potential for updating our data, especially regarding adverse events, quality of life and fertility outcomes (for detailed information on reporting of outcomes, please refer to [Table 7](#) and [Table 6](#)).

There are some limitations to the applicability of the findings regarding the staging and risk factors, the interim-PET cutoff score and technological advancements, and the evolving standard of care.

According to the classification of the European Organisation for Research and Treatment of Cancer/Lymphoma Study Association (EORTC/LYSA), stage IIB with bulky or extranodal disease was considered early unfavourable disease, while the German Hodgkin Study Group (GHSg) assigned these participants to the advanced-stage group. Although the definition of stage/risk groups defines the further treatment and study recruitment into the included studies, [Rossi 2022](#) concluded that upfront ABVD, as in early stages, and upfront BEACOPP^{escalated}, as currently recommended for advanced stages, result in similar patient outcomes.

Another important concern for applicability is the difference in definitions regarding the 5-point score according to the Lugano criteria cutoff scores for PET-positivity. These varied between studies included in this review, i.e. [HD16](#), [HD17](#), [HD18](#) and [UK NCRI RAPID](#) defined scores 3 to 5 as positive, whereas [AHL2011](#), [HD0607](#) and [RATHL](#) defined scores 4 and 5 as positive. [Picardi 2020](#), a prospective cohort study including individuals with advanced stage HL, examined the diagnostic accuracy of FDG-PET compared with core needle-cutting biopsy after six cycles of ABVD (end-of-chemotherapy treatment PET) for subsequent treatment escalation. From those with a score of 4, only 12.5% had a positive biopsy, while those rated as score 5 all had a positive biopsy ([Picardi 2007](#)). This large difference in positive predictive value between scores 4 and 5 may also be present after two cycles of chemotherapy, and may be even lower for score 3.

An additional difficulty in interpreting the 5-point score is interobserver discrepancies between raters for various reasons ([Georgi 2023](#)). In the included studies, subjectivity in ratings was prevented by installing a central, often blinded review by multiple nuclear physicians; however, this may not be feasible in daily routine. For the current review, we used the cutoff as reported in the studies.

Over the time the 10 included studies have been conducted, the current standard of care has developed and even been adapted according to the studies included in this review. Hence, some comparisons of primary studies included in this review do not represent clinical praxis any more. Currently, the standard of care for early-stage PET-negative Hodgkin lymphoma is a combined modality treatment of two cycles of ABVD and radiotherapy ([Engert 2010](#); [Follows 2022](#); [Sasse 2017](#)). [Picardi 2007](#) uses a chemotherapy regimen (VEBEP) that is not used any more; [H10](#) favourable HL and [UK NCRI RAPID](#) consider three cycles of ABVD plus radiotherapy as standard treatment. For intermediate-stage HL, the current standard of care is polychemotherapy consisting of two cycles of BEACOPP^{escalated} and two cycles of ABVD according to the [HD17](#) protocol ([Follows 2022](#); [HD17](#)), or according to the [H10](#) unfavourable HL protocol, where the control group receives four cycles of ABVD plus radiotherapy. Multiple treatment pathways according to [AHL2011/HD18](#) with initial BEACOPP^{escalated} or according to the [RATHL](#) protocol with initial ABVD have been suggested for advanced-stage HL ([Follows 2022](#)). Since the results of [HD15](#) have been published, six rather than eight cycles of BEACOPP would be considered standard concerning [HD18](#). The radiotherapy comparison may be less applicable to the current clinical routine; however, radiotherapy can be considered on an individual basis.

The first-line treatment options for advanced-stage Hodgkin lymphoma are currently developing, and new agents such as

monoclonal antibodies are being tested as first-line treatment options in addition to chemotherapy (Vassilakopoulos 2023). For example, immune checkpoint inhibitors such as pembrolizumab and nivolumab, which attach to and block anti-programmed cell death protein 1 receptors, have shown high response rates in relapsed and refractory HL (Ansell 2019). ECHELON-1 examined brentuximab vedotin in combination with AVD (BV-AVD) as first-line therapy, which showed improvements in overall and progression-free survival compared with ABVD (Ansell 2022). More recent initial results comparing BV-AVD with nivolumab and AVD (N-AVD) show a benefit for N-AVD in terms of PFS and OS (Herrera 2023). Interim HD21 results on 1482 participants followed up for a median of 40 months suggest that a remodelled BEACOPP^{escalated} regimen, the more intensive chemotherapy recommended in some countries, using brentuximab vedotin, etoposide, cyclophosphamide, doxorubicin, dacarbazine and dexamethasone (BrECADD) may be non-inferior to BEACOPP^{escalated} regarding progression-free survival with a potentially reduced and changed toxicity profile (Borchmann 2023). With changing first-line therapy options, the role of interim PET scans might change and will need to be re-evaluated for each regimen. This is especially true for checkpoint inhibitors due to immune activation due to the treatment that might get detected, instead of tumour activity (Ansell 2019).

Generally, participants in clinical trials, including in the RCTs included here, fulfil strict eligibility criteria regarding age and other conditions (usually less than 60 years old, non-pregnant, restrictions regarding comorbidities). In the current review, studies had varying inclusion criteria regarding age, some with an age cutoff of up to 60, 70 or a maximum of 75 years. Other, non-PET-adapted trials specifically include participants from 60 years onwards to explore treatment options specifically for older individuals (e.g. Böll 2019; CHL001; Lazarovici 2021; RATIFY), which include monoclonal antibodies. As Hodgkin lymphoma is a rare disease, valuable evidence may also result from non-randomised clinical trials.

Quality of the evidence

For interim-PET-based treatment adaptation in early-stage HL, our certainty in the evidence is very low for overall survival and progression-free survival, and low for safety outcomes such as treatment-related mortality, serious adverse events and incidence of secondary malignancy due to risk of bias and imprecision. In the intermediate stages, our certainty in the evidence is mostly low. Similarly, our certainty in the evidence for the combined early to intermediate stage is low.

The evidence in the main comparison on PET-based chemotherapy reduction in advanced-stage HL was of moderate certainty regarding overall survival and progression-free survival but of low certainty for treatment-related mortality and the incidence of secondary malignancy.

As only one study examined treatment escalation in early- to intermediate-stage PET-positive participants, the comparison escalation with intensified chemotherapy was limited by the risk of bias and small sample size; the certainty ranged between low and very low for all outcomes. In advanced-stage HL, our certainty of no difference between escalation with rituximab for overall survival and progression-free survival and an increase in serious adverse

events is moderate, while other safety outcomes were rated as low- to very low-certainty.

Potential biases in the review process

In the update of the current review, we adhered to Cochrane guidance and performed all relevant steps in duplicate (screening, data extraction). Where we did not follow the published protocol, we noted changes explicitly and transparently throughout the review and in the section [Differences between protocol and review](#). It is possible that although our search strategy was previously peer-reviewed and considered sensitive, studies published in non-English literature without an English abstract may have been missed. One author of this review has received payments from pharmaceutical companies (BvT); they predefined the relevant comparisons and provided clinical input, but were not involved in screening, data extraction, analysis and risk of bias assessment for the included studies.

As multiple studies reported data in formats that could not be used for meta-analysis, we contacted primary study authors for additional data and received responses from two teams, which especially improved the reporting on OS and PFS. Unfortunately, data on adverse events and quality of life remains limited, as the definition and timing of measuring adverse events across studies is heterogeneous and quality of life data are not (yet) published.

Agreements and disagreements with other studies or reviews

To our knowledge, there are no recent systematic reviews on interim-PET treatment adaptation covering all stages of HL and de-escalation for PET-negative participants as well as escalation for PET-positive participants. In a comprehensive narrative review (Trotman 2021), all studies included in this review except HD0801 and Picardi 2007 are covered; however, they are only described narratively and without assessment of the study's risk of bias. Since then, H10, HD16, RATHL and UK NCRI RAPID have published additional data and some data on NLPHL for HD16, HD17 and HD18 have been published.

AUTHORS' CONCLUSIONS

Implications for practice

In early-stage Hodgkin lymphoma (HL), we are uncertain regarding the effect of omission of radiotherapy after negative interim positron emission tomography (PET). In intermediate-stage HL, PET-adapted omission of radiotherapy in PET-negative participants may have little to no effect on overall survival and progression-free survival, opening the possibility for this de-intensification of treatment. There was no difference regarding serious adverse events and the incidence of secondary malignancy. In advanced-stage HL, PET-based de-intensification of chemotherapy probably results in an increased survival rate with no effect on progression-free survival; treatment-related mortality and the incidence of secondary malignancies may be reduced. Similarly, if using combined modality treatment in advanced-stage HL, interim-PET-based omission of radiotherapy may improve overall survival and progression-free survival.

In mixed early- and intermediate-stage HL, PET-based treatment escalation of PET-positive participants with intensified chemotherapy may increase overall survival and progression-free

survival, but the results are based on one trial with a small sample size. The addition of rituximab in advanced-stage HL based on positive interim PET did not increase overall and progression-free survival but increased the number of participants with serious adverse events and possibly treatment-related mortality.

With the currently available treatment modalities, the prognosis of HL has improved tremendously over time. There are still ongoing efforts to keep the high treatment success rate while reducing toxicities, for instance in [HD21](#) and [SWOG S0816](#), by modifying current treatment standards.

Implications for research

The current update fills gaps that were discovered by the last version of the review ([Sickinger 2015](#)) regarding comparisons, outcomes and long-term data. However, data on adverse events and quality of life are still heterogeneously reported or not at all, although they were assessed by some studies. Until now, older participants and those who are pregnant have not been represented in the trials included in this review, although other, non-PET-adapted studies are being conducted to cover the elderly. Routine data including PET scans may need to be consulted to fill the gap regarding the elderly and pregnant, and to shed light on the effect of certain PET-adapted treatment strategies on fertility, long-term adverse events and quality of life.

ACKNOWLEDGEMENTS

Editorial and peer-reviewer contributions

Cochrane Haematology supported the authors in the development of this update.

The following people conducted the editorial process for this article:

- Sign-off Editor (final editorial decision): Dr Lise J Estcourt, Cochrane Haematology;
- Managing Editor (selected peer reviewers, provided editorial guidance to authors, edited the article): Makalapua Motu'apuaka and Jenny Bellorini, Central Editorial Service;
- Editorial Assistant (conducted editorial policy checks, collated peer-reviewer comments and supported the editorial team): Sara Hales-Brittain, Central Editorial Service;
- Copy Editor (copy editing and production): Anne Lethaby, Cochrane Central Production Service
- Peer reviewers (provided comments and recommended an editorial decision): Nuala Livingstone, Cochrane Evidence Production and Methods Directorate (methods review), Jo Platt, Central Editorial Information Specialist (search review), Brian Duncan (consumer review). Three additional peer reviewers provided clinical/content peer review but chose not to be publicly acknowledged.

We would also like to thank the primary study authors, Chiara Pavoni (Department of Oncology-Hematology, Azienda Socio-Sanitaria Territoriale Papa Giovanni XXIII, Bergamo, Italy) and Michael Fuchs, Department of Internal Medicine, University Hospital Cologne, who kindly provided more data regarding their studies. In addition, we thank Peter Borchmann for his contributions to the previous version of this review.

REFERENCES

References to studies included in this review

AHL2011 {published data only}

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. [DOI: [10.1097/HS9.0000000000000060](https://doi.org/10.1097/HS9.0000000000000060)]

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**(Suppl 15):7503. [DOI: [10.1200/JCO.2018.36.15-suppl.7503](https://doi.org/10.1200/JCO.2018.36.15-suppl.7503)]

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: interim analysis of the AHL2011 LYSA study. *Blood* 2015;**126**(23):577. [DOI: [10.1182/blood.V126.23.577.577](https://doi.org/10.1182/blood.V126.23.577.577)]

Casasnovas O, Kanoun S, Tal I, Cottreau AS, Edeline V, Brice P, et al. Baseline total metabolic volume (TMTV) predicts the outcome of patients with advanced Hodgkin lymphoma (HL) enrolled in the AHL2011 LYSA trial. *Haematologica* 2016;**101**(Suppl 1):4. [PMID: PMC4894295]

* Casasnovas RO, Bouabdallah R, Brice P, Lazarovici J, Ghesquieres H, Stamatoulas 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**:202-15. [DOI: [10.1016/S1470-2045\(18\)30784-8](https://doi.org/10.1016/S1470-2045(18)30784-8)]

Casasnovas RO, Bouabdallah R, Brice P, Lazarovici J, Ghesquieres H, Stamatoulas A, et al. Positron emission tomography-driven strategy in advanced Hodgkin Lymphoma: prolonged follow-up of the AHL2011 phase III Lymphoma Study Association study. *Journal of Clinical Oncology* 2022;**40**(10):1091-101. [DOI: [10.1200/JCO.21.01777](https://doi.org/10.1200/JCO.21.01777)]

Casasnovas RO, Kanoun S, Tal I, Cottreau 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**(Suppl 15):7509. [DOI: [10.1200/JCO.2016.34.15_suppl.7509](https://doi.org/10.1200/JCO.2016.34.15_suppl.7509)]

Casasnovas RO, Meignan M, Reman O, Gaillard I, Stamatoulas 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):TPS8615. [DOI: [10.1200/jco.2013.31.15_suppl.tps8615](https://doi.org/10.1200/jco.2013.31.15_suppl.tps8615)]

Casasnovas O, Racape J, Dechene J, Bouabdallah R, Brice P, Lazarovici J, et al. PET-guided strategy improves the safety of BEACOPP-based treatment in advanced Hodgkin lymphoma: prolonged follow-up of the LYSA AHL 2011 phase 3 study. *Blood* 2020;**136**(Suppl 1):23-4. [DOI: [10.1182/blood-2020-136056](https://doi.org/10.1182/blood-2020-136056)]

Chevreaux S, De Barros S, Laurent C, Durand A, Delpierre C, Robert P, et al. Influence of sociodemographic determinants on the Hodgkin lymphoma baseline characteristics in long survivors patients enrolled in the prospective phase 3 trial AHL2011. *Cancers* 2022;**15**(1):22. [PMID: 36612050]

Demeestere I, Dechene J, Racape J, Bouabdallah R, Brice P, Ghesquieres H. Ovarian function recovery in newly diagnosed advanced Hodgkin lymphoma patients treated with PET-adapted regimen: prospective analysis of a multicentric randomized phase 3 trial (AHL2011). *Human Reproduction (Oxford, England)* 2020;**35**(Suppl 1):17-8. [DOI: [10.1093/humrep/35.Supplement_1.1](https://doi.org/10.1093/humrep/35.Supplement_1.1)]

Demeestere I, Racape J, Dechene J, Dupuis J, Morschhauser F, De Wilde V, et al. Gonadal function recovery in patients with advanced Hodgkin Lymphoma treated with a PET-adapted regimen: prospective analysis of a randomized phase III trial (AHL2011). *Journal of Clinical Oncology* 2021;**39**(29):3251-60. [DOI: [10.1200/JCO.21.00068](https://doi.org/10.1200/JCO.21.00068)]

Kanoun S, Berriolo-Riedinger A, Cottreau AS, Edeline V, Tal I, Brice P, et al. Total metabolic tumor volume and tumor dissemination are independent prognostic factors in advanced Hodgkin lymphoma. *Blood* 2021;**138** Suppl 1:880. [DOI: [10.1182/blood-2021-147177](https://doi.org/10.1182/blood-2021-147177)]

Kanoun S, Berriolo-Riedinger A, Tal I, Edeline V, Cottreau A, Brice P, et al. Prognosis value of baseline total metabolic tumor volume (TMTV) in advanced Hodgkin lymphoma (HL): ancillary study of AHL2011 LYSA trial. *European Journal of Nuclear Medicine and Molecular Imaging* 2016;**43** Suppl 1:S125-6. [PMID: 613164786]

Rossi C, Andre M, Dupuis J, Morschhauser F, Joly B, Lazarovici J, et al. High-risk stage IIB Hodgkin lymphoma treated in the H10 and AHL2011 trials: total metabolic tumor volume is a useful risk factor to stratify patients at baseline. *Haematologica* 2022;**107**(12):2897-904. [PMID: 35638548]

Rossi C, Andre M, Joubert C, Fortpied C, Brice P, Deau-Fisher B, et al. Stage IIB high risk Hodgkin Lymphoma treated in the H10 and AHL2011 trials: similar efficacy of both strategies and prognostic impact of baseline TMTV and PET2 response. *Blood* 2019;**134** Suppl 1:128. [PMID: 2013227008]

Traverse-Glehen A, Damotte D, Dartigues P, Martin L, Parrens M, Ghesquieres H, et al. The combination of high expression of CD68-positive tumor-associated macrophages and PET2-positivity predicts unfavorable outcome in patients with classical Hodgkin lymphoma treated in the LYSA AHL2011 study. *Blood* 2016;**128**(22):1825. [DOI: [10.1182/blood.V128.22.1825.1825](https://doi.org/10.1182/blood.V128.22.1825.1825)]

H10 {published data only}

Aleman BM, Ricardi U, Van Der Maazen RW, Meijnders P, Beijert M, Boros A, et al. Preliminary results of a quality control study on involved node radiotherapy in the EORTC/LYSA/FIL H10 trial on stages I/II Hodgkin lymphoma. *Hematological Oncology* 2019;**37** Suppl 2:487-8. [PMID: 628867652]

* Andre MP, 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. *Journal of Clinical Oncology* 2017;**35**(16):1786-96. [DOI: [10.1200/JCO.2016.68.6394](https://doi.org/10.1200/JCO.2016.68.6394)]

Andre MP, Reman O, Federico M, Brice P, Brusamolino E, Girinsky T, et al. First report on the H10 EORTC/GELA/III 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. [DOI: [10.1182/blood.V114.22.97.97](https://doi.org/10.1182/blood.V114.22.97.97)]

Andre MP, Reman O, Federico M, Girinsky 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**:549.

Andre MP. An update on the EORTC/LYSA/FIL H10 trial. 9th International Symposium on Hodgkin Lymphoma Cologne, 2013, Germany.

Cottreau A, Versari A, LoQ A, Casasnovas O, Bellei M, Ricci R, et al. Deauville score evaluation of PET2 positive patients included in the H10 trial. *European Journal of Nuclear Medicine and Molecular Imaging* 2018;**45** Suppl 1:S439. [DOI: [10.1007/s00259-018-4148-3](https://doi.org/10.1007/s00259-018-4148-3)]

Cottreau A, Versari A, LoQ A, Casasnovas R, Bellei M, Ricci R, et al. Prognostic value of baseline total metabolic tumor volume (TMTV) for patients with early stage Hodgkin lymphoma enrolled in the standard arm of the H10 (EORTC/LYSA/FIL) trial. *Hematological Oncology* 2017;**35** Suppl 2:35-6. [DOI: [10.1002/hon.2437_16](https://doi.org/10.1002/hon.2437_16)]

Cottreau AS, Versari A, LoQ A, Casasnovas 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. *Blood* 2018;**131**(13):1456-63. [DOI: [10.1182/blood-2017-07-795476](https://doi.org/10.1182/blood-2017-07-795476)]

Cottreau AS, Versari A, LoQ A, Casasnovas RO, Bellei M, Ricci R, et al. Prognostic value of baseline quantitative PET metrics for patients with unfavourable early stage Hodgkin lymphoma enrolled in the standard arm of the EORTC/LYSA/FIL H10 Trial. *Blood* 2016;**128**(22):184. [DOI: [10.1182/blood.V128.22.184.184](https://doi.org/10.1182/blood.V128.22.184.184)]

Edeline V, Ruelle C, Auperin A, Celebic A, Bonniaud G, Ferme C, et al. Contribution of PET /CT in the design of the involved node radiotherapy (INRT) concept for patients with localized Hodgkin

lymphoma: preliminary results on 117 patients entered in the H10 trial. *Haematologica* 2010;**95** Suppl 4:S13. [DOI: [10.3324/haematologica.2010.95.s13](https://doi.org/10.3324/haematologica.2010.95.s13)]

Federico M, Fortpied C, Stepanishyna Y, Gotti M, Van der Maazen R, Cristinelli C, et al. Long-term follow-up of the response-adapted intergroup EORTC/LYSA/FIL H10 trial for localized Hodgkin lymphoma. *Journal of Clinical Oncology* 2023;**42**(1):19-25. [DOI: [10.1200/JCO.23.01745](https://doi.org/10.1200/JCO.23.01745)]

Fiaccadori V, Neven A, Fortpied C, Aurer I, Andre M, Federico M, et al. Relapse patterns in early-PET negative, limited-stage Hodgkin lymphoma (HL) after ABVD with or without radiotherapy - a joint analysis of EORTC/LYSA/FIL H10 and NCRI RAPID trials. *British Journal of Haematology* 2023;**200**(6):731-9. [PMID: 36541117]

Filippi AR, Ciammella P, Namsyl-Kaletka A, Botticella A, Todisco L, Fiandra C, et al. Single institution preliminary experience on dose reduction to organs at risk in thoracic radiotherapy for patients enrolled in EORTC-GELA-III H10 study protocol on early stage Hodgkin's Lymphoma. *European Journal of Cancer* 2009;**7** Suppl:158-9. [DOI: [10.1016/S1359-6349\(09\)70541-5](https://doi.org/10.1016/S1359-6349(09)70541-5)]

Gac A, Chartier L, Girinsky T, Gotti M, Lazarovici J, Aurer I, et al. Outcome and treatment of relapsing early PET negative patients included in the EORTC/LYSA/FIL H10 trial on stages I/II Hodgkin lymphoma. *Hematological Oncology* 2019;**37** Suppl 2:294-6. [DOI: [10.1002/hon.104_2630](https://doi.org/10.1002/hon.104_2630)]

Girinsky T, Auperin A, Ribrag V, Elleuch M, Ferme C, Bonniaud G, et al. Role of FDG-PET in the implementation of involved-node radiation therapy for Hodgkin lymphoma patients. *International Journal of Radiation Oncology Biology Physics* 2014;**89**(5):1047-52. [DOI: [10.1016/j.ijrobp.2014.04.026](https://doi.org/10.1016/j.ijrobp.2014.04.026)]

Jullian A, Meignan M, Itti E, Bardet S, Edeline V, Vanderborght T, et al. Comparison of visual and quantitative assessment for FDG-PET interpretation in Hodgkin lymphoma after 2 cycles of ABVD. The EORTC-GELA-III H10 trial experience. *Haematologica* 2010;**95** Suppl 4:49. [DOI: [10.3324/haematologica.2010.95.s49](https://doi.org/10.3324/haematologica.2010.95.s49)]

Phillips EH, Counsell N, Illidge T, Andre M, Aurer I, Fiaccadori V, et al. Baseline maximum tumour diameter is associated with event-free survival for PET-negative patients with limited-stage Hodgkin lymphoma: analysis of the H10 and rapid trials. *Hematological Oncology* 2021;**39** Suppl 2:284-5. [DOI: [10.1002/hon.2880](https://doi.org/10.1002/hon.2880)]

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**(12):1188-94. [PMID: 24637998]

Yeddes I, Meignan M, Bardet S, Vander Borght T, Ricci R, Itti E, et al. Positive interim PET after 2 ABVD is more frequent in patients with unfavourable stage I/II Hodgkin lymphoma and bulky mediastinum than in non-bulky. The EORTC-GELA-III H10 trial experience. *Haematologica* 2010;**95** Suppl 4:49. [DOI: [10.3324/haematologica.2010.95.s49](https://doi.org/10.3324/haematologica.2010.95.s49)]

HD0607 {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** Suppl 1:44. [DOI: [10.1007/s40336-017-0227-x](https://doi.org/10.1007/s40336-017-0227-x)]

Biggi A, Chauvie S, Fallanca F, Guerra L, Bergesio F, Menga M, et al. Predictive value on advanced Hodgkin lymphoma treatment outcome of end-of treatment FDG PET/CT in the HD0607 clinical trial. *Hematological Oncology* 2023;**41**(3):415-23. [DOI: [10.1002/hon.3117](https://doi.org/10.1002/hon.3117)]

Biggi A. Concordance between six Italian centres of PET reporting criteria in the prospective clinical trial HD 0607 in advanced-stage Hodgkin Lymphoma. *European Journal of Nuclear Medicine and Molecular Imaging* 2011;**38** Suppl 2:164. [DOI: [10.1007/s00259-011-1908-8](https://doi.org/10.1007/s00259-011-1908-8)]

Buschiazio A, Bergesio F, Bianchi A, Menga M, Fallanca F, Chauvie S, et al. Diagnostic accuracy of FDG PET/CT at the end of treatment of Hodgkin lymphoma in the HD0607 trial. *Clinical and Translational Imaging* 2019;**7** Suppl 1:84. [DOI: [10.1007/s40336-019-00318-3](https://doi.org/10.1007/s40336-019-00318-3)]

Gallamini A, Biggi A, Chauvie S, Stancu A, Fallanca F, Chiaravallotti A, et al. Interim-PET scan interpretation in the ongoing prospective clinical trial HD 0607, in advanced-stage Hodgkin lymphoma: results of the expert panel review. *Blood* 2010;**116**(21):3891. [DOI: [10.1182/blood.V116.21.3891.3891](https://doi.org/10.1182/blood.V116.21.3891.3891)]

Gallamini A, Rambaldi A, Patti C, Romano A, Viviani S, Bolis S, et al. Baseline metabolic tumor volume and IPS predict ABVD failure in advanced-stage Hodgkin lymphoma with a negative interim PET scan after 2 chemotherapy cycles. A retrospective analysis from the GITIL/FIL HD0607 trial. *Hematological Oncology* 2021;**39** Suppl 2:47-8. [DOI: [10.1002/hon.2879](https://doi.org/10.1002/hon.2879)]

Gallamini A, Rambaldi A, Patti C, Romano A, Viviani S, Silvia B, et al. Lesion dissemination in baseline PET/CT (D-MAX) and IPS score predict ABVD treatment outcome in PET-2 negative advanced-stage Hodgkin lymphoma patients enrolled in the prospective GITIL/FIL HD0607 trial. *Blood* 2021;**138** Suppl 1:2443. [PMID: [10.1182/blood-2021-152292](https://pubmed.ncbi.nlm.nih.gov/3522292/)]

Gallamini A, Rossi A, Patti C, Picardi M, Di Raimondo F, Cantonetti M, et al. Early treatment intensification in advanced-stage high-risk Hodgkin Lymphoma (HL) patients, with a positive FDG-PET scan after two ABVD courses - second interim analysis of the GITIL/FIL HD0607 clinical trial. *Haematologica* 2013;**98** Suppl 2:3. [DOI: [10.3324/haematol.2013.25346](https://doi.org/10.3324/haematol.2013.25346)]

Gallamini A, Rossi A, Patti C, Picardi M, Di Raimondo F, Cantonetti M, et al. Early treatment intensification in advanced-stage high-risk Hodgkin lymphoma (HL) patients, with a positive FDG-PET scan after two ABVD courses-first interim analysis of the GITIL/FIL HD0607 clinical trial. *Blood* 2012;**120**(21):550. [DOI: [10.1182/blood.V120.21.550.550](https://doi.org/10.1182/blood.V120.21.550.550)]

Gallamini A, Rossi A, Patti C, Picardi M, Di Raimondo F, Cantonetti M, et al. Interim PET-adapted chemotherapy in advanced Hodgkin lymphoma (HL). Results of the second

interim analysis of the Italian GITIL/FIL HD0607 trial.

Hematological Oncology 2015;**33**:163-4. [DOI: [10.1002/hon.2227](https://doi.org/10.1002/hon.2227)]

Gallamini A, Rossi A, Patti C, Picardi M, Romano A, Cantonetti M, et al. Consolidation radiotherapy could be omitted in advanced Hodgkin lymphoma with large nodal mass in complete metabolic response after ABVD. Final analysis of the randomized HD0607 trial. *Hematological Oncology* 2019;**37** Suppl 2:147-8. [DOI: [10.1002/hon.105_2629](https://doi.org/10.1002/hon.105_2629)]

* Gallamini A, Rossi A, Patti C, Picardi M, Romano A, Cantonetti M, et al. Consolidation radiotherapy could be safely omitted in advanced Hodgkin lymphoma with large nodal mass in complete metabolic response after ABVD: final analysis of the randomized GITIL/FIL HD0607 trial. *Journal of Clinical Oncology* 2020;**38**(33):3905-13. [DOI: [10.1200/JCO.20.00935](https://doi.org/10.1200/JCO.20.00935)]

Gallamini A, Rossi A, Patti K, Picardi M, Romano A, Cantonetti M, 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. *Haematologica* 2017;**102** Suppl 2:145-6. [DOI: [10.3324/haematol.2017.25346](https://doi.org/10.3324/haematol.2017.25346)]

Gallamini A, Tarella C, Patti C, Gianni A, Bolis S, Trentin L, et al. Multicentre clinical study with early treatment intensification in high-risk Hodgkin lymphoma (HL) patients, with a positive FDG-PET scan after two ABVD courses - GITIL HD0607 study. *Annals of Oncology* 2011;**22** Suppl 4:139. [DOI: [10.1093/annonc/mdr212](https://doi.org/10.1093/annonc/mdr212)]

* 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**:454-62. [DOI: [10.1200/JCO.2017.75.2543](https://doi.org/10.1200/JCO.2017.75.2543)]

Marini C, Bauckneht M, Borra A, Lai R, Donegani MI, Miceli A, et al. Myocardial metabolic response predicts chemotherapy curative potential on Hodgkin Lymphoma: a proof-of-concept study. *Biomedicines* 2021;**9**(8):971. [DOI: [10.3390/biomedicines9080971](https://doi.org/10.3390/biomedicines9080971)]

Romano A, Di Raimondo F, Pavoni C, Tarella C, Viviani S, Rossi A, et al. Prognostic relevance of neutrophil to lymphocyte ratio (NLR) and lymphocyte to monocyte ratio (LMR) in newly diagnosed advanced-stage Hodgkin lymphoma patients treated up-front with a PET-2 risk adapted protocol. *Blood* 2018;**132** Suppl 1:2934. [DOI: [10.1182/blood-2018-99-115346](https://doi.org/10.1182/blood-2018-99-115346)]

Romano A, Pavoni C, Di Raimondo F, Tarella C, Viviani S, Rossi A, et al. The neutrophil to lymphocyte ratio (NLR) and the presence of large nodal mass are independent predictors of early response: a subanalysis of the prospective phase II PET-2-adapted HD0607 trial. *Cancer Medicine* 2020;**9**(23):8735-46. [DOI: [10.1002/cam4.3396](https://doi.org/10.1002/cam4.3396)]

Viviani S, Mazzocchi A, Pavoni C, Taverna F, Guidetti A, Rossi A, et al. Early reduction of serum TARC levels predicts prognosis of patients treated with a pet-adapted strategy for advanced stage Hodgkin lymphoma (HL) in the GITIL/FIL HD0607 trial. *Blood* 2017;**130** Suppl 1:1511. [DOI: [10.1182/blood.V130.Suppl_1.1511.1511](https://doi.org/10.1182/blood.V130.Suppl_1.1511.1511)]

Viviani S, Mazzocchi A, Pavoni C, Taverna F, Rossi A, Patti C, et al. Early serum TARC reduction predicts prognosis in advanced-stage Hodgkin lymphoma patients treated with a PET-adapted strategy. *Hematological Oncology* 2020;**38**:501-8. [DOI: [10.1002/hon.2775](https://doi.org/10.1002/hon.2775)]

HD0801 {published data only}2008-002684-14

Bari A, Marcheselli R, Sacchi S, Re A, Pagani C, Tucci A, et al. The classic prognostic factors in advanced Hodgkin's lymphoma patients are losing their meaning at the time of PET-guided treatments. *Annals of Hematology* 2020;**99**(2):277-82. [DOI: [10.1007/s00277-019-03893-7](https://doi.org/10.1007/s00277-019-03893-7)]

Ricardi U, Levis M, Evangelista A, Gioia DM, Rigacci L, Botto B, et al. Role of consolidation RT to bulky lesions of advanced Hodgkin lymphoma: results of FIL HD0801 trial. *Radiotherapy and Oncology* 2019;**133** Suppl 1:S258-9. [DOI: [10.1016/S0167-8140%2819%2930922-3](https://doi.org/10.1016/S0167-8140%2819%2930922-3)]

* Ricardi U, Levis M, Evangelista A, Gioia DM, Sacchetti GM, Gotti M, et al. Role of radiotherapy to bulky sites of advanced Hodgkin lymphoma treated with ABVD: final results of FIL HD0801 trial. *Blood Advances* 2021;**5**(21):4504-14. [DOI: [10.1182/bloodadvances.2021005150](https://doi.org/10.1182/bloodadvances.2021005150)]

Rigacci L, Puccini B, Broccoli A, Dona M, Gotti M, Evangelista A, et al. Clinical characteristics of interim-PET negative patients with a positive end PET from the prospective HD08-01 FIL study. *Annals of Hematology* 2020;**99**(2):283-91. [DOI: [10.1007/s00277-019-03889-3](https://doi.org/10.1007/s00277-019-03889-3)]

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. [DOI: [10.1002/hon.2437](https://doi.org/10.1002/hon.2437)]

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. [DOI: [10.1200/JCO.2015.63.0699](https://doi.org/10.1200/JCO.2015.63.0699)]

HD16 {published data only}

Baues C, Goergen H, Fuchs M, Kobe C, Dietlein M, Rosenbrock J, et al. Consolidating involved field radiotherapy prevents early and local recurrences in early stage Hodgkin Lymphoma. *International Journal of Radiation Oncology Biology Physics* 2019;**105** Suppl 1:32-3. [DOI: [10.1016/j.ijrobp.2019.06.445](https://doi.org/10.1016/j.ijrobp.2019.06.445)]

Baues C, Goergen H, Fuchs M, Rosenbrock J, Celik E, Eich H, et al. Involved-field radiation therapy prevents recurrences in the early stages of Hodgkin Lymphoma in PET-negative patients after ABVD chemotherapy: relapse analysis of GHSG phase 3 HD16 trial. *International Journal of Radiation Oncology, Biology, Physics* 2021;**111**(4):900-6. [DOI: [10.1016/j.ijrobp.2021.07.1697](https://doi.org/10.1016/j.ijrobp.2021.07.1697)]

Eich HT, Baues C, Fuchs M, Kobe C, Greil R, Sasse S, et al. PET-guided treatment of early-stage favorable Hodgkin Lymphoma: final results of the international, randomized phase 3 trial HD16 by the GHSG. *International Journal of Radiation*

Oncology Biology Physics 2019;**105** Suppl 1:S1. [DOI: [10.1016/j.ijrobp.2019.06.380](https://doi.org/10.1016/j.ijrobp.2019.06.380)]

Eichenauer DA, Buhnen I, Baues C, Kobe C, Kaul H, Greil R, et al. Interim PET-guided treatment for early-stage NLPHL: a subgroup analysis of the randomized GHSG HD16 and HD17 studies. *Blood* 2023;**142**(6):553-60. [PMID: 37257195]

Eichenauer DA, Buhnen I, Fuchs M, Greil R, Moccia A, Zijlstra JM, et al. Interim PET-guided treatment of early-stage nodular lymphocyte-Ppredominant Hodgkin lymphoma: a subgroup analysis of the GHSG HD16 and HD17 studies. *HemaSphere* 2022;**6** Suppl 5:10. [PMID: 639542851]

Eichenauer DA, Buhnen I, Fuchs M, Kaul H, Kobe C, Hartmann S, et al. Treatment of early-stage nodular lymphocyte-predominant Hodgkin Lymphoma: a subgroup analysis of the randomized German Hodgkin Study Group HD16 study. *Blood* 2021;**138** Suppl 1:2439. [DOI: [10.1182/blood-2021-145427](https://doi.org/10.1182/blood-2021-145427)]

* Fuchs M, Goergen H, Kobe C, Kuhnert G, Lohri A, Greil R, et al. Positron emission tomography-guided treatment in early-stage favorable Hodgkin Lymphoma: final results of the international, randomized phase III HD16 trial by the German Hodgkin Study Group. *Journal of Clinical Oncology* 2019;**37**(31):2835-45. [DOI: [10.1200/JCO.19.00964](https://doi.org/10.1200/JCO.19.00964)]

Fuchs M, Gorgen H, Kobe C, Kuhnert G, Hitz F, Greil R, et al. PET positivity after 2 cycles of ABVD is a risk factor in patients with early-stage favorable Hodgkin lymphoma treated in the phase 3 GHSG HD16 study. *Oncology Research and Treatment* 2019;**42**:297. [PMID: CN-02517489]

Fuchs M, Jacob AS, Kaul H, Kobe C, Kuhnert G, Pabst T, et al. Follow-up of the GHSG HD16 trial of PET-guided treatment in early-stage favorable Hodgkin lymphoma. *Leukemia* 2023;**16**:16. [PMID: 37845285]

Fuchs M, Jacob AS, Kaul H, Kobe C, Pabst T, Greil R, et al. PET-guided treatment in patients with early-stage favorable Hodgkin Lymphoma: follow-up analysis of the HD16 trial by the German Hodgkin Study Group. *Blood* 2021;**138** Suppl 1:2438. [DOI: [10.1182/blood-2021-152260](https://doi.org/10.1182/blood-2021-152260)]

Jacob A, Fuchs M, Kaul H, Kobe C, Pabst T, Greil R, et al. PET-guided treatment in patients with early-stage favorable Hodgkin lymphoma: follow-up analysis of the HD16 trial by the German Hodgkin Study Group. *Oncology Research and Treatment* 2022;**45**:167. [PMID: CN-02517589]

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. *Nuklearmedizin [Nuclear Medicine]* 2012;**51** (2):A39-40. [DOI: [10.1055/s-007-36540](https://doi.org/10.1055/s-007-36540)]

Van Heek L, Stuka C, Kaul H, Mettler J, Hitz F, Baues C, et al. Role of baseline metabolic tumor volume in early-stage favorable Hodgkin Lymphoma - data from the prospective, multicenter phase III HD16 trial. *Nuklearmedizin [Nuclear Medicine]* 2022;**61**(2):150-1. [DOI: [10.1055/s-0042-1745979](https://doi.org/10.1055/s-0042-1745979)] [PMID: 638260816]

Van Heek L, Stuka C, Kaul H, Muller H, Mettler J, Hitz F, et al. Predictive value of baseline metabolic tumor volume in early-stage favorable Hodgkin Lymphoma - data from the prospective, multicenter phase III HD16 trial. *HemaSphere* 2022;**6**(5):12.

