

Aus dem Institut für Versorgungsforschung und Gesundheitsökonomie
der Heinrich-Heine-Universität Düsseldorf
Leiterin: Univ.-Prof. Dr. Dr. Andrea Icks

**Evidenzbasierung der Instrumente zur Erhebung der
gesundheitsbezogenen Lebensqualität im Rahmen der
frühen Nutzenbewertung nach § 35a SGB V**

Am Beispiel von Verfahren in der Onkologie

Dissertation

zur Erlangung des Grades eines Doktors der Public Health
der Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Holger Bender

2021

Als Inauguraldissertation gedruckt mit der Genehmigung der
Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf

gez.:

Dekan: Prof. Dr. med. Nikolaj Klöcker

Erstgutachterin: Prof. Dr. med. Dr. PH Andrea Icks

Zweitgutachter: Prof. Dr. sc. hum. Adrian Loerbroks

Nur das Denken, das wir leben, hat einen Wert.

Hermann Hesse

Zusammenfassung

Die gesundheitsbezogene Lebensqualität ist ein zunehmend wichtiger Endpunkt in klinischen Therapiestudien. Idealerweise werden validierte Instrumente zur Erhebung der gesundheitsbezogenen Lebensqualität verwendet. Validierungsstudien dienen dabei der Bestimmung der psychometrischen Güte und der Definition von Kriterien für die klinische Relevanz. Die Frage ist, in welchem Maße Charakteristika wie beispielsweise das Alter oder der Erkrankungsgrad der Studienpopulation von Therapie- und Validierungsstudien übereinstimmen. Nichtübereinstimmung würde die Validität herangezogener psychometrischer Gütekriterien einschränken, ebenso die Belastbarkeit verwendeter klinischer Relevanzkriterien.

Ziel der Arbeit ist die Bestimmung von Umfang und Wertigkeit der Informationen zur gesundheitsbezogenen Lebensqualität im Kontext der frühen Nutzenbewertung nach § 35a SGB V. Im Fokus stehen die verbreitetsten Instrumente zur Erhebung der gesundheitsbezogenen Lebensqualität bei onkologischen Erkrankungen. Anhand zentraler Studien- und Patientenmerkmale werden spezifische Indikatoren zur Bewertung der Evidenzbasierung beschrieben. Daneben ist die Arbeit darauf gerichtet, Abweichungen zwischen Therapie- und Validierungsstudien einerseits und zwischen den Tumorentitäten andererseits anzugeben. Bestimmt durch Unterschiede in den Behandlungssituationen von Therapie- und Validierungsstudien sowie den Informationen zum Güteprofil eingesetzter Fragebögen aus Validierungsstudien wird die Robustheit der Evidenz beurteilt. Festgestellte Limitationen in Bezug auf die Verwertbarkeit der Evidenz in den Nutzenbewertungsverfahren werden einem dort beanspruchten Zusatznutzen gegenübergestellt und diskutiert.

Die Arbeit basiert auf den abgeschlossenen Nutzenbewertungsverfahren zu onkologischen Erkrankungen mit einem Bewertungsbeginn bis Januar 2016 (N = 59). Zunächst werden Auszüge zum Geltungsgrad von Endpunkten der gesundheitsbezogenen Lebensqualität, zum Design der klinischen Prüfungen, zum konzeptionellen Stellenwert sowie zu Art und Umfang der Informationen zur Zielgröße gesammelt und die Befunde deskriptiv (relative Häufigkeiten, Durchschnittsvergleiche je Tumorentität) ausgewertet. Zusätzlich werden Angaben zu Alter, Geschlecht, Grad der Erkrankungsschwere, Ethnizität und geographische Region inhaltsanalytisch erfasst. Die Ergebnisse werden deskriptiv (absolute Häufigkeiten, relative Abweichungen) und inferentiell (ungepaarter t-Test im Falle intervallskaliert operationalisierter Variablen, exakter Test nach Fisher im Falle nominalskaliert Variablen) aufbereitet. Überdies werden Beschreibungen der Validität, der Reliabilität und der Änderungssensitivität bezüglich der verbreitetsten Erhebungsinstrumente aus den Unterlagen extrahiert und qualitativ interpretiert. Zuletzt werden die Ergebnisse einzelner Betrachtungsebenen katalogisiert.

In 80 % der Verfahren wird die gesundheitsbezogene Lebensqualität erhoben und zur Ableitung eines Zusatznutzens berichtet. Zu 88 % kommen randomisierte kontrollierte Studien zum Einsatz. Diese sind zu 52 % doppelblind designt. Etwa jede zweite Studie untersucht den Interventionseffekt gegenüber einem aktiven Komparator. In den Therapiestudien sind Endpunkte der gesundheitsbezogenen Lebensqualität überwiegend als sekundäre Zielkriterien definiert. Für die krankheitsspezifischen Profilinstrumente ergibt sich in der Gegenüberstellung von Therapie- und Validierungsstudien zu 54 % eine umfassende und zu 18 % eine anteilige Übereinstimmung der Tumorentitäten. Im Hinblick auf soziodemographische und klinische Merkmale finden sich signifikante Unterschiede zwischen den Populationen der Therapie- und Validierungsstudien: in 83 % für das Alter, 54 % für das Geschlecht, 83 % für den Grad der Erkrankungsschwere, 59 % für die Ethnizität und 71 % für die geographische Region. In acht in den Nutzenbewertungsverfahren häufig angeführten Validierungsstudien sind in 2 Fällen Subskalen bzw. Komponenten unberücksichtigt, in 2 weiteren Fällen eingeschränkt bzw. in 4 Fällen ungleich belastbar. Übersichten fassen die einbezogenen Patientencharakteristika von Validierungsstudien, die Abweichungen zu Therapiestudien und die Einschätzung des Verfassers zur Aussagesicherheit von Skalen ausgewählter Fragebögen zusammen.

