

Aus dem Institut für Medizinsoziologie, Versorgungsforschung  
und Rehabilitationswissenschaft (IMVR)

Gemeinsames Institut der Medizinischen Fakultät und der  
Humanwissenschaftlichen Fakultät der Universität zu Köln

Direktor: Universitätsprofessor Dr. phil. H. Pfaff

# **Patient-reported Outcomes für die Versorgung von Krebspatient\*innen**

Inaugural-Dissertation zur Erlangung der Doktorwürde  
der Medizinischen Fakultät  
der Universität zu Köln

vorgelegt von  
Nora Tabea Sibert  
aus Düsseldorf

promoviert am 22.Mai 2023



Dekanin/Dekan: Universitätsprofessor Dr. med. G. R. Fink

1. Gutachterin oder Gutachter: Universitätsprofessor Dr. phil. H. Pfaff
2. Gutachterin oder Gutachter: Professor Dr. med. M. Kochanek

## Erklärung

Ich erkläre hiermit, dass ich die vorliegende Dissertationsschrift ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskriptes habe ich Unterstützungsleistungen von Herrn PD Dr. rer. medic. Christoph Kowalski erhalten.

Weitere Personen waren an der Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich nicht die Hilfe einer Promotionsberaterin/eines Promotionsberaters in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertationsschrift stehen.

Die Dissertationsschrift wurde von mir bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Dieser Arbeit liegen unter anderem zwei Datensätze zugrunde. Für die Arbeiten 1 und 3 wurde der Datensatz der „Prostate Cancer Outcome“ Studie (PCO Studie) genutzt. Dafür wurden in teilnehmenden Studienzentren ohne meine Mitarbeit (Befragungs-) Daten erhoben. Diese wurden anschließend ohne meine Mitarbeit von der OnkoZert GmbH gebündelt und mir über die Deutsche Krebsgesellschaft zur Verfügung gestellt. Der so erhaltene Datensatz wurde von mir für die weiteren statistischen Auswertungen aufbereitet. Für die Arbeit 4 wurde der Datensatz der „Ergebnisqualität bei Darmkrebs: Identifikation von Ursachen und Maßnahmen zur flächendeckenden Qualitätsentwicklung“ Studie (EDIUM Studie) genutzt, wobei die Datenerhebung und -bündelung analog zur PCO Studie ablief (eigenständige Erhebung in den EDIUM-Studienzentren, Datenbündelung durch OnkoZert). Als wissenschaftliche Mitarbeiterin bei der Deutschen Krebsgesellschaft war ich dabei auch für die EDIUM-Projektunterstützung zuständig. Der EDIUM-Datensatz wurde von mir für die weiteren statistischen Auswertungen aufbereitet. Die Projektleitung der PCO und der EDIUM Studie haben PD Dr. rer. medic. Christoph Kowalski und PD Dr. med. Simone Wesselmann (beide Deutsche Krebsgesellschaft e.V.) inne.

Die statistischen Auswertungen der Arbeiten 1 und 3 wurden von mir eigenständig geplant und durchgeführt mit Unterstützung von PD Dr. rer. medic. Christoph Kowalski. Für die Arbeit 4 wurden die für diese Promotionsschrift relevanten statistischen Auswertungen (Adjustierung der patient-reported Outcomes) eigenständig von mir geplant und ausgeführt, die Zusammenhangsanalysen wurden von Dr. rer. med. Rebecca Roth und Anna Hagemann (Universität Köln) als Konsortialpartnerinnen der EDIUM-Studie in enger Absprache mit mir durchgeführt. PD Dr. rer. medic. Christoph Kowalski unterstützte die Auswertungen und Dateninterpretation. Insbesondere entstand der erste Manuskriptentwurf für Arbeit 4 durch PD Dr. rer. medic. Christoph Kowalski als Studienleitung der EDIUM-Studie; die weitere Manuskripterstellung fand in enger Absprache mit mir statt.

Das systematische Review (Arbeit 2) wurde von mir konzeptualisiert und mit Unterstützung von PD Dr. rer. medic. Christoph Kowalski als zweiten Reviewer durchgeführt. Die Datenanalyse und -interpretation fand primär durch mich statt, ebenso wie die Manuskripterstellung.

Die Planung und Durchführung der qualitativen Analyse (Arbeit 5) fand im Rahmen der EDIUM Studie statt und wurde von mir sowie Clara Breidenbach (wissenschaftliche Mitarbeiterin Deutsche Krebsgesellschaft e.V.) ausgeführt: Der Leitfaden für die Interviews wurde dabei von

Clara Breidenbach und mir erstellt, die Interviews wurden von uns beiden durchgeführt und transkribiert, die Transkripte wurden von uns beiden codiert. Die anschließende Analyse der hier vorgestellten Forschungsfrage wurde von mir eigenständig durchgeführt und anschließend mit Clara Breidenbach und Dr. rer. medic. Christoph Kowalski diskutiert.

Alle Manuskripte (Ausnahme: Arbeit 4, siehe oben) wurden zunächst eigenständig von mir entworfen und anschließend von den entsprechenden Ko-Autor\*innen gegengelesen und kommentiert. Dabei handelt es sich im Speziellen um folgende Publikationen: „Psychometric validation of the German version of the EPIC-26 questionnaire for patients with localized and locally advanced prostate cancer“ veröffentlicht im *World Journal of Urology* (Impact Factor: 3,7), „Different Approaches for Case-Mix Adjustment of Patient-Reported Outcomes to Compare Healthcare Providers – Methodological Results of a Systematic Review“ veröffentlicht in *Cancers* (Impact Factor: 6,6), „Variation across operating sites in urinary and sexual outcomes after radical prostatectomy in localized and locally advanced prostate cancer“ veröffentlicht im *World Journal of Urology* (Impact Factor: 3,7), „Ergebnisqualität in zertifizierten Darmkrebszentren nach Resektion eines kolorektalen Karzinoms“ veröffentlicht im *Deutschen Ärzteblatt* (Impact Factor: 8,3) sowie „Clinicians' knowledge and attitudes towards patient reported outcomes in colorectal cancer care – insights from qualitative interviews“ veröffentlicht in *BMC Health Services Research* (Impact Factor: 2,9).

In meiner Erklärung über den Eigenanteil an den Publikationen findet sich eine detaillierte Beschreibung der Anteile aller Ko-Autor\*innen.

Erklärung zur guten wissenschaftlichen Praxis:

Ich erkläre hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten (Amtliche Mitteilung der Universität zu Köln AM 132/2020) der Universität zu Köln gelesen habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen.

Berlin, den 21. Dezember 2022

Unterschrift:  .....

# Danksagung

An dieser Stelle möchte ich mich bei all denjenigen bedanken, die mich bei der vorliegenden Dissertation unterstützt haben.

Zunächst gilt mein großer Dank Prof. Dr. Holger Pfaff für die Möglichkeit der Promotionsbetreuung und die wertvolle inhaltliche Unterstützung, den konstruktiven Ideenaustausch und das Aufzeigen neuer versorgungswissenschaftlicher Perspektiven.

PD Dr. Christoph Kowalski danke ich herzlichst, mir die Tür geöffnet zu haben zur spannenden und höchst relevanten Disziplin der Versorgungsforschung, für die vielen Denkanstöße und äußerst fruchtbaren Diskussionen über den Sinn (und gelegentlichen Unsinn) verschiedenster Methodiken und Forschungsansätze, für die stetige Unterstützung und Ermutigung, für die Offenheit für neue Ideen, für das äußerst aufmerksame Lektorat – und vor allem für die spannende, vertrauensvolle und stets erheiternde gemeinsame Arbeit.

Ich danke meinem Team bei der Deutschen Krebsgesellschaft: Allen voran PD Dr. Simone Wesselmann als Leiterin der Abteilung Zertifizierung, durch deren Arbeit erst die Bedingungen geschaffen wurden, auf denen die Versorgungsforschung in Krebszentren aufbaut; und vor allem auch herzlichst Clara Breidenbach, mit der ich gemeinsam die ersten Schritte auf dem Wissenschaftsparkett beschreiten durfte und deren Unterstützung und den gemeinsamen Austausch ich sehr schätze.

An den dieser Promotionsschrift zugrunde liegenden Publikationen waren zahlreiche Ko-Autor\*innen beteiligt, denen ich für diese gemeinsame Arbeit, die Datenerhebung und die kritische Kommentierung der Manuskripte sehr danke.

Von tiefstem Herzen dankbar bin ich meiner Mutter und meiner Schwester.

Zu guter Letzt möchte ich mich besonders bedanken bei den teilnehmenden Patient\*innen der EDIUM- und PCO-Studien: Ohne diese rege Beteiligung – auch ein Jahr nach Akutbehandlung im Zentrum – wären die in dieser Arbeit vorgestellten Überlegungen und Schlussfolgerungen nicht möglich gewesen. Ich danke ihnen deswegen für das Vertrauen in und die Unterstützung von onkologischer Versorgungsforschung und hoffe sehr, diese Arbeit kann einen Beitrag dazu leisten, die onkologische Versorgung in Deutschland im Sinne der Patient\*innen zu verbessern.

# Inhaltsverzeichnis

<b>ABKÜRZUNGSVERZEICHNIS</b>	<b>8</b>
<b>1. ZUSAMMENFASSUNG</b>	<b>9</b>
<b>2. EINLEITUNG</b>	<b>11</b>
2.1. <i>Outcome</i> - und Patient*innenorientierung in der Versorgungsforschung	12
2.2. <i>Outcomes</i> aus Sicht der Betroffenen – Nutzung von <i>patient-reported Outcomes</i> in der Onkologie auf verschiedenen Analyseebenen	13
2.3. PROs im Kontext der onkologischen Versorgung in Deutschland	16
2.3.1. Qualitätssicherung in der Onkologie – das Zertifizierungsverfahren der DKG	17
2.4. Fallbeispiel 1: Ergebnisqualität in zertifizierten Prostatakrebszentren – die PCO-Studie	18
2.5. Fallbeispiel 2: Ergebnisqualität in zertifizierten Darmkrebszentren – die EDIUM-Studie	19
2.6. Zielsetzung dieser Arbeit: Anwendungsmöglichkeiten von PROs auf verschiedenen Ebenen der Versorgung sowie hemmende und fördernde Faktoren für eine erfolgreiche Implementierung	20
<b>3. PUBLIKATIONEN</b>	<b>22</b>
3.1. Arbeit 1: Validierung des Messinstrumentes - Voraussetzung für eine sinnvolle Integration von PROs in die Versorgung	22
3.2. Arbeit 2: Systematic Review zu Methoden der Casemix-Adjustierung bei PROs – Übersicht über genutzte Methoden zum Vergleich von Versorgungsorganisationen	23
3.3. Arbeit 3: Funktionelles Outcome ein Jahr nach Radikaler Prostatektomie in 88 Prostatakrebszentren – Vergleich der Ergebnisqualität auf Versorgungsorganisationsebene (Fallbeispiel 1)	23
3.4. Arbeit 4: Funktionelles Outcome ein Jahr nach elektiver Tumorsektion beim kolorektalen Karzinom in 102 Darmkrebszentren – Vergleich der Ergebnisqualität auf Versorgungsorganisationsebene (Fallbeispiel 2)	24
3.5. Arbeit 5: Einstellungen zu und Wissen über PROs von Behandelnden – Möglichkeiten und Limitationen der Implementierung von PROs auf patient*innenindividueller Ebene	24

<b>4.</b>	<b>PUBLIKATIONEN DIESER ARBEIT</b>	<b>26</b>
4.1.	Psychometric validation of the German version of the EPIC-26 questionnaire for patients with localized and locally advanced prostate cancer	26
4.2.	Different Approaches for Case-Mix Adjustment of Patient-Reported Outcomes to Compare Healthcare Providers – Methodological Results of a Systematic Review	42
4.3.	Variation across operating sites in urinary and sexual outcomes after radical prostatectomy in localized and locally advanced prostate cancer	61
4.4.	Ergebnisqualität in zertifizierten Darmkrebszentren nach Resektion eines kolorektalen Karzinoms	72
4.5.	Clinicians‘ knowledge and attitudes towards patient reported outcomes in colorectal cancer care – insights from qualitative interviews	84
<b>5.</b>	<b>DISKUSSION: PATIENT-REPORTED OUTCOMES UND IHRE NUTZUNGSMÖGLICHKEITEN AUF VERSCHIEDENEN VERSORGUNGSEBENEN IN DER ONKOLOGIE</b>	<b>95</b>
5.1.	Forschungsfrage 1: Wie lassen sich PROs als relevante Outcomes für den Vergleich von Versorgungsorganisationen unter der Berücksichtigung des Patient*innenkollektivs adjustieren?	96
5.2.	Forschungsfrage 2: Wie unterscheiden sich Krebszentren bezüglich der <i>patient-reported Outcomes</i> ihrer behandelten Patient*innen voneinander?	97
5.3.	Forschungsfrage 3: Wie ist das Wissen über und die Haltung zu PROs von onkologisch Behandelnden in deutschen Krebszentren für die patient*innenindividuelle sowie die Versorgungsorganisations-übergreifende Nutzung?	99
5.4.	PROs für die Versorgung von Krebspatient*innen – Integration und Synthese der Ergebnisse auf verschiedenen Ebenen	100
5.5.	Fazit	102
<b>6.</b>	<b>LITERATURVERZEICHNIS</b>	<b>105</b>
<b>7.</b>	<b>ANHANG</b>	<b>110</b>
7.1.	Abbildungsverzeichnis	110



## Abkürzungsverzeichnis

<b>addz</b>	Arbeitsgemeinschaft deutscher Darmkrebszentren
<b>BMG</b>	Bundesministerium für Gesundheit
<b>BPS</b>	Bundesverband der Prostatakrebs Selbsthilfe
<b>CASP</b>	Critical Appraisal Skills Program
<b>DKG</b>	Deutsche Krebsgesellschaft e.V.
<b>DRKS</b>	Deutsches Register klinischer Studien
<b>EDIUM</b>	Ergebnisqualität bei Darmkrebs: Identifikation von Ursachen und Maßnahmen zur flächendeckenden Qualitätsentwicklung
<b>EORTC</b>	European Organization for Research and Treatment of Cancer
<b>EPIC-26</b>	Expanded Prostate Cancer Composite 26-items Short Form
<b>ePRO</b>	elektronisch erfasste patient-reported Outcomes
<b>ESMO</b>	European Society of Medical Oncology
<b>ICHOM</b>	International Consortium for Health Outcomes Measurement
<b>ILCO</b>	Solidargemeinschaft für Stomaträger und Menschen mit Darmkrebs
<b>LL</b>	Leitlinie
<b>MEDline</b>	Medical Literature Analysis and Retrieval System Online
<b>MEDTEP</b>	Medical Treatment Effectiveness Program
<b>MID</b>	minimally important difference
<b>NHS</b>	National Health Service
<b>NIH</b>	National Institute of Health
<b>NKP</b>	Nationaler Krebsplan
<b>OECD</b>	Organisation for Economic Co-operation and Development
<b>PCO</b>	Prostate Cancer Outcomes
<b>PICOS</b>	Population, Intervention, Comparator, Outcome, Setting
<b>PORT</b>	Patient Outcomes Research Teams
<b>PRO</b>	patient-reported Outcome
<b>PROM</b>	patient-reported outcome measure/measurement
<b>PROMIS</b>	Patient-Reported Outcomes Measurement Information System
<b>QLQ</b>	quality of life questionnaire
<b>RPE</b>	Radikale Prostatektomie
<b>TNGR</b>	TrueNorth Global Registry

## 1. Zusammenfassung

Sowohl Outcome- als auch Patient\*innenorientierung bilden (gleichermaßen) die Grundpfeiler der Versorgungsforschung. Diese beiden wichtige Aspekte kommen zusammen bei der Anwendung und Nutzung von sog. patient-reported Outcomes; dies sind von Patient\*innen selbst (das heißt ohne Zwischeninterpretation von bspw. klinischem Personal) zumeist auf Basis standardisierter Messungen (wie Fragebogen) berichtete Outcomes. Vor allem für die Onkologie mit ihrer medizinischen Komplexität sowie der hohen Anzahl an Betroffenen können patient-reported Outcomes vielfältig eingesetzt werden. Ziel der vorliegenden Arbeit ist es deshalb, die Möglichkeiten der Nutzung von patient-reported Outcomes für onkologische Patient\*innen in bereits bestehende, für den deutschen Versorgungskontext relevante Versorgungsstrukturen (zertifizierte Krebszentren) aufzuzeigen und zu erforschen. Ein besonderer Fokus liegt dabei auf den folgenden drei Fragestellungen: 1. Wie lassen sich patient-reported Outcomes als relevante Outcomes für den Vergleich von Versorgungsorganisationen unter der Berücksichtigung des Patient\*innenkollektivs adjustieren?, 2. Wie unterscheiden sich Krebszentren bezüglich der patient-reported Outcomes ihrer behandelten Patient\*innen voneinander? 3. Wie ist das Wissen über und die Haltung zu patient-reported Outcomes von onkologisch Behandelnden in deutschen Krebszentren für die patient\*innenindividuelle sowie die Versorgungsorganisationsübergreifende Nutzung? Als theoretischer Rahmen dieser Arbeit dient dabei zum einen Donabedians Qualitätsbegriff (als Sinne des Struktur-Prozess-Outcome-Modells) und die daraus resultierenden Möglichkeiten der Qualitätssicherung und -entwicklung sowie die vielfach geforderte Patient\*innenzentrierung in der Versorgungsforschung (unter Bezugnahme auf das integrative Modell der Patient\*innenzentrierung von Scholl et al.). Zusätzlich kommt das Throughput-Modell nach Pfaff und Schrappe zur Anwendung.

Zur Beantwortung der formulierten Forschungsfragen kamen mehrere unterschiedliche qualitative sowie quantitative methodische Ansätze zur Anwendung. Zum einen wurde ein Systematic Review durchgeführt, um bereits publizierte Adjustierungsmöglichkeiten von patient-reported Outcomes zu identifizieren und zu analysieren. Dabei konnte festgestellt werden, dass es derzeit keinen einheitlichen Standard zur Adjustierung in der Literatur gibt. Nichtsdestotrotz konnten unterschiedliche weitere Forschungsfelder identifiziert und Vorschläge für statistische Adjustierungsverfahren formuliert werden. Darauf aufbauend wurde anhand von zwei Fallstudien (zum einen für das Prostatakarzinom im Rahmen der PCO- und zum anderen für das kolorektale Karzinom im Rahmen der EDIUM-Studie) die Ergebnisqualität in zertifizierten Krebszentren anhand von für das Patient\*innenkollektiv adjustierten patient-reported Outcomes verglichen. Dafür konnte auch das für das Prostatakarzinom genutzte Instrument (EPIC-26) zunächst in seiner deutschen Übersetzung psychometrisch validiert werden. Es zeigten sich Unterschiede der Ergebnisqualität zwischen

den Versorgungsorganisationen. Um diese zu quantifizieren, wird in dieser Arbeit eine neue Kennzahl vorgeschlagen, die die Rate derjenigen Versorgungsorganisationen, die adjustierte, mindestens eine minimally important difference vom Median abweichende patient-reported Outcomes vorweisen, erfasst.

Zusätzlich zu diesen quantitativen Analysen wurden qualitative Interviews mit klinischem Personal aus den zertifizierten Krebszentren durchgeführt und anschließend inhaltsanalytisch nach Kuckartz ausgewertet, um das Wissen über und die Haltung zu patient-reported Outcomes aus der Behandelndensicht besser untersuchen zu können. Hierbei konnte erfasst werden, dass Kliniker\*innen durchaus verschiedene potentielle Anwendungsmöglichkeiten von PROs identifizieren (z. B. Verlaufskontrolle, Verbesserung der Ärzt\*innen-Patient\*innen-Kommunikation), jedoch auch – vor allem technisch-organisatorische – Hürden sehen, die bei einer nachhaltigen Implementierung berücksichtigt werden müssen.

Neben den hier aufgezeigten Möglichkeiten für den Einsatz von PROs in klinischen und qualitätsentwickelnden Kontexten offenbart diese Arbeit auch einen großen Bedarf an weiterer (Versorgungs-) Forschung: Beispielsweise können das beschriebene Adjustierungsverfahren und die darauf basierenden Vergleiche von Versorgungsorganisationen auch auf andere Tumorentitäten übertragen werden, um die Auswirkungen der Rückspiegelung von patient-reported Outcome-Ergebnissen (z. B. in Form von jährlichen Qualitätsberichten) auf die Ergebnisqualität in den Zentren und ggf. die dafür eingeführten Maßnahmen zu untersuchen und um internationale Vergleiche der Gesundheitsversorgung auf der Grundlage von patient-reported Outcomes zu fördern. So werden aus der Messung und Rückspiegelung von patient-reported Outcomes aus reinen Outcomes auch Interventionen mit der Möglichkeit der Versorgungsverbesserung/-veränderung und damit Input-Faktoren 2. Ordnung nach dem Throughput-Modell nach Pfaff und Schrappe.

Diese Arbeit bietet dafür sowohl bezüglich methodischer Ansätze als auch auf struktureller Ebene mehrere Referenzpunkte und -möglichkeiten, wie und in welchem Rahmen patient-reported Outcomes für die Versorgung von Krebspatient\*innen sinnvoll nutzbar sind.

## 2. Einleitung

Die Versorgungsforschung hat nach Pfaff als Gegenstand die „letzte Meile“ der medizinischen Versorgung - also die wissenschaftliche Untersuchung „konkrete[r] Kranken- und Gesundheitsversorgung in den Krankenhäusern, Arztpraxen und sonstigen Gesundheitseinrichtungen, in deren Rahmen die Versorgungsleistungen gemeinsam mit dem Patienten erbracht werden“<sup>1</sup>. Für die Versorgung von Krebspatient\*innen bedeutet dies zum Beispiel die Untersuchung und Evaluation von onkologischen Versorgungsleistungen unter Alltagsbedingungen. Die hohe Public-Health-Relevanz der Onkologie ergibt sich dabei durch die besondere Komplexität des Fachgebiets und die hohe Anzahl von Betroffenen<sup>2</sup>: Die Diagnose Krebs ist sowohl für Betroffene als auch für Versorgungsorganisationen herausfordernd und oft mit weitreichenden Konsequenzen verbunden. Auf der einen Seite steht die potentielle Lebensbedrohung durch eine schwerwiegende, den menschlichen Körper auf vielfache Weise beeinträchtigende Erkrankung, auf der anderen Seite steht der Beginn eines komplexen Behandlungspfades, der viele verschiedene Professionen einschließt und damit das Gesundheitssystem auf multiple Weise herausfordert. In Deutschland gehören onkologische Erkrankungen zu einer der häufigsten Todesursachen: im Jahr 2020 waren bösartige Neubildungen mit 23,5 % die zweithäufigste Todesursache<sup>3</sup>. Aber nicht nur die besondere Mortalität von Krebserkrankungen (die sich durchaus je nach Tumorentität sehr stark unterscheidet) sticht hervor. Auch das Leben mit und nach einer Krebserkrankung unterscheidet sich häufig deutlich vom Leben vor der Diagnose. Durch medizinische Fortschritte und auch den soziodemographischen Wandel gibt es immer mehr *cancer survivors* (Überlebende nach einer Krebsdiagnose) in Deutschland: Ende des Jahres 2017 lebten insgesamt 4,65 Millionen Menschen mit oder nach Krebs – das sind etwas mehr als 5 % der gesamtdeutschen Bevölkerung<sup>4</sup>. Damit rücken neben der ersten – im besten Falle kurativen – Behandlung und Beseitigung der Krebserkrankung auch (Spät-) Folgen der Erkrankungen sowie behandlungsassoziierte Lebenseinschränkungen in den Mittelpunkt einer versorgungswissenschaftlichen Perspektive.

Eckpfeiler der Versorgungsforschung sind die Outcome- und Patient\*innenzentrierung<sup>5</sup>, woraus auch für die Onkologie die Notwendigkeit von Messinstrumenten, die auf der einen Seite die Perspektive der Krebsbetroffenen widerspiegeln und auf der anderen Seite verlässliche, quantifizierbare Aussagen über die Qualität der onkologischen Versorgung ermöglichen, ergibt<sup>2</sup>. Besondere Beachtung haben seit mehreren Jahren die sog. *patient-reported Outcomes* als Ergebnisqualitätsparameter erhalten und deren Nutzung in der Onkologie wird national wie international vielfach gefordert<sup>6,7</sup>.

Diese Nutzungsmöglichkeiten von *patient-reported Outcomes* sowohl auf der Patient\*innenebene als auch zur Evaluation der Qualität onkologisch behandelnder

Versorgungsorganisationen und die daraus resultierenden methodischen Herausforderungen sind Gegenstand dieser Arbeit.

Im Folgenden wird zunächst der Begriff des *Outcomes* und dessen zentrale Bedeutung für die Versorgungsforschung hergeleitet und in Verbindung gesetzt zum Prinzip der Patient\*innenorientierung (2.1). Darauf aufbauend werden *patient-reported Outcomes* als Patient\*innen-relevante Instrumente vorgestellt sowie Möglichkeiten, aber auch Herausforderungen der Nutzung von *patient-reported Outcomes* aufgezeigt, wobei exemplarisch bereits auf den onkologischen Versorgungskontext Bezug genommen wird (2.2). Dieser komplexe Kontext mit seinen verschiedenen Versorgungsebenen wird anschließend eingehender beleuchtet, mit einem besonderen Fokus auf die von der Deutschen Krebsgesellschaft (DKG) zertifizierten Krebszentren als spezialisierte Versorgungsnetzwerke (2.3). Es werden zwei Studien vorgestellt, die als Fallbeispiele dieser Arbeit dienen (2.4 und 2.5), anhand derer die Möglichkeiten der Nutzung von *patient-reported Outcomes* für onkologische Patient\*innen aufgezeigt werden sollen (siehe Forschungsfragen und Zielsetzung der Arbeit in 2.6).

## **2.1. Outcome- und Patient\*innenorientierung in der Versorgungsforschung**

Die Fokussierung auf *Outcomes* ist ein zentraler Bestandteil der Versorgungsforschung, sehen beispielsweise Pfaff et al. in ihm die „zentrale abhängige Variable der Versorgungsforschung“<sup>5</sup>. So ist ein großer Anteil der (angloamerikanischen) Versorgungsforschung aus der *Outcomes Research* hervorgegangen<sup>1</sup>. Bereits der unter anderem in Harvard praktizierende Chirurg Codman (1869-1940) postulierte das Konzept der *end results* und kämpfte Zeit seines Lebens – zumeist erfolglos – für dessen Implementierung in das amerikanische Gesundheitssystem<sup>8</sup>. Codmans Idee basiert darauf, dass medizinische Fälle systematisch und strukturiert dokumentiert werden sollten von ihren Behandelnden, um diese Dokumentationen (sog. *end result cards*) anschließend zur Qualitätsüberprüfung und ggfs. -verbesserung zu nutzen. Er fordert dabei bereits 1914, dass Behandelnde (er bezog sich dabei vor allem auf Krankenhäuser) Berichte erstellen sollten „showing as nearly as possible what are the results of the treatment obtained at different institutions. This report must be made out and published by each hospital in a uniform manner, so that comparison will be possible“<sup>9</sup>. Als *end result* versteht Codman vor allem “the satisfied and relieved patient” (ebd.).

Donabedian, der das Konzept der *quality in health care* bis heute zentral prägt, hat seit den 1960ern die Idee der *end results* aufgegriffen und weitergeführt. Als Pionier der

---

<sup>1</sup> Zum Teil werden bzw. wurden die Begriffe *Outcomes Research* und *Health Services Research* sogar synonym verwendet und zahlreiche Autor\*innen machen auf die Schwierigkeit, diese Begriffe voneinander abzugrenzen, aufmerksam<sup>5,7</sup>.

Qualitätsforschung in der Medizin zeigt er zunächst, dass Qualität medizinischer Versorgung durchaus quantifizierbar und damit messbar ist. Qualität fasst er dabei zunächst auf als „the product of two factors. One is the science and technology of health care, and the second is the application of that science and technology in actual practice“<sup>10</sup>. Um die Qualität medizinischer Versorgung zu untersuchen und vor allem zu quantifizieren, schlägt Donabedian die Triade aus Struktur, Prozess und *Outcome* vor und definiert letzteres dabei als „mean [change] (desirable or undesirable) in individuals and populations that can be attributed to health care“ mit der weiteren Kategorisierung in „1. changes in health status, 2. changes in knowledge acquired by patients and family members that may influence future care, 3. changes in the behaviour of patients or family members that may influence future health, 4. satisfaction of patients and their family members with the care received and its outcomes“<sup>11</sup>. Das Struktur-Prozess-Outcome-Modell wird im deutschsprachigen Raum teilweise verwendet zur Definition des Konzepts der Ergebnisqualität, wobei *Outcome* also nur einen Teilbereich dessen umfasst. Der Begriff des *end results* taucht auch nach der Qualitätsdefinition von Donabedian weiterhin auf und bleibt zentral für die *Outcomes Research*, die 1988 von Clancy und Eisenberg wie folgt definiert wird: „the study of the end results of health services that take patients' experiences, preferences and values into account“<sup>5</sup>.

In Amerika – und nachfolgend auch in Deutschland – wird *Outcomes Research* besonders intensiv seit den 1990er betrieben, vor allem gefördert durch die Gründung des Medical Treatment Effectiveness Program (MEDTEP), durch das mehrere Patient Outcomes Research Teams (PORTs) initiiert wurden<sup>7,12</sup>. In der ersten Förderungsphase des MEDTEPs wurde neben den Indikationen Herzinfarkt oder Schmerzen des unteren Rückens mit dem Prostatakarzinom auch bereits eine onkologische Erkrankung in den Blick genommen. Ziel der PORTs war dabei unter anderem die „analysis of variations in medical practice associated with patient outcomes“ (zitiert ebd.). Auch hier wird – wie bereits bei Clancy und Eisenberg - der Begriff des *Outcomes* mit dem\*der Patient\*in explizit verknüpft. Dies unterstreicht die Patient\*innenorientierung der Versorgungsforschung, die neben der Outcomeorientierung von Pfaff et al. als zentral und disziplinedefinierend genannt wird<sup>5</sup>. Ebenda wird deshalb wiederholt betont, dass durch *Outcomes* die „Wirksamkeit unter Alltagsbedingungen“ primär aus Perspektive der Patienten beurteilt werden soll“<sup>5</sup>.

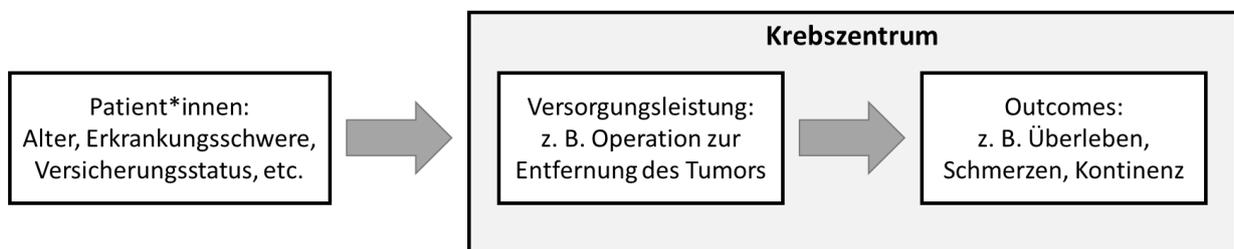
## **2.2. Outcomes aus Sicht der Betroffenen – Nutzung von *patient-reported Outcomes* in der Onkologie auf verschiedenen Analyseebenen**

Ein häufig genutztes Instrument, um *Outcomes* aus der Patient\*innenperspektive zu erheben, ist das der *patient-reported Outcomes* (PROs). PROs werden von der Organisation for Economic Co-operation and Development (OECD) wie folgt durch ihre Funktion definiert:

„[PROMs] measure patients' perceptions of their health status, clinical outcomes, mobility and quality of life“<sup>13</sup>. Zumeist handelt es sich dabei um standardisierte Fragebogen, die eine Quantifizierung der Ergebnisse ermöglichen, wodurch PROs für die quantitative Forschung sowie die darauf aufbauende Nutzung z. B. in der Qualitätssicherung besonders wertvoll sind (ebd.). PROs werden grob in generische (z. B. SF-36<sup>14</sup>) sowie krankheitsspezifische Instrumente (z. B. die EORTC QLQ-Fragebogen für Krebsbetroffene<sup>15</sup>) eingeteilt und können auf vielfältige Weise in der Gesundheitsversorgung eingesetzt werden. Für onkologische Versorgungsstrukturen konnte in Studien gezeigt werden, dass Patient\*innen von einer regelhaften PRO-Befragung profitieren, da Behandelnde Symptome weniger unterschätzen<sup>16,17</sup>, Behandlungen zielgerichteter geplant werden können<sup>18</sup> und sogar die Überlebenszeit verlängert werden kann<sup>19,20</sup>. Um den besonderen Nutzen von PROs vor allem für eine verbesserte Patient\*innenorientierung aufzuzeigen, kann das integrative Modell der Patient\*innenorientierung von Scholl et al. herangezogen werden, das unter den drei Hauptkategorien *principals*, *enablers* sowie *activities* unter anderem die folgenden, durch PROs adressierten, für eine Patient\*innenorientierung relevanten Dimensionen nennt: *patient as unique person*, *patient involvement in care* und *patient empowerment*<sup>21</sup>. Durch die Integration von PROs in die klinische Versorgung wird also der Fokus auf die Patient\*innen gesetzt und gleichzeitig eine standardisierte Outcomeerfassung ermöglicht.

Die standardisierte Erfassung von PROs wird seit Ende der 1990er Jahren von einigen Initiativen aktiv vorangetrieben, vor allem im angloamerikanischen Raum. So wurde 2004 vom National Institute of Health (NIH, USA) das *Patient-Reported Outcomes Measurement Information System* (PROMIS) gegründet, das als Zielsetzung folgendes angibt: „to develop an efficient state-of-the-art assessment system for self-reported health“<sup>22</sup>. Auch im Vereinigten Königreich wurden vom National Health Service (NHS) mehrere Projekte initiiert, die die flächendeckende PRO-Erhebung vorantreiben sollten. So wurden für vier elektive chirurgische Eingriffe (Knie- und Hüftersatz, Hernienchirurgie, Kataraktoperation) landesweite Programme aufgesetzt, durch die der Vergleich der Ergebnisqualität einzelner NHS-Versorgungsorganisationen ermöglicht werden sollte<sup>23,24</sup>. Außer solchen staatlichen Initiativen existieren zahlreiche weitere Netzwerke, die sich stark für die Einbettung von PROs im Gesundheitswesen einsetzen. So sind verschiedene PRO-Instrumente beispielsweise fester Bestandteil der *standard sets*, die vom International Consortium for Health Outcomes Measurement (ICHOM) für bestimmte Krankheiten (z. B. Diabetes, Schlaganfall, aber auch mehrere onkologische Erkrankungen) zusammengestellt werden und international Anwendung finden<sup>25</sup>. Auch die European Organization for Research and Treatment of Cancer (EORTC) treibt die Nutzung von PROs aktiv voran, indem sie beispielsweise eigene PRO-Instrumente entwickelt und zur Verfügung stellt<sup>15</sup>.

Neben der patient\*innenindividuellen Nutzung bieten sich PROs vor allem auch an, um die Versorgungsqualität unterschiedlicher Versorgungsorganisationen zu vergleichen. Die Qualität einer (onkologischen) Behandlung wird meistens entweder anhand von Überlebensraten oder anhand des Auftretens unerwünschter Komplikationen als „traditionelle“ Outcomes evaluiert<sup>26</sup>. Zusätzlich ist es für potentielle Patient\*innen, aber auch für Einweisende, Kostenträger, etc. von Bedeutung, wie eine Behandlung einer speziellen Versorgungsorganisation sich aus Sicht der Patient\*innen auswirkt. Vor allem bei Tumorentitäten mit einer relativ hohen Überlebenswahrscheinlichkeit (z. B. beim Prostatakarzinom: relative 5-Jahres-Überlebensrate von 89% im Jahr 2017<sup>27</sup>) können PRO-basierte Versorgungsorganisations-Vergleiche dabei helfen, ein versorgungsnäheres Bild der Behandlungsqualität zu geben, wenn diese Vergleiche für relevante Patient\*innenfaktoren adjustiert werden (sog. *Casemix*-Adjustierung). Hierbei stellt sich jedoch die methodische Frage, wie PROs im Sinne eines angemessenen Vergleichs, der mögliche verzerrende Patient\*innenmerkmale berücksichtigt, am besten für den *Casemix* adjustiert werden sollten. Dieses Problem wird schon von Donabedian erwähnt: „Even when a large number of cases has been observed, there is a further difficulty to overcome. This is because patients vary in their medical, social, psychological, and genetic characteristics; features that can, in themselves, influence outcomes either independently of process or by interacting with process. This means that before some outcomes can stand for quality, one must correct for differences among patients in such characteristics, extraneous to the process of care—a procedure known as "case-mix adjustment"<sup>10</sup>. Eine *Casemix*-Adjustierung ist also unerlässlich, da Vergleiche von Versorgungsorganisationen, die auf Ergebnissen des jeweiligen Patient\*innenkollektivs (engl. casemix) beruhen, auch von nicht von den Behandelnden beeinflussbaren Patient\*innenmerkmalen (bspw. Alter, soziodemographische Merkmale, Krankheitsstadium vor Beginn einer Therapie) verzerrt werden können (vgl. Abbildung 1). Um diese Effekte der Patient\*innen beim Vergleich der Versorgungsorganisationen zu berücksichtigen, müssen die Ergebnisse also zunächst für den *Casemix* adjustiert werden<sup>28</sup>.



**Abbildung 1:** Beispiellustration für die Notwendigkeit einer *Casemix*-Adjustierung: Das Krebszentrum kann die tatsächlich erbrachte Versorgungsleistung (und assoziierte Leistungen) und daraus resultierende Outcomes beeinflussen, jedoch nicht (oder nur teilweise) die Merkmale der Patient\*innen (nach Iezzoni: „intrinsic patient characteristics“<sup>28</sup> genannt), die behandelt werden müssen. Deshalb müssen Outcomes auf Krebszentrumsebene für diese Merkmale adjustiert werden, wenn sie für einen Vergleich von Versorgungsorganisationen herangezogen werden sollen.

PROs lassen sich also nicht nur für individuelle Patient\*innen nutzen, sondern auch auf einer übergeordneten Versorgungsorganisationsebene. Trotz dieser beiden wichtigen, die onkologische Versorgung potentiell verbessernden Anwendungsbereiche, werden PROs in der onkologischen Routineversorgung noch kaum genutzt. Zugleich gibt es noch erheblichen Forschungsbedarf im Zusammenhang mit der Erhebung von PROs, darunter die sinnvolle Einbindung von Behandelnden sowie des methodischen Vorgehens zum umfassenden Versorgungsorganisationsvergleich, da sich hierbei noch kein bestimmtes Verfahren durchsetzen konnte.

### **2.3. PROs im Kontext der onkologischen Versorgung in Deutschland**

Versorgungsforschung bezieht sich immer auf einen konkreten Versorgungskontext, der zumeist als „komplexer Kontext“ aufgefasst wird<sup>5</sup>. Diese Arbeit fokussiert sich deswegen in ihrer Analyse von PROs auf den speziellen Versorgungskontext der zertifizierten Krebszentren in Deutschland. Zur besseren Einordnung dieser wird nun zunächst die onkologische Versorgungslandschaft in Deutschland grob umrissen, um dann den Kontext der Krebszentren näher zu beleuchten. Die Versorgung von Krebspatient\*innen in Deutschland schließt viele verschiedene Versorgungsorganisationen und -leistungen ein. Deshalb kann dieser Abschnitt nur einen groben Überblick über maßgebliche, für diese Arbeit relevante Akteur\*innen und Strukturen geben.

Die Grundpfeiler der aktuellen Krebsversorgung in Deutschland beruhen auf dem Nationalen Krebsplan (NKP) von 2008<sup>29</sup>, der vom Bundesministerium für Gesundheit, der DKG, der Deutschen Krebshilfe sowie der Arbeitsgemeinschaft Deutscher Tumorzentren initiiert wurde. In diesem Programm wurden vier Haupthandlungsfelder identifiziert:

1. Weiterentwicklung der Krebsfrüherkennung
2. Weiterentwicklung der onkologischen Versorgungsstrukturen und der Qualitätssicherung
3. Sicherstellung einer effizienten onkologischen Behandlung
4. Stärkung der Patient\*innenorientierung

Zu den Leistungserbringern von onkologischen Behandlungen im Sinne des Handlungsfelds 3 gehören neben niedergelassenen Onkolog\*innen und Strahlentherapeut\*innen, entitätsspezifischen Fachärzt\*innen (z. B. Gynäkolog\*innen, Urolog\*innen, Gastroenterolog\*innen) der akut-stationäre sowie der Rehabilitationssektor. Die akute Behandlung einer Krebserkrankung erfolgt dabei meist primär stationär – z. B. im Rahmen einer operativen Entfernung des Karzinoms – und wird anschließend durch den ambulanten Sektor weitergeführt (z. B. anschließende Systemtherapie, Nachsorge). Während und nach der Krebsbehandlung sollen die Patient\*innen Zugang zu psychoonkologischer Betreuung

haben (Ziel 9 des NKPs). Die besondere Bedeutung der „sektorenübergreifenden, integrierten onkologischen Versorgung“ wird durch das Ziel 7 des NKPs unterstrichen, wodurch der hohe Stellenwert der Krebszentren (vgl. 2.3.1) in den Fokus genommen wird.

### **2.3.1. Qualitätssicherung in der Onkologie – das Zertifizierungsverfahren der DKG**

Ein wichtiger Baustein in der Versorgung von Krebspatient\*innen stellen Versorgungsnetzwerke dar, die sog. Krebszentren, die die gesamte Versorgungskette von Krebsbetroffenen abdecken sollen. Krebszentren gehen auf das von der DKG 2003 initiierte Zertifizierungssystem für onkologische Leistungserbringende zurück<sup>30</sup>. Ziel des Zertifizierungssystems ist die Schaffung von strukturierten Versorgungsnetzwerken, um die Evidenz-basierte, (bedarfs-)gerechte Versorgung von Krebsbetroffenen in Deutschland zu gewährleisten. Die Notwendigkeit und Förderung solcher Versorgungsstrukturen findet sich auch im Nationalen Krebsplan (Handlungsfeld 2, Ziel 5) wieder: „Es existieren einheitliche Konzepte und Bezeichnungen für die Qualitätssicherung und Qualitätsförderung und Zertifizierung onkologischer Behandlungseinrichtungen“<sup>29</sup>. Im Jahr 2019 wurden ca. 55 % der neu-diagnostizierten Krebspatient\*innen in Deutschland in einem zertifizierten Krebszentrum behandelt (Berechnungen auf Grundlage der Krebsstatistik des Robert-Koch-Instituts und Daten der DKG).

Um die Ergebnisqualität (im Sinne Donabedians Struktur-Prozess-Outcomes-Modell, vgl. auch Abschnitt 2.1) in den an der Zertifizierung teilnehmenden Versorgungsnetzwerken sicherzustellen, werden die Krebszentren jährlich vor Ort auditiert. Grundlage der Audits sind die von den entitätsspezifischen Zertifizierungskommissionen (die zusammengesetzt sind aus Mandatsträger\*innen aus den betroffenen Fachgesellschaften) festgelegten Zertifizierungsanforderungen, die beispielsweise Mindestmengen (z. B. > 99 Radikale Prostatektomien pro Jahr<sup>31</sup>) oder aus den S3-Leitlinien abgeleitete Qualitätsindikatoren (z. B. Anastomosensuffizienzrate  $\leq 6\%$  für das Kolonkarzinom<sup>32</sup>) beinhalten. Dafür muss eine umfassende Dokumentation erfolgen, die neben Zentrumsprozessen beispielsweise auch patient\*innenindividuelle klinische Parameter umfasst. Durch diese systematische Erfassung der Kennzahlen werden wichtige *Outcomes* quantifizierbar und damit auch vergleichbar. Zusätzlich stehen vor allem für die Entitäten Prostata, Brust und Darm große, patient\*innenindividuelle Datensätze in einem einheitlichen Datenformat – auch für die Versorgungsforschung – zur Verfügung, die wichtige Parameter der onkologischen Versorgung in Krebszentren abdecken<sup>33</sup>. Aus den Möglichkeiten der Nutzung dieser Datensätze und der durch die Zertifizierung geschaffenen, Versorgungsnetzwerke übergreifenden Strukturen ging die vom Bundesverband Prostatakrebs Selbsthilfe initiierte

PRO-Studie PCO (vgl. 2.4) hervor. Das resultierende Studienkonzept wurde dann anschließend auch in Darmkrebszentren umgesetzt (vgl. 2.5). Hieraus gehen die beiden nun vorgestellten Fallbeispiele hervor.

#### **2.4. Fallbeispiel 1: Ergebnisqualität in zertifizierten Prostatakrebszentren – die PCO-Studie**

Seit 2016 werden in teilnehmenden Prostatakrebszentren neben der Erhebung von Routine- und zertifizierungsrelevanten Daten auch PROs zum funktionellen *Outcome* nach Prostatakrebsbehandlung erhoben. Die dafür aufgesetzte „Prostate Cancer Outcomes“ Studie (PCO) wurde aktiv durch die Selbsthilfegruppe Bundesverband Prostatakrebs Selbsthilfe (BPS) initiiert und ist Teil des TrueNorth Global Registries (TNGR)<sup>34</sup>. TNGR ist ein Zusammenschluss von weltweiten Studienzentren, die sich zum Ziel gesetzt haben, die Ergebnisqualität der Prostatakrebsbehandlung nachhaltig zu verbessern durch systematische Messung und Vergleich der Ergebnisse<sup>35</sup>. Zurzeit nehmen 14 Länder an dem von der Organisation Movember Foundation geleiteten Vorhaben teil. In Deutschland, Österreich und der Schweiz konnten bis zum Jahr 2022 mehr als 130 Prostatakrebszentren mit ca. 45.000 behandelten Prostatakrebspatienten an der Teilnahme zur Studie motiviert werden. Deutschland ist das mit Abstand bestrekrutierende Land im TNGR.

Hauptfragestellung der PCO-Studie ist: „Zeigen sich Unterschiede in der Ergebnisqualität nach lokal begrenztem Prostatakarzinom zwischen DKG-zertifizierten Zentren, zwischen diesen und internationalen Institutionen und falls ja, lassen sich Struktur- und Prozessmerkmale identifizieren, die diese Unterschiede erklären?“<sup>36</sup>.

Für die PCO-Studie werden teilnehmende Patienten gebeten, vor und ein Jahr nach Beginn einer kurativen Behandlung<sup>2</sup> ihres Prostatakarzinoms Fragen zu Prostatakrebs-spezifischen Funktionseinschränkungen und Symptomen sowie allgemeine Fragen zur Soziodemographie (Versicherungsstatus, Staatsbürgerschaft, höchster Bildungsabschluss) zu beantworten. Als PRO-Instrument kommt hier der international etablierte Fragebogen EPIC-26 zur Anwendung (erste validierte Fassung von 2010). Dieser erfasst in 26 Items fünf Domänen: „Inkontinenz“, „irritative/obstruktive Symptomatik“, „gastrointestinale Symptomatik“, „Sexualität“ und „Vitalität/hormonelle Symptomatik“<sup>37–39</sup>. Die Angaben der Patienten zu diesen fünf Domänen werden dann zusammen mit ihren soziodemographischen Angaben verknüpft mit im Rahmen des Zertifizierungsprozess routinemäßig erhobenen Daten. Dadurch entsteht ein Datensatz mit klinischen (z. B. Tumorstatus, Komorbiditäten, Auftreten von Komplikationen), soziodemographischen (Versicherungsstatus, Staatsbürgerschaft, Bildungsabschluss),

---

<sup>2</sup> Neben der operativen Therapieoption und auch der strahlentherapeutischen Behandlung können auch Patienten mit Active Surveillance oder Watchful Waiting als Behandlungsstrategie eingeschlossen werden.

prozeduralen (z. B. erhaltene Behandlungen, Inanspruchnahme von Psychoonkologie) und PRO-Angaben zu jedem Patienten.