Van Heek L, Stuka C, Kaul H, Muller H, Mettler J, Hitz F, et al. Predictive value of baseline metabolic tumor volume in early-stage favorable Hodgkin lymphoma - data from the prospective, multicenter phase III HD16 trial. *BMC Cancer* 2022;**22**(1):672. [DOI: [10.1186/s12885-022-09758-z](https://doi.org/10.1186/s12885-022-09758-z)]

HD17 (116)

* Borchmann P, Plutschow A, Kobe C, Greil R, Meissner J, Topp MS, et al. PET-guided omission of radiotherapy in early-stage unfavourable Hodgkin lymphoma (GHSG HD17): a multicentre, open-label, randomised, phase 3 trial. *Lancet Oncology* 2021;**22**(2):223-34. [DOI: [10.1016/S1470-2045\(20\)30601-X](https://doi.org/10.1016/S1470-2045(20)30601-X)]

Borchmann P. Positron emission tomography guided omission of radiotherapy in early-stage unfavorable Hodgkin lymphoma: final results of the international, randomized phase III HD17 trial by the GHSG. *HemaSphere* 2020;**4**:1-2. [DOI: [10.1097/HS9.0000000000000404](https://doi.org/10.1097/HS9.0000000000000404)]

Eichenauer DA, Buhnen I, Baues C, Kobe C, Kaul H, Greil R, et al. Interim PET-guided treatment for early-stage NLPHL: a subgroup analysis of the randomized GHSG HD16 and HD17 studies. *Blood* 2023;**142**(6):553-60. [DOI: [10.1182/blood.20230119939](https://doi.org/10.1182/blood.20230119939)]

Eichenauer DA, Buhnen I, Fuchs M, Greil R, Moccia A, Zijlstra JM, et al. Interim PET-guided treatment of early-stage nodular lymphocyte-predominant Hodgkin lymphoma: a subgroup analysis of the GHSG HD16 and HD17 studies. *HemaSphere* 2022;**6** Suppl 5:10. [PMID: 639542851]

Kriz J, Baues C, Oertel M, Engenhart-Cabillic R, Herfarth K, Lukas P, et al. Quality assurance of involved-node radiotherapy: first results of the HD17-trial of the GHSG. *HemaSphere* 2018;**2** Suppl 3:14. [DOI: [10.1097/01.HS9.0000547879.40964.5c](https://doi.org/10.1097/01.HS9.0000547879.40964.5c)]

Oertel M, Hering D, Baues C, Kittel C, Fuchs M, Kriz J, et al. Radiation doses to mediastinal organs at risk in early-stage unfavorable Hodgkin lymphoma- a risk stratified analysis of the GHSG HD17 trial. *Frontiers in Oncology* 2023;**13**:1183906. [DOI: [10.3389/fonc.2023](https://doi.org/10.3389/fonc.2023)]

Oertel M, Hering D, Nacke N, Kittel C, Kroger K, Kriz J, et al. Estimating the dosimetric benefit of involved node radiotherapy in comparison to involved field radiotherapy - implications from the GHSG HD17 trial. *HemaSphere* 2022;**6** suppl 5:42. [DOI: [10.1097/01.HS9.0000890936.73031.da](https://doi.org/10.1097/01.HS9.0000890936.73031.da)]

Oertel M, Hering D, Nacke N, Kittel C, Kroger K, Kriz J, et al. From involved-field to involved-node - quality analysis of the radiation therapy in HD 17 by the expert panel of the German Hodgkin Study Group. *HemaSphere* 2022;**6** Suppl 5:42-3. [PMID: 639542912]

Oertel M, Hering D, Nacke N, Kittel C, Kröger K, Kriz J, et al. Radiation therapy in the German Hodgkin Study Group HD 16 and HD 17 trials: quality assurance and dosimetric analysis for

Hodgkin lymphoma in the modern era. *Advances in Radiation Oncology* 2023;**8**(3):101169. [DOI: [10.1016/j.adro.2022.101169](https://doi.org/10.1016/j.adro.2022.101169)]

Van Heek L, Weindler J, Gorniak C, Kaul H, Muller H, Mettler J, et al. Prognostic value of baseline metabolic tumor volume (MTV) for forecasting chemotherapy outcome in early-stage unfavorable Hodgkin lymphoma: data from the phase III HD17 trial. *European Journal of Haematology* 2023;**111**(6):881-7. [DOI: [10.1111/ejh.14093](https://doi.org/10.1111/ejh.14093)]

HD18 {published data only}

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. [DOI: [10.1002/hon.2437](https://doi.org/10.1002/hon.2437)]

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** Suppl 2:24-5. [DOI: [10.3324/%25x](https://doi.org/10.3324/%25x)]

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 1:737. [DOI: [10.1182/blood.V130.Suppl_1.737.737](https://doi.org/10.1182/blood.V130.Suppl_1.737.737)]

* 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**:2790-802. [DOI: [10.1016/S0140-6736\(17\)32134-7](https://doi.org/10.1016/S0140-6736(17)32134-7)]

Borchmann P, Haverkamp H, Lohri A, Kreissl S, Greil R, Markova J, et al. Addition of rituximab to BEACOPP escalated 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. [DOI: [10.1182/blood.V124.21.500.500](https://doi.org/10.1182/blood.V124.21.500.500)]

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 BEACOPP escalated 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. [DOI: [10.1016/S1470-2045\(17\)30103-1](https://doi.org/10.1016/S1470-2045(17)30103-1)]

Eichenauer DA, Buhnen I, Kreissl S, Goergen H, Fuchs M, Von Tresckow B, et al. Histopathological growth patterns in patients with advanced nodular lymphocyte-predominant Hodgkin lymphoma treated within the randomized HD18 study: a report from the German Hodgkin Study Group. *British Journal of Haematology* 2021;**196**(1):99-104. [DOI: [10.1111/bjh.17770](https://doi.org/10.1111/bjh.17770)]

Eichenauer DA, Kreissl S, Buhnen I, Baues C, Kobe C, Van Heek L, et al. PET-2-guided escalated BEACOPP for advanced nodular lymphocyte-predominant Hodgkin lymphoma: a subgroup

analysis of the randomized German Hodgkin Study Group HD18 study. *Annals of Oncology* 2021;**32**(6):807-10. [DOI: [10.1016/j.annonc.2021.02.018](https://doi.org/10.1016/j.annonc.2021.02.018)]

Eichenauer DA, Kreissl S, Buhnen I, Gorgen H, Hartmann S, Fuchs M, et al. PET-2-guided escalated BAECOPP for patients with advanced nodular lymphocyte-predominant Hodgkin lymphoma: an analysis from the German Hodgkin Study Group. *HemaSphere* 2020;**4** Suppl 1:68. [DOI: [10.1097/HS9.0000000000000404](https://doi.org/10.1097/HS9.0000000000000404)]

Ferdinandus J, Van Heek L, Roth K, Dietlein M, Eich HT, Baues C, et al. Patterns of PET-positive residual tissue at interim restaging and risk of treatment failure in advanced-stage Hodgkin's lymphoma: an analysis of the randomized phase III HD18 trial by the German Hodgkin Study Group. *European Journal of Nuclear Medicine and Molecular Imaging* 2023;**51**(22):490-5. [DOI: [10.1007/s00259-023-06431-w](https://doi.org/10.1007/s00259-023-06431-w)]

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. [DOI: [10.1182/blood-2018-05-852129](https://doi.org/10.1182/blood-2018-05-852129)]

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 2:312. [DOI: [10.1007/s00259-017-3822-1](https://doi.org/10.1007/s00259-017-3822-1)]

Kreissl S, Goergen H, Buehnen I, Kobe C, Moccia A, Greil R, et al. PET-guided eBAECOPP treatment of advanced-stage Hodgkin lymphoma (HD18): follow-up analysis of an international, open-label, randomised, phase 3 trial. *Lancet Haematology* 2021;**8**(6):398-409. [DOI: [10.1016/S2352-3026\(21\)00101-0](https://doi.org/10.1016/S2352-3026(21)00101-0)]

Kreissl S, Goergen H, Kobe C, Fuchs M, Greil R, Huttman 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 Group. *Oncology Research and Treatment* 2017;**40** Suppl 3:201. [DOI: [10.1159/000479566](https://doi.org/10.1159/000479566)]

Kreissl S, Goergen H, Kobe C, Lohri A, Greil R, Eichenauer D, et al. PET-guided treatment in patients with advanced-stage Hodgkin lymphoma: follow-up analysis of PET-2-negative patients in the HD18 trial by the German Hodgkin Study Group. *HemaSphere* 2020;**4** Suppl 1:69. [DOI: [10.1097/HS9.0000000000000404](https://doi.org/10.1097/HS9.0000000000000404)]

Kreissl S, Voltin CA, Kaul H, Buhnen I, Mettler J, Pabst T, et al. Impact of bone marrow involvement on early positron emission tomography response and progression-free survival in the HD18 trial for patients with advanced-stage Hodgkin lymphoma. *British Journal of Haematology* 2022;**197**(1):5-8. [DOI: [10.1111/bjh.17997](https://doi.org/10.1111/bjh.17997)]

Mettler J, Muller H, Voltin CA, Baues C, Klaeser B, Moccia A, et al. Metabolic tumour volume for response prediction in advanced-stage Hodgkin lymphoma. *Journal of Nuclear Medicine* 2018;**60**(2):207-11. [DOI: [10.2967/jnumed.118.210047](https://doi.org/10.2967/jnumed.118.210047)]

Voltin CA, Kreissl S, Kaul H, Buhnen I, Mettler J, Pabst T, et al. Impact of bone marrow involvement on early PET response and progression-free survival in the HD18 trial for patients with advanced-stage Hodgkin lymphoma. *HemaSphere* 2022;**6** Suppl 5:4. [PMID: 639542805]

Picardi 2007 {published data only}

Picardi M, De Renzo A, Pane F, Nicolai E, Pacelli R, Salvatore M, et al. Randomized comparison of consolidation radiation versus observation in bulky Hodgkin's lymphoma with post-chemotherapy negative positron emission tomography scans. *Leukemia & Lymphoma* 2007;**48**:1721-7. [DOI: [10.1080/10428190701559140](https://doi.org/10.1080/10428190701559140)]

RATHL {published data only}

Abro EU, Law SC, Keane C, Birch S, Sabdia MB, Tobin JW, et al. A critical role for intratumoral and circulating LAG3 in classical Hodgkin lymphoma: analysis from the RATHL prospective phase III international clinical trial. *Blood* 2018;**132** Suppl 1:1621. [DOI: [10.1182/blood-2018-99-112008](https://doi.org/10.1182/blood-2018-99-112008)]

Anderson RA, Remedios R, Kirkwood AA, Patrick P, Stevens L, Clifton-Hadley L, et al. Determinants of ovarian function after response-adapted therapy in patients with advanced Hodgkin's lymphoma (RATHL): a secondary analysis of a randomised phase 3 trial. *Lancet Oncology* 2018;**19**(10):1328-37. [DOI: [10.1016/S1470-2045\(18\)30500-X](https://doi.org/10.1016/S1470-2045(18)30500-X)]

Barrington SF, Kirkwood AA, Franceschetto A, Fulham MJ, Roberts TH, Almquist H, et al. PET-CT for staging and early response: results from the Response-Adapted Therapy in Advanced Hodgkin Lymphoma study. *Blood* 2016;**127**(12):1531-8. [DOI: [10.1182/blood-2015-11-679407](https://doi.org/10.1182/blood-2015-11-679407)]

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;**2**:115. [DOI: [10.1007/s00259-011-1908-8](https://doi.org/10.1007/s00259-011-1908-8)]

Barrington SF, O'Doherty MJ, Roberts TH, Kirkwood A, Viney ZN, Franceschetto A, et al. PET-CT for staging and early response - results from the response adapted therapy in advanced Hodgkin lymphoma (RATHL) study. *Hematological Oncology* 2013;**1**:101-2. [DOI: [10.1002/hon.2057](https://doi.org/10.1002/hon.2057)]

Burton CH, Scott D, Kirkwood AA, Gandhi M, Radford J, Barrington S, et al. Application of a gene expression-based model in combination with FDG-PET imaging to predict treatment response in advanced Hodgkin lymphoma in the RATHL study (CRUK/07/033). *Hematological Oncology* 2017;**35** Suppl 2:91-2. [DOI: [10.1002/hon.2437](https://doi.org/10.1002/hon.2437)]

Hague C, Johnson P, Federico M, Kirkwood A, Fossa A, O'Doherty M, et al. Pulmonary function and grade 3/4 clinical events in PET negative patients taking part in the international RATHL trial (CRUK/07/033): a comparison of 12 vs 4 doses of bleomycin. *Hematological Oncology* 2015;**33** Suppl 1:121. [DOI: [10.1002/hon.2227](https://doi.org/10.1002/hon.2227)]

Johnson P, Federico M, Fossa A, Barrington S, Kirkwood A, Roberts T, et al. Response rates and toxicity of response-adapted therapy in advanced Hodgkin lymphoma: initial

results. From the international RATHL study. *Haematologica* 2013;**2**:2. [DOI: [10.3324/hon.2057](https://doi.org/10.3324/hon.2057)]

Johnson P, Federico M, Fossa A, O'Doherty M, Roberts T, Stevens L, et al. Responses and chemotherapy dose adjustment determined by PET-CT imaging: first results from the international response adapted therapy in advanced Hodgkin lymphoma (RATHL) study. *Hematological Oncology* 2013;**31** Suppl 1:138. [DOI: [10.1002/hon.2057](https://doi.org/10.1002/hon.2057)]

* 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. [DOI: [10.1056/NEJMoa1510093](https://doi.org/10.1056/NEJMoa1510093)]

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 de-escalation and efficacy of escalation in the international RATHL study (CRUK/07/033). *Hematological Oncology* 2015;**33** Suppl 1:102-3. [DOI: [10.1002/hon.2227](https://doi.org/10.1002/hon.2227)]

Lim SH, Burton C, Barrans S, Kirkwood A, Patrick P, Stevens L, et al. Transcriptional networks associated with treatment failure in advanced-stage Hodgkin lymphoma: data from the RATHL trial (CRUK/07/033). *Hematological Oncology* 2019;**37** Suppl 2:293-4. [DOI: [10.1002/hon.103_2630](https://doi.org/10.1002/hon.103_2630)]

Luminari S, Fossa A, Trotman J, Molin D, D'Amore F, Enblad G, et al. Long-term follow-up of the response-adjusted therapy for advanced Hodgkin lymphoma trial. *Journal of Clinical Oncology* 2023;**42**(1):13-8. [DOI: [10.1200/JCO.23.01177](https://doi.org/10.1200/JCO.23.01177)]

Pike LC, Kirkwood AA, Patrick P, Radford J, Burton C, Stevens L, et al. Can baseline PET-CT features predict outcomes in advanced Hodgkin lymphoma? A prospective evaluation of UK patients in the RATHL trial (CRUK/07/033). *Hematological Oncology* 2017;**35** Suppl 2:37-8. [DOI: [10.1002/hon.2437_18](https://doi.org/10.1002/hon.2437_18)]

Preston A, Kirkwood A, Federico M, Trotman J, Fossa A, Stevens L, et al. Outcomes of treatment for obese patients with advanced Hodgkin lymphoma in the RATHL Trial (CRUK/07/033). *Journal of Oncology Pharmacy Practice* 2017;**23** Suppl 8:51. [DOI: [10.1177/1078155217729776](https://doi.org/10.1177/1078155217729776)]

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** Suppl 2:65-7. [DOI: [10.1002/hon.2437](https://doi.org/10.1002/hon.2437)]

UK NCRI RAPID {published data only}

Aurer I, Neven A, Fiaccadori V, Counsell N, Phillips E, Clifton-Hadley L, et al. Relapses in interim PET negative limited-stage Hodgkin lymphoma patients receiving ABVD with or without radiotherapy - analysis of EORTC/FIL/LYSA H10 and UK NCRI rapid trials. *Hematological Oncology* 2021;**39** Suppl 2:119-20. [DOI: [10.1002/hon.2879](https://doi.org/10.1002/hon.2879)]

Barrington S, Phillips EH, Counsell N, Hancock B, Pettengell R, Johnson P, et al. Treatment escalation in patients with early stage Hodgkin lymphoma and a positive pet scan after initial chemotherapy is not always required: subsidiary analysis of the

UK NCRI rapid study. *British Journal of Haematology* 2018;**181** Suppl 1:24. [DOI: [10.1111/bjh.15226](https://doi.org/10.1111/bjh.15226)]

Barrington SF, Phillips EH, Counsell N, Hancock B, Pettengell R, Johnson P, et al. Positron emission tomography score has greater prognostic significance than pretreatment risk stratification in early-stage Hodgkin lymphoma in the UK RAPID study. *Journal of Clinical Oncology* 2019;**37**(20):1732-41. [DOI: [10.1200/JCO.18.01799](https://doi.org/10.1200/JCO.18.01799)]

Cutter DJ, Diez P, Illidge T, Buckle A, Popova B, Darby SC, et al. Cardiac radiation dose and predicted cardiac mortality in the UK NCRI rapid trial in early stage Hodgkin lymphoma. *Hematological Oncology* 2015;**33**:192. [DOI: [10.1002/hon.2227](https://doi.org/10.1002/hon.2227)]

Cutter DJ, Ramroth J, Diez P, Buckle A, Kennedy D, Popova B, et al. Cardiovascular radiation dosimetry and predicted cardiovascular risks from involved field radiotherapy within the UK 'RAPID' trial in early low-risk Hodgkin lymphoma. *HemaSphere* 2018;**2** Suppl 1:13. [DOI: [10.1097/01.HS9.0000547877.95222.2e](https://doi.org/10.1097/01.HS9.0000547877.95222.2e)]

Cutter DJ, Ramroth J, Diez P, Buckle A, Ntents G, Popova B, et al. Predicted risks of cardiovascular disease following chemotherapy and radiotherapy in the UK NCRI RAPID trial of Positron Emission Tomography-directed therapy for early-stage Hodgkin Lymphoma. *Journal of Clinical Oncology* 2021;**39**(32):3591-601. [DOI: [10.1200/JCO.21.00408](https://doi.org/10.1200/JCO.21.00408)]

Fiaccadori V, Neven A, Fortpied C, Aurer I, Andre M, Federico M, et al. Relapse patterns in early-PET negative, limited-stage Hodgkin lymphoma (HL) after ABVD with or without radiotherapy - a joint analysis of EORTC/LYSA/FIL H10 and NCRI RAPID trials. *British Journal of Haematology* 2023;**200**(6):731-9. [DOI: [10.1111/bjh.18594](https://doi.org/10.1111/bjh.18594)]

Georgi T, Kurch L, Hasenclever D, Warbey V, Radford J, Sabri O, et al. Quantitative assessment of interim PET in Hodgkin lymphoma - an evaluation of the qPET method in adult patients in the RAPID trial. *Journal of Nuclear Medicine* 2019;**60** Suppl 1:140. [https://jnm.snmjournals.org/content/60/supplement_1/140]

Georgi TW, Kurch L, Hasenclever D, Warbey VS, Pike L, Radford J, et al. Quantitative assessment of interim PET in Hodgkin lymphoma: an evaluation of the qPET method in adult patients in the RAPID trial. *PLoS One* 2020;**15**(4):e0231027. [DOI: [10.1371/journal.pone.0231027](https://doi.org/10.1371/journal.pone.0231027)]

Illidge T, Phillips E, Counsell N, Pettengell R, Johnson P, Culligan D, et al. Maximum tumour dimension at baseline is associated with event-free survival in pet negative patients with stage IA/IIA Hodgkin lymphoma in the UK NCRI rapid trial. *HemaSphere* 2018;**2** Suppl 3:10. [DOI: [10.1097/01.HS9.0000547870.79974.6d](https://doi.org/10.1097/01.HS9.0000547870.79974.6d)]

Illidge TM, Phillips EH, Counsell N, Pettengell R, Johnson PW, Culligan DJ, et al. Maximum tumor diameter is associated with event-free survival in PET-negative patients with stage I/IIA Hodgkin lymphoma. *Blood Advances* 2020;**4**(1):203-6. [DOI: [10.1182/bloodadvances.2019001023](https://doi.org/10.1182/bloodadvances.2019001023)]

Linton KM, Oladipo JM, Counsell N, Taylor J, Burton C, Oakley R, et al. Enhanced outcome prediction in early stage

classical Hodgkin Lymphoma using pre-treatment biomarkers and interim PET (BioPET); a sub-analysis of the UK NCRI RAPID trial. *Blood* 2020;**136** Suppl 1:18-9. [DOI: [10.1182/blood-2020-136318](https://doi.org/10.1182/blood-2020-136318)]

Phillips EH, Counsell N, Illidge T, Andre M, Aurer I, Fiaccadori V, et al. Baseline maximum tumour diameter is associated with event-free survival for PET-negative patients with limited-stage Hodgkin lymphoma: analysis of the H10 and rapid trials. *Haematological Oncology* 2021;**39** Suppl 2:284-5. [DOI: [10.1002/hon.2880](https://doi.org/10.1002/hon.2880)]

Radford J, Barrington S, Counsell N, Pettengell R, Johnson P, Wimperis J, et al. Involved field radiotherapy versus no further treatment in patients with clinical stages IA and IIA Hodgkin lymphoma and a 'negative' PET scan after 3 cycles ABVD. Results of the UK NCRI RAPID trial. *Blood* 2012;**120** (21):547. [DOI: [10.1182/blood.V120.21.547.547](https://doi.org/10.1182/blood.V120.21.547.547)]

Radford J, Barrington S, Counsell N, Pettengell R, Johnson P, Wimperis J, et al. Prognostic performance of pre-treatment EORTC, GHSG and IPI risk factors and post-chemotherapy PET response in the UK NCRI RAPID trial in early stage Hodgkin Lymphoma (HL). *Haematologica* 2013;**98** Suppl 2:13. [DOI: [10.3324/hon.25x](https://doi.org/10.3324/hon.25x)]

Radford J, Illidge T, Counsell N, Hancock B, Pettengell R, Johnson P, et al. PET score following 3 cycles ABVD has greater prognostic value than pre-treatment risk stratification in the rapid trial in early stage Hodgkin lymphoma (HL). *Hematological Oncology* 2015;**33** Suppl 1:144. [DOI: [10.1002/hon.2227](https://doi.org/10.1002/hon.2227)]

* 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**(17):1598-607. [DOI: [10.1056/NEJMoa1408648](https://doi.org/10.1056/NEJMoa1408648)]

Radford J, O'Doherty M, Barrington S, Qian W, Patrick P, Coltart S, et al. Results of the 2nd planned interim analysis of the Rapid trial (involved field radiotherapy versus no further treatment) in patients with clinical stages 1a and 2a Hodgkin lymphoma and a 'negative' FDG-PET scan after 3 cycles ABVD. *Blood* 2008;**112** (11):143-4. [DOI: [10.1182/blood.V112.11.369.369](https://doi.org/10.1182/blood.V112.11.369.369)]

Radford J, O'Doherty M, Barrington S, Qian W, Popova B, Pettengell R, et al. Results of the 3rd planned interim analysis of the UK NCRI Rapid trial (involved field radiotherapy versus no further treatment) in patients with clinical stages IA/IIA Hodgkin lymphoma and a 'negative' 18 FDG-PET scan after 3 cycles ABVD. *Haematologica* 2010;**95** Suppl 4:16-7. [DOI: [10.3324/hon.25x](https://doi.org/10.3324/hon.25x)]

Radford JA, Barrington SF, O'Doherty MJ, Qian W, Mouncey P, Pettengell R, et al. Interim results of a UK NCRI randomized trial comparing involved field radiotherapy with no further treatment after 3 cycles ABVD and a negative pet scan in clinical stages IA/IIA Hodgkin lymphoma. *Haematologica* 2007;**92**:32.

References to studies excluded from this review

AHOD0031 {published data only}

* Friedman DL, Chen L, Wolden S, Buxton A, McCarten K, FitzGerald TJ, et al. Dose-intensive response-based chemotherapy and radiation therapy for children and adolescents with newly diagnosed intermediate-risk Hodgkin lymphoma: a report from the Children's Oncology Group study AHOD0031. *Journal of Clinical Oncology* 2014;**32**(32):3651-8. [DOI: [10.1200/JCO.2013.52.5410](https://doi.org/10.1200/JCO.2013.52.5410)]

NCT00025259. Chemotherapy with or without additional chemotherapy and/or radiation therapy in treating children with newly diagnosed Hodgkin's disease. <https://clinicaltrials.gov/study/NCT00025259> (first submitted 11 October 2001).

CALGB 50604 {published data only}

NCT01132807. Chemotherapy based on positron emission tomography scan in treating patients with stage I or stage II Hodgkin lymphoma. <https://clinicaltrials.gov/study/NCT01132807> (first submitted 27 May 2010).

Straus D, Pitcher B, Kostakoglu L, Grecula JC, Hsi ED, Schoder H, et al. Results of US intergroup trial of response-adapted chemotherapy or chemotherapy/radiation therapy based on pet for non-bulky stage I and II Hodgkin lymphoma (CALGB/Alliance 50604). *Haematologica* 2016;**101** Suppl 5:13.

* 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.

Straus DJ, Pitcher B, Kostakoglu L, Grecula JC, Hsi ED, Schoder H, et al. Initial results of US intergroup trial of response-adapted chemotherapy or chemotherapy/radiation therapy based on PET for non-bulky stage I and II Hodgkin lymphoma (HL) (CALGB/alliance 50604). *Blood* 2015;**126** (23):578.

CALGB50801 {published data only}

* LaCasce AS, Dockter T, Ruppert AS, Kostakoglu L, Schoder H, Hsi E, et al. Positron Emission Tomography-adapted therapy in bulky stage I/II classic Hodgkin lymphoma: CALGB 50801 (Alliance). *Journal of Clinical Oncology* 2023;**41**(5):1023-34. [DOI: [10.1200/JCO.22.00947](https://doi.org/10.1200/JCO.22.00947)]

NCT01118026. Response-based therapy assessed by PET scan in treating patients with bulky stage I and stage II classical Hodgkin lymphoma. <https://ClinicalTrials.gov/show/NCT01118026> (first submitted 5 May 2010).

ECHELON-1 {published data only}

Ansell SM, Connors JM, Radford JA, Kim WS, Gallamini A, Ramchandren R, et al. First-line brentuximab vedotin plus chemotherapy to improve overall survival in patients with stage III/IV classical Hodgkin lymphoma: an updated analysis of ECHELON-1. *Journal of Clinical Oncology* 2022;**40**(16):7503. [DOI: [10.1200/JCO.2022.40.16_suppl.7503](https://doi.org/10.1200/JCO.2022.40.16_suppl.7503)]

Ansell SM, Younes A, Connors JM, Gallamini A, Kim WS, Friedberg JW, et al. Phase 3 study of brentuximab vedotin

plus doxorubicin, vinblastine, and dacarbazine (A+AVD) versus doxorubicin, bleomycin, vinblastine, and dacarbazine (ABVD) as front-line treatment for advanced classical Hodgkin lymphoma (HL): Echelon-1 study. *Journal of Clinical Oncology* 2014;**32**(Suppl 15):TPS8613.

Bartlett NL, Straus DJ, Dlugosz-Danecka M, Alekseev S, Illes A, Lech-Maranda E, et al. Brentuximab Vedotin with chemotherapy for stage 3/4 classical Hodgkin lymphoma (cHL): 4-Year update of the Echelon-1 study. *Blood* 2019;**134** Suppl 1:4026. [DOI: [10.1182/blood-2019-124116](https://doi.org/10.1182/blood-2019-124116)]

Connors JM, Jurczak W, Straus DJ, Ansell SM, Kim WS, Center M, et al. Brentuximab vedotin plus doxorubicin, vinblastine, dacarbazine (A+AVD) as frontline therapy demonstrates superior modified progression-free survival versus ABVD in patients with previously untreated stage III or IV Hodgkin lymphoma (HL): the phase 3 Echelon-1 study. *Blood* 2017;**130**(Suppl 1):6. [DOI: [10.1182/blood.V130.Suppl_1.6.6](https://doi.org/10.1182/blood.V130.Suppl_1.6.6)]

Crosswell HE, LaCasce AS, Bartlett NL, Straus DJ, Savage KJ, Zinzani PL, et al. Brentuximab vedotin with chemotherapy in adolescents and young adults (AYAS) with stage III or IV Hodgkin lymphoma: a subgroup analysis from the phase 3 ECHELON-1 study. *Hematological Oncology* 2021;**39** Suppl 2:76-7. [DOI: [10.1002/hon.2879](https://doi.org/10.1002/hon.2879)]

Crosswell HE, LaCasce AS, Bartlett NL, Straus DJ, Savage KJ, Zinzani PL, et al. Brentuximab vedotin with chemotherapy in adolescents and young adults (AYAs) with stage III or IV Hodgkin lymphoma: a subgroup analysis from the phase 3 Echelon-1 study. *Journal of Clinical Oncology* 2021;**39**(Suppl 15):7528. [DOI: [10.1200/JCO.2021.39.15_suppl.7528](https://doi.org/10.1200/JCO.2021.39.15_suppl.7528)]

Evens AM, Connors JM, Younes A, Ansell SM, Kim WS, Radford J, et al. Older patients (aged ≥ 60 years) with previously untreated advanced-stage classical Hodgkin lymphoma: detailed study analysis from the phase III ECHELON-1 study. *Haematologica* 2022;**107** (5):1086-94. [DOI: [10.3324/haematol.2021.278438](https://doi.org/10.3324/haematol.2021.278438)]

Friedman J, Lee HJ, Ho L, Flinn IW. Brentuximab Vedotin in combination with nivolumab, doxorubicin, and dacarbazine in newly diagnosed patients with advanced stage Hodgkin lymphoma. *Blood* 2019;**134** Suppl 1:2836. [DOI: [10.1182/blood-2019-122969](https://doi.org/10.1182/blood-2019-122969)]

Gallamini A, Radford J, Jurczak W, Straus D, Ansell S, Kim WS, et al. Frontline brentuximab vedotin plus chemotherapy exhibits superior modified progression-free survival vs chemotherapy alone in patients with stage III or IV Hodgkin lymphoma: phase 3 Echelon-1 study. *HemaSphere* 2018;**2** Suppl 2:535. [DOI: [10.1097/HS9.000000000000060](https://doi.org/10.1097/HS9.000000000000060)]

Grigg AP, Lacasce AS, Bartlett NL, Straus DJ, Savage KJ, Zinzani PL, et al. Brentuximab vedotin with chemotherapy in adolescents and young adults (AYAS) with stage iii or iv Hodgkin lymphoma: a subgroup analysis from the phase 3 Echelon-1 study. *HemaSphere* 2021;**5** Suppl 2:351-2. [DOI: [10.1097/HS9.0000000000000566](https://doi.org/10.1097/HS9.0000000000000566)]

Hutchings M, Ansell SM, Straus DJ, Connors JM, Kim WS, Gallamini A, et al. Improved overall survival with first-line brentimab vedotin plus chemotherapy in patients with stage III/IV classical Hodgkin lymphoma: 6-year

analysis of ECHELON-1. *HemaSphere* 2022;**6**:207-8. [DOI: [10.1097/01.HS9.0000852292.38263.b8](https://doi.org/10.1097/01.HS9.0000852292.38263.b8)]

Hutchings M, Radford J, Dlugosz-Danecka M, Connors JM, Illes A, Picardi M, et al. Brentuximab vedotin with chemotherapy for patients with previously untreated, stage III or IV classical Hodgkin Lymphoma: a 5-year update of the phase 3 Echelon-1 study (NCT01712490). *HemaSphere* 2021;**5** Suppl 2:59. [DOI: [10.1097/HS9.0000000000000566](https://doi.org/10.1097/HS9.0000000000000566)]

Phillips T, Migliaccio-Walle K, Yu K, Bloudek B, Liu N, Burke J. An oncology simulation model to estimate 10-year overall survival, progression-free survival and stem cell transplantation for frontline, stage III or IV classical Hodgkin lymphoma based on overall survival update of the ECHELON-1 trial: a United States perspective. *Clinical Lymphoma, Myeloma & Leukemia* 2022;**22** Suppl 2:349-50. [DOI: [10.1016/S2152-2650%2822%2901482-3](https://doi.org/10.1016/S2152-2650%2822%2901482-3)]

Phillips T, Migliaccio-Walle K, Yu KS, Bloudek B, Liu N, Fanale M, et al. Estimating long-term outcomes in classic Hodgkin lymphoma: a United States population-based oncology simulation model based on overall survival from the ECHELON-1 trial. *Leukemia & Lymphoma* 2023;**64**(5):1017-25. [DOI: [10.1080/10428194.2023.2193854](https://doi.org/10.1080/10428194.2023.2193854)]

Phillips T, Migliaccio-Walle K, Yu KS, Bloudek B, Liu N, Fanale MA, et al. An oncology simulation model to estimate 10-year progression-free survival and stem cell transplantation for frontline, stage III or IV classical Hodgkin lymphoma based on the 5-year update of the ECHELON-1 trial: a United States perspective. *Blood* 2021;**138**:2440. [DOI: [10.1182/blood-2021-147308](https://doi.org/10.1182/blood-2021-147308)]

Picardi M, Straus DJ, Dlugosz-Danecka M, Alekseev S, Illes A, Lech-Maranda E, et al. Brentuximab vedotin with chemotherapy for stage III/IV classical Hodgkin lymphoma: 4-year update of the ECHELON-1 study. *HemaSphere* 2020;**4** Suppl 1:526. [DOI: [10.1097/HS9.0000000000000404](https://doi.org/10.1097/HS9.0000000000000404)]

Radford J, Bartlett N, Straus D, Dlugosz-Danecka M, Alekseev S, Illes A, et al. Brentuximab vedotin plus chemotherapy for stage III/IV classical Hodgkin lymphoma: 4-year update of the ECHELON-1 study. *British Journal of Haematology* 2020;**189** Suppl 1:101-2. [DOI: [10.1111/bjh.16638](https://doi.org/10.1111/bjh.16638)]

Radford J, Dugosz-Danecka M, Connors JM, Ills R, Picardi M, Lech-Maranda E, et al. Brentuximab vedotin plus chemotherapy for patients with previously untreated, stage III or IV classical Hodgkin lymphoma: 5-year update of the phase 3 ECHELON-1 study (NCT01712490). *British Journal of Haematology* 2021;**193** Suppl 1:145-7. [DOI: [10.1111/bjh.17492](https://doi.org/10.1111/bjh.17492)]

Ramchandren R, Dfugosz-Danecka M, Connors JM, Radford J, Illes A, Picardi M, et al. HL-147: brentuximab vedotin with chemotherapy for patients with previously untreated stage III/IV classical Hodgkin lymphoma: 5-year update of the ECHELON-1 study. *Clinical Lymphoma, Myeloma & Leukemia* 2021;**21**:371. [DOI: [10.1016/S2152-2650\(21\)01849-8](https://doi.org/10.1016/S2152-2650(21)01849-8)]

Steiner RE, Becnel MR, Cuglievan B, Nair R, Singh P, Strati P, et al. Brentuximab vedotin with chemotherapy in frontline treatment of classic Hodgkin lymphoma nodular sclerosis syncytial variant. *Blood* 2020;**136** Suppl 1:28-9. [DOI: [10.1182/blood-2020-138996](https://doi.org/10.1182/blood-2020-138996)]

Straus DJ, Dlugosz-Danecka M, Alekseev S, Illes A, Picardi M, Lech-Maranda E, et al. Brentuximab vedotin with chemotherapy for stage 3/4 classical Hodgkin lymphoma: three-year update of the ECHELON-1 study. *Journal of Clinical Oncology* 2019;**37**(Suppl 15):7532. [DOI: [10.1200/JCO.2019.37.15_suppl.7532](https://doi.org/10.1200/JCO.2019.37.15_suppl.7532)]

Straus DJ, Dlugosz-Danecka M, Alekseev S, Illes A, Picardi M, Lech-Maranda E, et al. Brentuximab vedotin with chemotherapy for stage III/IV classical Hodgkin lymphoma: 3-year update of the ECHELON-1 study. *Blood* 2020;**135**(10):735-42. [DOI: [10.1182/blood.2019003127](https://doi.org/10.1182/blood.2019003127)]

* Straus DJ, Dlugosz-Danecka M, Connors JM, Alekseev S, Illes A, Picardi M, et al. Brentuximab vedotin with chemotherapy for stage III or IV classical Hodgkin lymphoma (ECHELON-1): 5-year update of an international, open-label, randomised, phase 3 trial. *Lancet Haematology* 2021;**8**(6):410-21. [DOI: [10.1016/S2352-3026\(21\)00102-2](https://doi.org/10.1016/S2352-3026(21)00102-2)]

Straus DJ, Dlugosz-Danecka M, Connors JM, Illes A, Picardi M, Lech-Maranda E, et al. Brentuximab vedotin with chemotherapy for patients with previously untreated, stage III/IV classical Hodgkin lymphoma: 5-year update of the ECHELON-1 study. *Blood* 2020;**136** Suppl 1:26-8. [DOI: [10.1182/blood-2020-137089](https://doi.org/10.1182/blood-2020-137089)]

Sureda A, Connors JM, Younes A, Gallamini A, Ansell SM, Kim WS, et al. Serum sCD30 and Tarc do not correlate with pet-based response assessment in patients (PTS) with stage III or IV classical Hodgkin lymphoma (CHL): Phase 3 ECHELON-1 study of brentuximab vedotin plus chemotherapy vs chemotherapy alone. *HemaSphere* 2018;**2** Suppl 3:35-6. [DOI: [10.1097/01.HS9.0000547933.40071.ae](https://doi.org/10.1097/01.HS9.0000547933.40071.ae)]

EORTC-COBRA {published data only}

Diepstra A, Visser L, Fortpied C, Noordzij W, LoQ A, Arens A, et al. FDG-PET and serum Tarc levels after one cycle of Bv-AVD in advanced stage Hodgkin lymphoma patients: results from the very early PET-response adapted EORTC-COBRA trial. *HemaSphere* 2022;**6** Suppl 5:1.

NCT03517137. Very early PET-response adapted targeted therapy for advanced Hodgkin lymphoma: a single -arm phase II study. <https://clinicaltrials.gov/study/NCT03517137> (first submitted 24 April 2018).

FIL {published data only}2016-002509-21-IT

EUCTR2016-002509-21-IT. A randomized, open-label, multicenter, phase III, 2-arm study comparing efficacy and tolerability of the intensified variant 'dose-dense/dose-intense ABVD' (ABVD DD-DI) with an interim PET response-adapted ABVD program as upfront therapy in advanced-stage classical Hodgkin Lymphoma (HL) [Studio randomizzato, open-label, multicentrico, di fase III a 2 bracci di confronto dell'eMicacia e della tollerabilità della variante intensificata 'dose-dense/dose-intense ABVD' (ABVD DD-DI) e un programma terapeutico con ABVD a dosi standard per 2 cicli e successivamente orientato in base alla risposta PET, come trattamento di prima linea di pazienti con Linfoma di Hodgkin classico (HL) in stadio avanzato]. <https://www.clinicaltrialsregister.eu/ctr-search/trial/2016-002509-21/IT> (first posted 2017).

NCT03159897. FIL study on ABVD DD-DI as upfront therapy in HL. <https://clinicaltrials.gov/show/NCT03159897> (first submitted 18 May 2017).

GATLA LH-05 {published data only}

Ciliberti E, Fernandez I, Kurgansky N, Prates V, Zoppegno L, Negri P, et al. Safety and efficacy analysis of elderly patients treated within the GATLA HL-05 clinical trial: PET adapted therapy after 3 cycles of ABVD for all stages of Hodgkin lymphoma. *Hematological Oncology* 2017;**35** Suppl 2:168-9. [DOI: [10.1002/hon.2438](https://doi.org/10.1002/hon.2438)]

Iorio J, Eleta M, Plates V, Fiad L, Riddick M, Pavlovsky A. MTV, TLG and Suv max as promising parameters to predict early response in patients with HL treated with ABVD. A retrospective sub-analysis of the GATLA LH-05 trial. *Hematological Oncology* 2021;**39** Suppl 2:283-4. [DOI: [10.1002/hon.2880](https://doi.org/10.1002/hon.2880)]

Pavlovsky A, Fernandez I, Kurgansky N, Prates V, Zoppegno L, Negri P, et al. PET-CT adapted therapy after 3 cycles of ABVD to all stages of Hodgkin lymphoma. GATLA trial HL-05. *Haematologica* 2016;**101** Suppl 1:5. [PMID: 615548899]

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. [DOI: [10.1111/bjh.15838](https://doi.org/10.1111/bjh.15838)]

Pavlovsky A, Fiad L, Fernandez I, Prates V, Kurgansky N, Cerutti A, et al. PET-adapted therapy after three cycles of ABVD for all stages of Hodgkin lymphoma: long term follow up of the GATLA Lh-05 trial. *Hematological Oncology* 2021;**39** Suppl 3:276-7. [DOI: [10.1002/hon.2880](https://doi.org/10.1002/hon.2880)]

Pavlovsky A, Fiad NL, Prates MV, Fernandez I, Kurgansky N, Cerutti A, et al. PET-adapted therapy after three cycles of ABVD for all stages of Hodgkin lymphoma: long term follow up of the GATLA LH-05 trial. *HemaSphere* 2022;**6** Suppl 3:1870-1. [DOI: [10.1097/01.HS9.0000852292.38263.b8](https://doi.org/10.1097/01.HS9.0000852292.38263.b8)]

H11 {published data only}2011-005473-22

NCT01652261. Very early FDG-PET/CT-response adapted therapy for advanced Hodgkin lymphoma (H11). <https://clinicaltrials.gov/search?term=NCT01652261> (first received 2012).

HD15 {published data only}

Buerkle C, Goergen H, Muller H, Brockelmann P, Kreissl S, Fuchs M, et al. Patient preferences - survey of Hodgkin lymphoma survivors on treatment-associated burden and side effects. *Haematologica* 2016;**101** Suppl 1:146-7.

Eich HT, Kobe C, Dietlein M, Kriz J, Haverkamp H, Fuchs M, et al. Role of radiotherapy after assessment of residual bulky tumor using FDG-PET in patients with advanced-stage Hodgkin lymphoma: final report of the GHSG HD15 trial. *International Journal of Radiation Oncology, Biology, Physics* 2011;**1**:2-3.

Eich HT, Kobe C, Kriz J, Baues C, Haverkamp H, Fuchs M, et al. Assessment of residual bulky tumor using FDG-PET in patients with advanced-stage Hodgkin Lymphoma after completion of chemotherapy: final report of the GHSG HD15

trial. *Strahlentherapie und Onkologie* 2011;**1**:48. [DOI: [10.1007/s00066-011-1001-z](https://doi.org/10.1007/s00066-011-1001-z)]

Eich HT, Kriz J, Reinartz G, Kobe C, Kuhnert G, Haverkamp U, et al. Relapse analysis after radiation therapy of PET-positive residual tumors of patients with advanced stage Hodgkin lymphoma treated in the HD15 trial of the German Hodgkin Study Group (GHSG). *International Journal of Radiation Oncology, Biology, Physics* 2013;**1**:5-6. [DOI: [10.1016/j.ijrobp.2013.06.021](https://doi.org/10.1016/j.ijrobp.2013.06.021)]

Eichenauer DA, Haverkamp H, Behringer K, Halbsguth T, Fuchs M, Diehl V, et al. Secondary MDS/AML in Hodgkin lymphoma patients treated within German Hodgkin Study Group (GHSG) clinical trials after introduction of the BEACOPP protocol. *Blood* 2010;**116**:21. [PMID: 70775304]

Engert A, Goergen H, Markova J, Pabst T, Meissner J, Zijlstra JM, et al. Reduced-intensity chemotherapy in patients with advanced-stage Hodgkin lymphoma: updated results of the open-label, international, randomised phase 3 HD15 trial by the German Hodgkin Study Group. *HemaSphere* 2017;**1**(1):e5. [DOI: [10.1097/HS9.0000000000000005](https://doi.org/10.1097/HS9.0000000000000005)]

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. [DOI: [10.1016/S0140-6736%2811%2961940-5](https://doi.org/10.1016/S0140-6736%2811%2961940-5)]

Engert A, Haverkamp H, Kobe C, Markova J, Renner C, Ho AD, et al. Reduced intensity of chemotherapy and PET-guided radiotherapy in patients with advanced stage Hodgkin lymphoma: the GHSG HD15 final results. *Blood* 2011;**118**:21.

Engert A, Kobe C, Markova J, Haverkamp H, Borchmann P, Hitz F, et al. Assessment of residual bulky tumor using FDG-PET in patients with advanced-stage Hodgkin lymphoma after completion of chemotherapy: final report of the GHSG HD15 trial. *Blood* 2010;**116**:21.

Engert A, Kobe C, Markova J, Haverkamp H, Borchmann P, Hitz F, et al. FDG-PET to guide radiotherapy in advanced-stage Hodgkin lymphoma patients with residual bulky disease after chemotherapy: results of the GHSG HD15 trial. *Annals of Oncology* 2011;**4**:96.

Engert M, Haverkamp H, Kobe C, Markova J, Renner C, Ho A, et al. Comparing intensity of chemotherapy followed PET-guided radiotherapy in patients with advanced stage Hodgkin lymphoma: final results of the GHSG HD15 trial. *Haematologica* 2012;**97** Suppl 1:455.

Kobe C, Dietlein M, Borchmann P, Engert A, Haverkamp H, Fuchs M, et al. Assessment of residual bulky tumor using FDG-PET in patients with advanced stages after completion of chemotherapy. Final report of the GHSG HD15 trial. *European Journal of Nuclear Medicine and Molecular Imaging* 2011;**38**(Suppl 2):165-6. [DOI: [10.1007/s00259-011-1908-8](https://doi.org/10.1007/s00259-011-1908-8)]

Kobe C, Dietlein M, Franklin J, Markova J, Lohri A, Amthauer H. 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. [DOI: [10.1182/blood-2008-06-155820](https://doi.org/10.1182/blood-2008-06-155820)]

Kobe C, Engert A, Markova J, Haverkamp H, Eich HT, Muller RP, et al. Assessment of residual bulky tumor using FDG-PET in patients with advanced stages after completion of chemotherapy. Final report of the GHSG HD15 trial. *Fortschritte Auf Dem Gebiet der Rontgenstrahlen und der Bildgebenden Verfahren* 2012;**184**(Suppl 1):no pagination found. [DOI: [10.1055/s-0032-1311119](https://doi.org/10.1055/s-0032-1311119)]

Kobe C, Kuhnert G, Haverkamp H, Fuchs M, Kahraman D, Eich HT, et al. Concordance in the interpretation of PET after chemotherapy in advanced stage Hodgkin lymphoma [Ubereinstimmung von lokaler und zentraler interpretation der PET nach chemotherapie beim fortgeschrittenen Hodgkin-lymphom]. *Nuklearmedizin [Nuclear Medicine]* 2015;**54**(6):241-6. [DOI: [10.3413/Nukmed-0746-15-06](https://doi.org/10.3413/Nukmed-0746-15-06)]

Kriz J, Reinartz G, Kobe C, Kuhnert G, Haverkamp H, Haverkamp U, et al. Relapse analysis of irradiated patients within the HD15 trial of the German Hodgkin study group. *Haematologica* 2013;**98**(Suppl 2):3. [ONLINE ISSN: 1592-8721]

Kriz J, Reinartz G, Kobe C, Kunert G, Haverkamp H, Haverkamp U, et al. Analysis of relapses of irradiated patients within the HD15 trial of the German Hodgkin Study Group. *Strahlentherapie und Onkologie* 2013;**189**(Suppl 1):7-8. [DOI: [10.1007/s00066-013-0339-9](https://doi.org/10.1007/s00066-013-0339-9)]

Park SI, Shea TC, Olajide O, Reddy NM, Budde LE, Ghosh N, et al. ABVD followed by BV consolidation in risk-stratified patients with limited-stage Hodgkin lymphoma. *Blood Advances* 2020;**4**(11):2548-55. [DOI: [10.1182/bloodadvances.2020001871](https://doi.org/10.1182/bloodadvances.2020001871)]

HD21 {published data only}ACTRN12617000087358

ACTRN12617000087358. Treatment optimisation trial in the first-line treatment of advanced stage Hodgkin lymphoma: comparison of 6 cycles of escalated BEACOPP with 6 cycles

of BrECADD. <https://anzctr.org.au/Trial/Registration/TrialReview.aspx?ACTRN=ACTRN12617000087358> (date submitted 12 June 2016).

Borchmann P, Moccia A, Greil R, Herzberg M, Fossa A, Huttman A, et al. Treatment related morbidity in patients with classical Hodgkin lymphoma: results of the ongoing, randomized phase III HD21 trial by the German Hodgkin Study Group. *HemaSphere* 2022;**6** Suppl 5:1-2. [PMID: 639542766]

Greil R, Melchardt T. HD21 Treatment optimization trial in the first-line treatment of advanced-stage Hodgkin lymphoma: comparison of 6 cycles of escalated BEACOPP with 6 cycles of BrECADD. *Magazine of European Medical Oncology* 2017;**10** Suppl 1:39-40. [PMID: 615258535]

Greil R. GHSG-HD21: treatment optimization 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. *Magazine of European Medical Oncology* 2022;**15**:S61. [PMID: CN-02401457]

NCT02661503. HD21 for advanced stages. <https://clinicaltrials.gov/study/NCT02661503> (first submitted 19 January 2016).