Angesichts substanzieller Abweichungen zwischen Therapie- und Validierungsstudien in zentralen Studiencharakteristika sowie mitunter fehlenden oder minder belastbaren Evidenzen ist die Aussagekraft von Ergebnissen zu Endpunkten der gesundheitsbezogenen Lebensqualität in den zu bewertenden Studien der Nutzenbewertungsverfahren limitiert. Dies steht im Widerspruch zur beigemessenen Bedeutung der Zielgröße im Prozess der frühen Nutzenbewertung. Zu fordern ist daher die Einführung strenger Gütekriterien für die Verwertung von Validierungsstudien in den Nutzenbewertungsverfahren. Perspektivisch ist die Durchführung adäquater – u. a. an den „State-of-the-Art“-Interventionen sowie den Alters-, Geschlecht- und Erkrankungsschwereverteilungen klinischer Therapiestudien ausgerichteter – Validierungsstudien obligat. Diesbezüglich stehen neben den Sponsoren klinischer Studien im Besonderen jene Forschergruppen in der Verantwortung, die weit verbreitete Erhebungsinstrumente (fort-)entwickeln.

Summary

Health-related quality of life is an increasingly important endpoint in clinical therapy studies. Ideally, validated instruments are used to measure health-related quality of life. Validation studies are conducted to determine the psychometric properties and to define criteria for clinical relevance. The question is to what extent characteristics such as age or disease severity of the study population of therapy studies and validation studies match. Non-matching would limit the validity of psychometric properties considered, as well as the robustness of applied clinical relevance criteria.

The objective of this work is to determine the extent and quality of information on health-related quality of life in the context of early benefit assessment according to § 35a SGB V. The focus is on the most prevalent instruments for measuring health-related quality of life in oncological diseases. Based on key characteristics of studies and patients, particular indicators for the rating of the level of evidence are described. In addition, the work aims to report differences between therapy studies and validation studies on the one hand and between tumor entities on the other hand. The robustness of the evidence is assessed on the basis of differences in the treatment situations of therapy studies and validation studies, as well as information obtained from validation studies on the quality profile of the questionnaires used. Limitations found concerning the applicability of the evidence in the benefit assessment proceedings are contrasted with an additional benefit claimed there and discussed.

The work is based on the completed benefit assessment proceedings for oncological diseases with an evaluation start date up to January 2016 (N = 59). First, excerpts on presence of health-related quality of life endpoints, the design of clinical trials, the status of the conceptual significance and the type and level of information on the outcome measure are collected. The findings are analyzed descriptively (relative frequencies, average comparisons per tumor entity). Furthermore, data on age, gender, disease severity, ethnicity and geographic region are recorded by content analysis. The results are presented descriptively (absolute frequencies, relative deviations) and inferentially (unpaired t-test in the case of interval-scaled operationalized variables, Fisher's exact test in the case of nominal-scaled variables). Moreover, explanations of validity, reliability and responsiveness for the most prevalent outcome measures are extracted from the documents and qualitatively interpreted. Finally, the results of single observation levels are cataloged.

In 80 % of proceedings, health-related quality of life is ascertained and reported to derive additional benefit. Randomized controlled trials are applied in 88 % of cases. These are double-blind in design in 52 % of cases. About every second study examines the effect of intervention versus an active comparator. In therapy studies, endpoints of health-related quality of life are predominantly defined as secondary outcome variables. For the disease-specific profile instruments, the comparison of therapy studies and validation studies showed 54 % complete congruence and 18 % partial congruence of tumor entities. With regard to sociodemographic and clinical characteristics, significant differences are found between the populations of the therapy studies and validation studies: in 83 % for age, 54 % for gender, 83 % for disease severity, 59 % for ethnicity and 71 % for geographic region. In eight frequently cited validation studies regarding benefit assessment proceedings, subscales or components are omitted in 2 cases, limited in 2 other cases and unequally valid in 4 cases. Overviews summarize the included patient characteristics of validation studies, deviations from therapy studies and the author's rating of the confidence level of scales of selected questionnaires.

Given substantial differences between therapy studies and validation studies in key study characteristics as well as occasionally missing or less robust evidence, the explanatory power of results on endpoints of health-related quality of life in the studies to be assessed within the benefit assessment proceedings is limited. This is in contradiction to the declared importance of the outcome measure in the process of early benefit assessment. Therefore, the implementation of stringent quality criteria for the use of validation studies concerning the benefit assessment proceedings is required. In future, it is mandatory to conduct adequate validation studies geared to state of the art interventions along with the age, gender and disease severity distributions of clinical therapy studies. In this regard, besides the sponsors of clinical trials, the research groups that develop widely-used outcome measures have a primary responsibility.