Die hier vorgestellte Arbeit konzentriert sich auf operativ behandelte Patienten mit einem lokalen oder lokal fortgeschrittenen Prostatakarzinom: die zur Anwendung kommende Versorgungsleistung ist dabei die sog. Radikale Prostatektomie (RPE). Dabei wird die Prostata und nebenstehendes Gewebe vollständig („radikal“) entfernt mit dem kurativen Ziel der Entfernung jeglichen malignen Gewebes. Auch wenn es nach der Behandlung meistens zur vollständigen Remission kommt, ist die RPE mit teils erheblichen funktionellen Nebenwirkungen verbunden. Vor allem sind dabei die Kontinenz und Potenz der Patienten betroffen<sup>40</sup>. Die Erhebung von Funktionseinschränkungen und Symptomen aus Sicht der Patienten ist also besonders relevant sowohl für die Betroffenen selbst als auch für behandelnde Versorgungsorganisationen und Kostenträger.

Die PCO-Studie ist im DRKS unter der Nummer DRKS00010774 registriert; es liegt unter anderem ein positives Ethikvotum der Ärztekammer Berlin vor (Eth-12/16).

## **2.5. Fallbeispiel 2: Ergebnisqualität in zertifizierten Darmkrebszentren – die EDIUM-Studie**

Die EDIUM-Studie (Ergebnisqualität bei Darmkrebs: Identifikation von Unterschieden und Maßnahmen zur flächendeckenden Qualitätsentwicklung) verfolgt ein ähnliches Ziel wie die PCO-Studie, wobei es hierbei nun um die Versorgung von Darmkrebspatient\*innen geht. Die teilnehmenden Patient\*innen beantworten vor und anschließend ein Jahr nach Beginn ihrer Behandlung (operativ-kurativ oder palliativ-systemisch) die Fragebogen EORTC QLQ-C30 und -CR29 – ebenfalls international etablierte PRO-Instrumente zur Erfassung von (darm-) krebsspezifischen Symptomen und Funktionseinschränkungen (bspw. Lebensqualität, physische Funktion, Bauchschmerzen, Haarausfall<sup>41,42</sup>)<sup>3</sup>. Auch hier werden im Baseline-Fragebogen soziodemographische Angaben von den Patient\*innen gemacht (Versicherungsstatus, höchster Bildungsabschluss, Staatsbürgerschaft).

Neben dem eigentlichen Vergleich der Ergebnisqualität wird im Rahmen der EDIUM-Studie auch untersucht, welche fördernden und hemmenden Faktoren auf Versorgungsorganisationsebene vorliegen, um PROs flächendeckend in der onkologischen Routineversorgung in Darmkrebszentren zu implementieren. Die Synthese dieser beiden

---

<sup>3</sup> Eine Beispieldimension für den EORTC QLQ-C30 mit den dazugehörigen Items ist die Domäne „physische Funktion“ (Skalierung 0 – 100, wobei 100 der bestmögliche Wert ist) mit den folgenden Fragen: 1. Bereitet es Ihnen Schwierigkeiten, sich körperlich anzustrengen (z. B. eine schwere Einkaufstasche oder einen Koffer zu tragen)? 2. Bereitet es Ihnen Schwierigkeiten, einen längeren Spaziergang zu machen? 3. Bereitet es Ihnen Schwierigkeiten, eine kurze Strecke außer Haus zu gehen? 4. Müssen Sie tagsüber im Bett liegen oder in einem Sessel sitzen? 5. Brauchen Sie Hilfe beim Essen, Anziehen, Waschen oder Benutzen der Toilette? Antwortmöglichkeiten: Überhaupt nicht – Wenig – Mäßig – Sehr

Hauptziele (der Vergleich der Ergebnisqualität und die Evaluation der Implementierung) soll eine – wenn nötige – Veränderung der Versorgungsrealität ermöglichen.

Die EDIUM-Studie wurden von 2018-2021 vom Innovationsausschuss des Gemeinsamen Bundesausschusses als Versorgungsforschungsprojekt gefördert (Fördernummer: 01VSF17040) und wird seit Anfang 2022 aus Eigenmitteln der Projektbeteiligten (DKG, OnkoZert, Universität Köln, addz, ILCO, teilnehmenden Darmkrebszentren) weitergeführt. Sie ist im DRKS unter der Nummer DRKS00008724 registriert mit einem positiven Ethikvotum der Ärztekammer Berlin (Eth-19/18)<sup>43</sup>.

## **2.6. Zielsetzung dieser Arbeit: Anwendungsmöglichkeiten von PROs auf verschiedenen Ebenen der Versorgung sowie hemmende und fördernde Faktoren für eine erfolgreiche Implementierung**

Diese Arbeit setzt sich mit den verschiedenen Ebenen der Nutzung von PROs für onkologische Patient\*innen auseinander. Dabei ist als konkreter Versorgungskontext die Struktur der zertifizierten Krebszentren (im Speziellen: Prostata- und Darmkrebszentren) gewählt. Dadurch sind die Ergebnisse auf eine große Anzahl an Krebspatient\*innen anwendbar und durch die bereits bestehenden Strukturen können auch Änderungen der Versorgung schneller und flächendeckender erreicht und (langfristig) kontinuierlich evaluiert werden.

Um Möglichkeiten der Nutzung von PROs für onkologische Patient\*innen in bereits bestehenden Versorgungsstrukturen (zertifizierten Krebszentren) aufzuzeigen und zu erforschen, werden in dieser Arbeit die folgenden Forschungsfragen adressiert:

1. Wie lassen sich PROs als relevante *Outcomes* für den Vergleich von Versorgungsorganisationen unter der Berücksichtigung des Patient\*innenkollektivs adjustieren?

Hypothese: keine – explorative Fragestellung

2. Wie unterscheiden sich Krebszentren bezüglich der *patient-reported Outcomes* ihrer behandelten Patient\*innen voneinander?

Hypothese: Die durch PROs gemessene Ergebnisqualität unterscheidet sich zum Teil erheblich zwischen Krebszentren unter angemessener Berücksichtigung der Patient\*innenkollektive.

3. Wie ist das Wissen über und die Haltung zu PROs von onkologisch Behandelnden in deutschen Krebszentren für die patient\*innenindividuelle sowie die Versorgungsorganisations-übergreifende Nutzung?

Hypothese: keine – explorative Fragestellung

Dabei stehen also mehrere Ebenen der PRO-Nutzung im Fokus; dies resultiert daraus, dass PROs bereits so angelegt sind, dass sie – wurden sie einmal erfasst – auf vielfältige Weise

genutzt werden können. Als theoretischer Rahmen dient hierbei zum einen Donabedians Qualitätsbegriff und die daraus resultierenden Möglichkeiten der Qualitätssicherung und -entwicklung (vgl. Abschnitt 2.1) sowie die vielfach geforderte Patient\*innenzentrierung in der Versorgungsforschung<sup>2,21</sup>. Auch wenn der Fokus dieser Arbeit vor allem auf der Nutzung von PROs auf Versorgungsorganisationsebene liegt – da vor allem hierfür noch erheblicher Forschungsbedarf besteht – soll durch die dritte Forschungsfrage auch die Patient\*innenebene adressiert werden. Dadurch sollen die Möglichkeiten der Mehrfachnutzung von PROs in der Onkologie aufgezeigt werden, da dies vor dem Hintergrund des stetig steigenden Dokumentationsaufwands für Behandelnde in der Onkologie für den Routineeinsatz höchst relevant ist<sup>44,45</sup>: ein möglichst datensparender Ansatz für die Erfassung von relevanten *Outcomes* für verschiedene Versorgungsebenen kann die Akzeptanz aller beteiligten Akteur\*innen erhöhen und dadurch zu einer nachhaltigen und sinnvollen PRO-Implementierung beitragen. Diese Arbeit soll somit neben der Untersuchung von Nutzungsmöglichkeiten von PROs für die onkologische Versorgung auf verschiedenen Ebenen und den daraus resultierenden methodischen Herausforderungen auch Wege aufzeigen, in welchem Rahmen solch eine nachhaltige PRO-Implementierung möglich und sinnvoll ist.

### 3. Publikationen

Für die Zielsetzung dieser Arbeit mit ihren unterschiedliche Ebenen adressierenden Fragestellungen (vgl. 2.6) waren methodische Ansätze, die sowohl qualitative inhaltsanalytische, wissenssynthetisierende wie auch quantitativen Methoden beinhalten, erforderlich. So bauen die quantitativen Datenanalysen zum Versorgungsorganisationsvergleich (Arbeit 3 und 4) auf einem Systematic Review zu statistischen Adjustierungsmethoden (Arbeit 2) sowie der zuvor erfolgten psychometrischen Validierung des Messinstrumentes (Arbeit 1) auf. Zusätzlich kamen qualitative Methoden zum Einsatz, um mehr Einblicke in das Verständnis über und der Einstellung zu PROs von Behandelnden in Krebszentren zu erhalten (Arbeit 5). Es folgen Zusammenfassungen dieser fünf in peer-reviewten Zeitschriften veröffentlichten Arbeiten<sup>4</sup>.

#### 3.1. Arbeit 1: Validierung des Messinstrumentes - Voraussetzung für eine sinnvolle Integration von PROs in die Versorgung

Für die Nutzung von PROs in der onkologischen Versorgung müssen zunächst Messinstrumente zur Verfügung stehen, die die erforderlichen Konstrukte (generisch und auch krebspezifisch) zuverlässig messen. Dafür wird in der ersten Arbeit ein bereits international etablierter PRO-Fragebogen für das Prostatakarzinom (EPIC-26, „Expanded Prostate Cancer Composite 26-items Short Form“<sup>37</sup>) in seiner deutschen Übersetzung<sup>39</sup> psychometrisch untersucht und validiert<sup>5</sup>. Der EPIC-26 ist ein international etablierter PRO-Fragebogen und wird bspw. vom ICHOM für das lokal begrenzte Prostatakarzinom empfohlen<sup>46</sup>. Gemessen werden mithilfe von 26 Items fünf für das Prostatakarzinom spezifische Symptome und Funktionseinschränkungen: „Inkontinenz“, „irritative/obstruktive Symptomatik“, „gastrointestinale Symptomatik“, „Sexualität“ und „Vitalität/hormonelle Symptomatik“. Für die psychometrische Validierung wurden folgende statistische Methoden verwendet (unter anderem nach Cronbach<sup>47</sup>): Konstruktvalidität (Faktorenanalyse), interne Konsistenz

---

<sup>4</sup> Bis auf die Arbeit 4 (Funktionelles Outcome nach elektiver Tumoresektion in Darmkrebszentren), die auf Deutsch im Deutschen Ärzteblatt veröffentlicht wurde, wurden alle Arbeiten in internationalen Zeitschriften veröffentlicht. Die Originalmanuskripte sind ab Kapitel 3 zu finden (zum Teil ohne Appendices).

<sup>5</sup> Als Beispiel seien hier die zur Domäne Inkontinenz (Skalierung 0 – 100, mit 100 als bestmöglichen Wert) gehörigen Items und die dazugehörigen Fragen aufgeführt: Urinverlust (Wie häufig haben Sie in den letzten 4 Wochen Urin verloren? Antwortmöglichkeiten: öfter als einmal am Tag – etwa einmal am Tag – öfter als einmal die Woche – etwa einmal die Woche – selten oder nie), Kontrolle Harnentleerung (Welche der folgenden Aussagen beschreibt am besten, wie Ihre Kontrolle über die Harnentleerung in den letzten 4 Wochen war? Antwortmöglichkeiten: Keinerlei Kontrolle über die Harnentleerung - Häufiges Tröpfeln - Gelegentliches Tröpfeln - Vollständige Kontrolle), Gebrauch Einlagen (Wie viele Einlagen oder Inkontinenzvorlagen haben Sie normalerweise in den letzten 4 Wochen am Tag gebraucht, um den Urinverlust unter Kontrolle zu haben? Antwortmöglichkeiten: Keine – Eine Einlage am Tag – Zwei Einlagen am Tag – Drei oder mehr Einlagen am Tag), Probleme durch Tröpfeln/Urinverlust (Wie sehr hat Ihnen Tröpfeln oder Urinverlust in den letzten 4 Wochen Probleme bereitet? Antwortmöglichkeiten: Kein Problem – sehr kleines Problem – kleines Problem – mäßiges Problem – großes Problem)

(Cronbach's Alpha) und Diskriminanzvalidität (Unterscheidbarkeit prä- und posttherapeutische Ergebnisse). Es konnte gezeigt werden, dass die deutsche Übersetzung des EPIC-26 zufriedenstellende psychometrische Eigenschaften aufweist, die mit dem Originalfragebogen vergleichbar sind.

### **3.2. Arbeit 2: Systematic Review zu Methoden der Casemix-Adjustierung bei PROs – Übersicht über genutzte Methoden zum Vergleich von Versorgungsorganisationen**

Zur Beantwortung der Frage: „Welche statistischen Methoden werden für die Casemix-Adjustierung von PROs für den Vergleich von Versorgungsorganisationen genutzt?“ wurde eine systematische Literaturrecherche (in MEDLine durchgeführt unter Berücksichtigung des PICOS-Prinzips (*Population, Intervention, Comparator, Outcome, Setting* wie beispielweise von der Cochrane Collaboration vertreten<sup>48</sup>) und der qualitativen Evaluierung mithilfe der Critical Appraisal Skills Programme (CASP) Checkliste<sup>49</sup>. Aus 640 gescreenten Studien konnten elf in die Analyse eingeschlossen werden. Es konnte gezeigt werden, dass es bis jetzt keine einheitlichen Standards gibt zur Casemix-Adjustierung von PROs. Zumeist wurden Regressionsmodelle für die Adjustierung der PROs genutzt. Als Adjustoren kamen unterschiedlichste Patient\*innencharakteristiken zum Einsatz; alle eingeschlossenen Studien nutzten Baseline-PROs als Adjustoren, die meisten schlossen soziodemographische Angaben der Patient\*innen mit ein. Aufgrund der Heterogenität der Studienlage schließt die Arbeit mit der Identifizierung von wichtigen weiteren Forschungsfeldern sowie Empfehlungen für die Durchführung einer angemessenen Casemix-Adjustierung.

### **3.3. Arbeit 3: Funktionelles Outcome ein Jahr nach Radikaler Prostatektomie in 88 Prostatakrebszentren – Vergleich der Ergebnisqualität auf Versorgungsorganisationsebene (Fallbeispiel 1)**

Aufbauend auf den Ergebnissen des Systematic Reviews (vgl. Abschnitt 3.2) wurden für die Radikale Prostatektomie relevante PROs (Inkontinenz, irritative/obstruktive Symptomatik, Sexualität – gemessen durch den EPIC-26) von 7065 Prostatakrebspatienten aus 88 Prostatakrebszentren Casemix-adjustiert. Dafür wurde ein vom NHS genutzter Adjustierungsansatz (ursprünglich für den Einsatz von PROs bei elektiven orthopädischen Operationen<sup>23</sup>) für das Prostatakarzinom adaptiert und auf Patienten und Zentren, die an der PCO-Studie teilnehmen (vgl. Abschnitt 2.4), angewandt. Die Casemix-adjustierte Ergebnisqualität in den Prostatakrebszentren zeigte erhebliche Unterschiede vor allem für die Domänen Inkontinenz und Sexualität. Der Adjustierungseffekt wurde ferner untersucht und ein

Maß vorgeschlagen, um dessen Größe zu analysieren. Dafür wurden *minimally important differences* (MIDs) für die Domänen genutzt: die Anzahl derjenigen Krebszentren, die mehr als eine MID unterhalb des Medians der entsprechenden Domäne fielen, wurde bestimmt. Für Inkontinenz konnte so gezeigt werden, dass 26 % der eingeschlossenen Zentren unterdurchschnittlich schlechte Ergebnisse hatten. Es konnte ein kleiner bis mittlerer Adjustierungseffekt für die untersuchten PROs gezeigt werden (bezogen auf Cohen's  $d^{50}$ ).

### **3.4. Arbeit 4: Funktionelles Outcome ein Jahr nach elektiver Tumorresektion beim kolorektalen Karzinom in 102 Darmkrebszentren – Vergleich der Ergebnisqualität auf Versorgungsorganisationsebene (Fallbeispiel 2)**

Ähnlich dem Vorgehen der Arbeit 3 für Prostatakrebszentren wurde auch für 102 zertifizierte Darmkrebszentren ein Casemix-adjustierter Ergebnisqualitätsvergleich mit insgesamt 4239 eingeschlossenen Patient\*innen durchgeführt. Dabei wurde als PRO-Instrument zum einen das generische, tumorentitätsübergreifende Fragebogenmodul EORTC QLQ-C30<sup>42</sup> genutzt sowie auch das darmkrebsspezifische Modul EORTC QLQ-CR29<sup>41</sup>. Für die Publikation wurden fünf in einer Vorarbeit identifizierten klinisch relevante PRO Scores ausgewählt und untersucht (allgemeine Lebensqualität, Schmerzen, Obstipation, physische Funktion, Abdominalschmerzen)<sup>51</sup>. Vor allem für physische Funktion konnten größere Unterschiede zwischen den Zentren nach erfolgter Casemix-Adjustierung beobachtet werden. Darüber hinaus untersucht die Publikation durch Zusammenhangsanalysen, welche patient\*innenseitigen Charakteristiken die ausgewählten PROs vorhersagen können, wodurch auch eine patient\*innenindividuelle Nutzung von PROs durch Kliniker\*innen ermöglicht werden kann (beispielsweise durch Identifizierung von besonderen Risikogruppen). Für die Ebene der Versorgenden (also der Zentren) zeigt die Arbeit vor allem das Potential auf, auf die Ergebnisqualität bezogene „Ausreißer“ zu identifizieren, da in der Gesamtschau die Unterschiede gemessen durch PROs für diese Stichprobe der Darmkrebszentren eher gering ausfielen.

### **3.5. Arbeit 5: Einstellungen zu und Wissen über PROs von Behandelnden – Möglichkeiten und Limitationen der Implementierung von PROs auf patient\*innenindividueller Ebene**

Um herauszufinden, wie das Wissen über und die Haltung zu PROs von onkologisch Behandelnden in deutschen Krebszentren ist, wurde eine qualitative Befragung mit Kliniker\*innen aus den EDIUM-Studienzentren durchgeführt (vgl. Abschnitt 2.4). An der qualitativen Befragung nahmen 12 Kliniker\*innen (fünf Chirurg\*innen, zwei Onkolog\*innen,

eine Psychoonkologin, zwei onkologische Fachpfleger\*innen, eine Stomatherapeutin, eine medizinische Fachangestellte) teil; zur Auswertung wurde ein inhaltsanalytischer Ansatz nach Kuckartz gewählt<sup>52</sup>. Die Analysen zeigen, dass die Befragten wenig Wissen über und praktische Erfahrung mit PROs hatten. Trotzdem konnten die meisten Kliniker\*innen Vorteile von PROs für den klinischen Einsatz erkennen. Die meisten dieser Vorteile wurden in der Verwendung von PROs als strukturierenden Ansatz zur Quantifizierung darmkrebspezifischer Symptome und Funktionen gesehen, die den Kliniker\*innen mehr Informationen über Krebspatient\*innen und ihre Behandlung zur Verfügung stellen. Die Teilnehmenden sahen mehrere Nutzungsmöglichkeiten für PROs - nicht nur als klinisches Messinstrument, sondern auch als wichtiger wissenschaftlicher Indikator für Behandlungsergebnisse und für die Qualitätssicherung. Die zahlreichen Anwendungsfälle können bei künftigen Bemühungen um eine erfolgreiche Implementierung von PROs im klinischen Kontext sehr hilfreich sein, da sie mehrere Stärken von PROs aufzeigen.

## 4. Publikationen dieser Arbeit

### 4.1. Psychometric validation of the German version of the EPIC-26 questionnaire for patients with localized and locally advanced prostate cancer

Veröffentlicht als: **Sibert, N. T.**, Dieng, S., Oesterle, A., Feick, G., Carl, G., Steiner, T., Minner, J., Roghmann, F., Kaftan, B., Zengerling, F., Hinkel, A., Beyer, B., Heidenreich, A., Harke, N., Brehmer, B., Pfitzenmaier, J., Fichtner, J., Neisius, A., Hammerer, P., ... Kowalski, C. (2019). Psychometric validation of the German version of the EPIC-26 questionnaire for patients with localized and locally advanced prostate cancer. *World Journal of Urology*. <https://doi.org/10.1007/s00345-019-02949-7>



## Psychometric validation of the German version of the EPIC-26 questionnaire for patients with localized and locally advanced prostate cancer

Nora Tabea Sibert<sup>1</sup> · Sebastian Dieng<sup>2</sup> · Alisa Oesterle<sup>2</sup> · Günter Feick<sup>3</sup> · Günther Carl<sup>4</sup> · Thomas Steiner<sup>5</sup> · Jörg Minner<sup>6</sup> · Florian Roghmann<sup>7</sup> · Björn Kaftan<sup>8</sup> · Friedemann Zengerling<sup>9</sup> · Andreas Hinkel<sup>10</sup> · Burkhard Beyer<sup>11</sup> · Axel Heidenreich<sup>12</sup> · Nina Harke<sup>13</sup> · Bernhard Brehmer<sup>14</sup> · Jesco Pfitzenmaier<sup>15</sup> · Jan Fichtner<sup>16</sup> · Andreas Neisius<sup>17</sup> · Peter Hammerer<sup>18</sup> · Simone Wesselmann<sup>1</sup> · Christoph Kowalski<sup>1</sup>

Received: 20 May 2019 / Accepted: 4 September 2019  
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

### Abstract

**Purpose** For patients with prostate cancer, validated and reliable instruments are essential for measuring patient-reported outcomes. The aim of this study was to validate the German version of the widely established Expanded Prostate Cancer Index Composite with 26 items (EPIC-26).

**Methods** A German translation of the original questionnaire was tested in 3094 patients with localized or locally advanced (any T, any N and M0) prostate cancer with treatment intent (including radical prostatectomy, brachytherapy, active surveillance, watchful waiting). They completed the EPIC-26 questionnaire before treatment. A total of 521 of them also completed a questionnaire 12 months afterward. Internal consistency, sensitivity to change, and construct validity were assessed.

**Results** The internal consistency of all domains was sufficient (Cronbach's alpha between 0.64 and 0.93). Item-to-scale correlation coefficients showed acceptable associations between items and their domain score (all > 0.30), with the lowest scores for "bloody stools" ( $r=0.37$ ) and "breast problems" ( $r=0.32$ ). Confirmatory and exploratory factor analysis confirmed the five-dimension structure of the EPIC-26 (comparative fit index 0.95).

**Conclusions** Psychometric evaluation suggests that the German version of the EPIC-26 is a well-constructed instrument for measuring patient-reported health-related symptoms in patients with prostate cancer.

**Keywords** Prostate cancer · Patient-reported outcomes · EPIC-26 · Validation · Health-related Quality of Life

✉ Nora Tabea Sibert  
sibert@krebsgesellschaft.de

<sup>1</sup> German Cancer Society, Kuno-Fischer-Strasse 8,  
Berlin 14057, Germany

<sup>2</sup> OnkoZert, Neu-Ulm, Germany

<sup>3</sup> Federal Association of German Prostate Cancer Patient  
Support Groups, Bonn, Germany

<sup>4</sup> Help for Prostate Cancer Patients (Förderverein Hilfe bei  
Prostatakrebs e.V.–FHbP), Tornesch, Germany

<sup>5</sup> Helios Klinikum Erfurt, Erfurt, Germany

<sup>6</sup> Hegau-Bodensee-Klinikum Singen, Singen, Germany

<sup>7</sup> Marien Hospital, Ruhr-University Bochum, Herne, Germany

<sup>8</sup> Städtisches Klinikum Lüneburg, Lüneburg, Germany

<sup>9</sup> Universitätsklinikum Ulm, Ulm, Germany

<sup>10</sup> Franziskus Hospital, Bielefeld, Germany

<sup>11</sup> Martini-Klinik Prostate Cancer Center Hamburg, Hamburg,  
Germany

<sup>12</sup> Universitätsklinikum Köln, Cologne, Germany

<sup>13</sup> University Hospital Essen, Essen, Germany

<sup>14</sup> Diaconal Hospital of Schwäbisch Hall, Schwäbisch Hall,  
Germany

<sup>15</sup> Evangelisches Klinikum Bethel Johannesstift, Bielefeld,  
Germany

<sup>16</sup> Johanniter Krankenhaus Oberhausen, Oberhausen, Germany

<sup>17</sup> Barmherzige Brüder Trier, Trier, Germany

<sup>18</sup> Städtisches Klinikum Braunschweig, Brunswick, Germany

## Introduction

Prostate cancer (PCa) is the most common cancer among men in Germany, with 57,370 newly diagnosed cases in 2014 [1]. As the survival rates are increasing [2] and the 5-year as well as the 10-year overall survival rates are at a high level (91% and 90%, respectively in 2014) [1], patient-reported outcome (PRO) measurements of PCa-related symptoms and functions are an important part of clinical outcome measurement. PCa patients often describe specific cancer-related and therapy-related adverse effects (AEs). Disease-related symptom measurement in PCa patients should therefore differentiate between illness-specific domains such as urinary incontinence, bowel function, sexual function, and endocrine function [3–5]. Measurement of these PCa-related symptoms is of interest not only in relation to investigation of individual patients' AEs during PCa treatment, for example, but also for comparing patient outcomes across providers.

A widely established instrument for health-related quality of life (HRQoL) assessment in men with PCa is the Expanded Prostate Cancer Index Composite with 26 items (EPIC-26). The EPIC-26 questionnaire is an abbreviated version of the EPIC-50 questionnaire [6], which is used in many completed and ongoing studies. The original EPIC included a set of 50 questions. As there was a need to facilitate HRQoL measurement to allow more comprehensive research and clinical approaches, Szymanski et al. developed the EPIC-26 [7]. Subsequently, Schmidt et al. identified the EPIC-26 as being the HRQoL instrument with the best characteristics for managing change in PCa care [8]. Psychometric test results are available for several translations of the EPIC-26 (e.g., into Chinese, Norwegian, and Canadian French) [9–11].

A German translation of the EPIC-26 was published by Beyer et al. [12]. Umbehr et al. [13] validated a German version of the EPIC questionnaire with 50 items (including the items in the shorter EPIC-26) using a translation different from that used by Beyer et al. as the authors were focusing more on a cultural-adaptational approach resulting in a questionnaire applicable not only in Germany, but also in the German-speaking parts of Austria and Switzerland. No results from psychometric testing of the validity and reliability of the German version of EPIC-26 as translated by Beyer et al. have yet been published. The aim of the present study was therefore to test the German version of EPIC-26 for reliability and validity using responses from patients with PCa who answered the questionnaire before and 1 year after definitive treatment. Data from the ongoing prospective, multicenter Prostate Cancer Outcomes (PCO) study ([www.pco-study.com](http://www.pco-study.com)) [Kowalski et al. submitted] were used. The German PCO study is part of an ongoing international study (the TrueNTH Global Registry) that was initiated by the

Movember Foundation. The aim of the study is to compare clinical and patient-reported outcomes internationally for patients with localized PCa [14]. Fourteen countries—Austria, Australia, Canada, the Czech Republic, Denmark, Finland, Germany, Ireland, Italy, Spain, the Netherlands, New Zealand, the UK, and the USA—are currently participating in the TrueNTH Global Registry.

## Ethical approval

The PCO study was approved by the local ethics committee in Berlin (Eth-12/16). All of the patients included provided written informed consent.

## Materials and methods

### Patients

Patients with localized or advanced PCa (any T, any N, M0) are asked to complete the EPIC-26 questionnaire before and 1 year after definitive treatment. Patients treated with active surveillance (AS) or watchful waiting (WW) are asked to complete the follow-up questionnaire 12 months after being included in the study. PCa-specific clinical parameters (e.g., cancer stage according to the German Guideline for Prostate Cancer presented in Appendix 4, type of therapy, comorbidities) as well as patient-reported outcomes (PROs) measured using the EPIC-26 were collected in a setting of consecutive inclusion of patients between July 2016 and March 2018, in accordance with the reference guide of the International Consortium for Health Outcomes Measurement (ICHOM) standard set for localized prostate cancer [15, 16]. The study includes PCa patients receiving radical prostatectomy including cystoprostatectomy (RPE), radiation therapy (RT), watchful waiting (WW), active surveillance (AS), and other treatments (OT, such as androgen deprivation or focal therapy). All risk groups (low, medium, high, T3/T4, N1) are represented.

Questionnaire validation was conducted on the basis of the responses given by these patients in the participating German prostate cancer centers. Since this ongoing study started in July 2016, post-treatment data (12 months after study inclusion) were already available for patients who were recruited early during the study.

### The German version of the EPIC-26

The EPIC-26 is a set of 26 items, which are categorized into one single item (overall urinary problems) and five subdomains consisting of four to six items: incontinence (four items), irritative and obstructive urinary problems (four items), bowel function (six items), sexuality (six items),

and vitality/hormonal function (five items). Responses are measured on a Likert-type scale. The original and German versions of the EPIC-26 are included in the appendixes (Appendixes 2 and 3).

The German version of the EPIC-26 published by Beyer et al. [12] is used in this study. The process of translation and the evaluation of the translation are described in that paper.

**Statistical analysis**

Reliability was measured using the responses before definitive treatment by calculating interdomain correlations and Cronbach's alpha coefficients [17] as an indicator of internal consistency. Item-to-scale correlations for each subdomain were carried out as described by Cohen [18]. Sensitivity to change was examined by calculating correlations between EPIC-26 scores before and 12 months after treatment for those patients with available data. Paired Student's two-sample *t* test was performed.

Construct validity was analyzed on the basis of responses before definitive treatment by using a confirmatory factor analysis to examine the five-subdomain structure and the scales' dimensionality in the EPIC-26. Beforehand, an exploratory factor analysis was performed to analyze whether a five-subdomain structure is statistically tenable. Maximum likelihood with robust correction was used.

The open-source software program RStudio 1.1.383 was used for all statistical analyses.

**Results**

**Sample**

A total of 3094 PCa patients from 44 German PCa centers were included. At the time of data transfer (April 2018), post-treatment data were already available for 521 of these patients (within the 1-year follow-up time frame, 839 patients were asked to answer the post-treatment questionnaire). The mean age at diagnosis in the total patient sample was 66 years. In all, 2608 patients were treated with RPE alone; 221 with RT; 172 patients with a combination of RPE and RT; 79 patients with AS or WW; and 14 patients received other treatment options, including radical cystectomy (cf. Kowalski et al. submitted) (Table 1).

**Statistics**

**Reliability**

Reliability was investigated before treatment (T0) on the basis of the total patient sample. Item-to-scale correlation ranged from 0.32 to 0.95, with the lowest coefficients for

**Table 1** Characteristics of the patients (*n*=3094)

	All patients ( <i>n</i> =3094)	12 months after treatment ( <i>n</i> =521)
<b>Age</b>		
Mean (SD)	66.0 (7.4)	66.0 (7.4)
Median	66	66
<b>Risk groups</b>		
Localized, high risk <sup>a</sup>	991 (32%)	162 (31%)
Localized, medium risk <sup>a</sup>	1399 (45%)	247 (47%)
Localized, low risk <sup>a</sup>	521 (17%)	93 (18%)
Localized advanced (T3/T4)	144 (5%)	17 (3%)
Advanced (N1)	39 (1%)	2 (<1%)
<b>Type of therapy</b>		
RPE alone	2608 (84%)	452 (87%)
Radiotherapy alone	221 (7%)	33 (6%)
RPE and RT	172 (6%)	28 (5%)
AS/WW	79 (3%)	8 (2%)
Other	14 (<1%)	0

AS active surveillance, RPE radical prostatectomy including cystoprostatectomy, RT radiotherapy, SD standard deviation, WW watchful waiting

<sup>a</sup>Risk class in accordance with the German Guideline for Prostate Cancer (cf. Appendix 4) [22]

the items "bloody stools (#6d)" (*r*=0.37) and "breast problems (#13b)" (*r*=0.32) (Table 2). Internal consistency was assessed by calculating Cronbach's alpha for the EPIC-26 subdomains, where a Cronbach's alpha > 0.65 was considered sufficient [19]. This was possible for all but one subdomain; for the urinary irritative/obstructive subdomain score alone, Cronbach's alpha was 0.64 (Table 3). Table 2 provides an overview of the characteristics of the subdomains (in comparison with selected findings from Szymanski et al. [7]).

Table 4 shows the interdomain correlation matrix. The low correlation scores (range between 0.11 and 0.32) indicate sufficient discrimination between the single scores and thus confirm the five-dimension structure of the EPIC-26 (all *P*<0.001).

**Sensitivity to change**

Sensitivity to change was investigated in 521 patients for whom data were available for questionnaire responses 12 months after treatment. For all subdomains, significant changes over time were found (*P*<0.001), with the greatest difference being seen in the sexual domain (Table 3).

**Table 2** Item-to-scale correlation coefficients, mean scores, and factor loading for the German and American versions of the EPIC-26 (before treatment)

Domains and items	EPIC-26 item number	Mean score for item (SD)	Item-to-scale correlation (Germany)	Item-to-scale correlation (America)	Factor loading
Urinary incontinence					
Leaking > 1 time	1	90.4 (25.8)	0.87	0.75	0.69
Dribbling	2	87.1 (22.5)	0.74	0.77	0.66
Pad use	3	97.4 (12.4)	0.62	0.66	0.48
Leaking problem	4a	90.6 (19.6)	0.85	0.83	0.86
Urinary irritative/obstructive					
Dysuria	4b	94.2 (17.2)	0.61	0.65	0.45
Hematuria	4c	97.5 (12.5)	0.42	0.32	0.22
Weak stream	4d	75.4 (29.1)	0.85	0.67	0.76
Frequency	4e	71.7 (31.2)	0.84	0.61	0.80
Overall urinary problem	5	80.6 (27.9)			
Bowel					
Urgency	6a	93.7 (17.7)	0.83	0.77	0.86
Frequency	6b	94.9 (15.6)	0.80	0.81	0.78
Fecal incontinence	6c	98.8 (7.8)	0.54	0.65	0.45
Bloody stools	6d	99.2 (6.52)	0.37	0.55	0.19
Rectal pain	6e	94.0 (17.1)	0.66	0.65	0.41
Overall bowel problem	7	93.2 (18.2)	0.82	0.83	0.80
Sexual					
Poor erections	8a	50.4 (32.1)	0.94	0.86	0.95
Difficulty with orgasm	8b	53.0 (32.5)	0.92	0.68	0.92
Erections not firm	9	63.6 (40.1)	0.86	0.79	0.84
Erections not reliable	10	56.1 (38.7)	0.92	0.81	0.89
Poor sexual function	11	48.6 (31.5)	0.95	0.80	0.95
Overall sexual problem	12	71.6 (32.7)	0.59	0.50	0.49
Vitality/hormonal					
Hot flashes	13a	93.8 (18.1)	0.58	0.38	0.33
Breast problems	13b	98.7 (8.36)	0.32	0.31	0.12
Depression	13c	83.3 (27.0)	0.84	0.62	0.79
Lack of energy	13d	81.4 (27.0)	0.85	0.58	0.93
Weight change	13e	91.7 (20.7)	0.65	0.42	0.41

**Construct validity**

Construct validity was investigated before treatment using the total patient sample. After factor extraction using Kaiser's criterion (all factors with an eigenvalue greater than one are retained), solutions with up to six factors were possible. Exploratory factor analysis was therefore performed with four, five, and six latent variables. After extraction of five factors and factor rotation using the varimax method, the results showed a fit at  $\chi^2=1664.44$  ( $n=3094$ , degrees of freedom 185,  $P<0.001$ ), a standardized root mean square residual (SRMR) of 0.02, and a Tucker–Lewis index of 0.94. This five-factor solution was superior to the four-factor and six-factor solutions (cf. Appendix 1). For all exploratory factor analyses, item #5 was not taken into account, as it is

not used in the original EPIC-26 or any validated translation of it. A confirmatory factor analysis with five latent variables (the five domains of the EPIC-26) was thus estimated using the robust maximum likelihood estimator. After 68 iterations, the results showed a fit at  $\chi^2=991.72$  ( $n=3094$ , degrees of freedom 265 after 60 iterations,  $P<0.001$ ), a comparative fit index (CFI) of 0.95, a root mean square error of approximation (RMSEA) of 0.04, and an SRMR of 0.04. Factor loadings ranged between 0.12 and 0.95, with the weakest association with their latent variable for the items "hematuria (#4c)" ( $B=0.22$ ), "bloody stools (#6d)" ( $B=0.19$ ) and "breast problems (#13b)" ( $B=0.12$ ) (Table 2).

**Table 3** Characteristics of the domains in the German version of the EPIC-26 questionnaire in comparison with selected American findings (T0:  $n=3094$ , T1:  $n=521$ )

	Mean (SD)		Mean (SD) America <sup>a</sup>	Mean difference (T1 - T0) as reported by <i>t</i> test <sup>b</sup>	<i>t</i> test, comparison of T0 and T1, <i>t</i> , df ( <i>P</i> value)	Median Germany (T0)	Skewness Germany (T0)	Kurtosis Germany (T0)	Cronbach's alpha	
	Germany (T0)	Germany (T1)							Germany (T0)	America
Urinary incontinence	92.2 (14.8)	74.1 (27.2)	83.2 (22.9)	-16.9	14.2, 477 (<0.001)	100	-2.29	4.28	0.75	0.86
Urinary irritative/ obstructive	85.3 (16.2)	89.8 (11.9)	80.5 (20.2)	4.2	-5.4, 436 (<0.001)	87.5	-1.75	2.14	0.64	0.74
Bowel	95.8 (9.7)	93.0 (12.9)	85.0 (19.3)	-2.8	5.0, 452 (<0.001)	100	-2.55	5.06	0.78	0.89
Sexual	58.2 (29.5)	24.3 (22.9)	34.4 (28.1)	-34.8	25.4, 495 (<0.001)	61.2	-0.23	-1.23	0.93	0.90
Vitality/hormonal	89.9 (14.6)	84.0 (17.9)	87.1 (15.0)	-5.7	7.1, 476 (<0.001)	95	-2.24	4.35	0.70	0.70

<sup>a</sup> After treatment (range of time between questionnaire and primary intervention: 3-51 months); Szymanski et al. [7]

<sup>b</sup> *t* test based on  $n=521$  patients for whom T0 and T1 data were available

### Discussion

The aim of this study was to validate the German version of the EPIC-26 questionnaire. In general, the results of the psychometric tests described are similar to those for the original American questionnaire [7]. As a result, the findings suggest that the German version of the EPIC-26 can be used as a reliable and valid instrument to analyze symptoms and function in PCa patients.

The mean scores for the different time points were comparable to previously published results [7, 10, 11] (Table 3). However, differences between the study samples and timing of measurements in comparison with validations carried out for other languages need to be carefully discussed, as some of the published studies report results of questioning before treatment and some after it. Eighty-five percent of the patients included in the present sample, compared with 76% in the source population for these patients, were treated with RPE. Common AEs consequently include erectile dysfunction, for example, which is less prevalent in patients who undergo RT. The relatively low score for the sexual domain 12 months after treatment (mean = 24.28) in the sample therefore needs to be interpreted with these circumstances being taken into account. However, it may be assumed that the over-representation of patients with RPE in the study does not impede the quality of the questionnaire validation presented here.

The findings showed that the German version of the EPIC-26 questionnaire has sufficient reliability. The internal consistency of the scales (with Cronbach's alpha ranging between 0.64 and 0.93) and item-to-scale correlation coefficients (all  $\geq 0.30$ ) were acceptable. A Cronbach's alpha  $> 0.65$  was considered satisfactory. Only the urinary irritative/obstructive subdomain had a Cronbach's alpha below this threshold ( $\alpha=0.64$ ). This has also been reported for other versions of the EPIC-26 [10, 11], suggesting weaker internal consistency for this subdomain in general that is mainly due to the item "hematuria."

Psychometric testing commonly requires item-to-scale correlation coefficients to be over 0.40, or at least not less than 0.30, to confirm adequate reliability. As reported above, item-to-scale coefficients of between 0.30 and 0.40 were found for the items "bloody stools" and "breast problems." This is consistent with previously published results [10, 11]. For the "breast problems" item, this low correlation with the vitality/hormonal domain was also reported for the original EPIC-26 questionnaire. As these findings have now been confirmed by several studies, it may be suggested that the construction of the vitality/hormonal domain should be revised, and providers may be advised to interpret the single items rather than the total score.

**Table 4** Interdomain correlation matrix,  $P < 0.001$

	Urinary incontinence	Urinary irritative/obstructive	Bowel	Sexual	Vitality/hormonal
Urinary incontinence	1				
Urinary irritative/obstructive	0.32	1			
Bowel	0.17	0.27	1		
Sexual	0.19	0.20	0.11	1	
Vitality/hormonal	0.15	0.23	0.26	0.13	1

An acceptable level of sensitivity to change was found, analyzed by comparing the responses of patients before and 12 months after treatment with only the bowel and vitality/hormonal subdomain, showing low mean differences which are still statistically significant. This corresponds to clinically known AEs in PCa patients after treatment [4, 20]. Since only 33 of the 521 patients included in the analyses of sensitivity to change were treated with RT, the differences in EPIC-26 reported mostly involve RPE-related side effects.

Four-dimension, five-dimension, and six-dimension structures in EPIC-26 were compared using exploratory factor analysis. Kaiser's criterion suggested six factors. However, Kaiser's criterion often overestimates the number of factors [21]. After comparison of the three possible solutions, and without taking into consideration the item "overall urinary problem" (item #5), which is not used for domain construction in any of the existing translations, extracting five factors appeared to be the most convincing approach in relation to factor loadings, discriminatory power, and the authors' intention to cover the five domains in the instrument. Exclusion of the item "overall urinary problem" was based on the questionnaire's construction process. If the item is permitted for the factor analysis, a four-factor solution combining items #1 to #5 (all urinary items) is also statistically approved. The five-dimension structure of the EPIC-26 was also confirmed using confirmatory factor analysis. The weakest loading of an item and its postulated factor (e.g., subdomain) were found for the items "hematuria" and "breast problems." As these findings are also reported in other studies, the importance of these two items for the domain may be questionable.

**Limitations**

Test-retest reliability and convergent validity were not examined in the present study. The large sample size ( $n = 3094$ ) and the multicenter approach may be regarded as strengths of the study.

**Conclusion**

Measuring patient-reported outcomes is attracting increasing interest in oncological research and also in health-care provision. Valid and reliable psychometric instruments are therefore needed to evaluate disease-related symptoms in patients with prostate cancer. An international consensus group has strongly recommended using the EPIC-26 questionnaire for comparable and standardized evaluation of PROs and AEs in patients with localized prostate cancer [15]. The psychometric findings presented here mean that the German version of EPIC-26 is now available as a well-validated and reliable instrument for further PRO research and can be recommended for German-speaking patients with prostate cancer.

**Author contributions** NT Sibert: data analysis, data collection and management, manuscript writing/editing; S Dieng: protocol/project development, data collection and management, manuscript writing/editing; A Oesterle: data analysis, data collection and management; G Feick: protocol/project development, manuscript writing/editing; G Carl: protocol/project development, manuscript writing/editing; T Steiner: data collection and management, manuscript writing/editing; J Minner: data collection and management, manuscript writing/editing; F Roghmann: data collection and management, manuscript writing/editing; B Kaftan: data collection and management, manuscript writing/editing; F Zengerling: B Kaftan: data collection and management, manuscript writing/editing; A Hinkel: data collection and management, manuscript writing/editing; B Beyer: data collection and management, manuscript writing/editing; A Heidenreich: data collection and management, manuscript writing/editing; N Harke: data collection and management, manuscript writing/editing; B Brehmer: data collection and management, manuscript writing/editing; J Pfitzmaier: data collection and management, manuscript writing/editing; J Fichtner: data collection and management, manuscript writing/editing; A Neisius: data collection and management, manuscript writing/editing; P Hammerer: data collection and management, manuscript writing/editing; S Wesselmann: protocol/project development, manuscript writing/editing; C Kowalski: protocol/project development, data analysis, manuscript writing/editing.

**Funding** This study was funded by the Movember Foundation.

**Compliance with ethical standards**

**Conflicts of interest** GC, GF, TS, JM, FR, FZ, AHi, BB, AHe, NH, BB, JP, JF, AN, and PH hereby declare that they have no potential conflicts of interest. NTS, CK, SD, AO, and SW are employees of the two institutions in charge of the certification system. BK receives consulting and lecturing honoraria from various companies that do not involve any conflicts with the research presented here. A comprehensive list is available on request.

**Research involving human participants and/or animals** All procedures performed in studies involving human participants were in accordance with the ethical standards of the ethics committee of the Medical Association of Berlin (Eth-12/16) and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

**Informed consent** Informed consent was obtained from all individual participants included in the study.

**Appendix 1: Comparison of the four-, five- and six-factor solutions**

See Tables 5, 6 and 7.