JCOG1305 {published data only}

UMIN000019868. JCOG1305: Interim response adapted strategy for the advanced stage of Hodgkin Lymphoma. <https://jrct.niph.go.jp/latest-detail/jRCTs031180218> (date of registration 8 March 2019).

NCT03226249 {published data only}

Allen PB, Lu X, Chen Q, O'Shea K, Chmiel JS, Slonim LB, et al. Sequential pembrolizumab and AVD are highly effective at any PD-L1 expression level in untreated Hodgkin lymphoma. *Blood Advances* 2023;**7**(12):2570-6. [DOI: [10.1182/bloodadvances.2022008116](https://doi.org/10.1182/bloodadvances.2022008116)]

Allen PB, Savas H, Evens AM, Advani R, Palmer B, Pro B, et al. Pembrolizumab followed by AVD in untreated early unfavorable and advanced stage classical Hodgkin lymphoma. *Blood* 2020;**137**(10):1318-26. [DOI: [10.1182/blood.2020007400](https://doi.org/10.1182/blood.2020007400)]

* NCT03226249. PET-directed therapy with pembrolizumab and combination chemotherapy in treating patients with previously untreated classical Hodgkin lymphoma. <https://ClinicalTrials.gov/show/NCT03226249> (first submitted 19 July 2017).

NCT03712202 {published data only}

NCT03712202. Brentuximab vedotin and nivolumab in treating participants with early stage classic Hodgkin lymphoma. <https://clinicaltrials.gov/show/NCT03712202> (first submitted 6 August 2018).

NCT05922904 {published data only}

NCT05922904. PET adapted brentuximab vedotin and pembrolizumab in combination with doxorubicin and dacarbazine in classic Hodgkin lymphoma. <https://classic.clinicaltrials.gov/show/NCT05922904> (first submitted 28 June 2023).

RADAR {published data only}

NCT04685616. Brentuximab vedotin in early stage Hodgkin lymphoma. <https://clinicaltrials.gov/show/NCT04685616> (first submitted 27 November 2020).

Radford J, Adedayo T, Ardavan A, Barrington SF, Berkahn L, Chauvie S, et al. RADAR: an international phase III, PET response adapted, randomised trial in progress, comparing ABVD +/- LSRT with brentuximab vedotin +AVD +/- LSRT in patients with previously untreated limited-stage classical Hodgkin lymphoma. *HemaSphere* 2022;**6 Suppl 5**:12-3.

RAFTING {published data only}

Bergesio F, Meignan M, Cottreau AS, Guerra L, Versari A, Maggi AD, et al. An example of prospective application of imaging biomarker: set-up of the first worldwide prospective phase III Onco-Hematological clinical trial using TMTV for patient stratification. *Physica Medica* 2022;**104 Suppl 1**:126-7.

Gallamini 2023. Revisiting the predictive role of 18F-fluorodeoxyglucose positron emission tomography/

computerized tomography on treatment outcome in early-stage favorable Hodgkin lymphoma. *Hematological Oncology* 2023;**441**(4):608-11. [DOI: [10.1002/hon.3158](https://doi.org/10.1002/hon.3158)]

Gallamini A, Picardi M, Filonenko K, Gotti M, Cespedes JN, Rossi A, et al. Radiation-free therapy as the initial treatment of good-prognosis early non-bulky Hodgkin lymphoma, defined by a low metabolic tumor volume and a negative PET-2 - RAFTING trial. *HemaSphere* 2022;**6 Suppl 5**:10-1. [PMID: 639542854]

NCT04866654. Radiation free chemotherapy for early Hodgkin lymphoma. <https://ClinicalTrials.gov/show/NCT04866654> (first submitted 22 April 2021).

SWOG S0816 {published data only}

Danilov AV, Li H, Press OW, Shapira I, Swinnen LJ, Noy A, et al. Feasibility of interim PET-adapted therapy in HIV-positive patients with advanced Hodgkin lymphoma (HL): sub-analysis of SWOG S0816 phase 2 trial. *Blood* 2015;**126**(23):1498. [PMID: 72172042]

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. [PMID: 611182522]

NCT00822120. S0816 Fludeoxyglucose F 18-PET/CT imaging and combination chemotherapy with or without additional chemotherapy and G-CSF in treating patients with stage III or stage IV Hodgkin lymphoma. <https://ClinicalTrials.gov/show/NCT00822120> (first submitted 13 January 2009).

Press OW, LeBlanc M, Rimsza LM, Schoder H, Friedberg JW, Evens AM, et al. A phase II trial of response-adapted therapy of stages III-IV Hodgkin lymphoma using early interim FDG-PET imaging: US Intergroup S0816. *Hematological Oncology* 2013;**1**:137-8. [DOI: [10.1002/hon.2057](https://doi.org/10.1002/hon.2057)]

Press OW, Li H, Schoder H, Straus DJ, Moskowitz CH, LeBlanc M, et al. US intergroup trial of response-adapted therapy for stage III to IV Hodgkin lymphoma using early interim fluorodeoxyglucose-positron emission tomography imaging: Southwest Oncology Group S0816. *Journal of Clinical Oncology* 2016;**34**(17):2020-7. [PMID: 610575791]

Scott DW, Li H, Harvey Y, Chan F, Mottok A, Boyle M, et al. The 23-gene gene expression-based assay does not predict interim PET scan results after ABVD in advanced stage classical Hodgkin lymphoma in the US intergroup S0816 trial. *Hematological Oncology* 2017;**35 Suppl 2**:92-3. [PMID: 617729883]

Stephens DM, Li H, Schoder H, Straus DJ, Moskowitz CH, LeBlanc M, et al. Five-year follow-up of SWOG S0816: limitations and values of a PET-adapted approach with stage III/IV Hodgkin lymphoma. *Blood* 2019;**134**(15):1238-46. [DOI: [10.1182/blood.2019000719](https://doi.org/10.1182/blood.2019000719)]

Additional references

Aldin 2020

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. *Cochrane Database of Systematic Reviews* 2020, Issue 1. Art. No: CD012643. [DOI: [10.1002/14651858.CD012643.pub3](https://doi.org/10.1002/14651858.CD012643.pub3)]

Ansell 2019

Ansell SM. Immunotherapy in Hodgkin lymphoma: the road ahead. *Trends in Immunology* 2019;**40**(5):380-6.

Ansell 2022

Ansell SM, Radford J, Connors JM, Długosz-Danecka M, Kim WS, Gallamini A, et al. Overall survival with brentuximab vedotin in stage III or IV Hodgkin's lymphoma. *New England Journal of Medicine* 2022;**387**(4):310-20. [DOI: [10.1056/NEJMoa2206125](https://doi.org/10.1056/NEJMoa2206125)]

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]

Borchmann 2023

Borchmann P, Moccia AA, Greil R, Schneider G, Hertzberg M, Schaub V, et al. BrECADD is non-inferior to eBEACOPP in patients with advanced stage classical Hodgkin lymphoma: efficacy results of the GHSG phase III HD21 trial. *Hematological Oncology* 2023;**41**:881-2. [PMID: [10.1002/hon.3196_LBA5](https://pubmed.ncbi.nlm.nih.gov/4013196/)]

Böll 2019

Boll B, Fossa A, Gorgen H, Kamper P, Leppa S, Molin D, et al. B-CAP (brentuximab vedotin, cyclophosphamide, doxorubicin and prednisolone) in older patients with advanced-stage Hodgkin lymphoma: results of a phase II intergroup trial by the German Hodgkin study group (GHSG) and the Nordic Lymphoma Group (NLG). *Oncology Research and Treatment* 2019;**42** Suppl 4:161. [DOI: [10.1159/000502425](https://doi.org/10.1159/000502425)] [CLINICALTRIALS-GOV: NCT02191930]

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]

cHL001

NCT05404945. Fitness-adapted, pembrolizumab-based therapy for untreated classical Hodgkin lymphoma patients 60 years of age and above. <https://clinicaltrials.gov/study/NCT05404945> (first submitted 22 May 2022).

Cooper 2019

Cooper C, Varley-Campbell J, Carter P. Established search filters may miss studies when identifying randomized controlled trials. *Journal of Clinical Epidemiology* 2019;**112**:12-9. [DOI: [10.1016/j.jclinepi.2019.04.002](https://doi.org/10.1016/j.jclinepi.2019.04.002)]

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, et al, editors(s). *Cochrane Handbook for Systematic Reviews of Interventions*. 6.3 (updated February 2022) edition. London: Wiley, 2022. [www.training.cochrane.org/handbook]

Diaz 2011

Diaz T, Navarro A, Ferrer G, Gel B, Gaya A, Artells R, et al. Lestaurtinib inhibition of the Jak/STAT signalling pathway in Hodgkin lymphoma inhibits proliferation and induces apoptosis. *PloS One* 2011;**6**(4):e18856. [PMID: 21533094]

Eichenauer 2018

Eichenauer DA, Aleman BM, Andre M, Federico M, Hutchings M, Illidge T, et al. Hodgkin lymphoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology* 2018;**29**(Suppl 4):19-29. [DOI: [10.1093/annonc/mdy080](https://doi.org/10.1093/annonc/mdy080)]

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]

Follows 2022

Follows GA, Barrington SF, Bhuller KS, Culligan DJ, Cutter DJ, Gallop-Evans E, et al. Guideline for the first-line management of classical Hodgkin lymphoma - a British Society for Haematology guideline. *British Journal of Haematology* 2022;**197**:558-72. [DOI: [10.1111/bjh.18083](https://doi.org/10.1111/bjh.18083)]

Fraga 2007

Fraga M, Forteza J. Diagnosis of Hodgkin's disease: an update on histopathological and immunophenotypical features. *Histology and Histopathology* 2007;**22**(8):923-35. [PMID: 17503349]

Georgi 2023

Georgi TW, Kurch L, Hasenclever D, Warbey VS, Pike L, Radford J, et al. Interobserver variability in interim PET assessment in Hodgkin lymphoma - reasons and solutions. *PLoS One* 2023;**18**(3):e0283694. [DOI: [10.1371/journal.pone.0283694](https://doi.org/10.1371/journal.pone.0283694)]

Glanville 2019

Glanville J, Foxlee R, Wisniewski S, Noel-Storr A, Edwards M, Dooley G. Translating the Cochrane EMBASE RCT filter from the Ovid interface to Embase.com: a case study. *Health Information and Libraries Journal* 2019;**36**(3):264-77. [DOI: [10.1111/hir.12269](https://doi.org/10.1111/hir.12269)]

Herrera 2023

Herrera AF, LeBlanc ML, Castellino SM, Li H, Rutherford SC, Evens AM, et al. SWOG S1826, a randomized study of nivolumab(N)-AVD versus brentuximab vedotin(BV)- AVD in advanced stage (AS) classic Hodgkin lymphoma (HL). *Journal of Clinical Oncology* 2023;**41**(Suppl 17):LBA4. [DOI: [10.1200/JCO.2023.41.17_suppl.LBA4](https://doi.org/10.1200/JCO.2023.41.17_suppl.LBA4)] [CLINICALTRIALS.GOV: NCT03907488]

Higgins 2017

Higgins JP, Altman DG, 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, et al, editors(s). *Cochrane Handbook for Systematic Reviews of Interventions*. 5.2 (updated June 2017) edition. London: Wiley, 2017.

Higgins 2023b

Higgins JP, Li T, Deeks JJ (editors). Chapter 6: Choosing effect measures and computing estimates of effect. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors(s). *Cochrane Handbook for Systematic Reviews of Interventions*. 6.4 (updated August 2023) edition. London: Wiley, 2023.

Huang 2022

Huang J, Pang WS, Lok V, Zhang L, Lucero-Prisno III DE, Xu W, et al. Incidence, mortality, risk factors, and trends for Hodgkin lymphoma: a global data analysis. *Journal of Hematology & Oncology* 2022;**15**:57. [DOI: [10.1186/s13045-022-01281-9](https://doi.org/10.1186/s13045-022-01281-9)]

Klimm 2005

Klimm B, Engert A, Diehl V. First-line treatment of Hodgkin's lymphoma. *Current Hematology Reports* 2005;**4**(1):15-22. [PMID: 15610655]

Lazarovici 2021

Lazarovici J, Amorim S, Bouabdallah K, Guidez S, Molina L, Morschhauser F, et al. Nivolumab first-line therapy for elderly, frail Hodgkin lymphoma patients: NIVINIHO, a LYSA phase II study. *Blood* 2021;**138** Suppl 1:232. [DOI: [10.1182/blood-2021-147863](https://doi.org/10.1182/blood-2021-147863)] [CLINICALTRIALS.GOV: NCT03580408]

Lefebvre 2022

Lefebvre C, Glanville J, Briscoe S, Featherstone R, Littlewood A, Marshall C, et al. Chapter 4: Searching for and selecting studies. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors(s). *Cochrane Handbook for Systematic Reviews of Interventions*. 6.3 (updated February 2022) edition. London: Wiley, 2022.

Li 2023

Li T, Higgins JP, Deeks JJ (editors). Chapter 5: collecting data. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors(s). *Cochrane Handbook for Systematic Reviews of Interventions*. 6.4 (updated August 2023) edition. London: Wiley, 2023.

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. [DOI: [10.1200/JCO.1989.7.11.1630](https://doi.org/10.1200/JCO.1989.7.11.1630)]

Metzger 2019

Metzger M, Mauz-Körholz C. Epidemiology, outcome, targeted agents and immunotherapy in adolescent and young adult non-Hodgkin and Hodgkin lymphoma. *British Journal of Haematology* 2019;**185**(6):1142-57. [DOI: [10.1111/bjh.15789](https://doi.org/10.1111/bjh.15789)]

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. [PMID: 19631508]

Oken 1982

Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, et al. Toxicity and response criteria of the Eastern Cooperative Oncology group. *American Journal of Clinical Oncology* 1982;**5**(6):649-55. [PMID: 7165009]

Page 2024

Page MJ, Higgins JP, Sterne JA. Chapter 13: Assessing risk of bias due to missing results in a synthesis. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors(s). *Cochrane Handbook for Systematic Reviews of Interventions*. 6.5 (updated August 2024) edition. London: Wiley, 2024.

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. [PMID: 9921604]

Picardi 2020

Picardi M, Fonti R, Della Pepa R, Giordano C, Pugliese N, Nicolai E, et al. 2-deoxy-2[F-18] fluoro-D-glucose positron emission tomography Deauville scale and core-needle biopsy to determine successful management after six doxorubicin, bleomycin, vinblastine and dacarbazine cycles in advanced-stage Hodgkin lymphoma. *European Journal of Cancer* 2020;**132**:85-97. [DOI: [10.1016/j.ejca.2020.03.008](https://doi.org/10.1016/j.ejca.2020.03.008)]

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](https://doi.org/10.1002/14651858.CD009411.pub2)] [PMID: 23784872]

RATIFY

NCT05627115. Response adapted incorporation of tislelizumab into the front-line treatment of older patients with Hodgkin lymphoma (RATIFY). <https://clinicaltrials.gov/study/NCT05627115> (first submitted 7 November 2022).

Rossi 2022

Rossi C, Andre M, Dupuis J, Morschhauser F, Joly B, Lazarovici J, et al. High-risk stage IIB Hodgkin lymphoma treated in the H10 and AHL2011 trials: total metabolic tumor volume is a useful risk factor to stratify patients at baseline. *Haematologica* 2022;**107**(12):2897-904. [DOI: [10.3324/haematol.2021.280004](https://doi.org/10.3324/haematol.2021.280004)]

Santesso 2020

Santesso N, Glenton C, Dahm P, Garner P, Akl E, Alper B, et al. GRADE guidelines 26: Informative statements to communicate the findings of systematic reviews of interventions. *Journal of Clinical Epidemiology* 2020;**119**:126-35. [DOI: [10.1016/j.jclinepi.2019.10.014](https://doi.org/10.1016/j.jclinepi.2019.10.014)]

Sasse 2017

Sasse S, Brockelmann PJ, Goergen H, Plutschow A, Muller H, Kreissl S, et al. Long-term follow-up of contemporary treatment in early-stage Hodgkin lymphoma: updated analyses of the German Hodgkin study group HD7, HD8, HD10, and HD11 trials. *Journal Of Clinical Epidemiology* 2017;**35**(18):1999-2007. [DOI: [10.1200/JCO.2016.70.9410](https://doi.org/10.1200/JCO.2016.70.9410)]

Schünemann 2024

Schunemann HJ, Higgins JP, Vist GE, Glasziou P, Akl EA, Skoetz N, et al. Chapter 14: Completing 'Summary of findings' tables and grading the certainty of the evidence. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors(s). *Cochrane Handbook for Systematic Reviews of Interventions*. 6.5 (updated August 2023) edition. London: Wiley, 2024.

SEER 2022

Surveillance Research Program, National Cancer Institute. SEER*Explorer: an interactive website for SEER cancer statistics [Internet]. Data source(s): SEER Incidence Data, November 2022 Submission (1975-2020), SEER 22 registries. <https://seer.cancer.gov/statistics-network/explorer/> (accessed 19 April 2023).

Skoetz 2017

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 8. Art. No: CD007941. [DOI: [10.1002/14651858.CD007941.pub3](https://doi.org/10.1002/14651858.CD007941.pub3)]

Strati 2019

Strati P, Fanale MA, Oki Y, Turturro F, Fayad LE, Bartlett NL, et al. ABVD plus rituximab versus ABVD alone for advanced stage, high-risk classical Hodgkin lymphoma: a randomized phase 2 study. *Haematologica* 2019;**104**(2):65-7. [DOI: [10.3324/haematol.2018.199844](https://doi.org/10.3324/haematol.2018.199844)]

Swerdlow 2017

Swerdlow SH, Camp E, Harris NL, JaMe ES, Pileri SA, Stein H, et al. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. Lyon: International Agency for Research on Cancer (IARC), 2017. [<https://archive.org/details/whoclassificatio00swer/page/292/mode/2up>]

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. [PMID: 17555582]

Trotman 2021

Trotman J, Barrington SF. The role of PET in first-line treatment of Hodgkin lymphoma. *Lancet Haematology* 2021;**8**(1):e67-79. [DOI: [10.1016/S2352-3026\(20\)30357-4](https://doi.org/10.1016/S2352-3026(20)30357-4)]

Vassilakopoulos 2023

Vassilakopoulos TP, Liaskas A, Pereyra P, Panayiotidis P, Angelopoulou MK, Gallamini A. Incorporating monoclonal antibodies into the first-line treatment of classical Hodgkin lymphoma. *International Journal of Molecular Sciences* 2023;**24**(17):13187. [DOI: [10.3390/ijms241713187](https://doi.org/10.3390/ijms241713187)]

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]

References to other published versions of this review

Sickinger 2013

Sickinger MT, von Tresckow B, Kobe C, Engert A, Skoetz N. Positron emission tomography (PET)-adapted therapy for Hodgkin lymphoma patients. *Cochrane Database of Systematic Reviews* 2013, Issue 6. Art. No: CD010533. [DOI: [10.1002/14651858.CD010533](https://doi.org/10.1002/14651858.CD010533)]

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](https://doi.org/10.1002/14651858.CD010533.pub2)]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

AHL2011

Study characteristics

Methods	<u>Study design</u> Randomised controlled trial to test PET-based de-escalation from BEACOPP_{escalated} to ABVD after two cycles in advanced stage classic HL • Randomisation before any treatment and PET-scan • Intervention arm with FDG-PET-based treatment adaptation: de-escalation for PET-negative individuals, no change for PET-positive participants • Control arm: standard treatment for both arms • Analysed mostly together for PET-negative and PET-positive participants <u>Sample size</u> • Number randomised: 823 ◦ PET-adaptation: 410 ◦ Standard: 413 • 23 did not receive PET-2; 8 due to toxicity, 4 died, 3 withdrew consent, 3 concurrent illnesses, 1 PET-refusal, 2 other reasons PET-negative participants (number included for ITT analysis) • PET-adaptation: 346 • Standard: 349 PET-positive participants (number included for ITT analysis) • PET-adaptation: 51 • Standard: 49
Participants	<u>Setting:</u> Multicentre study on advanced stage HL including 90 study locations in France and Belgium with enrolment between 19 May 2011 and 29 April 2014 <u>Eligibility criteria</u> Inclusion criteria • Participants with a first diagnosis of classic HL according to WHO criteria, excluding the nodular lymphocyte predominant subtype • Age 16-60 years • No previous treatment for HL • Ann Arbor stages: IIB with mediastinum/thorax ≥ 0.33 or extranodal localisation III IV • Baseline [18F]FDG-PET scan (PET 0) performed before any treatment with at least one hypermetabolic lesion • ECOG performance status < 3 (Oken 1982) • With a minimum life expectancy of 3 months • Having previously signed a written informed consent • Participants must be covered by a social security system (in France) Exclusion criteria • Any history of cancer or cancer treatment during the last 5 years except for non-melanoma skin tumours or stage 0 (in situ) cervical carcinoma

AHL2011 (Continued)

- Uncontrolled infectious disease, including active HBV infection defined by either detection of HBs antigen or presence of anti-HBc antibody without detectable anti-HBs antibody
- HIV, HCV or HTLV serology positivity - abnormal liver (bilirubin > 2.5 N) function unless abnormalities are due to Hodgkin lymphoma
- Pregnant or lactating women, men and women of childbearing potential not practising an adequate method of contraception
- Abnormal renal (creatinin > 150 emol/L) function unless abnormalities are due to Hodgkin lymphoma
- Leukopenia < 2 G/L or thrombopenia < 100 G/L unless abnormalities are due to Hodgkin lymphoma
- Severe cardio-pulmonary, or metabolic disease interfering with normal application of protocol treatment
- Incapable person

Participant characteristics (PET-negative and -positive participants reported together)

Age (median, IQR)

- PET-adapted: 29 (24-40) years
- Standard: 31 (23-41) years

Sex (female)

- PET-adapted: 62%
- Standard: 64%

Stage (Ann Arbor)

- PET-adapted: I: 2%, II: 11%, III: 28%, IV: 59%
- Standard: I: < 1%, II: 42%, III: 28%, IV: 62%

B-symptoms

- Not reported

Bulky mass ≥ 10 cm

- PET-adapted: 37%
- Standard: 38%

Interventions

Initial therapy before PET-scan (same for all groups)

- 2 cycles of BEACOPP_{escalated}

FDG-PET scan

- Timing: after 2 cycles
- Scoring criteria: 5-point score according to Lugano criteria (negative: 1–3; positive: 4 or 5 positive)
- Other information:
 - Independent central review by three reviewers; the final decision was based on at least two concordant responses
 - Further PET after 4 and 6 cycles

Interventions:

- **PET-negative participants** (median follow-up time for both groups: 67 months, IQR 55 to 78)
 - De-escalation: + 4 cycles of ABVD
 - Standard: + 4 cycles of BEACOPP_{escalated}
- **PET-positive participants**: 2 cycles BEACOPP_{escalated}

Outcomes

Overall survival (secondary endpoint)

- Defined as the time from random assignment to death of any cause or last follow-up

AHL2011 (Continued)

- Sensitivity analysis PET-negative participants only

Progression-free survival (primary endpoint)

- Defined as the time from randomisation assignment to first progression or relapse or death of any cause or last follow-up
- Sensitivity analysis PET-negative participants only

Treatment response: reported as complete and partial response for both PET-negative and PET-positive participants together

Treatment-related mortality: reported for PET-positive and -negative together

Quality of life: not reported (not planned in protocol)

Adverse events:

- Adverse events: reported per grade and for both PET-positive and negative together
- Serious adverse events: only SAE related to treatment reported

Long-term toxicities

- Secondary malignancies reported

Other outcomes

- Event-free survival (EFS)
- Disease-free survival (DFS)
- SUVmax reduction between PET0 and PET2/PET4

Notes

- **Funding source**: French government Programme, Hospitalier de Recherche Clinique 2010
- **Declaration of interest**: ROC: Roche/Genentech, Takeda, Gilead Sciences, BMS, Merck, AbbVie, Celgene, Janssen, Amgen, Incyte; PB: Takeda, BMS, Roche, Amgen, AbbVie/Genentech; JL: MundipharmaHG: Gilead Sciences, Janssen, Celgene, Roche, Mundipharma, Takeda; AS: PfizerJD: AstraZeneca, Gilead Sciences; TG: Pfizer, Takeda, Gilead Sciences, Roche; KB: Roche, Takeda Science Foundation, AbbVie, Kite/Gilead; PF: Roche/Genentech, Janssen, AbbVie, Gilead Sciences, Amgen, AstraZeneca; FM: Roche/Genentech, Gilead Sciences, Celgene, Bristol Myers Squibb, AbbVie, Epizyme, ServierDS: Takeda, Iqone healthcare, Janssen, Roche, AbbVie; CB: Roche; DD: AstraZeneca/MedImmune; MA: Takeda, Roche, Celgene, Gilead Sciences; MM: Roche; CR: Janssen, AbbVie, Roche, Takeda; no other potential conflicts of interest were reported.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Randomisation was done centrally with the permuted block method and stratified according to Ann Arbor stage (IIB vs III–IV) and international prognosis score (IPS; 0–2 vs ≥ 3). The randomisation list was generated by LYSARC."
Allocation concealment (selection bias)	Low risk	Central randomisation; protocol: "The study site will receive back the randomization number and arm for the randomized patient."
Blinding (performance bias): All outcomes	Low risk	"Patients and investigators were not masked to treatment allocation". Although participants and personnel were not blinded to treatment, this did probably not cause systematic differences in other treatment or follow-up of participants, as the follow-up schedules were preplanned for both groups.
Blinding of outcome assessment (detection bias) Mortality outcomes	Low risk	Although this was an open-label trial, we assumed that an unblinded outcome assessor would have no influence on the assessment of objective outcomes.

AHL2011 (Continued)

Blinding of outcome assessment (detection bias) Subjective outcomes (all except mortality)	High risk	Investigators, who were probably also the outcome assessors, were not blinded to treatment allocation. This may have affected the assessment of subjective outcomes.
Incomplete outcome data (attrition bias) Efficacy outcomes	Low risk	The number of participants censored from the analysis was balanced between groups.
Incomplete outcome data (attrition bias) Safety outcomes	Low risk	Data were assessed for all participants formally randomised and who had received at least one dose of treatment (safety population).
Selective reporting (reporting bias)	Low risk	All reported outcomes were predefined at the protocol stage.
Other bias	Low risk	No other potential sources of bias identified

H10

Study characteristics

Methods

Study design

Randomised controlled trial to test FDG-PET-based treatment adaptation after 2 cycles of ABVD in stage I and II, supra-diaphragmatic classic HL

- Randomisation before any treatment and FDG-PET scan
- Intervention arm with FDG-PET-based treatment adaptation: de-escalation for PET-negative individuals, escalation for PET-positive patients
- Standard: standard treatment for both arms
- Separately analysed for early favourable and early unfavourable stages

Favourable group

Sample size

- Number randomised:
 - o PET-adaptation: 379
 - o Standard: 375
- 7 did not start, complete first two ABVD cycles or did not receive a PET-scan

PET-negative participants

- Number completed before safety amendment:
 - o De-escalation: 238
 - o Standard: 227
- Number completed after safety amendment:
 - o De-escalation: 95
 - o Standard: 90

Unfavourable group

Sample size:

- Number randomised:
 - o 595 to de-escalation

H10 (Continued)

- o 583 to standard
- 14 did not start, complete the first 2 ABVD cycles, or did not receive a PET-scan

PET-negative participants

- Number completed before safety amendment:
 - o 302 in de-escalation
 - o 292 in standard
- Number completed after safety amendment:
 - o 167 in de-escalation
 - o 153 in standard

PET-positive participants, favourable and unfavourable

- Number completed:
 - o 169 in escalation group
 - o 192 in standard group

Participants

Setting: Multicentre study in several European countries (France, Belgium, Italy, Netherlands) on early and intermediate stage HL between November 2006 and June 2011

Eligibility criteria

Inclusion criteria

- Previously untreated, histologically proven classic HL
- Supradiaphragmatic Ann Arbor stage I or II
- Clinical stage I disease OR stage II disease with ≤ 3 involved areas
- Age < 50 years
- ESR < 50 mm/hr (no B symptoms) OR ESR < 30 mm/hr (B symptoms present)
- MT ratio < 0.35
- WHO performance status of 0 to 3
- Bilirubin ≤ 2.5 times upper limit of normal (ULN); creatinine ≤ 2.5 times ULN; not pregnant or nursing and fertile patients must use effective contraception
- Written informed consent

Exclusion criteria

- Severe cardiac, pulmonary, neurologic, psychiatric, or metabolic disease
- Unstable diabetes mellitus
- Other malignancies within the past 5 years except for basal cell skin cancer or adequately treated carcinoma in situ of the cervix
- Known HIV infection
- Psychological, familial, sociological, or geographical conditions that would preclude study compliance

Participant baseline characteristics(before safety amendment only)

Age (median, range)

- **PET-negative participants, favourable**
 - o De-escalation: 30 (15 to 49 years)
 - o Standard: 31 (15 to 49) years
- **PET-negative participants, unfavourable**
 - o Standard: 32 (15 to 70) years
 - o De-escalation: 31 (16 to 70) years
- **PET-positive participants:**
 - o Escalation: 30 (15 to 70) years
 - o Standard: 30 (15 to 66) years

H10 (Continued)

Sex (female)

- **PET-negative participants, favourable**
 - o De-escalation: 52.1%
 - o Standard: 42.3%
- **PET-negative participants, unfavourable**
 - o De-escalation: 54.6%
 - o Standard: 51.7%
- **PET-positive participants**
 - o Escalation: 43.8%
 - o Standard: 49.0%

Stage (Ann Arbor stage I)

- **PET-negative participants, favourable**
 - o De-escalation: 31.9%
 - o Standard: 28.6%
- **PET-negative participants, unfavourable**
 - o De-escalation: 19.5%
 - o Standard: 19.2%
- **PET-positive participants**
 - o Escalation: 18.3%
 - o Standard: 24.5%

B-symptoms

- **PET-negative participants**
 - o De-escalation: 8.4%
 - o Standard: 7.9%
- **PET-negative participants, unfavourable**
 - o De-escalation: 40.1%
 - o Standard: 37.0%
- **PET-positive participants**
 - o Escalation: 37.3%
 - o Standard: 34.9%

Bulky mediastinum

- **PET-negative participants**
 - o De-escalation: 0.4%
 - o Standard: 0.4%
- **PET-negative participants, unfavourable**
 - o De-escalation: 39.4%
 - o Standard: 41.4%
- **PET-positive participants**
 - o Escalation: 40.8%
 - o Standard: 37.0%

Interventions

Initial therapy before PET-scan (same for all groups)

- 2 cycles of ABVD

FDG-PET scan

- Timing: between days 22 and 25 of the second cycle of ABVD
- Scoring criteria: International Harmonization Project criteria (a negative PET corresponds to scores 1 (no uptake) and 2 (uptake mediastinum) on the 5-point score according to Lugano)
- Other information: 75% of scans were centrally reviewed, 93% concordance with local assessment (Cohen's $k = 0.78$; 95% CI, 0.74 to 0.82).

H10 (Continued)

Interventions
PET-negative participants, favourable

- De-escalation: + 2 cycles of ABVD
 - o Median follow-up time 5.0 years (range 1.0 to 7.1)
 - o After safety amendment: 1 additional cycle of ABVD + 30 Gy (+6 Gy) INRT (median follow-up time 3.3 years (range 0.5 to 4.2))
- Standard: + 1 cycle of ABVD + 30 Gy (+6 Gy) INRT
 - o Median follow-up time 5.0 years (range 0.5 to 7.3)

PET-negative participants, unfavourable

- De-escalation: + 4 cycles of ABVD, no RT
 - o Median follow-up time 5.1 years (range 0.2 to 7.4)
 - o After safety amendment: + 2 cycles of ABVD + 30 Gy (+ 6 Gy) INRT (median follow-up time 3.5 years (range 0.8 to 4.3))
- Standard: + 2 cycles of ABVD + 30 Gy (+ 6 Gy) INRT
 - o Median follow-up time 5.1 years (range 0.7 to 7.1)

PET-positive participants

- **Favourable**
 - o Escalation: + 2 cycles of BEACOPP_{esc} + 30 Gy (+ 6 Gy) INRT
 - o Standard: + 1 cycles of ABVD + 30 Gy (+ 6 Gy) INRT
- **Unfavourable**
 - o Escalation: + 2 cycles of BEACOPP_{esc} + 30 Gy (+ 6 Gy) INRT
 - o Standard: + 2 cycle of ABVD + 30 Gy (+ 6 Gy) INRT

Follow-up time PET-positive participants

- Control group: median follow-up time 4.6 years (range 0.5 to 6.9) for both favourable and unfavourable
- Intervention group: median follow-up time 4.4 years (range 0.2 to 7.2) for both favourable and unfavourable

Outcomes
Overall survival

- Secondary outcome, defined as the time from entry into the clinical trial until death as a result of any cause

Progression-free survival

- Primary outcome, defined as the time 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 disease (50% increase from nadir of any previous partial remission lesions or appearance of new lesions) on computed tomography scan measurements during protocol treatment; or death from any cause, whichever occurred first

Treatment response: not reported (planned in protocol, CR, CRu, PR, no change, or PD)

Treatment-related mortality: not reported (not planned in protocol)

Quality of life: not reported (not planned in protocol)

Adverse events: partly reported (for PET-positive participants only)

Long-term toxicities: The authors reported the pre-planned toxicities in their long-term follow-up:

- Secondary malignancies
- Cardiovascular events
- Pulmonary events

H10 (Continued)

Other outcomes:

- Event-free survival (planned in protocol, not reported)

Notes

- **Funding source:** European Organisation for Research and Treatment of Cancer - EORTC
- **Declaration of interest:** MPEA, TG, MF, OR, CF, MG, RvdM, AR, VE, CF, GvI, FM, RB, CS, AS, VF, MB, TR, AV, MM and JR have no relationships to disclose; OC: Genentech, Takeda, Gilead Sciences, Sanofi; PB: MSD, Takeda; LS: Takeda, VMS; RD: Servier, Gilead Sciences, Roche, Celgene, Takeda, Karyopharm Therapeutics; MH: Takeda, Genentech, Celgene, Bayer, Janssen-Cilag, BMS

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"[The interactive randomization program]", centrally randomised by EORTC headquarters
Allocation concealment (selection bias)	Low risk	"Centrally randomly assigned to receive either..."
Blinding (performance bias): All outcomes	Low risk	Although participants and personnel were not blinded to treatment, this did probably not cause systematic differences in other treatment or follow-up of participants, as the follow-up schedules were preplanned for both groups.
Blinding of outcome assessment (detection bias) Mortality outcomes	Low risk	The study reports did not explicitly report blinding investigators, who were probably also the outcome assessors. We assumed that an unblinded outcome assessor would have no influence on the assessment of objective outcomes.
Blinding of outcome assessment (detection bias) Subjective outcomes (all except mortality)	High risk	The study reports did not explicitly report blinding investigators, who were probably also the outcome assessors. This may affect the assessment of subjective outcomes.
Incomplete outcome data (attrition bias) Efficacy outcomes	High risk	In the long-term follow-up, a substantial proportion of participants were missing without further elaboration.
Incomplete outcome data (attrition bias) Safety outcomes	High risk	In the long-term follow-up, a substantial proportion of participants were missing without further elaboration.
Selective reporting (reporting bias)	Unclear risk	Study protocol available. Not all the study's prespecified secondary outcomes reported: • Event-free survival Definitions between protocol and publication differed; that which was originally defined as EFS has been called PFS, and the original definition as in the protocol was not reported.
Other bias	Unclear risk	Based on the advice of an independent data monitoring board, a safety amendment was put in place in August 2010 for PET-negative participants only, which resulted in the closure of the chemotherapy-only arm (instead: AB-VD + INRT for favourable PET-negative participants). The included analysis was based on data from participants included before the safety amendment. Early stopping of clinical studies is known to increase the effect estimate of trials.

HD0607

Study characteristics

Methods	<u>Study design</u> Randomised controlled trial to investigate FDG-PET-based adaptation after initial treatment with ABVD in advanced stage Hodgkin lymphoma - Initial treatment for all participants with 2 cycles of ABVD - Interim PET after 2 cycles - Random assignment of PET-positive participants to BEACOPP_{escalated} followed by baseline BEACOPP, with or without rituximab on the first day of each cycle - PET-negative participants continued with 4 cycles of ABVD treatment; PET-6 negative participants with large nodal mass at baseline were randomly assigned to receive RT or no further RT <u>Sample size</u> - Number enrolled: 783 participants 1 patient was re-diagnosed with composite lymphoma 2 participants died during the first two cycles of ABVD before PET-scan <u>PET-negative participants</u> 630 participants interim-PET-2 negative 320 fulfilled randomisation criteria, 296 randomised - De-escalation: 148 - Standard: 148 - Not randomised: 47 PET-6 positive 3 withdrew study consent 260 without baseline LNM 320 participants fulfilled randomisation criteria 24 not allocated (patient or medical condition) <u>PET-positive participants</u> 150 participants had a proven PET-positive interim PET (PET-2) 149 randomised - Standard: 76 - Escalation: 72ACOPP + rituximab) - Not randomised: 7 participants withdrew consent 3 participants died due to disease progression or infection 1 progressed 2 stopped due to toxicity
Participants	<u>Setting:</u> Multicentre study including 24 study locations in Italy and Israel with enrolment of advanced-stage HL between June 2008 and June 2014 <u>Eligibility criteria</u> <u>Inclusion criteria</u> - Patients with advanced classical Hodgkin Lymphoma according to the World Health Organization classification - Aged 18-60 - Not previously treated - Stages IIB to IV B - All IPS prognostic groups

HD0607 (Continued)

- Patients who have signed an informed consent form

Exclusion criteria

- Concomitant or previously treated neoplastic disorder less than 5 years before the diagnosis of Hodgkin's lymphoma
- Psychiatric disorders
- Uncontrolled infectious disease
- Impaired cardiac (EF < 50%), renal (creatinine clearance < 60 mL/m)
- HIV, HBV DNA, HCV RNA positive markers
- Pregnancy and lactation
- Patients with uncompensated diabetes mellitus and fasting glucose levels over 200 mg/dL

Participant baseline characteristics

Age (median, range)

- **PET-negative participants**
 - De-escalation: 31 (16 to 59) years
 - Standard: 30 (18 to 60) years
- **PET-positive participants**
 - Both treatment groups: 30.5 (18 to 60) years

Sex (female)

- **PET-negative participants**
 - De-escalation: 55.4%
 - Standard: 58.8%
- **PET-positive participants**
 - Both treatment groups: 56.7%

Stage (Ann Arbor stage)

- **PET-negative participants**
 - De-escalation: II: 48.6%; III: 23.6%; IV: 27.7%
 - Standard: II: 45.9%; III: 29.7%; IV: 24.3%
- **PET-positive participants**
 - Both treatment groups: II: 33.3%, III: 29.3%, IV: 37.3%

B-symptoms

- **PET-negative participants**
 - De-escalation: 85.8%
 - Standard: 83.1%
- **PET-positive participants**
 - Both treatment groups: 80.7%

Large nodal mass > 10 cm

- **PET-negative participants**
 - De-escalation: 33.1%
 - Standard: 33.8%
- **PET-positive participants**
 - Both treatment groups together: 28.0%

Interventions

Initial therapy before PET-scan (same for all groups)

- 2 cycles of ABVD

FDG-PET scan

HD0607 (Continued)

- **Timing**
 - Interim PET-scan after 2 cycles of ABVD, PET-positive participants were randomised after 2 cycles
 - End-of-treatment PET scan after 6 cycles of ABVD for previously PET-negative participants, randomisation after PET-6-negativity
- **Scoring criteria:** 5-point score according to Lugano criteria (negative: 1-3; positive: 4 or 5)
- **Other information**
 - Central review: PET-2 images considered positive or minimally positive by the local PET site were uploaded on the study website and distributed to six independent reviewers
 - 281/630 PET-2 negative scans centrally reviewed
 - All PET-2-positive scans were centrally reviewed

Interventions

- **PET-negative participants** (median follow-up time for both groups 5.9 years, range 0.5 to 10 years)
 - **De-escalation:** no further treatment
 - **Standard:** + 30 Gy IFRT on the nodal region where an LNM was recorded at diagnosis (at least 4 weeks after completion of chemotherapy)
- **PET-positive participants**
 - **Escalation:** + 4 cycles of BEACOPP_{escalated} + 4 cycles of baseline BEACOPP + rituximab (IV 375 mg/m²) given on first day of each BEACOPP course
 - **Standard:** + 4 cycles of BEACOPP_{escalated} + 4 cycles of baseline BEACOPP

Outcomes
Overall survival

- 3-year OS rate for PET-positive participants
- Measured from the date of registration to the date of death as a result of any cause (two calculations: from PET-2 and end of chemotherapy)

Progression-free survival

- 3-year PFS rate for PET-positive participants
- Measured from the date of trial registration to the date of the first appearance of disease progression, relapse, or death as a result of any cause (two calculations: from PET-2 and end of chemotherapy)

Treatment response

- PET-positive participants: complete response
- PET-negative participants: complete response measured after 6 chemotherapy cycles for patients randomly assigned to NFT and after RT for patients randomly assigned to cRT

Treatment-related mortality: not reported

Quality of life: not assessed

Adverse events:

- Not reported separately per treatment arm, no comparison possible

Long-term toxicities:

- Data on secondary malignancies for PET-negative participants

Other outcomes

- 3-year relapse-free survival rate for PET-negative participants
- Disease-free survival for PET-negative participants
 - Measured from the date of complete remission (CR) to the date of relapse

Notes

- **Funding source:** Associazione Italiana per la Ricerca sul Cancro I.G. No. 203 (A.G.); Associazione Italiana Lotta alla Leucemia, Bergamo and Cuneo sections; research grant from the Bank Foundation Cassa di Risparmio of Cuneo

HD0607 (Continued)

- **Declaration of interest:** AG: Takeda Pharmaceuticals; ARossi: Takeda Pharmaceuticals, Roche; AROmano: Takeda Science Foundation, Incyte, Takeda Pharmaceuticals; SV: Takeda Italia; Millenium Pharmaceuticals, Seattle Genetics, Roche Italia; LT: AbbVie, Janssen Pharmaceuticals, Janssen Scientific Affairs; SC: DiXit srl, Terumo, Sirtex, Hoffmann la Roche; RZ: Deciphera, Novartis; CT: ADC Therapeutics, TG Therapeutics; LG: GE Healthcare; ARambaldi: Amgen, Omeros, Novartis, Astellas Pharma, Jazz Pharmaceuticals, Celgene. No other potential conflicts of interest were reported.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Protocol: "I pazienti vengono assegnati ad uno dei due trattamenti previsti in base ad un programma informatico del tutto indipendente da qualsiasi influenza esterna."
Allocation concealment (selection bias)	Unclear risk	No information on concealment of treatment allocation; probably central allocation
Blinding (performance bias): All outcomes	Low risk	Although participants and personnel were not reported to be blinded to treatment, this did probably not cause systematic differences in other treatment or follow-up of participants, as the follow-up schedules were preplanned for both groups.
Blinding of outcome assessment (detection bias) Mortality outcomes	Low risk	The outcome assessor was probably not blinded. We assumed that an unblinded outcome assessor would have no influence on the assessment of objective outcomes.
Blinding of outcome assessment (detection bias) Subjective outcomes (all except mortality)		High risk Investigators, who were probably also the outcome assessors, were probably not blinded to treatment allocation. This may have affected the assessment of subjective outcomes.
Incomplete outcome data (attrition bias) Efficacy outcomes	Low risk	No missing outcome data
Incomplete outcome data (attrition bias) Safety outcomes	Low risk	No missing outcome data, except for the outcome secondary malignancies: data available only for 75% (N = 223 out of 296 PET-2 and PET-6 negative participants); no reasons for missing data provided Data on AE, AE grade 3-4 and SAE were kindly provided by the primary study authors.
Selective reporting (reporting bias)	Low risk	Results on all protocol-planned outcomes reported or provided by the authors (information on safety). CR was reported for the combined treatment arms.
Other bias	Low risk	No other risk of bias identified

HD0801

Study characteristics

Methods

Study design

Randomised controlled trial to test whether advanced stage cHL participants with interim-PET2 and PET6 negativity benefit from consolidation radiotherapy of areas of bulky lesions after six cycles of AB-VD

HD0801 (Continued)

- Randomisation after completion of six cycles of ABVD
- FDG-PET-based treatment adaptation: de-escalation for PET-negative individuals
- Control arm: standard treatment for both arms
- PET-positive participants: salvage treatment with high-dose chemotherapy and subsequent autologous bone marrow transplant

Sample size

- Number enrolled: 520
- Number with PET2-scan: 512
 - PET-positive: 103
 - PET-negative: 409
- Number with PET6 negative scan and bulky masses at baseline: 125
- Number randomised:
 - PET-adaptation: 58
 - Standard: 58
- 9 eligible participants were not assigned.