Abkürzungsverzeichnis

<i>AE</i>	<i>Adverse Events</i>
AMNOG	Gesetz zur Neuordnung des Arzneimittelmarktes in der gesetzlichen Krankenversicherung (Arzneimittelmarktneuordnungsgesetz)
AM-NutzenV	Verordnung über die Nutzenbewertung von Arzneimitteln nach § 35a Absatz 1 SGB V für Erstattungsvereinbarungen nach § 130b SGB V (Arzneimittel-Nutzenbewertungsverordnung)
<i>AUC</i>	<i>Area Under the Curve</i>
<i>BAP</i>	<i>Bone Alkaline Phosphatase</i>
<i>BI</i>	<i>(European Organization for Research and Treatment of Cancer Quality of Life Questionnaire - Multiple Myeloma Module 20 [EORTC QLQ-MY20]) Body Image Subscale</i>
<i>CER</i>	<i>Comparative Effectiveness Research</i>
<i>CF</i>	<i>(European Organization for Research and Treatment of Cancer Quality of Life Questionnaire - Core 30 [EORTC QLQ-C30]) Cognitive Functioning Subscale</i>
<i>CONSORT</i>	<i>Consolidated Standards of Reporting Trials</i>
<i>DS</i>	<i>(European Organization for Research and Treatment of Cancer Quality of Life Questionnaire - Multiple Myeloma Module 20 [EORTC QLQ-MY20]) Disease Symptoms Subscale</i>
<i>ECOG PS</i>	<i>Eastern Cooperative Oncology Group Performance Status</i>
<i>EF</i>	<i>(European Organization for Research and Treatment of Cancer Quality of Life Questionnaire - Core 30 [EORTC QLQ-C30]) Emotional Functioning Subscale</i>
<i>EMA</i>	<i>European Medicines Agency</i>
<i>EORTC</i>	<i>European Organization for Research and Treatment of Cancer</i>
<i>EORTC QLQ-C30</i>	<i>European Organization for Research and Treatment of Cancer Quality of Life Questionnaire - Core 30</i>
<i>EORTC QLQ-MY20</i>	<i>European Organization for Research and Treatment of Cancer Quality of Life Questionnaire - Multiple Myeloma Module 20</i>
<i>EQ-5D (VAS)</i>	<i>European Quality of Life Questionnaire 5 Dimensions (Visual Analogue Scale)</i>

<i>EWB</i>	<i>(Functional Assessment of Cancer Therapy [FACT]) Emotional Well Being Subscale</i>
<i>FACIT</i>	<i>Functional Assessment of Chronic Illness Therapy</i>
<i>FACT-Lym</i>	<i>Functional Assessment of Cancer Therapy - Lymphoma</i>
<i>FACT-M</i>	<i>Functional Assessment of Cancer Therapy - Melanoma</i>
<i>FACT-P</i>	<i>Functional Assessment of Cancer Therapy - Prostate</i>
<i>FP</i>	<i>(European Organization for Research and Treatment of Cancer Quality of Life Questionnaire - Multiple Myeloma Module 20 [EORTC QLQ-MY20]) Future Perspective Subscale</i>
<i>FWB</i>	<i>(Functional Assessment of Cancer Therapy [FACT]) Functional Well Being Subscale</i>
<i>G-BA</i>	<i>Gemeinsamer Bundesausschuss</i>
<i>GKV</i>	<i>Gesetzliche Krankenversicherung</i>
<i>HAD</i>	<i>Hospital Anxiety and Depression (Scale)</i>
<i>HAQ</i>	<i>Health Assessment Questionnaire</i>
<i>HTA</i>	<i>Health Technology Assessment</i>
<i>ISOQOL</i>	<i>International Society for Quality of Life Research</i>
<i>IQWiG</i>	<i>Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen</i>
<i>KPS</i>	<i>Karnofsky Performance Status</i>
<i>LymS</i>	<i>(Functional Assessment of Cancer Therapy - Lymphoma [FACT-Lym]) Lymphoma Subscale</i>
<i>MC-SDS</i>	<i>Marlowe-Crowne Social Desirability Scale</i>
<i>MCID</i>	<i>Minimal Clinically Important Difference</i>
<i>MID</i>	<i>Minimal Important Difference</i>
<i>MS</i>	<i>(Functional Assessment of Cancer Therapy – Melanoma [FACT-M]) Melanoma Subscale</i>
<i>NHL</i>	<i>Non-Hodgkin-Lymphom</i>
<i>PCOR</i>	<i>Patient-Centered Outcomes Research</i>
<i>PCS</i>	<i>(Functional Assessment of Cancer Therapy - Prostate [FACT-P]) Prostate Cancer Subscale</i>