**Table 5** Factor loadings > 0.30 of the four-factor solution ( $p < 0.001$ ,  $\chi^2 = 2865.08$ ,  $df = 206$ ,  $RMSA = 0.04$ ,  $TLI = 0.90$ )

Domains and items	EPIC-26 item number	Factor 1	Factor 2	Factor 3	Factor 4
Leaking > 1 time	1			- 0.68	
Dribbling	2			- 0.62	
Pad use	3			0.48	
Leaking problem	4a			0.83	
Dysuria	4b			0.31	
Hematuria	4c				
Weak stream	4d			0.38	
Frequency	4e			0.48	
Urgency	6a		0.86		
Frequency	6b		0.73		
Fecal incontinence	6c		0.43		
Bloody stools	6d				
Rectal pain	6e		0.39		
Overall bowel problem	7		0.80		
Poor erections	8a	0.94			
Difficulty with orgasm	8b	0.91			
Erections not firm	9	0.82			
Erections not reliable	10	0.88			
Poor sexual function	11	0.93			
Overall sexual problem	12	- 0.43			
Hot flashes	13a				0.35
Breast problems	13b				
Depression	13c				0.83
Lack of energy	13d				0.89
Weight change	13e				0.40

**Table 6** Factor loadings > 0.30 for the five-factor solution ( $p < 0.001$ ,  $\chi^2 = 1605.99$ ,  $df = 185$ , RMSA 0.03, TLI = 0.94)

Domains and items	EPIC-26 item number	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Leaking > 1 time	1				- 0.71	
Dribbling	2				- 0.59	
Pad use	3				0.50	
Leaking problem	4a				0.78	
Dysuria	4b					0.33
Hematuria	4c					
Weak stream	4d					0.85
Frequency	4e					0.65
Urgency	6a		0.86			
Frequency	6b		0.73			
Fecal incontinence	6c		0.43			
Bloody stools	6d					
Rectal pain	6e		0.38			
Overall bowel problem	7		0.80			
Poor erections	8a	0.94				
Difficulty with orgasm	8b	0.91				
Erections not firm	9	0.83				
Erections not reliable	10	0.88				
Poor sexual function	11	0.93				
Overall sexual problem	12	- 0.43				
Hot flashes	13a			0.35		
Breast problems	13b					
Depression	13c			0.83		
Lack of energy	13d			0.88		
Weight change	13e			0.39		

**Table 7** Factor loadings > 0.30 for the six-factor solution ( $p < 0.001$ ,  $\chi^2 = 1186.32$ ,  $df = 165$ , RMSA 0.02, TLI = 0.95)

Domains and items	EPIC-26 item number	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Leaking > 1 time	1				- 0.71		
Dribbling	2				- 0.59		
Pad use	3				- 0.71		
Leaking problem	4a				- 0.59		
Dysuria	4b				0.49		
Hematuria	4c				0.79		
Weak stream	4d						0.45
Frequency	4e						0.42
Urgency	6a					0.78	
Frequency	6b					0.67	
Fecal incontinence	6c		0.86				
Bloody stools	6d		0.73				
Rectal pain	6e		0.43				
Overall bowel problem	7						0.32
Poor erections	8a		0.31				0.39
Difficulty with orgasm	8b		0.77				
Erections not firm	9	0.94					
Erections not reliable	10	0.91					
Poor sexual function	11	0.83					
Overall sexual problem	12	0.88					
Hot flashes	13a				0.94		
Breast problems	13b	- 0.43					
Depression	13c			0.36			
Lack of energy	13d						
Weight change	13e			0.84			

**Appendix 2: The EPIC-26—Expanded Prostate Cancer Index Composite Short Form [23]**

1. Over the past 4 weeks, how often have you leaked urine? (Circle one number)

- More than once a day..... 1
- About once a day..... 2
- More than once a week..... 3
- About once a week..... 4
- Rarely or never..... 5

2. Which of the following best describes your urinary control during the last 4 weeks? (Circle one number)

- No urinary control whatsoever.....1
- Frequent dribbling..... 2
- Occasional dribbling..... 3
- Total control..... 4

3. How many pads or adult diapers per day did you usually use to control leakage during the last 4 weeks? (Circle one number)

- None ..... 0
- 1 pad per day..... 1
- 2 pads per day..... 2
- 3 or more pads per day..... 3

4. How big a problem, if any, has each of the following been for you during the last 4 weeks? (Circle one number on each line)

	No problem	Very small problem	Small problem	Moderate problem	Big problem
a. Dripping or leaking urine	0	1	2	3	4
b. Pain or burning on urination	0	1	2	3	4
c. Bleeding with urination	0	1	2	3	4
d. Weak urine stream or incomplete emptying	0	1	2	3	4
e. Need to urinate frequently during the day	0	1	2	3	4

5. Overall, how big a problem has your urinary function been for you during the last 4 weeks? (Circle one number)

- No problem..... 1
- Very small problem..... 2
- Small problem..... 3
- Moderate problem..... 4
- Big problem..... 5

6. How big a problem, if any, has each of the following been for you? (Circle one number on each line)

	No problem	Very small problem	Small problem	Moderate problem	Big problem
a. Urgency to have a bowel movement	0	1	2	3	4
b. Increased frequency of bowel movements	0	1	2	3	4
c. Losing control of your stools	0	1	2	3	4
d. Bloody stools	0	1	2	3	4
e. Abdominal/Pelvic/Rectal pain	0	1	2	3	4

7. Overall, how big a problem have your bowel habits been for you during the last 4 weeks? (Circle one number)

- No problem..... 1
- Very small problem..... 2
- Small problem..... 3
- Moderate problem..... 4
- Big problem..... 5

8. How would you rate each of the following during the last 4 weeks?(Circle one number on each line)

	Very poor to none	Poor	Fair	Good	Very good
a. Your ability to have an erection?	1	2	3	4	5
b. Your ability to reach orgasm (climax)?	1	2	3	4	5

9. How would you describe the usual QUALITY of your erections during the last 4 weeks? (Circle one number)

- None at all..... 1
- Not firm enough for any sexual activity..... 2
- Firm enough for masturbation and foreplay only..... 3
- Firm enough for intercourse..... 4

10. How would you describe the FREQUENCY of your erections during the last 4 weeks? (Circle one number)

- I NEVER had an erection when I wanted one..... 1
- I had an erection LESS THAN HALF the time I wanted one..... 2
- I had an erection ABOUT HALF the time I wanted one ..... 3
- I had an erection MORE THAN HALF the time I wanted one..... 4
- I had an erection WHENEVER I wanted one..... 5

11. Overall, how would you rate your ability to function sexually during the last 4 weeks? (Circle one number)

- Very poor..... 1 Poor..... 2  
Fair..... 3 Good..... 4  
Very good..... 5

12. Overall, how big a problem has your sexual function or lack of sexual function been for you during the last 4 weeks? (Circle one number)

- No problem..... 1  
Very small problem..... 2  
Small problem..... 3  
Moderate problem..... 4  
Big problem..... 5

13. How big a problem during the last 4 weeks, if any, has each of the following been for you? (Circle one number on each line)

	No problem	Very small problem	Small problem	Moderate problem	Big problem
a. Hot flashes	0	1	2	3	4
b. Breast tenderness/ enlargement	0	1	2	3	4
c. Feeling depressed	0	1	2	3	4
d. Lack of energy	0	1	2	3	4
e. Change in body weight	0	1	2	3	4

**Appendix 3: German translation of the EPIC-26 [12]**

1. Wie oft haben Sie **in den letzten 4 Wochen** Urin verloren? (Bitte nur eine Zahl einkreisen)

- Öfter als einmal am Tag .....1
- Etwa einmal am Tag .....2
- Öfter als einmal in der Woche.....3
- Etwa einmal in der Woche .....4
- Selten oder nie .....5

2. Welche der folgenden Aussagen beschreibt am besten, wie Ihre Kontrolle über die Harnentleerung **in den letzten 4 Wochen** war? (Bitte nur eine Zahl einkreisen)

- Keinerlei Kontrolle über die Harnentleerung .....1
- Häufiges Tröpfeln .....2
- Gelegentliches Tröpfeln .....3
- Vollständige Kontrolle.....4

3. Wie viele Einlagen oder Inkontinenzvorlagen haben Sie normalerweise **in den letzten 4 Wochen** am Tag gebraucht, um den Urinverlust unter Kontrolle zu haben? (Bitte nur eine Zahl einkreisen)

- Keine .....0
- Eine Einlage am Tag .....1
- Zwei Einlagen am Tag .....2
- Drei oder mehr Einlagen am Tag .....3

4. Wie sehr hat Ihnen Folgendes **in den letzten 4 Wochen** Probleme bereitet? (Bitte in jeder Zeile eine Zahl einkreisen)

	Kein Problem	Sehr kleines Problem	Kleines Problem	Mäßiges Problem	Großes Problem
a. Tröpfeln oder Urinverlust	0	1	2	3	4
b. Schmerzen oder Brennen beim Wasserlassen	0	1	2	3	4
c. Blut im Urin	0	1	2	3	4
d. Schwacher Harnstrahl oder unvollständige Blasenentleerung	0	1	2	3	4
e. Häufiger Harndrang tagsüber	0	1	2	3	4

5. Alles in allem, wie sehr haben Ihnen Beschwerden im Zusammenhang mit dem Wasserlassen **in den letzten 4 Wochen** Probleme bereitet? (Bitte nur eine Zahl einkreisen)

- Kein Problem .....1
- Sehr kleines Problem .....2
- Kleines Problem .....3
- Mäßiges Problem .....4
- Großes Problem .....5

6. Wie sehr hat Ihnen Folgendes **in den letzten 4 Wochen** Probleme bereitet? (Bitte in jeder Zeile eine Zahl einkreisen)

	Kein Problem	Sehr kleines Problem	Kleines Problem	Mäßiges Problem	Großes Problem
a. Stuhldrang	0	1	2	3	4
b. Vermehrter Stuhlgang	0	1	2	3	4
c. Verlust der Stuhlkontrolle	0	1	2	3	4
d. Blutiger Stuhl	0	1	2	3	4
e. Schmerzen im Bauch/im Becken/im Rektum	0	1	2	3	4

7. Alles in allem, wie sehr hat Ihnen Ihr Stuhlgang **in den letzten 4 Wochen** Probleme bereitet? (Bitte nur eine Zahl einkreisen)

- Kein Problem .....1
- Sehr kleines Problem .....2
- Kleines Problem .....3
- Mäßiges Problem .....4
- Großes Problem .....5

8. Wie würden Sie Folgendes, bezogen **auf die letzten 4 Wochen**, einschätzen? (Bitte in jeder Zeile eine Zahl einkreisen)

	Very poor to none	Poor	Fair	Good	Very good
a. Ihre Fähigkeit, eine Erektion zu haben	1	2	3	4	5
b. Ihre Fähigkeit, zum Orgasmus (Höhepunkt) zu kommen	1	2	3	4	5

9. Wie würden Sie die übliche **QUALITÄT** Ihrer Erektionen **in den letzten 4 Wochen** beschreiben? (Bitte nur eine Zahl einkreisen)

- Hatte keine Erektion ..... 1
- Nicht fest genug für irgendeine Form der sexuellen Aktivität ..... 2
- Nur fest genug für Masturbation und Vorspiel ..... 3
- Fest genug für Geschlechtsverkehr ..... 4

10. Wie würden Sie die **HÄUFIGKEIT** Ihrer Erektionen **in den letzten 4 Wochen** beschreiben? (Bitte nur eine Zahl einkreisen)

- Ich hatte **NIE** eine Erektion, wenn ich eine wollte ..... 1
- Ich hatte **WENIGER ALS HALB SO OFT** eine Erektion wie ich haben wollte.. 2
- Ich hatte **ETWA HALB SO OFT** eine Erektion wie ich haben wollte ..... 3
- Ich hatte **MEHR ALS HALB SO OFT** eine Erektion wie ich haben wollte ..... 4
- Ich hatte eine Erektion **WANN IMMER** ich sie haben wollte ..... 5

11. Alles in allem, wie würden Sie Ihre sexuelle Funktionsfähigkeit in den letzten 4 Wochen einschätzen? (Bitte nur eine Zahl einkreisen)

- Sehr schlecht .....1
- Schlecht .....2
- Einigermaßen .....3
- Gut .....4
- Sehr gut .....5

12. Alles in allem, wie sehr hat Ihnen Ihre sexuelle Funktionsfähigkeit oder deren Fehlen in den letzten 4

Wochen Probleme bereitet? (Bitte nur eine Zahl einkreisen)

- Kein Problem .....1
- Sehr kleines Problem .....2
- Kleines Problem .....3
- Mäßiges Problem .....4
- Großes Problem .....5

13. Wie sehr hat Ihnen Folgendes in den letzten 4 Wochen Probleme bereitet? (Bitte in jeder Zeile eine Zahl einkreisen)

	Kein Problem	Sehr kleines Problem	Kleines Problem	Mäßiges Problem	Großes Problem
a. Hitzewallungen	0	1	2	3	4
b. Empfindliche/vergrößerte Brüste	0	1	2	3	4
c. Niedergeschlagenheit	0	1	2	3	4
d. Mangel an Energie	0	1	2	3	4
e. Veränderung des Körpergewichts	0	1	2	3	4

VIELEN DANK.

#### Appendix 4: Staging according to the German Guideline Prostate Cancer (version April 2018)

In this study we use the staging proposed by the German Guideline for Prostate Cancer [22]:

- Localized prostate cancer: T1-2, N0 M0.
  - Localized prostate cancer with low risk: PSA ≤ 10 ng/ml and Gleason-Score 6 and cT1c or cT2a.
  - Localized prostate cancer with intermediate risk: PSA > 10 ng/ml–30 ng/ml or Gleason 7 or cT 2b.
  - Localized prostate cancer with high risk: PSA > 20 ng/ml or Gleason ≥ 8 or cT2c.
- Locally advanced prostate cancer: T3-4 N0 M0.
- Advanced prostate cancer: any T N1 and M0.
- Metastasized prostate cancer: any T any N and M1.

Furthermore, localized prostate cancer with cT1a or cT1b is classified as localized prostate cancer with low risk.

#### References

1. Zentrum für Krebsregisterdaten and Gesellschaft der epidemiologischen Krebsregister in Deutschland (2017) Krebs in Deutschland für 2013/2014. Robert Koch-Institut, Berlin
2. Trama A et al (2015) Survival of male genital cancers (prostate, testis and penis) in Europe 1999-2007: results from the EURO-CARE-5 study. Eur J Cancer 51(15):2206–2216
3. Litwin MS et al (1995) Quality-of-life outcomes in men treated for localized prostate cancer. JAMA 273(2):129–135
4. Madalinska JB et al (2001) Health-related quality-of-life effects of radical prostatectomy and primary radiotherapy for screen-detected or clinically diagnosed localized prostate cancer. J Clin Oncol 19(6):1619–1628
5. Sanda MG et al (2008) Quality of life and satisfaction with outcome among prostate-cancer survivors. N Engl J Med 358(12):1250–1261
6. Wei JT et al (2000) Development and validation of the expanded prostate cancer index composite (EPIC) for comprehensive

- assessment of health-related quality of life in men with prostate cancer. *Urology* 56(6):899–905
7. Szymanski KM et al (2010) Development and validation of an abbreviated version of the expanded prostate cancer index composite instrument for measuring health-related quality of life among prostate cancer survivors. *Urology* 76(5):1245–1250
  8. Schmidt S et al (2014) Assessing quality of life in patients with prostate cancer: a systematic and standardized comparison of available instruments. *Qual Life Res* 23(8):2169–2181
  9. Lam WWT et al (2017) Psychometric assessment of the Chinese version of the abbreviated expanded prostate cancer index composite (EPIC-26) and the clinical practice version (EPIC-CP) in chinese men with prostate cancer. *J Pain Symptom Manage* 53(6):1085–1090
  10. Fossa SD et al (2016) Psychometric testing of the Norwegian version of the expanded prostate cancer index composite 26-item version (EPIC-26). *Scand J Urol* 50(4):280–285
  11. Vigneault E et al (2017) Validation of the French-Canadian version of the expanded prostate cancer index composite (EPIC) in a French-Canadian population. *Can Urol Assoc J* 11(12):404–410
  12. Beyer B et al (2015) “Expanded prostate cancer index composite” (EPIC-26): Funktionelles Behandlungsergebnis bei Patienten mit lokalisiertem Prostatakarzinom. *Urologe A* 54(11):1591–1595
  13. Umbehr MH et al (2018) The German version of the Expanded Prostate Cancer Index Composite (EPIC): translation, validation and minimal important difference estimation. *Health Qual Life Outcomes* 16(1):36
  14. Evans SM et al (2017) Cohort profile: the TrueNTH Global Registry: an international registry to monitor and improve localised prostate cancer health outcomes. *BMJ Open* 7(11):e017006
  15. Martin NE et al (2015) Defining a standard set of patient-centered outcomes for men with localized prostate cancer. *Eur Urol* 67(3):460–467
  16. ICHOM (2017) Localized prostate cancer data collection reference guide. <https://ichom.org/files/medical-conditions/localized-prostate-cancer/localized-prostate-cancer-reference-guide.pdf>. Accessed 23 Sep 2019
  17. Cronbach LJ (1951) Coefficient alpha and the internal structure of tests. *Psychometrika* 16:297–334
  18. Cohen J (1988) *Statistical power analysis for the behavioral sciences*. N.J.L. Erlbaum Associates, Hillsdale
  19. Nunnally CJ, Bernstein I (1994) *Psychometric Theory*. McGraw-Hill, New York
  20. Reeve Bryce B et al (2018) Symptom and function profiles of men with localized prostate cancer. *Cancer* 124(13):2832–2840
  21. Nunnally JC, Bernstein IH (1994) *Psychometric theory*, 3rd edn. McGraw-Hill, New York
  22. German Guideline Program in Oncology (2018) ed. Interdisziplinäre Leitlinie der Qualität S3 zur Früherkennung, Diagnose und Therapie der verschiedenen Stadien des Prostatakarzinoms, Version 5.0, 1. Aktualisierung, AWMF-Registernummer: 043/022OL
  23. Wei J, Dunn R, Litwin M, Sandler H, and Sanda M (2000) Development and validation of the expanded prostate cancer index composite (EPIC) for comprehensive assessment of health-related quality of life in men with prostate cancer. *Urology* 56:899–905

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### **4.2. Different Approaches for Case-Mix Adjustment of Patient-Reported Outcomes to Compare Healthcare Providers – Methodological Results of a Systematic Review**

Veröffentlicht als: **Sibert, N. T.**, Pfaff, H., Breidenbach, C., Wesselmann, S., & Kowalski, C. (2021). Different Approaches for Case-Mix Adjustment of Patient-Reported Outcomes to Compare Healthcare Providers-Methodological Results of a Systematic Review. *Cancers*, 13(16), 3964. <https://doi.org/10.3390/cancers13163964>

Systematic Review

# Different Approaches for Case-Mix Adjustment of Patient-Reported Outcomes to Compare Healthcare Providers—Methodological Results of a Systematic Review

Nora Tabea Sibert <sup>1,\*</sup>, Holger Pfaff <sup>2</sup>, Clara Breidenbach <sup>1</sup>, Simone Wesselmann <sup>1</sup> and Christoph Kowalski <sup>1</sup>

<sup>1</sup> German Cancer Society, Kuno-Fischer-Str. 8, 14507 Berlin, Germany; breidenbach@krebsgesellschaft.de (C.B.); wesselmann@krebsgesellschaft.de (S.W.); kowalski@krebsgesellschaft.de (C.K.)

<sup>2</sup> Institute of Medical Sociology, Health Services Research, and Rehabilitation Science, Faculty of Medicine and University Hospital Cologne, Faculty of Human Sciences, University of Cologne, 50933 Cologne, Germany; holger.pfaff@uk-koeln.de

\* Correspondence: sibert@krebsgesellschaft.de

**Simple Summary:** Patient-reported outcomes need to be reported with case-mix adjustment in order to allow fair comparison between healthcare providers. This systematic review identified different approaches to case-mix adjustment, with wide variation between the various approaches.

**Abstract:** Patient-reported outcomes (PROs) are increasingly being used to compare the quality of outcomes between different healthcare providers (medical practices, hospitals, rehabilitation facilities). However, such comparisons can only be fair if differences in the case-mix between different types of provider are taken into account. This can be achieved with adequate statistical case-mix adjustment (CMA). To date, there is a lack of overview studies on current CMA methods for PROs. The aim of this study was to investigate which approaches are currently used to report and examine PROs for case-mix-adjusted comparison between providers. A systematic MEDLINE literature search was conducted (February 2021). The results were examined by two reviewers. Articles were included if they compared (a) different healthcare providers using (b) case-mix-adjusted (c) patient-reported outcomes (all AND conditions). From 640 hits obtained, 11 articles were included in the analysis. A wide variety of patient characteristics were used as adjusters, and baseline PRO scores and basic sociodemographic and clinical information were included in all models. Overall, the adjustment models used vary considerably. This evaluation is an initial attempt to systematically investigate different CMA approaches for PROs. As a standardized approach has not yet been established, we suggest creating a consensus-based methodological guideline for case-mix adjustment of PROs.

**Keywords:** patient-reported outcomes; case-mix adjustment; healthcare provider comparison; quality improvement



**Citation:** Sibert, N.T.; Pfaff, H.; Breidenbach, C.; Wesselmann, S.; Kowalski, C. Different Approaches for Case-Mix Adjustment of Patient-Reported Outcomes to Compare Healthcare Providers—Methodological Results of a Systematic Review. *Cancers* **2021**, *13*, 3964. <https://doi.org/10.3390/cancers13163964>

Academic Editor: Massimo Di Maio

Received: 1 July 2021  
Accepted: 2 August 2021  
Published: 5 August 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

There is a growing body of evidence that patient-reported outcomes (PROs) are effective in improving cancer patients' health when used for patient monitoring or as a tool during patient consultations for better evaluation of patients' symptoms or functional impairment [1–4]. Following landmark studies by Basch et al. [5,6] and Denis et al. [7], many approaches for incorporating (electronic) PROs into routine clinical cancer care have recently been developed—although some approaches also date back 15 years or more [8,9]. Due to cost and organizational barriers [10], only a small proportion of patients are currently benefiting from these approaches, but PROs are on the wish list of many service providers and healthcare systems and can be expected to find their way into more and more national cancer plans in the near future [11].

In addition to these approaches based on individual patients, PROs can also be used for another purpose in routine clinical care—quality assurance—in the sense of quality

improvement initiatives. For this purpose, aggregated data at the level of healthcare providers are compared with those of other providers, resulting in rankings or benchmark reports. PROs are then regarded in a similar manner as other quality metrics: if suboptimal outcomes are observed at the provider level, the providers are typically asked to use these comparisons as part of their quality management in order to improve care. This function has been prominently advocated by, among others, the International Consortium on Health Outcomes Management (ICHOM), for whom PROs are a central part of the standard datasets used to allow comparisons. Building on the ICHOM standard dataset for localized prostate cancer [12], for example, the multinational TrueNorth Global Registry has been set up [13], with over 200 participating sites in 15 countries and with national substudies, e.g., [14]. It has been argued in the past that these two functions (with approaches at the individual patient level and at the provider level) can work hand in hand, but examples of this are as yet scarce [15].

When PROs are used as quality metrics to compare healthcare providers, special attention needs to be given to the different case-mixes present in order to allow fair reporting [16]. If PROs were to be reported without adjustment, services with less severely affected patients or otherwise favorable patient characteristics, such as fewer comorbidities or earlier disease stages, might appear to be better than they actually are, and vice versa. This applies to quality metrics in general and may lead to inappropriate incentives such as refusal to treat specific patients—especially if the providers are ranked and the results are publicly reported, with the providers' names not being concealed. The practice of refusing patients in order to avoid costs or unfavorable outcomes in competitive healthcare markets is often referred to as “cream skimming” and is highly applicable to PRO reporting. According to Iezzoni's framework of case-mix adjustment (CMA), a healthcare outcome can be described as a function of “intrinsic patient-related risk factors, treatment effectiveness, quality of care” and “random chance” [16]. Thus, when PROs are used to compare providers, effects summarized as “intrinsic patient-related risk factors” must not distort the comparison results. CMA (which is sometimes more generally referred to as “risk adjustment”) aims to account for such differences in providers' case-mixes using a statistical procedure. The goal of CMA is to allow fair and reliable comparison of providers relative to the given outcome(s). Its aim is to allow “like-to-like comparisons,” as described by Iezzoni in her introductory chapter describing the purpose of CMA [16]. Ultimately, CMA is essential for the credibility and usefulness of provider comparisons.

Different approaches to CMA have been used in the past for quality metrics other than those based on PROs, and differences in approaches may lead to substantial differences in the reported outcomes and in the ranking of providers and services. There have been few publications on how to adjust the case-mix for PROs, and there is a lack of systematic evaluation of published and proposed CMA approaches. The aim of the present systematic review was to identify methods used for CMA in provider comparisons using PROs, with a particular focus on the adjusters included in the CMA models and the statistical approaches used. Recommendations on the choice of future CMA methodology—as suggestions for a possible CMA guideline—are also derived from this.

## 2. Methods

### 2.1. Studies and Methodological Decisions of Interest

This systematic review adapted the participants, interventions, comparisons, outcomes, study designs (PICOS) template, with omission of the intervention (I) and comparator (C) categories, which were not applicable here. Any patients treated by various healthcare providers were therefore examined as participants (P), PROs as outcomes (O), and observational studies and secondary analyses of interventional studies were included as study designs (S).

Iezzoni [16] proposed four main questions that can be used for guidance in finding the “right” CMA approach:

1. Risk of what outcome?
2. Over what time frame?
3. For what population?
4. For what purpose?

These questions (referred to as “Iezzoni’s four”) are therefore used in this review to evaluate the different CMA approaches together with the adjustors used, the criteria with which the adjustors were chosen, the statistical model used, and the way in which the results were reported.

## 2.2. Eligibility Criteria and Information Sources

This review includes observational studies or secondary analyses of interventional studies published between January 1992 and January 2021 written in English, French, or German, in which PROs are used to compare the quality of care between recognizable healthcare providers and in which any kind of case-mix adjustment of PROs is used. PROs were considered to be “instruments to elicit information from patients on self-reported health status [and] outcomes” in accordance with the definition provided by the Organization for Economic Cooperation and Development (OECD) [17]. Studies were included if they used at least one adjustor that was not the PRO baseline score. Exclusion criteria were: studies that did not use (case-mix-adjusted) PROs for provider comparison, or had no provider comparison or no CMA or did not explain their CMA; those that used patient-reported experience measures (including satisfaction surveys) as opposed to the above definition of PROs (patient-reported *experience* measurements measure—in contrast to PROs—“how patients experience health care and refer to practical aspects of care, such as accessibility, care coordination and provider-patient communication” [17] and are as such not healthcare outcomes). The electronic database MEDLINE was used as the information source. Citations and references in the studies included were also screened for eligibility. The review was registered at PROSPERO (CRD42020180420).

## 2.3. Search and Study Selection

The search string was developed iteratively, with test searches being conducted in the MEDLINE database by N.T.S. and C.K. The study group developed the final search strategy with support from an experienced systematic reviewer, combining independent and medical subject headings (MeSH) terms in accordance with the PRESS guideline statement [18]. The final search string contained variations of the terms: “humans” (P), “patient-reported outcomes” or “patient-reported outcome measures” (O), and “risk adjustment” or “case-mix adjustment.” The full search string can be found in Appendix A.

C.K. and N.T.S. first independently screened all titles and abstracts. N.T.S. searched for available full texts, which were then independently reviewed by C.K. and N.T.S. relative to the inclusion criteria. Any divergent results were discussed by C.K. and N.T.S. to obtain a consensus.

## 2.4. Data Collection Process, Data Items, and Risk of Bias in Individual Studies

C.K. and N.T.S. independently used a standardized data extraction form that was developed deductively using Iezzoni’s earlier research on risk adjustment [16]. Data items included in the extraction form were: type of provider, Iezzoni’s four, information on adjustors (which ones, and how they were chosen), CMA methodology, and reporting of case-mix-adjusted PROs. In addition, quality assessment was carried out using the Critical Appraisal Skills Programme (CASP) checklist for cohort studies [19].

## 2.5. Summary Measures and Synthesis of Results

Data synthesis was carried out by qualitatively summarizing the extracted information on statistical methods used for CMA, adjustors, and reporting of provider comparisons. Since there was wide heterogeneity among the CMA approaches and PROs applied, it was not possible to pool the findings quantitatively as a meta-analysis. However, where

possible, similar case-mix adjustment approaches are presented in bundled form. No additional subgroup analysis was carried out in any of the studies.

### 2.6. Risk of Bias across Studies

As the purpose of this methodological review was to provide an overview of the statistical methods used for adjustment, and the studies included were conducted in different populations, over different time periods, and with different PROs, it was not possible to carry out a meta-analysis. Evaluation of the risk of bias among the studies included was therefore not applicable.

## 3. Results

### 3.1. Study Selection and Characteristics

A total of 640 studies were identified in MEDLINE using the search string. After screening of the titles/abstracts and full text, nine studies were included for the qualitative synthesis (Figure 1: PRISMA flowchart). Two studies were added after the references from these nine studies had been screened. Thus, a total of eleven studies were included in this analysis. An overview on study characteristics can be found in Table 1.

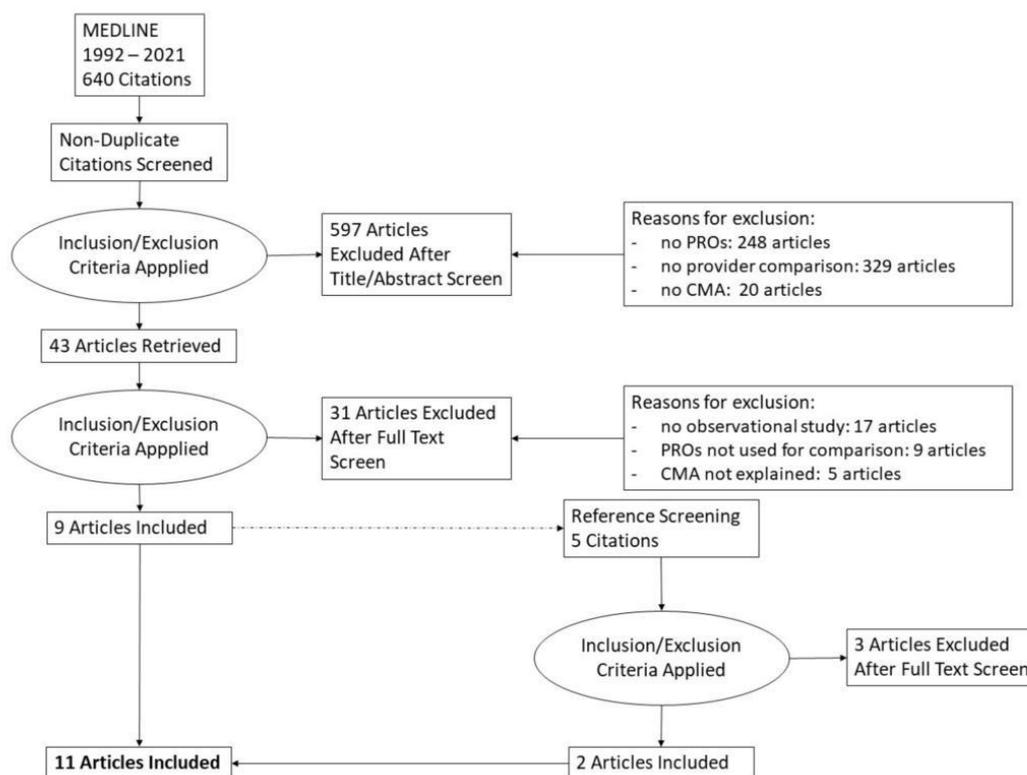


Figure 1. PRISMA flowchart. PRO, patient-reported outcome; CMA, case-mix adjustment.

### 3.2. Quality of Included Studies (CASP Evaluation)

All studies had a clear focus, precise and mostly applicable results, and fitted with other evidence. Follow-up information was partly insufficient or could not be examined. Details are provided in the CASP results Table A1 in Appendix B.

### 3.3. Provider Settings

Five inpatient settings [20–24], five outpatient settings [25–29], one mental health services setting [26], and three rehabilitation settings [20,28,29] were identified. Ten studies included patients with orthopedic conditions [20–25,27–30]. Three of the studies included were set in National Health Service-funded hospitals in the United Kingdom [21,23,30]. Three studies were set in hospitals in the United States, taking part in the Focus on Therapeutic Outcomes (FOTO) initiative [25,28,29].

### 3.4. “Iezzoni’s Four”

This synthesis follows Iezzoni’s approach on analyzing risk adjustment. Iezzoni’s four main questions are therefore covered in the following sections [16]. Detailed information is provided in Table 2.

#### 3.4.1. Risk of What Outcome?

The studies included a total of 17 different PRO instruments: seven generic PRO questionnaires (PCS [31], PF-10 [32], SF-12 [31], ADL [33], Lehman Quality of Life Interviews [34], EQ-5D [35], including one rehabilitation-specific one: IRES [36]) and ten condition-specific instruments (VAS-BP, VAS-LP [37], OHS, OKS [38], MDQ [39], NDI [40], ODI [41], MacNew [42], SMFA [43], and LCAT [44]).

#### 3.4.2. For What Population?

All of the studies included explored PROs for an adult population. Ten studies were conducted in orthopedic settings [20–25,27–30], and one investigated patients with psychiatric conditions [26]. None of the studies that were identified investigated PROs in populations of cancer patients.

#### 3.4.3. Over What Time Period?

Most of the studies examined the development of PRO scores from baseline (e.g., at admission, before the start of therapy) and a predefined or treatment success-defined post-therapeutic time point. In some studies, the time period of PRO data collection was not reported or specified [24,27,28]—for example, if the post-therapeutic end point was measured at the end of an intervention, of which the duration could vary, e.g., [28].

#### 3.4.4. For What Purpose?

The main purpose of the studies included was to obtain inter-provider comparisons to measure performance, in some cases resulting in a (clinical) ranking [25] or in the identification of special expertise [28].

### 3.5. Selection of Adjustors

There are several approaches for choosing adjustors for CMA. The studies included only partly reported on the selection process. Two studies stated that they chose the adjustors that were used on the basis of statistical parameters: Resnik and Hart used bivariate analysis to identify suitable adjustors [28], and Deutscher et al. employed a back-step regression model [25]. Farin et al. and Khor et al. based their choice of adjustors on earlier findings by their research groups [20,22]. Table 2 provides further information.

### 3.6. Patient Characteristics Included in the CMA Models

All of the studies included sociodemographic characteristics (such as sex, age) of their patients and baseline PRO scores as adjustors in the CMA models used. All of the

studies chose comorbidities and/or a measure of the patient's general physical status (i.e., specific comorbidities, number of comorbidities, Charleston Comorbidity Index, American Society of Anesthesiologists score, etc.) as adjustors. Most of the studies that were included featured disease-specific information or information about treatments previously received by the patients.

### 3.7. Provider Characteristics

Lutz et al. included provider characteristics (the American state in which the study was located) as an adjustor and Sivaganesan et al. included information about surgeons (years in practice, fellowship training) [24,27].

### 3.8. Statistical Approaches

The statistical approaches used differed in relation to the intended presentation of the adjusted results. Some studies calculate scores that keep the scale of the initial PRO score, while others keep the scale but present “performance indicators”—typically the difference between the observed score and the score expected on the basis of the patient case-mix. These approaches are referred to in this article as “unit/scale-preserving approaches.” Other studies compute ratios of the observed scores divided by the expected scores, or only calculate expected scores. These approaches thus do not present results that are consistent with the units of the original PRO score and are referred to in this article as “unit/scale-aberrant approaches.”

A common differentiation made in the statistical models used for CMA is that between *direct* and *indirect* standardization [45]. Briefly, direct approaches use a matched subsample from each provider to compare the outcomes for those patients with those in a standardized population, whereas indirect approaches use the whole sample from the providers and adjust the outcomes using a modeling procedure, typically regression analyses. All of the studies included in this review used indirect standardization methods. However, they differ in the way in which the adjusted scores are calculated and presented. These differences are therefore examined in greater depth in the following paragraphs.

Since both the topic itself and differences in CMA methods are quite complex, the different statistical approaches are illustrated by using an example based on the Prostate Cancer Outcomes (PCO) Study data [14] (*described in italics*):

*To compare outcomes of prostate cancer centers one year after treatment—e.g., surgery—certified prostate cancer centers take part in a survey collecting information on EPIC-26 scores. The EPIC-26 is a prostate cancer-specific PRO questionnaire resulting in five scores, among them an incontinence score—with 0 being the lowest and 100 the highest possible score [46]. For our example, we would like to compare Center A, Center B, and Center C using the following metrics:*

*Unadjusted—i.e., observed—mean incontinence scores (1 year after surgery): 26 (Center A), 37 (Center B), 31 (Center C), 30 (all patients in all centers)*

*Expected mean incontinence scores based on linear regression models (using the following adjustors: baseline EPIC-26 scores, age, tumor stage, etc.): 20 (Center A), 33 (Center B), 40 (Center C).*

#### 3.8.1. Calculation of Transformed Adjusted Scores (Unit/Scale-Preserving Approaches)

Two studies used the observed and previously calculated expected values to transform them into adjusted scores that can be interpreted using the initial PRO score's unit [23,28]: Nuttal et al. firstly calculated the ratio  $\frac{\text{observed}}{\text{expected}}$  for every provider and multiplied those ratios with the overall population's observed average [23]:

$$\text{adjusted score} = \frac{\text{observed}}{\text{expected}} * \text{overall observed average.} \quad (1)$$

*In our example, the adjusted scores were:*

$$\text{Center A: } \frac{26}{20} * 30 \approx 39$$

$$\text{Center B: } \frac{37}{33} * 30 \approx 33$$

$$\text{Center C: } \frac{30}{40} * 30 \approx 26$$

These scores are all within the original EPIC-26 limits (0–100) and are interpreted identically. The adjusted scores provide information about what the incontinence score of an “average” prostate cancer patient would be in the respective center.

Resnik and Hart calculated the differences between observed and expected values [28]. The resulting adjusted (or “performance”) scores were not within the limits of the original PRO scale (e.g., for a positive PRO scale, the differences might become negative), but they are interpretable using the original units:

$$\text{adjusted score} = \text{expected} - \text{observed} \quad (2)$$

Deutscher et al. and Farin et al. used the standardized residuals of the final case-mix regression models as “adjusted outcomes” [20,25]. This approach is therefore very similar to the adjustment method used by Resnik and Hart (residuals are defined as the difference between the observed and expected (“fitted”) values in regression analyses).

Using the Resnik and Hart approach, we obtain the following adjusted scores using our example:

$$\text{Center A: } +6$$

$$\text{Center B: } +4$$

$$\text{Center C: } -10$$

These adjusted scores provide information about how the centers perform in relation to their expected outcomes (e.g., Center C is performing less well than expected, whereas Center A is achieving better results) using the same unit as in EPIC-26. It would, for example, be possible to use the EPIC-26 minimally important differences (MIDs; lower threshold for incontinence: 6 [47]) to interpret these results better: in Center B, the “better performance” would be less than one MID, and therefore not perceptible for the patients, whereas a patient treated in Center C would notice the poorer performance.

Varagunam et al. used a similar approach to Nuttal et al., but based on the residuals, firstly, they calculated the difference in the observed and the expected value (ergo, the residuals) for every provider and then added this to the overall population’s observed average [30]:

$$\text{adjusted score} = (\text{expected} - \text{observed}) + \text{overall observed average}. \quad (3)$$

In our example, the adjusted scores were:

$$\text{Center A: } (26 - 20) + 30 = 36$$

$$\text{Center B: } (37 - 33) + 30 = 34$$

$$\text{Center C: } (40 - 30) + 30 = 20$$

These scores are all within the original EPIC-26 limits (0–100) and are interpreted identically. The adjusted scores provide information about what the incontinence score of an “average” prostate cancer patient would be in the respective center.

### 3.8.2. Calculation of Transformed Adjusted Scores (Unit/Scale-Aberrant Approaches)

Two studies calculated the ratio  $\frac{\text{observed}}{\text{expected}}$  and used this “ratio indicator” as an adjusted outcome [26,27].

The ratio indicators for the PCO examples are:

$$\text{Center A: } \frac{26}{20} \approx 1.3$$

$$\text{Center B: } \frac{37}{33} \approx 1.12$$

$$\text{Center C: } \frac{30}{40} \approx 0.75$$

These adjusted scores are not interpretable using the original EPIC-26 unit. However, they quantify the deviation from the expected outcome: Centers A and B are both performing better than expected, resulting in a ratio > 1, whereas Center C is performing less well than expected, with a ratio < 1.

One study investigated the proportion of patients who experienced an improvement of at least one minimally important difference (MID) as the outcome of interest—i.e., they dichotomized the original PRO score [22]. They then calculated the expected number of events per provider on the basis of a previously published risk calculator, calculated the observed/expected ratio, and finally calculated the case-mix-adjusted event rate by multiplying the O/E ratio by the overall state-wide average.

*The MID for the EPIC-26 incontinence score is 6, as described above. Let us assume that 20% of the patients of Center A, 40% of the patients of Center B, and 30% of the patients of Center C had an improvement of at least 6 points concerning their incontinence score after surgery. Applying the risk calculator for the centers' populations, for Center A, 18% would have been expected, for Center B 35%, and for Center C 50%. For the state-wide population, an improvement of 30% is already known (this is a theoretical example). Using the "MID approach," the adjusted scores would be:*

$$\begin{aligned} \text{Center A: } & \frac{20}{18} * 30 \% \approx 27 \% \\ \text{Center B: } & \frac{40}{35} * 30 \% \approx 34 \% \\ \text{Center C: } & \frac{30}{50} * 30 \% = 18 \% \end{aligned}$$

### 3.8.3. Other

Sivaganesan et al. calculated the expected values for the providers examined, on the basis of their case-mix. However, the authors did not seem to use the expected values to compute any further "transformed" adjusted scores. They used a multilevel model [24].

*For our example, the "adjusted scores" would be the same as the expected ones:*

*Center A: 20*

*Center B: 33*

*Center C: 40*

*Center C, which performed 10 points less well than expected, would be top ranked following this approach.*

Gutacker et al. used another approach: based on a multilevel model using an hierarchical Bayesian approach with the PRO score (EQ-5D) used as an ordinal scaled variable, the predicted probabilities for each outcome per provider were calculated and reported as a quality metric [21].

Using a multilevel model as well, Gozalo et al. reported a provider ranking based on the predicted providers' random intercepts [29].

### 3.8.4. Minimum Case Number Requirements for Statistical Modeling/ Adjustment

Seven studies did not report any minimum case number requirements at the provider level for adjustment. The other studies differed widely in the minimum numbers of patients per provider that were required for adjustment. Gozalo et al. and Resnik and Hart set a minimum of eight patients per provider [28,29], whereas Farin et al. required at least 200 patients per provider [20]. Deutscher et al. only included providers in which at least 50% of the patient records were complete [25]. Most of the studies did not provide any explanations for the thresholds chosen.

Table 1. Characteristics of eligible studies.

Author	Study Design	Participants/Setting	Data Sources	Study Size (Complete/Included Datasets)	Outcome (PRO)	No. of Variables Used for Case-Mix Adjustment (and Specific Adjustors)	Model R <sup>2</sup>
Deutscher et al., 2018 [25]	Retrospective analysis of prospectively collected data	Adult patients with lower back pain treated in outpatient physical therapy clinics (2014–2016)	FOTO	n = 341,642 patients n = 6934 clinicians n = 2107 hospitals	LCAT	11: LCAT score at admission, age, sex, acuity, payer, surgical history, physical exercise history, medication/previous treatment, type of surgery, comorbidities	0.373
Farin et al., 2009 [20]	Analysis of quality assurance data	Quality assurance report for statutory rehabilitation institutions in Germany on: 1. musculoskeletal; 2. cardiological; and 3. neurological diseases	QS Reha®	1: n = 4045 patients in n = 27 hospitals 2: n = 2503 patients in n = 25 hospitals 3: n = 1477 patients in n = 12 hospitals	1: IRES 2: SMFA 3: MacNew	Reported elsewhere [48]: 10: age, sex, somatic, functional and psychosocial status at admission, diagnosis, multimorbidity, treatment motivation, application for pension, education	Not reported
Gozalo et al., [29]	Retrospective analysis of prospectively collected data	Outpatient rehabilitation clinics for patients with lower back pain	FOTO	n = 90,392 patients n = 2040 therapists n = 538 clinics	FOTO overall health status measure: OHS, PCS, PF-10	8: PRO at intake, age, gender, onset, surgery count, functional comorbidity index, fear avoidance level, payer type	OHS: 0.416 PCS: 0.345 PF-10: 0.421
Gutacker et al., 2013 [21]	Retrospective analysis of prospectively collected data	Analysis of hospital performance variation in hip replacement surgery	NHS PROMs and routinely collected inpatient data (HES)	n = 21,565 patients n = 153 hospitals	EQ-5D	10: pre-treatment PROs, age, gender, deprivation index, weighted Charlson Index of comorbidities, number of additional comorbidities (not included in Charlson Index), time between pre-operative assessment and admission, primary surgery, revision surgery, treatment	Not reported
Hendryx et al., 1999 [26]	Retrospective analysis of prospectively collected data	Development of outcome risk adjustment models for public mental health outpatient treatment programs	Patient-reported, case manager ratings, management information system data	n = 289 patients n = 6 mental health agencies	SF-12, ADL, Lehman Quality of Life Interviews	8: sex, age, race, presence of severe primary diagnosis (major depression, schizophrenia, bipolar disorder), baseline levels of substance abuse, baseline PRO, baseline quality of life, baseline satisfaction with services	0.34
Khor et al., 2020 [22]	Retrospective analysis of prospectively collected data	Analysis of variation in PROs 1 year after elective lumbar fusion surgery across surgeons and hospitals	Spine Care and Outcomes Assessment Program (Washington State, USA)	n = 737 patients n = 17 hospitals n = 58 surgeons	ODI	11: age, sex, insurance status, race, ASA, smoking status, prior spine surgery, diagnosis, opiate use, asthma, baseline PRO	Not reported

Table 1. Cont.

Author	Study Design	Participants/Setting	Data Sources	Study Size (Complete/Included Datasets)	Outcome (PRO)	No. of Variables Used for Case-Mix Adjustment (and Specific Adjustors)	Model R <sup>2</sup>
Lutz et al., 2020 [27]	Retrospective analysis of prospectively collected data	Benchmarking physical therapist study	ATI Patient Outcomes Registry	n = 182,276 patients n = 2799 physical therapists	MDQ, NDI	10: sex, age, BMI, initial PRO, payer type, physical component score and mental component score of the Veterans RAND 12-Item Health Survey, state of physical therapy services, type of pain (acute/chronic), comorbidities	MDQ: 0.19 NDI: 0.19
Nuttall et al., 2015 [23]	Retrospective analysis of prospectively collected data	Feasibility study to analyze case-mix adjustment methodology for elective surgery	NHS PROMs and HES	n = 30,555 patients n = 237 providers	OKS	11: baseline PRO, sex, ethnicity, age, deprivation index, assistance for completing questionnaires (baseline and follow-up), disability, previous surgery, comorbidities, type of surgery, length of post-operative stay	0.258 (ordinary least square model) 0.257 (fixed effects model)
Resnik & Hart 2003 [28]	Retrospective analysis of prospectively collected data	Outpatient rehabilitation clinics for patients with lower back pain	FOTO	n = 24,276 patients n = 930 therapists n = 354 hospitals	FOTO overall health status measure (OHS), PCS, PF-10	8: age, employment, exercise history, sex, intake PRO, onset, reimbursement, surgery	OHS: 0.416 PCS: 0.345 PF-10: 0.421
Sivaganesan et al., 2018 [24]	Retrospective analysis of prospectively collected data	Report of a risk-adjusted ranking of spine surgeons and sites performing elective lumbar surgery	Quality and Outcome Database (QOD)	n = 8834 patients n = 124 surgeons n = 21 sites	ODI, EQ-5D, VAS-BP/VAS-LP	22 (19 at patient level: age, BMI, ethnicity, education, smoking status, opioid use, comorbidities, pre-operative symptoms, motor deficit, ASA, symptom duration, interbody graft placement, worker's compensation, liability claims, insurance status, employment, baseline PRO); 3 constructs at surgeon level: site ID, years in practice, fellowship training)	Not applicable (Bayesian model)
Varagunam et al., 2015 [30]	Retrospective analysis of prospectively collected data	Feasibility study analyzing different approaches on how to adjust PROM scores for individual consultant comparison for elective surgery	NHS PROMs	N = 65,465 (hip), 68,107 (knee) and 38,965 (hernia) patients N = 948 (hip), 1130 (knee) and 974 (hernia) consultants N = 183 (hip), 188 (knee) and 197 (hernia) clinics	OHS, OKS, EQ-5D	6: age, sex, deprivation index, comorbidities, previous surgery, baseline PRO	Not reported

ADL, Activities of Daily Living; ASA, American Society of Anesthesiologists score; ATI, Athletic Therapeutic Institute; BMI, Body Mass Index; EQ-5D, European Quality of Life 5 Dimensions; FOTO, Focus on Therapeutic Outcomes; HES, Hospital Episode Statistics; IRES, Indicators des Reha-Status (Indicators for Rehabilitation Status); LCAT, Lumbar Computerized Adaptive Test; MDQ, Modified Low Back Pain Disability Questionnaire; NDI, Neck Disability Index; NHS, National Health Service (UK); ODI, Oswestry Disability Index; OHS, Oxford Hip Score; OKS, Oxford Knee Score; PCS, Physical Component Summary; PF-10, Physical Functioning Scale; PRO, Patient-Reported Outcome; PROM, Patient-Reported Outcome Measure; SF-12, Short Form 12; SMFA, Short Musculoskeletal Function Assessment questionnaire; VAS-BP/VAS-LP, Visual Analog Scale for Back Pain/for Leg Pain.

Table 2. Synthesis of case-mix adjustment methods.