Participants

Setting: Multicentre study on advanced stage HL including 50 study locations in Italy with enrolment between September 2008 and April 2013

Eligibility criteria Inclusion criteria

- Histologically confirmed Hodgkin's lymphoma of the classical type (nodular lymphocyte predominance excluded)
- Stage IIB-IV
- Age 18-70 years
- No prior therapy for Hodgkin's lymphoma
- Written informed consent
- ECOG performance status grades 0-3 (see Appendix E)
- FDG-PET scan before the initiation of treatment

Exclusion criteria

- Prior therapy for Hodgkin's lymphoma
- Age less than 18 or more than 70
- Other concomitant or prior malignancies, except basal cell skin carcinoma, or adequately treated carcinoma in situ of the cervix, or any cancer in complete remission for more than 5 years
- HIV infection
- Pregnancy or breastfeeding
- Renal failure (creatinine ≥ 2 times the normal value), liver failure (AST/ALT or bilirubin ≥ 2.5 times the normal value) or heart failure (NYHA class ≥ 2 or FEV $< 45\%$)

Participant characteristics (PET-negative participants only)

Age (median, IQR)

- PET-adapted: 29.5 (25.0 to 37.0) years
- Standard: 31.5 (26.0 to 39.8) years

Sex (female)

- PET-adapted: 48%
- Standard: 41%

Stage (Ann Arbor)

HD0801 (Continued)

- PET-adapted: II: 26%, III: 34%, IV: 40%
- Standard: II: 33%, III: 36%, IV: 31%

B-symptoms (%)

- PET-adapted: 64%
- Standard: 76%

Bulky mass $\geq$ 10 cm

- PET-adapted: 36%
- Standard: 40%

Interventions

Initial therapy before PET-scan (same for all groups)

- 2 cycles of ABVD

FDG-PET scan

- Timing: after 2 cycles
- Scoring criteria: positive: scan with FDG-uptake focus exceeding the normal FDG uptake considering anatomy/physiology in previously diagnosed sites or new sites, which is considered to represent malignancy. A positive PET scan is also considered mild and diffusely increased FDG uptake in nodal residual masses less than 2 cm with intensity lower than or equal to that of mediastinum.
- Other information:
 - o Central review within five days from scan by 5 reviewers
 - o PET after six cycles had to be negative as well

Interventions:

- **PET-negative participants** (median follow-up time for both groups: 67 months, IQR 55 to 78)
 - o De-escalation: + 4 cycles of ABVD
 - o Standard: + 4 cycles of ABVD + 30 Gy RT
- **PET-positive participants**: salvage therapy with high-dose chemotherapy and bone marrow transplant; no randomisation

Outcomes

Overall survival (secondary endpoint)

- Defined as the time from entry into the clinical trial (registration) until death from any cause or the date of the last follow-up

Progression-free survival (primary endpoint)

- Defined as the time from entry into the study (registration) until lymphoma progression or death from any cause

Treatment response: assessed, not reported

Treatment-related mortality: not reported

Quality of life: not assessed

Adverse events:

- Adverse events: assessed and determined by version 3.0 of the National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE); reported per AE
- Serious adverse events: assessed, not reported

Long-term toxicities

- Secondary malignancies

Other outcomes

HD0801 (Continued)

- Event-free survival (EFS)
- Relapse-free survival (RFS)

Notes

- **Funding source:** Italian Lymphoma Foundation
- **Declaration of interest:** UR: AstraZeneca, Accuracy; AR: Takeda, Italfarmaco, Clinigen Healthcare; LN: Takeda, Janssen-Cilag; AP: Roche, MSD, Pfizer, Sandoz, Gilead, BMS; AS: AbbVie, Amgen, AstraZeneca, Arqule, Bayer, Bristol Myers Squibb, Celgene, Eisai, Gilead, Lilly, MSD, Novartis, Pfizer, Roche, Sandoz, Sanofi, Takeda, and Servier; MT: Janssen-Cilag, Incyte, and AstraZeneca; FZ: AbbVie, Takeda, Roche, Amgen, Novartis, Gilead, Grifols, Celgene, and Janssen Cilag; EA: Bristol, Incyte, Novartis, Pfizer, and Takeda; ARF: AstraZeneca, Roche, MSD, Ipsen; PLZ: Verastem, Celltrion, Gilead, Janssen-Cilag, BMS, Servier, Sandoz, MSD, Immune Design, Celgene, Portola, Roche, Eusapharma, Kyowa Kirin, and Sanofi. The remaining authors declare no competing financial interests.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Both enrolment and randomization will be centralized via online procedures from Epidemiologia dei Tumori – università e ASO San Giovanni Battista - Torino through the Epiclin web site (http://www.epiclin.cpo.it). Responsible for randomization procedures and data collection is Dr. Giovannino Ciccone".
Allocation concealment (selection bias)	Low risk	Central randomisation
Blinding (performance bias): All outcomes	Unclear risk	Participants and personnel were not reported to be blinded to treatment; nine participants in the consolidation RT arm had treatment interruptions without further information provided on the nature of these interruptions.
Blinding of outcome assessment (detection bias) Mortality outcomes	Low risk	Although this was an open-label trial, we assumed that an unblinded outcome assessor would have no influence on the assessment of objective outcomes.
Blinding of outcome assessment (detection bias) Subjective outcomes (all except mortality)		High risk Investigators, who were probably also the outcome assessors, were not blind- ed to treatment allocation.
Incomplete outcome data (attrition bias) Efficacy outcomes	Unclear risk	It is unclear whether the two groups were followed up for a similar amount of time and whether censoring was balanced (not reported on KM curve).
Incomplete outcome data (attrition bias) Safety outcomes	Low risk	No missing outcome data
Selective reporting (reporting bias)	Low risk	Analysed as in protocol
Other bias	Low risk	No other potential sources of bias identified

HD16

Study characteristics

Methods

Study design

HD16 (Continued)

Randomised controlled trial comparing combined-modality therapy (CMT) with chemotherapy alone participants with early-stage HL based on FDG-PET scan after two cycles of ABVD

- Randomisation before any treatment and PET-2
- Intervention: De-escalation for PET-negative individuals, standard treatment for PET-positive patients
- Control is standard treatment for both arms.

Sample size

- Number randomised: 1150
 - PET-adaptation: 575, ITT population: 566
 - Standard: 575, ITT population: 573
 - 10 participants had a disconfirmed HL diagnosis, 1 withdrew consent.
- Per status after PET-2:
 - PET-positive: 340
 - PET-negative, de-escalation: 314
 - PET-negative, standard: 353
 - 110 participants with restaging or other eligibility criteria issue, 12 PET-2 too early or no PET-2, 4 patient wish, 2 withdrew consent, 2 independent disease entity, 1 inadequate response, 1 treatment in non-trial centre

Participants

Setting: Multicentre, international phase III trial on early stage HL conducted across 250 sites in Germany, Switzerland, Austria, and the Netherlands between November 2009 and December 2015 (median follow-up time 47 months)

Inclusion criteria

- Histologically proven primary diagnosis of Hodgkin lymphoma
- Classical Hodgkin lymphoma of clinical stage I or II or nodular lymphocyte-predominant Hodgkin lymphoma of Ann Arbor stage IB, IIA or IIB without any of the risk factors: large mediastinal mass ($\geq$ a third of the maximum transverse diameter of the thorax), extranodal lesions, elevated erythrocyte sedimentation rate ($\geq$ 50 mm/h without B symptoms, $\geq$ 30 mm/h with B symptoms), or three or more involved lymph node areas
- No previous treatment for Hodgkin lymphoma
- Age 18–75 years at inclusion
- Written informed consent to participate in the trial and consent to storage of data and tissue samples
- Normal organ function (except for Hodgkin lymphoma-related impairments)
- Life expectancy $>$ 3 months

Exclusion criteria

- Incomplete diagnosis of the disease stage
- Prior or concurrent disease that prevents treatment according to protocol (please refer to trial protocol for details) or Hodgkin lymphoma as part of a composite lymphoma or malignant disease within the last five years (exceptions: basal cell carcinoma, carcinoma in situ of the cervix uteri, completely resected melanoma TNMpT1)
- Pregnancy, lactation
- Eastern Cooperative Oncology Group (ECOG) performance status $>$ 2
- Long-term ingestion of corticosteroids (e.g. for chronic polyarthritis) or antineoplastic drugs (e.g. methotrexate); antiepileptic treatment
- 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
- General intolerance of any protocol medication
- Unsafe contraceptive methods
- Relationship of dependence or employer-employee relationship to the sponsor or the investigator

HD16 (Continued)

- Commitment to an institution on judicial or official order
- Participation in another interventional trial that could interact with this trial

Participant baseline characteristics (PP population):
Age (median, range)

- De-escalation: 39 (18 to 75) years
- Standard: 38 (18 to 75) years

Sex (female)

- De-escalation: 44%
- Standard: 42%

Stage (Ann Arbor)

- De-escalation: IA: 31%, IB: 5%, IIA: 58%, IIB: 5%
- Standard: IA: 32%, IB: 5%, IIA: 58%, IIB: 5%

Bulky disease

- De-escalation: 18%
- Standard: 16%

NLPHL:

- De-escalation: 10%
- Standard: 8%

Interventions
Initial therapy before PET-scan (same for all groups)

- 2 cycles ABVD

FDG-PET scans

- Timing: between day 22 and day 35 of the second ABVD cycle
- Scoring criteria: 5-point score according to Lugano criteria (negative: 1-2, positive: 3-5)
- Other information: Central review of all imaging at staging and after two cycles ABVD. The panel was masked to treatment group allocation and local findings.

Interventions

- **PET-negative participants**
 - o De-escalation: no further therapy (N = 314)
 - o Standard: 20 Gy IF-RT (N = 353)
- **PET-positive participants**
 - o De-escalation: + 20 Gy IF-RT
 - o Standard: 20 Gy IF-RT

Outcomes
Overall survival (secondary outcome)

- Defined as time from the date noted on the staging form until the date of death, or (if the patient is still alive) until the last date for which information exists

Progression-free survival (primary outcome)

- Defined as the time from the date of staging until the date of the first occurrence of one of the below-mentioned events, or (if none of these occurred) until the last documented date of continuous remission: progression of disease, relapse, death from any cause

Treatment response: assessed, not reported

HD16 (Continued)

Treatment-related mortality: not reported

Quality of life: assessed, not reported

Adverse events: reported for the overall PET-negative group, no comparison possible

Long-term toxicities:

- Secondary malignancies
- No other late toxicities reported

Notes

- **Funding source:** Deutsche Krebshilfe (108556, 111744) and the Swiss State Secretariat for Education, Research and Innovation (SERI)
- **Declaration of interest:** MF: Amgen, Affimed, Celgene, Takeda; RG: Celgene, Roche, Merck, Takeda, AstraZeneca, Novartis, Amgen, BMS, MSD Oncology, Sandoz, AbbVie, Gilead Sciences, Daiichi Sankyo, Janssen Pharmaceuticals; MST: Amgen, Regeneron, Roche, Novartis, Kite Pharmaceuticals, Gilead Sciences, Boehringer Ingelheim; ES: Celgene, Roche; UBK: Gilead Sciences, Amgen, BMS, Roche, Takeda, Merck Serono, Janssen-Cilag, Hexal, AstraZeneca, Celgene; SWK: MSD Oncology, Takeda; Siemens Healthcare Diagnostics; Gilead Sciences, Celgene; GM: Gilead Sciences, Bristol-Myers Squibb, Janssen-Cilag, AstraZeneca, Merck Serono; UD: Roche, Amgen; JM: MSD Oncology, Bristol-Myers Squibb, Takeda, Hexal, Celgene; AV: Roche, Pfizer, Amgen, Kite Pharmaceuticals, Gilead Sciences, Janssen, AbbVie, BMS; BvT: Roche, Takeda, MSD Oncology, Amgen, Pfizer, Novartis; AE: Hexal, Chugai Pharma, MSD Oncology, Takeda, BMS, Affimed Therapeutics.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Protocol: "..., the patient will be randomized by the GHSG Trial Coordination Center via telephone." . "The treating trial center will not yet be informed about the randomization result at this point in time."
Allocation concealment (selection bias)	Low risk	"Patients and investigators are blinded to treatment group allocation until documentation of PET-2 results."
Blinding (performance bias): All outcomes	Unclear risk	Participants and personnel were not blinded to treatment. This may have caused systematic differences in treatment or follow-up of participants: some patients (N = 18 in standard, N = 7 in experimental) swapped treatment based on personal preference, which impacts the effect of randomisation. The ITT and PP analyses reached different conclusions for this non-inferiority trial; therefore, we judged the RoB to be high.
Blinding of outcome assessment (detection bias) Mortality outcomes	Low risk	Although this was an open-label trial, we assumed that an unblinded outcome assessor would have no influence on the assessment of objective outcomes.
Blinding of outcome assessment (detection bias) Subjective outcomes (all except mortality)	High risk	Investigators, who were probably also the outcome assessors, were not blinded to treatment allocation. This may affect the assessment of subjective outcomes.
Incomplete outcome data (attrition bias) Efficacy outcomes	Low risk	55 participants dropped out after PET-2; the number and reasons were unbalanced across treatment groups. Analysis according to ITT and PP reached different conclusions. For this meta-analysis, ITT was used.
Incomplete outcome data (attrition bias) Safety outcomes	High risk	Secondary malignancies were reported only according to per-protocol population.

HD16 (Continued)

Selective reporting (re-reporting bias)	Unclear risk	Toxicities were reported only for the overall group; no comparison possible. Late toxicities such as cardiotoxicity and pulmonary toxicity were not (yet) reported. CR rate was not reported. The protocol on JCO website also mentioned QoL and partial response.
Other bias	Low risk	The study seemed to be free of other sources of bias. Although 143 participants dropped out before PET2, these were mainly due to restaging.

HD17

Study characteristics

Methods	<u>Study design</u> Randomised controlled trial to test PET-adapted treatment de-escalation • Randomisation before any treatment and PET-scan • Intervention arm with FDG-PET-based de-escalation for PET-negative participants and standard treatment for PET-positive participants • Control arm: Standard treatment for both arms • Analysed together for both PET-negative and -positive participants <u>Sample size</u> • Number randomised: - o PET-adaptation: 550 - o Standard: 546 • 3 HL diagnoses disconfirmed, 1 withdrew consent before start of treatment <u>PET-negative participants (ITT)</u> • PET-adaptation: 333 • Standard: 318
Participants	<u>Setting:</u> Multicentre study on early unfavourable HL including 224 study locations in Germany, Switzerland, Austria, and the Netherlands with enrolment between 13 January 2012 and 21 March 2017 <u>Eligibility criteria</u> <u>Inclusion criteria</u> • Newly diagnosed and histologically proven Hodgkin lymphoma • Stage: - o Stages IA, IB, IIA with one or more of the risk factors a-d, or - o Stage IIB with one or both of the risk factors c and d - o Risk factors: a) Large mediastinal mass ($\geq 1/3$ of the maximum transverse diameter of the thorax); b) extranodal diseases; c) High ESR (≥ 50 mm/h with A symptoms, ≥ 30 mm/h with B symptoms); d) 3 or more involved lymph node areas • Patient has not had any previous treatment for HL • Age at entry: 18-60 years • Written informed consent to participate in the trial • Agreement that personal data and tissue material will be used for the study, with due regard to data protection • Normal organ function (except HL-related) • Life expectancy > 3 months. <u>Exclusion criteria</u>

HD17 (Continued)

- Incomplete diagnosis of the disease stage
- Prior or concurrent disease that prevents treatment according to protocol (i.e. COPD with global respiratory insufficiency; symptomatic coronary heart disease; cardiomyopathy or cardiac insufficiency; serious uncontrolled hypertension; uncontrollable infection; poorly controlled diabetes mellitus; HIV infection; chronic active hepatitis B and/or C)
- Hodgkin lymphoma as part of a composite lymphoma
- Prior chemotherapy or radiation
- Malignant disease within the last five years (exceptions: basal cell carcinoma, carcinoma in situ of the cervix uteri, completely resected melanoma TNMpT1)
- Pregnancy, lactation
- Eastern Cooperative Oncology Group (ECOG) performance status > 2
- Long-term administration of corticosteroids (e.g. for chronic polyarthritis) or antineoplastic drugs (e.g. methotrexate)
- Patient's lack of accountability, inability to comprehend the nature, meaning and consequences of the trial and to express his/her own wishes accordingly
- 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 to any protocol medication
- Unsafe contraceptive methods
- Relationship of dependence or employer-employee relationship to the sponsor or the investigator
- Participation in another interventional trial that could interact with this trial

Participant characteristics (PET-negative and positive participants reported together)

Age (median, IQR)

- PET-adaptation: 31 (24 to 40) years
- Standard: 30 (24 to 39) years

Sex (female)

- PET-adaptation: 53%
- Standard: 54%

Stage (Ann Arbor)

- PET-adaptation: IA: 3%, IB: 2%, IIA: 67%, IIB: 26%, IIIA: < 1%
- Standard: IA: 4%, IB: 3%, IIA: 64%, IIB: 25%, IIIA: 0%

Bulky disease

- PET-adaptation: 55%
- Standard: 52%

Large mediastinal disease

- PET-adaptation: 18%
- Standard: 18%

NLPHL

- PET-adaptation: 2%
- Standard: 2%

Interventions

Initial therapy before PET-scan (same for all groups)

- 2 cycles of BEACOPP^{escalated}
- 2 cycles of ABVD

HD17 (Continued)

FDG-PET scan

- Timing: after 4 cycles of chemotherapy
- Scoring criteria: 5-point score according to Lugano criteria (negative: 1 or 2, positive: 3 to 5)
- Other information: Masked central review by a multidisciplinary expert panel (medical oncology, nuclear medicine, radiation oncology, radiology)

Interventions:

- **PET-negative participants** (median follow-up time)
 - o De-escalation: no further therapy
 - o Standard: + 30 Gy IF-RT
- **PET-positive participants** (median follow-up time)
 - o Both groups: + 30 Gy IN-RT
- Median follow-up time for both PET-positive and PET-negative participants
 - o PET-adaptation: 46.4 months (32.9 to 61.2)
 - o Standard: 45.8 months (32.4 to 61.0)

Outcomes

Overall survival (secondary outcome)

- Calculated from the date noted on the staging form until the date of death, or (if the patient is still alive) until the last date for which information exists

Progression-free survival (primary outcome)

- Defined as the time from the date of staging until the date of the first occurrence of progression of disease, relapse, or death from any cause; or (if none of these occurred) until the last documented date of continuous remission

Treatment response: Assessed, not reported

Treatment-related mortality: Reported for PET-positive and negative participants together

Quality of life: Assessed, not reported

Adverse events: AE grade 3-4 and SAE reported for both PET-positive and PET-negative together, AE grade 3-4 separate for chemotherapy and radiotherapy phase, therefore not analysed

Long-term toxicities:

- Second primary malignant neoplasm, any event
- Pulmonary and cardiac toxicity assessed, not yet published

Other outcomes:

- Fertility outcomes assessed, not yet published

Notes

- **Funding source**: Deutsche Krebshilfe
- **Declaration of interest**: BvT: Amgen, Pfizer, Gilead, Roche, Merck Sharpe & Dohme, Takeda, and Novartis; JM: Merck Sharpe & Dohme, Bristol Myers Squibb, Takeda, Hexal, and Celgene; PJB: Bristol Myers Squibb, Merck Sharpe & Dohme, Takeda, Bristol-Myers Squibb and Takeda; MH: AbbVie, Celgene, Gilead Science, Janssen, Mundipharma, Pharmacyclics, and Roche; RG: Celgene, Roche, Merck, Takeda, Astra Zeneca, Novartis, Amgen, Bristol Myers Squibb, Merck Sharpe & Dohme, Sandoz, Abbvie, Gilead, Daiichi Sankyo, and Janssen; All other authors declared no competing interests.
- The primary study authors provided additional data on the PET-negative ITT population.

Risk of bias

Bias

Authors' judgement

Support for judgement

HD17 (Continued)

Random sequence generation (selection bias)	Low risk	"Randomisation was done centrally using the minimisation method with a random component."
Allocation concealment (selection bias)	Low risk	Protocol: "The trial is blinded in the beginning, i.e. until PET-4 the treating physician will not be informed about which trial arm his patient was allocated to."
Blinding (performance bias): All outcomes	High risk	The trial was open-label. After PET-4 review, the number of treatment switchers/dropouts for PP analysis based on patient wish was not balanced between groups (58/486 patients [12%] in the standard arm with RT, 16/493 [3%] in the experimental arm). Of these, 33/318 [10%] of PET-4 negative patients in the standard arm refused IF-RT; 7/333 [2%] PET-4 negative patients in the experimental arm requested RT.
Blinding of outcome assessment (detection bias) Mortality outcomes	Low risk	Although this was an open-label trial, we assumed that an unblinded outcome assessor would have no influence on the assessment of objective outcomes.
Blinding of outcome assessment (detection bias) Subjective outcomes (all except mortality)	High risk	Investigators, who were probably also the outcome assessors, were not blinded to treatment allocation any more after PET-4. This may have affected the assessment of subjective outcomes.
Incomplete outcome data (attrition bias) Efficacy outcomes	Low risk	Censoring seemed balanced between treatment groups for PFS.
Incomplete outcome data (attrition bias) Safety outcomes	Low risk	No missing outcome data, safety set. Safety outcomes were reported according to treatment group for both PET-negative and -positive participants together; per phase and per grade.
Selective reporting (reporting bias)	Unclear risk	Not all outcomes were reported as planned in the protocol; treatment response, QoL and fertility outcomes were assessed but not (yet) reported.
Other bias	Unclear risk	The standard arm used IF-RT for both PET-positive and PET-negative participants, while the experimental arm had no further RT for PET-negative and IN-RT for PET-positive patients. This may affect comparability between PET-adapted and standard treatment arms.

HD18

Study characteristics

Methods

Study design

Randomised controlled trial to evaluate treatment de-escalation after negative PET-2 and treatment escalation with rituximab after positive PET-2 in advanced-stage HL

- Randomisation between the end of the 2nd and the beginning of the 4th cycle of chemotherapy
- Intervention arm with FDG-PET-based treatment adaptation
- De-escalation for PET-negative participants from 8 (or 6) to 4 cycles BEACOPP_{escalated}
- Escalation for PET-positive participants - addition of rituximab to 8 cycles of BEACOPP_{escalated}
- Control arm: Standard treatment of 8 cycles BEACOPP_{escalated}
- Safety amendment based on HD15 in June 2011: reduction of standard therapy from 8 to 6 cycles; PET-positive participants no longer randomised (simply received 6 cycles of BEACOPP_{escalated})

HD18 (Continued)

Sample size

- Number enrolled: 1100, number with PET-2: 1038

PET-negative participants randomised

- De-escalation: 505 (287 before amendment, 218 after amendment)
- Standard: 508 (290 before amendment, 218 after amendment)
 - 3 withdrawn due to HL diagnosis disconfirmation, 2 in de-escalation, 1 in standard

PET-positive participants randomised before the amendment

- Escalation: 220
- Standard: 220, 1 withdrew due to disconfirmation of HL diagnosis, 2 due to randomisation error

Participants

Setting: Multicentre study on advanced stage HL including 301 study locations in Germany, Switzerland, Austria, the Netherlands, and the Czech Republic with enrolment between 14 May 2008 and 18 July 2014

Eligibility criteria
Inclusion criteria

- Histologically proven primary diagnosis of Hodgkin lymphoma
- Stage: Clinical stage IIB with one or both of the following risk factors: a) large mediastinal mass (at least 1/3 of the maximum transverse diameter of the thorax), b) extranodal lesions Clinical stage III, IV.
- No previous treatment for Hodgkin lymphoma
- Age at inclusion: 18-60 years
- Written informed consent, consent to storage of data and tissue samples
- Normal organ function (except Hodgkin lymphoma-related)
- Negative HIV test
- Life expectancy > 3 months

Exclusion criteria

- Incomplete diagnosis of the disease stage
- Prior or concurrent disease that prevents treatment according to protocol (i.e. COPD with global respiratory insufficiency; symptomatic coronary heart disease; cardiomyopathy or cardiac insufficiency; serious uncontrolled hypertension; uncontrollable infection; poorly controlled diabetes mellitus; HIV infection; chronic active hepatitis B and/or C)
- Hodgkin lymphoma as composite lymphoma
- Malignant disease within the last 5 years (exceptions: basalioma, carcinoma in situ of the cervix uteri, completely resected melanoma TNMpT1)
- Pregnancy, lactation, unsafe contraceptive methods
- ECOG performance status > 28
- Long-term ingestion of corticosteroids (e.g. for chronic polyarthritis) or antineoplastic drugs (e.g. methotrexate)
- Lack of accountability, inability to appreciate the nature, meaning and consequences of the trial and to formulate their own wishes correspondingly
- Non-compliance: Refusal of blood products during treatment, epilepsy, drug dependency, change of residence to abroad, prior cerebral injury or similar circumstances
- General intolerance of any protocol medication
- Participation in another interventional trial that could interact with this trial

Participant characteristics
Age (median, IQR)

- **PET-negative (combined pre- and post-amendment)**
 - De-escalation: 33 (25 to 44) years

HD18 (Continued)

- o Standard: 32 (24 to 44) years
- **PET-positive (pre-amendment)**
 - o Escalation: 29 (23 to 41) years
 - o Standard: 30 (23 to 40) years

Sex (female)

- **PET-negative**
 - o De-escalation: 38%
 - o Standard: 37%
- **PET-positive (pre-amendment)**
 - o Escalation: 40%
 - o Standard: 41%

Stage (Ann Arbor)

- **PET-negative (pre- and post-amendment)**
 - o De-escalation: IIB: 8%, IIIA: 31%, IIIB: 24%, IVA: 12%, IVB: 24%
 - o Standard:
- **PET-positive (pre-amendment)**
 - o Escalation: IIB: 25%, IIIA: 13%, IIIB: 23%, IVA: 12%, IVB: 27%
 - o Standard: IIB: 22%, IIIA: 29%, IIIB: 27%, IVA: 9%, IVB: 22%

B-symptoms

- **PET-negative (pre- and post-amendment)**
 - o De-escalation: 56%
 - o Standard: 57%
- **PET-positive (pre-amendment)**
 - o Escalation: 75%
 - o Standard: 72%

Large mediastinal mass

- **PET-negative (pre- and post-amendment)**
 - o De-escalation: 18%
 - o Standard: 16%
- **PET-positive (pre-amendment)**
 - o Escalation: 45%
 - o Standard: 42%

NLPHL

- **PET-negative (pre- and post-amendment)**
 - o De-escalation: 5%
 - o Standard: 7%
- **PET-positive (pre-amendment)**
 - o Escalation: 6%
 - o Standard: 2%

Interventions

Initial therapy before PET-scan (same for all groups)

- 2 cycles of BEACOPP_{escalated}

FDG-PET scan

- Timing: between day 17 and day 21 of the second course of BEACOPP_{escalated} (3–7 days between the last day of treatment and the PET-2 imaging)
- Scoring criteria: 5-point score according to Lugano criteria (negative: 1 or 2, positive: 3 to 5)

HD18 (Continued)

- **Other information:** Central review by a multidisciplinary expert panel (medical oncology, nuclear medicine, radiation oncology, and radiology) masked to local findings

Interventions

- **PET-negative participants**
 - o De-escalation: + 2 cycles of BEACOPP_{escalated}
 - o Standard: + 6 cycles of BEACOPP_{escalated}; after safety amendment in June 2011: + 4 cycles of BEACOPP_{escalated}
- **PET-positive participants**
 - o Escalation: + 6 cycles of BEACOPP_{escalated} + rituximab
 - o Standard arm: + 6 cycles of BEACOPP_{escalated}
- PET-positive participants were randomised into two arms up to amendment 2 in June 2011. Afterwards, only 1 arm (+4 cycles BEACOPP_{escalated}) remained. These participants were not included in analysis.

Outcomes

Overall survival (secondary outcome)

- Calculated from the date noted on the staging form until the date of death, or (if the patient is alive) until the last date on which information exists

Progression-free survival (primary outcome)

- Defined as the time from the date of staging until the date of the first occurrence of progression of the disease, relapse, or death from any cause, or (if none of these occurred) until the last documented date of continuous remission

Treatment response: assessed, not reported

Treatment-related mortality: Cause of death toxicity of study treatment

Quality of life: assessed, not (yet) reported

Adverse events: AE grade 3-4 reported; other AE reported per specific AE

Long-term toxicities

- Secondary primary malignancies

Other outcomes

- Fertility outcomes reported

Notes

- **Funding:** Deutsche Krebshilfe, Swiss State Secretariat for Education and Research, and Roche Pharma AG
- **Declaration of interest:** SK, HG, IB, CK, DAE, JMZ, JMa, MFB, MS, H-JB, WW, W-DL, TP, MST, FH, MB, UBK, DK, HO, BH, WA, TV, HE, CB, HS, MF, VD, MD, and PB declared no competing interests. AM: Roche, Janssen, and Takeda outside the submitted work; RG: Celgene, Roche, Merck, Takeda, AstraZeneca, Novartis, Amgen, Bristol Myers Squibb, MSD, Sandoz, Abbvie, Gilead, Daiichi Sankyo, and Janssen outside the submitted work. JMe: MSD, Bristol Myers Squibb, Takeda, Celgene, and Hexal outside the submitted work; GM: Amgen, Bristol Myers Squibb, Janssen, AstraZeneca, Gilead, and Merck Serono, outside the submitted work; AE: Takeda/ Millennium, Bristol Myers Squibb, Hexal, Janssen, AstraZeneca, and Merck outside the submitted work

Risk of bias

Bias

Authors' judgement

Support for judgement

HD18 (Continued)

Random sequence generation (selection bias)	Low risk	"For both study questions the chance allocation of patients to treatment arms is/was effected by the trial coordination center according to the minimization method by Pocock (1983)."
Allocation concealment (selection bias)	Low risk	Central randomisation; information to physicians by fax or telephone
Blinding (performance bias): All outcomes	Low risk	"There is no blinding in this trial"; although participants and personnel were not blinded to treatment, this probably did not cause systematic differences in other treatment or follow-up of participants, as the follow-up schedules were preplanned for both groups.
Blinding of outcome assessment (detection bias) Mortality outcomes	Low risk	Although this was an open-label trial, we assumed that an unblinded outcome assessor would have no influence on the assessment of objective outcomes.
Blinding of outcome assessment (detection bias) Subjective outcomes (all except mortality)	High risk	There was no blinding in the study. This may have affected the assessment of subjective outcomes.
Incomplete outcome data (attrition bias) Efficacy outcomes	Low risk	Censoring seemed balanced between treatment groups for efficacy outcomes.
Incomplete outcome data (attrition bias) Safety outcomes	Low risk	No missing outcome data reported
Selective reporting (reporting bias)	Unclear risk	Some preplanned outcomes were not (yet) reported: quality of life, response assessments, and late toxicities (QoL and toxicities reported to be published later).
Other bias	Low risk	No other potential sources of bias identified

Picardi 2007

Study characteristics

Methods

Study design

Randomised trial to test if individuals with early-to-intermediate stage HL with a negative FDG-PET scan after 6 cycles of chemotherapy require further treatment with RT

- Randomisation after 6 cycles of chemotherapy and end-of-therapy negative PET-scan
- Intervention arm was de-escalated and received VEBEP (PET-adapted arm)
- Standard arm received VEBEP + 32 Gy radiotherapy to bulky area

Sample size

- Number enrolled:
 - o 260 consecutive participants

PET-positive participants

- 100 participants, no randomisation
- 70 with complete remission (CR) received no further treatment

Picardi 2007 (Continued)

- 30 were classified as non-responders and received 2 cycles of IGEV followed by high-dose chemotherapy and autologous PBSC transplantation.

PET-negative (160 participants randomised):

- 80 in control group
- 80 in intervention group

Participants

Setting: Single-centre study in Italy with recruitment between 2000 and 2006

Eligibility criteria

Inclusion criteria

- Newly diagnosed bulky HL (single lymph node or conglomerate nodal mass of ≥ 5 cm long axis)

For randomisation:

- Complete responders to induction chemotherapy
- Residual masses (i.e. $\geq 75\%$ reduction in tumour size at CT scans at the initial sites of bulky disease)
- PET-negative

Exclusion criteria

- Not reported

Participant baseline criteria (PET-negative only)

Age (median, range)

- De-escalation: 31 (15 to 70) years
- Standard: 30 (15 to 70) years

Sex (female)

- De-escalation: 45%
- Control group: 43.8%

Stage (Ann Arbor)

- De-escalation: I: 3.8%; II: 61.3%; III: 22.5%; IV: 12.5%
- Standard: I: 3.8%; II: 62.5%; III: 21.3%; IV: 12.5%

B-symptoms

- De-escalation: 55.0%
- Standard: 56.3%

Bulky disease

- De-escalation: supradiaphragmatic: 68.8%; subdiaphragmatic: 31.3%
- Standard: supradiaphragmatic: 70%; subdiaphragmatic: 30%

Interventions

Initial therapy before PET-scan

- 6 cycles VEBEP repeated every 4 weeks (vinblastine, etoposide, bleomycin, epirubicin, prednisone)

FDG-PET scans

Timing: after 6 chemotherapy cycles

Scoring criteria: negative: no evidence of uptake; positive: increased uptake in a focus within abnormal area

Picardi 2007 (Continued)

Other information

- Read by two nuclear medicine physicians and decided by consensus

Interventions

PET-negative participants

- o Intervention group: no further therapy
- o Standard: + 32 Gy RT (without boost on residual masses, fraction size being about 1.6 Gy five times a week)
- Median follow-up time 40 months (range 13-72 months) for both arms

Outcomes	<u>Overall survival</u>: no events <u>Progression-free survival</u>: was calculated from EFS (only one secondary malignancy event) <u>Treatment response</u>: response after PET-6 was inclusion criterion for randomisation. <u>Treatment-related mortality</u>: not reported <u>Quality of life</u>: not reported <u>Adverse events</u>: reported as haematological adverse events $\geq$ grade 2, non-haematological adverse events $\geq$ grade 2; not analysed as not comparable with other studies <u>Long-term toxicities</u>: not reported, follow-up too short <u>Other outcomes</u>: • Event-free survival (EFS, primary outcome), from start of therapy up to an event; event was defined as relapse, secondary malignancies, or death from any cause.
Notes	• Funding source: Supported by grants from Associazione Italiana contro le Leucemie (Salerno and benevento Sections) • Declaration of interest: The authors did not provide information on potential conflicts of interest. • Note: The study did not report hazard ratios. These were recalculated according to the methods of Tierney 2007.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Insufficient information about the sequence generation process ("They were randomly divided into two well-matched groups ...")
Allocation concealment (selection bias)	Unclear risk	No information on the method of allocation concealment
Blinding (performance bias): All outcomes	Low risk	The study did not address blinding of participants, physicians or outcome assessors. This probably did not cause systematic differences in other treatment or follow-up of participants, as the follow-up schedules were preplanned for both groups.
Blinding of outcome assessment (detection bias) Mortality outcomes	Low risk	The study publication did not explicitly report blinding. We assumed that an unblinded outcome assessor would have no influence on the assessment of objective outcomes.
Blinding of outcome assessment (detection bias)	High risk	The study publication did not explicitly report blinding. This may have affected the assessment of subjective outcomes.

Picardi 2007 (Continued)

Subjective outcomes (all except mortality)

Incomplete outcome data (attrition bias) Efficacy outcomes	Low risk	No missing outcome data reported
Incomplete outcome data (attrition bias) Safety outcomes	Low risk	No missing outcome data reported
Selective reporting (reporting bias)	Unclear risk	The study protocol was not available.
Other bias	Low risk	No other potential sources of bias identified

RATHL

Study characteristics

Methods	<u>Study design</u> Randomised controlled trial to evaluate FDG-PET-based de-escalation in PET-negative participants after 2 cycles of ABVD in advanced stage HL • Randomisation after negative PET-scan • After 2 cycles of ABVD, PET-negative participants were randomised to de-escalation arm receiving 4 cycles AVD as compared to 4 cycles of ABVD. • PET-positive participants received standard treatment. <u>Sample size</u> • Registered: 1201 participants* • Interim-PET-positive: 182 • Interim-PET-negative: 937 • Randomised: - o De-escalation: 464 - o Standard: 469 * Luminary 2023, 7-year update. Slightly more participants in the primary publication
Participants	<u>Setting:</u> Multicentre study on advanced stage HL including 138 study locations in the United Kingdom, Italy, Australia, New Zealand, Norway, Sweden, and Denmark, with enrolment between August 2008 and December 2012 <u>Eligibility criteria</u> <u>Inclusion criteria</u> • Histologically confirmed classical HL, clinical stage IIB, IIIA, IIIB or IV, or clinical stage IIA with adverse features (bulk mediastinal disease, outside the mediastinum, lymph node or lymph node mass greater than 10 cm in diameter, > two sites of disease, other poor risk features) • Aged 18 or above • No previous chemotherapy, radiotherapy or other investigational drug for HL • Performance status 0-3 • Adequate bone marrow function with platelets > 100x10⁹/L; neutrophils > 1.5x10⁹/L at the time of study entry unless lower numbers are attributed to bone marrow infiltration by lymphoma

RATHL (Continued)

- Serum creatinine less than 150% of the upper limit of normal, serum bilirubin less than twice the upper limit of normal and transaminases $< 2.5\times$ upper limit of normal unless attributed to lymphoma
- Patients with a significant history of ischaemic heart disease or hypertension must have an acceptable left ventricular ejection fraction (LVEF) $\geq 50\%$.
- Lung function tests showing diffusion capacity within 25% of normal predicted value for the patient
- Life expectancy > 3 months
- All patients of childbearing potential are willing to use adequate contraceptive precautions.
- Written, informed consent
- Access to PET-CT scanning

Exclusion criteria

- Poorly controlled diabetes mellitus
- Other concurrent uncontrolled medical condition
- Pregnant or lactating
- Central nervous system or meningeal involvement by lymphoma
- Cardiac contra-indication to doxorubicin: abnormal contractility on echocardiography or nuclear medicine examination (MUGA)
- Neurological contra-indication to chemotherapy (e.g. pre-existing neuropathy)
- General status that does not allow the administration of a full course of chemotherapy, according to the investigator
- Previous history of active malignant disease other than fully excised basal or squamous cell carcinoma of the skin or carcinoma in situ of the uterine cervix in the last 10 years
- Known positive serology for HIV, Hepatitis B or Hepatitis C
- Medical or psychiatric conditions that compromise the patient's ability to give informed consent

Participant characteristics

Age (median, range)

- **PET-negative participants**
 - De-escalation: 32 (18 to 76) years
 - Standard: 32 (18 to 79) years
- **PET-positive participants**
 - 32 (18 to 70) years

Sex (female)

- **PET-negative participants**
 - De-escalation: 46.3%
 - Standard: 45%
- **PET-positive participants**
 - 46.5%

Stage (Ann Arbor)

- **PET-negative participants**
 - De-escalation: II: 41.9%, III: 30.7%, IV: 27.4%
 - Standard: II: 41.4%, III: 33.6%, IV: 24.9%
- **PET-positive participants**
 - II: 42.4%, III: 19.8%, IV: 37.8%

B-symptoms

- **PET-negative participants**
 - De-escalation: 59.4%
 - Standard: 61.1%
- **PET-positive participants**
 - 70.3%

RATHL (Continued)

Bulky disease

- **PET-negative participants**
 - o De-escalation: 32.4%
 - o Standard: 28.35
- **PET-positive participants**
 - o 46.2%

Interventions

Initial therapy before PET-scan (same for all groups)

- 2 cycles of ABVD

FDG-PET scan

- Timing: between 7 and 10 days from the second dose of chemotherapy in cycle 2
- Scoring criteria: negative: score 1 to 3 (no uptake, slight uptake but lower than uptake in the normal mediastinal blood pool, or uptake equal to or slightly above uptake in the blood pool but less than uptake in the liver); positive: 4 or 5 (uptake moderately or markedly higher than uptake in the liver)
- Other information: Scans were scored by two readers at each core laboratory who were unaware of the patient's clinical status.

Interventions

- **PET-negative participants**
 - o De-escalation: + 4 cycles of AVD
 - o Standard: + 4 cycles of ABVD
- Median follow-up time for both groups: 41.2 months (range 2.0 to 79.7)
- **PET-positive participants**
 - o 3 cycles BEACOPP_{escalated} or 4 cycles BEACOPP14; PET-scan; PET-negative + 1 cycle BEACOPP_{escalated} or 2 cycles BEACOPP14; PET-positive + radiation or salvage

Outcomes

Overall survival (secondary outcome)

- Measured from date of randomisation to date of death from any cause; surviving patients will be censored at the date last known to be alive.

Progression-free survival (primary outcome)

- Measured from date of randomisation to date of first appearance of disease progression, relapse or death from any cause. Patients alive without progression or relapse will be censored at date last known to be alive.

Treatment response: assessed, not reported

Treatment-related mortality: Death owing to initial therapy

Quality of life: not assessed

Adverse events: Acute and chronic using NCI criteria; AE grade 3 to 4 reported

Long-term toxicities

- Secondary primary malignancies

Other outcomes: none

Notes

- **Funding source**: Cancer Research UK (reference CRUK/07/033), Leukaemia and Blood Cancer New Zealand, and Cancer Australia
- **Declaration of interest**: SB: National Institute for Health Research in England, and the Departments of Health for Scotland, Wales and Northern Ireland, Cancer Research UK; JR: UK National Cancer Re-

RATHL (Continued)

search Institute; LB, AMC, FdA, GE, MF, AF, AF; MJF, PJ, AK, SL, MOD, PP, TR, GS, PS, LS, JT, ZV: nothing to disclose

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	From Protocol: "Following central review of the images, patients should be randomised by the Lymphoma Trials Office."; from publication: "Randomization was performed by the Cancer Research UK and University College London Cancer Trials Centre, with stratification according to PET score and center".
Allocation concealment (selection bias)	Low risk	Central randomisation
Blinding (performance bias): All outcomes	Low risk	Although participants and personnel were not reported to be blinded to treatment, this probably did not cause systematic differences in other treatment or follow-up of participants (preplanned schedules); only two participants were reported to have changed their minds regarding the treatment assignment (randomised to AVD, wanted to receive ABVD).
Blinding of outcome assessment (detection bias) Mortality outcomes	Low risk	Although this was an open-label trial, we assumed that an unblinded outcome assessor would have no influence on the assessment of objective outcomes.
Blinding of outcome assessment (detection bias) Subjective outcomes (all except mortality)	High risk	Investigators, who were probably also the outcome assessors, were not blinded to treatment allocation. This may have affected the assessment of subjective outcomes.
Incomplete outcome data (attrition bias) Efficacy outcomes	Unclear risk	It is unclear whether the two groups were followed up for a similar amount of time, i.e. if censoring was balanced (not reported on KM curve).
Incomplete outcome data (attrition bias) Safety outcomes	Low risk	For short-term AE, no missing outcome data reported
Selective reporting (reporting bias)	Low risk	Outcomes were analysed as described in the protocol.
Other bias	Low risk	No other potential sources of bias identified

UK NCRI RAPID

Study characteristics

Methods

Study design

Randomised controlled trial to evaluate PET-3-based treatment de-escalation in early-stage classic HL

- Randomisation after 3 cycles of ABVD and PET-3 scan; only PET-negative participants were randomised
- Comparison of de-escalation (no further therapy) with standard IF-RT
- PET-positive participants received one more cycle of ABVD and follow-up as necessary.