<i>PF</i>	<i>(European Organization for Research and Treatment of Cancer Quality of Life Questionnaire - Core 30 [EORTC QLQ-C30]) Physical Functioning Subscale</i>
<i>POMS</i>	<i>Profile of Mood States</i>
<i>PRO</i>	<i>Patient-Reported Outcome</i>
<i>PSA</i>	<i>Prostate-specific Antigen</i>
<i>PSR</i>	<i>Performance Status Rating</i>
<i>pU</i>	<i>pharmazeutischer Unternehmer</i>
<i>PWB</i>	<i>(Functional Assessment of Cancer Therapy [FACT]) Physical Well Being Subscale</i>
<i>QL</i>	<i>(European Organization for Research and Treatment of Cancer Quality of Life Questionnaire - Core 30 [EORTC QLQ-C30]) Global Health Status Subscale</i>
<i>RCT</i>	<i>Randomized Controlled Trial</i>
<i>RF</i>	<i>(European Organization for Research and Treatment of Cancer Quality of Life Questionnaire - Core 30 [EORTC QLQ-C30]) Role Functioning Subscale</i>
<i>rrMM</i>	<i>rezidiviertes oder refraktäres Multiples Myelom</i>
<i>SCLC</i>	<i>Small Cell Lung Cancer</i>
<i>SD</i>	<i>Standard Deviation</i>
<i>SE</i>	<i>(European Organization for Research and Treatment of Cancer Quality of Life Questionnaire - Multiple Myeloma Module 20 [EORTC QLQ-MY20]) Side Effects of Treatment Subscale</i>
<i>SEM</i>	<i>Standard Error of Measurement</i>
<i>SF</i>	<i>(European Organization for Research and Treatment of Cancer Quality of Life Questionnaire - Core 30 [EORTC QLQ-C30]) Social Functioning Subscale</i>
<i>SF-36</i>	<i>Short Form-36 Health Survey Questionnaire</i>
<i>SmPCs</i>	<i>Summaries of Product Characteristics</i>
<i>SPIRIT</i>	<i>Standard Protocol Items: Recommendations for Interventional Trials</i>
<i>SRM</i>	<i>Standardized Response Mean</i>

SSQ	<i>Subjective Significance Questionnaire</i>
SWB	<i>(Functional Assessment of Cancer Therapy [FACT]) Social Well Being Subscale</i>
TOI	<i>(Functional Assessment of Cancer Therapy [FACT]) Trial Outcome Index</i>
TTP	<i>Time to Disease Progression</i>
VAS	Visuelle Analogskala
WHO	<i>World Health Organization</i>

Eigene Abkürzungen:

(Reihung entsprechend dem Vorkommen in der Arbeit)

frühe Nutzenbewertung	Bewertung des Nutzens von Arzneimitteln mit neuen Wirkstoffen gemäß § 35a SGB V.
Nutzenbewertungs- verfahren	Verfahren zur frühen Nutzenbewertung.
Referenzstudie	In einem Nutzenbewertungsverfahren angeführte Studie zum Nachweis der Verwendung eines validierten Instrumentes beziehungsweise eines validierten Responsekriteriums.
zu bewertende Studie	In die frühe Nutzenbewertung eingebrachte Studie.
ECOG PS gw.	Häufigkeitsgewichtete Maßzahl zum <i>ECOG PS</i> . Der Gewichtungsfaktor für die Stufen 0 bis 4 des <i>ECOG PS</i> entspricht dem jeweils relativen Anteil von Patienten einer Studie.
Standardreferenz	In Nutzenbewertungsverfahren wiederkehrend zitierte Studie zum Nachweis der Verwendung eines validierten Instrumentes beziehungsweise eines validierten Responsekriteriums.
Referenz mit AWG- Übereinstimmung	In einem Nutzenbewertungsverfahren zitierte Studie zum Nachweis der Verwendung eines validierten Instrumentes beziehungsweise eines validierten Responsekriteriums mit einem Patientenkollektiv, welches hinsichtlich des onkologischen Anwendungsgebiets ganz oder teilweise jenem der zu bewertenden Studie entspricht.

Abbildungsverzeichnis

Abb. 1: Rechercheoptionen im Informationsarchiv des G-BA	15
Abb. 2: Modularer Aufbau des Dossiers zur Nutzenbewertung gemäß § 35a SGB V	16
Abb. 3: Flussdiagramm zum allgemeinen Aufbau der Arbeit.....	18
Abb. 4: Berücksichtigung der gesundheitsbezogenen Lebensqualität in den Therapiegebieten.....	25
Abb. 5: Onkologische Entitäten mit Endpunktbedeutung für die gesundheitsbezogene Lebensqualität.....	26
Abb. 6: Design klinischer Studien mit Endpunktbedeutung für die gesundheitsbezogene Lebensqualität.....	27
Abb. 7: Konzeptioneller Stellenwert von Endpunkten der gesundheitsbezogenen Lebensqualität.....	28
Abb. 8: Anzahl randomisierter Patienten je zu bewertender Studie.....	29
Abb. 9: Anzahl randomisierter Patienten je Nutzenbewertungsverfahren.....	29
Abb. 10: Behandlungsdauer je zu bewertender Studie	30
Abb. 11: Anzahl Messungen je zu bewertender Studie	31
Abb. 12: Bedeutungsgrad eingesetzter Messinstrumente	33
Abb. 13: Ermittlung der Standardreferenzen; hier: <i>EORTC QLQ-C30</i>	35
Abb. 14: Entsprechung nach onkologischer Entität; hier: Aaronson et al., 1993	36
Abb. 15: Entsprechung nach onkologischer Entität; hier: Osoba et al., 1998.....	36
Abb. 16: Ermittlung der Standardreferenzen; hier: <i>EQ-5D (VAS)</i>	38
Abb. 17: Entsprechung nach onkologischer Entität; hier: Rabin und de Charro, 2001.....	39
Abb. 18: Ermittlung der Standardreferenzen; hier: <i>FACT-P</i>	40
Abb. 19: Entsprechung nach onkologischer Entität; hier: Esper et al., 1997.....	41
Abb. 20: Entsprechung nach onkologischer Entität; hier: Cella et al., 2009	41
Abb. 21: Ermittlung der Standardreferenzen; hier: <i>FACT-M</i>	42
Abb. 22: Entsprechung nach onkologischer Entität; hier: Cormier et al., 2008.....	43
Abb. 23: Ermittlung der Standardreferenzen; hier: <i>EORTC QLQ-MY20</i>	44