Author	What Type of Provider?	Lezzoni's Four				Which Adjustors?				Are Criteria for the Selection of Adjustors Met? <sup>1</sup>		How Is Case-Mix Adjustment Statistically Performed?	How Are Case-Mix Adjustment Results Reported?
		Risk of What Outcome?	For What Population?	Over What Time Period?	For What Purpose?	Socio-Economic Information?	Baseline PROM Included?	Comor-Bidities?	Adjustment for Provider Characteristics?	Are the Adjustors Associated with the Outcome?	Do Adjustors Vary Across Providers?		
Deutscher et al., 2018 [25]	Outpatient physical therapist clinics (US)	LCAT	Lower back pain patients treated in outpatient setting	Time between admission and discharge	Clinical ranking for performance measuring	x	x	x	-	Yes, assessed via back-step regression model approach	Not analyzed	Multiple regression, >50% complete patients per hospital and ≥10 complete cases per clinician per year required	Comparison between adjusted and unadjusted scores (agreement: 30%); ranking, changed 70% of the hospitals, reported as percentile ranking
Farin et al., 2009 [20]	In- and outpatient rehab. institutions (Germany)	IRES, SMFA, MacNew	Patients in statutory rehab. institutions	Time between start and end of rehab.	Quality assurance	x	x	x	-	Not reported	Not reported	Multilevel regression, taking the standardized residuals in the prediction as the case-mix-adjusted outcome, >199 patients per institution required	95% confidence intervals reported graphically
Gozalo et al., 2016 [29]	Outpatient rehabilitation clinics (US)	FOTO CAIs, either generic or body impairment specific	Patients treated in outpatient rehabilitation clinics (for orthopedic conditions)	At intake and discharge from outpatient rehabilitation	Identify expert physical therapists	x	x	x	-	Not assessed	Not reported	Generalized linear model with >7 patients per clinic required	Ranking of clinics based on predicted clinic random intercept (including 95% confidence interval)
Gutacker et al., 2013 [21]	NHS-funded hospitals (UK)	EQ-5D	Patients receiving hip replacement in an NHS-funded hospital	Time between before/day of admission and 6 months after surgery	Inter-provider comparison for performance measuring	x	x	x	-	Yes, assessed via regression modeling	Not reported	Multilevel linear regression and probit model, no minimum number of patients per hospital required	95% confidence intervals reported graphically
Hendryx et al., 1999 [26]	Outpatient mental health services (US)	SF-12 (functional status domain), Lehman Quality of Life Interview (Quality of Life domain)	Adult patients treated in outpatient mental health agencies	At baseline, 5 and 10 months	Inter-provider comparison for performance measuring	x	x	x	-	Not reported	Not assessed	Multiple linear regression, then calculating the ratio: observed/expected, no minimum number of patients per agency required	Observed score ranks and observed/expected ratio ranks are compared

Table 2. Cont.

Author	What Type of Provider?	Lezzoni's Four				Which Adjustors?				Are Criteria for the Selection of Adjustors Met? <sup>1</sup>		How Is Case-Mix Adjustment Statistically Performed?	How Are Case-Mix Adjustment Results Reported?
		Risk of What Outcome?	For What Population?	Over What Time Period?	For What Purpose?	Socio-Economic Information?	Baseline PROM Included?	Comor-Bidities?	Adjustment for Provider Characteristics?	Are the Adjustors Associated with the Outcome?	Do Adjustors Vary Across Providers?		
Khor et al., 2020 [22]	In-patient hospitals performing elective spine surgery (USA)	ODI (probability of improvement beyond MID and reaching minimal disability level)	Adults treated with elective lumbar fusion surgery	At 0-60 days before surgery and at 12-month follow-up	Inter-provider comparison for performance measuring	x	x	x	-	Yes (association based on findings from earlier studies by the research group)	For all adjustors except for age, sex, and smoking status	1. Multiple logistic model 2. Calculation of an expected number of events per provider based on a previously published risk calculator 3. Calculation of observed/expected ratio 4. Calculation of a case-mix-adjusted event rate by multiplying the O/E ratio by the overall state-wide average >9 patients per provider required	Adjusted and unadjusted scores, ICCs to describe the proportion of total variability accounted for by between-provider variance
Lutz et al., 2020 [27]	Outpatient physical therapists (USA)	MDQ, NDI	Adults with episodes related to the lower back and neck with no history of related surgery	Not specified	Benchmarking physical therapists	x	x	x	x (state of physical therapy services)	Not reported	Not assessed	1. Multiple-linear regression 2. Calculation of an observed/expected ratio; >39 patients per physical therapist required	Reporting of counts of "outperforming, meeting, and underperforming" physical therapists (adjusted and unadjusted) based on 95% confidence intervals
Nuttall et al., 2015 [23]	NHS-funded hospitals (UK)	OKS	Patients treated by NHS providers for unilateral knee replacement	Before and 3 or 6 months after surgery	Inter-provider comparison for performance measuring	x	x	x	-	Yes (association based on findings from earlier studies)	Not analyzed	Multilevel linear regression, generalized least squares with fixed effects, general least squares with random effects, ordinary least squares, then calculating (observed/expected) * observed mean overall, no minimum number of patients per hospital required	Differences between different adjustment scores reported; ranking reported as funnel plots
Resnik & Hart 2003 [28]	Outpatient rehab. clinics (USA)	OHS, PCS, PF-10	Patients treated in outpatient rehab. clinics for lower back pain	At intake and discharge from outpatient rehabilitation	Identify expert physical therapists	x	x	x	-	Yes, bivariate analysis to identify adjustors	Not reported	Generalized linear model with >7 patients per therapist required, then calculating the difference: observed-expected; no minimum number of patients per clinic required	Percentiles of differences reported

Table 2. Cont.

Author	What Type of Provider?	Iezzoni's Four				Which Adjustors?			Are Criteria for the Selection of Adjustors Met? <sup>1</sup>			How Is Case-Mix Adjustment Statistically Performed?	How Are Case-Mix Adjustment Results Reported?
		Risk of What Outcome?	For What Population?	Over What Time Period?	For What Purpose?	Socio-Economic Information?	Baseline PROM Included?	Comor-Bidiities?	Adjustment for Provider Characteristics?	Are the Adjustors Associated with the Outcome?	Do Adjustors Vary Across Providers?		
Sivaganesan et al., 2018 [24]	Sites performing spine surgery (USA)	ODI, EQ-5D, VAS-BP/VAS-LP	Patients treated with spine surgery	Before and 1 year after surgery	Inter-provider comparison for performance measuring	x	x	x	x	Yes, but no association coefficients are reported	Not reported	Random effects regression models and multilevel hierarchical Bayesian models (3 levels); no minimum number of patients per hospital required	Box plots for adjusted ranks per surgeon per site reported
Vargunam et al., 2015 [30]	NHS-funded hospitals (UK)	OHS, OKS, EQ-5D	Patients treated by consultant in NHS providers	Before and three or six months after surgery	Inter-consultant comparison for performance measuring	x	x	x	-	Not assessed	Not reported	Multiple linear regression, >39 patients per NHS provider and >9 patients per consultant required; multilevel models	Number of consultants reported that perform better than expected, as expected, and worse than expected

EQ-5D, European Quality of Life 5 Dimensions; IRES, *Indikatoren des Reha-Status* (Indicators for Rehabilitation Status); MDQ, Modified Low Back Pain Disability Questionnaire; MID, Minimally Important Difference; NDI, Neck Disability Index; NHS, National Health Service (UK); ODI, Oswestry Disability Index; OHS, Oxford Hip Score; OKS, Oxford Knee Score; PCS, Physical Component Summary; PF-10, Physical Functioning Subscale (of Short Form-36 questionnaire); SMFA, Short Musculoskeletal Function Assessment; VAS-BP, Visual Analog Scale for Back Pain; VAS-LP, Visual Analog Scale for Leg Pain.  
<sup>1</sup> According to Iezzoni, an eligible adjustor should 1. be associated with the outcome and 2. vary across providers [16].

#### 4. Discussion

The results of the present systematic literature review on how to case-mix adjust PROs for provider comparison show that there is substantial variation in the literature with regard to several methodological criteria.

Firstly, the review shows that a uniform approach to case-mix adjustment of patient-reported outcomes does not currently exist. In most cases, regression models were used to adjust for relevant patient characteristics that differed between providers, with expected outcomes being calculated. Five of the eleven studies that were included made comparisons between expected and observed values, and three of these studies ultimately calculated an outcome that was evaluable in terms of the scale of the original one. One study seemed to only calculate fitted values and based the resulting provider ranking on them, which seems to be a questionable adjustment approach since providers with a “favorable” case-mix (e.g., younger or healthier patients) will always perform better than providers with a “difficult” case-mix if using this CMA approach. This can result in a biased ranking and does not provide information on the actual quality of care.

Secondly, there were also considerable differences between the studies included in the way in which expected values were calculated, and thus in the selection of adjusters. Relevant sociodemographic information and important clinical parameters (e.g., comorbidities) were included in most studies, and baseline patient-reported outcomes were included in all of them. There is little debate on including baseline PROs, as it is known from many observational and validation studies that PROs can vary considerably between care providers before the start of treatment even when adjustment is made for other patient characteristics [49]. It is thus essential to take baseline PROs into account in the adjustment.

Thirdly, nine of the studies included only reported patient characteristics, while two also included the characteristics of the providers that were being compared, such as information about the training of surgeons. Not taking provider characteristics into consideration is consistent with the purpose of CMA, since if outcomes are a function of patient characteristics, treatment effectiveness, and quality of care, and if the aim of the survey is to compare treatment quality, then considering provider characteristics may lead to over-adjustment. However, there are situations in which it might be justifiable to take provider characteristics into account—for example, if some contextual factors (such as state legislation) affect the outcomes that are exogenous, beyond the control of the provider. In principle, it is preferable to investigate provider characteristics in a further step in order to explain any differences in treatment quality in greater detail. This was done, for example, by Resnik and Hart [28], who looked at associations between provider characteristics and whether a practitioner was classified as an “expert physical therapist.” They found no association between years of experience and this expert status, for example.

Surprisingly, the MEDLINE search did not identify any cancer-related studies. However, in addition to the peer-reviewed studies that were included, there are also public quality initiatives designed to inform the general population about the quality of care in cancer settings, some of which include PROs. For example, in the National Prostate Cancer Audit (NPCA) in the United Kingdom, post-therapy PROs (EPIC-26) are adjusted for case-mix and reported using funnel plots [50]. This initiative, however, does not take baseline PROs into account for the adjustment, and the results have not yet been reported in peer-reviewed journals. Moreover, we excluded five studies that did use case-mix-adjusted PROs to compare providers, but unfortunately did not explain their adjustment methods [51–55]. However, those studies highlight that PROs—when case-mix adjusted—are valid and valuable parameters to compare different providers.

##### *Limitations*

We are aware of several limitations of our study: to begin with, only the MEDLINE database was used for the literature search. To find any other peer-reviewed research, all references of eligible studies were therefore scanned, and two more articles were included in this systematic review.

Moreover, articles not published in English, German, or French were excluded (language bias), which may result in a selection bias. Although a search strategy strictly following the PRISMA recommendations was used and an additional reference screening was performed, it is possible that some relevant studies were missed if the specific keywords were not used by either the authors or the MEDLINE database.

Since the included studies showed a huge variety in terms of the statistical methods used, it was not possible to perform a meta-analysis. However, as those statistical approaches themselves were the research subject, this was not expected.

For assessing the quality and risk of bias of the included studies, the CASP tool results are presented. The included studies are all retrospective analyses of prospectively collected data and the appraisal tools for cohort studies were mostly applicable. However, most of those appraisal tools incorporate items that are not applicable for our purposes, such as the assessment of the exposure of the outcome (e.g., [56]) and we decided to only use the CASP tool, which does not provide a numerical scale for assessing the quality of the studies, and thus the rating of each study is subjectively dependent on the reader's main focus. This subjectivity may be a limitation, however, the use of the CASP tool results in a greater flexibility and usability of the risk of bias evaluation.

The review was conducted strictly according to PRISMA. We employed a concise and literature-based search string that was critically reviewed by multiple peers. The methodology used for this review was assessed for validity by the PROSPERO Foundation, where a protocol was pre-registered. A thorough quality assessment focusing on methodological issues was performed for each included study.

## 5. Conclusions and Recommendations

All in all, the variety of the approaches and the small numbers of published articles on the topic underscore the uncertainties concerning the best way of carrying out case-mix adjustment of PROs. To report interpretable PROs for all stakeholders (patients, clinicians, decision makers, researchers, etc.), it would be desirable for a more consistent methodology to be used across approaches.

Moreover, taking the heterogeneity of our findings into account, we would like to point out two additional major challenges of CMA of PROs—especially in contrast to non-PRO quality metrics—for which we see a need for further research:

Firstly, while most other quality metrics are rate based, with only two possible response categories (such as dead/alive or complication/no complication), PROs typically come in the form of continuous measures. PROs thus require data collection before and after an intervention (such as a hospital stay), because a patient's functional status can change gradually, whereas in the other example, he/she is necessarily alive and could not have had an intervention-related complication before an intervention. Continuous measures with two points of measurement, however, require a number of analytical decisions associated with statistical pitfalls. For example: should *changes* from before and after an intervention be reported, or should adjusted *scores after* an intervention be presented that include the pre-intervention score as an adjustor? The included studies in this review mostly adjusted the scores after intervention. However, some study groups decided to report changes (e. g., [52]) and qualitative research on how to present PROs to clinicians often highlights the importance of presenting trends over time [56,57].

Secondly, while traditional quality metrics data are typically collected without—or with little—active contribution from the patient and are thus mostly documented for all patients (although with varying documentation quality), PRO data collection requires active involvement of the patients, who have to complete the questionnaires. This leads to patient selection and may create substantial bias. For example, patients with some characteristics are less likely to complete questionnaires (due to factors such as age, educational level, or their degree of trust in the healthcare system) [58]. Similarly, providers may be reluctant to include some patients because they expect them to have a negative impact on their results—e.g., patients with more comorbidities tend to have poorer PRO scores [59]. Correcting

such potential biases is difficult because of the unknown characteristics of the unsurveyed patients, so weighting is scarcely possible. In practice, this may lead to a decision to include only those providers in comparisons for which a minimum number (or proportion) of patients are included in the analytical dataset. Thus, in our opinion, quality improvement initiatives based on PRO comparisons should be encouraged to include all eligible patients and report on potential selection biases.

Due to the lack of comparisons between different adjustment methods for the same datasets, there is no strong scientific evidence for favoring one approach over the others. Recommendations on what a “gold standard” should include are therefore difficult to make at present. We would therefore cautiously derive three recommendations as a starting point for further discussion, on the basis of what appear to be successful practices in the studies included in this review. In order to keep the threshold for comparisons between treatment outcomes with the help of PROs as low as possible, so that they are accessible to a broad public, we would recommend the following on the basis of the present findings:

Firstly, with regard to baseline PROs, it appears to be common methodological practice to include them. All of the studies feature them either as an adjustor in the model, or as part of the dependent variable when the difference between baseline and follow-up PRO is used. The reasons for doing so are well elaborated on in some of the studies included [23,27].

Secondly, a number of the studies emphasized that the CMA should be accessible by different stakeholders (e.g., clinicians, patients) and should therefore be easily interpretable and offer additional information that could be used for better selection, or to identify providers with good performance [20,23,25]. In our view, this notion favors unit/scale-preserving approaches, so that a new “unit” does not need to be learned.

Thirdly, in line with most of the studies included, provider characteristics should not be included unless there are strong reasons for this such as contextual factors that are beyond the providers’ control.

In the absence of strong evidence favoring one approach over any of the others, we would suggest that further research should be carried out, including comparison of approaches both in terms of the statistics used and also the comprehensibility of the reported results for the target audiences of the provider comparisons. Among the many open questions are minimum case thresholds, whether to take the hierarchical structure of the data into account using multilevel models, as well as measuring the impact that CMA has on rankings and the threshold at which case-mix-adjusted differences in PROs should be considered relevant at all. The variation in CMA approaches underlines the importance of establishing a methodological guideline or “checklist” that could be used, for example, for quality improvement initiatives aiming to use PROs for provider comparisons. As the evidence is still scarce, we would recommend—alongside further research on the benefits and pitfalls of different CMA methods—a consensus-based approach to the development of such a guideline.

**Author Contributions:** Conceptualization: N.T.S.; Methodology: N.T.S., C.K.; Formal Analysis: N.T.S., C.K.; Investigation: N.T.S., C.K.; Data Curation: N.T.S.; Writing—Original Draft Preparation: N.T.S., C.K.; Writing—Review and Editing: N.T.S., C.K., H.P., S.W., C.B.; Visualization: N.T.S.; Supervision: H.P., S.W.; Project Administration: S.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### Appendix A. Search String for MEDLine, Including Number of Hits of MEDLine Search

(“patient reported outcome measures”[MeSH Terms] OR (“patient”[All Fields] AND “reported”[All Fields] AND “outcome”[All Fields] AND “measures”[All Fields]) OR “patient reported outcome measures”[All Fields] OR (“patient”[All Fields] AND “re-

ported"[All Fields] AND "outcome"[All Fields]) OR "patient reported outcome"[All Fields]): 747 hits

AND

("casemix"[All Fields] OR "casemixes"[All Fields] OR ("risk"[MeSH Terms] OR "risk"[All Fields])) AND ("social adjustment"[MeSH Terms] OR ("social"[All Fields] AND "adjustment"[All Fields]) OR "social adjustment"[All Fields] OR "adjustment"[All Fields]): 704 hits

AND

("hasabstract"[Text Word] AND "humans"[MeSH Terms] AND ("English"[Language] OR "French"[Language] OR "German"[Language])): 640 hits

## Appendix B

Table A1. CASP tool for included studies.

Author	Clearly Focused	Recruitment Acceptable	Exposure Accurately Measured	Outcome Accurately Measured	Identified Confounding	Accounted for Confounding	Subject Follow-Up Complete Enough	Subject Follow-Up Long Enough	Results Precise	Believe Results	Applicable Results	Fit with Other Evidence	Implications for Practice
Deutscher et al., 2018 [25]	x	can't tell	x	x	x	x	can't tell	can't tell	x	x	x	x	x
Farin et al., 2009 [20]	x	x	can't tell	x	can't tell	can't tell	can't tell	can't tell	x	x	x	x	x
Gozaló et al., [29]	x	can't tell	x	x	x	x	can't tell	x	x	x	x	x	x
Gutacker et al., 2013 [21]	x	x	x	x	x	x	x	x	x	x	x	x	x
Hendryx et al., 1999 [26]	x	x	x	x	can't tell	x	x	can't tell	x	x	can't tell	x	x
Khor et al., 2020 [22]	x	x	x	x	x	x	x	x	x	x	x	x	x
Lutz et al., 2020 [27]	x	x	x	x	x	x	x	can't tell	x	x	x	x	x
Nuttall et al., 2015 [25]	x	can't tell	x	x	x	x	can't tell	x	x	x	x	x	x
Resnik & Hart 2003 [28]	x	can't tell	x	x	x	x	can't tell	x	x	x	x	x	x
Sivaganesan et al., 2018 [24]	x	can't tell	x	x	x	x	x	x	x	x	x	x	x
Varagunam et al., 2015 [30]	x	can't tell	x	x	x	x	can't tell	x	x	x	x	x	x

## References

- Klinkhammer-Schalke, M.; Koller, M.; Steinger, B.; Ehret, C.; Ernst, B.; Wyatt, J.C.; Hofstadter, F.; Lorenz, W.; Regensburg Qo, L.S.G. Direct improvement of quality of life using a tailored quality of life diagnosis and therapy pathway: Randomised trial in 200 women with breast cancer. *Br. J. Cancer* **2012**, *106*, 826–838. [\[CrossRef\]](#) [\[PubMed\]](#)
- Klinkhammer-Schalke, M.; Steinger, B.; Koller, M.; Zeman, F.; Fürst, A.; Gump, J.; Obermaier, R.; Piso, P.; Lindberg-Scharf, P.; Altendorfer, A.; et al. Diagnosing deficits in quality of life and providing tailored therapeutic options: Results of a randomised trial in 220 patients with colorectal cancer. *Eur. J. Cancer* **2020**, *130*, 102–113. [\[CrossRef\]](#) [\[PubMed\]](#)
- Fromme, E.K.; Eilers, K.M.; Mori, M.; Hsieh, Y.C.; Beer, T.M. How accurate is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the Quality-of-Life Questionnaire C30. *J. Clin. Oncol.* **2004**, *22*, 3485–3490. [\[CrossRef\]](#) [\[PubMed\]](#)
- Laugsand, E.A.; Sprangers, M.A.; Bjordal, K.; Skorpén, F.; Kaasa, S.; Klepstad, P. Health care providers underestimate symptom intensities of cancer patients: A multicenter European study. *Health Qual. Life Outcomes* **2010**, *8*, 104. [\[CrossRef\]](#) [\[PubMed\]](#)
- Basch, E.; Deal, A.M.; Kris, M.G.; Scher, H.I.; Hudis, C.A.; Sabbatini, P.; Rogak, L.; Bennett, A.V.; Dueck, A.C.; Atkinson, T.M.; et al. Symptom Monitoring with Patient-Reported Outcomes during Routine Cancer Treatment: A Randomized Controlled Trial. *J. Clin. Oncol.* **2015**, *34*, 557. [\[CrossRef\]](#)
- Basch, E.; Deal, A.M.; Dueck, A.C.; Scher, H.I.; Kris, M.G.; Hudis, C.; Schrag, D. Overall Survival Results of a Trial Assessing Patient-Reported Outcomes for Symptom Monitoring During Routine Cancer Treatment. *JAMA* **2017**, *318*, 197–198. [\[CrossRef\]](#) [\[PubMed\]](#)

7. Denis, F.; Yossi, S.; Septans, A.-L.; Charron, A.; Voog, E.; Dupuis, O.; Ganem, G.; Pointreau, Y.; Letellier, C. Improving survival in patients treated for a lung cancer using self-evaluated symptoms reported through a web application. *Am. J. Clin. Oncol.* **2017**, *40*, 464–469. [CrossRef]
8. Zeegers, M.P.; Bryan, R.T.; Langford, C.; Billingham, L.; Murray, P.; Deshmukh, N.S.; Hussain, S.; James, N.; Wallace, D.M.A.; Cheng, K.K. The West Midlands Bladder Cancer Prognosis Programme: Rationale and design. *BJU Int.* **2010**, *105*, 784–788. [CrossRef]
9. Engelen, V.; Detmar, S.; Koopman, H.; Maurice-Stam, H.; Caron, H.; Hoogerbrugge, P.; Egeler, R.M.; Kaspers, G.; Grootenhuys, M. Reporting health-related quality of life scores to physicians during routine follow-up visits of pediatric oncology patients: Is it effective?: Patient Reported Outcomes in Pediatric Clinic. *Pediatr. Blood Cancer* **2012**, *58*, 766–774. [CrossRef]
10. Trautmann, F.; Hentschel, L.; Hornemann, B.; Rentsch, A.; Baumann, M.; Ehninger, G.; Schmitt, J.; Schuler, M. Electronic real-time assessment of patient-reported outcomes in routine care—First findings and experiences from the implementation in a comprehensive cancer center. *Support. Care Cancer* **2016**, *24*, 3047–3056. [CrossRef]
11. Slovenian National Institute of Public Health. Report on the Basis of the Analysis of Data from the Survey on National Cancer Control. Programmes/Cancer Documents in EU—An IPAAC Report; Ljubljana: 2019. Available online: <https://www.ipaac.eu/res/file/outputs/wp10/national-cancer-control-plans-survey.pdf> (accessed on 2 August 2021).
12. ICHOM—International Consortium for Health Outcomes Measurement Localized Prostate Cancer Data Collection Reference Guide. Localized Prostate Cancer Data Collection Reference Guide. 2015. Available online: <http://ichom.org/files/medical-conditions/localized-prostate-cancer/localized-prostate-cancer-reference-guide.pdf> (accessed on 4 August 2021).
13. Evans, S.M.; Millar, J.L.; Moore, C.M.; Lewis, J.D.; Huland, H.; Sampurno, F.; Connor, S.E.; Villanti, P.; Litwin, M.S. Cohort profile: The TrueNTH Global Registry—An international registry to monitor and improve localised prostate cancer health outcomes. *BMJ Open* **2017**, *7*, e017006. [CrossRef] [PubMed]
14. Kowalski, C.; Roth, R.; Carl, G.; Feick, G.; Oesterle, A.; Hinkel, A.; Steiner, T.; Brock, M.; Kaftan, B.; Borowitz, R.; et al. A multicenter paper-based and web-based system for collecting patient-reported outcome measures in patients undergoing local treatment for prostate cancer: First experiences. *J. Patient-Rep. Outcomes* **2020**, *4*, 56. [CrossRef] [PubMed]
15. Scheibe, M.; Herrmann, A.; Schmitt, J.; Einhart, N.; Sedlmayr, B.; Kowalski, C. Implementation of patient-reported outcome assessment in routine cancer care: A systematic review of multicentric programs in Europe. *Z. Für Evidenz Fortbild. Qual. Im Gesundheitswesen* **2020**, *156–157*, 11–23. [CrossRef] [PubMed]
16. Iezzoni, L. *Risk Adjustment for Measuring Healthcare Outcomes*, 4th ed.; Health Administration Press: Chicago, IL, USA, 2012.
17. OECD. *Measuring What Matters: The Patient-Reported Indicator Surveys*; OECD: Paris, France, 2019.
18. McGowan, J.; Sampson, M.; Salzwedel, D.M.; Cogo, E.; Foerster, V.; Lefebvre, C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *J. Clin. Epidemiol.* **2016**, *75*, 40–46. [CrossRef]
19. Critical Appraisal Skills Programme CASP Cohort Study Checklist. 2018. Available online: [https://casp-uk.net/wp-content/uploads/2018/01/CASP-Cohort-Study-Checklist\\_2018.pdf](https://casp-uk.net/wp-content/uploads/2018/01/CASP-Cohort-Study-Checklist_2018.pdf) (accessed on 18 March 2021).
20. Farin, E.; Jäckel, W.H.; Schalaster, V. Projektgruppe QS-Reha-Verfahren in der AQMS The quality assurance programme of the statutory health insurance funds in medical rehabilitation: Results and further developments. *Gesundh. Bundesverb. Arzt Offentlichen Gesundh.* **2009**, *71*, 163–174. [CrossRef]
21. Gutacker, N.; Bojke, C.; Daidone, S.; Devlin, N.; Street, A. Hospital variation in patient-reported outcomes at the level of EQ-5D dimensions: Evidence from England. *Med. Decis. Mak. Int. J. Soc. Med. Decis. Mak.* **2013**, *33*, 804–818. [CrossRef]
22. Khor, S.; Lavalley, D.C.; Cizik, A.M.; Bellabarba, C.; Dagal, A.; Hart, R.A.; Howe, C.R.; Martz, R.D.; Shonnard, N.; Flum, D.R. Hospital and Surgeon Variation in Patient-reported Functional Outcomes After Lumbar Spine Fusion: A Statewide Evaluation. *Spine* **2020**, *45*, 465–472. [CrossRef]
23. Nuttall, D.; Parkin, D.; Devlin, N. Inter-provider comparison of patient-reported outcomes: Developing an adjustment to account for differences in patient case mix. *Health Econ.* **2015**, *24*, 41–54. [CrossRef]
24. Sivaganesan, A.; Asher, A.L.; Bydon, M.; Khan, I.; Kerezoudis, P.; Foley, K.T.; Nian, H.; Harrell, F.E.; Archer, K.R.; Devin, C.J. A Strategy for Risk-adjusted Ranking of Surgeons and Practices Based on Patient-reported Outcomes after Elective Lumbar Surgery. *Spine* **2019**, *44*, 670–677. [CrossRef]
25. Deutscher, D.; Werneke, M.W.; Hayes, D.; Mioduski, J.E.; Cook, K.F.; Fritz, J.M.; Woodhouse, L.J.; Stratford, P.W. Impact of Risk Adjustment on Provider Ranking for Patients With Low Back Pain Receiving Physical Therapy. *J. Orthop. Sports Phys. Ther.* **2018**, *48*, 637–648. [CrossRef]
26. Hendryx, M.S.; Dyck, D.G.; Srebnik, D. Risk-adjusted outcome models for public mental health outpatient programs. *Health Serv. Res.* **1999**, *34*, 171–195. [PubMed]
27. Lutz, A.D.; Brooks, J.M.; Chapman, C.G.; Shanley, E.; Stout, C.E.; Thigpen, C.A. Risk Adjustment of the Modified Low Back Pain Disability Questionnaire and Neck Disability Index to Benchmark Physical Therapist Performance: Analysis From an Outcomes Registry. *Phys. Ther.* **2020**, *100*, 609–620. [CrossRef]
28. Resnik, L.; Hart, D.L. Using clinical outcomes to identify expert physical therapists. *Phys. Ther.* **2003**, *83*, 990–1002. [CrossRef]
29. Gozalo, P.L.; Resnik, L.J.; Silver, B. Benchmarking Outpatient Rehabilitation Clinics Using Functional Status Outcomes. *Health Serv. Res.* **2016**, *51*, 768–789. [CrossRef]
30. Varaganam, M.; Hutchings, A.; Black, N. Do patient-reported outcomes offer a more sensitive method for comparing the outcomes of consultants than mortality? A multilevel analysis of routine data. *BMJ Qual. Saf.* **2015**, *24*, 195–202. [CrossRef] [PubMed]

31. Ware, J.; Kosinski, M.; Keller, S.D. A 12-Item Short-Form Health Survey: Construction of scales and preliminary tests of reliability and validity. *Med. Care* **1996**, *34*, 220–233. [CrossRef]
32. Haley, S.M.; McHomey, C.A.; Ware, J.E. Evaluation of the MOS SF-36 physical functioning scale (PF-10): I. Unidimensionality and reproducibility of the Rasch item scale. *J. Clin. Epidemiol.* **1994**, *47*, 671–684. [CrossRef]
33. McGee, J.; Goldfield, N.; Riley, K.; Morton, J. *Collecting Information from Health Care Consumers: A Resource Manual of Tested Questionnaires and Practical Advice*; Apsen Publications: Gaithersburg, MD, USA, 1996.
34. Lehman, A.F.; Postrado, L.T.; Rachuba, L.T. Convergent validation of quality of life assessments for persons with severe mental illnesses. *Qual. Life Res. Int. J. Qual. Life Asp. Treat. Care Rehabil.* **1993**, *2*, 327–333. [CrossRef] [PubMed]
35. Rabin, R.; de Charro, F. EQ-5D: A measure of health status from the EuroQol Group. *Ann. Med.* **2001**, *33*, 337–343. [CrossRef] [PubMed]
36. Bührlen, B.; Gerdes, N.; Jäckel, W.H. Development and psychometric testing of a patient questionnaire for medical rehabilitation (IRES-3). *Die Rehabil.* **2005**, *44*, 63–74. [CrossRef]
37. Chiarotto, A.; Maxwell, L.J.; Ostelo, R.W.; Boers, M.; Tugwell, P.; Terwee, C.B. Measurement Properties of Visual Analogue Scale, Numeric Rating Scale, and Pain Severity Subscale of the Brief Pain Inventory in Patients With Low Back Pain: A Systematic Review. *J. Pain* **2019**, *20*, 245–263. [CrossRef]
38. Murray, D.W.; Fitzpatrick, R.; Rogers, K.; Pandit, H.; Beard, D.J.; Carr, A.J.; Dawson, J. The use of the Oxford hip and knee scores. *J. Bone Jt. Surg. Br.* **2007**, *89*, 1010–1014. [CrossRef] [PubMed]
39. Fritz, J.M.; Irrgang, J.J. A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale. *Phys. Ther.* **2001**, *81*, 776–788. [CrossRef] [PubMed]
40. Vernon, H.; Mior, S. The Neck Disability Index: A study of reliability and validity. *J. Manip. Physiol. Ther.* **1991**, *14*, 409–415.
41. Fairbank, J.C.; Pynsent, P.B. The Oswestry Disability Index. *Spine* **2000**, *25*, 2940–2952. [CrossRef]
42. Dempster, M.; Donnelly, M.; O’Loughlin, C. The validity of the MacNew Quality of Life in heart disease questionnaire. *Health Qual. Life Outcomes* **2004**, *2*, 6. [CrossRef]
43. Swiontkowski, M.F.; Engelberg, R.; Martin, D.P.; Agel, J. Short musculoskeletal function assessment questionnaire: Validity, reliability, and responsiveness. *J. Bone Jt. Surg. Am.* **1999**, *81*, 1245–1260. [CrossRef] [PubMed]
44. Hart, D.L.; Werneke, M.W.; Wang, Y.-C.; Stratford, P.W.; Mioduski, J.E. Computerized adaptive test for patients with lumbar spine impairments produced valid and responsive measures of function. *Spine* **2010**, *35*, 2157–2164. [CrossRef]
45. Schokkaert, E.; Van de Voorde, C. Direct versus indirect standardization in risk adjustment. *J. Health Econ.* **2009**, *28*, 361–374. [CrossRef]
46. Sibert, N.T.; Dieng, S.; Oesterle, A.; Feick, G.; Carl, G.; Steiner, T.; Minner, J.; Roghmann, F.; Kaftan, B.; Zengerling, F.; et al. Psychometric validation of the German version of the EPIC-26 questionnaire for patients with localized and locally advanced prostate cancer. *World J. Urol.* **2019**, *39*, 11–25. [CrossRef]
47. Skolarus, T.A.; Dunn, R.L.; Sanda, M.G.; Chang, P.; Greenfield, T.K.; Litwin, M.S.; Wei, J.T. Minimally important difference for the Expanded Prostate Cancer Index Composite Short Form. *Urology* **2015**, *85*, 101–105. [CrossRef]
48. Farin, E.; Follert, P.; Gerdes, N.; Jäckel, W.H.; Thalau, J. Quality assessment in rehabilitation centres: The indicator system “Quality Profile”. *Disabil. Rehabil.* **2004**, *26*, 1096–1104. [CrossRef]
49. Roth, R.; Dieng, S.; Oesterle, A.; Feick, G.; Carl, G.; Hinkel, A.; Steiner, T.; Kaftan, B.T.; Kunath, F.; Hadaschik, B.; et al. Determinants of self-reported functional status (EPIC-26) in prostate cancer patients prior to treatment. *World J. Urol.* **2021**, *39*, 27–36. [CrossRef]
50. The Royal College of Surgeons of England. National Prostate Cancer Audit Annual Report 2020. 2021. Available online: [https://www.npca.org.uk/content/uploads/2021/01/NPCA-Annual-Report-2020\\_Final\\_140121.pdf](https://www.npca.org.uk/content/uploads/2021/01/NPCA-Annual-Report-2020_Final_140121.pdf) (accessed on 2 August 2021).
51. Neuburger, J.; Cromwell, D.A.; Hutchings, A.; Black, N.; van der Meulen, J.H. Funnel plots for comparing provider performance based on patient-reported outcome measures. *BMJ Qual. Saf.* **2011**, *20*, 1020–1026. [CrossRef]
52. Waljee, J.F.; Ghaferi, A.; Cassidy, R.; Varban, O.; Finks, J.; Chung, K.C.; Carozzi, N.E.; Dimick, J.B. Are Patient-reported Outcomes Correlated With Clinical Outcomes After Surgery?: A Population-based Study. *Ann. Surg.* **2016**, *264*, 682–689. [CrossRef] [PubMed]
53. Waljee, J.F.; Ghaferi, A.; Finks, J.F.; Cassidy, R.; Varban, O.; Carlin, A.; Carozzi, N.; Dimick, J. Variation in Patient-reported Outcomes Across Hospitals Following Surgery. *Med. Care* **2015**, *53*, 960–966. [CrossRef] [PubMed]
54. Desai, A.; Bekelis, K.; Ball, P.A.; Lurie, J.; Mirza, S.K.; Tosteson, T.D.; Zhao, W.; Weinstein, J.N. Outcome Variation Across Centers after Surgery for Lumbar Stenosis and Degenerative Spondylolisthesis. *Spine* **2013**, *38*, 678–691. [CrossRef] [PubMed]
55. Yen, S.-C.; Corkery, M.B.; Chui, K.K.; Manjourides, J.; Wang, Y.-C.; Resnik, L.J. Risk Adjustment for Lumbar Dysfunction: Comparison of Linear Mixed Models With and Without Inclusion of Between-Clinic Variation as a Random Effect. *Phys. Ther.* **2015**, *95*, 1692–1702. [CrossRef]
56. Brundage, M.D.; Smith, K.C.; Little, E.A.; Bantug, E.T.; Snyder, C.F. The PRO Data Presentation Stakeholder Advisory Board Communicating patient-reported outcome scores using graphic formats: Results from a mixed-methods evaluation. *Qual. Life Res.* **2015**, *24*, 2457–2472. [CrossRef] [PubMed]

57. Sibert, N.T.; Kowalski, C.; Pfaff, H.; Wesselmann, S.; Breidenbach, C. Clinicians' knowledge and attitudes towards patient reported outcomes in colorectal cancer care—Insights from qualitative interviews. *BMC Health Serv. Res.* **2021**, *21*, 366. [[CrossRef](#)]
58. Nguyen, H.; Butow, P.; Dhillon, H.; Sundaresan, P. A review of the barriers to using Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs) in routine cancer care. *J. Med. Radiat. Sci.* **2021**, *68*, 186–195. [[CrossRef](#)] [[PubMed](#)]
59. Luque Ramos, A.; Redeker, I.; Hoffmann, F.; Callhoff, J.; Zink, A.; Albrecht, K. Comorbidities in Patients with Rheumatoid Arthritis and Their Association with Patient-reported Outcomes: Results of Claims Data Linked to Questionnaire Survey. *J. Rheumatol.* **2019**, *46*, 564–571. [[CrossRef](#)] [[PubMed](#)]

#### **4.3. Variation across operating sites in urinary and sexual outcomes after radical prostatectomy in localized and locally advanced prostate cancer**

Veröffentlicht als: **Sibert, N. T.**, Pfaff, H., Breidenbach, C., Wesselmann, S., Roth, R., Feick, G., Carl, G., Dieng, S., Gaber, A. A., Blana, A., Darr, C., Distler, F., Kunath, F., Bedke, J., Erdmann, J., Minner, J., Simon, J., Kwiatkowski, M., Burchardt, M., ... Kowalski, C. (2022). Variation across operating sites in urinary and sexual outcomes after radical prostatectomy in localized and locally advanced prostate cancer. *World Journal of Urology*. <https://doi.org/10.1007/s00345-022-03985-6>



## Variation across operating sites in urinary and sexual outcomes after radical prostatectomy in localized and locally advanced prostate cancer

Nora Tabea Sibert<sup>1</sup> · Holger Pfaff<sup>2</sup> · Clara Breidenbach<sup>1</sup> · Simone Wesselmann<sup>1</sup> · Rebecca Roth<sup>3</sup> · Günther Feick<sup>4</sup> · Günter Carl<sup>4</sup> · Sebastian Dieng<sup>5</sup> · Amr A. Gaber<sup>6</sup> · Andreas Blana<sup>7</sup> · Christopher Darr<sup>8</sup> · Florian Distler<sup>9</sup> · Frank Kunath<sup>10</sup> · Jens Bedke<sup>11</sup> · Jörg Erdmann<sup>12</sup> · Jörg Minner<sup>13</sup> · Jörg Simon<sup>14</sup> · Maciej Kwiatkowski<sup>15</sup> · Martin Burchardt<sup>16</sup> · Nino Harz<sup>17</sup> · Stefan Conrad<sup>18</sup> · Thomas Höfner<sup>19</sup> · Thomas Knoll<sup>20</sup> · Burkhard Beyer<sup>21</sup> · Peter Hammerer<sup>22</sup> · Christoph Kowalski<sup>1</sup>

Received: 19 November 2021 / Accepted: 7 March 2022  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

### Abstract

**Purpose** The extent of variation in urinary and sexual functional outcomes after radical prostatectomy (RPE) between prostate cancer (PC) operating sites remains unknown. Therefore, this analysis aims to compare casemix-adjusted functional outcomes (EPIC-26 scores incontinence, irritative/obstructive function and sexual function) between operating sites 12 months after RPE.

**Materials and methods** Analysis of a cohort of 7065 men treated with RPE at 88 operating sites (prostate cancer centers, “PCCs”) between 2016 and 2019. Patients completed EPIC-26 and sociodemographic information surveys at baseline and 12 months after RPE. Survey data were linked to clinical data. EPIC-26 domain scores at 12 months after RPE were adjusted for relevant confounders (including baseline domain score, clinical and sociodemographic information) using regression analysis. Differences between sites were described using minimal important differences (MIDs) and interquartile ranges (IQR). The effects of casemix adjustment on the score results were described using Cohen’s *d* and MIDs.

**Results** Adjusted domain scores at 12 months varied between sites, with IQRs of 66–78 (incontinence), 89–92 (irritative/obstructive function), and 20–29 (sexual function). Changes in domain scores after casemix adjustment for sites  $\geq 1$  MID were noted for the incontinence domain (six sites). Cohen’s *d* ranged between  $-0.07$  (incontinence) and  $-0.2$  (sexual function), indicating a small to medium effect of casemix adjustment.

**Conclusions** Variation between sites was greatest in the incontinence and sexual function domains for RPE patients. Future research will need to identify the factors contributing to this variation.

**Trial Registry.** The study is registered at the German Clinical Trial Registry ([https://www.drks.de/drks\\_web/](https://www.drks.de/drks_web/)) with the following ID: DRKS00010774.

**Keywords** Center comparison · Patient-reported outcomes · Prostate cancer · Quality assurance · Radical prostatectomy

### Introduction

Prostate cancer (“PC”) is the most common cancer among men in Western countries [1, 2]. The survival rates have been increasing and are relatively high, with 5-year and 10-year overall survival rates of 91% and 90%, respectively, in 2014

[3]. Functional status after PC treatment is consequently an important outcome in addition to survival [4, 5]. Functional status includes disease-specific symptoms and impairment, as well as health-related quality of life. Patient-reported outcome (“PRO”) questionnaires are used to measure health-related outcomes directly reported by patients themselves. Combining PROs and clinical outcomes such as (relapse-free) survival to evaluate patients’ functional status is now a widespread method, not only in health-care research and clinical trials [6] but also in routine clinical care [7]. For

✉ Nora Tabea Sibert  
sibert@krebsgesellschaft.de

Extended author information available on the last page of the article

cancer care, in particular, the importance of PROs has been increasing in recent years [8, 9].

PROs are particularly valuable in PC, since recovery and improvement in functional status are important indicators for treatment success as well as treatment regret [10–12], for radical prostatectomy (“RPE”) especially urological (incontinence, irritative/obstructive symptoms) and sexual outcome (erectile dysfunction). PROs can be used not only at the individual patient level, but also bundled to provide information about health outcomes with different providers. PROs can thus serve as a quality assurance instrument to compare performance between healthcare providers [13]. When this is done, however, the results have to be carefully reported, with the different providers’ casemixes being taken into account [14]. Otherwise, unadjusted provider comparisons may remain unfair and misleading.

Recently, Nossiter et al. reported the relationship between PROs after RPE on a hospital level and hospital volume [15] and thus used PROs to evaluate the hospitals’ performance. However, baseline PROs were not included in the models but are known to be important adjusters for follow-up PROs [16]. The study focuses rather on hospital volume than on evaluating how PROs can be used for comparing different PC-care providers.

Therefore, the aims of this study were to report on casemix-adjusted prostate cancer-specific PRO scores (urinary incontinence, irritative/obstructive function, sexual function) for comparing different operating sites and to present a casemix adjustment methodology for PROs.

## Materials and methods

### Data collection

Prostate cancer centers (“PCCs”) in Germany recruit patients for the Prostate Cancer Outcomes (“PCO”) Study since 2016. Before definitive treatment for PC, patients with clinically localized or (locally) advanced PC (any T, any N, M0) are requested to complete the Expanded Prostate Cancer Index Composite-26 (“EPIC-26”) questionnaire—either using a web interface or a paper–pencil version—along with three additional sociodemographic questions after providing informed consent. Patients are asked to complete the EPIC-26 questionnaire again one year after the start of treatment.

Questionnaire responses are bundled with clinical information documented for certification purposes and according to the International Consortium for Health Outcomes Measurement (“ICHOM”) standard set [17] by the PCCs, including disease-specific information and treatment information, using the OncoBox [18]. This tool harmonizes output formats from different tumor documentation systems.

The PCO is an ongoing study and forms part of the TrueNTH Registry launched by the Movember Foundation. The ethics committee of the Medical Association of Berlin has approved the study (Eth-12/16). All of the participating PCCs are surgical sites and are referred to here as “sites” or “operating sites.”

### Measures

The analysis presented here includes data for PC patients taking part in the PCO study who underwent (any kind of) RPE as the first definitive treatment. Patients receiving additional (salvage) treatments like androgen deprivation therapy (“ADT”) or radiation before completion of the follow-up questionnaire were excluded from the analytical data set. Sites with fewer than 10 patients were not included in the analysis.

### Dependent variables

The EPIC-26 domain scores [19] 1 year after RPE ( $t_1$ ) were used as outcome measurements. The EPIC-26 is a well-established, PC-specific PRO questionnaire recommended by the ICHOM [20] that summarizes responses in five domains: incontinence, irritative/obstructive function, bowel function, sexual function, and vitality/hormonal function. The validated German translation of the EPIC-26 was used [21]. For this article, only the three most RPE-relevant domains urinary incontinence, irritative/obstructive function and sexual function are reported. However, all results for bowel function and vitality/hormonal function can be found in Supplementary Material 6.

All EPIC-26 domain scores can range between 0 and 100, with 0 indicating the poorest functional outcome. They are calculated using a scoring manual [22].

### Adjustors

The following disease-specific and sociodemographic patient characteristics were used to adjust for different casemixes at the different operating sites (reference category in **bold**, selection based on prior research [23] and iteratively):

- Baseline EPIC-26 domain score before surgery ( $t_0$ ).
- Age at diagnosis (categories < 60, **60–69**, 70–79, > 79).
- Risk classification according to the German PC guideline [24], which follows the D’Amico risk classification [25] (categories: localized low risk, **localized intermediate risk**, localized high risk, locally advanced, advanced).
- Number of comorbidities (categories: **none**, 1–2, > 2, unknown).
- Educational level (measured as highest school qualification, categories: **lower secondary school**, interme-

diate secondary school west, intermediate secondary school east [the school-leaving certificate of the German Democratic Republic up to 1990], entrance certificate for university of applied science, entrance certificate for university, other, none).

- Health insurance type (categories: **statutory**, private, other/none).
- Nationality (categories: **German**, other).
- Hormone therapy before surgery (ADT, categories: **no**, yes).
- Active surveillance (AS) before surgery (categories: **no**, yes).

#### Missing values

Multiple imputations by chained equations and k-nearest neighbor imputation was used for handling missing values for all variables expect for numbers of comorbidities. More information can be found in supplementary material 4.

#### Casemix adjustment of EPIC-26 scores 1 year after RPE

Casemix adjustment is a statistical method to account for different casemixes when comparing different healthcare providers [16]. The following casemix adjustment approach was used for this analysis (similar to the National Health Service England approach [26]):

1. Calculation of  $e = \text{expected}_{i_1}$  EPIC-26 scores ( $t_1$ ) for each patient, using a multiple regression model with the EPIC-26 score ( $t_1$ ) as dependent variable and patients' disease-specific socioeconomic characteristics and baseline PROs  $t_0$  (compare the "Adjustors" section above).
2. Calculation of a performance indicator for each site as the mean of the difference between observed and expected EPIC-26 scores one year after RPE:

$$\text{performance}_j = \frac{1}{n} \sum_{i=1}^n (\text{observed}_{i,j} - \text{expected}_{i,j})$$

with  $n$  : number of patients treated by site  $j$ .

3. Calculation of the adjusted score ( $t_1$ ) for each site:

$$\text{adjusted}_j = \text{observed}_{i,\text{all}} + \text{performance}_j$$

with  $\text{observed}_{i,\text{all}}$  : mean of  $\text{observed}_{i_1}$  of patients from all sites.

In addition, minimal important difference (MID) ranges were calculated: MIDs are the smallest change in a treatment outcome—as PROs—that a patient would identify as important. For a better comparison between sites, MID ranges for

every site were included in the graphics with the adjusted EPIC-26 scores:

$$\text{MID range}_j = [\text{adjusted}_j - \text{MID}; \text{adjusted}_j + \text{MID}].$$

Skolarus et al. determined MIDs for each EPIC-26 score. [27]. Those MIDs are currently frequently used in urological research to assess differences in EPIC-26 results (i.e. [28]). The lower threshold for each EPIC-26 domain is listed below:

- Incontinence: 6
- Irritative/obstructive function: 5
- Bowel function: 4
- Sexual function: 10
- Vitality/hormonal function: 4

$R^2$  and adjusted  $R^2$  were calculated for all regression models. The models were checked for heteroscedasticity, multicollinearity, and normality using the R package "performance" (results are available upon request).

Effects of casemix adjustment were measured using Cohen's  $d$  when comparing unadjusted and adjusted scores. A Cohen's  $|d| > 0.5$  indicates a large effect.

The following formula was used to measure the number of poorer-performing sites:

$$\text{MID}_{\text{lower}} \text{ratio} = \frac{N_{\text{adjusted} < \bar{x} - \text{MID}}}{N}$$

for each EPIC-26 domain with  $N_{\text{adjusted} < \bar{x} - \text{MID}}$ : number of sites, with an adjusted score less than  $\bar{x} - \text{MID}$ ,  $\bar{x}$ : median of adjusted scores and  $N$ : total number of sites.