Sample size

UK NCRI RAPID (Continued)

- Number screened: 602
 - 145 PET-positive, 31 without PET scan, 6 not randomised

PET-negative randomised:

- De-escalation: 211
- Standard: 209

Participants

Setting: Multicentre study on early stage classic HL including 94 study locations in the UK with enrolment between October 2003 and August 2010

Eligibility criteria

Inclusion criteria

- Age 16-75 years
- Histologically confirmed Hodgkin's disease
- Clinical stages IA/IIA above the diaphragm with no mediastinal bulk. Bulky disease at other sites (defined as nodal mass with transverse diameter > 10 cm) is acceptable.
- No previous treatment for Hodgkin's disease
- No previous malignancy excepting appropriately treated basal cell carcinoma of the skin or carcinoma in situ of the cervix
- No contraindications to chemotherapy or radiotherapy
- Patients with reproductive potential have a willingness to use contraception from entry into the study for a period of 6 months.
- Written informed consent

Exclusion criteria

- Patients with clinical stage IA Hodgkin's disease with no clinical or CT evidence of disease after diagnostic biopsy
- Pregnant or lactating women
- Severe underlying illness considered likely to make the trial therapy hazardous (for example, severe heart disease or lung fibrosis)
- Patients unwilling to travel to the nearest PET centre
- Patients unable to comply with follow-up arrangements

Participant characteristics(PET-negative only)

Age (median, range)

- De-escalation: 34 (16 to 75) years
- Standard: 24 (16 to 74) years

Sex (female)

- De-escalation: 49.3%
- Standard: 50.7%

Stage (Ann Arbor)

- De-escalation: IA: 65.9%, IIA: 66.8%
- Standard: IA: 33.0%, IIA: 67.0%

Nonmediastinal bulk present

- De-escalation: 0.5%
- Standard: 1%

Interventions

Initial therapy before PET-scan (same for all groups)

UK NCRI RAPID (Continued)

- 2 cycles of BEACOPP_{escalated}

FDG-PET scan

- Timing: after 2 cycles
- Scoring criteria: 5-point score according to Lugano criteria (negative: 1–3; positive: 4 or 5 positive)
- Other information:

Interventions (PET-negative participants; median follow-up 41.2 months (range 2.0 to 79.7))

- De-escalation: 4 cycles of AVD
- Standard: 4 cycles of ABVD

Outcomes	<u>Overall survival</u> • Defined as time from date of registration/randomisation to date of death from any cause; and cause of death <u>Progression-free survival</u> • Defined as time from the date of randomisation to first progression, relapse, or death, whichever occurred first <u>Treatment response</u>: Assessed, not reported <u>Treatment-related mortality</u>: Not reported <u>Quality of life</u>: Not assessed <u>Adverse events</u>: Not reported <u>Long-term toxicities</u>: Not reported <u>Other outcomes</u>: None
Notes	• Funding source: Supported by Leukaemia and Lymphoma Research, the Lymphoma Research Trust, Teenage Cancer Trust, and the U.K. Department of Health. • Declaration of interest: SB, AB, NC, DC, BH, PH, TI, AK, AM, MO, RP, BP, PS, JW: nothing to disclose; PJ: Millenium-Takeda; TAL: Gilead, Roche, Millenium, Spectrum, Celgene, GaxoSmithKline, Pfizer, Novartis, Johnson & Johnson; JR: Leukemia and Lymphoma research, Lymphoma Research Trust, Teenage Cancer Trust

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Protocol: "patients with a negative FDG-PET scan after 3 cycles of ABVD should be randomised by the Haematology Trials Group, UCL CTC"; central randomisation
Allocation concealment (selection bias)	Low risk	Central randomisation
Blinding (performance bias): All outcomes	Unclear risk	Although participants and personnel were not reported to be blinded to treatment, this probably did not cause systematic differences in other treatment options or follow-up of participants (preplanned schedules). This may have been a cause for 20 participants to change their minds regarding the treatment assignment (randomised to RT, refused RT).

UK NCRI RAPID *(Continued)*

Blinding of outcome assessment (detection bias) Mortality outcomes	Low risk	Although this was probably an open-label trial, we assumed that an unblinded outcome assessor would have no influence on the assessment of objective outcomes.
Blinding of outcome assessment (detection bias) Subjective outcomes (all except mortality)	High risk	Investigators, who were probably also the outcome assessors, were not blinded to treatment allocation. This may have affected the assessment of subjective outcomes.
Incomplete outcome data (attrition bias) Efficacy outcomes	Low risk	It is unclear whether the two groups were followed up for a similar amount of time, and thus whether censoring was balanced (not reported on KM curve).
Selective reporting (reporting bias)	Unclear risk	Outcome "incidence and type of second cancers" was not reported and there was no reference to future publications.
Other bias	Low risk	No other potential sources of bias identified

ABVD: adriamycin, bleomycin, vinblastine, and dacarbazine

AE: adverse events

ALT: alanine transaminase

AST: aspartate aminotransferase

AVD: adriamycin, vinblastine, and dacarbazine

BEACOPP: bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine, and prednisone

cHL: classic Hodgkin lymphoma

CI: confidence interval

CMT: combined-modality therapy

COPD: chronic obstructive pulmonary disease

CR: complete remission

cRT: consolidation radiotherapy

CRu: complete remission unconfirmed

CT: computed tomography

CTCAE: Common Terminology Criteria for Adverse Events

DFS: disease-free survival

DNA: Deoxyribonucleic acid

ECOG PS: Eastern Cooperative Oncology Group performance status

EF: ejection fraction

EFS: event-free survival

EORTC: European Organisation for Research and Treatment of Cancer

esc: escalated

ESR: erythrocyte sedimentation rate

FDG: fluorodeoxy-D-glucose

FEV: forced expiratory volume

Gy: Gray

HBc: Hepatitis B virus core

HBs: hepatitis B surface

HBV: hepatitis B virus

HCV: hepatitis C virus

HIV: human immunodeficiency virus

HL: Hodgkin lymphoma

HTLV: human T-lymphotropic virus

IFRT: Involved-field radiotherapy

IGEV: Ifosfamide, gemcitabine, vinorelbine

INRT: involved-node radiotherapy

IPS: international prognostic score

IQR: interquartile range

ITT: intention-to-treat

IV: intravenous

KM: Kaplan Meier

LNM: lymph node metastasis
 LVEF: left ventricular ejection fraction
 LYSARC: The Lymphoma Academic Research Organisation
 MT: magnetisation transfer
 MUGA: Multigated acquisition
 NCI: National Cancer Institute
 NFT: no further treatment
 NLPHL: nodular lymphocyte predominant Hodgkin lymphoma
 NYHA: New York Heart Association
 OS: overall survival
 PBSC: peripheral blood stem cell
 PD: progressive disease
 PET: positron emission tomography
 PFS: progression-free survival
 PP: per protocol
 PR: partial remission
 QoL: quality of life
 RNA: ribonucleic acid
 RFS: relapse-free survival
 RT: radiotherapy
 QoL: quality of life
 SAE: serious adverse events
 SUV_{max}: maximum standardised uptake values
 TNM_{pT1}: Tumour, Node, Metastasis (international tumour classification system)
 ULN: upper limit of normal
 VEBEP: vinblastine, etoposide, bleomycin, epirubicin, prednisone
 WHO: World Health Organization

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
AHOD0031	< 80% adults older than 18 years old
CALGB 50604	Non-randomised PET-adaptation trial
CALGB50801	Non-randomised PET-adaptation trial
ECHELON-1	Randomised but not PET-adapted
EORTC-COBRA	Non-randomised PET-adaptation trial
FIL	Initial therapy differs; ABVD vs. dose-dense ABVD
GATLA LH-05	Non-randomised PET-adaptation trial
H11	Early trial termination; no data available
HD15	Randomised but not PET-adapted
HD21	Randomised but not PET-adapted
JCOG1305	Non-randomised PET-adaptation trial
NCT03226249	Non-randomised PET-adaptation trial
NCT03712202	Initial therapy differed; brentuximab vedotin and nivolumab vs. ABVD and nivolumab vs. AVD and brentuximab vedotin and nivolumab

Study	Reason for exclusion
NCT05922904	Non-randomised PET-adaptation trial
RADAR	Initial therapy differed; ABVD vs. A2VD
RAFTING	Non-randomised PET-adaptation trial
SWOG S0816	Non-randomised PET-adaptation trial

A2VD: doxorubicin, brentuximab vedotin, vinblastine, dacarbazine

ABVD: adriamycin, bleomycin, vinblastine, dacarbazine

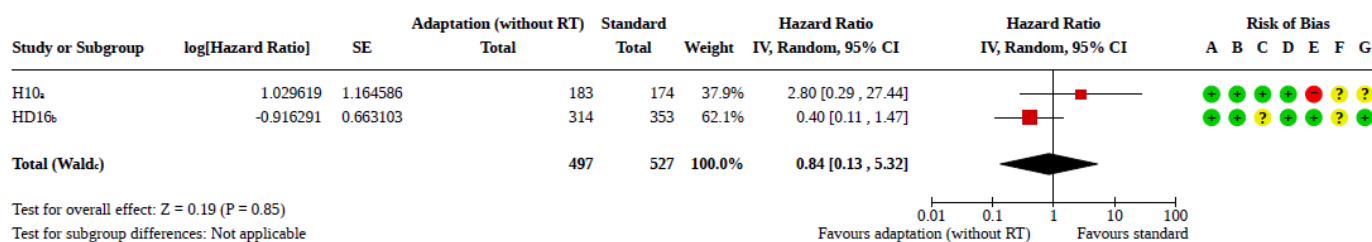
PET: positron emission tomography

vs.: versus

DATA AND ANALYSES

Comparison 1. PET-adapted treatment adaptation in PET-negative early stage HL (omission of radiotherapy, with or without more chemotherapy)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.1 Overall survival (OS)	2	1024	Hazard Ratio (IV, Random, 95% CI)	0.84 [0.13, 5.32]
1.2 Progression-free survival (PFS)	2	1034	Hazard Ratio (IV, Random, 95% CI)	4.52 [0.72, 28.41]
1.3 Treatment-related mortality	1	465	Risk Ratio (M-H, Random, 95% CI)	Not estimable
1.4 Incidence of second primary malignancy	2	983	Risk Ratio (M-H, Random, 95% CI)	0.83 [0.46, 1.49]

Analysis 1.1. Comparison 1: PET-adapted treatment adaptation in PET-negative early stage HL (omission of radiotherapy, with or without more chemotherapy), Outcome 1: Overall survival (OS)**Footnotes**

^aFavourable; de-escalation: 4 x ABVD, Standard: 3 x ABVD + 30 Gy IN-RT; ITT before amendment; 10-year FU

^bDe-escalation: 2 x ABVD, Standard: 2 x ABVD + 20 Gy IF-RT; median follow-up 66 months

^cCI calculated by Wald-type method.

^d Tau^2 calculated by DerSimonian and Laird method.

Risk of bias legend

(A) Random sequence generation (selection bias)

(B) Allocation concealment (selection bias)

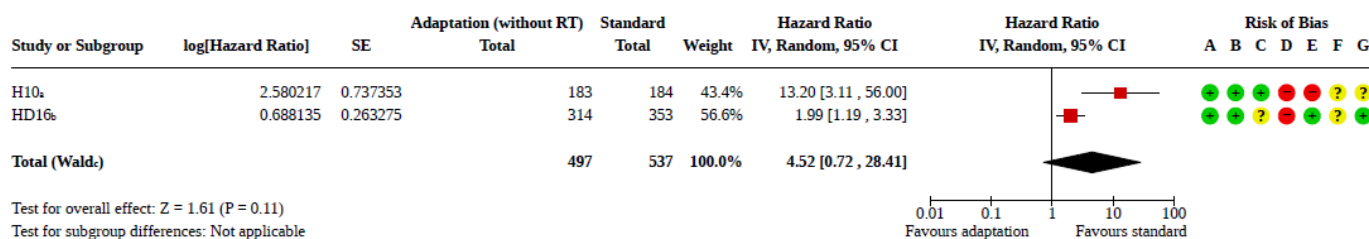
(C) Blinding (performance bias): All outcomes

(D) Blinding of outcome assessment (detection bias): Mortality outcomes

(E) Incomplete outcome data (attrition bias): Efficacy outcomes

(F) Selective reporting (reporting bias)

(G) Other bias

Analysis 1.2. Comparison 1: PET-adapted treatment adaptation in PET-negative early stage HL (omission of radiotherapy, with or without more chemotherapy), Outcome 2: Progression-free survival (PFS)**Footnotes**

^aFavourable; de-escalation: 4 x ABVD, Standard: 3 x ABVD + 30 Gy IN-RT; 10-year FU

^bDe-escalation: 2 x ABVD, Standard: 2 x ABVD + 20 Gy IF-RT; median follow-up 64 months

^cCI calculated by Wald-type method.

^d Tau^2 calculated by DerSimonian and Laird method.

Risk of bias legend

(A) Random sequence generation (selection bias)

(B) Allocation concealment (selection bias)

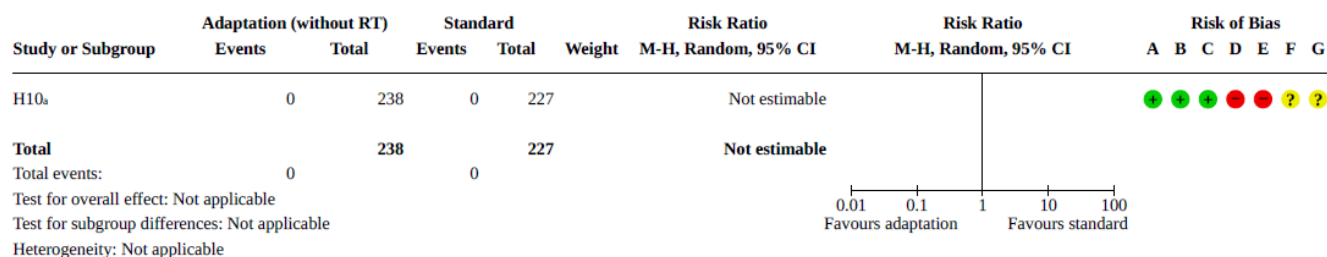
(C) Blinding (performance bias): All outcomes

(D) Blinding of outcome assessment (detection bias): Subjective outcomes (all except mortality)

(E) Incomplete outcome data (attrition bias): Efficacy outcomes

(F) Selective reporting (reporting bias)

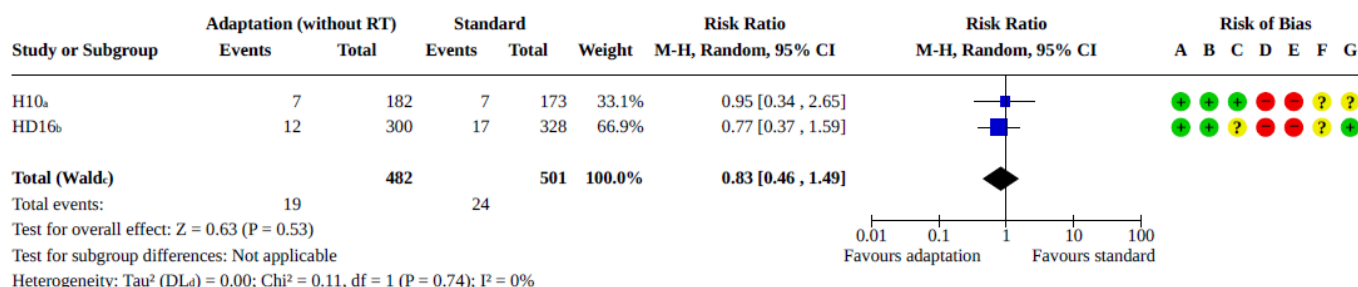
(G) Other bias

Analysis 1.3. Comparison 1: PET-adapted treatment adaptation in PET-negative early stage HL (omission of radiotherapy, with or without more chemotherapy), Outcome 3: Treatment-related mortality**Footnotes**

^aFavourable; de-escalation: 4 cycles ABVD, Standard: 3 cycles ABVD + 30 Gy IN-RT; 5-year FU

Risk of bias legend

- (A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding (performance bias): All outcomes
(D) Blinding of outcome assessment (detection bias): Subjective outcomes (all except mortality)
(E) Incomplete outcome data (attrition bias): Safety outcomes
(F) Selective reporting (reporting bias)
(G) Other bias

Analysis 1.4. Comparison 1: PET-adapted treatment adaptation in PET-negative early stage HL (omission of radiotherapy, with or without more chemotherapy), Outcome 4: Incidence of second primary malignancy**Footnotes**

^aFavourable; de-escalation: 4 x ABVD, Standard: 3 x ABVD + 30 Gy IN-RT; 10-year FU

^bDe-escalation: 2 x ABVD, Standard: 2 x ABVD + 20 Gy IF-RT; per protocol population, median follow-up 64 months

^cCI calculated by Wald-type method.

^dTau² calculated by DerSimonian and Laird method.

Risk of bias legend

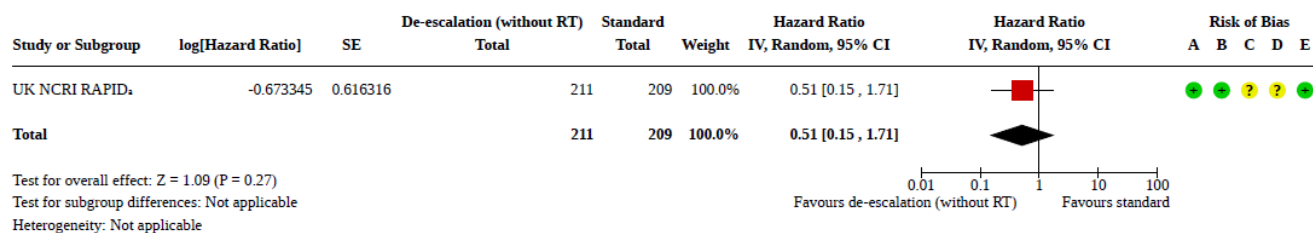
- (A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding (performance bias): All outcomes
(D) Blinding of outcome assessment (detection bias): Subjective outcomes (all except mortality)
(E) Incomplete outcome data (attrition bias): Safety outcomes
(F) Selective reporting (reporting bias)
(G) Other bias

Comparison 2. PET-adapted de-escalation in PET-negative, mixed early-and-intermediate-stage PET-negative HL (omission of radiotherapy)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
2.1 Overall survival (OS)	1	420	Hazard Ratio (IV, Random, 95% CI)	0.51 [0.15, 1.71]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
2.2 Progression-free survival (PFS)	2	580	Hazard Ratio (IV, Random, 95% CI)	2.00 [1.00, 3.97]

Analysis 2.1. Comparison 2: PET-adapted de-escalation in PET-negative, mixed early-and-intermediate-stage PET-negative HL (omission of radiotherapy), Outcome 1: Overall survival (OS)



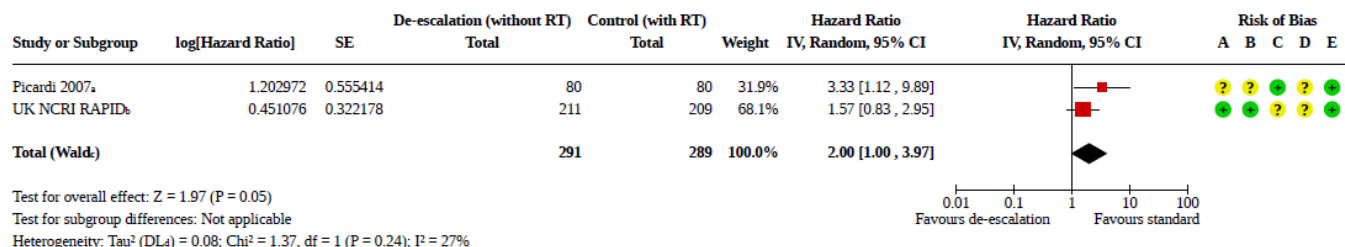
Footnotes

^aDe-escalation: 4 x ABVD, Standard: 3 x ABVD + 30 Gy IF-RT

Risk of bias legend

- (A) Random sequence generation (selection bias)
- (B) Allocation concealment (selection bias)
- (C) Blinding (performance bias): All outcomes
- (D) Selective reporting (reporting bias)
- (E) Other bias

Analysis 2.2. Comparison 2: PET-adapted de-escalation in PET-negative, mixed early-and-intermediate-stage PET-negative HL (omission of radiotherapy), Outcome 2: Progression-free survival (PFS)



Footnotes

- ^aDe-escalation: 6 x VEBEP, Standard: 6 x VEBEP + 32 Gy RT
- ^bDe-escalation: 4 x ABVD, Standard: 3 x ABVD + 30 Gy IF-RT
- ^cCI calculated by Wald-type method.
- ^d Tau^2 calculated by DerSimonian and Laird method.

Risk of bias legend

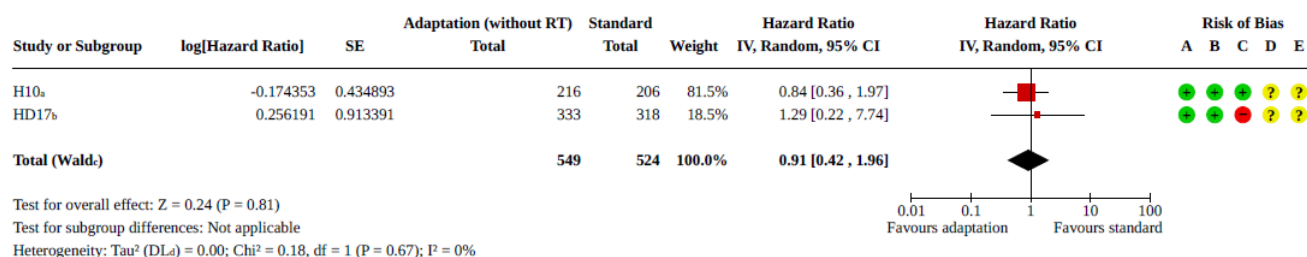
- (A) Random sequence generation (selection bias)
- (B) Allocation concealment (selection bias)
- (C) Blinding (performance bias): All outcomes
- (D) Selective reporting (reporting bias)
- (E) Other bias

Comparison 3. PET-adapted de-escalation in PET-negative intermediate-stage HL (ITT, except where noted)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
3.1 Overall survival (OS)	2	1073	Hazard Ratio (IV, Random, 95% CI)	0.91 [0.42, 1.96]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
3.2 Progression-free survival (PFS)	2	1073	Hazard Ratio (IV, Random, 95% CI)	1.59 [0.95, 2.67]
3.3 Treatment-related mortality	2	1690	Risk Ratio (M-H, Random, 95% CI)	2.94 [0.31, 28.20]
3.4 Serious adverse events	1	1096	Risk Ratio (M-H, Random, 95% CI)	1.01 [0.84, 1.21]
3.5 Incidence of second primary malignancy	2	1515	Risk Ratio (M-H, Random, 95% CI)	1.01 [0.57, 1.82]

Analysis 3.1. Comparison 3: PET-adapted de-escalation in PET-negative intermediate-stage HL (ITT, except where noted), Outcome 1: Overall survival (OS)

**Footnotes**

^aUnfavorable; de-escalation: 6 x ABVD, Standard: 4 x ABVD + 30 Gy IN-RT; 10-year FU

^bDe-escalation: 2 x BEACOPP^{escalated}, 2 x of ABVD, Standard: 2 x BEACOPP^{escalated}, 2 x of ABVD, + 30 Gy IF-RT

^cCI calculated by Wald-type method.

^d Tau^2 calculated by DerSimonian and Laird method.

Risk of bias legend

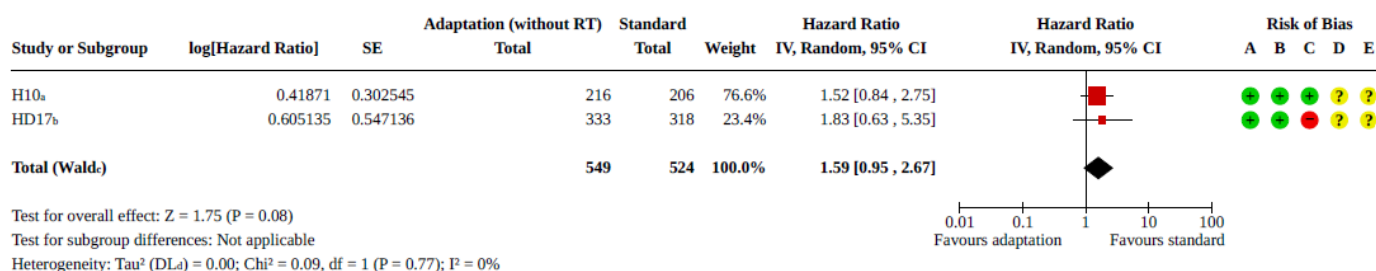
(A) Random sequence generation (selection bias)

(B) Allocation concealment (selection bias)

(C) Blinding (performance bias): All outcomes

(D) Selective reporting (reporting bias)

(E) Other bias

Analysis 3.2. Comparison 3: PET-adapted de-escalation in PET-negative intermediate-stage HL (ITT, except where noted), Outcome 2: Progression-free survival (PFS)**Footnotes**^aUnfavourable; de-escalation: 6 x ABVD, Standard: 4 x ABVD + 30 Gy IN-RT; 10-year FU^bDe-escalation: 2 x eBEACOPP, 2 x ABVD, Standard: 2 x eBEACOPP, 2 x ABVD, + 30 Gy IF-RT; ITT; PET-negative only, unpublished data^cCI calculated by Wald-type method.^d Tau^2 calculated by DerSimonian and Laird method.**Risk of bias legend**

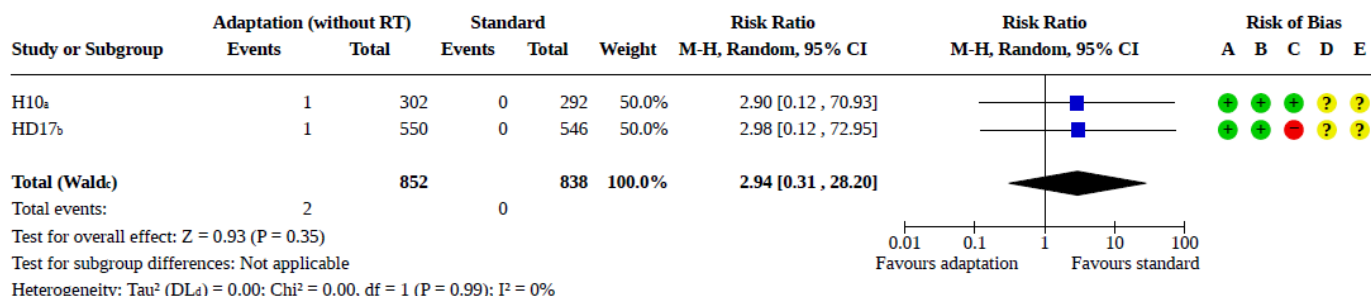
(A) Random sequence generation (selection bias)

(B) Allocation concealment (selection bias)

(C) Blinding (performance bias): All outcomes

(D) Selective reporting (reporting bias)

(E) Other bias

Analysis 3.3. Comparison 3: PET-adapted de-escalation in PET-negative intermediate-stage HL (ITT, except where noted), Outcome 3: Treatment-related mortality**Footnotes**^aUnfavourable; de-escalation: 6 cycles ABVD, Standard: 4 cycles ABVD + 30 Gy IN-RT; ITT before amendment, 5-year FU^bDe-escalation: 2 cycles BEACOPP^{escalated}, 2 cycles of ABVD, Standard: 2 cycles BEACOPP^{escalated}, 2 cycles of ABVD, + 30 Gy IF-RT; includes PET-positive participants^cCI calculated by Wald-type method.^d Tau^2 calculated by DerSimonian and Laird method.**Risk of bias legend**

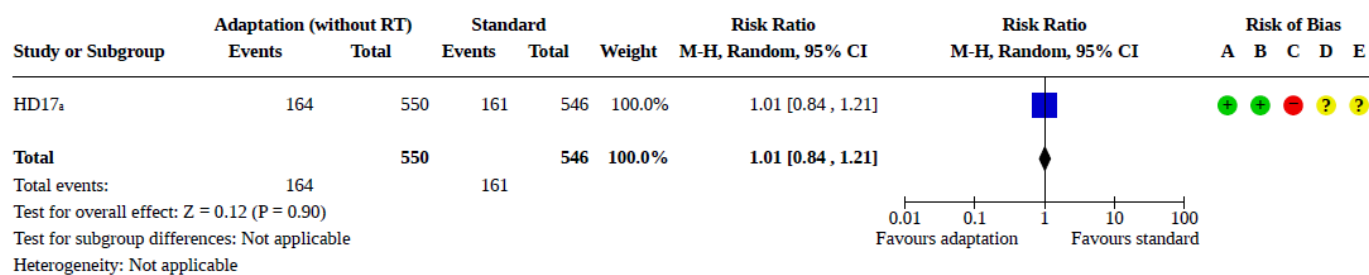
(A) Random sequence generation (selection bias)

(B) Allocation concealment (selection bias)

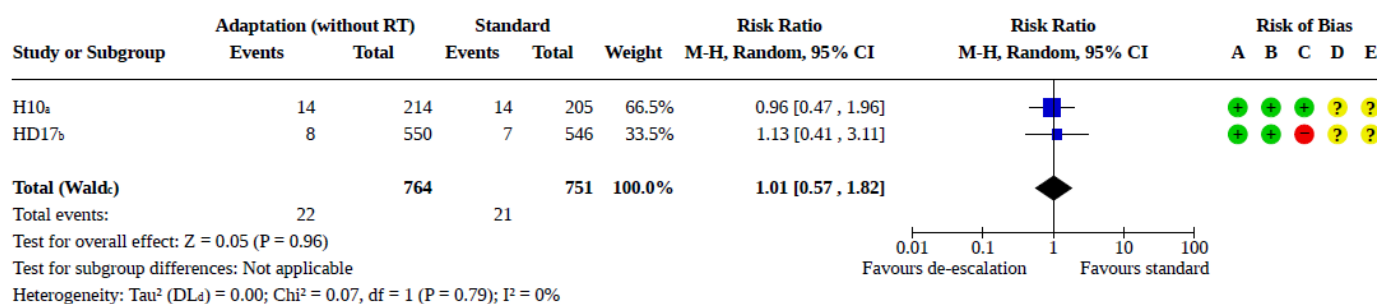
(C) Blinding (performance bias): All outcomes

(D) Selective reporting (reporting bias)

(E) Other bias

Analysis 3.4. Comparison 3: PET-adapted de-escalation in PET-negative intermediate-stage HL (ITT, except where noted), Outcome 4: Serious adverse events**Footnotes**^aDe-escalation: 2 x BEACOPPescalated, 2 x ABVD, Standard: 2 x BEACOPPescalated, 2 x ABVD, + 30 Gy IF-RT; PET-negative + positive**Risk of bias legend**

- (A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding (performance bias): All outcomes
(D) Selective reporting (reporting bias)
(E) Other bias

Analysis 3.5. Comparison 3: PET-adapted de-escalation in PET-negative intermediate-stage HL (ITT, except where noted), Outcome 5: Incidence of second primary malignancy**Footnotes**^aUnfavourable; de-escalation: 6 x ABVD, Standard: 4 x ABVD + 30 Gy IN-RT; 10-year FU^bDe-escalation: 2 x BEACOPPescalated, 2 x ABVD, Standard: 2 x BEACOPPescalated, 2 x ABVD + 30 Gy IF-RT; includes PET-positive participants^cCI calculated by Wald-type method.^dTau² calculated by DerSimonian and Laird method.**Risk of bias legend**

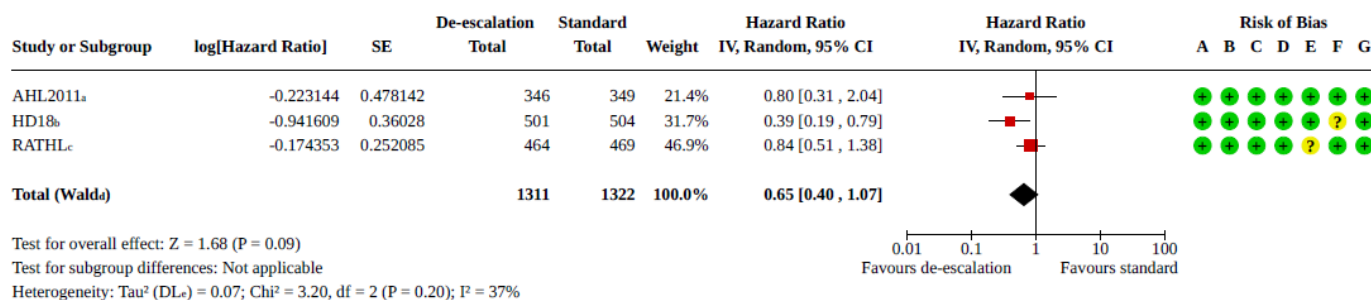
- (A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding (performance bias): All outcomes
(D) Selective reporting (reporting bias)
(E) Other bias

Comparison 4. PET-adapted de-escalation in PET-negative, advanced stage HL (ITT) - de-escalation of chemotherapy

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
4.1 Overall survival	3	2633	Hazard Ratio (IV, Random, 95% CI)	0.65 [0.40, 1.07]
4.2 Progression-free survival	3	2633	Hazard Ratio (IV, Random, 95% CI)	0.98 [0.78, 1.25]

Outcome or subgroup title	No. of studies	No. of partici-	Statistical method	Effect size
4.3 Treatment-related mortality	3	2761	Risk Ratio (M-H, Random, 95% CI)	0.21 [0.06, 0.73]
4.4 Adverse events, grade 3-4	2	1930	Risk Ratio (M-H, Random, 95% CI)	0.94 [0.92, 0.97]
4.5 Incidence of second primary malignancy	3	2757	Risk Ratio (M-H, Random, 95% CI)	0.87 [0.60, 1.26]

Analysis 4.1. Comparison 4: PET-adapted de-escalation in PET-negative, advanced stage HL (ITT) - de-escalation of chemotherapy, Outcome 1: Overall survival



Footnotes

^aDe-escalation: 2 cycles BEACOPP^{escalated}, 4 cycles ABVD; Standard: 6 cycles BEACOPP^{escalated}; HR calculated based on Tierney 2007, method 7

^bDe-escalation: 4 cycles BEACOPP^{escalated}; Standard: 8 or 6 cycles of BEACOPP^{escalated}

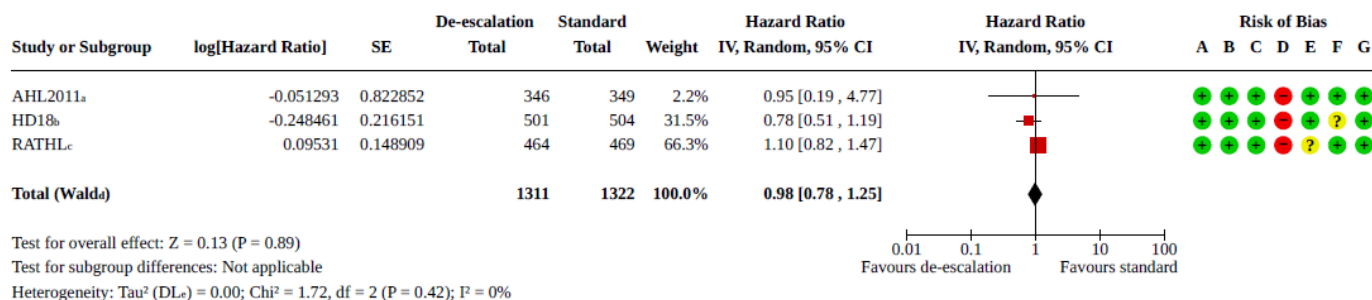
^cDe-escalation: 2 cycles ABVD, 4 cycles AVD; Standard: 6 cycles ABVD; 7-year FU

^dCI calculated by Wald-type method.

^e Tau^2 calculated by DerSimonian and Laird method.

Risk of bias legend

- (A) Random sequence generation (selection bias)
- (B) Allocation concealment (selection bias)
- (C) Blinding (performance bias): All outcomes
- (D) Blinding of outcome assessment (detection bias): Mortality outcomes
- (E) Incomplete outcome data (attrition bias): Efficacy outcomes
- (F) Selective reporting (reporting bias)
- (G) Other bias

Analysis 4.2. Comparison 4: PET-adapted de-escalation in PET-negative, advanced stage HL (ITT) - de-escalation of chemotherapy, Outcome 2: Progression-free survival**Footnotes**

^aDe-escalation: 2 cycles BEACOPP^{escalated}, 4 cycles ABVD; Standard: 6 cycles BEACOPP^{escalated}; HR calculated based on Tierney 2007, method 7

^bDe-escalation: 4 cycles BEACOPP^{escalated}; Standard: 8 or 6 cycles of BEACOPP^{escalated}

^cDe-escalation: 2 cycles ABVD, 4 cycles AVD; Standard: 6 cycles ABVD; 7-year follow-up

^aCI calculated by Wald-type method.

^e τ^2 calculated by DerSimonian and Laird method.

Risk of bias legend

(A) Random sequence generation (selection bias)

(B) Allocation concealment (selection bias)

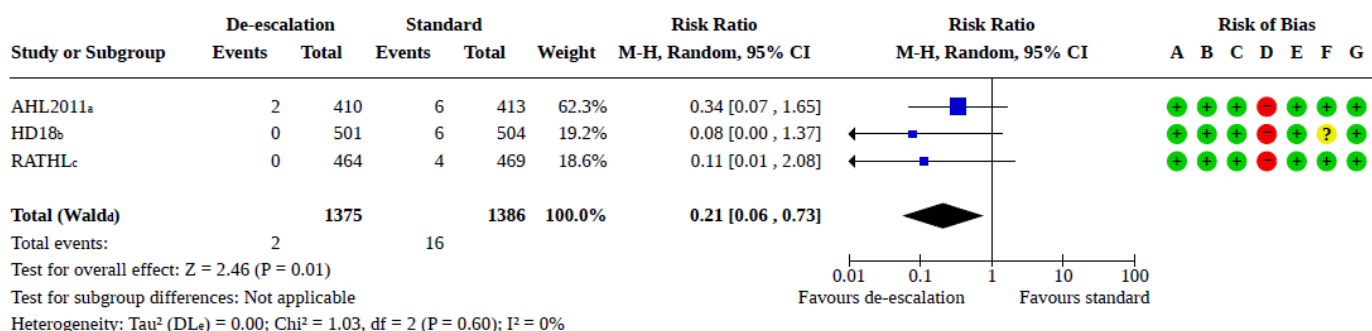
(C) Blinding (performance bias): All outcomes

(D) Blinding of outcome assessment (detection bias): Subjective outcomes (all except mortality)

(E) Incomplete outcome data (attrition bias): Efficacy outcomes

(F) Selective reporting (reporting bias)

(G) Other bias

Analysis 4.3. Comparison 4: PET-adapted de-escalation in PET-negative, advanced stage HL (ITT) - de-escalation of chemotherapy, Outcome 3: Treatment-related mortality**Footnotes**

^aDe-escalation: 2 cycles BEACOPP^{escalated}, 4 cycles ABVD; Standard: 6 cycles BEACOPP^{escalated}; the population also includes PET-positive participants

^bDe-escalation: 4 cycles BEACOPP^{escalated}; Standard: 8 or 6 cycles of BEACOPP^{escalated}

^cDe-escalation: 2 cycles ABVD, 4 cycles AVD; Standard: 6 cycles ABVD; treatment-related death due to first line treatment; 7-year FU

^aCI calculated by Wald-type method.

^e τ^2 calculated by DerSimonian and Laird method.

Risk of bias legend

(A) Random sequence generation (selection bias)

(B) Allocation concealment (selection bias)

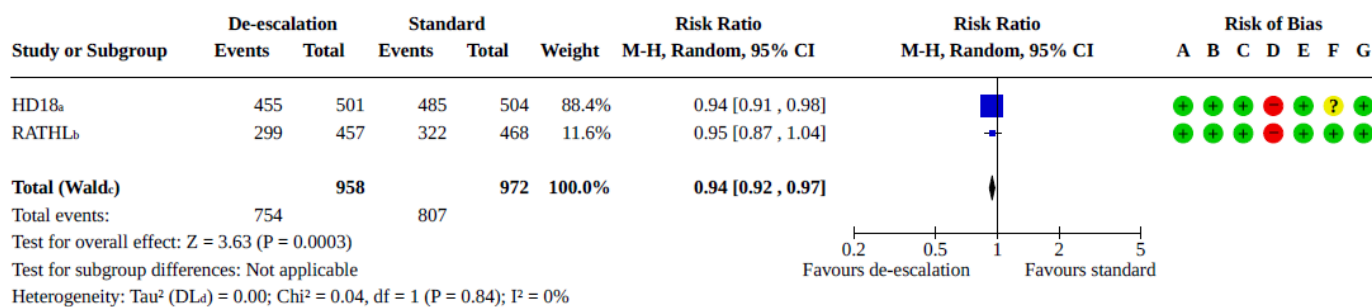
(C) Blinding (performance bias): All outcomes

(D) Blinding of outcome assessment (detection bias): Subjective outcomes (all except mortality)

(E) Incomplete outcome data (attrition bias): Safety outcomes

(F) Selective reporting (reporting bias)

(G) Other bias

Analysis 4.4. Comparison 4: PET-adapted de-escalation in PET-negative, advanced stage HL (ITT) - de-escalation of chemotherapy, Outcome 4: Adverse events, grade 3-4**Footnotes**

aDe-escalation: 4 cycles BEACOPP^{escalated}; Standard: 8 or 6 cycles of BEACOPP^{escalated}

bDe-escalation: 2 cycles ABVD, 4 cycles AVD; Standard: 6 cycles ABVD

cCI calculated by Wald-type method.

dTau² calculated by DerSimonian and Laird method.

Risk of bias legend

(A) Random sequence generation (selection bias)

(B) Allocation concealment (selection bias)

(C) Blinding (performance bias): All outcomes

(D) Blinding of outcome assessment (detection bias): Subjective outcomes (all except mortality)

(E) Incomplete outcome data (attrition bias): Safety outcomes

(F) Selective reporting (reporting bias)

(G) Other bias

Analysis 4.5. Comparison 4: PET-adapted de-escalation in PET-negative, advanced stage HL (ITT) - de-escalation of chemotherapy, Outcome 5: Incidence of second primary malignancy**Footnotes**

aDe-escalation: 2 cycles BEACOPP^{escalated}, 4 cycles ABVD; Standard: 6 cycles BEACOPP^{escalated}; the population also includes PET-positive participants

bDe-escalation: 4 cycles BEACOPP^{escalated}; Standard: 8 or 6 cycles of BEACOPP^{escalated}

cDe-escalation: 2 cycles ABVD, 4 cycles AVD; Standard: 6 cycles ABVD; 7-year FU

dCI calculated by Wald-type method.

eTau² calculated by DerSimonian and Laird method.