Abb. 24: Entsprechung nach onkologischer Entität; hier: Cocks et al., 2007.....	44
Abb. 25: Ermittlung der Standardreferenzen; hier: <i>FACT-Lym</i>	45
Abb. 26: Entsprechung nach onkologischer Entität; hier: Hlubocky et al., 2013	46

Tabellenverzeichnis

Tabelle 1: Beispielrechnung zu Katalogteil 2 ([gemittelte] Abweichungen und Rangfolgen)	24
Tabelle 2: Zusammenschau bezüglich der Standardreferenzen; hier: <i>EORTC QLQ-C30</i>	53
Tabelle 3: Zusammenschau bezüglich der Standardreferenzen; hier: <i>EQ-5D (VAS)</i>	56
Tabelle 4: Zusammenschau bezüglich der Standardreferenzen; hier: <i>FACT-P</i>	59
Tabelle 5: Zusammenschau bezüglich der Standardreferenzen; hier: <i>FACT-M</i>	61
Tabelle 6: Zusammenschau bezüglich der Standardreferenzen; hier: <i>EORTC QLQ-MY20</i>	63
Tabelle 7: Zusammenschau bezüglich der Standardreferenzen; hier: <i>FACT-Lym</i>	65
Tabelle 8: Zusammenschau hinsichtlich der Referenzen mit AWG-Übereinstimmung; hier: <i>EORTC QLQ-C30</i>	73
Tabelle 9: Zusammenschau hinsichtlich der Referenzen mit AWG-Übereinstimmung; hier: <i>EQ-5D (VAS)</i>	79
Tabelle 10: Zusammenschau hinsichtlich der Referenzen mit AWG-Übereinstimmung; hier: <i>FACT-P</i>	80
Tabelle 11: Zusammenschau hinsichtlich der Referenzen mit AWG-Übereinstimmung; hier: <i>FACT-M</i>	81
Tabelle 12: Zusammenschau hinsichtlich der Referenzen mit AWG-Übereinstimmung; hier: <i>EORTC QLQ-MY20</i>	82
Tabelle 13: Zusammenschau hinsichtlich der Referenzen mit AWG-Übereinstimmung; hier: <i>FACT-Lym</i>	83
Tabelle 14: Zusatznutzenbeanspruchung und Bezugnahme in den Dossiers; hier: Aaronson et al., 1993	85
Tabelle 15: Zusatznutzenbeanspruchung und Bezugnahme in den Dossiers; hier: Esper et al., 1997	87
Tabelle 16: Zusatznutzenbeanspruchung und Bezugnahme in den Dossiers; hier: Osoba et al., 1998	90
Tabelle 17: Zusatznutzenbeanspruchung und Bezugnahme in den Dossiers; hier: Cella et al., 2009	91
Tabelle 18: Katalogteil 1 zur Charakteristik von Studienpopulationen.....	94
Tabelle 19: Katalogteil 2 zu (gemittelten) Abweichungen und Rangfolgen.....	99

Inhaltsverzeichnis

1	Einleitung	1
2	Stand der Forschung.....	6
3	Ziele der Arbeit	11
4	Material und Methoden	14
4.1	Ebenen-Betrachtung (I).....	19
4.2	Kongruenz-Betrachtung (II)	19
4.3	Güte-Betrachtung (III).....	22
4.4	Katalogisierung (IV)	23
5	Ergebnisse.....	25
5.1	Verfahrensebene (I.1).....	26
5.2	Studienebene (I.2).....	26
5.3	Endpunktebene (I.3).....	28
5.4	Kongruenz innerhalb der zu bewertenden Studien (II.1)	32
5.5	Kongruenz hinsichtlich der regelhaft angeführten Validierungsstudien (II.2).....	32
5.6	Kongruenz hinsichtlich der Validierungsstudien mit Übereinstimmung in den Anwendungsgebieten (II.3)	66
5.7	Psychometrische Güte und klinische Relevanzkriterien (III.1-III.3).....	84
5.8	Charakteristik der Studienpopulationen bezüglich der Verweise (IV.1)	92
5.9	(Gemittelte) Abweichungen und Rangfolgen bei Referenzstudien (IV.2).....	95
5.10	Kategorisierung nach Belastbarkeit einzelner Skalen von Erhebungsinstrumenten (IV.3)	100
6	Diskussion.....	101
7	Fazit	108
	Anhang	110
	Literatur	175
	Danksagung	

1 Einleitung

„Die vorgelegten Responsekriterien werden als nicht validiert eingeschätzt“ [1].

Mit diesem Fazit leitet das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) eine Wende in der Wertung klinischer Relevanzkriterien in Bezug auf patientenberichtete Endpunkte (*Patient-Reported Outcomes, PROs*) ein – und konsekutiv eine methodische Diskussion zum Stellenwert von Responderanalysen im Rahmen der Bewertung des Nutzens von Arzneimitteln mit neuen Wirkstoffen gemäß § 35a SGB V (im Folgenden: frühe Nutzenbewertung), welche programmatisch in die Novellierung der Allgemeinen Methoden des Instituts eingeht [2].