In addition, absolute differences between unadjusted and adjusted scores were analyzed and compared with the corresponding MIDs: for each EPIC-26 domain, the number of operating sites for which this absolute difference was greater than or equal to the corresponding MID was calculated as an indicator of adjustment effects.

Operating sites with fewer than 10 patients were not included in the analysis. The TRIPOD Statement advice was followed for model development [29]. All statistical analyses were performed using the R statistical package, version 4.0.2 (2020-06-22).

## Results

### Participants and operating sites

Between July 2016 and July 2019, 13 218 men participated in PCO. Data for 7 065 patients who underwent radical prostatectomy between July 2016 and 2019 with no additional (salvage) treatment between baseline and answering the

EPIC-26 questionnaire one year after surgery at 88 different sites were included in the analysis (see Supplementary Material 1 for a more detailed description of the data sample). The mean age of the patients included was 65.5 (SD 7.2). Most of the patients had localized PC with intermediate risk. Table 1 lists all of the patients' characteristics.

Table 2 presents the interquartile ranges (IQRs), means, and standard deviation (SD) of the EPIC-26 scores before and one year after RPE. At both measurement time points, sexual function showed the lowest score (62.9 ( $t_0$ ), 26.7 ( $t_1$ )), whereas urinary incontinence had the highest at baseline (93.45 ( $t_0$ ), ( $t_1$ )) and irritative/obstructive function had the highest after RPE (86.61 ( $t_0$ ), 91.12 ( $t_1$ )) among the selected EPIC-26 domains.

### Functional outcome in PCCs 1 year after RPE

#### Regression models for casemix adjustment

Table 1 of the Supplementary Material 5 shows the results for each regression model (one per EPIC-26 domain), including estimates and  $p$  values. Goodness of fit for the models ranged between  $R^2=0.1$  (irritative/obstructive function),  $R^2=0.11$  (incontinence), and  $R^2=0.22$  (sexual function).

Supplementary Material 2 provides the results of the following sensitivity analysis: only sites with more than 49 patients and only sites that documented comorbidities. Both showed the same trends as the models presented in this paper.

#### Comparison of adjusted and unadjusted EPIC-26 scores

Differences between adjusted and unadjusted EPIC-26 site scores were evaluated using Cohen's  $d$  as effect sizes. Cohen's  $d$  was 0.03 (irritative/obstructive function),  $-0.08$  (urinary incontinence), and  $-0.2$  (sexual function).

The median absolute difference between adjusted and unadjusted EPIC-26 site scores ranged between one domain score point (irritative/obstructive function) and three domain score points (sexual function). Differences in adjusted and unadjusted scores barely reached MIDs, only for urinary incontinence, the score after RPE changed by six points or more (MID for incontinence) at six sites. For the remaining domains, no sites were identified that had an absolute difference greater than or equal to the corresponding MID (see Supplementary Material 3: observed and adjusted site-specific EPIC-26 scores).

#### Casemix-adjusted EPIC-26 scores one year after RPE

Figure 1 shows casemix-adjusted EPIC-26 scores one year after RPE for the sites, including MID ranges. The mean

**Table 1** Characteristics of the 7065 patients

Characteristic	<i>n</i>	%
Age (yr)		
<60	1475	21
60–69	3443	49
70–79	2097	30
>79	50	1
Type of RPE		
Open	2466	35
Robotic	3090	44
Laparoscopic	612	9
Unknown	897	13
Comorbidities		
None	4596	65
1–2	1919	27
3 or more	130	2
Unknown	420	6
Risk classification		
Localized, intermediate risk	3649	52
Localized, high risk	2049	29
Localized, low risk	1140	16
Locally advanced	174	3
Advanced (N1)	53	1
Type of health insurance		
Statutory	4819	69
Private	2088	30
None or other	32	1
Unknown	126	
Nationality		
German	6752	97
Other	199	3
Unknown	114	
Educational level *		
Lower secondary school	2244	32
Entrance certificate for university	1957	28
Intermediate secondary school, west	1343	19
Entrance certificate for applied science college	930	13
Intermediate secondary school, east	331	5
Other	101	2
None	25	0
Unknown	134	
AS before RPE	104	2
ADT before RPE	101	1

ADT androgen deprivation therapy, AS active surveillance, RPE radical prostatectomy

\*Highest school-leaving qualification

adjusted score was lowest for the EPIC-26 domain of the sexual function (24.6); by contrast, the mean adjusted score was highest for irritative/obstructive function (90.4).

**Table 2** EPIC-26 scores before ( $t_0$ ) and after ( $t_1$ ) radical prostatectomy: unadjusted on patient-level data, adjusted ( $t_1$ ) on-site level

	Incontinence			Irritative/obstructive function			Sexual function		
	$t_0$	$t_1$ (unadjusted)	$t_1$ (adjusted)	$t_0$	$t_1$	$t_1$ (adjusted)	$t_0$	$t_1$	$t_1$ (adjusted)
IQR	91–100	52–100	66–78	81–100	87–100	89–92	40–87	8–40	20–29
Mean (SD)	93.45 (13.22)	74.05 (27.25)	71.35 (8.39)	86.61 (15.1)	91.12 (11.47)	90.42 (2.79)	62.87 (28.93)	26.7 (24.33)	24.58 (6.14)
Data missing	362	193	n. a.	469	442	n. a.	256	140	n. a.

Data missing only applicable for unadjusted patient-level scores

IQR, interquartile range; SD, standard deviation

Table 2 shows means and IQRs for all selected adjusted EPIC-26 scores.

The  $MID_{lower}$  – ratio values (see the “Materials and methods” section above) were 0.26 (incontinence), 0.06 (irritative/obstructive function), and 0.01 (sexual function).

## Discussion

These results show that urinary and sexual functional outcomes one year after RPE differ between operating sites before and after adjustment for relevant confounders such as patients’ baseline functional status, disease information (risk for recurrence), age, and socioeconomic status. For patients who undergo RPE, incontinence and impaired sexual function—which are characteristic adverse effects after surgery—are by far the most important PC-specific PROs, and these two differ most between sites. For incontinence, 26% of the operating sites had a poorer adjusted score than the sites’ median minus the corresponding MID (as measured by the  $MID_{lower}$  – ratio). In addition, the functional outcome for PC patients improves one year after RPE only for the irritative/obstructive domain score.

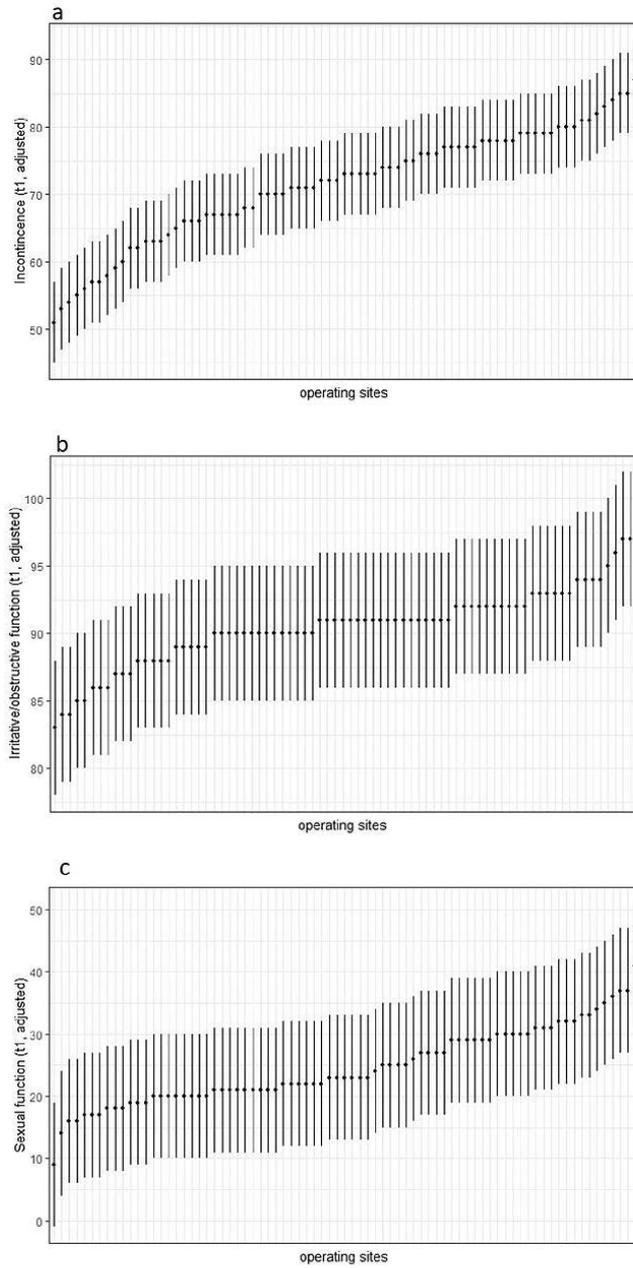
Variation in PROs after RPE between sites has recently been investigated based on the NHS National Prostate Cancer Audit database [15]. The main focus of the study was to understand the impact of hospital volume on functional outcomes. Since the casemix adjustment methodology did not include baseline PROs we did not replicate their approach but instead carefully adapted the casemix adjustment methodology proposed by the NHS England for elective surgery, with particular emphasis on choosing PC-specific, PRO-relevant and accessible patients’ characteristics as adjusters. Specifically, baseline PRO scores were included in all models and showed the strongest predictive value for PROs after one year (compare Table 1, Supplementary Material 5).  $R^2$ , as a goodness-of-fit criterion, ranged between 0.1 (irritative/obstructive function) and 0.22 (sexual function) for the

different domain scores, suggesting that the adjusters used do not work equally well in the models. This finding is consistent with previous research by Laviana et al. [23]. Thus, some variance remains for all the analyzed PROs that the variables in our models cannot explain. This was expected, as it was hypothesized that there would be differences in PC care between the participating operating sites. According to Iezonni, a health-care outcome can be expressed as a function of patients’ characteristics, treatment effectiveness, and quality of care [30]. Since patients’ characteristics were included in the models and only the same treatment method (RPE) was compared,<sup>1</sup> the “unexplained” variance in the models may reflect differences in the choice of procedure (open, laparoscopic, robotic), different levels of surgical experience or differences in follow-up care and this should be a topic for further research. For different surgical procedures, though, Haese et al. could show that functional outcomes do not differ between open and robotic approaches [31], whereas Nyberg et al. report a “moderate advantage for the robotic technique regarding erectile dysfunction” [32]. Research by Vickers et al. already strengthens that surgery outcome is strongly influenced by the individual surgeon [33]. However, since patients mostly choose sites rather than surgeons, variation on-site level is important to focus on.

Casemix adjustment is essential when comparing sites. To the best of the authors’ knowledge, there is no established method for analyzing the effects of adjustment—i.e., for describing how large a difference the adjustment makes. We, therefore, propose two different measurements: Firstly, taking effect sizes measured by Cohen’s  $d$  into account, there are no substantial differences between the observed

<sup>1</sup> Although, all patients received an RPE, the kind of surgery (open, robotic or laparoscopic) varied within the study cohort and even those different surgical modalities differ within their procedures. The surgery type however is mostly affected by the provider’s choice. Consequently, for those analysis it is considered as a provider characteristic and as such not included in the casemix adjustment models (otherwise the EPIC-26 scores would be over-adjusted).

**Fig. 1** Adjusted EPIC-26 scores ( $t_1$ ) 12 months after radical prostatectomy. EPIC-26 scores per site, including the minimum important differences (MID) range. **a** Incontinence, **b** Irritative/Obstructive Symptoms, **c** Sexual function



and casemix-adjusted PRO scores in the analysis (largest Cohen's  $d$  for sexual function:  $-0.2$ )—i.e., the casemix adjustment does not “change” the sites' scores to any large extent. Secondly, we propose an indicator for measuring adjustment effects, counting the sites for which the EPIC-26 scores changed more than the corresponding MID. The results show that the adjusted score differed to an extent greater than or equal to the corresponding MID only for urinary incontinence. Hence, both measurements reflect the fact that, although casemix adjustment is essential, for most sites the scores do not change to a clinically perceptible extent (hence MIDs).

This is the first comparison between so many different operating sites ( $n = 88$  PCCs). However, we are aware of some limitations. Firstly, although the study sample is relatively large, the casemix adjustment methodology needs to be tested for robustness using larger samples. Furthermore, more detailed information on the patients' development of functional outcomes, i.e., by more frequent follow-up questionnaires, would be desirable.

Another limitation of the study is the measurement of comorbidities which are important confounders for e. g. incontinence or erectile dysfunction. Since comorbidities do not yet form part of data collection for the quality assurance (certification) data on which the study is based, clinicians were asked to document them for the study additionally, resulting in relatively high missing rates. On a positive note, all of the other clinical and treatment information was close to complete.

Thirdly, this analysis is limited to the first description of differences between operating sites. Since different surgical approaches and follow-up care are not included in the analysis, the differences in adjusted outcomes cannot be explained in depth because it is beyond the scope of this article.

Moreover, as a key indicator to evaluate the variation of functional outcome, MIDs were proposed. For this analysis, validated MIDs from Skolarus et al. were available based on a US PC population [27]. However, as Revicki et al. pointed out, MIDs are best applicable when developed using a validation population which is similar to the target population [34].

## Conclusion

The present results show differences in the PC care provided for patients who undergo radical prostatectomy. Hence, the choice of an operating site for RPE has an impact on the patient's outcome. However, the reasons for these differences were not analyzed. The results thus raise questions about firstly, why there are so many differences in PROs between operating sites and secondly, how these differences could be reduced. As a first step, all of the results were transparently

presented to and discussed with clinicians from the participating PCCs. Additionally, PROs could be reported annually in audit reports and thus be used as a benchmarking instrument for PCCs. For this purpose, a restriction to specific EPIC-26 domains such as urinary incontinence and sexual function being relevant for surgical treatment may be useful. The use of PROs for benchmarking will only be possible if PROs are carefully and robustly casemix-adjusted and thus accepted by the clinical stakeholders. The casemix adjustment methodology described here appears to be a promising and practicable approach for using patient-reported outcomes to compare prostate cancer operating sites.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00345-022-03985-6>.

**Author contributions** *Study concept and design:* CK, SW. *Acquisition of data:* AAG, AB, CD, FD, FK, JB, JE, JM, JS, MK, NH, SC, TH, TK, PH, BB. *Analysis and interpretation of data:* NTS, CK, RR. *Drafting of the manuscript:* NTS. *Critical revision of the manuscript for important intellectual content:* NTS, CK, CB, SW, HP, GF, GC, MB, PH. *Statistical analysis:* NTS. *Obtaining funding:* GC, GF, SD, SW, CK. *Administrative, technical, or material support:* SD. *Supervision:* CK, SW, HP. *Other:* None.

**Funding** Funding was received from the Movember Foundation through Help for Prostate Cancer Patients, the collaborating partner of the Movember Foundation in Germany. The analysis and interpretations are those of the authors and not of the Movember Foundation.

## Declarations

**Conflict of interest** Breidenbach, Dieng, Kowalski, Sibert, Wesselmann are employees of the two institutions in charge of the certification system. Conrad reports personal fees from Janssen Cilag GmbH outside the submitted work. Edeling reports grants and personal fees from Proctoring Intuitive Surgical outside the submitted work. All other authors have nothing to declare.

**Research involving human participants and/or animals** All procedures performed in studies involving human participants were in accordance with the ethical standards of the ethics committee of the Medical Association of Berlin (Eth-12/16) and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

## References

1. Zentrum für Krebsregisterdaten, Gesellschaft der epidemiologischen Krebsregister in Deutschland (2017) Krebs in Deutschland für 2013/2014. Robert Koch-Institut, Berlin
2. Ferlay J, Colombet M, Soerjomataram I, Dyba T, Randi G, Bettio M et al (2018) Cancer incidence and mortality patterns in Europe: estimates for 40 countries and 25 major cancers in 2018. *Eur J Cancer* 103:356–387
3. De Angelis R, Sant M, Coleman MP, Francisci S, Baili P, Pierannunzio D et al (2014) Cancer survival in Europe 1999–2007 by country and age: results of EURO CARE–5—a population-based study. *Lancet Oncol* 15(1):23–34

4. Zaider T, Manne S, Nelson C, Mulhall J, Kissane D (2012) Loss of masculine identity, marital affection, and sexual bother in men with localized prostate cancer. *J Sex Med* 9(10):2724–2732
5. Carlsson S, Drevin L, Loeb S, Widmark A, Lissbrant IF, Robinson D et al (2016) Population-based study of long-term functional outcomes after prostate cancer treatment. *BJU Int* 117(6B):E36–E45
6. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD (2013) Reporting of patient-reported outcomes in randomized trials. *JAMA* 309:814–822
7. Basch E, Torda P, Adams K (2013) Standards for patient-reported outcome-based performance measures. *JAMA* 310(2):139–140
8. Resnick MJ, Barocas DA, Morgans AK, Phillips SE, Koyama T, Albertsen PC et al (2015) The evolution of self-reported urinary and sexual dysfunction over the last two decades: implications for comparative effectiveness research. *Eur Urol* 67(6):1019–1025
9. Basch E, Abernethy AP (2011) Supporting clinical practice decisions with real-time patient-reported outcomes. *J Clin Oncol* 29(8):954–956
10. Hartz A, He T, Stroppe S, Cutler DR, Andriole G, Dechet C (2013) Surgeon variation in patient quality of life after radical prostatectomy. *J Urol* 189(4):1295–1301
11. Rnic K, Linden W, Tudor I, Pullmer R, Vodermaier A (2013) Measuring symptoms in localized prostate cancer: a systematic review of assessment instruments. *Prostate Cancer Prostatic Dis* 16(2):111–122
12. Hoffman KE, Penson DF, Zhao Z, Huang L-C, Conwill R, Laviana AA et al (2020) Patient-reported outcomes through 5 years for active surveillance, surgery, brachytherapy, or external beam radiation with or without androgen deprivation therapy for localized prostate cancer. *JAMA* 323(2):149–163
13. National Health Service (2019) Patient Reported Outcome Measures (PROMs) [Internet]. Available from: <https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/patient-reported-outcome-measures-proms>
14. Waljee JF, Ghaferi A, Finks JF, Cassidy R, Varban O, Carlin A et al (2015) Variation in patient-reported outcomes across hospitals following surgery. *Med Care* 53(11):960–966
15. Nossiter J, Morris M, Cowling TE, Parry MG, Sujenthiran A, Aggarwal A et al (2021) Hospital volume and outcomes after radical prostatectomy: a national population-based study using patient-reported urinary continence and sexual function. *Prostate Cancer Prostatic Dis*. <https://doi.org/10.1038/s41391-021-00443-z>
16. Sibert NT, Pfaff H, Breidenbach C, Wesselmann S, Kowalski C (2021) Different approaches for case-mix adjustment of patient-reported outcomes to compare healthcare providers—methodological results of a systematic review. *Cancers (Basel)* 13(16):3964
17. ICHOM - International Consortium for Health Outcomes Measurement. Localized Prostate Cancer Data Collection Reference Guide (2015)
18. Kowalski C, Ferencz J, Albers P, Fichtner J, Wiegel T, Feick G et al (2016) Quality assessment in prostate cancer centers certified by the German Cancer Society. *World J Urol* 34(5):665–672
19. Szymanski KM, Wei JT, Dunn RL, Sanda MG (2010) Development and validation of an abbreviated version of the expanded prostate cancer index composite instrument for measuring health-related quality of life among prostate cancer survivors. *Urology* 76(5):1245–1250
20. Martin NE, Massey L, Stowell C, Bangma C, Briganti A, Bill-Axelsson A et al (2014) Defining a standard set of patient-centered outcomes for men with localized prostate cancer. *Eur Urol* 67(3):460–467
21. Sibert NT, Dieng S, Oesterle A, Feick G, Carl G, Steiner T et al (2021) Psychometric validation of the German version of the EPIC-26 questionnaire for patients with localized and locally advanced prostate cancer. *World J Urol* 39(1):11–25
22. Sanda MG, Wei MG, Litwin MS (2002) Scoring instructions for the expanded prostate cancer index composite short form (EPIC-26). The University of Michigan. Ann Arbor
23. Laviana AA, Zhao Z, Huang L-C, Koyama T, Conwill R, Hoffman K et al (2020) Development and internal validation of a web-based tool to predict sexual, urinary, and bowel function longitudinally after radiation therapy, surgery, or observation. *Eur Urol* 78(2):248–255
24. German Guideline Program in Oncology (2018) Interdisziplinäre Leitlinie der Qualität S3 zur Früherkennung, Diagnose und Therapie der verschiedenen Stadien des Prostatakarzinoms, Version 5.0, 1. Aktualisierung, AWMF-Registernummer: 043/022OL
25. D'Amico AV, Whittington R, Malkowicz SB, Schultz D, Blank K, Broderick GA et al (1998) Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA* 280(11):969–974
26. NHS England Analytical Team. Patient Reported Outcome Measures (PROMs) An alternative aggregation methodology for case-mix adjustment [Internet]. 2013 [cited 2020 Oct 7]. Available from: <https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/07/proms-agg-meth-adju.pdf>
27. Skolarus TA, Dunn RL, Sanda MG, Chang P, Greenfield TK, Litwin MS et al (2015) Minimally important difference for the Expanded Prostate Cancer Index Composite Short Form. *Urology* 85(1):101–105
28. Tyson MD, Koyama T, Lee D, Hoffman KE, Resnick MJ, Wu X-C et al (2018) Effect of prostate cancer severity on functional outcomes after localized treatment: comparative effectiveness analysis of surgery and radiation study results. *Eur Urol* 74(1):26–33
29. Collins GS, Reitsma JB, Altman DG, Moons KGM, members of the TRIPOD group. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. *Eur Urol*. 2015 Jun;67(6):1142–51.
30. Iezzoni L (2012) Risk adjustment for measuring healthcare outcomes, 4th edn. Health Administration Press, Chicago
31. Haese A, Knipper S, Isbarn H, Heinzer H, Tilki D, Salomon G et al (2019) A comparative study of robot-assisted and open radical prostatectomy in 10 790 men treated by highly trained surgeons for both procedures. *BJU Int* 123(6):1031–1040
32. Nyberg M, Hugosson J, Wiklund P, Sjöberg D, Wilderäng U, Carlsson SV et al (2018) Functional and oncologic outcomes between open and robotic radical prostatectomy at 24-month follow-up in the Swedish LAPPRO trial. *Eur Urol Oncol* 1(5):353–360
33. Vickers A, Savage C, Bianco F, Mulhall J, Sandhu J, Guillonneau B et al (2011) Cancer control and functional outcomes after radical prostatectomy as markers of surgical quality: analysis of heterogeneity between surgeons at a single cancer center. *Eur Urol* 59(3):317–322
34. Revicki D, Hays RD, Cella D, Sloan J (2008) Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 61(2):102–109

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

Nora Tabea Sibert<sup>1</sup> · Holger Pfaff<sup>2</sup> · Clara Breidenbach<sup>1</sup> · Simone Wesselmann<sup>1</sup> · Rebecca Roth<sup>3</sup> · Günther Feick<sup>4</sup> · Günter Carl<sup>4</sup> · Sebastian Dieng<sup>5</sup> · Amr A. Gaber<sup>6</sup> · Andreas Blana<sup>7</sup> · Christopher Darr<sup>8</sup> · Florian Distler<sup>9</sup> · Frank Kunath<sup>10</sup> · Jens Bedke<sup>11</sup> · Jörg Erdmann<sup>12</sup> · Jörg Minner<sup>13</sup> · Jörg Simon<sup>14</sup> · Maciej Kwiatkowski<sup>15</sup> · Martin Burchardt<sup>16</sup> · Nino Harz<sup>17</sup> · Stefan Conrad<sup>18</sup> · Thomas Höfner<sup>19</sup> · Thomas Knoll<sup>20</sup> · Burkhard Beyer<sup>21</sup> · Peter Hammerer<sup>22</sup> · Christoph Kowalski<sup>1</sup>

Holger Pfaff  
holger.pfaff@uk-koeln.de

Clara Breidenbach  
breidenbach@krebsgesellschaft.de

Simone Wesselmann  
wesselmann@krebsgesellschaft.de

Rebecca Roth  
rebecca.roth@uk-koeln.de

Günther Feick  
g-feick@gmx.de

Günter Carl  
guenther.carl@prostatakrebs-bps.de

Sebastian Dieng  
s.dieng@onkozert.de

Amr A. Gaber  
a.gaber@ctk.de

Andreas Blana  
Andreas.Blana@klinikum-fuerth.de

Christopher Darr  
christopher.darr@uk-essen.de

Florian Distler  
Florian.Distler@klinikum-nuernberg.de

Frank Kunath  
frank.kunath@uk-erlangen.de

Jens Bedke  
jens.bedke@med.uni-tuebingen.de

Jörg Erdmann  
joerg.erdmann@ckbm.de

Jörg Minner  
joerg.minner@glkn.de

Jörg Simon  
joerg.simon@ortenau-klinikum.de

Maciej Kwiatkowski  
maciej.kwiatkowski@ksa.ch

Martin Burchardt  
martin.burchardt@meduni-greifswald.de

Nino Harz  
nino.harz@klinikumdo.de

Stefan Conrad  
Stefan.Conrad@diakovere.de

Thomas Höfner  
Thomas.Hoefner@unimedizin-mainz.de

Thomas Knoll  
T.Knoll@klinikverbund-suedwest.de

Burkhard Beyer  
beyer.burkhard@gmail.com

Peter Hammerer  
p.hammerer@klinikum-braunschweig.de

Christoph Kowalski  
kowalski@krebsgesellschaft.de

<sup>1</sup> German Cancer Society, Kuno-Fischer-Straße 8, 14057 Berlin, Germany

<sup>2</sup> Faculty of Human Sciences, Faculty of Medicine and University Hospital Cologne, Institute of Medical Sociology, Health Services Research and Rehabilitation Science, University of Cologne, Eupener Str. 129, 50933 Cologne, Germany

<sup>3</sup> Faculty of Medicine and University Hospital Cologne, Institute of Medical Statistics and Computational Biology, University of Cologne, Robert-Koch-Str. 10, 50931 Cologne, Germany

<sup>4</sup> Federal Association of German Prostate Cancer Patient Support Groups, Bonn, Germany

<sup>5</sup> OnkoZert GmbH, Neu-Ulm, Germany

<sup>6</sup> Urologische Klinik, Carl-Thiem-Klinikum Cottbus, Thiemstr. 111, 03048 Cottbus, Germany

<sup>7</sup> Klinik für Urologie und Kinderurologie, Klinikum Fürth, Jakob-Henle-Strasse 1, 90766 Fürth, Germany

<sup>8</sup> Klinik für Urologie, Universitätsklinikum Essen, Hufelandstraße 55, 45147 Essen, Germany

<sup>9</sup> Klinik für Urologie, Universitätsklinik der Paracelsus Medizinischen Privatuniversität, Standort Klinikum Nürnberg, Prof.-Ernst-Nathan-Straße 1 (Haus 22), 90419 Nuremberg, Germany

<sup>10</sup> Department of Urology and Pediatric Urology, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), University Hospital Erlangen, Krankenhausstrasse 12, 91054 Erlangen, Germany

<sup>11</sup> Klinik für Urologie, Universität Tübingen, Hoppe-Seyler-Str. 3, 72076 Tübingen, Germany

<sup>12</sup> Prostatakarzinomzentrum Tauber-Franken, Umlandstr. 7, 97980 Bad Mergentheim, Germany

<sup>13</sup> Hegau-Bodensee-Klinikum GmbH, Virchowstraße 10, 78224 Singen, Germany

<sup>14</sup> Klinik für Urologie und Kinderurologie, Ortenau Klinikum, Ebertplatz 12, 77654 Offenburg, Germany

<sup>15</sup> Kantonsspital Aarau AG, Onkologiezentrum Mittelland, Tellstrasse 25, 5001 Aarau, Switzerland

<sup>16</sup> Klinik und Poliklinik für Urologie, Universitätsmedizin Greifswald, Ferdinand-Sauerbruch-Straße, 17475 Greifswald, Germany

<sup>17</sup> Klinikum Dortmund, Münsterstraße 240, 44145 Dortmund, Germany

<sup>18</sup> DIAKOVERE Friederikenstift, Humboldtstraße 5, 30169 Hannover, Germany

<sup>19</sup> Klinik und Poliklinik für Urologie und Kinderurologie, UNIVERSITÄTSMEDIZIN der Johannes

Gutenberg-Universität Mainz, Langenbeckstraße 1, 55131 Mainz, Germany

<sup>20</sup> Kliniken Sindelfingen, Arthur-Gruber-Str. 70, 71065 Sindelfingen, Germany

<sup>21</sup> Martini-Klinik Prostate Cancer Center Hamburg, Martinistraße 52, 20246 Hamburg, Germany

<sup>22</sup> Städtisches Klinikum Braunschweig, Freisestraße 9/10, 38118 Braunschweig, Germany

#### 4.4. Ergebnisqualität in zertifizierten Darmkrebszentren nach Resektion eines kolorektalen Karzinoms

Veröffentlicht als: Kowalski, C., **Sibert, N. T.**, Breidenbach, C., Hagemeyer, A., Roth, R., Seufferlein, T., Benz, S., Post, S., Siegel, R., Wiegering, A., Winkels, R., Bieck-Messemer, S., Fahlke, J., Reissfelder, C., Fuchs, M., Herzog, T., Weihrauch, R., Faber-Mertens, J., ... Wesselmann, S. (2022). Ergebnisqualität in zertifizierten Darmkrebszentren nach Resektion eines kolorektalen Karzinoms—Patient-Reported und kurzfristig-klinische Outcomes. Deutsches Ärzteblatt. <https://doi.org/10.3238/arztebl.m2022.0325>

## Originalarbeit

# Ergebnisqualität in zertifizierten Darmkrebszentren nach Resektion eines kolorektalen Karzinoms

Patient-Reported und kurzfristig-klinische Outcomes

Christoph Kowalski, Nora Tabea Sibert, [...] Simone Wesselmann

Deutsche Krebsgesellschaft, Berlin; PD Dr. Christoph Kowalski, Nora Tabea Sibert, PD Dr. Simone Wesselmann

\* An dieser Publikation sind weitere Autorinnen und Autoren beteiligt. Sie sind in der Zitierweise und am Ende des Beitrags aufgeführt. Dort befinden sich auch ihre Affiliationen.

## Zusammenfassung

**Hintergrund:** In dieser Beobachtungsstudie wurden Patient-Reported Outcomes (PROs) und kurzfristig-klinische Ergebnisqualitätsparameter bei Darmkrebs 12 Monate nach Behandlungsbeginn untersucht und die Ergebnisse von zertifizierten Darmkrebszentren miteinander verglichen.

**Methode:** Die Datenbasis bildeten die Angaben von 4 239 Patientinnen und Patienten mit Darmkrebs, die eine elektive Tumoresektion in einem von 102 Darmkrebszentren erhielten und prätherapeutisch in einem Fragebogen Auskunft zu ihrer Lebensqualität gaben (EORTC QLQ-C30 und -CR29). 3 142 (74,1 %) dieser Personengruppe füllten 12 Monate später einen posttherapeutischen Fragebogen aus. Für ausgewählte PROs, Anastomosensuffizienz und 30-Tage-Mortalität wurden Zusammenhangsanalysen berechnet und Casemix-adjustierte Zentrumsvergleiche vorgenommen.

**Ergebnisse:** Nach 12 Monaten zeigten sich im Mittel leichte Verbesserungen der Lebensqualität (66 versus 62 Punkte), Obstipation (16 versus 19) und bei Abdominalschmerzen (15 versus 17). Verschlechterungen erwiesen sich bei der physischen Funktion (75 versus 82) und bei Schmerzen (22 versus 19). Bessere PROs nach 12 Monaten waren mit besseren prätherapeutischen Scores assoziiert. Bei mindestens 3 der 5 genannten Scores waren bessere Ergebnisse mit dem männlichen Geschlecht, einer höheren Schulbildung, einem höheren Alter und einer privaten Krankenversicherung assoziiert. Hinsichtlich der fäkalen Inkontinenz zeigten sich große Verschlechterungen im Fall eines Rektumkarzinoms (ohne Stoma). Die größten Unterschiede zwischen Zentren fanden sich bei der physischen Funktion. 4,3 (Kolon) beziehungsweise 8,2 % (Rektum) der Patienten hatten eine Anastomosensuffizienz, 1,9 % starben innerhalb von 30 Tagen nach Resektion.

**Schlussfolgerung:** Klinikerinnen und Kliniker können die Ergebnisse nutzen, um Patientengruppen mit einem höheren Risiko für schlechte PROs zu identifizieren. Es ergaben sich Unterschiede zwischen Darmkrebszentren, die Maßnahmen zur Qualitätsentwicklung anregen können.

## Zitierweise

Kowalski C, Sibert NT, Breidenbach C, Hagemeyer A, Roth R, Seufferlein T, Benz S, Post S, Siegel R, Wiegering A, Winkels R, Bieck-Messemer S, Fahlke J, Reissfelder C, Fuchs M, Herzog T, Wehrauch R, Faber-Mertens J, Rudolph H, Puskás L, Kohlhaw K, Szczerbinska M, Scheuerlein H, Neumann PA, Hollerbach S, Riechmann M, Kolbe EW, Weigert N, König J, Klink C, Mussa S, Horn AK, Staib L, Werner J, Jähne J, Aly M, Mörk H, Grützmann R, Piso P, Dieng S, Wesselmann S: Outcome quality after colorectal cancer resection in Certified Colorectal Cancer Centers—patient-reported and short-term clinical outcomes. Dtsch Arztebl Int 2022; 119: 821–8. DOI: 10.3238/arztebl.m2022.0325

**D**armkrebs ist mit etwa 60 000 jährlichen Neuerkrankungen die bei Männern dritt- und bei Frauen zweithäufigste Krebserkrankung in Deutschland (1). Neben dem Überleben sind möglichst geringe Symptome und funktionelle Einschränkungen wesentliche Ziele der Leitlinie ([2], zum Beispiel Empfehlungen Kapitel 7.5 und folgende). Zur Erfassung von Symptomen und funktionellen Einschränkungen sind Patient-Reported Outcomes (PROs) besonders geeignet (3, 4). PROs messen „jegliche patientenseitige Ein-

schätzung ihres Gesundheitszustands, ihrer klinischen Ergebnisse, Mobilität und Lebensqualität“ ([5] S. 25, Übersetzung nach [6]). Ein Beispiel für ein PRO ist „fakale Inkontinenz“, die im auch hier verwendeten Fragebogen EORTC QLQ-CR29 mit der Einzelfrage „Während der letzten Woche: Hatten Sie ungewollte Stuhlabgänge aus Ihrem Darmausgang?“ und den Antwortmöglichkeiten „überhaupt nicht“, „wenig“, „mäßig“, „sehr“ abgefragt wird (7). Randomisierte Studien konnten den Nutzen von PROs zur Behandlungspla-

TABELLE 1

Stichprobenbeschreibung			
	Kolon (n = 2859)	Rektum (n = 1 380)	Total (N = 4 239)
<b>Alter</b>			
Mittelwert (SD)	70,8 (11,6)	67,7 (11,2)	69,8 (11,5)
Median (IQR)	73 (63–79)	68 (60–77)	71 (62–79)
<b>Alter (gruppiert)</b>			
≤ 39	35 (1,2 %)	9 (0,7 %)	44 (1,0 %)
40–49	94 (3,3 %)	59 (4,3 %)	153 (3,6 %)
50–59	357 (12,5 %)	275 (19,9 %)	632 (14,9 %)
60–69	686 (24,0 %)	413 (29,9 %)	1 099 (25,9 %)
70–79	983 (34,4 %)	385 (27,9 %)	1 368 (32,3 %)
≥ 80	704 (24,6 %)	239 (17,3 %)	943 (22,2 %)
<b>Geschlecht</b>			
männlich	1 580 (55,3 %)	891 (64,6 %)	2 471 (58,3 %)
weiblich	1 279 (44,7 %)	489 (35,4 %)	1 768 (41,7 %)
<b>ASA-Score</b>			
ASA 1	113 (4,0 %)	57 (4,1 %)	170 (4,0 %)
ASA 2	893 (31,2 %)	439 (31,8 %)	1 332 (31,4 %)
ASA 3	707 (24,7 %)	331 (24,0 %)	1 038 (24,5 %)
ASA 4	43 (1,5 %)	10 (0,7 %)	53 (1,3 %)
ASA 5	0 (0,0 %)	0 (0,0 %)	0 (0,0 %)
fehlend	1 103 (38,6 %)	543 (39,3 %)	1 646 (38,8 %)
<b>Staatsbürgerschaft</b>			
deutsch	2 729 (95,5 %)	1 321 (95,7 %)	4 050 (95,5 %)
andere	112 (3,9 %)	43 (3,1 %)	155 (3,7 %)
fehlend	18 (0,6 %)	16 (1,2 %)	34 (0,8 %)
<b>Versicherung</b>			
gesetzlich	2 449 (85,7 %)	1 171 (84,9 %)	3 620 (85,4 %)
privat	367 (12,8 %)	183 (13,3 %)	550 (13,0 %)
andere/nicht versichert	20 (0,7 %)	10 (0,7 %)	30 (0,7 %)
fehlend	23 (0,8 %)	16 (1,2 %)	39 (0,9 %)
<b>Bildung</b>			
Haupt-/Volksschule	1 312 (45,9 %)	596 (43,2 %)	1 908 (45,0 %)
Realschule/Mittlere Reife	560 (19,6 %)	306 (22,2 %)	866 (20,4 %)
polytechnische Oberschule	152 (5,3 %)	99 (7,2 %)	251 (5,9 %)
Fachhochschulreife	249 (8,7 %)	114 (8,3 %)	363 (8,6 %)
Abitur/allg. Hochschulreife	443 (15,5 %)	202 (14,6 %)	645 (15,2 %)
andere	67 (2,3 %)	27 (2,0 %)	94 (2,2 %)
kein Abschluss	25 (0,9 %)	16 (1,2 %)	41 (1,0 %)
fehlend	51 (1,8 %)	20 (1,4 %)	71 (1,7 %)
<b>UICC-Stadium</b>			
UICC I	812 (28,4 %)	498 (36,1 %)	1 310 (30,9 %)
UICC II	961 (33,6 %)	332 (24,1 %)	1 293 (30,5 %)
UICC III	773 (27,0 %)	395 (28,6 %)	1 168 (27,6 %)
UICC IV	313 (10,9 %)	155 (11,2 %)	468 (11,0 %)

nung oder zum Monitoring von Patientinnen und Patienten mit einer Krebserkrankung zeigen: Ein besseres funktionales Outcome und längeres Überleben zeigte sich nach intensiviertem PRO-Monitoring (8–10), eine Verbesserung beeinträchtigter Lebensqualität zeigte sich durch eine maßgeschneiderte interdisziplinäre PRO-Intervention (4, 11).

Obwohl bei der Entwicklung von PROs bereits intendiert (12) und von vielen Fachgruppen befürwortet, (zum Beispiel 13, 14) werden PROs zur Beurteilung der Ergebnisqualität und dem Vergleich von Leistungserbringenden bislang wenig genutzt. Die während der Therapie auftretenden Symptome und Funktionseinschränkungen (15) sind jedoch so bedeutsam, dass PROs als eine wichtige patientenzentrierte Ergänzung von Qualitätssicherungs- und -entwicklungsverfahren anzusehen sind (16). Es gibt bereits einige vielversprechende Initiativen im Ausland, beispielsweise die standardisierte Erfassung von PROs in amerikanischen Physiotherapiepraxen, die Praxisvergleiche ermöglichen (17). In Großbritannien gibt es das National Prostate Cancer Audit des National Health Service (NHS) (18). Beispiele aus Deutschland sind das QS-Reha-Verfahren (19) und die PCO-Studie beim Prostatakrebs (20). Insgesamt werden PROs in der Qualitätsforschung in Deutschland noch zu wenig berücksichtigt (13).

Dieser Beitrag stellt Ergebnisse der EDIUM-Studie („Ergebnisqualität bei Darmkrebs: Identifikation von Unterschieden und Maßnahmen zur flächendeckenden Qualitätsentwicklung“) vor, in der PROs und kurzfristige klinische Ergebnisqualitätsparameter von über 100 zertifizierten Darmkrebszentren verglichen werden. Die Studie wurde vom Innovationsausschuss beim gemeinsamen Bundesausschuss von Juli 2018 bis Dezember 2021 gefördert (Förderkennzeichen VSF1\_2017–169). Zunächst werden Zusammenhänge zwischen Patienten- und Zentrummerkmalen und ausgewählten Dimensionen des EORTC QLQ-C30 und -CR29 (7, 21) beschrieben. Anschließend werden die auf diesen Analysen aufbauenden Casemix-adjustierten Vergleiche zwischen den Zentren berichtet. Zusätzlich werden Ergebnisse für die kurzfristig-klinischen Parameter Anastomosensuffizienz und 30-Tage-Mortalität vorgestellt.

**Methode**

**Patientenkollektiv und Datenerhebung**

Knapp die Hälfte der jährlich neuerkrankten Patienten mit Darmkrebs werden in von der Deutschen Krebsgesellschaft zertifizierten Zentren behandelt (22). Aus den deutschen Darmkrebszentren wurde eine stratifizierte Zufallsstichprobe mit Nachrückverfahren gezogen. Alle Primärfälle mit kolorektalem Karzinom der initial 106 teilnehmenden Zentren, die 2019 (in sechs Pilotzentren ab Oktober 2018) eine elektive Tumorresektion erhielten oder nichtoperativ palliativ behandelt wurden, sollten von den Zentren zur Studienteilnahme eingeladen werden. In der hier vorgestellten Analyse wurden lediglich Patienten mit elektiver Tumorresektion mit und ohne (neo)adjuvanter Therapie berücksichtigt.

Notfallpatienten waren ausgeschlossen, weil sie keinen prätherapeutischen Fragebogen ausfüllen konnten. Nichtoperativ palliativ versorgte Patienten wurden gesondert ausgewertet, die Ergebnisse wurden im öffentlich zugänglichen Ergebnisbericht an den Förderer (G-BA) berichtet. Vor Therapiebeginn (Tumorresektion beziehungsweise neoadjuvante Therapie) füllten die Teilnehmenden den Baseline-Fragebogen (T0, EORTC-QLQ und soziodemografische Angaben) wahlweise auf Papier oder per Web-Anwendung aus, wobei nicht alle Zentren die Online-Option ermöglichten. Die Nachbefragung (T1, EORTC-QLQ) erfolgte 12 Monate nach Tumorresektion. Die Zentren dokumentierten zusätzlich im Rahmen der Zertifizierung klinische und Behandlungsmerkmale, die im Zentrum mittels der Software OncoBox mit den Befragungsdaten verknüpft wurden. Die einzelnen Zentrumsdatensätze wurden dem Zertifizierungsinstitut OnkoZert zur Qualitätssicherung und der Studienleitung zur Auswertung übermittelt. Das Studienprotokoll (DRKS00008724) sieht die Auswertung der EORTC-Dimensionen nach 12 Monaten als primäre Endpunkte und die Anastomoseninsuffizienz getrennt nach Kolon- und Rektumkarzinom sowie die 30-Tage-Mortalität als sekundäre Endpunkte vor. Aus Platzgründen wird hier eine Auswahl von EORTC-Dimensionen (siehe folgender Abschnitt), die Anastomoseninsuffizienz und 30-Tage-Mortalität nach Resektion berichtet. Die übrigen Endpunkte werden gemäß Protokoll im öffentlich zugänglichen Ergebnisbericht vorgestellt.

**Variablen und statistische Analysen**

Untersucht wurden zunächst je fünf für das Kolon- und Rektumkarzinom relevante Dimensionen („Lebensqualität“, „physische Funktion“, „Schmerzen“, „Obstipation“, „Abdominalschmerzen“) des EORTC QLQ-C30 und -CR29. Die Scores wurden aufgrund einer Befragung der beteiligten Zentren danach ausgewählt, wie relevant die Dimensionen für ihre Behandlungsplanung sind (23). Die Scorebildung erfolgte gemäß den Manualen, wobei Werte von 0–100 angenommen werden können. Für die drei C30-Dimensionen „physische Funktion“ (Schwellenwert: 83), „Schmerzen“ (Schwellenwert: 25), „Obstipation“ (Schwellenwert: 50) liegen Schwellenwerte für die klinische Relevanz vor (24). Bei Funktionsskalen („Lebensqualität“, „physische Funktion“) deuten höhere Werte auf eine bessere Funktion hin, bei Symptomskalen („Schmerzen“, „Obstipation“, „Abdominalschmerzen“) deuten sie auf schwere Symptome hin. Aufgrund der Bedeutung insbesondere für das Rektumkarzinom (23) wurden zusätzlich Veränderungen im Score „fäkale Inkontinenz“ von T0 zu T1 betrachtet. Abweichend vom Vorgehen für die fünf anderen Scores erfolgte dies aufgrund von Besonderheiten des Scores getrennt nach Lokalisation (Kolon/Rektum) und Stoma nach 12 Monaten nur kreuztabellarisch (*eMethodenteil*). Die sekundären Endpunkte Anastomoseninsuffizienz (bei Patienten mit Anlage einer Anastomose) und 30-Tage-Mortalität nach Resektion wurden im Rahmen der Zertifizierung dokumentiert.

	Kolon (n = 2859)	Rektum (n = 1 380)	Total (N = 4 239)
<b>Stoma (primär angelegt)</b>			
ja	192 (6,7 %)	1 010 (73,2 %)	1 202 (28,4 %)
nein	2 667 (93,3 %)	370 (26,8 %)	3 037 (71,6 %)
<b>Stoma (nach 12 Monaten vorhanden)</b>			
nein	2 025 (70,8 %)	632 (45,8 %)	2 657 (62,7 %)
ja	102 (3,6 %)	325 (23,6 %)	427 (10,1 %)
fehlend	732 (25,6 %)	423 (30,7 %)	1 155 (27,2 %)
<b>verstorben</b>			
nicht verstorben/unbekannt	2 607 (91,2 %)	1 265 (91,7 %)	3 872 (91,3 %)
verstorben	252 (8,8 %)	115 (8,3 %)	367 (8,7 %)
<b>verstorben (30 Tage)</b>			
mehr als 30 Tage überlebt	2 800 (97,9 %)	1 359 (98,5 %)	4 159 (98,1 %)
30 oder weniger Tage überlebt	59 (2,1 %)	21 (1,5 %)	80 (1,9 %)

Fehlende Werte sind nur angegeben für Variablen, bei denen fehlende Werte vorkommen. ASA, American Society of Anesthesiologists; IQR, Interquartilsabstand; SD, Standardabweichung; UICC, „union internationale contre le cancer“

Die für die Zusammenhangsanalysen und die Casemix-Adjustierung genutzten Kovariaten werden im *eMethodenteil* aufgeführt und neben anderen in *Tabelle 1* und *Tabelle 2* beschrieben.

Die primären Zusammenhangsanalysen für die fünf oben genannten EORTC-Scores erfolgten entsprechend dem Vorgehen zur Casemix-Adjustierung als multivariable lineare Regressionen, dargestellt mithilfe von Forest Plots (*eTabellen 1*). Die Casemix-Adjustierung wurde angelehnt an das Vorgehen des NHS (25) durchgeführt. Pro Score wurden dabei nur Zentren berücksichtigt, die posttherapeutische Fragebogen von mindestens zehn Patienten aufwiesen. Grafisch wurden die adjustierten Werte in Beziehung zu behelfsweisen „minimally important differences“ (MIDs) (26–28) gesetzt. MIDs bezeichnen die kleinsten Veränderungen eines Scores, die Patienten als wichtig wahrnehmen. Details zum statistischen Vorgehen für die primären und sekundären Endpunkte und zu zusätzlichen Analysen (gemischte und Tobit-Modelle) sind im *eMethodenteil* dargestellt.