Risk of bias legend

(A) Random sequence generation (selection bias)

(B) Allocation concealment (selection bias)

(C) Blinding (performance bias): All outcomes

(D) Blinding of outcome assessment (detection bias): Subjective outcomes (all except mortality)

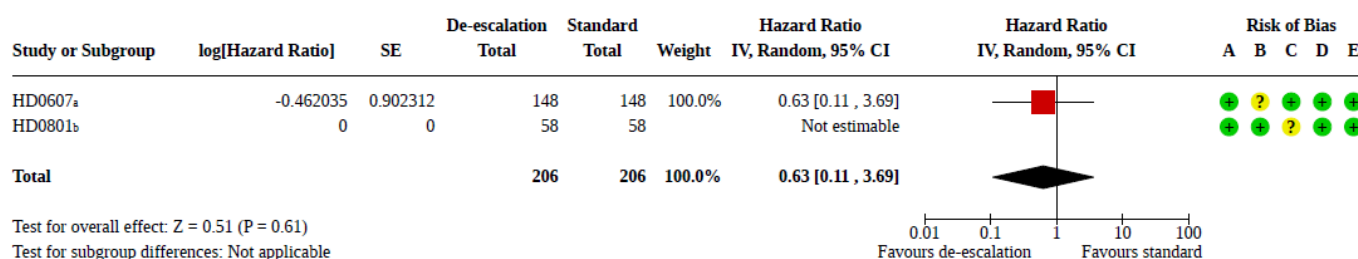
(E) Incomplete outcome data (attrition bias): Safety outcomes

(F) Selective reporting (reporting bias)

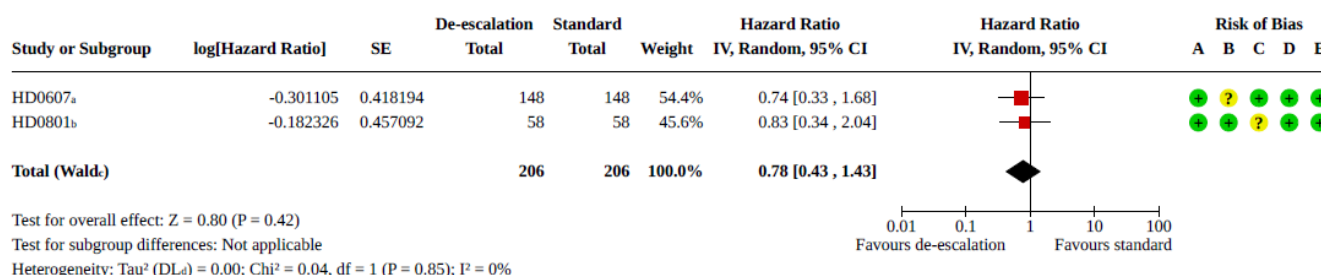
(G) Other bias

Comparison 5. PET-adapted de-escalation in PET-negative, advanced stage HL (ITT) - RT vs. no further RT

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
5.1 Overall survival	2	412	Hazard Ratio (IV, Random, 95% CI)	0.63 [0.11, 3.69]
5.2 Progression-free survival	2	412	Hazard Ratio (IV, Random, 95% CI)	0.78 [0.43, 1.43]
5.3 Incidence of second primary malignancy	2	349	Risk Ratio (M-H, Random, 95% CI)	0.41 [0.08, 2.09]

Analysis 5.1. Comparison 5: PET-adapted de-escalation in PET-negative, advanced stage HL (ITT) - RT vs. no further RT, Outcome 1: Overall survival**Footnotes**^aDe-escalation: 6 x ABVD, Standard: 6 x ABVD + 30 Gy RT, LNM at baseline and PET-2 and PET-6 negative; data provided by the primary study authors^bDe-escalation: 6 x ABVD, Standard: 6 x ABVD + 30 Gy RT, bulky mass at baseline and PET-2 and PET-6 negative; only one participant died up to data cutoff**Risk of bias legend**

- (A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding (performance bias): All outcomes
(D) Selective reporting (reporting bias)
(E) Other bias

Analysis 5.2. Comparison 5: PET-adapted de-escalation in PET-negative, advanced stage HL (ITT) - RT vs. no further RT, Outcome 2: Progression-free survival**Footnotes**^aDe-escalation: 6 x ABVD, Standard: 6 x ABVD + 30 Gy RT, LNM at baseline and PET-2 and PET-6 negative; data provided by the primary study authors^bDe-escalation: 6 x ABVD, Standard: 6 x ABVD + 30 Gy RT, bulky mass at baseline and PET-2 and PET-6 negative^cCI calculated by Wald-type method.^d τ^2 calculated by DerSimonian and Laird method.**Risk of bias legend**

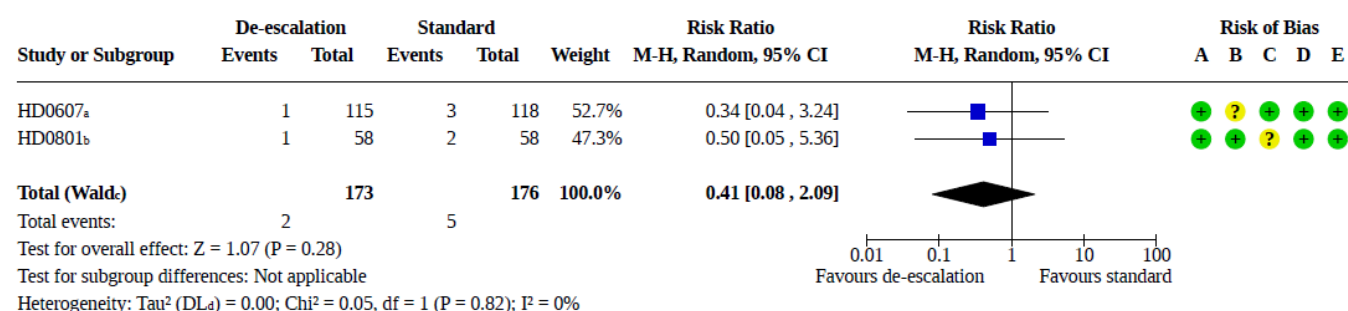
(A) Random sequence generation (selection bias)

(B) Allocation concealment (selection bias)

(C) Blinding (performance bias): All outcomes

(D) Selective reporting (reporting bias)

(E) Other bias

Analysis 5.3. Comparison 5: PET-adapted de-escalation in PET-negative, advanced stage HL (ITT) - RT vs. no further RT, Outcome 3: Incidence of second primary malignancy**Footnotes**^aDe-escalation: 6 x ABVD, Standard: 6 x ABVD + 30 Gy RT, LNM at baseline and PET-2 and PET-6 negative^bDe-escalation: 6 x ABVD, Standard: 6 x ABVD + 30 Gy RT, bulky mass at baseline and PET-2 and PET-6 negative^cCI calculated by Wald-type method.^d τ^2 calculated by DerSimonian and Laird method.**Risk of bias legend**

(A) Random sequence generation (selection bias)

(B) Allocation concealment (selection bias)

(C) Blinding (performance bias): All outcomes

(D) Selective reporting (reporting bias)

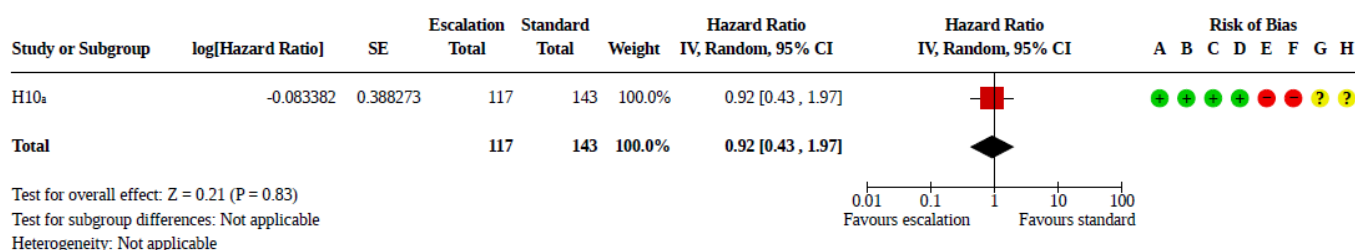
(E) Other bias

Comparison 6. PET-adapted escalation in PET-positive, mixed early-and-intermediate-stage HL (ITT, except where noted)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
6.1 Overall survival	1	260	Hazard Ratio (IV, Random, 95% CI)	0.92 [0.43, 1.97]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
6.2 Progression-free survival	1	260	Hazard Ratio (IV, Random, 95% CI)	0.67 [0.29, 1.55]
6.3 Incidence of second primary malignancy	1	234	Risk Ratio (M-H, Random, 95% CI)	1.23 [0.43, 3.55]

Analysis 6.1. Comparison 6: PET-adapted escalation in PET-positive, mixed early-and-intermediate-stage HL (ITT, except where noted), Outcome 1: Overall survival



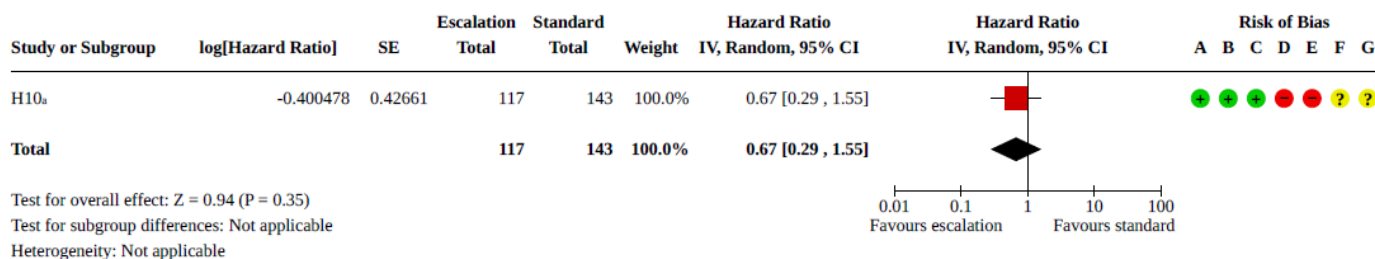
Footnotes

^aEscalation: 2 x ABVD + 2 x eBEACOPP + IN-RT, Standard F: 3 x ABVD + IN-RT, Standard U: 4 x ABVD + IN-RT; 10-year FU

Risk of bias legend

- (A) Random sequence generation (selection bias)
- (B) Allocation concealment (selection bias)
- (C) Blinding (performance bias): All outcomes
- (D) Blinding of outcome assessment (detection bias): Mortality outcomes
- (E) Blinding of outcome assessment (detection bias): Subjective outcomes (all except mortality)
- (F) Incomplete outcome data (attrition bias): Efficacy outcomes
- (G) Selective reporting (reporting bias)
- (H) Other bias

Analysis 6.2. Comparison 6: PET-adapted escalation in PET-positive, mixed early-and-intermediate-stage HL (ITT, except where noted), Outcome 2: Progression-free survival

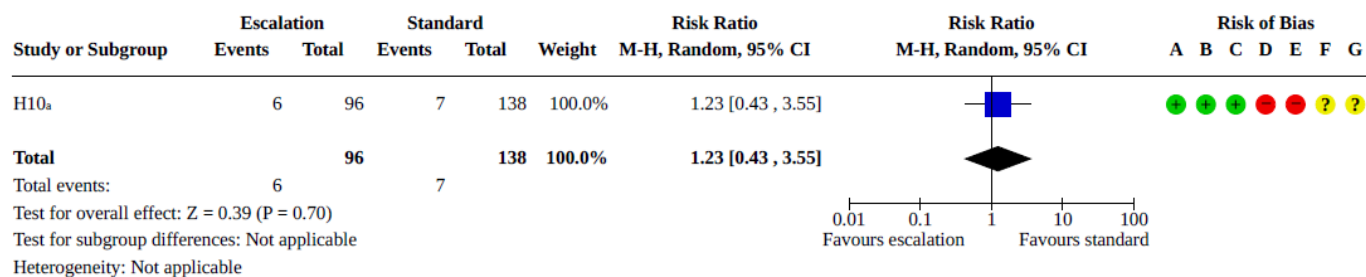


Footnotes

^aEscalation: 2 x ABVD + 2 x eBEACOPP + IN-RT, Standard F: 3 x ABVD + IN-RT, Standard U: 4 x ABVD + IN-RT; 10-year FU

Risk of bias legend

- (A) Random sequence generation (selection bias)
- (B) Allocation concealment (selection bias)
- (C) Blinding (performance bias): All outcomes
- (D) Blinding of outcome assessment (detection bias): Subjective outcomes (all except mortality)
- (E) Incomplete outcome data (attrition bias): Efficacy outcomes
- (F) Selective reporting (reporting bias)
- (G) Other bias

Analysis 6.3. Comparison 6: PET-adapted escalation in PET-positive, mixed early-and-intermediate-stage HL (ITT, except where noted), Outcome 3: Incidence of second primary malignancy**Footnotes**

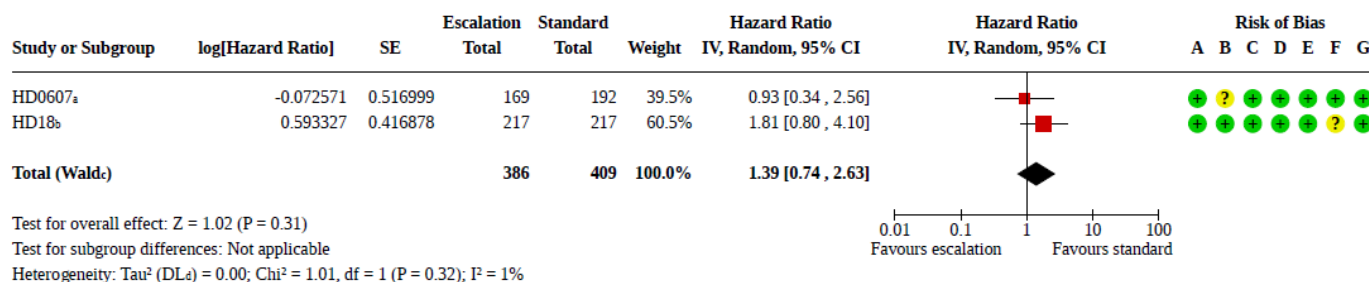
aEscalation: 2 x ABVD + 2 x eBEACOPP + IN-RT, Standard F: 3 x ABVD + IN-RT, Standard U: 4 x ABVD + IN-RT; 10-year FU

Risk of bias legend

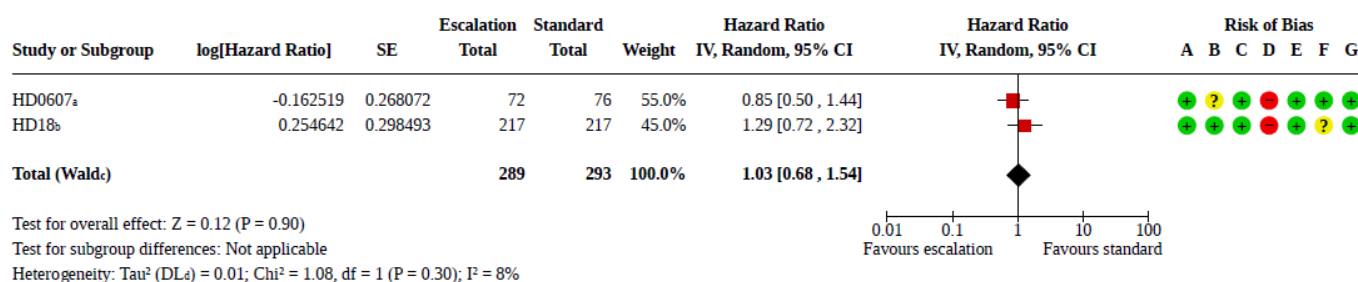
- (A) Random sequence generation (selection bias)
- (B) Allocation concealment (selection bias)
- (C) Blinding (performance bias): All outcomes
- (D) Blinding of outcome assessment (detection bias): Subjective outcomes (all except mortality)
- (E) Incomplete outcome data (attrition bias): Safety outcomes
- (F) Selective reporting (reporting bias)
- (G) Other bias

Comparison 7. PET-adapted escalation in PET-positive, advanced-stage HL (ITT, except where noted)

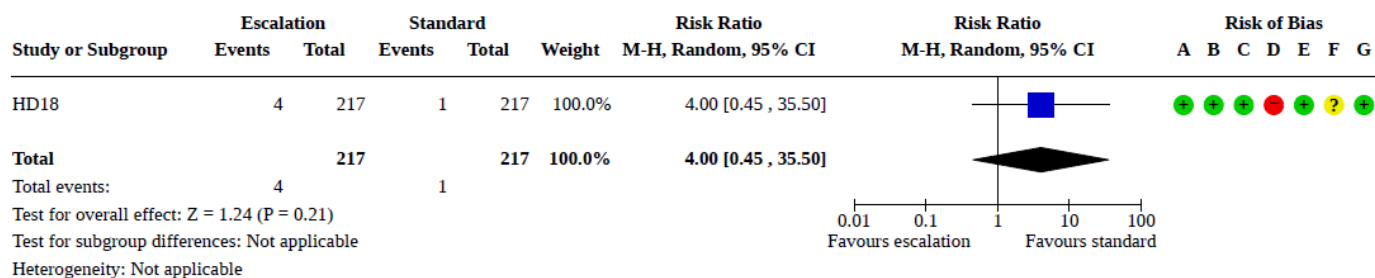
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
7.1 Overall survival	2	795	Hazard Ratio (IV, Random, 95% CI)	1.39 [0.74, 2.63]
7.2 Progression-free survival	2	582	Hazard Ratio (IV, Random, 95% CI)	1.03 [0.68, 1.54]
7.3 Treatment-related mortality	1	434	Risk Ratio (M-H, Random, 95% CI)	4.00 [0.45, 35.50]
7.4 Serious adverse events	1	148	Risk Ratio (M-H, Random, 95% CI)	1.61 [1.00, 2.60]
7.5 Adverse events, grade 3-4	2	586	Risk Ratio (M-H, Random, 95% CI)	1.06 [0.77, 1.45]
7.6 Any grade adverse event	1	148	Risk Ratio (M-H, Random, 95% CI)	1.16 [0.89, 1.51]
7.7 Incidence of secondary malignancy	2	582	Risk Ratio (M-H, Random, 95% CI)	0.67 [0.28, 1.60]

Analysis 7.1. Comparison 7: PET-adapted escalation in PET-positive, advanced-stage HL (ITT, except where noted), Outcome 1: Overall survival**Footnotes**^aEscalation: 2 x ABVD + 4 x BEACOPP^{escalated} + 4 x BEACOPP, + rituximab, Standard: 2 x ABVD + 4 x BEACOPP^{escalated} + 4 x BEACOPP^bEscalation: 8 x BEACOPP^{escalated} + rituximab, Standard: 8 x BEACOPP^{escalated}, median FU 73 months; pre-amendment cohort^cCI calculated by Wald-type method.^d τ^2 calculated by DerSimonian and Laird method.**Risk of bias legend**

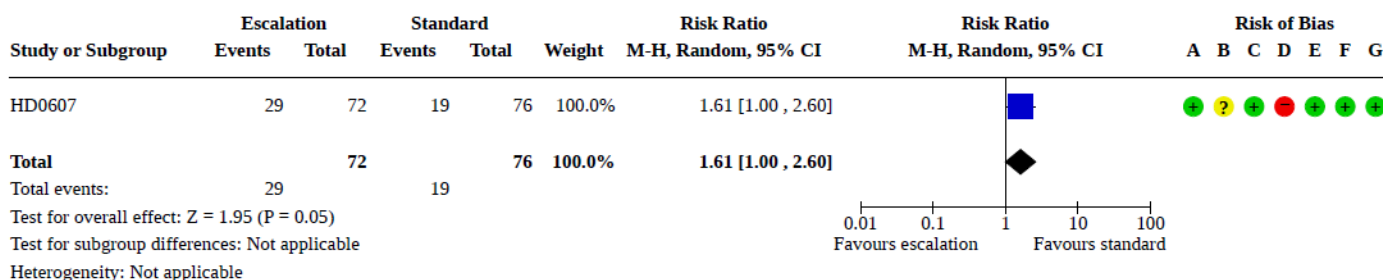
- (A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding (performance bias): All outcomes
(D) Blinding of outcome assessment (detection bias): Mortality outcomes
(E) Incomplete outcome data (attrition bias): Efficacy outcomes
(F) Selective reporting (reporting bias)
(G) Other bias

Analysis 7.2. Comparison 7: PET-adapted escalation in PET-positive, advanced- stage HL (ITT, except where noted), Outcome 2: Progression-free survival**Footnotes**^aEscalation: 2 x ABVD + 4 x BEACOPP^{escalated} + 4 x BEACOPP + R, Standard: 2 x ABVD + 4 x BEACOPP^{escalated} + 4 x BEACOPP; median FU 43.2 months^bEscalation: 8 x BEACOPP^{escalated} + rituximab, Standard: 8 x BEACOPP^{escalated}, median FU 73 months; pre-amendment cohort^cCI calculated by Wald-type method.^d τ^2 calculated by DerSimonian and Laird method.**Risk of bias legend**

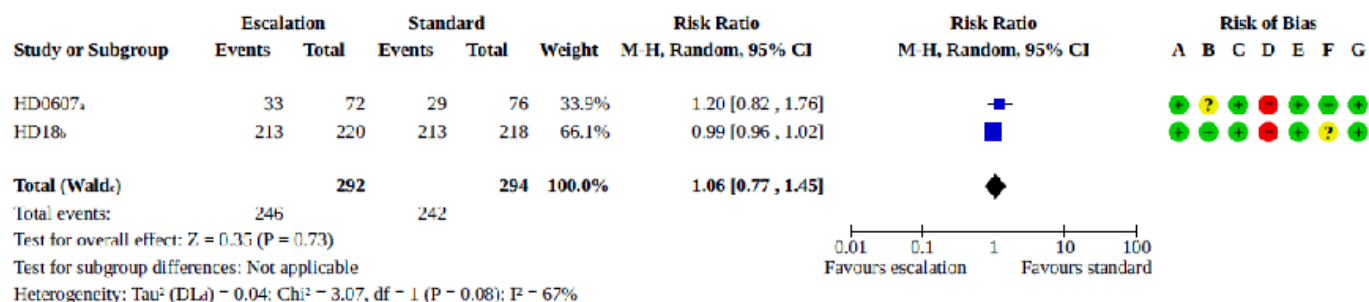
- (A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding (performance bias): All outcomes
(D) Blinding of outcome assessment (detection bias): Subjective outcomes (all except mortality)
(E) Incomplete outcome data (attrition bias): Efficacy outcomes
(F) Selective reporting (reporting bias)
(G) Other bias

Analysis 7.3. Comparison 7: PET-adapted escalation in PET-positive, advanced-stage HL (ITT, except where noted), Outcome 3: Treatment-related mortality**Risk of bias legend**

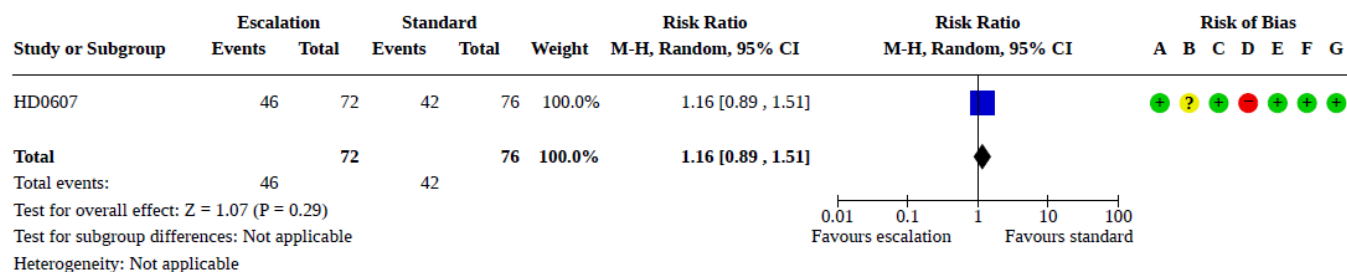
- (A) Random sequence generation (selection bias)
- (B) Allocation concealment (selection bias)
- (C) Blinding (performance bias): All outcomes
- (D) Blinding of outcome assessment (detection bias): Subjective outcomes (all except mortality)
- (E) Incomplete outcome data (attrition bias): Safety outcomes
- (F) Selective reporting (reporting bias)
- (G) Other bias

Analysis 7.4. Comparison 7: PET-adapted escalation in PET-positive, advanced-stage HL (ITT, except where noted), Outcome 4: Serious adverse events**Risk of bias legend**

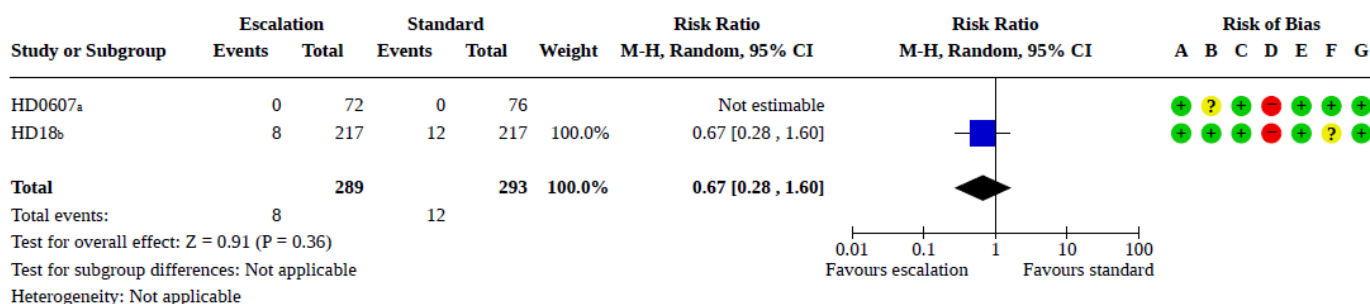
- (A) Random sequence generation (selection bias)
- (B) Allocation concealment (selection bias)
- (C) Blinding (performance bias): All outcomes
- (D) Blinding of outcome assessment (detection bias): Subjective outcomes (all except mortality)
- (E) Incomplete outcome data (attrition bias): Safety outcomes
- (F) Selective reporting (reporting bias)
- (G) Other bias

Analysis 7.5. Comparison 7: PET-adapted escalation in PET-positive, advanced-stage HL (ITT, except where noted), Outcome 5: Adverse events, grade 3-4**Footnotes**^aEscalation: 2 x ABVD + 4 x BEACOPP^{escalated} + 4 x BEACOPP + rituximab, Standard: 2 x ABVD + 4 x BEACOPP^{escalated} + 4 x BEACOPP, unpublished data^bEscalation: 8 x BEACOPP^{escalated} + rituximab, Standard: 8 x BEACOPP^{escalated}, pre-amendment cohort^cCI calculated by Wald-type method.^dTau² calculated by DerSimonian and Laird method.**Risk of bias legend**

- (A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding (performance bias): All outcomes
(D) Blinding of outcome assessment (detection bias): Subjective outcomes (all except mortality)
(E) Incomplete outcome data (attrition bias): Safety outcomes
(F) Selective reporting (reporting bias)
(G) Other bias

Analysis 7.6. Comparison 7: PET-adapted escalation in PET-positive, advanced-stage HL (ITT, except where noted), Outcome 6: Any grade adverse event**Risk of bias legend**

- (A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding (performance bias): All outcomes
(D) Blinding of outcome assessment (detection bias): Subjective outcomes (all except mortality)
(E) Incomplete outcome data (attrition bias): Safety outcomes
(F) Selective reporting (reporting bias)
(G) Other bias

Analysis 7.7. Comparison 7: PET-adapted escalation in PET-positive, advanced-stage HL (ITT, except where noted), Outcome 7: Incidence of secondary malignancy**Footnotes**^aEscalation: 2 x ABVD + 4 x BEACOPP_{escalated} + 4 x BEACOPP + R, Standard: 2 x ABVD + 4 x BEACOPP_{escalated} + 4 x BEACOPP; median FU 43.2 months^bEscalation: 8 x BEACOPP_{escalated} + rituximab, Standard: 8 x BEACOPP_{escalated}, median FU 73 months; pre-amendment cohort**Risk of bias legend**

(A) Random sequence generation (selection bias)

(B) Allocation concealment (selection bias)

(C) Blinding (performance bias): All outcomes

(D) Blinding of outcome assessment (detection bias): Subjective outcomes (all except mortality)

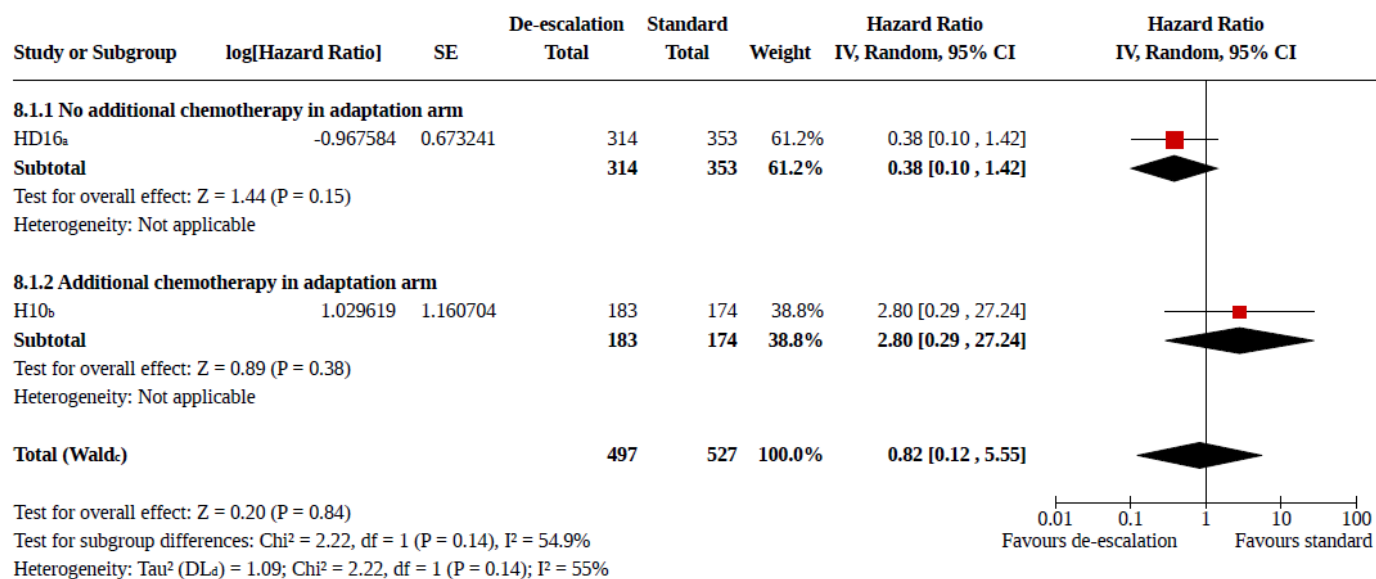
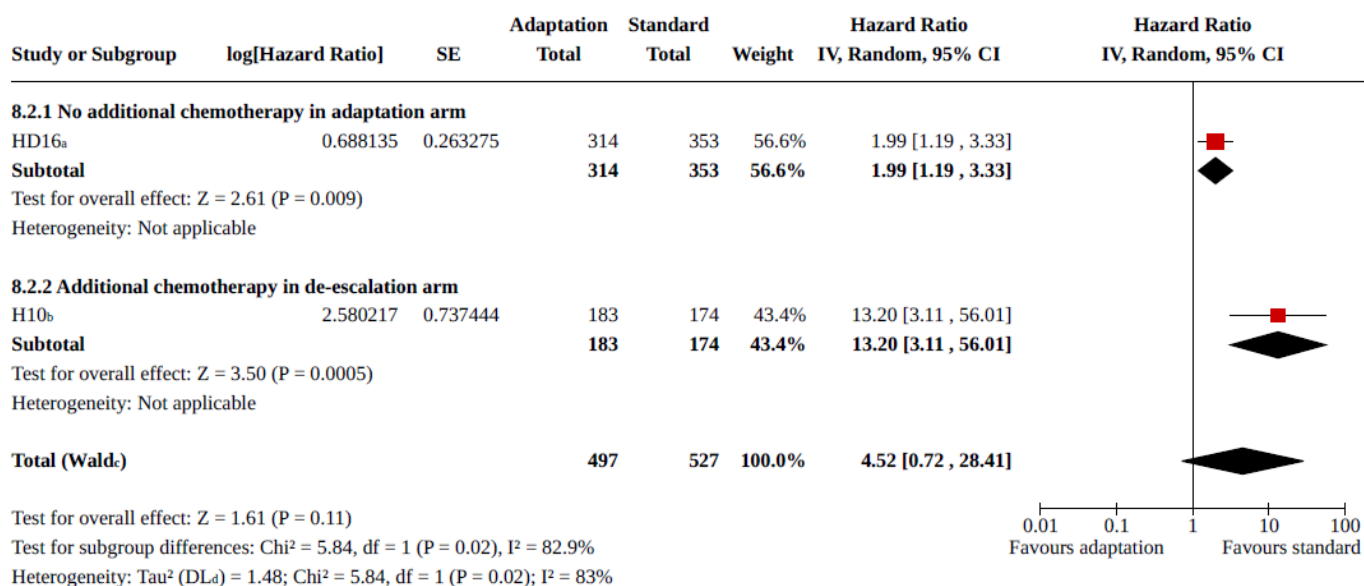
(E) Incomplete outcome data (attrition bias): Safety outcomes

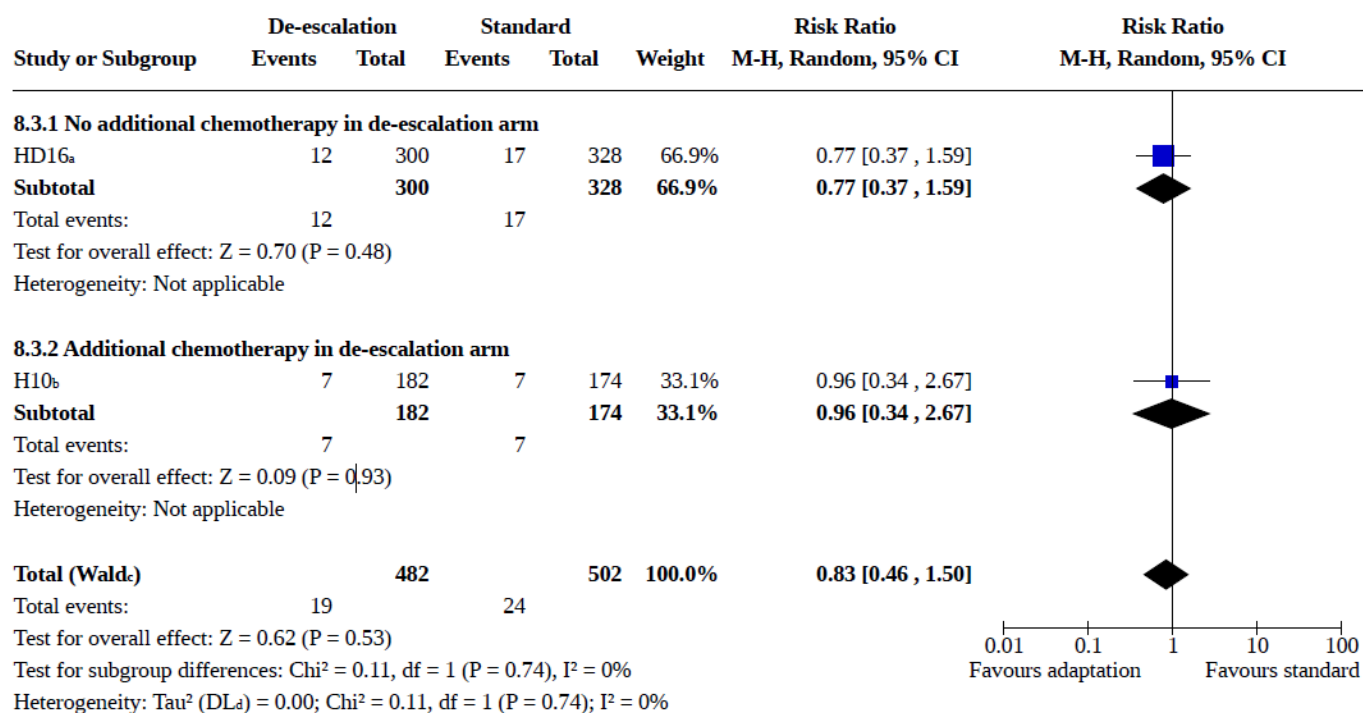
(F) Selective reporting (reporting bias)

(G) Other bias

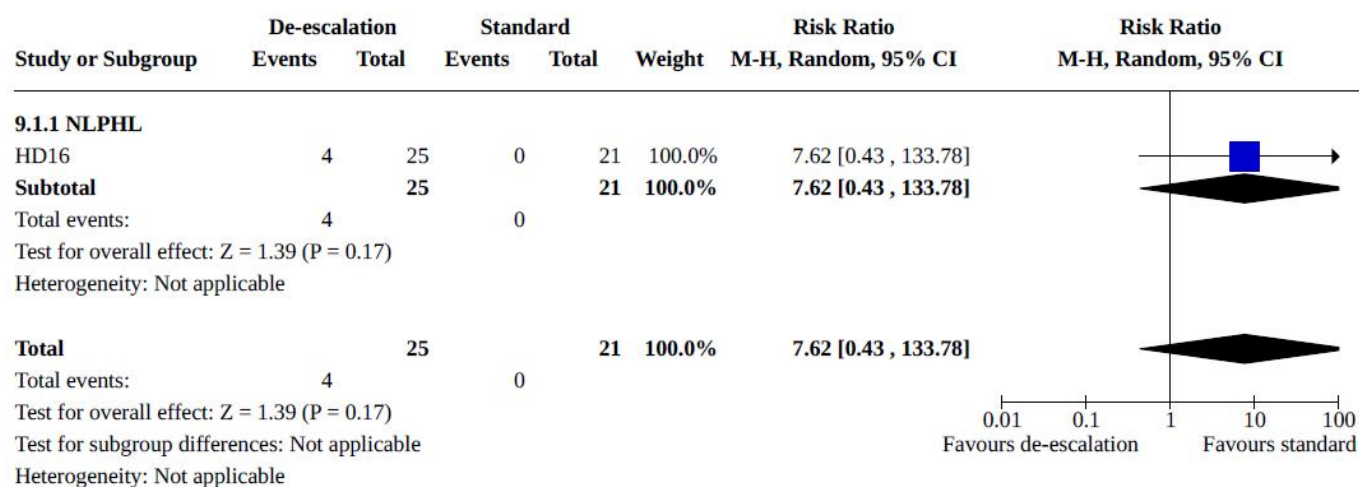
Comparison 8. Subgroup analyses: PET-adapted de-escalation in PET-negative early-stage HL (additional chemotherapy)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
8.1 Overall survival	2	1024	Hazard Ratio (IV, Random, 95% CI)	0.82 [0.12, 5.55]
8.1.1 No additional chemotherapy in adaptation arm	1	667	Hazard Ratio (IV, Random, 95% CI)	0.38 [0.10, 1.42]
8.1.2 Additional chemotherapy in adaptation arm	1	357	Hazard Ratio (IV, Random, 95% CI)	2.80 [0.29, 27.24]
8.2 Progression-free survival	2	1024	Hazard Ratio (IV, Random, 95% CI)	4.52 [0.72, 28.41]
8.2.1 No additional chemotherapy in adaptation arm	1	667	Hazard Ratio (IV, Random, 95% CI)	1.99 [1.19, 3.33]
8.2.2 Additional chemotherapy in de-escalation arm	1	357	Hazard Ratio (IV, Random, 95% CI)	13.20 [3.11, 56.01]
8.3 Incidence of second primary malignancy	2	984	Risk Ratio (M-H, Random, 95% CI)	0.83 [0.46, 1.50]
8.3.1 No additional chemotherapy in de-escalation arm	1	628	Risk Ratio (M-H, Random, 95% CI)	0.77 [0.37, 1.59]
8.3.2 Additional chemotherapy in de-escalation arm	1	356	Risk Ratio (M-H, Random, 95% CI)	0.96 [0.34, 2.67]

Analysis 8.1. Comparison 8: Subgroup analyses: PET-adapted de-escalation in PET-negative early-stage HL (additional chemotherapy), Outcome 1: Overall survival**Footnotes**^aDe-escalation: 2 x ABVD, Standard: 2 x ABVD + 20 Gy IF-RT; median follow-up 66 months^bDe-escalation: 4 x ABVD, Standard: 3 x ABVD + 30 Gy IN-RT; 10-year FU^cCI calculated by Wald-type method.^d Tau^2 calculated by DerSimonian and Laird method.**Analysis 8.2. Comparison 8: Subgroup analyses: PET-adapted de-escalation in PET-negative early-stage HL (additional chemotherapy), Outcome 2: Progression-free survival****Footnotes**^aDe-escalation: 2 x ABVD, Standard: 2 x ABVD + 20 Gy IF-RT; median follow-up 64 months^bDe-escalation: 4 x ABVD, Standard: 3 x ABVD + 30 Gy IN-RT; ITT before amendment^cCI calculated by Wald-type method.^d Tau^2 calculated by DerSimonian and Laird method.