Die frühe Nutzenbewertung geht zurück auf das mit Wirkung zum 01. Januar 2011 in Kraft getretene Gesetz zur Neuordnung des Arzneimittelmarktes in der gesetzlichen Krankenversicherung (Arzneimittelmarktneuordnungsgesetz, AMNOG) [3]. Ausweislich der Gesetzesbegründung [4] sollten die identifizierten und insbesondere auf kostenintensive Spezialpräparate zurückzuführenden Ausgabenanstiege im Wege nutzenadaptierter Erstattungsbeträge reguliert und in der Folge gemindert werden (Kostendämpfungsziel). Weitere Absichten des Gesetzgebers stellen auf die Verfügbarkeit der besten und wirksamsten Arzneimittel im Krankheitsfall (Qualitätsziel) und die Schaffung verlässlicher Rahmenbedingungen für Innovationen, die Versorgung der Versicherten und die Sicherung von Arbeitsplätzen (Innovationsziel) ab.

Entscheidend für die Bewertung neuer Arzneimittel ist laut Rechtsnorm [5] das vom Gemeinsamen Bundesausschuss (G-BA) attestierte komparative Nutzensmaß. Dabei wird der Zusatznutzen eines Arzneimittels in der Verordnung über die Nutzenbewertung von Arzneimitteln nach § 35a Absatz 1 SGB V für Erstattungsvereinbarungen nach § 130b SGB V (Arzneimittel-Nutzenbewertungsverordnung, AM-NutzenV) definiert als

„der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der Verbesserung des Gesundheitszustands, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder **einer Verbesserung der Lebensqualität** [...] der quantitativ oder qualitativ höher ist als der Nutzen, den die zweckmäßige Vergleichstherapie aufweist [Hervorhebung durch den Verfasser]“ [5].

Im Sinne einer Graduierung werden solche (Zusatz-)Effekte wie folgt eingeordnet: (i) erheblicher Zusatznutzen, (ii) beträchtlicher Zusatznutzen, (iii) geringer Zusatznutzen, (iv) nicht quantifizierbarer Zusatznutzen, (v) kein Zusatznutzen und (vi) geringerer Nutzen [5]. Die Bewertung fußt auf der arzneimittelrechtlichen Zulassung, den im Dossier des pharmazeutischen Unternehmers (pU) aufbereiteten Evidenzen (Daten, die idealerweise auf eine Verbesserung der Beeinflussung relevanter Endpunkte schließen lassen, z. B. eine Verlängerung der Überlebenszeit oder die Vermeidung unerwünschter Ereignisse) sowie etwaigen anwendungsbegleitenden Datenerhebungen nach § 35a Absatz 3b SGB V, welche, nach Festlegung der Verfahrensgrundzüge [6] zur Forderung solcher Erhebungen und Auswertungen, den Datenkörper ergänzen können [5]. Leitend für die Festlegungen zu Ausmaß und der als (i) Anhaltspunkt, (ii) Hinweis oder (iii) Beleg geschätzten Wahrscheinlichkeit [7] eines therapeutischen (Zusatz-)Nutzens in den jeweiligen Anwendungsgebieten sind – gemäß dem allgemein anerkannten Stand der medizinischen Erkenntnisse – die Evidenzstufe und die Qualität der Studien sowie die Validität und die Patientenrelevanz der Endpunkte [5]. Letztere sind den vorrangig zu berücksichtigenden Zielgrößen Mortalität, Morbidität, Lebensqualität und Nebenwirkungen zuzurechnen. Dort, wo quantifizierbar wird der Zusatznutzen zunächst endpunktbezogen bestimmt und schließlich im Wege einer Saldierung semi-quantitativ auf die Zielpopulation (bzw. relevante Teilpopulation) übertragen [7]. In Fällen, in denen ein Mehrwert angenommen wird, die wissenschaftliche Datenlage jedoch keine Quantifizierung erlaubt – im Besonderen gegeben bei Arzneimitteln mit Orphan-Drug-Status –, kann ersatzweise ein „nicht quantifizierbarer“ Zusatznutzen ausgesprochen werden [5].

Neben den tradierten Zielgrößen hat sich die gesundheitsbezogene Lebensqualität als im Wesentlichen komplementäre Bewertungskategorie betreffend den subjektiv erlebten Gesundheitszustand [8] etabliert. Ihr Leitgedanke lässt sich zurückführen auf die im Jahr 1946 verfasste Definition von Gesundheit der *World Health Organization (WHO)*, wonach gilt:

„Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity“ [9].

Gleichwohl die Formulierung von Idealismus geprägt ist, wird die – der Gesundheit immanente – Mehrdimensionalität deutlich. Entsprechendes gilt für das Konzept der gesundheitsbezogenen Lebensqualität. Mit dem sozialen Wohlbefinden, der kognitiven, emotionalen und körperlichen Funktion sowie der Rollenfunktion beschreiben spezifische Indikatoren (Teil-)Aspekte der gesundheitsbezogenen Lebensqualität, die selbst nicht unmittelbar beobachtbar ist [10, 11]. Für ein umfassendes Verständnis von Lebensqualität wird

an anderer Stelle [12] angeführt, dass ausgehend von den biologischen und physiologischen Parametern sich eine Symptomatik und in der Folge ein funktioneller Status unter Einwirkung personengebundener wie umweltbedingter Faktoren entwickelt, welcher bedingt durch die allgemeinen Gesundheitsperzeptionen die individuelle gesundheitsbezogene Lebensqualität schließlich determiniert. Dies veranschaulicht nicht zuletzt die Komplexität und mit ihr die Herausforderungen hinsichtlich einer adäquaten Operationalisierung dieses mehrdimensionalen und zugleich latenten Konstrukts [11]. So ist nachzuvollziehen, dass es bis etwa Mitte der 1990er-Jahre dauerte bis erste Erhebungsinstrumente in der Forschung Anwendung fanden, während bereits in den 1970er-Jahren die konzeptuellen Grundlagen zur gesundheitsbezogenen Lebensqualität geschaffen und diskutiert wurden [13]. In der jüngeren Entwicklungsphase beschäftigt sich die Lebensqualitätsforschung vermehrt mit den Implikationen für die Gesundheitsversorgung – auch und insbesondere im Kontext einer sich verändernden Behandlungsroutine in Zusammenhang mit der Zunahme chronischer Erkrankungen und palliativer Therapiesituationen [13].