**Ergebnisse**

**Stichprobe**

Zwischen Oktober 2018 und Dezember 2019 wurden 4 239 Patienten mit Tumorresektion aus 102 Zentren prätherapeutisch befragt (*Tabelle 1*). Knapp 7 % der Teilnehmenden füllte prätherapeutisch den Online-Bogen aus, die übrigen nutzten die Papierversion. Die mediane Ausschöpfung der Zentren in der Gesamtstudie betrug 51 %. 3 142 (74,1 %) der Patienten nahmen an der Nachbefragung teil und bilden die Grundlage für die Zusammenhangsanalysen. Die Drop-out-Quote liegt bei 25,9 %. Circa ein Drittel davon bestehen aus dokumentierten Todesfällen, die Gründe für die restlichen Fälle sind unbekannt (*Tabelle 1, eGrafik 1*).

**TABELLE 2**  
**EORTC-Scores für Patienten, die an der Nachbefragung teilgenommen haben (N = 3 142)**

		T0	T1
Lebensqualität	n fehlend	41	18
	Durchschnitt (SD)	62 (23)	66 (21)
	Median (IQR)	67 (50–83)	67 (50–83)
physische Funktion	n fehlend	7	9
	Durchschnitt (SD)	82 (22)	75 (24)
	Median (IQR)	93 (73–100)	80 (60–93)
Schmerzen	n fehlend	3	10
	Durchschnitt (SD)	19 (28)	22 (28)
	Median (IQR)	0 (0–33)	17 (0–33)
Obstipation	n fehlend	30	31
	Durchschnitt (SD)	19 (31)	16 (26)
	Median (IQR)	0 (0–33)	0 (0–33)
Abdominalschmerzen	n fehlend	23	19
	Durchschnitt (SD)	17 (27)	15 (24)
	Median (IQR)	0 (0–33)	0 (0–33)

IQR, Interquartilsabstand, SD, Standardabweichung, T0, Befragung vor Behandlungsbeginn; T1, Befragung 12 Monate nach Tumorresektion

**Patient-Reported Outcomes**

Gegenüber der Baseline-Erhebung zeigen sich, wie in *Tabelle 2* zu sehen, nach 12 Monaten im Mittel eine leicht verbesserte Lebensqualität (66 versus 62 Punkte), geringere Obstipation (16 versus 19) und Abdominalschmerzen (15 versus 17). Verschlechterungen zeigen sich im Mittel bei der physischen Funktion (75 versus 82) und bei den Schmerzen (22 versus 19). Die arithmetischen Mittel der drei Dimensionen mit vorliegenden Schwellenwerten für die klinische Relevanz betragen posttherapeutisch 75 (physische Funktion), 22 (Schmerzen) und 16 (Obstipation). Es wird also im Mittel bei der physischen Funktion eine Beeinträchtigung identifiziert, die bei der Behandlungsplanung relevant sein kann (24: 2).

*eTabelle 2* zeigt, dass es bei der fäkalen Inkontinenz große Verschlechterungen bei Patienten mit Kolonkarzinom (im Durchschnitt von 8 auf 26 Punkte, n = 63) und Patienten mit Rektumkarzinom ohne Stoma (im Durchschnitt von 14 auf 27 Punkte, n = 522) gibt.

Die multivariablen linearen Regressionsanalysen (*Grafik 1*, exemplarisch für Schmerzen, *eTabellen 1*, *eGrafik 2*) klären 11–45 % der Varianz (R<sup>2</sup>) der abhängigen Variablen auf.

Die prätherapeutischen Scores waren in allen Modellen positiv mit den PROs assoziiert. Bessere PROs nach 12 Monaten waren mit einem besseren Allgemeinzustand (Lebensqualität, Schmerzen), einem Kolonkarzinom (Obstipation, Lebensqualität), mit einer höheren Schulbildung (Schmerzen, physische Funktion, Lebensqualität), einer privaten Krankenversicherung

(Obstipation, Schmerzen, physische Funktion, Lebensqualität), dem männlichen Geschlecht (Abdominalschmerzen, physische Funktion, Obstipation) sowie eher niedrigerem Stadium der „union internationale contre le cancer“ (UICC) bei physischer Funktion und Lebensqualität, aber höherem Stadium bei Abdominalschmerz und Obstipation assoziiert. Über 79-Jährige haben in zwei, 50–59-Jährige in vier Scores schlechtere PROs als die Referenzgruppe der 70–79-Jährigen. Während ein 12 Monate nach Therapie vorhandenes Stoma eine Einschränkung für die physische Funktion und Lebensqualität bedeutet, ist es mit einem besseren Score im Bereich Obstipation assoziiert. Die Prüfung der Modellannahmen zeigte bezüglich der Verteilung der Residuen Abweichungen von der Normalverteilung insbesondere für Obstipation sowie Abdominalschmerzen. Tobit-Modelle zeigen größere Schätzer der Baseline-Scores und kommen ansonsten zu ähnlichen Ergebnissen (*eTabelle 3*). Aufgrund von Singularität konnten lediglich für die Lebensqualität Mehrebenenmodelle berechnet werden, wobei der Anteil der Varianz auf Zentrums Ebene klein war (ICC < 1 %) und Zentrumsmerkmale die Varianz im einzigen konvergierenden Mehrebenenmodell nicht erklärten (*eTabellen 4*).

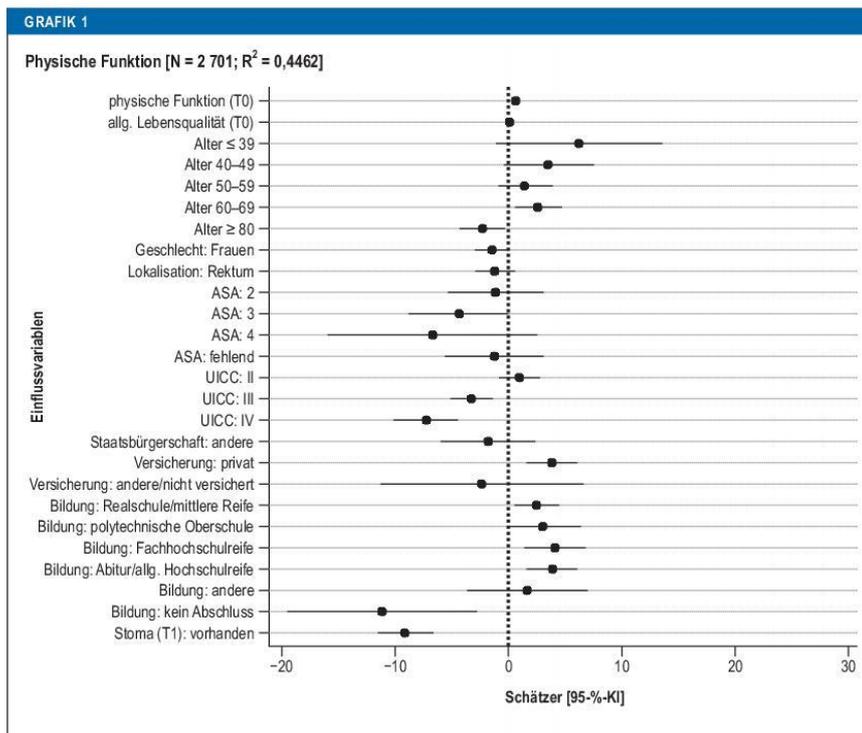
Beim Zentrumsvergleich zeigen sich nach Casemix-Adjustierung grafisch Unterschiede in den Scores zwischen den Zentren (*Grafik 2*, exemplarisch für Schmerzen und physische Funktion, *eGrafik 3*, beobachtete und adjustierte Werte auf Anfrage). Der größte Interquartilsabstand zeigt sich mit 69–79 beim Score physische Funktion, bei dem er größer ist als eine MID. Bei den anderen Scores betragen die Interquartilsabstände 4, 7, 5 und 7 Punkte und sind kleiner als eine MID (7, 10, 10, 9). Erkennbar werden Zentren mit deutlich unter- und überdurchschnittlichen Ergebnissen, insbesondere bei der physischen Funktion. Deskriptiv zeigen sich hinsichtlich der Zusammenhänge zwischen den adjustierten Scores der Zentren und den Zentrumsmerkmalen operative Fallzahl, Trägerschaft, Lehrstatus und Urbanitätsgrad keine deutlichen Muster (*eTabellen 5–8*).

**Anastomoseninsuffizienz und 30-Tage-Mortalität**

Von den Patienten mit angelegter Anastomose hatten 4,3 (Kolon, 121/2783) beziehungsweise 8,2 % (Rektum, 87/1061) eine Anastomoseninsuffizienz, 1,9 % (80/4239) starben innerhalb von 30 Tagen nach der Resektion. Für die multivariablen logistischen Regressionsanalysen der kurzfristigen klinischen Endpunkte Anastomoseninsuffizienz und 30-Tage-Mortalität ergeben sich R<sup>2</sup> < 0,03. Die Ergebnisse sind in *eGrafik 4* und *eGrafik 5* dargestellt. Aufgrund der geringen Varianzaufklärung für die sekundären Endpunkte wird auf eine Darstellung der Casemix-adjustierten Ergebnisse verzichtet.

**Diskussion**

Dieser Beitrag untersuchte Zusammenhänge zwischen Patientenmerkmalen und PROs sowie kurzfristigen klinischen Ergebnisqualitätsparametern bei Darmkrebs 12 Monate nach elektiver Tumorresektion und verglich die



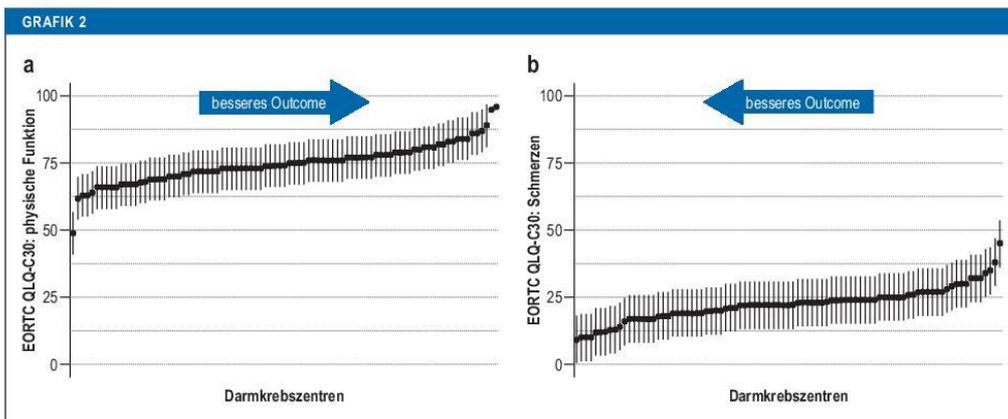
**Ergebnisse der multivariablen linearen Regressionsanalysen:** Schätzer (95%-Konfidenzintervalle [95%-KI])  
 ASA, American Society of Anesthesiologist; UICC, „union internationale contre le cancer“

Ergebnisse zwischen zertifizierten Darmkrebszentren. Die Zusammenhangsanalysen geben zunächst Hinweise auf bislang wenig untersuchte Korrelate posttherapeutischer PROs. Diese können von Klinikern und Klinikern genutzt werden, um Patienten mit einem höheren Risiko für schlechtere PROs nach einem Jahr zu identifizieren. Insbesondere sind die teils starken Zusammenhänge bei den sozioökonomischen Merkmalen bemerkenswert. Beispielsweise ist die allgemeine Lebensqualität unter Kontrolle der klinischen Merkmale und dem Baselinewert bei Privatversicherten um 5 und bei Personen mit Abitur um 4 Punkte höher als bei gesetzlich Versicherten beziehungsweise Personen mit Hauptschulabschluss. Hinweise auf Zusammenhänge zwischen der Ergebnisqualität und dem sozioökonomischen Status konnten zuvor bereits für Deutschland gezeigt werden (29). Hinsichtlich American Society of Anesthesiologists (ASA) zeigen sich erwartungsgemäß überwiegend schlechtere PROs mit schwerer Erkrankung. Bemerkenswert ist, dass es 50- bis 59-Jährigen bei vier Scores schlechter geht als den 70- bis 79-Jährigen.

Beim Zentrumsvergleich zeigen sich hinsichtlich der Interquartilsabstände Unterschiede in Höhe zumeist et-

wa einer MID, wobei die Spreizung insgesamt erheblich ist. Die Casemix-Adjustierungen und die grafischen Darstellungen erlauben die Identifikation einzelner Zentren mit deutlich schlechteren Ergebnissen, als aufgrund des Casemix zu erwarten war. Am größten sind die Unterschiede beim Score „physische Funktion“ (Grafik 2), wo der Interquartilsabstand zehn Punkte beträgt und damit größer als eine MID ist. Grundlage dieses Scores sind fünf Items und möglicherweise vermag er deshalb besser zu differenzieren als die übrigen Scores, die aus einem oder zwei Items errechnet werden. Die Items sind darmkrebsunspezifisch und von erheblicher Relevanz für die Patienten. Ein Beispiel dafür ist das Item „Bereitet es Ihnen Schwierigkeiten, einen längeren Spaziergang zu machen?“ mit den Antwortkategorien „überhaupt nicht“, „wenig“, „mäßig“, „sehr“. Die in Zusatzanalysen betrachteten Zentrumsmerkmale konnten nicht oder nicht wesentlich zur Varianzaufklärung beitragen.

Hinsichtlich der Veränderung von vor der Operation bis 12 Monate danach deuten niederländische Ergebnisse in eine ähnliche Richtung (30, 31). Auch dort zeigen sich nach 12 Monaten beispielsweise nur noch geringe Unterschiede zum Ausgangswert, allerdings mit



**Casemix-Adjustierung – Punktschätzer der adjustierten posttherapeutischen Patient-Reported Outcomes (PROs)-Scores der Zentren.** Dargestellt sind ausschließlich Zentren mit mindestens zehn posttherapeutischen Fragebogen. Jeder Punkt repräsentiert ein anonymisiertes Zentrum.  
 a) Funktionsscore (je höher, desto besser das Outcome): physische Funktion (75, 69–79)  
 b) Symptomsscore (je niedriger, desto besser das Outcome): Schmerzen (23, 19–26)  
 Die Striche (sogenannte „Antennen“) stellen nach oben und unten behelfsmäßige „minimally important differences“ (MIDs) als 1/3 der Standardabweichung auf Patientenebene dar: physische Funktion: 7; Schmerzen: 10

teils ausgeprägten Unterschieden bei der physischen Funktion. Insgesamt gibt es nur wenig vergleichbare Daten aus dem Ausland.

Die Häufigkeit der kurzfristigen klinischen Endpunkte Anastomoseninsuffizienz und 30-Tage-Mortalität in unserer Stichprobe ist fast identisch zu der Häufigkeit in den zertifizierten Zentren insgesamt (22). Abgesehen vom Geschlecht und der Lokalisation (Anastomoseninsuffizienz) beziehungsweise dem Alter und ASA (30-Tage-Mortalität) zeigten sich keine Faktoren, die mit den kurzfristigen klinischen Endpunkten zusammenhängen und die Varianzaufklärung der Modelle für diese Endpunkte ist insgesamt gering. Einige aus der Literatur bekannte Risikofaktoren, zum Beispiel der Body-Mass-Index, wurden im Rahmen dieser Studie nicht erhoben. Für den Casemix-adjustierten Vergleich der kurzfristigen klinischen Endpunkte scheint das hier vorgestellte Verfahren daher nicht geeignet, dafür wären unter anderem zusätzliche Adjustorvariablen erforderlich.

Nach Kenntnis der Autorinnen und Autoren handelt es sich um die erste Untersuchung, in der PROs zwischen Versorgungseinrichtungen bei Patienten mit Darmkrebs verglichen werden. In der Vergangenheit wurden für andere Erkrankungen Unterschiede zwischen Behandelnden gefunden, beispielsweise bei orthopädischen Diagnosen (15), Prostatakrebs (28) und weiteren chirurgischen Eingriffen (32). Dort sind die Unterschiede teils deutlich größer. Die Unterschiede zwischen den Zentren fallen möglicherweise aufgrund der relativen Homogenität zertifizierter Zentren geringer aus als dies bei einer Stichprobe weiterer Krankenhäuser gewesen wäre. Zertifizierte Zentren müssen zahlreiche Qualitätsanforderungen umsetzen, Mindest-

mengen erreichen und haben im Durchschnitt eine bessere onkologische Ergebnisqualität als nichtzertifizierte Einrichtungen (33–35). Sie stellen damit eine Selektion aller Darmkrebs behandelnden Einrichtungen dar. Dies trifft allerdings auch auf Prostatakrebszentren zu, bei welchen die Unterschiede in den PROs größer sind (28). Eine vergleichbare häufige Funktionseinschränkung wie Inkontinenz und erektile Dysfunktion beim Prostatakrebs gibt es nach Resektion beim Darmkrebs zwar nicht, allerdings lässt der anspruchsvollere Eingriff beim Rektum durchaus Unterschiede erwarten. Hier sind zukünftig getrennte Betrachtungen nach Lokalisation und Stomastatus sinnvoll.

EORTC QLQ-C30 und -CR29 werden unter anderem im ICHOM-Standardset für Darmkrebs empfohlen (36). Allerdings gibt es keine Empfehlungen zur Priorisierung der über 30 Scores bei Patienten mit Darmkrebs sowie darüber, ob stratifizierte Auswertungen erforderlich sind. Weitere Forschungen dazu sind nötig. Ebenso ist noch zu erforschen, welche Bedeutung unterschiedliche Operationsverfahren hinsichtlich PROs sowie Rehabilitation und Nachsorge haben, die nur teilweise durch die Darmkrebszentren beeinflusst werden können.

Bei der Interpretation der Ergebnisse sind einige Stärken und Schwächen zu berücksichtigen: Anders als bei vielen anderen Messungen der Versorgungsqualität sind bei Befragungsdaten Rekrutierungsanstrengungen seitens der Zentren und aktives Zutun der Patienten erforderlich. Im Median wurden 51 % der einschussfähigen Patienten befragt, was zufriedenstellend ist. Etwa ein Viertel der Patienten nahmen nicht an der Nachbefragung teil, für ein Drittel von ihnen ist ein Versterben dokumentiert. Unter den übrigen waren die Patienten,

die keine Angaben machten, zu T0 durchschnittlich kränker (auf Anfrage).

Die hier genutzte klinische Basisdokumentation erfolgte zentrumsübergreifend einheitlich, wird auch verpflichtend für das Zertifizierungsverfahren genutzt und ist von entsprechend hoher Vollständigkeit bei zugleich externer Validierung durch die Auditierenden während der Zertifizierung. Ein hoher Anteil fehlender Werte ist allerdings bei der ASA-Klassifikation zu finden (*eMethodenteil*). Auch wenn es sich im deutschen Vergleich um fallzahlstarke Zentren handelt, können kleine Fallzahlen in einigen Zentren (mediane Primärfallzahl elektiv operierter Patienten: 76) ebenso wie der selektive Patienteneinschluss Ergebnisse möglicherweise verzerren. Zentren, für die ausreichend Patientendaten für die Casemix-Adjustierung vorlagen, unterschieden sich nach Trägerschaft, Lehrstatus und Urbanisierung des Standorts nicht von denen, für die nicht ausreichend Daten vorlagen, hatten aber höhere Primärfallzahlen (auf Anfrage).

Abschließend lässt sich festhalten, dass ein Messen von PROs zum Ergebnisqualitätsvergleich insbesondere zur Identifikation von Zentren mit deutlich unter- und überdurchschnittlichen Ergebnissen sinnvoll sein kann. Den Behandelnden wird damit ein zusätzliches Instrument an die Hand gegeben, um zu verstehen, was Patienten brauchen und was getan werden kann, um die Versorgung weiter zu verbessern. Die hier dargestellten und zentrumspezifisch zur Verfügung gestellten Ergebnisse können dazu anregen, Maßnahmen zur Qualitätsverbesserung abzuleiten, beispielsweise die Entwicklung von Maßnahmen für Betroffene mit besonderem Risiko für reduzierte Funktion, die Betrachtung in Zertifizierungsaudits, in internen Qualitätszirkeln oder die Identifikation von Zentren mit besonders guten Ergebnissen. So wird gegenseitiges Lernen gefördert (37).

#### Weitere Autorinnen und Autoren

Clara Breidenbach, Anna Hagemeyer, Rebecca Roth, Thomas Seufferlein, Stefan Benz, Stefan Post, Robert Siegel, Armin Wiegering, Raphael Winkels, Stefanie Bieck-Messemer, Jörg Fahlke, Christoph Reissfelder, Martin Fuchs, Torsten Herzog, Richard Wehrauch, Julia Faber-Mertens, Hagen Rudolph, László Puskás, Kay Kohlhaw, Malgorzata Szczerbinska, Hubert Scheuerlein, Philipp-Alexander Neumann, Stephan Hollerbach, Maren Rechmann, Ernst W. Kolbe, Norbert Weigert, Jörg Königer, Christian Klink, Shueb Mussa, Anja-Kathrin Horn, Ludger Stab, Jens Werner, Joachim Jähne, Mohamed Aly, Hubert Mork, Robert Grützmann, Pompilio Piso, Sebastian Dieng

#### Affiliationen der weiteren Autorinnen und Autoren

Deutsche Krebsgesellschaft, Berlin: Clara Breidenbach  
 Institut für Medizinische Statistik und Bioinformatik, Medizinische Fakultät, Universität zu Köln, Uniklinikum Köln: Anna Hagemeyer, Dr. Rebecca Roth  
 Uniklinik Ulm: Prof. Dr. Thomas Seufferlein  
 Klinikverbund Südwest: Prof. Dr. Stefan Benz, Prof. Dr. Hubert Mork  
 Universitätsklinikum Mannheim: Prof. Dr. Stefan Post, Prof. Dr. Christoph Reissfelder  
 Helios Klinikum Berlin-Buch: PD Dr. Robert Siegel  
 Universitätsklinikum Würzburg: Prof. Dr. Armin Wiegering  
 Kliniken des Main-Taunus-Kreises, Bad Soden: Dr. Raphael Winkels  
 Westfälisch-Klinikum, Kaiserslautern: Dr. Stefanie Bieck-Messemer  
 Johanniter-Krankenhaus, Stendal: Prof. Dr. Jörg Fahlke  
 Klinik Bogenhausen, München: Dr. Martin Fuchs

Katholisches Klinikum Bochum: PD Dr. Torsten Herzog  
 Diakonissenkrankenhaus Dresden: Richard Wehrauch  
 St. Elisabeth-Krankenhaus Geilenkirchen: Dr. med. Julia Faber-Mertens  
 Klinikum Chemnitz: Dr. Hagen Rudolph  
 Klinikum Memmingen: Dr. László Puskás  
 Sana Klinikum Borna: Dr. Kay Kohlhaw  
 Johannes Wesling Klinikum Minden: Malgorzata Szczerbinska  
 St. Vincenz Krankenhaus Paderborn: PD Dr. Hubert Scheuerlein  
 Klinikum rechts der Isar, München: PD Dr. Philipp-Alexander Neumann  
 Klinik für Gastroenterologie, Celle: Prof. Dr. Stephan Hollerbach  
 Sana Klinikum Hof/Saale: Dr. Maren Rechmann  
 Klinikum Herford: Dr. Ernst W. Kolbe  
 Klinikum St. Elisabeth Straubing: Prof. Dr. Norbert Weigert  
 Kathannenhospital Stuttgart: Prof. Dr. Jörg Königer  
 Diakonissen-Stiftungs-Krankenhaus Speyer: Prof. Dr. Christian Klink  
 Ev. Diakonissenkrankenhaus Leipzig: Dr. Shueb Mussa  
 Sozialstiftung Bamberg: Dr. Anja-Kathrin Horn  
 Klinikum Esslingen: Prof. Dr. Ludger Stab  
 Klinikum der Universität München: Prof. Dr. Jens Werner  
 DIAKOVERE Henriettenstift, Hannover: Prof. Dr. Joachim Jähne  
 Krankenhaus Landshut-Achdorf: Mohamed Aly  
 Universitätsklinikum Erlangen: Prof. Dr. Robert Grützmann  
 Krankenhaus Barmherzige Brüder Regensburg: Prof. Dr. Pompilio Piso  
 OnkoZert GmbH: Sebastian Dieng

#### Studienregistrierung

Die Studie wurde durch die Ethikkommission der Ärztekammer Berlin zustimmend beraten (Eth19-18) (38) und prospektiv registriert (DRKS-ID: DRKS00006724). Alle Patientinnen und Patienten gaben eine schriftliche Einwilligung für die Studienteilnahme.

#### Data-Sharing-Statement

Es werden keine anonymisierten Patientendaten (Studienteilnehmerdaten) zur Verfügung gestellt. Zusätzlich, auf Anfrage bereitgestellte Daten (Auswertungen) sind: beobachtete und adjustierte Scores pro Zentrum, Zusatzanalysen für imputierte Modelle und stratifizierte Modelle nach Lokalisation. Weiterhin zur Verfügung gestellt werden: Studienprotokoll, Ergebnisbericht an den Forderer nach erfolgter Freigabe, Patienteninformation und Einwilligung, Fragebogen, beobachtete/adjustierte Werte, abrufbar auf der Webseite [www.edium-studie.de](http://www.edium-studie.de) oder über den korrespondierenden Autor. Die Daten sind ab sofort beziehungsweise ab Freigabe durch den Forderer für mindestens zwei Jahre verfügbar. Die Daten werden jedem mit begründetem Interesse zur Verfügung gestellt.

#### Förderung

Die EDIUM-Studie wurde gefördert durch den Innovationsausschuss beim Gemeinsamen Bundesausschuss.

#### Interessenkonflikt

Sebastian Dieng war im Rahmen der EDIUM-Studie für die OnkoZert GmbH tätig, die als Konsortialpartner für diese Studie Fördermittel des Innovationsfonds erhalten hat. Die OnkoZert GmbH betreut das Zertifizierungssystem der Deutschen Krebsgesellschaft (DKG) und führt in deren Auftrag Zertifizierungen von Onkologischen und Organkrebszentren (darunter Darmkrebszentren) durch.

PD Dr. Kohlhaw wurde für Gutachterleistung von der DKG und der OnkoZert GmbH honoriert.

PD Dr. Siegel wurde im Rahmen des Projektes „S3-Leitlinie Analkarzinom“ im Leitlinienprogramm Onkologie der AWMF/DiKG/Krebshilfe gefördert.

PD Dr. Christoph Kowalski, Nora Tabea Sibert, Clara Breidenbach und PD Dr. Simone Wesselmann sind Mitarbeiter der DKG.

Prof. Dr. Post ist Mitglied der Zertifizierungskommission Viszeralonkologie der DKG sowie des Lenkungs Ausschusses Leitlinie für die DKG.

Prof. Dr. Reissfelder ist Vorsitzender der Zertifizierungskommission Darmkrebszentren.

Die übrigen Autorinnen und Autoren erklären, dass kein Interessenkonflikt besteht.

**Manuskriptdaten**

eingereicht: 29.09.2021, revidierte Fassung angenommen: 15.09.2022

**Literatur**

1. Robert Koch-Institut: Krebs in Deutschland für 2015/2016. Berlin 2019.
2. Onkologisches Leitlinienprogramm: S3-Leitlinie Kolorektales Karzinom. AWMF-Registernummer: 021.007OL. Berlin 2019.
3. Basch E, Torda P, Adams K: Standards for patient-reported outcome-based performance measures. *JAMA* 2013, 310: 139–40.
4. Klinkhammer-Schalke M, Stenger B, Koller M, et al.: Diagnosing deficits in quality of life and providing tailored therapeutic options: results of a randomised trial in 220 patients with colorectal cancer. *Eur J Cancer* 2020, 130: 102–13.
5. OECD: Recommendations to OECD ministers of health from the high level reflection group on the future of health statistics. Strengthening the international comparison of health system performance through patient-reported indicators. *Ohne Ort* 2017.
6. Kowalski C, Hübner J.: „Patient-reported outcome measures“: Reif für die Routine? *Forum* 2020, 35: 401–5.
7. Aaronson NK, Ahmedzai S, Bergman B, et al.: The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993, 85: 365–76.
8. Basch E, Deal AM, Kris MG, et al.: Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. *J Clin Oncol* 2016, 34: 557–65.
9. Basch E, Deal AM, Dueck AC, et al.: Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. *JAMA* 2017, 318: 197–8.
10. Denis F, Lethrosne C, Pourel N, et al.: Randomized trial comparing a web-mediated follow-up with routine surveillance in lung cancer patients. *J Natl Cancer Inst* 2017, 109: dx029.
11. Klinkhammer-Schalke M, Koller M, Stenger B, et al.: Direct improvement of quality of life using a tailored quality of life diagnosis and therapy pathway: randomised trial in 200 women with breast cancer. *Br J Cancer* 2012, 106: 826–38.
12. Wintner LM, Sztrankay M, Gesinger JM, et al.: EORTC Quality of Life Group manual for the use of EORTC measures in daily clinical practice. *Brüssel* 2016.
13. Gerædts M, Drosler SE, Dobler K, et al.: DNVF-Memorandum III „Methoden für die Versorgungsforschung“, Teil 3: Methoden der Qualitäts- und Patientensicherheitsforschung. *Gesundheitswesen* 2017, 79: e65–124.
14. Di Maio M, Basch E, Denis F, et al.: The role of patient-reported outcome measures in the continuum of cancer clinical care. *ESMO Clinical Practice Guideline. Ann Oncol* 2022, 33: 878–892.
15. Gordon BE, Basak R, Carpenter WR, Usinger D, Godley PA, Chen RC: Factors influencing prostate cancer treatment decisions for African American and white men. *Cancer* 2019, 126: 1693–700.
16. IQTiG—Institut für Qualitätssicherung und Transparenz im Gesundheitswesen: *Methodische Grundlagen V1.1*. Berlin 2019.
17. Deutscher D, Werneke MW, Hayes D, et al.: Impact of risk adjustment on provider ranking for patients with low back pain receiving physical therapy. *J Orthop Sports Phys Ther* 2018, 48: 637–48.
18. National Prostate Cancer Audit: Results of the NPCA Prospective Audit in England and Wales for men diagnosed from 1 April 2018 to 31 March 2019. *The Royal College of Surgeons of England* 2021.
19. Farn E, Jackel WH, Schalaster V: Das Qualitätssicherungsverfahren der GKV in der medizinischen Rehabilitation: Ergebnisse und Weiterentwicklung. *Gesundheitswesen* 2009, 71: 163–74.
20. Kowalski C, Roth R, Carl G, et al.: A multicenter paper-based and web-based system for collecting patient-reported outcome measures in patients undergoing local treatment for prostate cancer: first experiences. *J Patient Rep Outcomes* 2020, 4: 56.
21. Whistance RN, Conroy T, Chie W, et al.: Clinical and psychometric validation of the EORTC QLQ-CR29 questionnaire module to assess health-related quality of life in patients with colorectal cancer. *Eur J Cancer* 2009, 45: 3017–26.
22. Deutsche Krebsgesellschaft: *Jahresbericht der zertifizierten Darmkrebszentren*. Berlin 2020.
23. Sibert NT, Breidenbach C, Wesselmann S, et al.: Which EORTC QLQ-C30 and -CR29 scores are relevant for clinicians for therapy planning and decisions? Results of an online survey. *Coloproctology* 2021, 43: 411–6.
24. Gesinger JM, Loth FLC, Aaronson NK, et al.: Thresholds for clinical importance were established to improve interpretation of the EORTC QLQ-C30 in clinical practice and research. *J Clin Epidemiol* 2020, 118: 1–8.
25. NHS England Analytical Team: *Patient Reported Outcome Measures (PROMs). An alternative aggregation methodology for case-mix adjustment*. London 2013.
26. Eton DT, Cella D, Yost KJ, et al.: A combination of distribution—and anchor-based approaches determined minimally important differences (MIDs) for four endpoints in a breast cancer scale. *J Clin Epidemiol* 2004, 57: 898–910.

27. Osoba D, Rodrigues G, Myles J, Zee B, Pater J: Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol* 1998, 16: 139–44.
28. Sibert NT, Pfaff H, Breidenbach C, et al.: Variation across operating sites in urinary and sexual outcomes after radical prostatectomy in localized and locally advanced prostate cancer. *World J Urol* 2022, 40: 1437–46.
29. Finke I, Behrens G, Maier W, et al.: Small area analysis on socioeconomic inequalities in cancer survival for 25 cancer sites in Germany. *Int J Cancer* 2021, 149: 561–72.
30. Couwenberg AM, Burbach JPM, van Grevenstein WMJ, et al.: Effect of neoadjuvant therapy and rectal surgery on health-related quality of life in patients with rectal cancer during the first 2 years after diagnosis. *Clin Colorectal Cancer* 2018, 17: e499–512.
31. Reudink M, Molenaar CJL, Bonhof CS, Janssen L, Mols F, Slooter GD: Evaluating the longitudinal effect of colorectal surgery on health related quality of life in patients with colorectal cancer. *J Surg Oncol* 2022, 125: 217–26.
32. Waljee JF, Ghafari A, Finks JF, et al.: Variation in patient-reported outcomes across hospitals following surgery. *Med Care* 2015, 53: 960–6.
33. Diers J, Baum P, Matthes H, Germer C-T, Wiegner A: mortality and complication management after surgery for colorectal cancer depending on the DKG minimum amounts for hospital volume. *Eur J Surg Oncol* 2021, 47: 850–7.
34. Trautmann F, Reißfelder C, Pecqueux M, Weitz J, Schmitt J: Evidence-based quality standards improve prognosis in colon cancer care. *Eur J Surg Oncol* 2018, 44: 1324–30.
35. Nimpitsch U, Mansky T: Hospital volume and mortality for 25 types of inpatient treatment in German hospitals: observational study using complete national data from 2009 to 2014. *BMJ Open* 2017, 7: e016184.
36. ICHOM—International Consortium for Health Outcomes Measurement: *Colorectal Cancer Data Collection Reference Guide*. Cambridge 2016.
37. Bradley EH, Curry LA, Ramanadhan S, Rowe L, Nembhard IM, Krumholz HM: *Research in action: using positive deviance to improve quality of health care. Implement Sci* 2009, 4: 25.
38. Breidenbach C, Sibert NT, Wesselmann S, Kowalski C: *Erratum: Die Beratung durch Ethikkommissionen bei einer multizentrischen Beobachtungsstudie in Deutschland—Aufwand und Kosten. Gesundheitswesen* 2021, 83: e50.

**Anschrift für die Verfasser**

PD Dr. Christoph Kowalski  
Deutsche Krebsgesellschaft e. V.  
Kuno-Fischer-Str. 8, 14057 Berlin  
kowalsk@krebsgesellschaft.de

**Zitierweise**

Kowalski C, Sibert NT, Breidenbach C, Hagemeyer A, Roth R, Seufferlein T, Benz S, Post S, Siegel R, Wiegner A, Winkels R, Beck-Messmer S, Fahlke J, Reissfelder C, Fuchs M, Herzog T, Weirauch R, Faber-Mertens J, Rudolph H, Puskás L, Kohlhaw K, Szczerbinska M, Scheuerlein H, Neumann PA, Hollerbach S, Riechmann M, Kolbe EW, Weigert N, König J, Klink C, Mussa S, Horn AK, Staib L, Werner J, Jähne J, Aly M, Mork H, Grutzmann R, Piso P, Dieng S, Wesselmann S: Outcome quality after colorectal cancer resection in Certified Colorectal Cancer Centers—patient-reported and short-term clinical outcomes. *Dtsch Arztebl Int* 2022, 119: 821–8. DOI: 10.3238/arztebl.m2022.0325

► Die englische Version des Artikels ist online abrufbar unter:  
[www.aerzteblatt-international.de](http://www.aerzteblatt-international.de)

**Zusatzmaterial**

eLiteratur, eMethodenteil, eTabellen, eGrafiken:  
[www.aerzteblatt.de/m2022.0325](http://www.aerzteblatt.de/m2022.0325) oder über QR-Code



**Berichtigung**

Im Beitrag „Hautinfektionen durch Panton-Valentine-Leukozidin-bildenden *Staphylococcus aureus*“ in Heft 45 ist in der Zusammenfassung des Artikels im Abschnitt Ergebnis ein Sachverhalt nicht korrekt wiedergegeben worden. Hier heißt es: „Haut- und Weichteilinfektionen mit PVL-SA führen dreimal häufiger zu Rezidiven als PVL-negativer *S. aureus*“. Korrekt ist: „Haut- und Weichteilinfektionen mit PVL-SA führen ungefähr doppelt so häufig zu Rezidiven wie PVL-negativer *S. aureus*.“ MWR

## Zusatzmaterial zu:

## Ergebnisqualität in zertifizierten Darmkrebszentren nach Resektion eines kolorektalen Karzinoms

Patient-Reported und kurzfristig-klinische Outcomes

Christoph Kowalski, Nora Tabea Sibert, Clara Breidenbach, Anna Hagemeyer, Rebecca Roth, Thomas Seufferlein, Stefan Benz, Stefan Post, Robert Siegel, Armin Wiegering, Raphael Winkels, Stefanie Bieck-Messemer, Jörg Fahlke, Christoph Reissfelder, Martin Fuchs, Torsten Herzog, Richard Weihrauch, Julia Faber-Mertens, Hagen Rudolph, László Puskás, Kay Kohlhaw, Malgorzata Szczerbinska, Hubert Scheuerlein, Philipp-Alexander Neumann, Stephan Hollerbach, Maren Riechmann, Ernst W. Kolbe, Norbert Weigert, Jörg Köninger, Christian Klink, Shueb Mussa, Anja-Kathrin Horn, Ludger Staib, Jens Werner, Joachim Jähne, Mohamed Aly, Hubert Mörk, Robert Grützmann, Pompilio Piso, Sebastian Dieng, Simone Wesselmann

Dtsch Arztebl Int 2022; 119: 821–8. DOI: 10.3238/arztebl.m2022.0325

### eLiteratur

- e1. NHS England Analytical Team. Patient Reported Outcome Measures (PROMs). An alternative aggregation methodology for case-mix adjustment. 2013.
- e2. Buuren S van, Buuren S. Flexible imputation of missing data, 2. edition. Boca Raton, 2018.
- e3. Nuttall D, Parkin D, Devlin N. Inter-provider comparison of patient-reported outcomes: developing an adjustment to account for differences in patient case mix. *Health Econ* 2015, 24: 41–54.
- e4. Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: data mining, inference, and prediction: with 200 full-color illustrations. New York 2001.
- e5. Eton DT, Cella D, Yost KJ, et al. A combination of distribution- and anchor-based approaches determined minimally important differences (MIDs) for four endpoints in a breast cancer scale. *J Clin Epidemiol* 2004, 57: 898–910.
- e6. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol* 1998, 16: 139–44.
- e7. Sibert NT, Pfaff H, Breidenbach C, et al. Variation across operating sites in urinary and sexual outcomes after radical prostatectomy in localized and locally advanced prostate cancer. *World J Urol*, DOI: 10.1007/s00345-022-03665-6
- e8. National Center for Emerging and Zoonotic Infectious Diseases, Division of Healthcare Quality Promotion. The NHSN standardized infection ratio (SIR)T—a Guide to the SIR. 2022.

## eMETHODENTEIL

### Zentrumsstichprobe

Aus allen 273 deutschen Darmkrebszentren, die im März 2018 ein gültiges Zertifikat hielten und spätestens 2016 erstmals zertifiziert worden waren, wurde eine stratifizierte Zufallsstichprobe mit Nachrückverfahren gezogen: Die 25 fallzahlstärksten und eine Zufallsauswahl der übrigen Zentren wurden zur Teilnahme an der Studie eingeladen. Die Studie wurde in 106 Zentren initiiert. Alle Primärfälle der teilnehmenden Zentren mit kolorektalem Karzinom, die im Zentrum 2019 (in sechs Pilotzentren ab Oktober 2018) eine elektive Tumorektomie erhielten oder nichtoperativ palliativ behandelt wurden, sollten von den Zentren zur Studienteilnahme aufgeklärt, eingeladen und eingeschlossen werden.

### Variablen und statistische Analysen

Als Kovariaten für die Zusammenhangsanalysen und die Casemix-Adjustierung wurden auf Ebene der Patientinnen und Patienten jeweils der prätherapeutische Symptom- oder Funktionsscore und die prätherapeutische allgemeine Lebensqualität betrachtet sowie Alter, Geschlecht, Stadium, American Society of Anesthesiologists (ASA)-Score, Lokalisation (Kolon/Rektum), Staatsangehörigkeit, Bildungsabschluss, Versicherungsstatus (Ausprägungen siehe *Tabelle 1*) sowie vorhandenes Stoma nach 12 Monaten. In zusätzlichen Analysen wurden in Mehrebenenmodellen die Zentrumsmerkmale Fallzahl (stetig), Lehrstatus (Uniklinik, akademisches Lehrkrankenhaus, kein Lehrkrankenhaus), Urbanität (Klein-, Mittel-, Groß-, Millionenstadt) und die Trägerschaft (freigemeinnützig, öffentlich-rechtlich, privat) des Standorts berücksichtigt (*eTabelle 4*, *eTabelle 9*, *eTabelle 10*). Zusätzlich wurden die Durchschnitte der adjustierten Scores kreuztabellarisch nach Zentrumsmerkmalen beobachtet (*eTabelle 5–8*).

Aufgrund der Bedeutung insbesondere für das Rektumkarzinom werden zusätzlich Veränderungen im Score „fäkale Inkontinenz“ vor und 12 Monate nach Resektion betrachtet. Abweichend vom Vorgehen für die fünf anderen Scores erfolgt dies aufgrund von Besonderheiten des Scores, stratifiziert nach Lokalisation (Kolon/Rektum) und Stoma (ja/nein), nach 12 Monaten nur kreuztabellarisch. Anders als bei den übrigen berichteten Scores unterscheidet sich das Item zur fäkalen Inkontinenz zwischen Patienten mit und ohne Stoma, weshalb die Auswertung getrennt erfolgt. Zu T0 wurde das Item nur von Personen ohne Stoma beantwortet, war dort also für alle Befragten gleich. Das Item für Personen ohne Stoma lautete: „Hatten Sie ungewollte Stuhlabgänge aus Ihrem Darmausgang?“, für Stomatragende lautete es: „Hatten Sie ungewollte Stuhlabgänge aus Ihrem Stomabeutel?“. Beide Fragen konnten jeweils mit den Kategorien „überhaupt nicht“, „wenig“, „mäßig“, „sehr“ beantwortet werden. Vergleiche der

Baseline- und 12-Monate-Werte des Scores „fäkale Inkontinenz“ sind bei Stomatragenden unter dieser Einschränkung zu betrachten.

Die Zusammenhangsanalysen erfolgten in Anlehnung an das Verfahren des National Healthcare Safety Network des Center for Disease Control (NHS) zur Casemix-Adjustierung (CMA) primär als multivariable lineare Regressionen ohne Berücksichtigung der Zentrumsmerkmale für jeden der fünf Scores (e1). Die Modellannahmen (Normalverteilung der Residuen, Linearität, Multikollinearität, Varianzhomogenität) wurden geprüft. Die primären Analysen erfolgten aufgrund vollständiger Fälle. Für den ASA-Score wurden aufgrund der Vielzahl der fehlenden Werte Effekte für fehlende Werte geschätzt. ASA ist zwar ein Feld des Moduls „kolorektales Karzinom“ des onkologischen Basisdatensatzes für die Krebsregistrierung, wird aber für die klinischen Krebsregister noch lückenhafter dokumentiert als in der EDIUM-Studie (Auskunft Arbeitsgemeinschaft Deutscher Tumorzentren). Da es sich um eine explorative Studie handelt, wurde nicht für multiples Testen korrigiert.

Die Darstellung erfolgte mithilfe von Forest Plots, die dazugehörigen Tabellen sind *eTabellen 1zu* entnehmen. Als Sensitivitätsanalysen wurden Tobit-Regressionen (*eTabelle 3*) sowie Auswertungen getrennt nach Lokalisation (Kolon/Rektum) durchgeführt (*eTabelle 11*, stratifizierte Regressionsanalysen auf Anfrage). In weiteren Sensitivitätsanalysen (auf Anfrage) wurden fehlende Werte der Baseline-PROs, der Staatsangehörigkeit, des Versicherungsstatus, des Stomastatus und des Schulabschlusses mittels multipler Imputation ersetzt (Anzahl Imputationen: maximaler Prozentsatz fehlender Werte aller Variablen) (e2). Weil die Anlage und das Vorhandensein eines Stomas stark von der Tumorklassifikation abhängen, wurde zusätzlich ein Interaktionseffekt für diesen Zusammenhang geschätzt (Tumorklassifikation  $\times$  Stoma). In die Modelle floss hierzu das durch Befragung erhobene Merkmal „Stoma“ nach 12 Monaten ein. Zu diesem Zeitpunkt ist damit zu rechnen, dass das Stoma bis auf wenige Ausnahmen permanent angelegt ist. Es ist auch möglich, dass das Stoma wegen einer anderen Indikation angelegt ist. In der klinischen Dokumentation ist die (zwischenzeitliche) Stomaanlage dokumentiert. Berücksichtigt wurde der Stomastatus nach 12 Monaten, weil dies maßgeblich für die PROs zu diesem Zeitpunkt sein sollte. Weitere Sensitivitätsanalysen wurden erstens zur Abbildung der Mehrebenenstruktur der Daten (Patienten in Zentren) mittels linearer Mehrebenenmodelle (unter Berücksichtigung der Zentrumsmerkmale) sowie zweitens zur Berücksichtigung der zensierten PRO-Skalen (mögliche Boden- und Deckeneffekte) mittels Tobit-Regression durchgeführt (e3). Die Ergebnisse der gemischten linearen Modelle werden nur berichtet, wenn die Varianz-Kovarianz-Matrix nicht singulär war

(eTabelle 4). Für die Nullmodelle und die vollständigen gemischten Modelle (adjustiert auf mögliche Einflussfaktoren) wurden Intraklassen-Korrelationskoeffizienten (ICCs) bestimmt. In weiteren Sensitivitätsanalysen wurde das Patientenkollektiv getrennt nach Lokalisation (Kolon/Rektum) untersucht. Zusätzlich wurde exemplarisch ein Modell für „Lebensqualität“ gerechnet, in dem geprüft wurde, ob der Modus (online/Papier) mit den PROs assoziiert ist (eTabelle 12). Die Verteilungen der Ausprägungen der posttherapeutischen Scores werden zusätzlich als „violin plots“ dargestellt (eGrafik 6). Zur Überprüfung der Modellgüte der Zusammenhangsanalysen wurde eine 10-fache Kreuzvalidierung durchgeführt (e4). Patienten eines Zentrums wurden dabei nicht getrennt. Die Ermittlung der Modellgüte erfolgte über  $R^2$  (erklärte Varianz). Da es sich bei EDIUM um eine explorative Studie handelt, werden die hier berichteten p-Werte nur zur Deskription dargestellt: p-Werte  $\leq 0,05$  werden als statistisch auffällig angesehen.

Die CMA erfolgte in Anlehnung an das Vorgehen des NHS (e1). Fehlende Werte bei der Prädiktion wurden mittels k-nearest neighbour (k = 6) ersetzt:

- Berechnung der Differenz vom beobachteten (O) und erwarteten (E; Prädiktion des Regressionsmodells) posttherapeutischen PRO pro Patient (O-E)
- Berechnung der Mittelwerte der Differenzen aus 1. pro Zentrum („Performance-Score“)
- Berechnung der Casemix-adjustierten Zentrumswerte durch Addition der Performance-Scores zum posttherapeutischen PRO-Durchschnittswert ( $\bar{o}$ ) aller Studienpatienten (adjustierter Wert = Performance-Score +  $\bar{o}$ ).

Für die Casemix-Adjustierung wurden pro Score nur Zentren berücksichtigt, die posttherapeutische Fragebogen von mindestens zehn Patienten eingebracht haben. In den Grafiken wurden die Punktschätzer (adjustierte Werte) in Beziehung zu behelfsweisen „minimally important differences“ (MIDs) gemäß (e5, e6) gesetzt (Standardabweichung Individualdaten  $\times 1/3$ ). MIDs sind die kleinsten Veränderungen des Scores, die Patienten als wichtig wahrnehmen. Diese visuelle Orientierung dient der besseren Einordnung der Zentrumsunterschiede und hat sich in ähnlichen Berichten bewährt (e7). Alle Analysen erfolgten mit der Software R Version 4.0.0.