Analysis 8.3. Comparison 8: Subgroup analyses: PET-adapted de-escalation in PET-negative early-stage HL (additional chemotherapy), Outcome 3: Incidence of second primary malignancy**Footnotes**^aDe-escalation: 2 x ABVD, Standard: 2 x ABVD + 20 Gy IF-RT; per protocol population, median follow-up 64 months^bDe-escalation: 4 x ABVD, Standard: 3 x ABVD + 30 Gy IN-RT; 10-year FU^cCI calculated by Wald-type method.^d Tau^2 calculated by DerSimonian and Laird method.**Comparison 9. Subgroup analysis: PET-adapted de-escalation in PET-negative early-stage HL (NLPHL vs. cHL)**

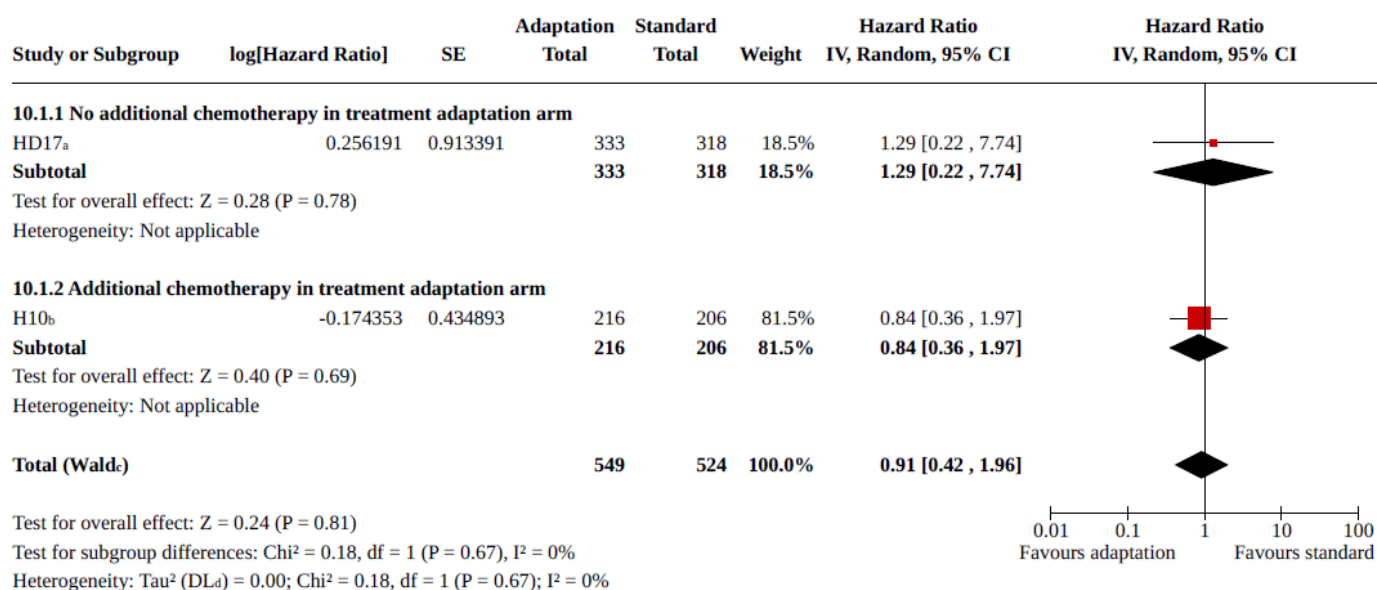
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
9.1 Progression-free survival	1	46	Risk Ratio (M-H, Random, 95% CI)	7.62 [0.43, 133.78]
9.1.1 NLPHL	1	46	Risk Ratio (M-H, Random, 95% CI)	7.62 [0.43, 133.78]

Analysis 9.1. Comparison 9: Subgroup analysis: PET-adapted de-escalation in PET-negative early-stage HL (NLPHL vs. cHL), Outcome 1: Progression-free survival**Comparison 10. Subgroup analyses: PET-adapted de-escalation in PET-negative intermediate stage HL (additional chemotherapy)**

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
10.1 Overall survival	2	1073	Hazard Ratio (IV, Random, 95% CI)	0.91 [0.42, 1.96]
10.1.1 No additional chemotherapy in treatment adaptation arm	1	651	Hazard Ratio (IV, Random, 95% CI)	1.29 [0.22, 7.74]
10.1.2 Additional chemotherapy in treatment adaptation arm	1	422	Hazard Ratio (IV, Random, 95% CI)	0.84 [0.36, 1.97]
10.2 Progression-free survival	2	1073	Hazard Ratio (IV, Random, 95% CI)	1.59 [0.95, 2.67]
10.2.1 No additional chemotherapy in treatment adaptation arm	1	651	Hazard Ratio (IV, Random, 95% CI)	1.83 [0.63, 5.35]
10.2.2 Additional chemotherapy in treatment adaptation arm	1	422	Hazard Ratio (IV, Random, 95% CI)	1.52 [0.84, 2.75]
10.3 Treatment-related mortality	2	1690	Risk Ratio (M-H, Random, 95% CI)	2.94 [0.31, 28.20]
10.3.1 No additional chemotherapy in treatment adaptation arm	1	1096	Risk Ratio (M-H, Random, 95% CI)	2.98 [0.12, 72.95]
10.3.2 Additional chemotherapy in treatment adaptation arm	1	594	Risk Ratio (M-H, Random, 95% CI)	2.90 [0.12, 70.93]
10.4 Incidence of second primary malignancy	2	1515	Risk Ratio (M-H, Random, 95% CI)	1.01 [0.57, 1.82]
10.4.1 No additional chemotherapy in treatment adaptation arm	1	1096	Risk Ratio (M-H, Random, 95% CI)	1.13 [0.41, 3.11]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
10.4.2 Additional chemotherapy in treatment adaptation arm	1	419	Risk Ratio (M-H, Random, 95% CI)	0.96 [0.47, 1.96]

Analysis 10.1. Comparison 10: Subgroup analyses: PET-adapted de-escalation in PET-negative intermediate stage HL (additional chemotherapy), Outcome 1: Overall survival



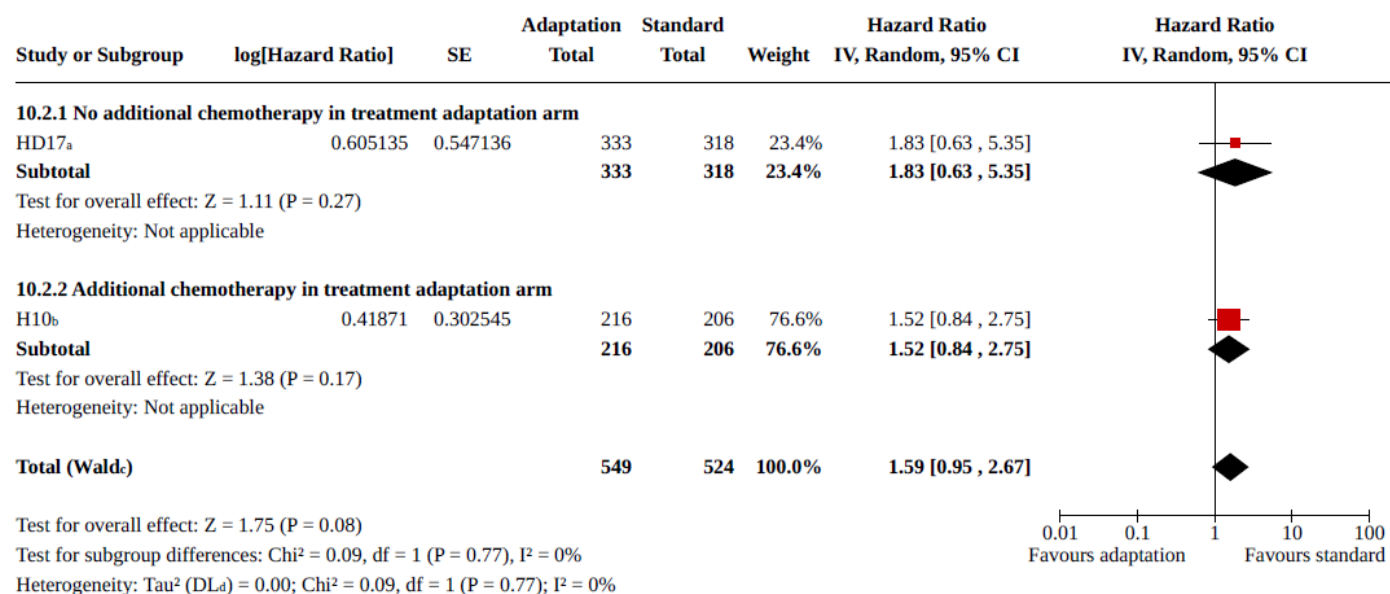
Footnotes

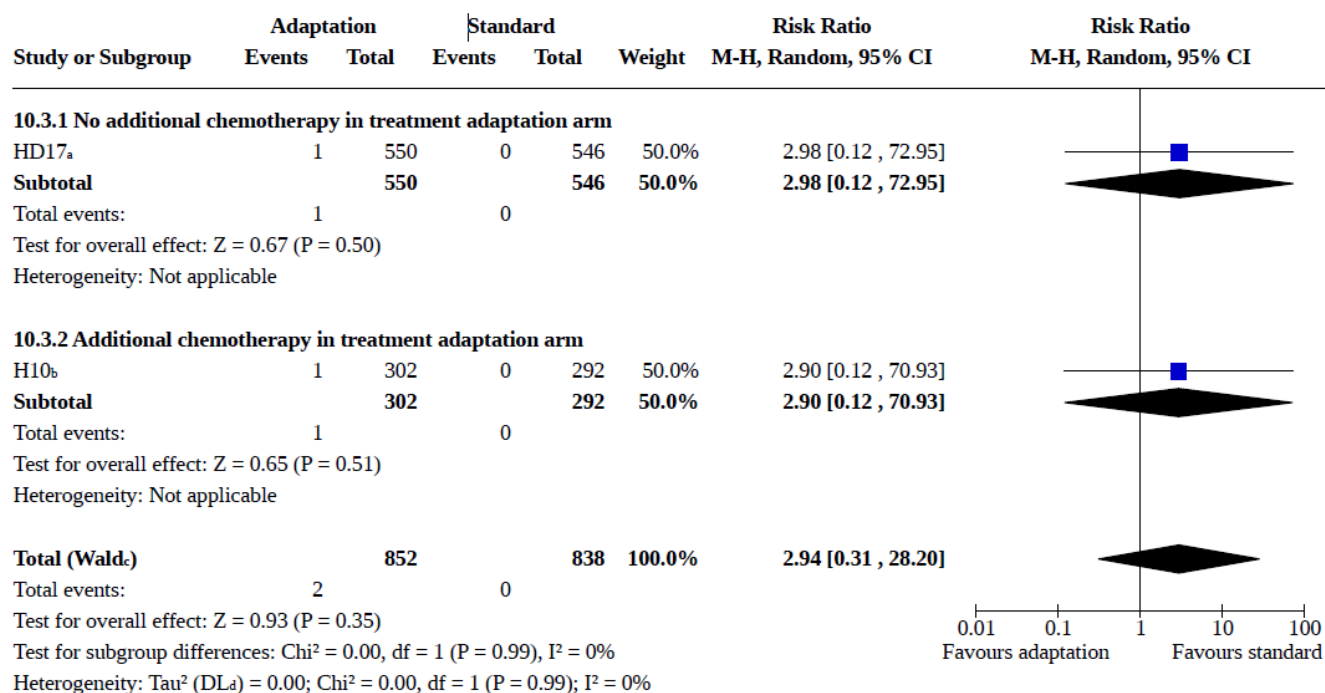
^aDe-escalation: 2 x BEACOPP^{escalated}, 2 x of ABVD, Standard: 2 x BEACOPP^{escalated}, 2 x of ABVD, + 30 Gy IF-RT

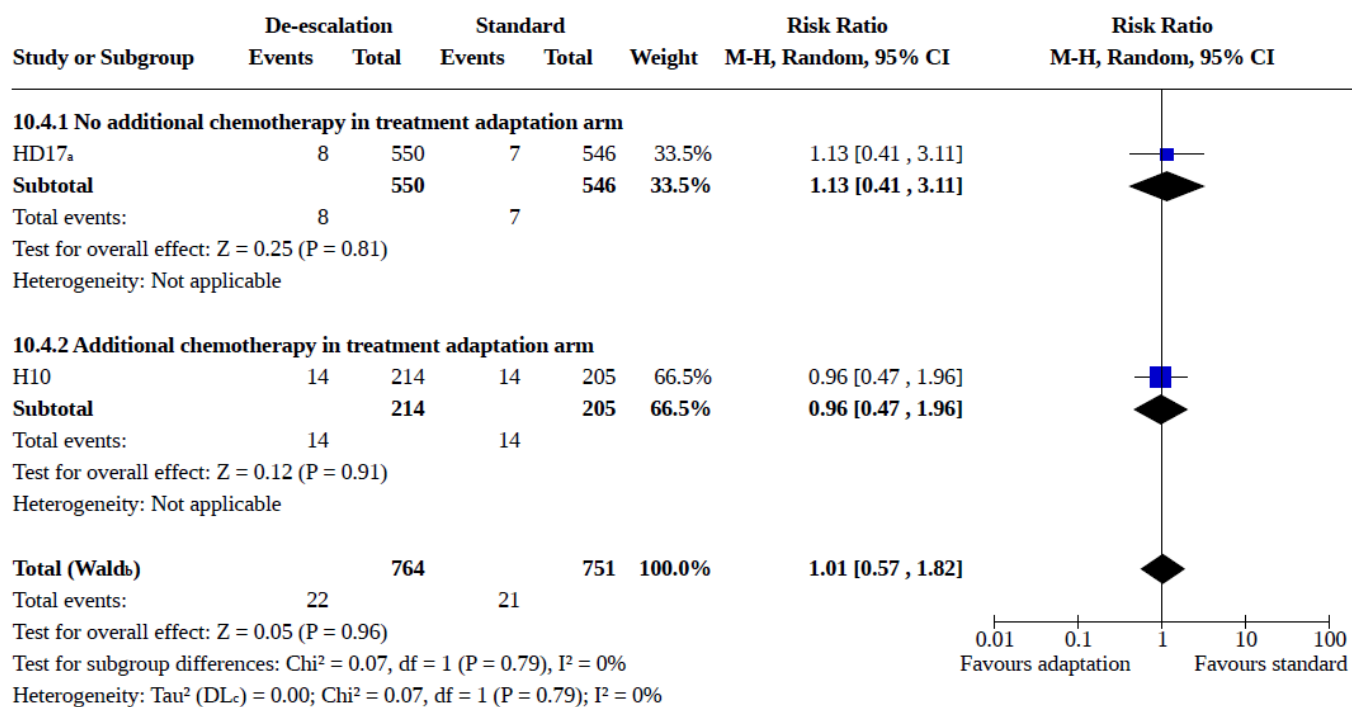
^bUnfavourable; de-escalation: 6 x ABVD, Standard: 4 x ABVD + 30 Gy IN-RT; 10-year FU.

^cCI calculated by Wald-type method.

^dTau² calculated by DerSimonian and Laird method.

Analysis 10.2. Comparison 10: Subgroup analyses: PET-adapted de-escalation in PET-negative intermediate stage HL (additional chemotherapy), Outcome 2: Progression-free survival**Footnotes**^aDe-escalation: 2 x BEACOPPescalated, 2 x ABVD, Standard: 2 x BEACOPPescalated, 2 x ABVD, + 30 Gy IF-RT; ITT; PET-negative only, unpublished data^bUnfavourable; de-escalation: 6 x ABVD, Standard: 4 x ABVD + 30 Gy IN-RT; ITT before amendment^cCI calculated by Wald-type method.^d Tau^2 calculated by DerSimonian and Laird method.

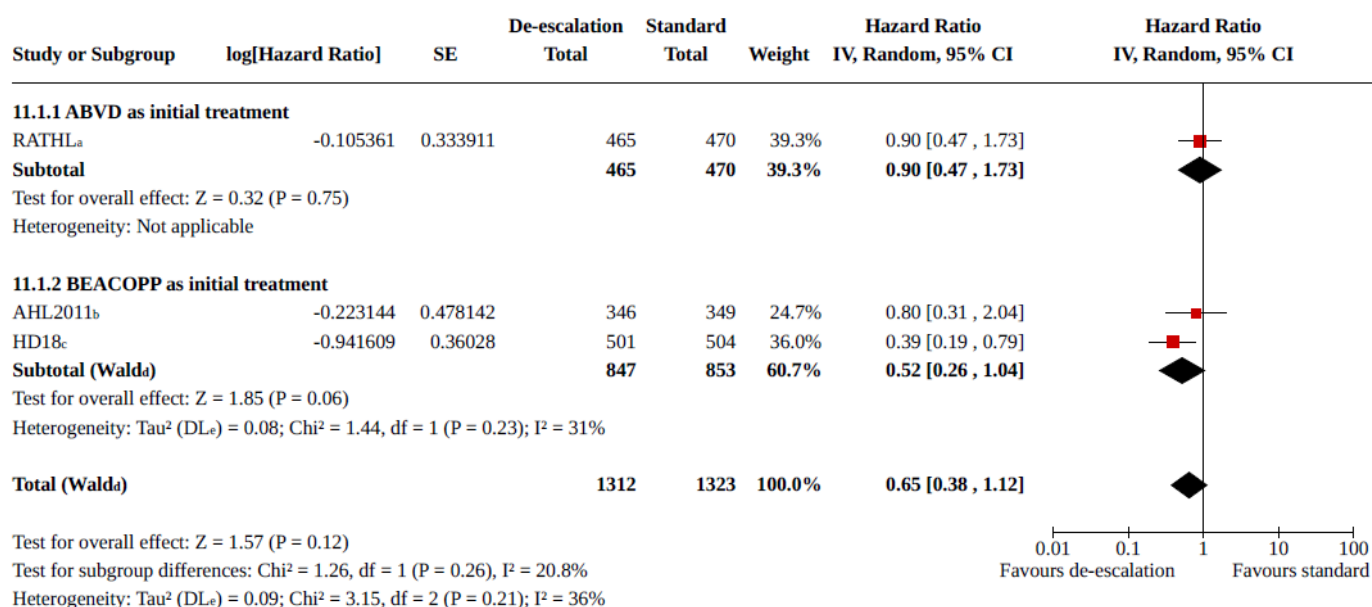
Analysis 10.3. Comparison 10: Subgroup analyses: PET-adapted de-escalation in PET-negative intermediate stage HL (additional chemotherapy), Outcome 3: Treatment-related mortality**Footnotes**^aDe-escalation: 2 cycles BEACOPP_{escalated}, 2 cycles of ABVD, Standard: 2 cycles BEACOPP_{escalated}, 2 cycles of ABVD, + 30 Gy IF-RT; includes PET-positive^bUnfavourable; de-escalation: 6 cycles ABVD, Standard: 4 cycles ABVD + 30 Gy IN-RT; 5-year FU.^cCI calculated by Wald-type method.^dTau² calculated by DerSimonian and Laird method.

Analysis 10.4. Comparison 10: Subgroup analyses: PET-adapted de-escalation in PET-negative intermediate stage HL (additional chemotherapy), Outcome 4: Incidence of second primary malignancy**Footnotes**^aDe-escalation: 2 x BEACOPP^{escalated}, 2 x ABVD, Standard: 2 x BEACOPP^{escalated}, 2 x ABVD + 30 Gy IF-RT; includes PET-positive participants^bCI calculated by Wald-type method.^c Tau^2 calculated by DerSimonian and Laird method.**Comparison 11. Subgroup analyses: PET-adapted de-escalation in PET-negative advanced stage HL (initial chemotherapy regimen; ITT)**

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
11.1 Overall survival	3	2635	Hazard Ratio (IV, Random, 95% CI)	0.65 [0.38, 1.12]
11.1.1 ABVD as initial treatment	1	935	Hazard Ratio (IV, Random, 95% CI)	0.90 [0.47, 1.73]
11.1.2 BEACOPP as initial treatment	2	1700	Hazard Ratio (IV, Random, 95% CI)	0.52 [0.26, 1.04]
11.2 Progression-free survival	3	2635	Hazard Ratio (IV, Random, 95% CI)	0.98 [0.76, 1.27]
11.2.1 ABVD as initial treatment	1	935	Hazard Ratio (IV, Random, 95% CI)	1.13 [0.81, 1.57]
11.2.2 BEACOPP as initial treatment	2	1700	Hazard Ratio (IV, Random, 95% CI)	0.79 [0.52, 1.19]
11.3 Treatment-related mortality	3	2763	Risk Ratio (M-H, Random, 95% CI)	0.21 [0.06, 0.72]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
11.3.1 ABVD as initial treatment	1	935	Risk Ratio (M-H, Random, 95% CI)	0.11 [0.01, 2.04]
11.3.2 BEACOPP as initial treatment	2	1828	Risk Ratio (M-H, Random, 95% CI)	0.24 [0.06, 0.96]
11.4 Adverse events, grade 3-4	2	1930	Risk Ratio (M-H, Random, 95% CI)	0.94 [0.92, 0.97]
11.4.1 ABVD as initial treatment	1	925	Risk Ratio (M-H, Random, 95% CI)	0.95 [0.87, 1.04]
11.4.2 BEACOPP as initial treatment	1	1005	Risk Ratio (M-H, Random, 95% CI)	0.94 [0.91, 0.98]
11.5 Incidence of second primary malignancy	3	2759	Risk Ratio (M-H, Random, 95% CI)	0.82 [0.54, 1.24]
11.5.1 ABVD as initial treatment	1	935	Risk Ratio (M-H, Random, 95% CI)	0.86 [0.39, 1.89]
11.5.2 BEACOPP as initial treatment	2	1824	Risk Ratio (M-H, Random, 95% CI)	0.80 [0.49, 1.32]

Analysis 11.1. Comparison 11: Subgroup analyses: PET-adapted de-escalation in PET-negative advanced stage HL (initial chemotherapy regimen; ITT), Outcome 1: Overall survival



Footnotes

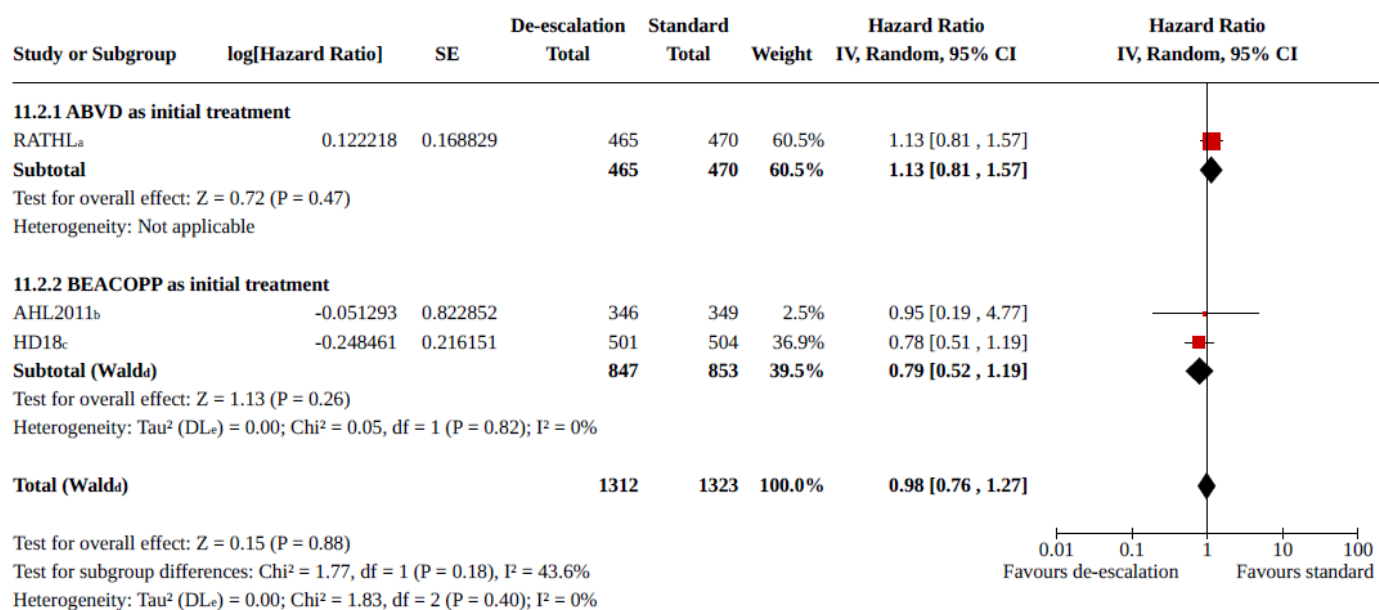
^aDe-escalation: 2 cycles ABVD, 4 cycles AVD; Standard: 6 cycles ABVD

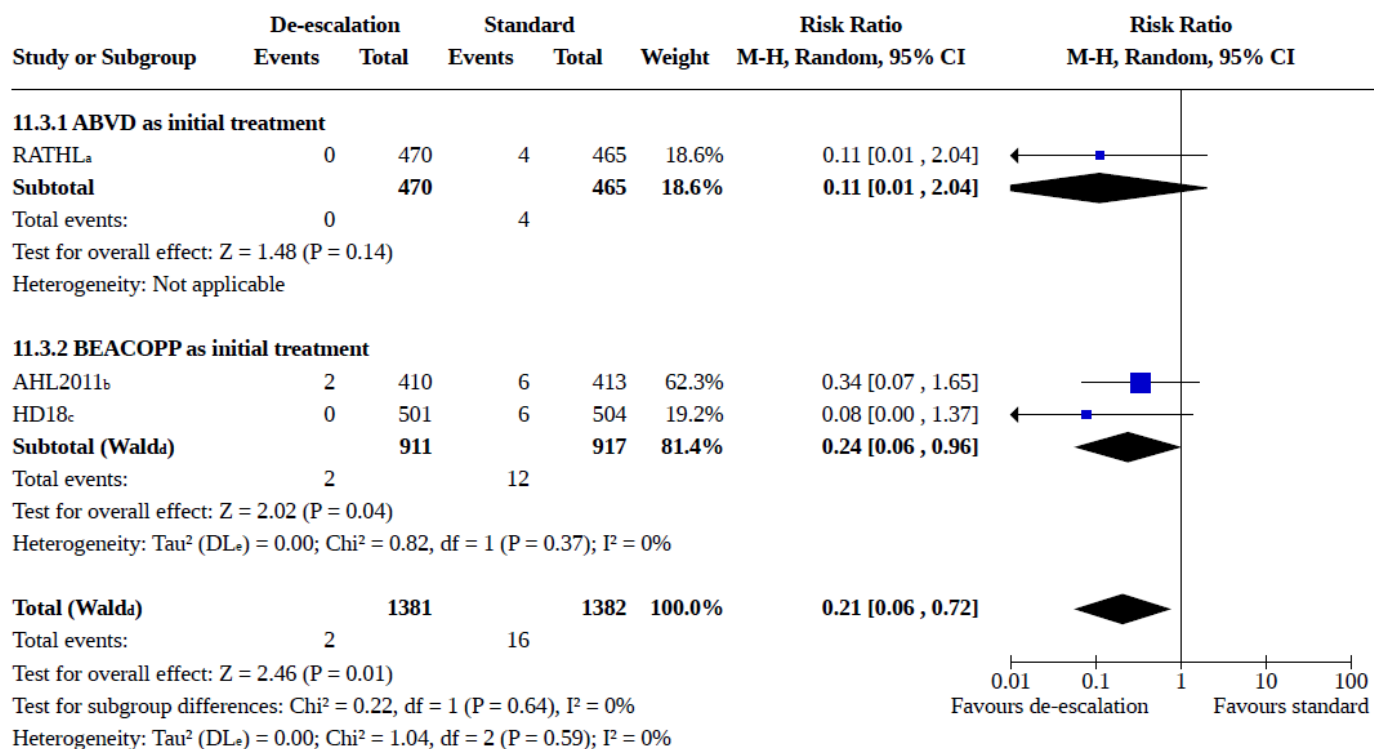
^bDe-escalation: 2 cycles BEACOPP^{escalated}, 4 cycles ABVD; Standard: 6 cycles BEACOPP^{escalated}; HR calculated based on Tierney 2007, method 7

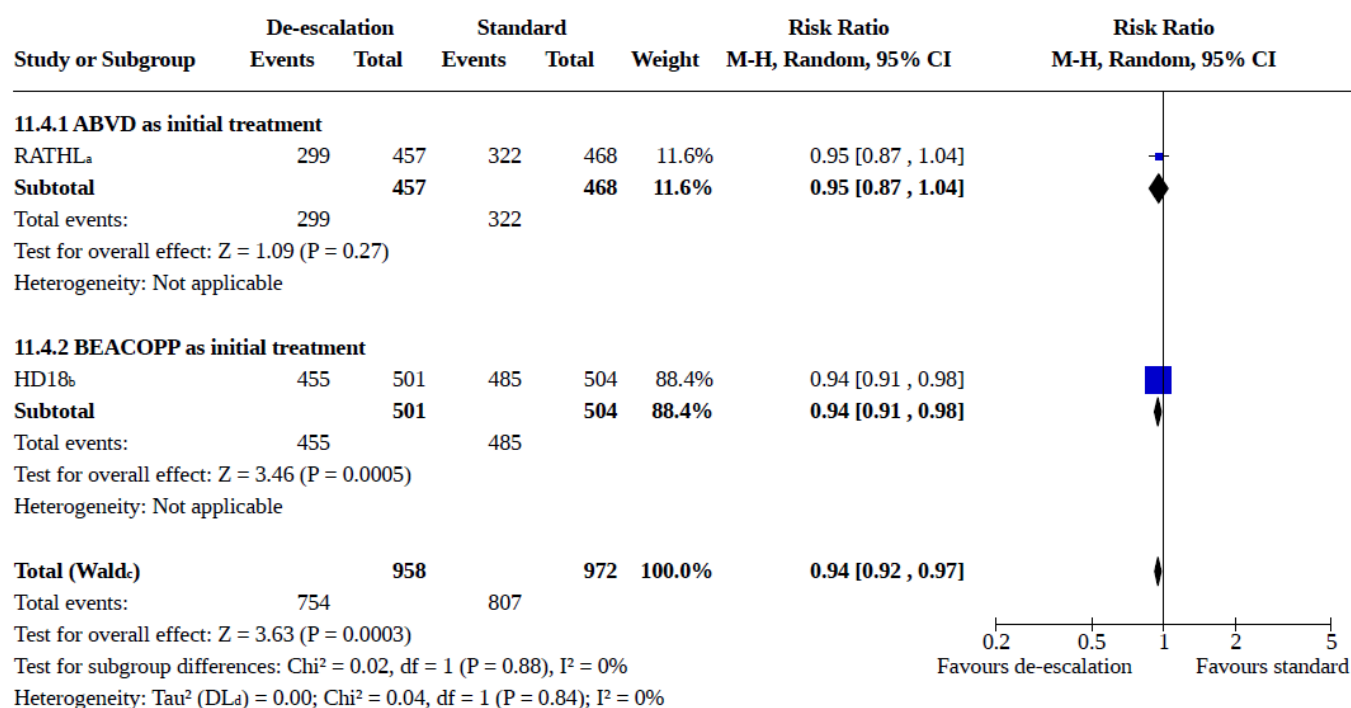
^cDe-escalation: 4 cycles BEACOPP^{escalated}; Standard: 8 or 6 cycles of BEACOPP^{escalated}

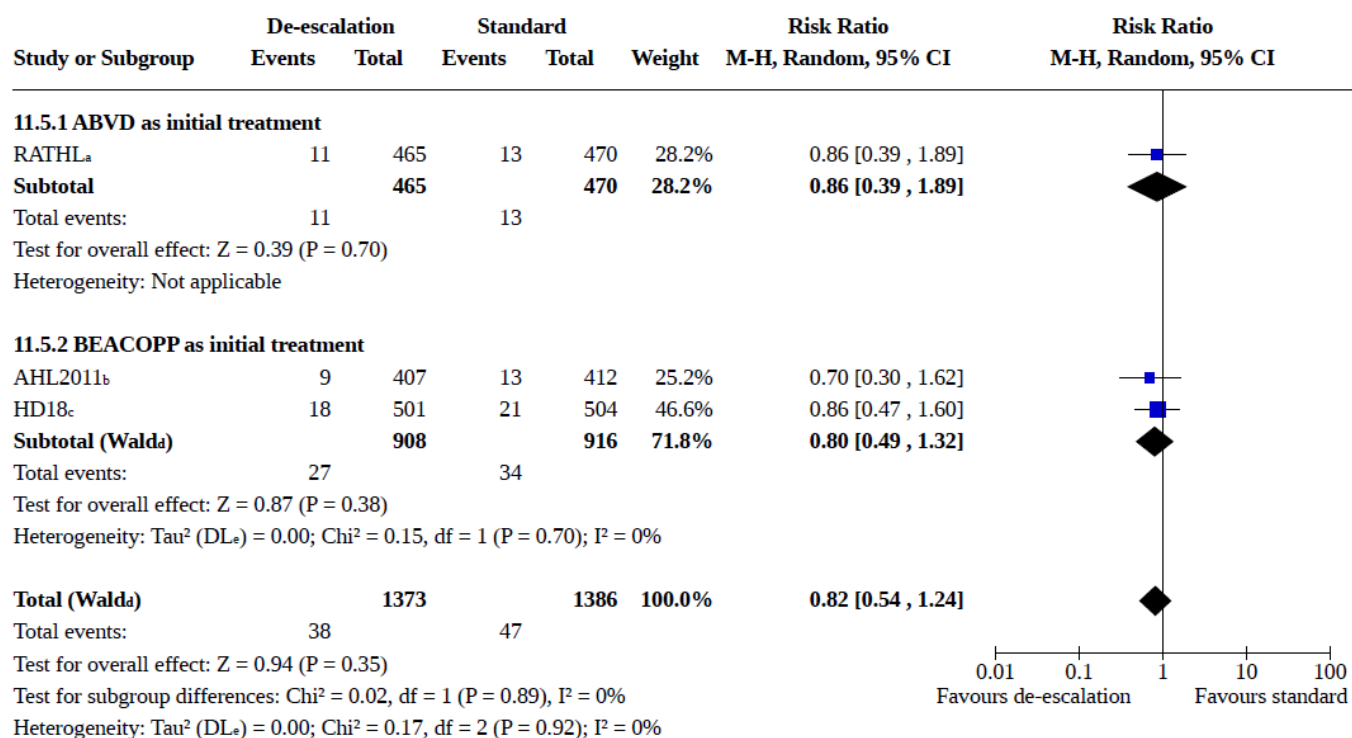
^dCI calculated by Wald-type method.

^eTau² calculated by DerSimonian and Laird method.

Analysis 11.2. Comparison 11: Subgroup analyses: PET-adapted de-escalation in PET-negative advanced stage HL (initial chemotherapy regimen; ITT), Outcome 2: Progression-free survival**Footnotes**^aDe-escalation: 2 cycles ABVD, 4 cycles AVD; Standard: 6 cycles ABVD^bDe-escalation: 2 cycles BEACOPP^{escalated}, 4 cycles ABVD; Standard: 6 cycles BEACOPP^{escalated}; HR calculated based on Tierney 2007, method 7^cDe-escalation: 4 cycles BEACOPP^{escalated}; Standard: 8 or 6 cycles of BEACOPP^{escalated}^dCI calculated by Wald-type method.^e Tau^2 calculated by DerSimonian and Laird method.

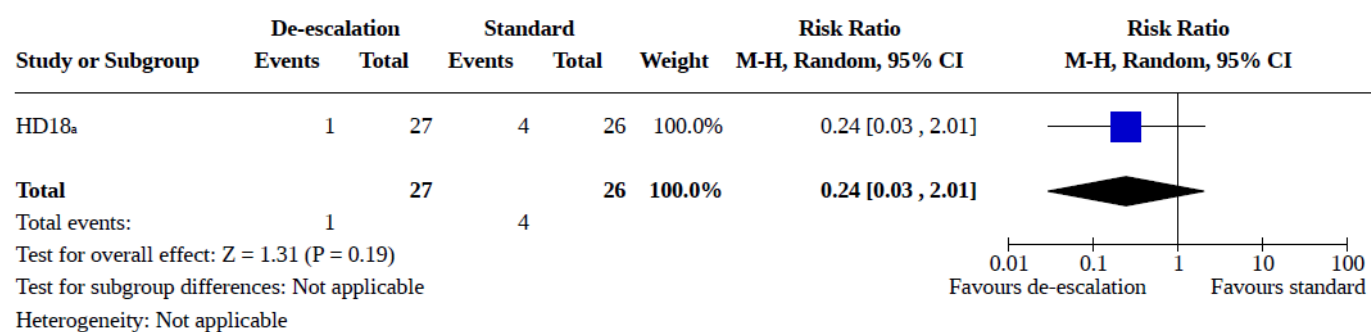
Analysis 11.3. Comparison 11: Subgroup analyses: PET-adapted de-escalation in PET-negative advanced stage HL (initial chemotherapy regimen; ITT), Outcome 3: Treatment-related mortality**Footnotes**^aDe-escalation: 2 cycles ABVD, 4 cycles AVD; Standard: 6 cycles ABVD; death owing to initial treatment^bDe-escalation: 2 cycles BEACOPP^{escalated}, 4 cycles ABVD; Standard: 6 cycles BEACOPP^{escalated}; the population also includes PET-positive participants^cDe-escalation: 4 cycles BEACOPP^{escalated}; Standard: 8 or 6 cycles of BEACOPP^{escalated}^dCI calculated by Wald-type method.^eTau² calculated by DerSimonian and Laird method.

Analysis 11.4. Comparison 11: Subgroup analyses: PET-adapted de-escalation in PET-negative advanced stage HL (initial chemotherapy regimen; ITT), Outcome 4: Adverse events, grade 3-4**Footnotes**^aDe-escalation: 2 cycles ABVD, 4 cycles AVD; Standard: 6 cycles ABVD^bDe-escalation: 4 cycles BEACOPP^{escalated}; Standard: 8 or 6 cycles of BEACOPP^{escalated}^cCI calculated by Wald-type method.^dTau² calculated by DerSimonian and Laird method.

Analysis 11.5. Comparison 11: Subgroup analyses: PET-adapted de-escalation in PET-negative advanced stage HL (initial chemotherapy regimen; ITT), Outcome 5: Incidence of second primary malignancy**Footnotes**^aDe-escalation: 2 cycles ABVD, 4 cycles AVD; Standard: 6 cycles ABVD^bDe-escalation: 2 cycles BEACOPP^{escalated}, 4 cycles ABVD; Standard: 6 cycles BEACOPP^{escalated}; the population also includes PET-positive participants^cDe-escalation: 4 cycles BEACOPP^{escalated}; Standard: 8 or 6 cycles of BEACOPP^{escalated}^aCI calculated by Wald-type method.^eTau² calculated by DerSimonian and Laird method.**Comparison 12. Subgroup analyses: PET-adapted de-escalation in PET-negative advanced stage HL (NLPHL vs. cHL)**

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
12.1 progression-free survival	1	53	Risk Ratio (M-H, Random, 95% CI)	0.24 [0.03, 2.01]

Analysis 12.1. Comparison 12: Subgroup analyses: PET-adapted de-escalation in PET-negative advanced stage HL (NLPHL vs. cHL), Outcome 1: progression-free survival



Footnotes

^aDe-escalation: 4 cycles BEACOPP_{escalated}; Standard: 8 or 6 cycles of BEACOPP_{escalated}



ADDITIONAL TABLES

Table 1. Assumptions and reporting of time-to-event outcomes of included trials

Study ID	Stage	Outcome	HR re-ported	Number of events reported	Stratification & adjust-ment	Survival curve re-ported	Assump-tion of propor-tional hazards	Censoring	Censoring marked?	Method-ological comment
FDG-PET-negative with treatment de-escalation										
HD16	early	OS	yes	yes	stratified randomisation (centre, age (> 45 vs ≥ 45 years), sex, B symp-toms, disease localisa-tion (supradiaphragmat-ic vs infradiaphragmat-ic), albumin level (< 4 g/dL vs ≥ 4 g/dL), and pres-ence versus absence of initial bulk (< 5 cm vs ≥ 5 cm in largest diameter)	yes	not report-ed	OS is censored at the date of last document-ed information on the patient.	yes, marked in KM-curve and num-ber cen-sored re-ported	-
		PFS	yes	no	stratified randomisation (centre, age (> 45 vs ≥ 45 years), sex, B symp-toms, disease localisa-tion (supradiaphragmat-ic vs infradiaphragmat-ic), albumin level (< 4 g/dL vs ≥ 4 g/dL), and pres-ence versus absence of initial bulk (< 5 cm vs ≥ 5 cm in largest diameter)	yes	not report-ed	PFS was cen-sored at the last documented date of continu-ous remission.	yes, marked in KM-curve and num-ber cen-sored re-ported	HR and 95% CI reported; number of events re-ported sep-arately for progression and death
UK NCRI RAPID	early-in-termedi-ate	OS	yes	yes	no stratification factors	yes	SAP only; "Time-to-event da-ta: e.g. use Schoen-feld resid-uals or a p-value to test the assump-tion of	not reported	not report-ed	-

Table 1. Assumptions and reporting of time-to-event outcomes of included trials (Continued)

		PFS	yes	yes	no stratification factors	yes	proportional hazards"	not reported	not reported	-
							SAP only; "Time-to-event data: e.g. use Schoenfeld residuals or a p-value to test the assumption of proportional hazards"			
H10	early-intermediate	OS	yes	yes	stratified randomisation (institution, Ann Arbor stage (I vs II), and availability of baseline PET)	yes	not reported	Alive patients will be censored at the date of last follow-up.	not reported	NA. The H10F effect was not estimable for OS due to no events in one group.
		PFS	yes	yes	stratified randomisation (institution, Ann Arbor stage (I vs II), and availability of baseline PET), log-rank test stratified by stage (I vs II) and availability of baseline PET	yes	not reported	Patients alive and free of progression will be censored at the date of last follow-up.	not reported	The H10U OS effect reported as PP was used for ITT analysis, as no participant was excluded due to protocol violation.
Picardi 2007	early-intermediate	OS	no	yes	NA	no	-	not reported	-	no events
	ate	PFS	no	yes	not reported	yes	not reported	not reported	yes, marked in cumulative in-	recalculated (EFS) based on log-rank P value and

Table 1. Assumptions and reporting of time-to-event outcomes of included trials (Continued)

									cidence curve	number of events/part- icipants; Tierney method 8
HD17	intermedi- ate	OS	yes	yes	stratified randomisa- tion (centre, age (< 45 years vs ≥ 5 years), sex, B-symptoms, disease localisation (supradi- aphragmatic vs infradi- aphragmatic), albumin concentration (< 40 g/ L vs ≥ 40 g/L), and ini- tial tumour bulk (pres- ence vs absence of ini- tial tumour bulk ≥ 5 cm in largest diameter)	yes	yes: "The propor- tional haz- ards as- sumption was tested by visual inspection of survival curves".	"or (if the pa- tient is still alive) until the last date for which informa- tion exists"	yes, num- ber cen- sored re- ported	unpublished data on PET- negative participants provided by authors; HR and 95% CI inverted for meta-analy- sis
		PFS	yes	yes	stratified randomisa- tion (centre, age (< 45 years vs ≥ 45 years), sex, B-symptoms, disease localisation (supradi- aphragmatic vs infradi- aphragmatic), albumin concentration (< 40 g/ L vs ≥ 40 g/L), and ini- tial tumour bulk (pres- ence vs absence of ini- tial tumour bulk ≥ 5 cm in largest diameter)	yes	yes: "The propor- tional haz- ards as- sumption was tested by visual inspection of survival curves"	"[T]he PFS cal- culation the patient is cen- sored accord- ing to the last known date of progres- sion-free sur- vival".	yes, num- ber cen- sored re- ported	unpublished data on PET- negative participants provided by authors; HR and 95% CI inverted for meta-analy- sis
AHL2011	advanced	OS	yes	yes	stratified randomisation (Ann Arbor stage (IIB vs IIIV) and international prognosis score (IPS: 0-2 vs. 3)), stratified analysis for the same factors	yes	not report- ed	"Patients who have not died at the time of the analysis or who are lost to fol- low-up will be censored at the date of last con- tact."	yes, marked in KM-curves	recalcula- tion based on P value, number of events (Tier- ney method no. 7)

Table 1. Assumptions and reporting of time-to-event outcomes of included trials (Continued)

		PFS	yes	yes	stratified randomisation (Ann Arbor stage (IIB vs IIIV) and international prognosis score (IPS: 0-2 vs. 3)), stratified analysis for the same factors	yes	not report- ed	"Patients alive and free of pro- gression or who are lost to fol- low-up will be censored at their last fol- low-up date."	yes, marked in KM-curves	recalcula- tion based on P value, number of events (Tier- ney method no. 7)
HD0607	advanced	OS	no	no	unadjusted analysis	yes	yes: "Pro- portional hazard as- sumption was veri- fied for all estimated models."	not reported	no, num- ber at risk only	HR, 95% CI and number of events provided by the authors
		PFS	no	no	unadjusted analysis	yes	yes: "Pro- portional hazard as- sumption was veri- fied for all estimated models."	Censoring rea- sons: failing pa- tients should be censored when starting a second-line of therapy.	no, num- ber at risk only	HR, 95% CI and number of events provided by the authors
HD0801	advanced	OS	no	yes	-	-	-	-	-	NA. The ef- fect was not es- timable for OS due to no event in one group.
		PFS	yes	yes, sepa- rate	stratified randomisation (HL stage), adjusted Cox regression analysis	yes	not report- ed	yes, "In the phase III study for PET-2 neg- ative patients, those with mi- nor or no re- sponse with- out progression at the end of planned thera-	no, num- ber at risk only	HR, 95% CI and number of events provided (separate for for death and pro- gression)