Weitgehend unstrittig ist, dass die gesundheitsbezogene Lebensqualität einzig aus der subjektiven Sicht der Betroffenen sinnvoll interpretiert werden kann [11]. Dies deckt sich mit der Begriffsdefinition der *WHO*, wonach die Lebensqualität das Folgende umfasst:

„individuals‘ perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns“ [14].

Hilfsweise kann ein Verfahren zur Fremdbeurteilung (Proxy-Befragungen) dann geeignet sein, sollten vulnerable Gruppen, zum Beispiel kleine Kinder oder kognitiv eingeschränkte Personen, im Fokus stehen [13]. Was die Messung subjektiv empfundener Gesundheitszustände über *PROs* anbelangt, lassen sich zwei Methodiken unterscheiden. Während sich der präferenzbasierte Ansatz der Ökonometrie bedient und von Befragten verlangt, Werte einem definierten Gesundheitszustand zuzuordnen, intendiert der psychometrische Ansatz das Erfragen einer subjektiven Zustandsbeschreibung aus Betroffenenperspektive [15]. Entsprechend ihrer theoretischen Fundierung finden präferenzbasierte Erhebungsinstrumente und aus diesen abgeleitete eindimensionale Indexwerte vorrangig im Bereich gesundheitsökonomischer Evaluationen Verwendung [16]. Mittels psychometrischer Verfahren lassen sich indes Interventionseffekte aus der Patientenperspektive detailliert abbilden, weshalb solche *PROs* zumeist im Kontext klinischer Therapiestudien als Endpunkte verortet sind [15]. Da die Approximation solcher Effekte eine hinreichende Sensitivität und

einen gemeinhin geringen Aggregationsgrad von Ergebnisdaten voraussetzen, dominieren diesbezüglich die an der Mehrdimensionalität der gesundheitsbezogenen Lebensqualität ausgerichteten Profilinstrumente [17]. Abseits der Granularität von Datenerhebung und -aufbereitung ist die Unterteilung von Verfahren zur Erhebung der gesundheitsbezogenen Lebensqualität hinsichtlich ihrer Anwendungsbreite gebräuchlich. Neben standardisierten Fragebögen zur Erfassung spezifischer Aspekte dieser Endpunktkategorie lassen sich generische und krankheitsspezifische Instrumente originär voneinander abgrenzen [18]. Während generische Konzeptionen eine Gültigkeit jenseits spezifischer Interventionen, Erkrankungs- und Therapiesituationen oder demographischer und kultureller Charakteristika postulieren, ist der Einsatz krankheitsspezifischer Instrumente auf Populationen mit definierten Diagnosen oder bestimmten Zustandsbeschreibungen beschränkt [19]. Da Restriktionen dieser Art die (Änderungs-)Sensitivität günstig beeinflussen, dienen krankheitsspezifische Erhebungen vielfach der Ermittlung klinisch relevanter Unterschiede [19].

Um jene in der Literatur zunächst als *Minimal Clinically Important Difference (MCID)* eingeführte [20] und später – einem generalisierten Verständnis folgend – zur *Minimal Important Difference (MID)* fortentwickelte [21] klinische Relevanzschwelle im Sinne einer als subjektiv bedeutsam – im therapeutischen Kontext bestenfalls nutzbringend – wahrgenommenen Änderung im Scorewert eines Fragebogens [20] qualifiziert detektieren zu können, bedarf es geeigneter Studien. In diesen werden die messmethodischen Eigenschaften der Instrumente hinsichtlich ihrer psychometrischen Gütekriterien – gegebenenfalls partiell – erfasst. Die in den Verfahren zur frühen Nutzenbewertung (im Folgenden: Nutzenbewertungsverfahren) zum geforderten Nachweis der Verwendung eines zum einen validierten respektive etablierten Instrumentes [22, 7] und zum anderen eines validierten respektive etablierten Responsekriteriums [7] angeführten Studien (im Folgenden: Referenzstudien) nehmen dabei typischerweise Bezug auf die testtheoretischen Kriterien Validität, Reliabilität und Änderungssensitivität. Substanziell sind in dieser Beziehung Aussagen zur kriterienbezogenen Validität, in dem Bestreben, erwartungskonforme Korrelationen hinsichtlich gleichartiger (konvergente Validität) beziehungsweise verschiedenartiger (diskriminante Validität) Konzepte zu erzielen [23]. Optional eignet sich die Betrachtung verschiedener Subgruppen von Patienten, sofern und soweit Merkmale bekannt sind, die eine geeignete Differenzierung erlauben – beispielsweise über das klinische Symptombild [24]. Mit der Reliabilität existiert zudem ein Maß für zufällige Messfehler und damit ein Indikator für die Reproduzierbarkeit von Messergebnissen [23]. Von Belang sind