Für die kurzfristigen klinischen sekundären Endpunkte Anastomoseninsuffizienz Kolon beziehungsweise Rektum sowie die 30-Tage-Mortalität wurden abweichend multivariable logistische Regressionen geschätzt. In den Zusammenhangsanalysen zur Untersuchung der sekundären Endpunkte wurden folgende Variablen informiert gewählt: Altersgruppe ( $\leq 49$ , 50–69, 70–89,  $\geq 90$ ), Geschlecht, Lokalisation (Kolon/Rektum), ASA-Score, Versicherungsstatus, Staatsbürgerschaft, Bildung (Realschule und polytechnische Oberschule sowie Fachabitur und Abitur zusammengefasst), Tumorstadium („union internationale contre le cancer“ [UICC] I+II/III/IV), Radiotherapie (ja/nein) und PRO-Score zur physischen Funktion (T0).

Basierend auf dem Vorgehen des National Healthcare Safety Network des Center for Disease Control (NHSN) für kurzfristige klinische Komplikationen (postoperative Infektionen, e8) wurde folgendes Vorgehen für die Casemix-Adjustierung der sekundären Outcomes gewählt (die Outcomes wurden jeweils binär operationalisiert):

- Berechnung von erwarteten Werten pro Patient (erwartete Wahrscheinlichkeit für das Outcome)  $p_{j,i}$  aufgrund der logistischen Regressionsmodelle (Outcomes: „Anastomoseninsuffizienz“ ja/nein beziehungsweise „gestorben 30 Tage nach Operation“ ja/nein; „i“ steht für „Patient in Zentrum j“)
- Berechnung der erwarteten Ereignisse pro Zentrum j durch Aufsummierung der erwarteten Werte der zugehörigen Patienten:

$$erwartet_j = \sum_{i=1}^{n_j} p_{j,i}$$

Dabei steht „ $n_j$ “ für die Anzahl Patienten, die im Zentrum j behandelt wurden.

- Berechnung des adjustierten Wertes durch Multiplikation des Quotienten aus tatsächlich beobachteten Ereignissen und den erwarteten Ereignissen für jedes Zentrum mit der Rate an beobachteten Ereignissen in allen Zentren:

$$adjustiert_j = beobachtet_{all} * \frac{beobachtet_j}{erwartet_j}$$

Mit „beobachtet<sub>all</sub>“ ist die Anzahl an beobachteten Ereignissen über alle Zentren gemeint. Falls der Wert für „erwartet<sub>j</sub>“ kleiner 1 ist, kann „adjustiert<sub>j</sub>“ nicht berechnet werden. Die Ergebnisse für Anastomoseninsuffizienz und die 30-Tage-Mortalität werden analog zu den primären Endpunkten stratifiziert nach Tumorklassifikation berichtet.

#### Hinweise zu ASA und Modellgüte

ASA ist ein Feld des Moduls „kolorektales Karzinom“ des onkologischen Basisdatensatzes und wird in klinischen Krebsregistern insgesamt noch deutlich schlechter dokumentiert als in der EDIUM-Studie (persönliche Kommunikation mit der Arbeitsgemeinschaft Deutscher Tumorzentren). Hier besteht Nachbesserungsbedarf. Während die linearen Regressionsmodelle zu den PROs eine meist zufriedenstellende Vorhersagekraft zeigten, trifft dies auf die Modelle zu den kurzfristigen klinischen (sekundären) Endpunkten nicht zu. Dies könnte zumindest teilweise auf das Fehlen von mit diesen Endpunkten assoziierten Parametern zurückgeführt werden. Es ist zu erwarten, dass eine adäquate Casemix-Adjustierung für diese Endpunkte weitere Adjustorvariablen erfordert, die zusätzlich dokumentiert werden müssten, darunter Nebenerkrankungen, Body-Mass-Index und Raucherstatus. Diese zusätzliche Erfassung bedeutet zusätzlichen Dokumentationsaufwand.

#### **4.5. Clinicians' knowledge and attitudes towards patient reported outcomes in colorectal cancer care – insights from qualitative interviews**

Veröffentlicht als: **Sibert, N. T.**, Kowalski, C., Pfaff, H., Wesselmann, S., & Breidenbach, C. (2021). Clinicians' knowledge and attitudes towards patient reported outcomes in colorectal cancer care—Insights from qualitative interviews. *BMC Health Services Research*, 21(1), 366. <https://doi.org/10.1186/s12913-021-06361-z>

RESEARCH ARTICLE

Open Access

# Clinicians' knowledge and attitudes towards patient reported outcomes in colorectal cancer care – insights from qualitative interviews



Nora Tabea Sibert<sup>1\*</sup>, Christoph Kowalski<sup>1</sup>, Holger Pfaff<sup>2</sup>, Simone Wesselmann<sup>1</sup> and Clara Breidenbach<sup>1</sup>

## Abstract

**Introduction:** Patient-reported outcomes (PROs) can be used in cancer care to monitor patients' disease-related symptoms and functional status. However, successful implementation of such instruments is only possible if clinical staff are convinced of the clinical benefits. It is therefore crucial to investigate the attitudes of clinical staff to PROs in routine cancer care.

**Methods:** Semi-structured, guideline-based interviews were held with 12 clinicians working in certified colorectal cancer centers in Germany who are taking part in an observational study on PROs (five surgeons, two oncologists, one psycho-oncologist, two oncological care nurses, one stoma therapist, and one physician assistant) in order to investigate firstly, how clinicians describe PRO instruments ("wording"); and secondly, the clinicians' general attitude toward PROs. A qualitative content analysis according to Kuckartz was performed.

**Results:** The wording used to describe PROs was not consistent. Statements on attitudes toward PROs were very heterogeneous and were therefore categorized into "(rather) positive" and "(rather) negative." The principal advantages of PROs mentioned by participants included broader, structured knowledge about patients and treatment, as well as relevance for patients. Subcategories for (rather) negative attitudes included statements expressing doubts about the questionnaires and "no need for PROs."

**Discussion:** The clinicians participating mainly expressed fairly positive attitudes toward PROs. However, they had little knowledge about PROs in general and the interviews therefore mainly reflect their expectations and assumptions about them. These initial impressions may be regarded as providing a basis for future implementation strategies and for training of clinicians on how to use PROs in routine cancer care.

**Keywords:** Patient-reported outcomes, Patient-reported outcome measures, Implementation, Routine care, EORTC, Cancer care

\* Correspondence: [sibert@krebsgesellschaft.de](mailto:sibert@krebsgesellschaft.de)

<sup>1</sup>German Cancer Society, Kuno-Fischer-Straße 8, 14057 Berlin, Germany

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Introduction

Cancer care is a prolonged and multidisciplinary task. Cancer patients often need not only definitive treatment, such as surgery, but also a more complex treatment strategy involving surgery, chemotherapy, and radiotherapy, or a combination of these. The various treatment pathways have different, but often severe side effects. As a result, health-related quality of life and functional status are often severely affected by oncological disease and may vary substantially with the treatment regimen chosen, the treating clinicians, and over time. Monitoring health-related quality of life and functional status regularly is therefore an important part of cancer care [1].

Traditionally, this monitoring involves the clinical staff taking the patient's history and carrying out medical assessments such as blood tests or medical imaging. Nevertheless, there is evidence that clinicians underestimate the symptom burden and adverse effects [2]. Standardized questionnaires filled out by the patients themselves are therefore instruments that can help clinical staff to identify disease and treatment-specific problems effectively. These instruments are called patient-reported outcomes (PROs), and many clinical trials have already shown that they have benefits for monitoring purposes [3]. To allow close monitoring of each case, the patients complete PRO questionnaires, and their answers and results are passed on to the clinicians treating them (and sometimes to the patients themselves). Using these results, clinicians can assess and reevaluate treatment choices where necessary and assess signs of disease progression. Several clinical trials that have focused on regular monitoring of patients' functional status using PROs have reported improvements in overall survival [4] [5]. In Germany, for example, Klinkhammer-Schalke et al. could show in a randomized clinical trial a systematic monitoring of PROs and – based on a so called “quality of life (QoL) profile” – initiating personalized treatment options can enhance routine colorectal cancer care. In this setting, PROs were collected using the EORTC QLQ-C30 and -CR29 questionnaires and summarizing specific scores to a set of 13 QoL domains based on an expert consensus. If a patient scored lower than a pre-defined threshold, their physicians were advised to react therapeutically (e. g. by offering a specialized pain treatment or psycho-oncological assistance) [6]. Implementing PROs in routine cancer care can therefore help clinical staff identify disease-related problems in cancer patients, and the patients can consequently receive more personalized therapy.

However, PROs have not yet become fully established in routine cancer care, and most clinical staff do not yet have any initial experience with them. In order to successfully implement PROs in routine clinical work,

clinicians — oncologists, surgeons, nurses, and psychotherapists — need to be involved as early as possible. The “knowledge-to-action” (KTA) framework described by Graham et al. [7] has previously been used for PRO implementation [8]. Key parts of the KTA framework are the knowledge creation – with its main aim to “tailor knowledge” – and the iterative action cycle. This action cycle reflects the processes needed to cause change in the behavior and attitudes of persons concerned. Key parts of the KTA action cycle are “identification of the problem and review and selection of knowledge”, “adaption of knowledge to local context”, “assessment of barriers to knowledge use”, “selection, tailoring and implementation of interventions”, “monitoring knowledge use”, “evaluation of outcomes” and “maintenance of knowledge use” [7]. Thus, the KTA framework can be used to identify crucial steps during the implementation process and has already been used in more than 140 implementation studies [9]. To identify appropriate implementation strategies, it is crucial to know whether and in what ways clinicians regard PROs as a useful tool for routine cancer care [10]. Eccles et al. have reported that Ajzen's theory of planned behavior [11] can be successfully used to analyze changes in clinicians' attitudes [12]. In addition to subjective standards and perceived behavioral control, an individual's attitude toward a specific behavior is a major predictor for changes in behavior, according to Ajzen. The theory can also be adapted for research questions regarding implementation strategies, and information about the attitudes of the individuals affected by a specific behavior should be collected. In this context, “attitude toward a behavior” is taken to mean “the degree to which a person has a favorable or unfavorable evaluation or appraisal of the behavior in question” [13]. Ibidem, “perceived behavioral control” is explained as “people's perception of the ease or difficulty of performing the behavior of interest” and thus can be analysed by exploring perceived facilitators and barriers for the use of PROs. The aims of the study were to 1) collect information about clinicians' general attitudes toward and knowledge about PROs in routine cancer care, and 2) identify potential factors facilitating or inhibiting the use of PROs in practice. This analysis addresses the first aim. Results from the second aim will be published in a subsequent paper.

## Material and methods

### Study design

For the “Ergebnisqualität bei Darmkrebs: Identifikation von Unterschieden und Maßnahmen zur flächendeckenden Qualitätsentwicklung”, (EDIUM) study, clinicians' from colorectal cancer centers (CCCs) certified by the German Cancer Society used the PRO questionnaires “European Organization for Research and Treatment of

Cancer" (EORTC) QLQ-C30 [14] and EORTC QLQ-CR29 [15] to inquire about the functional status of colorectal cancer patients before the start of treatment and 1 year afterward. In Germany, approximately half of all newly diagnosed colorectal cancer patients are treated in a CCC certified by the German Cancer Society [16]. Thus, cancer care in CCCs is an important part of routine cancer care in Germany. The aim of the EDIUM study was to assess and compare the quality of care between certified CCCs in Germany using PROs. It is possible, but not compulsory, for the participating CCCs to use their patients' baseline PROs for clinical decision-making. The study coordinators did not mandate which EORTC domains should be used nor thresholds for PROs if they were embedded in clinical decision-making because there is a lack of evidence regarding which EORTC domains are superior to others in colorectal cancer care [17]. However, the study coordinators provided a clinical guide on how to interpret the PRO results using the EORTC QLQ-C30 reference values manual [18]. Some CCCs therefore include PROs in decision-making processes, whereas others only use the PROs for comparison/benchmarking purposes. Nine months after the start of the study, clinicians from the participating CCCs were asked to take part in qualitative semi-structured interviews regarding their general attitudes to and initial experience with the PROs used in the study. The study was approved by the ethics committee of the Medical Association of Berlin (*Ärztelkammer Berlin*) as part of the EDIUM study (eth-19/18) and was funded by the Innovation Fund of the Federal Joint Committee (G-BA).<sup>1</sup>

#### Recruitment of participants

Interview participants were recruited from the CCCs participating in the EDIUM study using a mix of convenience and snowball sampling approaches: The study administrators in the participating CCCs were informed about the interviews and could nominate clinicians who were willing to answer questions about their use of PROs. Eligible interview participants had to be involved in in-patient colorectal cancer care in one of the centers — such as oncologists, surgeons, nurses, or psychotherapists. However, they did not have to be actively involved in the EDIUM study; only the center they were working for needed to be enrolled in the study. Nevertheless, study centres were asked during the sample recruitment process if they used PROs collected for EDIUM for

clinical decision making. Besides experience with PROs, sample saturation criteria also included profession and professional experience, participants' gender and age, as well as the academic status of the CCCs they were working for. All of the participants provided written informed consent.

#### Data collection

The interviews were conducted by C.B. or N.T.S., both female and aged between 25 and 27, and took place in the CCCs the clinicians were working for in November and December 2019. All of the interviews followed the same guideline (see additional file 1) and consisted of two parts. In each interview, the participants were provided with copies of the PRO questionnaire used for the EDIUM study. To begin with, the participants were asked about their profession and position in cancer care in their center, as well as their involvement in EDIUM. Questions about their general knowledge about PROs (including clarification of wording) followed. If the participants had not previously heard the term "PRO," the basic concept was briefly described to them and examples were given. The participants were then asked about their attitude toward PROs in general and specifically those used for the EDIUM study. Factors facilitating or inhibiting the use of PROs in cancer care were inquired about as well, but these are reported elsewhere. In addition, if participants were in favor of the clinical use of PROs, they were asked about which part of cancer care and at which time point they would include PROs in routine cancer care.

In the second part of the interview, different options for presenting PRO scores were discussed. For this purpose, five different approaches were presented to each participant. The results of this second part are reported elsewhere. The interviews took place in the centers where the participants were working and were conducted in German. All of the interviews were audio-recorded and transcribed in an anonymized format. The guideline and the quotations from the participants' responses used in this article were translated from German by a professional translation office.

#### Data analysis

The transcribed interviews were analyzed using Kuckartz content analysis [19]. Both interviewers (C.B., N.T.S.) started by coding the first three interviews deductively, using the main categories "wording," "general attitude," "data collection," "clinical use," and "result presentation," and developed additional categories and subcategories inductively. For the research question, N.T.S. then developed a code book using the main categories "wording" and "general attitude," in collaboration with C.B., and a qualitative content analysis of the resulting categories

<sup>1</sup>Further information about the EDIUM study is available at: <http://edium-studie.de> or <https://innovationsfonds.g-ba.de/projekte/versorgungsforschung/edium-ergebnisqualitaet-bei-darmkrebs-identifikation-von-unterschieden-und-massnahmen-zur-flaechendeckenden-qualitaetsentwicklung.157> (both in German; last accessed January 25, 2021).

was performed. In a second step, differences in the perception of PROs among the various specialties and professions represented were investigated (referred to as group comparison) in order to explore to what extent attitudes towards PROs varied across professional affiliation. Transcription and data analysis were carried out with the help of the software programs f4transcript and f4analysis, version 2.5.4.

## Results

### Sample

Twelve clinicians from eight CCCs were interviewed: five surgeons, two oncologists, one psycho-oncologist, two oncological care nurses, one stoma therapist, and one physician assistant. Their ages ranged from 31 to 58 years. Participants were eligible to participate regardless of the duration of their professional experience. For details, see Table 1.

### Qualitative content analysis

The main categories (referred to as “themes”) are described below, along with results from the interviews. The code book, with category definitions and anchor examples, is presented in additional file 2.

#### Theme 1: wording

At the beginning of the interview, each participant was asked what they called the questionnaires used for the EDIUM study or similar instruments. The way in which participants referred to PROs during the course of the interview was also analyzed. The participants mostly did not distinguish between the actual instrument and the

results when talking about PROs. PROs were consequently mainly referred to as “(result of the) questionnaire.” Some of the participants indicated their understanding of the results of PROs by referring to them as “(functional or symptom) scales” (as proposed by the EORTC) or simply “quality of life.” Three participants claimed not to have any specific term for PROs (“No, we don't have a name for it. Maybe we should think about one,” CH03f), and only two participants actually used the term “patient-reported outcomes.”

#### Theme 2: general attitudes toward PROs

In this category, the participants' attitudes toward PROs were analyzed, including their thoughts on general possible uses for PROs. Their statements were classified into “(rather) positive attitude” and “(rather) negative attitude,” following the bipolar conception of attitude of Ajzen's theory of planned behavior [11], and further information explaining the attitude was specified using subcategories.

#### Theme 2a: positive attitudes toward PROs

Statements that mainly highlighted possibilities and advantages of (clinically use of) PROs were classified as “(rather) positive.”

The participants mentioned three main possible uses. Firstly, many statements suggested that PROs could be used in a clinical setting:

- “I can imagine using this [the PRO instruments], or even having it included in treatment planning.” (CH01m)

The participants described different purposes and parts of cancer care for which PROs could be used clinically. To begin with, using PROs as a screening tool was mentioned:

- “If we weren't seeing all of the patients personally, then I could imagine the questionnaire could also be quite useful for filtering out where we should go.” (PO01w)

The participants were also able to imagine using PROs for “treatment planning” and for “treatment monitoring”:

- “But I'd mainly be interested first of all in the questionnaire from beforehand, because from that you can quickly find out a lot of things you need to pay attention to, for treatment planning as well.” (IM01m)
- “So of course you could pass this questionnaire back to the therapist, that would be an option.” (PA0102)

**Table 1** Interview sample: 12 participants from eight colorectal cancer centers

Profession		
Physicians		7
Surgery	5	
Internal medicine	2	
Nurses		4
With oncological specialization	2	
Physician assistant	1	
Stoma therapist	1	
Psychotherapist		1
Gender		
Female		9
Male		3
Age: mean (SD)		46 (9)
Colorectal cancer centers		
In a university hospital		2
Not in a university hospital		6

In addition to clinical uses, “scientific use” of PROs — reflecting their current use in the EDIUM study — was also mentioned:

- “That’s why I think the thing with the questionnaires is very, very good, and of course — and this is where the scientific approach comes in — you can also quantify it.” (CH02f)

Thirdly, some participants referred to the “quality assurance” aspects of PROs:

- “These are different quality data from the purely key figures provided during the certification procedures, so to speak, and together it also produces a much more specific picture of the center.” (CH01m)

Nevertheless, these different usage possibilities of PROs mentioned by the participants were not always clearly distinguishable, and some statements remained ambiguous, highlighting the multipurpose approach of PROs:

- “Because I think you’d want to pass it back. Whether to individuals or just to a larger group, you’d have to see.” (PA0102)

Participants mentioned several major advantages of using PROs. To begin with, many of the aspects mentioned can be summarized using the subcategory “additional information gained”:

- “Personally, I’ve actually already noticed that you do also find out a few things about the patients using questionnaires like this that you wouldn’t otherwise get out, even if you have a long discussion with them.” (IM01m)
- “A whole lot of things may possibly come out in the process that we might not grasp like that as therapists at all.” (PA0102)

The opportunity to “visualize disease progression” through routine and periodic use of PROs was also mentioned:

- “But I just think, if you just follow it up over the course, and supposing these questionnaires become established now, maybe at each follow-up appointment, then I’d have a course like that and I could see whether it’s getting better or worse, and then it would actually help, I think. Then for the next check-up, I’ll somehow know, ‘Ah, last time he was doing worse, now I’ll need to take a more careful look today.’” (CH02f)

The “relevance to patients” of PROs was also discussed by some of the participants:

- “I think it’s very positive, because it’s relevant for the patients. I mean, so if we say we’ve had a curative treatment approach, then the patients will have to or may be able to live with the situation for another 10, 20, 30 years. And that needs to go along with a reasonable quality of life — and particularly with rectal cancer patients, both of us know that isn’t always the case.” (CH02f)

For some of the participants, the opportunity to use PROs in face-to-face discussions with patients, or as a way of screening for communication needs, was important (“support for patient–clinician communications”):

- “And then you can also use it to offer a discussion or to arrange for someone who could then offer a discussion.” (IM01m)

Finally, “standardization/quantification” was also highlighted as an advantage with PROs:

- “Because in the end it’s a matter of standardization, so that things can be made comparable.” (CH01m)

#### **Theme 2b: negative attitudes toward PROs**

Statements that tended to express doubts about PROs were categorized as “(rather) negative.” Inductively, two subcategories of these statements were found: “doubts about questionnaire” and “doubts about the need for PROs.”

The reasons that participants gave for having concerns about the questionnaires were manifold and sometimes contradictory. On the one hand, some participants complained about “unspecific questions” in the EORTC QLQ-C30 and -CR29:

- “Otherwise I think some of the others are also very unspecific and may not give you any concrete help ... These are pretty standardized questionnaires ... so to that extent you need to see what the analysis actually shows, what’s useful and what isn’t.” (CH04f)

By contrast, other participants were concerned about there being too many questions (“questions too specific”):

- “So in that sense for colorectal patients, I think it would actually be quite good to focus a bit on the typical symptoms, such as dry mouth, hair loss is maybe more to do with chemotherapy and not so

much before the operation, that might be something I would tend to leave out. And maybe a few things could be summed up a bit. Pain here, for example, there are several forms here like abdominal pain, anal, rectal, and the incontinence stories could maybe tend to be summed up a bit.” (CH02f)

In addition, there were participants who tended to the view that the “questions are not relevant for in-patient cancer care”:

- “Well, I think that’s a problem of course for an acute-care hospital like this one ... When the patients are back home, of course, and everything has settled down. I think then the whole thing becomes even more important again ... But here [i.e. treatment in the center] it’s still a ‘worst case’ situation for lots of patients. And then of course it’s difficult to discuss things like that with them.” (PF01f)

The algorithm for scoring the questionnaire was also questioned (“unspecific scaling”):

- “What bothers me a bit, I mean, the scales, they’re not very differentiated at all, are they? ... And I think the scaling isn’t very detailed. I mean, it’s actually relatively rough in my view.” (PO01w)

The participants gave a few reasons for doubting the need for PROs in cancer care, some of them alongside the advantages mentioned above. The main reasons mentioned can be summed up as “no additional information” being provided when using PROs:

- “Lots of the things are the same as what you have in routine clinical work anyway. Especially if the patients are in surgery or gastroenterology, then it’s hair loss, abdominal pain, anorectal pain, lots of things that are part of normal data collection, if I can put it like that ... So that’s why I think it’s not really that relevant at the moment, because it’s actually in parallel with it, or maybe there’s some overlapping. So of course you could pass this questionnaire back to the therapist, that would be an option. So it’s already a bit superfluous.” (PA0102)

Some of the participants also emphasized that PROs “cannot replace face-to-face conversation” between clinicians and patients:

- “Well, it’s always — when it’s about very specific, personal areas, I think, then it’s, for many patients it gets difficult, it’s the financial side on the one hand,

where many people probably find it hard to make a statement, that’s probably better done in a personal conversation, I think.” (IM01m)

#### Comparison of professional groups

Overall, the different professional groups (physicians, nurses, psychotherapist) did not express entirely contrasting views. The physicians who were interviewed tended to describe multiple possible uses of PROs more often: they not only mentioned their clinical use, but also scientific benefits and opportunities for quality assurance provided by PROs. By contrast, the nurses who were participating mainly mentioned aspects of practical usage of PROs in routine clinical work.

It is also interesting that some of the participants tended to envisage possible clinical uses of PROs in specialties other than their own clinical work. Several participants, who were all involved in in-patient cancer care in CCCs, suggested that PROs could be useful for out-patient cancer care:

- “Well in the end, of course, the follow-up physician ... er ... would certainly also be a target, who’s still looking after them, because you already get a functional picture, and you should really take a look at that, because patients don’t always tell you everything, I mean just during the course of the follow-up, I think it can already, it can ask about relevant points, and the patient really has to properly write it down.” (CH01m)

It also appeared that specific professions were mentioned as the main intended users of PROs — e.g., psychotherapists or stoma therapists:

- “I mean, stoma therapists can benefit from it. It would be helpful for stoma therapists. It would be helpful for the psycho-oncologists, and in the end it would certainly be helpful for the physician providing further treatment.” (CH04f)

Despite this, neither the psychotherapist nor the stoma therapist who were interviewed expressed any need for PROs for their routine cancer care:

- “And also, we visit each patient personally. In other words, I have a conversation with every patient. In more or less detail, depending on the situation. That’s the one side that I have, I mean the direct patient contact. And on the other, I then also have a look at the EDIUM questionnaire and notice I already know the answers given in the questionnaire, because I had the personal contact and talked to the patients about their condition and everything

troubling them and so on. I do also try to go through it and see if there's anything to do, but at the moment I don't get very much new information from it." (PO01w)

**Summary: overview of clinicians' fears and hopes with PROs**

At the time of the interviews, the participating clinicians showed little or no experience with PROs in routine cancer care. Their statements can therefore be regarded as reflecting their fears and hopes with PROs. Due to the limited experience the clinicians had with PROs, the statements they gave are partly inconsistent and ambiguous for some of the participants. Figure 1 encapsulates these fears and hopes as a synthesis of the reports.

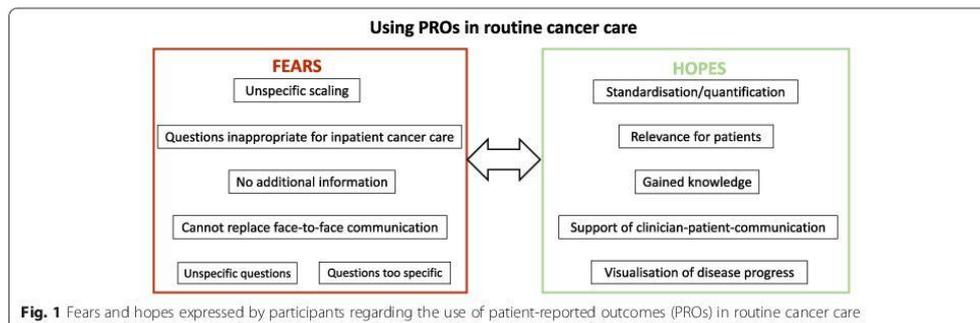
**Discussion**

Most of the participating clinicians were able to identify advantages provided by PROs in clinical use. Most of these advantages were seen in using PROs as a structured approach to quantifying cancer-related symptoms and functions, and hence to obtain more information about cancer patients and their treatment. The participants envisaged several use cases for PROs — not only as a clinical instrument, but also as an important scientific indicator of outcomes and for quality assurance. The multiple use cases may be very helpful in future efforts to successfully implement PROs in clinical contexts, as they highlight several strengths of PROs. If PROs are successfully implemented, they may make it possible to avoid multiple documentation of outcomes for different purposes in CCCs (e.g., certification, clinical documentation, observational or clinical studies). In combination with the quantifying aspects of PROs, this structuring and information bundling may help CCCs to work more effectively.

However, the clinicians participating in the present study mentioned reasonable doubts concerning PROs, including criticism of the questionnaire and scepticism

regarding the need for PROs in clinical use. These doubts highlight the importance of involving clinicians as early as possible in processes of implementation when planning to use PROs in CCCs. To begin with, clinicians can contribute their specialist knowledge to the development of appropriate instruments. For the EORTC, for instance, this was an important part of the questionnaire development process from the beginning: for each questionnaire, different professional specialists are involved in developing relevant disease-related quality of life questionnaires [20]. However, clinicians' involvement should not end with questionnaire development. If PRO scores tend to be too unspecific for some clinicians or departments, some authors have suggested using the item responses or including additional questionnaires for some indications [21, 22]. On the other hand, there are several implementation strategies that restrict the PRO scores to only the most relevant ones for a particular case [23].

Another important reason for the clinicians' doubts might be their uncertainty of possible clinical consequences of PRO results - like therapy modification or even precise therapy options. Thus, clinicians remain uncertain about the practical usefulness of PROs – although there is already evidence indicating that PROs can serve as predictors of survival of especially colorectal cancer patients [24, 25]. Our results highlight that knowledge of how important PROs can be for colorectal cancer care has yet not reached clinicians' treating colorectal cancer patients. In the late 1980s, Davis developed the "Technology Acceptance Model" (TAM) to describe how and if users accept and use technology [26]. Since then, TAM has widely been used in health care sciences, as well [27]. The "perceived usefulness" is an important factor for technology acceptance according to TAM and is defined as "perception that using system leads to enhanced personal performance" [27]. For implementing PROs, hence, concrete clinical implications or therapeutic options conditional upon particular



**Fig. 1** Fears and hopes expressed by participants regarding the use of patient-reported outcomes (PROs) in routine cancer care

PRO results could help to increase clinicians' opinion about the usefulness of PROs. For example, Klinkhammer-Schalke et al. measured PROs in colorectal cancer patients and proposed tailored patient pathways, as described in the introduction. For patients with PRO results above or below preassigned thresholds, particular therapy options were proposed to clinicians [6]. However, it remains unclear how these thresholds were determined. Thus, thresholds and clinical implications should be developed together with clinicians as early as possible and thus could enhance clinicians' acceptance of PROs.

In Germany, certified cancer centres are an important part in colorectal cancer care with approximately half of all colorectal cancer patients being treated in a CCC [16]. Regarding PROs, there is little to no experience in routine cancer care in Germany so far and PROs are only used in trials and outside routine care. For reaching as many colorectal cancer patients as possible in Germany, any attempts for implementing PROs in routine colorectal cancer care should involve cancer care in CCCs. From the findings of this analysis some practical implications might be derived: By showing that there is a lack of experience working with PROs in clinical-decision making which might trigger fears and hopes in clinicians, the present study supports the need to develop successful implementation strategies for PROs, taking clinicians' concerns and ideas into account from the very start. For the contextualization of our findings, we used the KTA framework by Graham et al. [7]. It emphasizes the importance of active collaboration among all stakeholders. In a systematic review of factors facilitating and inhibiting the implementation of PROs, Foster et al. also highlight the importance of involving clinicians at an early stage [28]. Following the KTA framework's action cycle as described above, our results on clinicians' attitude towards PROs are an important prerequisite when closing the knowledge-practice gap and should be used when developing concrete PRO implementation strategies for the CCCs (referred to as "adapting the knowledge to the local context" within the KTA framework). For instance, it seems reasonable to organize training events (e.g., webinars, question-and-answer sessions) for professionals as early as possible in order to enhance clinicians' knowledge about PROs [29]. The present findings may suggest possible aspects that should be discussed with clinicians in such training sessions. As Fig. 1 shows, the fears and hopes identified among clinicians are often parallel. Discussing these can thus enable clinicians to prioritize the purposes for which they want to use PROs. Moreover, results from the analysis of facilitators and barriers of PRO use in routine colorectal cancer care in our setting, as described in detail by Breidenbach et al., highlight the

importance of a robust and easily accessible technical PRO infrastructure, as well as some organizational aspects (e. g. clear responsibility/coordination of PROs within the CCC or precise dissociation from other clinical monitoring tools). This "assessment of barriers to using knowledge" is another important phase of the KTA framework's action cycle addressed in this study. With this knowledge about the prerequisites of PRO implementation in German CCCs, stakeholders like clinicians, as well as coordinators of CCCs or even medical societies as supporters, can start to tackle the next action cycle's phase ("selection, tailoring and implementation of interventions") with e. g. educational training as proposed above.

Our findings, however, are not only crucial when applying the KTA framework's action cycle, but also highlight knowledge gaps in PROs research which should be approached with more emphasis ("knowledge creation" within the KTA framework): The clinicians' response show that there is not enough evidence-based knowledge on what to do with PRO results, e. g. there are only few to none reference manuals with thresholds for PROs [7]. Hence, our findings emphasize the need for the development of practical guidelines for clinicians for the routine use of PROs. For example, the principal investigators of the EDIUM study are planning to develop minimal important differences for both the EORTC QLQ-C30 and -CR29 using the results of EDIUM.

The authors are aware of some limitations resulting from the recruitment methods used: As clinicians from study centers nominated themselves if they were interested in participating, the presented study is biased by self-selection. However, the inconsistency and ambiguity of their statements highlight the variety of clinicians' attitudes the study could carve out. The present study clearly has a small sample size, with only 12 participants. Although full content saturation cannot be claimed, the inclusion of different professions (physicians, nurses, psychologist) allows a more diverse view of the topic and hence conclusions that involve not just a single profession, but the whole CCC as a treatment team. This multidisciplinary approach may be regarded as a strength of the study. The comparison of different professional groups presented should not lead to any hasty conclusions being drawn and is only intended to show that more precise research questions need to be developed relative to the different views of PROs among the professions. For EDIUM, the findings form the basis of a quantitative questionnaire, the results of which will be reported elsewhere.

Although all the CCCs participating in this study are involved in the EDIUM study, not all of the interview partners had any initial experience with PROs yet. Some of them were therefore seeing a PRO questionnaire and

its results for the first time during the interview. Some statements may thus be first impressions rather than well-founded opinions. Nevertheless, as implementation processes for PROs usually start in a setting in which there is generally little knowledge about PROs, these statements are also informative in revealing potential barriers to and prejudices against PROs.

In conclusion, the interview partners mostly showed a rather positive general attitude toward PROs, and this is an important first step toward using them clinically. Implementation strategies should take into account the advantages of PROs already mentioned by clinicians, without ignoring the reasonable doubts also expressed.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12913-021-06361-z>.

**Additional file 1.**

**Additional file 2.**

### Acknowledgements

We would like to thank all study participants for their time and effort. We are grateful to the Innovation Fund of the Federal Joint Committee (G-BA) for funding this research.

### Authors' contributions

CK and SW designed and supervised the EDIUM study. NTS and CB planned this analysis, collected data and performed data analysis. NTS wrote the first draft of the manuscript. CB, CK, HP and SW discussed the results and commented on the manuscript. The author(s) read and approved the final manuscript.

### Funding

EDIUM was funded by the Innovation Fund of the Federal Joint Committee (Innovationsausschuss beim Gemeinsamen Bundesausschuss (G-BA)).

### Availability of data and materials

According to the patient consent form data is not available for scientific use by others than the project group members (contact: Nora Tabea Sibert, [sibert@krebsgesellschaft.de](mailto:sibert@krebsgesellschaft.de)).

### Declarations

#### Ethics approval and consent to participate

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Ethical approval was obtained by the Ethics Committees of the Berlin Chamber of Physicians (Eth-19/18). All participants gave written informed consent to participate in the study.

#### Consent for publication

Not applicable.

#### Competing interests

All authors expect for HP are employees of the German Cancer Society.

#### Author details

<sup>1</sup>German Cancer Society, Kuno-Fischer-Straße 8, 14057 Berlin, Germany.

<sup>2</sup>University of Cologne, Faculty of Human Sciences, Faculty of Medicine and University Hospital Cologne, Institute of Medical Sociology, Health Services Research and Rehabilitation Science, Eupener Str. 129, 50933 Köln, Germany.

Received: 3 December 2020 Accepted: 5 April 2021

Published online: 20 April 2021

### References

- Fromme EK, Eilers KM, Mori M, Hsieh YC, Beer TM. How accurate is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the quality-of-life questionnaire C30. *J Clin Oncol*. 2004;22(17):3485–90. <https://doi.org/10.1200/JCO.2004.03.025>.
- Laugsand EA, Sprangers MA, Bjordal K, Skorpén F, Kaasa S, Klepstad P. Health care providers underestimate symptom intensities of cancer patients: a multicenter European study. *Health Qual Life Outcomes*. 2010;8(1):104. <https://doi.org/10.1186/1477-7525-8-104>.
- Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. *J Clin Oncol*. 2015. <https://doi.org/10.1200/JCO.2015.63.0830>.
- Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. *JAMA*. 2017;318(2):197–8. <https://doi.org/10.1001/jama.2017.7156>.
- Denis F, Yossi S, Septans A-L, Charron A, Voog E, Dupuis O, et al. Improving survival in patients treated for a lung cancer using self-evaluated symptoms reported through a web application. *Am J Clin Oncol*. 2017;40(5):464–9. <https://doi.org/10.1097/JCO.000000000000189>.
- Klinkhammer-Schalke M, Steinger B, Koller M, Zeman F, Fürst A, Gump J, et al. Diagnosing deficits in quality of life and providing tailored therapeutic options: results of a randomised trial in 220 patients with colorectal cancer. *Eur J Cancer Oxf Engl*. 1990;130:102–13.
- Graham ID, Logan J, Harrison MB, Straus SE, Tetroe J, Caswell W, et al. Lost in knowledge translation: time for a map? *J Contin Educ Heal Prof*. 2006; 26(1):13–24. <https://doi.org/10.1002/chp.47>.
- Stover AM, Haverman L, van Oers HA, Greenhalgh J, Potter CM, Ahmed S, et al. Using an implementation science approach to implement and evaluate patient-reported outcome measures (PROM) initiatives in routine care settings. *Qual Life Res*. 2020. <https://doi.org/10.1007/s11136-020-02564-9>.
- Field B, Booth A, Illott J, Gerrish K. Using the knowledge to action framework in practice: a citation analysis and systematic review. *Implement Sci*. 2014; 9(1):172. <https://doi.org/10.1186/s13012-014-0172-2>.
- Devlin NJ, Appleby J, Buxton M. King Edward's Hospital Fund for London, Office of Health Economics London E. getting the most out of PROMs: putting health outcomes at the heart of NHS decision-making. London: King's Fund; 2010.
- Ajzen I. The theory of planned behavior. *Organ Behav Hum Decis Process*. 1991;50(2):179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T).
- Eccles MP, Grimshaw JM, Johnston M, Steen N, Pitts NB, Thomas R, et al. Applying psychological theories to evidence-based clinical practice: identifying factors predictive of managing upper respiratory tract infections without antibiotics. *Implement Sci IS*. 2007;2(1):26. <https://doi.org/10.1186/1748-5908-2-26>.
- Ajzen I. The theory of planned behavior. *Organ Behav Hum Decis Process*. 1991;50:188.
- Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European organization for research and treatment of cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst*. 1993;85(5):365–76. <https://doi.org/10.1093/jnci/85.5.365>.
- Whistance RN, Conroy T, Chie W, Costantini A, Sezer O, Koller M, et al. Clinical and psychometric validation of the EORTC QLQ-CR29 questionnaire module to assess health-related quality of life in patients with colorectal cancer. *Eur J Cancer Oxf Engl*. 1990. 2009;45:3017–26.
- German Cancer Society (DKG), Certification Committee Visceral Oncology Centres / Colorectal Cancer Centres, Seufferlein T, Post S, Wesselmann S, Rückher J, et al. Annual Report 2020 of the Certified Colorectal Cancer Centres (CRCCs). Audit year 2019 / Indicator. Berlin: German Cancer Society; 2018. p. 2020.
- van der Hout A, Neijenhuis KJ, Jansen F, van Uden-Kraan CF, Aaronson NK, Groenvold M, et al. Measuring health-related quality of life in colorectal cancer patients: systematic review of measurement properties of the EORTC QLQ-CR29. *Support Care Cancer*. 2019;27(7):2395–412. <https://doi.org/10.1007/s00520-019-04764-7>.

18. Scott NW, Fayers P, Aaronson NK, Bottomley A, de Graeff A, Groenvold M, et al. EORTC QLQ-C30 Reference Values Manual; 2008.
19. Kuckartz U. Qualitative text analysis. Thousand Oaks: SAGE Publications Ltd; 2014.
20. Sprangers MA, Cull A, Bjordal K, Groenvold M, Aaronson NK. The European Organization for Research and Treatment of Cancer. Approach to quality of life assessment: guidelines for developing questionnaire modules. EORTC study group on quality of life. *Qual Life Res Int J Qual Life Asp Treat Care Rehabil.* 1993;2(4):287–95. <https://doi.org/10.1007/BF00434800>.
21. van der Wees PJ, Verkerk EW, Verbiest MEA, Zuidgeest M, Bakker C, Braspenning J, et al. Development of a framework with tools to support the selection and implementation of patient-reported outcome measures. *J Patient-Rep Outcomes.* 2019;3(1):75. <https://doi.org/10.1186/s41687-019-0171-9>.
22. Sibert NT, Dieng S, Oesterle A, Feick G, Carl G, Steiner T, et al. Psychometric validation of the German version of the EPIC-26 questionnaire for patients with localized and locally advanced prostate cancer. *World J Urol.* 2019. <https://doi.org/10.1007/s00345-019-02949-7>.
23. Klinkhammer-Schalke M, Koller M, Steinger B, Ehret C, Ernst B, Wyatt JC, et al. Direct improvement of quality of life using a tailored quality of life diagnosis and therapy pathway: randomised trial in 200 women with breast cancer. *Br J Cancer.* 2012;106(5):826–38. <https://doi.org/10.1038/bjc.2012.4>.
24. Rutherford C, Campbell R, White K, King M. Patient-reported outcomes as predictors of survival in patients with bowel cancer: a systematic review. *Qual Life Res Int J Qual Life Asp Treat Care Rehabil.* 2019;28(11):2871–87. <https://doi.org/10.1007/s11136-019-02255-0>.
25. Hsu T, Speers CH, Kennecke HF, Cheung WY. The utility of abbreviated patient-reported outcomes for predicting survival in early stage colorectal cancer. *Cancer.* 2017;123(10):1839–47. <https://doi.org/10.1002/cncr.30511>.
26. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q.* 1989;13(3):319–40. <https://doi.org/10.2307/249008>.
27. HOLDEN RJ, B-T KARSH. The technology acceptance MODEL: its past and its future in health care. *J Biomed Inform.* 2010;43(1):159–72. <https://doi.org/10.1016/j.jbi.2009.07.002>.
28. Foster A, Croot L, Brazier J, Harris J, O’Cathain A. The facilitators and barriers to implementing patient reported outcome measures in organisations delivering health related services: a systematic review of reviews. *J Patient-Rep Outcomes.* 2018;2(1):46. <https://doi.org/10.1186/s41687-018-0072-3>.
29. Santana MJ, Haverman L, Absolom K, Takeuchi E, Feeny D, Grootenhuis M, et al. Training clinicians in how to use patient-reported outcome measures in routine clinical practice. *Qual Life Res Int J Qual Life Asp Treat Care Rehabil.* 2015;24(7):1707–18. <https://doi.org/10.1007/s11136-014-0903-5>.

### Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)



## **5. Diskussion: *Patient-reported Outcomes* und ihre Nutzungsmöglichkeiten auf verschiedenen Versorgungsebenen in der Onkologie**

Ziel dieser Arbeit war es, die Möglichkeiten der Nutzung von *patient-reported Outcomes* (PROs) für onkologische Patient\*innen in bereits bestehenden Versorgungsstrukturen (zertifizierten Krebszentren) aufzuzeigen und zu erforschen. Dafür wurde zunächst die deutsche Übersetzung eines international bereits etablierten PRO-Instruments (EPIC-26) für das lokalisierte Prostatakarzinom psychometrisch validiert, was eine Grundvoraussetzung für die klinisch relevante und wissenschaftlich fundierte Implementierung von PROs ist (Arbeit 1). Dann wurden mithilfe eines Systematic Reviews statistische Methoden zusammengetragen, die in anderen Untersuchungen zur Anwendung gekommen sind, um PROs für den Casemix-adjustierten Versorgungsorganisations-Vergleich zu nutzen (Arbeit 2). Aufbauend auf den Ergebnissen dieses Reviews konnte dann zum einen das funktionelle Outcome – gemessen durch den zuvor validierten Fragebogen – ein Jahr nach Radikaler Prostatektomie in unterschiedlichen Prostatakrebszentren untersucht werden (Arbeit 3). Hierbei konnte gezeigt werden, dass sich Prostatakrebszentren in ihren funktionellen Outcomes unterscheiden und diese Unterschiede unabhängig vom Patientenkollektiv sind. Auch für die Tumorentität Darmkrebs wurden PROs genutzt, um eventuell bestehende Ergebnisqualitätsunterschiede zwischen den Krebszentren zu untersuchen und zu quantifizieren (Arbeit 4). Dabei zeigten sich keine solch bedeutenden Unterschiede wie für Prostatakrebszentren. Um weitere, über den Qualitätsvergleich hinausgehende Nutzungsmöglichkeiten von PROs zu identifizieren, wurden anschließend qualitativ das Wissen über und die Einstellung zu PROs von onkologisch tätigen Kliniker\*innen untersucht (Arbeit 5). Hierbei konnte erfasst werden, dass Kliniker\*innen durchaus verschiedene potentielle Anwendungsmöglichkeiten von PROs identifizieren (z. B. Verlaufskontrolle, Verbesserung der Ärzt\*innen-Patient\*innen-Kommunikation), jedoch auch – vor allem technisch-organisatorische – Hürden sehen, die bei einer nachhaltigen Implementierung berücksichtigt werden müssen (vgl. auch 5.3)<sup>53</sup>.

Vor diesem Hintergrund lassen sich nun im Folgenden abschließend die in Abschnitt 2.6 aufgeworfenen Fragestellungen adressieren (5.1, 5.2, 5.3), die anschließend zusammengefasst die verschiedenen Nutzungsebenen von PROs und deren Möglichkeiten aufzeigen sollen (5.4).

## 5.1. Forschungsfrage 1: Wie lassen sich PROs als relevante Outcomes für den Vergleich von Versorgungsorganisationen unter der Berücksichtigung des Patient\*innenkollektivs adjustieren?

Eine Hauptherausforderung der Nutzung von Outcomes auf Versorgungsorganisationsebene ist die des angemessenen Vergleichs – dies trifft im Besonderen auch auf PROs zu (vgl. auch Abschnitte 2.2 und 2.3). Daraus ergab sich die erste Forschungsfrage dieser Arbeit. Durch die durchgeführte systematische Literaturrecherche konnte kein einheitliches Adjustierungsvorgehen für PROs identifiziert werden. Stattdessen existieren eine Vielzahl an unterschiedlichen Methodiken, um mithilfe von PROs Versorgungsorganisationen zu vergleichen. Neben erheblichen Differenzen (bspw. in der Skalierung der adjustierten PROs oder der Einbeziehung von Merkmalen der zu vergleichenden Versorgungsorganisationen) hatten die in der Literatur gefundenen Ansätze allerdings gemein, dass die Notwendigkeit einer Baseline-Befragung – also einer Befragung vor Initiierung der Versorgungsleistung – unterstrichen wurde: Alle identifizierten Studien ( $n = 11$ ) benutzen als Patient\*innenmerkmal für die Adjustierung in einer Baseline-Befragung erfasste PROs. Auf Grundlage der Literaturrecherche wurden vor allem die folgenden drei Forderungen für ein methodisch angemessenes und korrektes Adjustierungsverfahren von PROs aufgestellt – die so dann auch für die in den Arbeiten 3 und 4 durchgeführten Versorgungsorganisationsvergleiche angewandt wurden<sup>54</sup>:

1. Nutzung von Baseline-PROs für die Adjustierungsmodelle
2. Nutzung der ursprünglichen Skalierung des verwendeten PRO-Instruments<sup>6</sup>
3. Möglichst keine Einbeziehung von Versorgungsorganisationsmerkmalen in die Adjustierungsmodelle

Für die Casemix-Adjustierungen der PROs von Prostata<sup>57</sup> sowie Darmkrebszentren<sup>58</sup> wurde dann das methodische Vorgehen des NHS für elektive operative Eingriffe gewählt<sup>23</sup>, das sich bereits für den Vergleich einer großen Anzahl von Versorgungsorganisationen bewährt hat und fortlaufend methodisch evaluiert wird. Dabei ist das Vorgehen das folgende (und wird für jede Therapieform stratifiziert durchgeführt):

1. Identifizierung von für die Erkrankung relevanten (und aus der Dokumentation vorhandenen) individuellen Patient\*innenmerkmalen, die als Adjustoren  $a_k$  dienen ( $k$  ist die Anzahl der Adjustoren)

---

<sup>6</sup> Dies bezieht sich darauf, dass einige der im Systematic Review eingeschlossenen Studien für die Adjustierung Quotienten bilden (z. B.  $\frac{\text{erwartetes Ergebnis}}{\text{beobachtetes Ergebnis}}$ ) und diese dann berichten, bspw. <sup>55,56</sup>. Diese Quotienten sind dann nicht mehr in der ursprünglichen Skalierung des verwendeten PRO-Instruments verständlich, was die Akzeptanz bei relevanten Akteur\*innen verringern kann.