Table 1. Assumptions and reporting of time-to-event outcomes of included trials (Continued)

HD18	advanced	OS	yes	yes	stratified randomisation, unadjusted analysis	yes	not reported	py are considered failures and switched to second-line therapy. . . because in the calculation of PFS time, all failures are censored when starting second line of therapy"	yes, marked in KM-curve and number censored reported	-
		PFS	yes	no	stratified randomisation, unadjusted analysis	yes	not reported	"If none of these events [progression, relapse, death] had occurred, progression-free survival was censored at the date of last information on the disease status."	yes, marked in KM-curve and number censored reported	number of events reported separately for progression and death events
RATHL	advanced	OS	yes	yes	not reported	yes	not reported	"[S]urviving patients will be censored at date last known to be alive."	no, number at risk only	-
		PFS	yes	yes	The absolute risk difference was adjusted for stratification factors (PET score and centre).	yes	not reported	"Patients alive without progression or relapse will be	no, number at risk only	-

Table 1. Assumptions and reporting of time-to-event outcomes of included trials (Continued)

								censored at date last known to be alive."		
FDG-PET-positive with treatment escalation										
H10	early-in- intermedi- ate	OS	yes	yes	stratified randomisation (institution, Ann Arbor stage (I v II), and avail- ability of baseline PET)	yes	not report- ed	"Alive patients will be cen- sored at the date of last fol- low-up."	not report- - ed	-
		PFS	yes	yes	stratified randomisa- tion (institution, Ann Ar- bor stage (I vs II), and availability of baseline PET), log-rank test strati- fied by stage (I vs II) and availability of baseline PET	yes	not report- ed	"Patients alive and free of pro- gression will be censored at the date of last fol- low-up."	not report- ed	-
HD0607	advanced	OS	no	yes	unadjusted analysis	yes	yes: "Pro- portional hazard as- sumption was veri- fied for all estimated models."	not reported	number at risk and number of events per time point	HR and 95% CI provided by the au- thors
		PFS	no	yes	unadjusted analysis	yes	yes: "Pro- portional hazard as- sumption was veri- fied for all estimated models."	"Censoring rea- sons: failing pa- tients should be censored when starting a sec- ond line of ther- apy".	Number at risk and number of events per time point	HR and 95% CI provided by the au- thors
HD18	advanced	OS	yes	yes	stratified randomisa- tion, unadjusted analy- sis	yes	not report- ed	"or censored at the date of last information on the patient be- ing alive"	yes, marked in KM-curve and num- ber cen-	-

Table 1. Assumptions and reporting of time-to-event outcomes of included trials (Continued)

PFS	yes	no	stratified randomisation, unadjusted analysis	yes	not reported	"If none of these events [progression, relapse, death] had occurred, progression-free survival was censored at the date of last information on the disease status."	sored re-reported	number of events reported separately for progression and death events
							yes, marked in KM-curve and number censored re-reported	

CI: confidence interval
 EFS: event-free survival
 FDG: fluorodeoxyglucose
 H1OF: favourable
 H1OU: unfavourable
 HL: Hodgkin lymphoma
 HR: hazard ratio
 IPS: international prognostic score
 ITT: intention-to-treat
 KM: Kaplan-Meier
 NA: not applicable
 OS: overall survival
 PET: positron emission tomography
 PFS: progression-free survival
 PP: per protocol
 SAP: statistical analysis plan
 vs: versus

Table 2. General study characteristics and interventions

Study	Timing of randomisation	Stage	Treatment before treatment modification	PET scoring criteria	Experimental treatment	Control treatment	Median follow-up



Table 2. General study characteristics and interventions (Continued)

FDG-PET-negative participants with treatment de-escalation

H10, favourable HL	before interim-PET	early	2 cycles of ABVD	international harmonisation criteria; 0 = negative; 1 = minimal residual uptake; 2 = malignant	2 cycles ABVD	1 cycle ABVD + IN-RT (30 Gy + 6 Gy boost to areas with residual disease)	60 months/10 years ^b
HD16	before interim-PET, analysis mostly PET-negative only, together available for two outcomes	early	2 cycles of ABVD	5-point score according to Lugano criteria; negative: 1, 2; positive: 3 to 5	no further therapy	IF-RT (20 Gy)	46/64 ^a months
UK NCRI RAPID	after interim-PET	early-intermediate	3 cycles of ABVD	5-point scale; negative: 1, 2; positive: 3 to 5	no further therapy	IF-RT (30 Gy)	60 months
Picardi 2007	after end-of-treatment PET 6	early-intermediate	6 cycles of VE-BEP	Negative PET examination was defined as no evidence of uptake. Positive PET examination was defined as increased uptake in a focus within an abnormal area.	no further therapy	RT (32 Gy, without boost on residual masses)	40 months
H10, unfavourable HL	before interim-PET	intermediate	2 cycles of ABVD	international harmonisation criteria; 0 = negative; 1 = minimal residual uptake; 2 = malignant	4 cycles ABVD	2 cycles ABVD + INRT (30 Gy + 6 Gy boost to areas with residual disease)	61.2 months
HD17	before interim-PET, analysis mostly PET-negative and PET-positive together	intermediate	2 cycles of BEA-COPP ^{escalated} + 2 cycles of ABVD	5-point score according to Lugano criteria; negative: 1, 2; positive: 3 to 5	no further treatment	IF-RT (30 Gy)	46.2 months
AHL2011	before interim-PET, analysis of PET-negative and PET-positive together	advanced	2 cycles of BEA-COPP	5-point score according to Lugano criteria; negative: 1 to 3;	4 cycles ABVD	4 cycles BEA-COPP	67 months

Table 2. General study characteristics and interventions *(Continued)*

				positive: 4, 5			
HD0607	after end-of-treatment PET 6		advanced 6 cycles of ABVD; cycle 2 and 6 interim-PET had to be negative for randomisation.	5-point score according to Lugano criteria; negative: 1 to 3; positive: 4, 5	no further therapy	IF-RT (30.6 Gy on initial bulky lesions or residual disease)	70.8 months
HD0801	after end-of-treatment PET 6	advanced	6 cycles of ABVD, cycle 2 and 6 interim-PET had to be negative for randomisation.	two-point visual scoring system: 0: negative/normal/minimal residual uptake/benign; 1: positive/malignant	no further therapy	IF-RT (30 Gy)	71 months
HD18	after interim-PET	advanced	2 cycles of BEA-COPP	5-point score according to Lugano criteria; negative: 1, 2; positive: 3 to 5	2 cycles BEA-COPP _{escalated}	4 cycles BEA-COPP _{escalated}	53 months
RATHL	after interim-PET	advanced	2 cycles of ABVD	5-point score according to Lugano criteria; Negative: 1-3; Positive: 4, 5	4 cycles AVD (ABVD without bleomycin)	4 cycles ABVD	41/87 ^c months
FDG-PET-positive participants with treatment escalation							
H10	before interim-PET	early-intermediate	2 cycles of ABVD	international harmonisation criteria; 0 = negative; 1 = minimal residual uptake; 2 = malignant	2 cycles BEA-COPP _{escalated} + INRT (30 Gy + 6 Gy Boost to areas with residual disease)	2 cycles ABVD + INRT (30 Gy + 6 Gy boost to areas with residual disease)	10 years ^b
HD0607	after interim-PET	advanced	2 cycles of ABVD	5-point score according to Lugano criteria; negative: 1 to 3; positive: 4-5	4 cycles of BEA-COPP _{escalated} and 4 cycles of BEACOPP plus rituximab	4 cycles of BEA-COPP _{escalated} and 4 cycles of BEACOPP	43.2 months



Table 2. General study characteristics and interventions (Continued)

HD18	after interim-PET	advanced	2 cycles of BEA-COPP	5-point score according to Lugano criteria: negative: 1, 2; positive: 3 to 5	6 cycles of BEA-COPP _{escalated} plus rituximab	6 cycles of BEA-COPP _{escalated}	66 months

^a Secondary reference, Fuchs 2023.

^b Secondary reference, Federico 2023.

^c Secondary reference, Luminari 2023.

Abbreviations:

ABVD: doxorubicin (Adriamycin), bleomycin, vinblastine, dacarbazine

BEACOPP: bleomycin, etoposide, doxorubicin (Adriamycin), cyclophosphamide, vincristine (Oncovin), procarbazine, prednisolone (a steroid)

FDG: fluorodeoxyglucose

HL: Hodgkin lymphoma

IF-RT: involved-field radiotherapy

IN-RT: involved-node radiotherapy

IQR: interquartile range

PET: positron emission tomography

VEBEP: etoposide, epirubicin, bleomycin, cyclophosphamide, and prednisolone

Table 3. Participant characteristics

Study ID	Stage	Sample size	Recruitment period	Location	Centres	Age (years)	Sex (% female)	NLPHL (%)
FDG-PET-negative with treatment de-escalation								
H10, favourable HL	early	465	November 2006-June 2011	France, Belgium, Italy, Netherlands	multi-centre	median 30 (range 15 to 49)	47.3%	0.0%
HD16	early	667	November 2009-29 December 2015	Germany, Switzerland, Austria, Netherlands	multi-centre	median 39 (range 18 to 75)	43.0%	9.0%
UK NCRI RAPID	early-inter-mediate	420	October 2003-August 2010	United Kingdom	multi-centre	median 34 (range 16 to 75)	50.0%	NR
Picardi 2007	early-inter-mediate	160	2000-2006	Italy	single centre	median 30 (range 15 to 70)	44.4%	0.0%

Table 3. Participant characteristics (Continued)

H10, unfavourable HL	intermediate	594	November 2006-June 2011	France, Belgium, Italy, Netherlands	multi-centre	median 31 (range 15 to 70)	53.2%	0.0%
HD17	intermediate	651	13 January 2012-21 March 2017	Germany, Switzerland, Austria, Netherlands	multi-centre	median 31 (range 18 to 60)	54.0%	2.0%
AHL2011	advanced	695	19 May 2011-29 April 2014	France, Belgium	multi-centre	median 30 (range 16 to 60)	37.0%	0.0%
HD0607	advanced	296	June 2008-June 2014	Italy, Israel	multi-centre	median 31 years (range 16 to 60)	57.1%	0.0%
HD0801	advanced	116	September 2008-April 2013	Italy	multi-centre	median 31 years (IQR 25 to 39)	45%	0.0%
HD18	advanced	1005	14 May 2008-18 July 2014	Germany, Switzerland, Austria, the Netherlands, and the Czech Republic	multi-centre	median 33 (range 18 to 60)	38.0%	6.2%
RATHL	advanced	935	August 2008-December 2012	United Kingdom, Italy, Norway, Australia/New Zealand, Sweden, Denmark	multi-centre	median 33 (range 18 to 79)	54.7%	0.0%
FDG-PET-positive with treatment escalation								
H10	early-intermediate	361	November 2006-June 2011	France, Belgium, Italy, Netherlands	multi-centre	median 30 years (range 15 to 70)	46.5%	0.0%
HD0607	advanced	148	June 2008-June 2014	Italy, Israel	multi-centre	median 30.5 (range 18 to 60)	43.3%	0.0%
HD18	advanced	434	14 May 2008-18 July 2014	Germany, Switzerland, Austria, the Netherlands, and the Czech Republic	multi-centre	median 30 years (range 18 to 60)	40.0%	4.3%

FDG: fluorodeoxyglucose

HL: Hodgkin lymphoma

IQR: interquartile range

NLP HL: Nodular lymphocyte-predominant Hodgkin lymphoma

NR: not reported



PET: positron emission tomography

Positron emission tomography-adapted therapy for first-line treatment in adults with Hodgkin lymphoma (Review)

Table 4. Outcome definitions

Study ID	Overall survival	Progression-free survival
FDG-PET-negative with treatment de-escalation		
HD16	"OS is calculated from the date noted on the staging form until the date of death, or (if the patient is still alive) until the last date for which information exists."	"PFS is the time from the date of staging ... until the date of the first occurrence of one of the below-mentioned events, or (if none of these occurred) until the last documented date of continuous remission."
UK NCRI RAPID	overall survival time	PFS was defined as "time from the date of randomization to first progression, relapse, or death, whichever occurred first".
Picardi 2007	"death from any cause"	"EFS was computed in terms of months since the start of chemotherapy to one of the events or to the last follow-up." "Relapse, secondary malignancies, or death from any cause were considered as failure (events)."
H10	"OS will be counted from the date of randomization to the date of death (from any cause). Alive patients will be censored at the date of last follow-up."	"PFS is counted from the date of randomization to the date of documentation of objective progression or death (from any cause), whichever occurs first."
HD17	"OS is calculated from the date noted on the staging form until the date of death, or (if the patient is still alive) until the last date for which information exists."	"PFS is the time from the date of staging (date on the ST form) until the date of the first occurrence of one of the below-mentioned events, or (if none of these occurred) until the last documented date of continuous remission."
AHL2011	"OS was defined as the time from random assignment to death of any cause or last follow-up"	"PFS was defined as the time from randomization assignment to first progression or relapse or death of any cause or last follow-up."
HD0607	OS "was measured from the date of registration to the date of death as a result of any cause."	"PFS was measured from the date of trial registration to the date of first appearance of disease progression, relapse, or death as a result of any cause."
HD0801	OS was "defined as the time from entry into the clinical trial (registration) until death from any cause or the date of the last follow-up".	PFS was "defined as the time from entry into the study (registration) until lymphoma progression or death from any cause."
HD18	OS was "defined as the time from completion of staging until death from any cause, or censored at the date of last information on the patient being alive".	PFS was "defined as the time from completion of staging until progression, relapse, or death from any cause."
RATHL	OS was "measured from date of randomisation to date of death from any cause".	PFS was "measured from date of randomisation to date of first appearance of disease progression, relapse or death from any cause."
FDG-PET-positive with treatment escalation		
H10	"OS will be counted from the date of randomization to the date of death (from any cause). Alive patients will be censored at the date of last follow-up."	"PFS is counted from the date of randomization to the date of documentation of objective progression or death (from any cause), whichever occurs first."

Table 4. Outcome definitions (Continued)

HD0607	OS "was measured from the date of registration to the date of death as a result of any cause."	"PFS was measured from the date of trial registration to the date of first appearance of disease progression, relapse, or death as a result of any cause."
HD18	OS was "defined as the time from completion of staging until death from any cause, or censored at the date of last information on the patient being alive".	PFS was "defined as the time from completion of staging until progression, relapse, or death from any cause."

EFS: event-free survival

FDG: fluorodeoxyglucose

OS: overall survival

PET: positron emission tomography

PFS: progression-free survival

Table 5. Response outcomes

Study ID	Stage	Reporting	Outcome	Experimental	Standard
FDG-PET-negative with treatment de-escalation					
HD16	early	assessed, not reported			
UK NCRI RAPID	early-intermediate	assessed, not reported			
Picardi 2007	early-intermediate	yes	complete response ^a	80/80	80/80
H10	early-intermediate	assessed, not reported			
HD17	intermediate	assessed, not reported			
AHL2011	advanced	reported, including PET-positive who all received standard therapy, ITT	complete response	334/410	338/413
			partial response	39/410	34/413
HD0607^b	advanced	according to PET status, in end-of-treatment PET-negative	complete response	540/580	
HD0801	advanced	all randomised end-of-treatment PET negative			
HD18	advanced	assessed, not reported			
RATHL	advanced	assessed, not reported			
FDG-PET-negative with treatment escalation					
H10	early-intermediate	assessed, not reported			
HD0607^b	advanced	according to PET status, in end-of-treatment PET-negative	complete response	81/108	

Table 5. Response outcomes (Continued)

HD18 ^c	advanced	final treatment outcome	complete remission	31/208	37/202
			unconfirmed complete remission	162/208	159/202

^aOnly those with complete response after 6 cycles were randomised to radiotherapy.

^bSecondary publication, Gallamini 2018

^cSecondary publication, Borchmann 2017

Abbreviations:

FDG-PET: fluorodeoxyglucose-positron emission tomography

ITT: intention-to-treat

Table 6. Fertility outcomes

Study ID	Stage	Population	Outcome	Experimental	Standard
FDG-PET-negative with treatment de-escalation					
H10, favourable HL	early	assessed; not yet published			
HD16 ^a	early	women, aged 18 to 40 years	cumulative incidence of childbirth after 5 years; HR 0.66 (95% CI 0.39 to 1.13)	17.9% (95% CI 10.4% to 25.5%)	24.0% (95% CI 15.9% to 32.2%)
		men, aged 18 to 60 years	37 of 590 have had children.		
UK NCRI RAPID	early-intermediate	not assessed			
Picardi 2007	early-intermediate	not assessed			
H10, unfavourable HL	intermediate	assessed; not yet published			
HD17	intermediate	assessed; not yet published			
AHL2011	advanced; all study participants, including PET-positive receiving standard therapy	women, aged ≤ 45 years	with at least one pregnancy	38.6% (29/75)	41.4% (29/70)
			Premature ovarian insufficiency: OR 0.20, 95% CI 0.08 to 0.50 in favour of the experimental group		
			Low ovarian reserve: OR 0.15, 95% CI 0.04 to 0.56 in favour of the experimental group		
			Ovarian function recovery: HR 2.52, 95% CI 1.73 to 3.67 in favour of the experimental group		

Table 6. Fertility outcomes (Continued)

		men, aged ≤ 45 years	with at least one pregnancy	9.5% (20/210)	2.8% (6/214)
Severe gonadal dysfunction: OR 0.26, 95% CI 0.13 to 0.50 in favour of the experimental group					
HD0607	advanced	not assessed			
HD0801	advanced	not assessed			
HD18	advanced, median 66 months FU	women, aged 18-29 years	5-year probability of first childbirth	4 x BEACOPP _{escalated} : 15.3%	6 x BEACOPP _{escalated} : 17.7%
					8 x BEACOPP _{escalated} : 18.0%
		women, aged 30-40 years		4 x BEACOPP _{escalated} : 12.9%	6 x BEACOPP _{escalated} : 10.9%
					8 x BEACOPP _{escalated} : 6.0%
		men, aged 18-60 years		4% (45/1105) male participants reported childbirth during a median of 65 months FU.	
RATHL ^b	advanced	women, aged 18-35 years	follicle-stimulating hormone concentrations; comparison between ABVD-AVD and BEACOPP groups	HR 0.40, 95% CI 0.19 to 0.83 in favour of ABVD-AVD	
				% recovered at 2 years (ABDV-AVD): 96%	
		women aged 35-40		HR 0.38, 95% CI 0.23 to 0.62 in favour of ABVD-AVD	
				% recovered at 2 years (ABDV-AVD): 83%	
		all	pregnancies (ABVD-AVD vs. BEACOPP)	57 participants with 71 pregnancies (17%)	7 participants with 10 pregnancies (13%)
FDG-PET-negative with treatment escalation					
H10	early-intermediate	assessed; not yet published			
HD0607	advanced	not assessed			
HD18	advanced not reported				

^a Reported in secondary reference, Fuchs 2023.^b Reported in Anderson 2018**Abbreviations:**

ABV: adriamycin, bleomycin, vinblastine

ABVD: adriamycin, bleomycin, vinblastine, dacarbazine

BEACOPP: bleomycin, etoposide, doxorubicin (Adriamycin), cyclophosphamide, vincristine (Oncovin), procarbazine, prednisolone (a steroid)

CI: confidence interval

FDG-PET: fluorodeoxyglucose-positron emission tomography

FU: follow-up

HL: Hodgkin lymphoma

HR: hazard ratio

vs.: versus

Table 7. Reporting of adverse events

Study ID	Stage	AE, any ^a	AE, grade 3-4 ^a	SAE	Other re- porting of AE, short- term ^b	Long-term toxicity
FDG-PET-negative with treatment de-escalation						
H10 , favourable HL	early	no	no	no	yes	yes
HD16	early	no	no	no	yes	yes
UK NCRI RAPID	early-intermediate	no	no	no	no	no
Picardi 2007	early-intermediate	no	no	no	yes	yes
H10 , unfavourable HL	intermediate	no	no	no	yes	yes
HD17	intermediate	no	yes, for two timings	yes	yes	yes
AHL2011	advanced	no	no	no	yes	yes
HD0607	advanced	no	no	no	yes	yes
HD0801	advanced	no	no	no	yes	yes
HD18	advanced	no	yes	no	yes	yes
RATHL	advanced	no	yes	no	yes	yes
FDG-PET-negative with treatment escalation						
H10	early-intermediate	no	no	no	yes	yes
HD0607	advanced	provided by authors	provided by authors	provided by authors	yes	yes
HD18	advanced	no	yes	no	yes	yes

^a presented as required for analysis (number of participants with at least one AE per treatment group)

^b presented in any other way (i.e., each AE individually, per grade, AE as unit of analysis, for both treatment arms together, etc.)

Abbreviations:

AE: adverse event

FDG-PET: fluorodeoxyglucose-positron emission tomography

HL: Hodgkin lymphoma

SAE: serious adverse event

Table 8. Funding and potential conflicts of interest

Study	Funding	Potential conflicts of interest
H10	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 Publique des Hopitaux de Paris (France), Societe Française de Medecine Nucleaire et Imagerie Moleculaire (France), Associazione Angela Serra (Italy), van Vlissingen Lymfoom Fonds (the Netherlands), and Chugai Pharmaceutical (Japan)	Reported; see Characteristics of included studies
HD16	Deutsche Krebshilfe Grants No. 108556 and 111744 and the Swiss State Secretariat for Education, Research, and Innovation	Reported; see Characteristics of included studies
UK NCRI RAPID	Leukaemia and Lymphoma Research, the Lymphoma Research Trust, Teenage Cancer Trust, and the U.K. Department of Health	Reported; see Characteristics of included studies
Picardi 2007	Associazione Italiana contro le Leucemie (Salerno and Benevento Sections)	Not reported
HD17	Deutsche Krebshilfe (grant no. 109238 and 110318)	Reported; see Characteristics of included studies
AHL2011	French government Programme Hospitalier de Recherche Clinique 2010	Reported; see Characteristics of included studies
HD0607	Associazione Italiana per la Ricerca sul Cancro I.G. No. 203 (A.G.); the Associazione Italiana Lotta alla Leucemia, Bergamo and Cuneo sections; and a research grant from the Bank Foundation Cassa di Risparmio of Cuneo	Reported; see Characteristics of included studies
HD0801	Italian Lymphoma Foundation	Reported; see Characteristics of included studies
HD18	Deutsche Krebshilfe, Swiss State Secretariat for Education and Research, and Roche Pharma AG	"We declare no competing interests."
RATHL	Cancer Research UK (reference CRUK/07/033), Leukaemia and Blood Cancer New Zealand, and Cancer Australia	Reported; see Characteristics of included studies

Table 9. Treatment discontinuation

Study ID	Stage	Reason/definition	Events in experimental group	N in experimental group	Events in standard group	N in standard group
FDG-PET-negative with treatment de-escalation						
H10 , favourable HL	early	not reported				
HD16	early	not reported				

Table 9. Treatment discontinuation *(Continued)*

UK NCRI RAPID	early-intermedi- ate	not reported				
Picardi 2007	early-intermedi- ate	not reported				
H10, un- favourable HL	intermediate	not reported				
HD17	intermediate	discontinuation due to drug-relat- ed toxicity	2	550	1	546
AHL2011	advanced	discontinuation due to toxicity	27	413	4	410
		discontinuation due to disease progression	12	413	10	410
HD0607	advanced	not reported				
HD0801	advanced	treatment interruptions	0	58	9	58
HD18	advanced	discontinuation due to toxicity	11	504	2	501
RATHL	advanced	discontinuation due to toxicity	7	465	5	470
FDG-PET-positive with treatment escalation						
H10	early-intermedi- ate	not reported				
HD0607	advanced	stopped treatment as a result of treatment toxicity	2/148 participants, during BEACOPP _{escalated}			
HD18	advanced	not reported				

Abbreviations:

BEACOPP: bleomycin, etoposide, doxorubicin (Adriamycin), cyclophosphamide, vincristine (Oncovin), procarbazine, prednisolone (a steroid)

FDG-PET: fluorodeoxyglucose-positron emission tomography

HL: Hodgkin lymphoma

N: sample size

APPENDICES
Appendix 1. Search strategies

MEDLINE ALL (via OVID) 1946 to November 17, 2023

Searches

1 Lymphoma/

2 exp Hodgkin Disease/

3 germinoblastom*.tw,kf.

4 reticulolymphosarcom*.tw,kf.

5 hodgkin*.tw,kf.

6 (malignan* adj2 (lymphogranulom* or granulom*)).tw,kf.

7 or/1-6

8 exp Tomography, Emission-Computed/

9 exp Positron-Emission Tomography/

10 (pet* or (positron* and emission*) or (positron* and tomograph*) or (emission* and tomograph*)).tw,kf.

11 or/8-10

12 7 and 11

13 randomized controlled trial.pt.

14 controlled clinical trial.pt.

15 randomi?ed.ab.

16 placebo.ab.

17 drug therapy.fs.

18 randomly.ab.

19 trial.ab.

20 groups.ab.

21 or/13-20

22 exp animals/ not humans/

23 21 not 22

24 clinical trial, phase III/

25 ("Phase 3" or "phase3" or "phase III" or P3 or "PIII").ti,ab,kw.

26 (24 or 25) not 22

27 23 or 26

28 12 and 27

Line 13 to line 23: [Lefebvre 2022](#)

Line 24 to line 26: [Cooper 2019](#)

Embase (via OVID) 1974 to present

Searches

1 lymphoma/

2 exp 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 computer assisted emission tomography/

9 (tomograph* and emission*).tw,kw.

10 exp positron emission tomography/

11 (pet* or (positron* and emission*) or (positron* and tomograph*)).tw,kw.

12 or/8-11

13 7 and 12

14 Randomized controlled trial/

15 Controlled clinical study/

16 random*.ti,ab.

17 randomization/

18 intermethod comparison/

19 placebo.ti,ab.

20 (compare or compared or comparison).ti.

21 (open adj label).ti,ab.

22 ((double or single or doubly or singly) adj (blind or blinded or blindly)).ti,ab.

23 double blind procedure/

24 parallel group\$1.ti,ab.

25 (crossover or cross over).ti,ab.

26 ((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.

27 (controlled adj7 (study or design or trial)).ti,ab.

28 (volunteer or volunteers).ti,ab.

29 trial.ti.

30 or/14-29

31 phase 3 clinical trial/

32 ("Phase 3" or "phase3" or "phase III" or P3 or "PIII").ti,ab,kw.

33 31 or 32

34 (animal experiment/ or Animal experiment/) not (human experiment/ or human/)

35 (30 or 33) not 34

36 13 and 35

Line 14 to line 30 and line 33 and 34: [Glanville 2019](#)

Line 31 to line 35: [Cooper 2019](#)

Cochrane Central Register of Controlled Trials (Central, 2023, Issue 10) (via Cochrane Library)

ID Search

#1 MeSH descriptor: [Lymphoma] explode all trees

#2 MeSH descriptor: [Hodgkin Disease] explode all trees

#3 germinoblastom*

#4 reticulolymphosarcom*

#5 (hodgkin* or hogkin* or hodkin* or hodgin*):ti,ab,kw

#6 (malignan* near/2 (lymphogranulom* or granulom*))

#7 #1 or #2 or #3 or #4 or #5 or #6

#8 MeSH descriptor: [Tomography, Emission-Computed] explode all trees

#9 MeSH descriptor: [Positron-Emission Tomography] explode all trees

#10 (pet* or (positron* and emission*) or (positron* and tomograph*) or (tomograph* and emission*)):ti,ab,kw

#11 #8 or #9 or #10

#12 #7 and #11 in Trials

ClinicalTrials.gov

Expert search:

Hodgkin | Interventional Studies | (tomography, emission-computed OR positron-emission tomography OR PET OR FDG-PET OR PET-directed OR PET-response OR PET/CT)

ICTRP (<https://trialsearch.who.int/>)

advanced search

In the condition: Hodgkin

In the intervention: tomography OR positron-emission OR emission-computed OR PET OR FDG-PET OR PET-directed OR PET-response OR PET/CT

Recruitment status: all

Appendix 2. Sensitivity analyses (random-effects model vs. fixed effects model, OS and PFS)

Comparison	Outcome	REM	FEM
Early-stage PET-negative	OS	0.84 [0.13, 5.32]	0.64 [0.21, 1.99]
	PFS	4.52 [0.72, 28.41]	2.46 [1.52, 4.01]
Intermediate-stage PET-negative	OS	0.91 [0.42, 1.96]	0.91 [0.42, 1.96]
	PFS	1.59 [0.95, 2.67]	1.59 [0.95, 2.67]
Early-to-intermediate PET-negative	OS	NA, only one study	
	PFS	2.00 [1.00, 3.97]	1.90 [1.10, 3.28]
Advanced-stage PET-negative, chemotherapy	OS	0.65 [0.40, 1.07]	0.67 [0.46, 0.98]
	PFS	0.98 [0.78, 1.25]	0.98 [0.78, 1.25]
Advanced-stage PET-negative, RT	OS	0.63 [0.11, 3.69]	0.63 [0.11, 3.69]

(Continued)

	PFS	0.78 [0.43, 1.43]	0.78 [0.43, 1.43]
Early-to-intermediate PET-positive	NA, only one study		
Advanced PET-positive	OS	1.39 [0.74, 2.63]	1.39 [0.74, 2.63]
	PFS	1.03 [0.68, 1.54]	1.02 [0.69, 1.51]

Appendix 3. Summary of findings: PET-adapted treatment compared to standard of care in PET-negative early-to-intermediate-stage HL

PET-adapted de-escalation compared to standard of care in PET-negative, early-to-intermediate-stage HL Patient

or population: PET-negative, early-to-intermediate stage Hodgkin lymphoma

Setting: Inpatients and outpatients

Intervention: PET-adapted de-escalation (chemotherapy, no further radiotherapy)

Comparison: Standard of care (chemotherapy plus radiotherapy)

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	N of participants (studies)	Certainty of the evidence (GRADE)	Comment
	Frequency with standard of care	Frequency with PET-adapted de-escalation				
Overall survival (OS) at 3 years	971 per 1000 [§]	985 per 1000 (951 to 985)	HR 0.51 (0.15 to 1.71) [death]	420 (1 RCT)	⊕⊕⊕⊕ Low ^a	PET-adapted de-escalation may increase overall survival compared to standard therapy.
Progression-free survival (PFS) at 3 years	946 per 1000 [§]	895 per 1000 (802 to 946)	HR 2.00 (1.00 to 3.97) [progression, death]	580 (2 RCTs)	⊕⊕⊕⊕ Low ^{b,c}	PET-adapted de-escalation may decrease progression-free survival compared to standard therapy.
Treatment-related mortality (TRM) during treatment	-	-	-	-	-	There were no studies with published results on treatment-related mortality for this comparison.
Serious adverse events (SAE) Follow-up until 30 days after last protocol treatment, or longer if related	-	-	-	-	-	There were no studies with published results on serious adverse events for this comparison.

(Continued)

Incidence of secondary malignancy	-	-	-	-	-	There were no studies with published results on the incidence of secondary malignancies for this comparison.
Quality of life (QoL)	-	-	-	-	-	There were no studies with published results on quality of life for this comparison.

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

§ **Baseline risk** estimated based on the median 3-year overall survival and progression-free survival of the standard arm in RAPID.

CI: confidence interval; HR: hazard Ratio; RCT: randomised controlled trial; RR: risk 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.

a. Downgraded two levels for very serious imprecision due to wide confidence intervals and small sample size.

b. Downgraded one level for serious risk of bias (all trials were open-label, which may have affected subjective outcome assessments such as PFS, TRM, quality of life).

a. Downgraded one level for serious imprecision due to wide confidence intervals and small sample size.

WHAT'S NEW

Date	Event	Description
26 March 2025	New citation required and conclusions have changed	See event updated.
26 March 2025	New search has been performed	Authors: The authors have changed. Peter Borchmann and Andreas Engert were involved in the previous version of the review, but not in this update. Nina Kreuzbeger and Marius Goldkuhle were added to the review. All other review authors remained the same. Methods: - In comparison to the previous review version, we separated early-stage into early-and-intermediate stage Hodgkin lymphoma (HL) where primary studies allowed this. - We defined outcomes in more detail, i.e. adverse events, any grade, grade 3-4, serious adverse events, and second primary malignancy. - We did not handsearch conference proceedings, as these are included in Embase, which we have added according to Cochrane standards. The meta-register of controlled trials is no longer available.

Date	Event	Description
		- We used random-effects models, assuming that underlying effects can vary between studies; for example, due to the differences between study populations, health systems, etc. Included studies: The first publication of this review in 2015 included three studies with four comparisons (H10; Picardi 2007; UK NCRI RAPID). Since then, data from seven eligible studies on various comparisons have been published (AHL2011; HD0607; HD0801; HD16; HD17; HD18; RATHL), and H10 and UK NCRI RAPID (previously abstract only) include additional data. Overall, 667 participants with early-stage, 651 participants with intermediate-stage, 1639 participants with early-to-intermediate stage, and 3629 participants with advanced-stage HL are included in this version, compared to 1999 early-to-intermediate stage HL participants only in the previous version. Comparisons with previously no data that have been added in this version: - Separation of the treatment de-escalation comparison in PET-negative participants, early vs. intermediate stage; - Treatment de-escalation in PET-negative participants, advanced-stage HL; - Treatment escalation in PET-positive participants, early-to-intermediate HL; - Treatment escalation in PET-positive participants, advanced stage HL. Outcomes: The current version adds data on overall survival as time-to-event long-term follow-up data, the analysis of adverse events, separated into any, grade 3-4, serious AE, and secondary malignancies. Fertility and response outcomes are listed in table format due to heterogeneity. Main results: The current version of the review is differently structured regarding the HL stage. These are the main findings: <u>PET-negative participants</u> - In early-stage HL, data on OS and PFS are uncertain, while no treatment-related death was reported in either arm. Adverse events may not differ between arms. This comparison includes follow-up data of H10 that was previously included in the combined early-to-intermediate stage ("early-stage") HL comparison. - In intermediate-stage HL, PET-adapted de-escalation may increase OS but decrease PFS. Certainty in the evidence is low. This comparison includes follow-up data from H10 that was previously included in the combined early-to-intermediate stage ("early-stage") HL comparison. - In the early-to-intermediate stage, little to no effect was shown on OS, PFS, SAE, and secondary malignancy. Certainty in the evidence is low. Previously, where all studies were analysed together, data on OS and AE were uncertain, while data on PFS were rated as moderate. - In advanced-stage HL, a new comparison not covered previously, OS probably increases with treatment de-escalation while PFS shows no difference (moderate certainty). <u>PET-positive participants</u>

Dat	Eve	Descriptio
		Data for this comparison are new in this version, although the study (H10) was previously included. - In early-to-intermediate stage HL, chemotherapy intensification may increase OS and PFS, while we are uncertain about the effects on secondary malignancies. - In advanced stage HL, treatment intensification by adding rituximab probably does not increase OS and PFS, but probably increases SAEs (moderate certainty).

HISTORY

Protocol first published: Issue 6, 2013

Review first published: Issue 1, 2015

CONTRIBUTIONS OF AUTHORS

- Nina Kreuzberger: study screening, data extraction, risk of bias assessment, data analysis and interpretation, writing of the review;
- Marius Goldkuhle: study screening, data extraction, risk of bias assessment, data analysis check, proofread final version;
- Bastian von Tresckow: clinical expertise, defined the grouping of HL, contributed to introduction and discussion, proofread final version;
- Carsten Kobe: clinical expertise, defined the grouping of HL, contributed to introduction and discussion, proofread final version;
- Marie-Therese Sickinger: clinical expertise, screening, data extraction, risk of bias assessment, data analysis of studies included in the previous version, proofread final version;
- Ina Monsef: developed the search strategies, wrote the search section of the manuscript, proofread final version;
- Nicole Skoetz: study screening, data extraction, risk of bias assessment, data analysis and interpretation of the previous version of this review, proofreading of the protocol and review, statistical and methodological advice and supervision, content input, proofreading final version.

DECLARATIONS OF INTEREST

NK	None known. NK is a member of staff at Cochrane Haematology. She was not involved in the editorial process for this review.
MG	None known. MG is a member of staff at Cochrane Haematology. He was not involved in the editorial process for this review.
BvT	BvT declares grant/research funding from Esteve, Merck Sharp & Dohme Corporation, Novartis and Takeda Pharmaceutical Company (all paid to institution). He also declares (1) consulting fees from Allogene, Amgen, Bristol-Myers Squibb, Cerus Corporation, Gilead Sciences Inc, Incyte Corporation, IQVIA Ltd, Lilly Deutschland, Merck Sharp and Dohme, Miltenyi Biotec, Noscendo GmbH, Novartis Pharma, PentixaPharm GmbH, Pfizer Canada Inc, Pierre Fabre Pharmaceuticals, Inc, Qualworld, Quintiles Transnational Corp, Roche, Sobi, Inc and Takeda Pharmaceutical Company; (2) honoraria for speaking engagements from AbbVie, AstraZeneca, Bristol-Myers Squibb, Gilead Sciences Inc., Incyte Corporation, Lilly Deutschland, Merck Sharp and Dohme, Novartis, Roche and Takeda Pharmaceutical Company; and (3) travel support from AbbVie, AstraZeneca, Gilead Sciences Inc, Lilly Deutschland, Merck Sharp and Dohme, Novartis, Pierre Fabre Pharmaceuticals, Inc, Roche and Takeda Pharmaceutical Company (all personal payments). BvT is a health professional in haematology and stem cell transplantation at University Hospital Essen and an Editor with Cochrane Haematology. He was not involved in the editorial process for this protocol. BvT was engaged with studies included in this systematic review (HD16, HD17, HD18). He was therefore not involved in data extraction, risk of bias assessment and analysis.
CK	None known. CK is a health professional in Nuclear Medicine at the University Hospital Cologne.

CK was engaged with studies included in this systematic review (HD16, HD17). He was therefore not involved in data extraction, risk of bias assessment and analysis.

M-TS	None known. M-TS is a healthcare professional at University Hospital St. Pölten.
IM	None known. IM is the Information Specialist at Cochrane Haematology. She was not involved in the editorial process for this review.
NS	None known. She is the Co-ordinating Editor of Cochrane Haematology. She was not involved in the editorial process for this review.

SOURCES OF SUPPORT

Internal sources

- Cochrane Haematology, Institute of Public Health, University of Cologne, Germany
Income for employment
- Department I of Internal Medicine, University Hospital Cologne, Germany
Income for employment

External sources

- No external sources of support, Other
This publication was conducted as part of a doctoral thesis. No external funding was obtained.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

Differences between the protocol and the first version of the review

We planned to evaluate the role of PET-adapted therapy for the first-line treatment of individuals with all stages of HL, and in both PET-positive and PET-negative individuals. As only data for PET-negative individuals in early-stage HL were available, we were unable to evaluate the role of PET-adapted therapy in individuals with advanced-stage or PET-positive disease.

We planned to report the HR for OS but, as we were unable to pool the data, we reported the RR. Furthermore, we were unable to analyse QoL, TRM and response rates, as these outcomes were not reported in the included trials. Due to lack of data, we were also unable to perform subgroup analyses for age, stage of disease, type and intensity of chemotherapeutic regimen, and duration of follow-up, or sensitivity analyses for quality components and preliminary results versus mature results.

Due to the design of the studies, we added a subgroup analysis for PFS comparing 'radiotherapy versus no radiotherapy' with 'radiotherapy versus chemotherapy without radiotherapy'.

Differences between the first review and the current review update

Methods

Criteria for considering studies for this review

In contrast to the previous version and the protocol, we aimed to limit inclusion to adult participants with newly diagnosed HL (at least 80% of participants of a trial > 18 years), because the disease course differs between adults and children.

We defined the outcomes in more detail. For instance, we detailed that we aimed to meta-analyse any grade adverse events, adverse events grade 3 to 4, serious adverse events, and the incidence of secondary malignancy.

Search methods

We added Embase to our repertoire of databases of medical literature searched according to Cochrane's requirements.

Handling of missing data

According to the protocol, we intended to perform sensitivity analyses to estimate how sensitive the results were to reasonable changes in the assumptions regarding missing data. As missing individual data within available outcomes were limited, this was not necessary.

Data synthesis

At the protocol stage, it was decided to use fixed-effect models except in cases of heterogeneity. For the current version of this review, we decided to use random-effects models irrespective of the presence of statistical heterogeneity, as we assume that intervention effects are not the same across trials but rather follow the normal distribution.

In the protocol, it was specified that we would calculate numbers needed to treat for an additional beneficial outcome (NNTB) and numbers needed to treat for an additional harmful outcome (NNTH) with corresponding 95% CIs for clinical interpretation; however, based on the current format of summary of findings tables, we calculated absolute risks for clinical interpretation.

In contrast to the protocol, we treated the disease stages of early, intermediate, and advanced not only as subgroups but as entirely separate comparisons as, in clinical practice, even non-PET-adapted treatment modalities differ between those groups. We had to add a comparison for studies that included both early and intermediate-stage HL.

Summary of findings

We moved the section according to the up-to-date review format and added details.

Results

We were unable to analyse QoL and response rates, as these outcomes were not reported in the included trials, or were reported too heterogeneously for analysis. Due to the small number of studies per comparison and a similar risk of bias across studies (all open-label studies), we did not examine the effect of RoB on the effect estimates. All studies published mature results as full-text publications.

Due to the design of the studies, we created a subgroup analysis in early and intermediate-stage HL comparing the study replacing radiotherapy with additional chemotherapy with the study simply omitting radiotherapy. For advanced stages, a subgroup based on initial chemotherapy regimen was added; both fall under the preplanned subgroup "intensity of chemotherapeutic regimen".

Stage was examined as different comparisons. We could not perform subgroup analysis regarding age.

NOTES

Parts of this review text are taken over from the protocol and first full version of this Cochrane review ([Sickinger 2013](#); [Sickinger 2015](#)). The methods section was based on the templates of Cochrane Haematology.

INDEX TERMS

Medical Subject Headings (MeSH)

Antineoplastic Agents [therapeutic use]; Chemoradiotherapy [methods]; Disease-Free Survival; Fluorodeoxyglucose F18; Hodgkin Disease [*diagnostic imaging] [mortality] [pathology] [*therapy]; *Positron-Emission Tomography [methods]; Radiopharmaceuticals; Randomized Controlled Trials as Topic; Selection Bias

MeSH check words

Female; Humans; Male

Appendix V: Declaration of an Oath**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 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, 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(1), Art. No. CD012643. doi: 10.1002/14651858.CD012643.pub3. PMID: 31930780.
2. Kreuzberger N, Damen JA, Trivella M, Estcourt LJ, Aldin A, Umlauff L, Vazquez-Montes MD, Wolff R, Moons KG, Monsef I, Foroutan F, Kreuzer KA, Skoetz N. Prognostic models for newly-diagnosed chronic lymphocytic leukaemia in adults: a systematic review and meta-analysis. Cochrane Database Systematic Reviews 2020; Issue 7(7), Art. No. CD012022. doi: 10.1002/14651858.CD012022.pub2
3. Kreuzberger N, Goldkuhle M, von Tresckow B, Kobe C, Sickinger M, Monsef I, Skoetz N. Positron emission tomography-adapted therapy for first-line treatment in adults with Hodgkin lymphoma. Cochrane Database of Systematic Reviews 2025, Issue 3. Art. No.: CD010533. doi: 10.1002/14651858.CD010533.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.

Köln, 19. August 2025

Datum

Unterschrift