2. Aufstellung eines multiplen linearen Regressionsmodells  $PRO = \beta_0 + \beta_k * a_k + \varepsilon$ , das das erfasste PRO durch die Adjustoren erklärt
3. Berechnung eines erwarteten PROs  $e_i$  für jede\*n Patient\*in auf Grundlage ihrer\*seiner spezifischen Patient\*innenmerkmale ( $i$  ist die Gesamtanzahl der Patient\*innen aller am Vergleich teilnehmender Versorgungsorganisationen)
4. Berechnung der  $Performance_j = \frac{1}{n_j} \sum_{i=1}^{n_j} (e_i - o_i)$  für jede Versorgungsorganisation  $j$ , wobei  $n_j$  die Anzahl der Patient\*innen der Versorgungsorganisation  $j$  ist und  $o_i$  das beobachtete PRO der\*s  $i$ ten Patient\*in
5. Berechnung der adjustierten PROs  $adjustiert_j = Performance_j + \bar{o}$  für jede Versorgungsorganisation mit dem beobachteten PRO-Durchschnittswert  $\bar{o}$  aller Patient\*innen aller Versorgungsorganisationen

Das beschriebene Vorgehen hat dabei mehrere nennenswerte Vorteile: Zum einen wird mit der multiplen linearen Regression eine vergleichsweise niederschwellige statistische Methode gewählt, um für alle Beteiligten, insbesondere Behandelnde, das methodische Vorgehen nachvollziehbar zu machen. Dies kann die Akzeptanz bei den genannten Beteiligten erheblich erhöhen (vgl. auch Abschnitt 5.3). Des Weiteren sind die adjustierten Scores weiterhin so skaliert wie die ursprünglich gemessenen. Auch dies erleichtert die Interpretation der Ergebnisse für alle Akteur\*innen. Schließlich wird durch das Nutzen eines additiven Ansatzes (im Gegensatz zu Raten- bzw. Quotienten-nutzenden Verfahren, vgl. auch Fußnote 6) verhindert, dass vor allem sehr niedrige beobachtete oder erwartete PROs (z. B. Werte zwischen 0 und 1) die adjustierten Werte stark beeinflussen können.

## 5.2. Forschungsfrage 2: Wie unterscheiden sich Krebszentren bezüglich der *patient-reported Outcomes* ihrer behandelten Patient\*innen voneinander?

Die wissenschaftliche und empirische Relevanz der oben beantworteten Fragestellung und die daraus abgeleitete, weiterentwickelte Adjustierungsmethode (vgl. 5.1) wurde anschließend auf die zwei Fallbeispiele (vgl. 2.4 und 2.5) angewandt und darauf aufbauend wurden die untersuchten Krebszentren verglichen, wobei sich durchaus unterschiedliche Resultate ergaben: Für die Prostatakrebszentren konnten nach der Casemix-Adjustierung für die behandelten Patienten relevante Unterschiede in den untersuchten PROs (Inkontinenz, Sexualität, irritative/obstruktive Symptomatik) festgestellt werden<sup>57</sup>. Um diese Unterschiede quantifizierbar zu machen und besser einordnen zu können, wurde eine auf MIDs beruhende Kennzahl eingeführt ( $MID_{Performancerate} = \frac{N_{adjustiert_j < (\overline{adjustiert} - MID)}}{N} * 100$ , wobei  $\overline{adjustiert}$

der Mittelwert aller adjustierter PROs ist,  $N_{adjustiert_j < (\overline{adjustiert} - MID)}$  die Anzahl aller

Versorgungsorganisationen mit einem adjustierten PRO, der mehr als eine MID kleiner ist als der Mittelwert aller adjustierten PROs und  $N$  die Anzahl aller am Vergleich teilnehmender Versorgungsorganisationen): diese Kennzahl gibt also die Prozentzahl derjenigen Versorgungsorganisationen an, in denen das Outcome mehr als eine MID – also für die Patient\*innen bemerkbar – schlechter ist als im Durchschnitt aller zu vergleichenden Versorgungsorganisationen. Durch Anwendung der  $MID_{Performancerate}$  konnte gezeigt werden, dass 26 % der Prostatakrebszentren ein deutlich schlechteres und für Patienten spürbares Inkontinenz-Outcome ein Jahr nach Operation aufweisen. Solche Unterschiede und vor allem auch solch eine negative Abweichung vom „Durchschnittsoutcome“ wurde vormals – aufgrund fehlender Daten und Analysemethoden – für Prostatakrebszentren noch nicht beschrieben. Sie identifizieren – aufbauend auf einer strikten, transparent dargelegten und nachvollziehbaren Methodik – erhebliche Versorgungsungleichheiten in für die Betroffenen besonders relevanten Funktionen. Für Prostatakrebszentren erwiesen sich also die Nutzung des EPIC-26 als spezifisches PRO-Instrument und der darauf aufbauende, Casemix-adjustierte Vergleich als für verschiedene Gruppen (Patienten, Behandelnde, Forschende) wertvoll (vgl. auch Abschnitt 5.4).

Auch für Darmkrebszentren wurden die durch den EORTC QLQ-C30 und -CR29 erhobenen PROs adjustiert und anschließend zum Zentrumsvergleich genutzt. Zwar konnte zunächst festgestellt werden, dass die beschriebene Adjustierungsmethode auch hierbei sehr gut anwendbar war und zuverlässig zu auswertbaren Casemix-adjustierten PROs auf Versorgungsebene führte. Jedoch wurden geringere, durch MIDs gemessene Unterschiede zwischen den Darmkrebszentren identifiziert: für die Domäne „physische Funktion“ hatten auf Grundlage der  $MID_{Performancerate}$  ca. 14 % der Darmkrebszentren ein für Patient\*innen bemerkbares, schlechteres Outcome als der Median aller teilnehmenden Zentren. Physische Funktion ist dabei der Score mit der größten Varianz zwischen den Zentren.

Eine mögliche Erklärung der geringeren Unterschiede zwischen Darmkrebszentren im Vergleich zu Prostatakrebszentren ist die Abwesenheit von einigen wenigen, für die operative Entfernung eines kolorektalen Karzinoms typischen Funktionseinschränkungen oder Symptomen. Im Gegensatz zum Prostatakrebs und der Radikalen Prostatektomie, bei der aufgrund des Operationszugangs und -gebiets vor allem die Kontinenz und Potenz der Behandelten eingeschränkt werden können, können beim Darmkrebs viele verschiedene Funktionseinschränkungen oder auch gar keine auftreten und das Spektrum der Nebenwirkungen ist sehr viel breiter. Dadurch diversifizieren sich die erfassten Symptome stark und anstatt einzelner, sehr spezifischer Symptome eignet sich eher ein Outcome, das ein breiter gefasstes Konstrukt erfasst – so wie physische Funktion. Im EORTC QLQ-C30 – dem Fragebogen, der für den Vergleich der Darmkrebszentren zur Anwendung kommt – besteht physische Funktion aus den meisten Einzelitems (insgesamt 5, vgl. auch Fußnote 2)

und deckt damit ein breiter gefasstes Spektrum an Funktionseinschränkung ab. Der Vergleich der Ergebnisse der Arbeiten 3 und 4 unterstreicht also für die Verwendung von PROs zum Versorgungsorganisationsvergleich unter anderem, wie wichtig die Wahl angemessener Outcomes ist (noch vor der daran anschließenden Wahl des eigentlichen PRO-Messinstruments<sup>59,60</sup>): So scheinen PROs, die sehr krankheitsspezifische und die Behandelten stark beeinflussende Einschränkungen oder Nebenwirkungen erfassen, geeigneter, als die Erhebung einer große Bandbreite an PROs, die zwar alle für Betroffene potentiell relevant sein können, jedoch zumeist Symptome erfassen, die selten eintreffen oder nur zu einer geringen Einschränkung führen. Auch die 2022 von der European Society for Medical Oncology (ESMO) erstellte Clinical Practice Guideline weist darauf explizit hin: „Outcomes to be assessed in a routine clinical care setting should be meaningful in the target population (i.e. prevalent and/or impactful on function or QoL) and clinically actionable (i.e. a management approach exists for clinicians to address the problem(s) through action, such as modifying cancer treatment or adding a supportive therapy)“<sup>60</sup>. Für Patienten, die vor ihrer Radikalen Prostatektomie kontinent und potent waren (ca. 97 % der bis 2021 eingeschlossenen PCO-Patienten sind vor Beginn der Behandlung kontinent<sup>7</sup>) und deren Risiko, nach ihrer operativen Prostatakrebsbehandlung täglich Einlagen tragen oder auf (medikamentöse) potenzsteigernde Mittel zurück greifen zu müssen recht hoch liegt (ein Jahr nach Operation benötigen nur noch 57 % der Patienten keine Einlagen zum Kontinenzertalt), trifft dies beispielsweise zu. Wenn solche markanten behandlungsassoziierten Funktionseinschränkungen also vorliegen, diese auch (am besten) durch Patient\*innen selbst erfasst werden können<sup>61</sup> und die Versorgungsorganisationen diese auch wirklich selbst beeinflussen können (wie die erheblichen Unterschiede nach erfolgter Casemix-Adjustierung zwischen den Prostatakrebszentren aufzeigen), dann kann die Messung dieser PROs zum Qualitätsvergleich von Versorgungsorganisationen herangezogen werden und sollte es aus - im Sinne der Patient\*innenorientierung als eine der Grundpfeiler der Versorgungsforschung<sup>5</sup>.

### **5.3. Forschungsfrage 3: Wie ist das Wissen über und die Haltung zu PROs von onkologisch Behandelnden in deutschen Krebszentren für die patient\*innenindividuelle sowie die Versorgungsorganisationsübergreifende Nutzung?**

Zur Beantwortung der dritten Forschungsfrage wurde ein qualitativer Ansatz gewählt<sup>62</sup>. Aus den durchgeführten Interviews konnte abgeleitet werden, dass die befragten Kliniker\*innen aus den an der PRO-Studie teilnehmenden Darmkrebszentren generell eher wenig Wissen

---

<sup>7</sup> Berechnung durch Autorin, Artikel in Vorbereitung

über PROs hatten, jedoch diesen grundsätzlich positiv gegenüber eingestellt waren. Es wurden vor allem die Möglichkeiten einer standardisierten und quantifizierbaren Symptom- und Funktionserfassung als gewinnbringend genannt: für eine patient\*innenindividuelle Nutzung wurden darüber hinaus vor allem der Wissenszuwachs über Symptome und Funktionen, die besondere Patient\*innenrelevanz, die Verbesserung der Ärzt\*in-Patient\*innenkommunikation sowie eine verbesserte Verlaufskontrolle als sich aus der PRO-Nutzung ergebende Vorteile gesehen. Nichtsdestotrotz wurde auch kritisch angemerkt, dass eine systematische PRO-Erfassung das eigentliche Ärzt\*innen-Patient\*innen-Gespräch nicht ersetzen könne und die verwendeten Instrumente (in diesem Fall: EORTC QLQ-C30 und -CR29) entweder zu spezifisch oder zu unspezifisch seien.

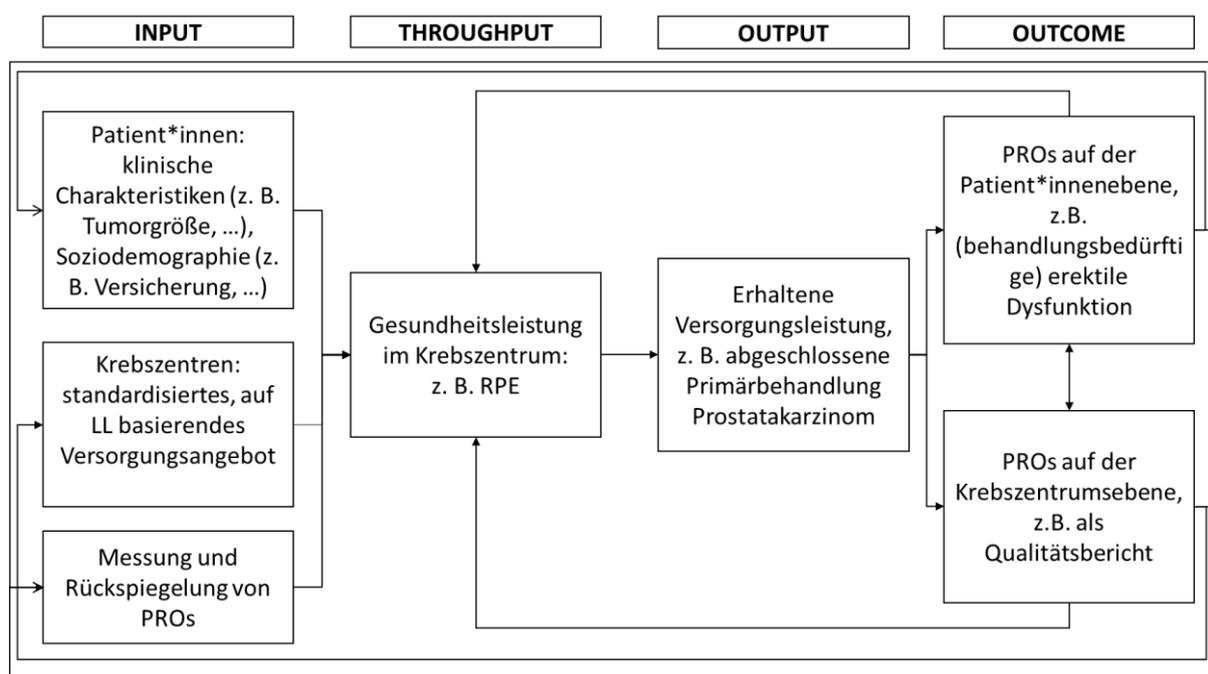
Neben der Untersuchung der Einstellung von Kliniker\*innen PROs gegenüber, wurden die qualitativen Interviews ferner genutzt, um erleichternde und erschwerende Faktoren für eine erfolgreiche PRO-Implementierung in der Praxis zu identifizieren<sup>53</sup>: Auf der technischen Ebene sind vor allem die Einbettung der PRO-Erfassung in ein elektronisches Krankenhausinformationssystem und eine klare Darstellung Patient\*innen-individueller Ergebnisse (bspw. durch farbliche Markierungen, Visualisierungen der Ergebnisse, etc.) wichtig für eine erfolgreiche Implementierung, wohingegen auf der organisatorischen Ebene die Einbettung in schon bestehende Strukturen, klare Verantwortlichkeiten (z.B. PRO-Koordinator\*innen), ein geringer Zeitverlust zwischen Erhebung und Ergebnismitteilung sowie die ausschließliche Nutzung von relevanten PRO-Messinstrumenten genannt wurden.

#### **5.4. PROs für die Versorgung von Krebspatient\*innen – Integration und Synthese der Ergebnisse auf verschiedenen Ebenen**

Die Zusammenstellung und Diskussion der Ergebnisse, die zur Beantwortung der Forschungsfragen herangezogen werden konnten, zeigen die verschiedenen Möglichkeiten der PRO-Nutzung auf unterschiedlichsten onkologischen Versorgungsebenen im konkreten Kontext der Krebszentren auf. Gerade für die Onkologie unterstreichen mehrere Übersichtsarbeiten den Nutzen von PROs für die klinische Routineversorgung und auch für Qualitätsentwicklungsinitiativen<sup>63,64</sup> und auch die ESMO Guideline zu PROs betont neben der klinischen, patient\*innenindividuellen Nutzung, dass PROs aggregiert als „quality metric“ zur „quality assurance and improvement in clinical care“<sup>60</sup> genutzt werden sollten. Diese Mehrfachnutzung von PROs wird also häufig adressiert und soll auch durch diese Arbeit unterstrichen werden.

Wenn die Ergebnisse der vorangestellten Publikationen also gemeinsam betrachtet werden, so fällt auf, dass die beiden als Fallbeispiele genutzten Studien (PCO- sowie EDIUM) eine Implementationsmöglichkeit dieser verschiedenen geforderten Ebenen liefern: Obwohl die

PROs (zum einen der EPIC-26, zum anderen die Fragebogen EORTC QLQ-C30 und -CR29) zunächst vor allem für den Ergebnisqualitätsvergleich erhoben wurden<sup>57</sup>, sehen Kliniker\*innen auch die Vorteile der patient\*innenindividuellen Nutzung<sup>62</sup> und können beispielweise auch explizit Scores nennen, die für die prätherapeutische Beurteilung von Darmkrebspatient\*innen relevant sind<sup>51</sup>. Für den konkreten Versorgungskontext (zertifizierte Krebszentren als integrierte Versorgungsnetzwerke) unterstreicht dies, wie die Messung von PROs, das patient\*innenindividuelle und Versorgungsorganisations-bezogene Rückspiegeln der daraus resultierenden Ergebnisse und dadurch die Versorgungs- und Qualitätsbewertung sich auf die verschiedenen Ebenen der Versorgung auswirken können.



**Abbildung 2:** Schematische Darstellung der Auswirkungen einer systematischen PRO-Erfassung auf unterschiedlichen Ebenen der onkologischen Versorgung am Beispiel des operativ behandelten Prostatakarzinoms. LL = Leitlinie, PRO = patient-reported Outcome, RPE = Radikale Prostatektomie

Helfen beim Verständnis dieser Zusammenhänge kann eine Betrachtung mithilfe des Throughput-Modells nach Pfaff und Schrappe<sup>5</sup> (vgl. Abbildung 2). Dabei fungieren sowohl die patient\*innenindividuellen als auch die adjustierten Zentrums-PROs als *Outcome*, das – einmal zurückgespiegelt an die Kliniker\*innen und potentiell auch an die Betroffenen selbst – den *Input* (Krebszentren, Patient\*innen, Intervention) beeinflussen kann, indem beispielsweise Strukturen in Krebszentren angepasst werden. Nach Donabedian kann man diese Input-Faktoren auch als Strukturfaktoren und damit zum ersten Teil des Struktur-Prozess-*Outcome*-Modells auffassen<sup>10</sup>. Die Intervention als „Input 2. Ordnung“, die nach Pfaff und Schrappe „die entscheidende Rolle im [...] Throughput [spielt]“<sup>45</sup>, ist bei der Implementierung von PROs das Messen der PROs (im Sinne eines Diagnostikums) sowie das Rückspiegeln der Ergebnisse

(auf der patient\*innenindividuellen Ebene und auch der Versorgungsorganisationsebene). Dabei kann das Rückspiegeln der PRO-Ergebnisse auf Versorgungsorganisationsebene wie für Darm- und Prostatakrebszentren, die an PCO oder EDIUM teilnehmen, durch Ergebnisqualitätsberichte erfolgen<sup>65,66</sup>. Dies wiederum hat Einfluss (z. B. durch Identifizierung von Versorgungslücken oder strukturell auftretenden spezifischen Operationsnebenwirkungen im eigenen Krebszentrum) auf die eigentliche Gesundheitsleistung (z. B. verbesserte Operationstechnik) als *Throughput* und damit auf den *Output* – die abgeschlossenen Krebsbehandlungen. Durch die zuvor erfolgten Anpassungen (also durch Struktur- und Prozessanpassungen nach Donabedian) kann so auch das *Outcome* potentiell verbessert werden – sowohl auf patient\*innenindividueller Ebene (z. B. weniger Patienten mit erektiler Dysfunktion) als auch auf Versorgungsorganisationsebene (bessere adjustierte Sexualitätsscores im Vergleich mit anderen Krebszentren). Hervorgehoben werden sollte hierbei, dass in der vorliegenden Arbeit keinerlei Faktoren des *Throughputs* (bspw. Einfluss von Operationstechnik auf Kontinenz) untersucht wurden. Dies sollte Gegenstand weiterer Forschung sein (bzw. ist zum Teil auch bereits aus der Literatur bekannt<sup>67</sup>) und kann Kliniker\*innen und anderen relevanten Akteur\*innen helfen, auf PROs aufbauende Initiativen zu nutzen, um den Versorgungsalltag aktiv besser zu gestalten – im Sinne der Betroffenen.

## 5.5. Fazit

*Patient-reported Outcomes* (PROs) und die für ihre Messung verwendeten Fragebogen sind Instrumente, die *Outcomes* aus der Patient\*innenperspektive quantifizierbar machen und zunehmend in der Forschung und Klinik genutzt werden<sup>60</sup>. In der vorliegenden Arbeit konnten viele mehrschichtige Anwendungsbereiche und -potentiale von PROs für die Onkologie aufgezeigt werden, um die Versorgung von Krebspatient\*innen nachhaltig zu verbessern: für die patient\*innenindividuellen Ebene identifizierten Behandelnde viele Chancen durch eine routinemäßige PRO-Erhebung (die zu einer Verbesserung der individuellen Versorgungsleistungen führen können) und auf der Versorgungsorganisationsebene konnten PROs erfolgreich zum Qualitätsvergleich eingesetzt werden. Die Einbettung der auf dem PRO-Vergleich basierenden Ergebnisse kann also – basierend auf den vorgestellten und diskutierten Arbeiten – vor allem unter folgenden Bedingungen gelingen: Einbettung in bereits bestehende Organisationsstrukturen (wie in den zertifizierten Krebszentren), Nutzung bereits bestehender Organisationsprozesse (wie die Tumordokumentation, die sowohl für die jährlichen Audits der Zentren als auch für EDIUM und PCO genutzt wird), Einbeziehung der relevanten Akteur\*innen (Betroffene, Behandelnde) und methodische Angemessenheit und Transparenz (durch Nutzung statistischer Adjustierungsmodelle, die so verständlich wie möglich den Anwendenden vermittelt werden). Damit aus der Nutzung von PROs in der Onkologie auch eine tatsächliche Intervention – also eine Maßnahme, die auch potentielle

Änderungen mit sich birgt – werden kann, können Donabedians Antworten auf die von ihm selbst gestellte Frage „What must be done to enhance [the power of quality monitoring] to do good and, perhaps, to reduce the likelihood of its doing harm?“<sup>10</sup> dienen. PROs werden hier also, wie in 2.2 und 2.3 dargelegt, als Qualitätsmessungen auf unterschiedlichen Ebenen aufgefasst. Donabedian nennt unter anderem folgende, auf PROs anwendbare Punkte: Zunächst unterstreicht er den Zusammenhang zwischen vorhandenen Strukturen, Prozessen und den gemessenen *Outcomes* und fordert „readjustments in the system“ (kurz- sowie langfristig) durch „restructuring of incentives that allow good performance to be recognized and rewarded while substandard performance is also recognized and, if possible, corrected“<sup>10</sup>. Dieser Forderung entsprechen die den Krebszentren zur Verfügung gestellten jährlichen Ergebnisberichte, in denen die Zentren ihre eigenen (adjustierten) PRO-Ergebnisse mit denen der anderen Zentren vergleichen können und so gute, aber auch schlechtere Versorgungsleistungen identifiziert werden können. Des Weiteren nennt Donabedian „facilitation, routinization, circumvention [and] direct intervention“ (ebd.) als mögliche Anpassungsstrategien einzelner Versorgungsorganisationen und Behandelnder. Die standardisierte Erfassung von PROs und zeitnahe Ergebnismitteilung dieser adressiert vor allem drei dieser Faktoren: durch die niederschwellige Symptomerfassung und Nutzung als Diagnostikum wird die Verbesserung der Versorgungsleistung vereinfacht (*facilitation*,<sup>17,19,68</sup>), die Nutzung von PROs kann in bestehende Abläufe standardmäßig integriert werden – vor allem, wenn „alert systems“ und eine elektronische PRO-Erhebung („ePROs“) genutzt werden (*routinization*,<sup>69,70</sup>), das Rückspiegeln der PRO-Ergebnisse ermöglicht die *direct intervention* bei Versorgungslücken, -ungleichheiten oder -auffälligkeiten. Dies trifft allerdings nur zu, wenn die gemessenen PROs den folgenden, von Donabedian geforderten Kriterien entsprechen: „1. The outcome selected should be relevant to the objective of care; it stands for what the clinician is aiming for. 2. The outcome must be achievable by good care. This means that the methods for doing this are available and under the control of the health-care system. 3. The outcome, whether good or bad, must be attributable first to health care, and then, to the contribution of the practitioner (or other person) whose performance is being assessed. 4. The duration of the outcome as well as its magnitude should be taken into account“<sup>10</sup>. Dies kann auch bedeuten, dass PROs, die auf der patient\*innenindividuellen Ebene (also für die individuelle onkologische Behandlung im Sinne einer Versorgungsleistung) höchst relevant sind, weil sie ggfs. seltene oder nur einige wenige Patient\*innen betreffende Symptome erfassen, die jedoch für die\*den einzelne\*n Patient\*in mit einem hohen Grad an Einschränkungen verbunden sind, nicht für die Nutzung auf der Versorgungsorganisationsebene geeignet sind. Bei der Wahl von geeigneten PRO-Instrumenten sollten also im besten Falle beide Ebenen adressiert werden können, um das Potential von PROs für den Versorgungskontext möglichst gut auszuschöpfen.

Neben den hier aufgezeigten Möglichkeiten der PRO-Nutzung im klinischen und qualitätsentwickelnden Kontext, deckt diese Arbeit auch viel weiteren (Versorgungs-) Forschungsbedarf auf: So lassen sich das oben beschriebene Adjustierungsverfahren und darauf aufbauende Versorgungsorganisationsvergleiche auch auf andere Tumorentitäten übertragen (unter Berücksichtigung der Anforderungen an geeignete PROs, s. o.), die Wirkung der PRO-Rückspiegelung (z. B. als jährliche Qualitätsberichte wie bei EDIUM<sup>65</sup> und PCO<sup>66</sup>) auf die Ergebnisqualität in den Zentren und ggfs. dafür eingeleitete Maßnahmen näher untersuchen sowie auch internationale Versorgungsvergleiche, die auf PROs basieren, vorantreiben (was im Falle des Prostatakarzinoms bspw. im Rahmen des TrueNorth Global Registries angestrebt wird<sup>35</sup>).

Das Ziel des Nationalen Krebsplans wird vom Bundesministerium für Gesundheit wie folgt beschrieben: „Wichtig sind [...] Maßnahmen zur Unterstützung von Krebskranken und ihren Angehörigen bei der Bewältigung der körperlichen, seelischen und sozialen Folgen der Erkrankung. Wir wollen die Aktivitäten aller an der Krebsbekämpfung Beteiligten wirksamer aufeinander abstimmen und ein zielorientiertes Vorgehen befördern.“<sup>71</sup> Durch eine Implementierung von PROs auf den verschiedenen Versorgungsebenen wird genau dieses Ziel adressiert und PROs können somit zur Verbesserung der onkologischen Versorgung nachhaltig beitragen. Durch die Identifizierung angemessener Adjustierungsmethodiken, deren explizite Anwendung auf zwei Versorgungskontexte (Prostata- und Darmkrebs), daraus abgeleiteten Empfehlungen für die Implementierung in anderen Kontexten (bspw. zur Wahl geeigneter PROs, Einbettung in bestehende Versorgungsstrukturen, Einbeziehung aller relevanter Beteiligter) und die theoretische Einordnung der PRO-Nutzung zum einen im Kontext der *Outcomes Research* im Sinne Donabedians<sup>11</sup>, zum anderen als Antwort auf eine zunehmend geforderte *patient-centredness* (bspw. bei Scholl<sup>21</sup>), hat diese Arbeit verschiedenste Wege sowohl bezüglich methodischer Ansätze als auch struktureller Ebenen aufzeigen wollen, wie und in welchem Rahmen *patient-reported Outcomes* für die Versorgung von Krebspatient\*innen sinnvoll nutzbar sind und zur Umsetzung der Forderungen und Handlungsfelder des Nationalen Krebsplans maßgeblich beitragen können, denn „wir fragen ja wirklich, wie es den Leuten geht“.<sup>8</sup>

---

<sup>8</sup> Zitat eines Chirurgen aus der Antwort auf die Frage „Wie würden Sie [PROs] benutzen?“, die im Rahmen der Interviews an den EDIUM-Zentren gestellt wurde (vgl. Arbeit 5)<sup>62</sup>

## 6. Literaturverzeichnis

- 1 Pfaff H. Versorgungsforschung - Begriffsbestimmung, Gegenstand und Aufgaben. In: Pfaff H, Schrappe M, Lauterbach KW, Engelmann U, Halber M, eds. Gesundheitsversorgung und Disease Management. Grundlagen und Anwendungen der Versorgungsforschung. Bern: Hans HÜber, 2003: 13–23.
- 2 Kowalski C, Albert U-S, Baumann W, *et al.* DNVF-Memorandum Versorgungsforschung in der Onkologie. *Gesundheitswesen* 2020; **82**: e108–21.
- 3 Todesursachen. Statistisches Bundesamt. 2022; published online June 6. [https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Gesundheit/Todesursachen/\\_inhalt.html](https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Gesundheit/Todesursachen/_inhalt.html) (accessed June 6, 2022).
- 4 Arndt V, Dahm S, Kraywinkel K. Krebsprävalenz in Deutschland 2017. *Onkologe* 2021; **27**: 717–23.
- 5 Pfaff H, Neugebauer EA, Glaeske G, Schrappe M. Lehrbuch Versorgungsforschung, 2nd edn. Stuttgart: Schattauer, 2017.
- 6 Wintner LM, Giesinger JM, Kemmler G, *et al.* Verwendung und Nutzen von Patient-Reported Outcomes in der onkologischen Behandlung: eine Übersicht. *Wien Klin Wochenschr* 2012; **124**: 293–303.
- 7 Lee SJ, Earle CC, Weeks JC. Outcomes research in oncology: history, conceptual framework, and trends in the literature. *J Natl Cancer Inst* 2000; **92**: 195–204.
- 8 Donabedian A. The End Results of Health Care: Ernest Codman's Contribution to Quality Assessment and Beyond. *The Milbank Quarterly* 1989; **67**: 233.
- 9 Codman EA. The product of a hospital. *Surg Gynecol Pbstet* 1914; **18**: 491–6.
- 10 Donabedian A. An Introduction to Quality Assurance in Health Care. Oxford University Press, 2002.
- 11 Donabedian A. Evaluating the quality of medical care. *The Milbank Memorial Fund quarterly* 1966; **44**: Suppl:166-206.
- 12 Karakiewicz PI, Briganti A, Chun FK-H, Valiquette L. Outcomes Research: A Methodologic Review. *European Urology* 2006; **50**: 218–24.
- 13 OECD. RECOMMENDATIONS TO OECD MINISTERS OF HEALTH FROM THE HIGH LEVEL REFLECTION GROUP ON THE FUTURE OF HEALTH STATISTICS. Strengthening the international comparison of health system performance through patient-reported indicators. 2017.
- 14 Brazier JE, Harper R, Jones NM, *et al.* Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ* 1992; **305**: 160–4.
- 15 Wintner LM, Sztankay M, Giesinger JM, *et al.* EORTC Quality of Life Group manual for the use of EORTC measures in daily clinical practice. 2016.
- 16 Fromme EK, Eilers KM, Mori M, Hsieh YC, Beer TM. How accurate is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the Quality-of-Life Questionnaire C30. *Journal of Clinical Oncology* 2004; **22**: 3485–90.

- 17 Laugsand EA, Sprangers MA, Bjordal K, Skorpen F, Kaasa S, Klepstad P. Health care providers underestimate symptom intensities of cancer patients: a multicenter European study. *Health and quality of life outcomes* 2010; **8**: 104.
- 18 Klinkhammer-Schalke M, Steinger B, Koller M, *et al.* Diagnosing deficits in quality of life and providing tailored therapeutic options: Results of a randomised trial in 220 patients with colorectal cancer. *Eur J Cancer* 2020; **130**: 102–13.
- 19 Basch E, Deal AM, Dueck AC, *et al.* Overall Survival Results of a Trial Assessing Patient-Reported Outcomes for Symptom Monitoring During Routine Cancer Treatment. *JAMA* 2017; **318**: 197–8.
- 20 Denis F, Lethrosne C, Pourel N, *et al.* Randomized Trial Comparing a Web-Mediated Follow-up With Routine Surveillance in Lung Cancer Patients. *Journal of the National Cancer Institute* 2017; **109**. DOI:10.1093/jnci/djx029.
- 21 Scholl I, Zill JM, Härter M, Dirmaier J. An Integrative Model of Patient-Centeredness – A Systematic Review and Concept Analysis. *PLoS One* 2014; **9**: e107828.
- 22 Cella D, Yount S, Rothrock N, *et al.* The Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care* 2007; **45**: S3–11.
- 23 NHS England Analytical Team. Patient Reported Outcome Measures (PROMs) An alternative aggregation methodology for case-mix adjustment. 2013. <https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/07/proms-agg-meth-adju.pdf> (accessed Oct 7, 2020).
- 24 NHS England. Changes to the National Patient Reported Outcomes Measures programme. 2022. <https://www.england.nhs.uk/ourwork/insight/promsconsultation/> (accessed June 13, 2022).
- 25 ICHOM. International Consortium for Health Outcomes Measurement. ICHOM. 1970; published online Jan 1. <https://www.ichom.org/> (accessed Nov 19, 2022).
- 26 DeMartino JK. Measuring quality in oncology: challenges and opportunities. *J Natl Compr Canc Netw* 2013; **11**: 1482–91.
- 27 Robert-Koch-Institut. Krebs in Deutschland für 2017/2018. Berlin, 2021.
- 28 Iezzoni L. Risk Adjustment for Measuring Healthcare Outcomes, 4th edn. Chicago, IL: Health Administration Press, 2012.
- 29 Bundesministerium für Gesundheit. Nationaler Krebsplan. Handlungsfelder, Ziele und Umsetzungsempfehlungen. Berlin, 2012.
- 30 Das Zertifizierungsprogramm der Deutschen Krebsgesellschaft | DKG. <https://www.krebsgesellschaft.de/deutsche-krebsgesellschaft/zertifizierung.html> (accessed May 4, 2021).
- 31 Deutsche Krebsgesellschaft. Erhebungsbogen für Prostatakrebszentren, Inkraftsetzung am 18.08.2020. Berlin, 2020 <https://www.onkozeit.de/organ/prostata/>.
- 32 Deutsche Krebsgesellschaft. Erhebungsbogen für Darmkrebszentren der Deutschen Krebsgesellschaft, Inkraftsetzung am 28.08.2014. Berlin, 2014.

- 33 Kowalski C, Kämmerle A, Ortmann O. Die OncoBox Research als Instrument für Versorgungsforschung. *Der Onkologe* 2017; **23**: 736–41.
- 34 Kowalski C, Carl G, Feick G, Wesselmann S. The Prostate Cancer Outcomes (PCO) study in prostate cancer centres certified by the German Cancer Society. *Strahlenther Onkol* 2020; published online June 16. DOI:10.1007/s00066-020-01655-3.
- 35 Evans SM, Millar JL, Moore CM, *et al.* Cohort profile: the TrueNTH Global Registry - an international registry to monitor and improve localised prostate cancer health outcomes. *BMJ Open* 2017; **7**: e017006.
- 36 Deutsche Krebsgesellschaft e. V., Bundesverband Prostatakrebs Selbsthilfe, OnkoZert GmbH. Studienprotokoll "Prostate Cancer Outcomes - Compare & Reduce Variation in DKG-zertifizierten Prostatakrebszentren ". 2018; published online June 14. [https://www.pco-study.com/DownloadData/Studienprotokoll%202.0%20Deutsch%20\(Version%2014.06.2018\).pdf](https://www.pco-study.com/DownloadData/Studienprotokoll%202.0%20Deutsch%20(Version%2014.06.2018).pdf).
- 37 Szymanski KM, Wei JT, Dunn RL, Sanda MG. Development and validation of an abbreviated version of the expanded prostate cancer index composite instrument for measuring health-related quality of life among prostate cancer survivors. *Urology* 2010; **76**: 1245–50.
- 38 Sibert NT, Dieng S, Oesterle A, *et al.* Psychometric validation of the German version of the EPIC-26 questionnaire for patients with localized and locally advanced prostate cancer. *World Journal of Urology* 2021; **39**: 11–25.
- 39 Beyer B, Huland H, Feick G, Graefen M. 'Expanded prostate cancer index composite' (EPIC-26): Funktionelles Behandlungsergebnis bei Patienten mit lokalisiertem Prostatakarzinom. *Urologe* 2015; **54**: 1591–5.
- 40 Bianco FJ, Scardino PT, Eastham JA. Radical prostatectomy: long-term cancer control and recovery of sexual and urinary function ('trifecta'). *Urology* 2005; **66**: 83–94.
- 41 Whistance RN, Conroy T, Chie W, *et al.* Clinical and psychometric validation of the EORTC QLQ-CR29 questionnaire module to assess health-related quality of life in patients with colorectal cancer. *European Journal of Cancer* 2009; **45**: 3017–26.
- 42 Sprangers MA, Cull A, Bjordal K, Groenvold M, Aaronson NK. The European Organization for Research and Treatment of Cancer. Approach to quality of life assessment: guidelines for developing questionnaire modules. EORTC Study Group on Quality of Life. *Qual Life Res* 1993; **2**: 287–95.
- 43 Breidenbach C, Sibert NT, Wesselmann S, Kowalski C. Die Beratung durch Ethikkommissionen bei einer multizentrischen Beobachtungsstudie in Deutschland – Aufwand und Kosten. *Gesundheitswesen* 2020; : a-1192-4946.
- 44 Cornelius F. Krebsregister: Bewertung der aktuellen Situation. *Uro-News* 2021; **25**: 34–9.
- 45 Blum TG, Schönfeld N, Jagota A, Klinkhammer-Schalke M. Integration und Steuerfunktion klinischer Krebsregister in der onkologischen Versorgung. *Forum* 2012; **27**: 431–5.
- 46 ICHOM - International Consortium for Health Outcomes Measurement. Localized Prostate Cancer Data Collection Reference Guide. 2015.

- 47 Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951; **16**: 297–334.
- 48 Higgins JPT, Thomas J, Chandler J, *et al.* Cochrane Handbook for Systematic Reviews of Interventions. John Wiley & Sons, 2019.
- 49 Critical Appraisal Skills Programme. CASP Systematic Review Checklist [online]. 2018. [https://casp-uk.net/wp-content/uploads/2018/01/CASP-Systematic-Review-Checklist\\_2018.pdf](https://casp-uk.net/wp-content/uploads/2018/01/CASP-Systematic-Review-Checklist_2018.pdf) (accessed March 18, 2021).
- 50 Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, N.J.: L. Erlbaum Associates, 1988.
- 51 Sibert NT, Breidenbach C, Wesselmann S, *et al.* Which EORTC QLQ-C30 and -CR29 scores are relevant for clinicians for therapy planning and decisions?: Results of an online survey. *coloproctology* 2021; published online Sept 8. DOI:10.1007/s00053-021-00560-2.
- 52 Kuckartz U. Qualitative text analysis: a guide to methods, practice & using software. Los Angeles: SAGE, 2014.
- 53 Breidenbach C, Kowalski C, Wesselmann S, Sibert NT. Could existing infrastructure for using patient-reported outcomes as quality measures also be used for individual care in patients with colorectal cancer? *BMC Health Services Research* 2021; **21**: 448.
- 54 Sibert NT, Pfaff H, Breidenbach C, Wesselmann S, Kowalski C. Different Approaches for Case-Mix Adjustment of Patient-Reported Outcomes to Compare Healthcare Providers-Methodological Results of a Systematic Review. *Cancers (Basel)* 2021; **13**: 3964.
- 55 Hendryx MS, Dyck DG, Srebnik D. Risk-adjusted outcome models for public mental health outpatient programs. *Health Serv Res* 1999; **34**: 171–95.
- 56 Lutz AD, Brooks JM, Chapman CG, Shanley E, Stout CE, Thigpen CA. Risk Adjustment of the Modified Low Back Pain Disability Questionnaire and Neck Disability Index to Benchmark Physical Therapist Performance: Analysis From an Outcomes Registry. *Phys Ther* 2020; **100**: 609–20.
- 57 Sibert NT, Pfaff H, Breidenbach C, *et al.* Variation across operating sites in urinary and sexual outcomes after radical prostatectomy in localized and locally advanced prostate cancer. *World J Urol* 2022; published online March 26. DOI:10.1007/s00345-022-03985-6.
- 58 Kowalski C, Sibert NT, Breidenbach C, *et al.* Ergebnisqualität in zertifizierten Darmkrebszentren nach Resektion eines kolorektalen Karzinoms - Patient-Reported und kurzfristig-klinische Outcomes. *Deutsches Ärzteblatt* 2022; published online Dec 2. DOI:10.3238/arztebl.m2022.0325.
- 59 Weldring T, Smith SMS. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). *Health Serv Insights* 2013; **6**: 61–8.
- 60 Di Maio M, Basch E, Denis F, *et al.* The role of patient-reported outcome measures in the continuum of cancer clinical care: ESMO Clinical Practice Guideline. *Annals of Oncology* 2022; **33**: 878–92.
- 61 Kielb S, Dunn RL, Rashid MG, *et al.* Assessment of early continence recovery after radical prostatectomy: patient reported symptoms and impairment. *J Urol* 2001; **166**: 958–61.

- 62 Sibert NT, Kowalski C, Pfaff H, Wesselmann S, Breidenbach C. Clinicians' knowledge and attitudes towards patient reported outcomes in colorectal cancer care - insights from qualitative interviews. *BMC Health Serv Res* 2021; **21**: 366.
- 63 Kotronoulas G, Kearney N, Maguire R, *et al.* What Is the Value of the Routine Use of Patient-Reported Outcome Measures Toward Improvement of Patient Outcomes, Processes of Care, and Health Service Outcomes in Cancer Care? A Systematic Review of Controlled Trials. *Journal of Clinical Oncology* 2014; **32**: 1480–501.
- 64 Strong LE. The past, present, and future of patient-reported outcomes in oncology. *Am Soc Clin Oncol Educ Book* 2015; : e616-620.
- 65 Deutsche Krebsgesellschaft (DKG). Ergebnisbericht an die EDIUM-Studienzentren. 2021.
- 66 Deutsche Krebsgesellschaft. Prostate Cancer Outcome (PCO) Studie - Ergebnisbericht an die Studienzentren Dezember 2020. Berlin, 2020.
- 67 Davis M, Egan J, Marhamati S, Galfano A, Kowalczyk KJ. Retzius-Sparing Robot-Assisted Robotic Prostatectomy: Past, Present, and Future. *Urol Clin North Am* 2021; **48**: 11–23.
- 68 Klinkhammer-Schalke M, Koller M, Steinger B, *et al.* Direct improvement of quality of life using a tailored quality of life diagnosis and therapy pathway: randomised trial in 200 women with breast cancer. *British journal of cancer* 2012; **106**: 826–38.
- 69 Basch E, Barbera L, Kerrigan CL, Velikova G. Implementation of patient-reported outcomes in routine medical care. *American Society of Clinical Oncology Educational Book* 2018; **38**: 122–34.
- 70 Basch E, Deal AM, Kris MG, *et al.* Symptom Monitoring With Patient-Reported Outcomes During Routine Cancer Treatment: A Randomized Controlled Trial. *Journal of Clinical Oncology* 2015; published online Dec 7. DOI:10.1200/jco.2015.63.0830.
- 71 Bundesministerium für Gesundheit. Nationaler Krebsplan - Handlungsfelder, Ziele, Umsetzungsempfehlungen und Ergebnisse. 2017.

## 7. Anhang

### 7.1. Abbildungsverzeichnis

<b>Abbildung 1:</b> Beispielillustration für die Notwendigkeit einer Casemix-Adjustierung.....	15
<b>Abbildung 2:</b> Schematische Darstellung der Auswirkungen einer systematischen PRO-Erfassung auf unterschiedlichen Ebenen der onkologischen Versorgung am Beispiel des operativ behandelten Prostatakarzinoms.....	101

## 8. Vorabveröffentlichungen von Ergebnissen

Diese Promotionsschrift basiert zum einen als kumulative Arbeit aus fünf Hauptpublikationen, die vorab veröffentlicht wurden:

- Arbeit 1: **Sibert, N. T.**, Dieng, S., Oesterle, A., Feick, G., Carl, G., Steiner, T., Minner, J., Roghmann, F., Kaftan, B., Zengerling, F., Hinkel, A., Beyer, B., Heidenreich, A., Harke, N., Brehmer, B., Pfitzenmaier, J., Fichtner, J., Neisius, A., Hammerer, P., Wesselmann, S., Kowalski, C. (2019). Psychometric validation of the German version of the EPIC-26 questionnaire for patients with localized and locally advanced prostate cancer. *World Journal of Urology*. <https://doi.org/10.1007/s00345-019-02949-7>
- Arbeit 2: **Sibert, N. T.**, Pfaff, H., Breidenbach, C., Wesselmann, S., & Kowalski, C. (2021). Different Approaches for Case-Mix Adjustment of Patient-Reported Outcomes to Compare Healthcare Providers-Methodological Results of a Systematic Review. *Cancers*, 13(16), 3964. <https://doi.org/10.3390/cancers13163964>
- Arbeit 3: **Sibert, N. T.**, Pfaff, H., Breidenbach, C., Wesselmann, S., Roth, R., Feick, G., Carl, G., Dieng, S., Gaber, A. A., Blana, A., Darr, C., Distler, F., Kunath, F., Bedke, J., Erdmann, J., Minner, J., Simon, J., Kwiatkowski, M., Burchardt, M., Harz, N., Conrad, S., Höfner, T., Knoll, T., Beyer, B., Hammerer, P., Kowalski, C. (2022). Variation across operating sites in urinary and sexual outcomes after radical prostatectomy in localized and locally advanced prostate cancer. *World Journal of Urology*. <https://doi.org/10.1007/s00345-022-03985-6>
- Arbeit 4: Kowalski, C., **Sibert, N. T.**, Breidenbach, C., Hagemeyer, A., Roth, R., Seufferlein, T., Benz, S., Post, S., Siegel, R., Wiegering, A., Winkels, R., Bieck-Messemer, S., Fahlke, J., Reissfelder, C., Fuchs, M., Herzog, T., Weihrauch, R., Faber-Mertens, J., Rudolph, H., Puskás, L., Kohlhaw, K., Szczerbinska, M., Scheuerlein, H., Neumann, P.-A., Hollerbach, S., Riechmann, M., Kolbe, E. W., Weigert, N., Königer, J., Klink, C., Mussa, S., Horn, A.-K., Staib, L., Werner, J., Jähne, J., Aly, M., Mörk, H., Grützmann, R., Piso, P., Dieng, S., Wesselmann, S. (2022). Ergebnisqualität in zertifizierten Darmkrebszentren nach Resektion eines kolorektalen Karzinoms—Patient-Reported und kurzfristig-klinische Outcomes. *Deutsches Ärzteblatt*. <https://doi.org/10.3238/arztebl.m2022.0325>
- Arbeit 5: **Sibert, N. T.**, Kowalski, C., Pfaff, H., Wesselmann, S., & Breidenbach, C. (2021). Clinicians' knowledge and attitudes towards patient reported outcomes in colorectal cancer care—Insights from qualitative interviews. *BMC Health Services Research*, 21(1), 366. <https://doi.org/10.1186/s12913-021-06361-z>

Zusätzlich dazu referenziert diese Arbeit die folgenden, aus den Projektvorhaben hervorgegangenen und für die Promotionsschrift relevanten Publikationen:

- Breidenbach C, Kowalski C, Wesselmann S, **Sibert NT**. Could existing infrastructure for using patient-reported outcomes as quality measures also be used for individual care in patients with colorectal cancer? *BMC Health Services Research* 2021; **21**: 448.
- **Sibert NT**, Breidenbach C, Wesselmann S, Schult S, Benz S, Post S, Seufferlein T, Schloss P, Kowalski C. Which EORTC QLQ-C30 and -CR29 scores are relevant for clinicians for therapy planning and decisions?: Results of an online survey. *coloproctology* 2021; DOI:10.1007/s00053-021-00560-2.