insbesondere Ergebnisse zur internen Konsistenz und zur Test-Retest-Reliabilität, mit einem kritischen Wert von 0,7 (Koeffizienten α , r) [24]. Die Änderungssensitivität prüft schließlich die Eignung eines Instrumentes zur Erfassung klinisch bedeutsamer interventionsinduzierter Veränderungen der gesundheitsbezogenen Lebensqualität anhand anker- oder verteilungsbasierter Verfahren [25]. Während Erstere Messwerte und ihre Änderungen mit einem externen Kriterium (z. B. Erkrankungsschwere) in Beziehung setzen und inhaltlich zu interpretieren sind, basieren Letztere auf statistischen Kenngrößen der jeweiligen Stichprobe (z. B. Effektstärke-Indizes) [25]. Ihre Güte wird maßgeblich beeinflusst von der gegebenen Validität und Reliabilität des Messverfahrens sowie dem Bestehen etwaiger Boden- oder Deckeneffekte im Zuge einer Häufung von Messergebnissen an den Extremen einer Skala [23].

Der G-BA stellt mit Verweis auf die Tendenz zur Stratifizierung und Individualisierung der Pharmakotherapie den besonderen Stellenwert der gesundheitsbezogenen Lebensqualität im Kontext onkologischer Erkrankungen fest und plädiert für eine stärkere Berücksichtigung der Zielgröße in den betreffenden Nutzenbewertungsverfahren [26]. Auch die Europäische Arzneimittel-Agentur (*European Medicines Agency, EMA*), deren Standards die Planung klinischer Studien prägen und damit auf die in den Nutzenbewertungsprozess einzubringende Evidenz einwirken, thematisiert im Rahmen eines *Reflection Paper* [27] den Nutzen von PRO-Messungen speziell im Bereich onkologischer Studien und attestiert der gesundheitsbezogenen Lebensqualität eine im Allgemeinen supportive Wertigkeit, was die regulatorische Entscheidungsfindung anbelangt (im Unterschied zur Annahme der Gleichwertigkeit zu Morbiditätsendpunkten in der Logik der frühen Nutzenbewertung [7]). Überdies kann im palliativen Setting die Berücksichtigung der gesundheitsbezogenen Lebensqualität als primärer Endpunkt in konfirmatorischen Studien erwogen werden [28].

Abseits der Bewertungsschemata von *Health Technology Assessment (HTA)*- und regulatorischen Institutionen und in Rekurs auf die Intention des AMNOG stehen Onkologika hinsichtlich ihrer Ausgabenrelevanz im Fokus. Ausweislich des Arzneiverordnungs-Reports 2020 [29] bedingen antineoplastische Arzneimittel jährliche Zahlungen der Gesetzlichen Krankenversicherung (GKV) von zuletzt 8,2 Milliarden Euro (plus 12,1 Prozent im Vergleich zum Vorjahr). Als Treiber dieses Zuwachses werden vorrangig monoklonale Antikörper und die Substanzklasse der Proteinkinaseinhibitoren angeführt. An anderer Stelle [30] findet der ausgeprägte Struktureffekt bei antineoplastischen Mitteln Erwähnung, welcher eine Verschiebung der Absatzmengen innerhalb einer Arzneimittelgruppe hin zu höherpreisigen Arzneimitteln beschreibt und folglich den Umsatzanstieg der Wirkstoffgruppe maßgeblich

bedingt. Im Ergebnis bilden onkologische Arzneimitteltherapien die umsatzstärkste Indikationsgruppe des GKV-Arzneimittelmarktes im Jahr 2019 [29, 30].

Für die frühe Nutzenbewertung gilt, dass Verfahren im Feld der Onkologie nicht nur numerisch führend sind, sondern gleichermaßen in Bezug auf aus G-BA-Sicht verwertbarer Daten zur gesundheitsbezogenen Lebensqualität [31]. Entsprechendes lässt sich mit Blick auf das vorliegend in die Analyse einbezogene Datenset ableiten, welches etwa jedes dritte Nutzenbewertungsverfahren im Zeitraum Januar 2011 bis Januar 2016 in die Betrachtung einbezieht.

2 Stand der Forschung

Wertigkeit der gesundheitsbezogenen Lebensqualität im AMNOG-Kontext

Aktuelle Publikationen belegen den kontextspezifischen Rang der gesundheitsbezogenen Lebensqualität im Bewertungsgefüge. Hinsichtlich der frühen Nutzenbewertung ist zugleich erkennbar, dass selbst bei prinzipieller Bedeutsamkeit des subjektiven Empfindens, Angaben zur gesundheitsbezogenen Lebensqualität selten Entscheidungsrelevanz erlangen. Übereinstimmend werden (erhebungs-)methodische Limitationen angeführt, welche die Eignung und Übertragbarkeit relativieren und damit gegebene Evidenzen in Frage stellen. Dazu im Detail:

Ruof et al. [32] führen an, dass Messinstrumente der gesundheitsbezogenen Lebensqualität in geringerem Maße in den *Summaries of Product Characteristics (SmPCs)*¹ der EMA berichtet werden, verglichen mit Angaben dazu in den Dossiers der pUs im Rahmen der Nutzenbewertungsverfahren – was den differenten Stellenwert dieser Zielgröße offen legt. Begründen lässt sich die marginale Geltung aus regulatorischer Sicht mit jenen den PROs, mithin der gesundheitsbezogenen Lebensqualität, anhaftenden methodischen Unsicherheiten, die beispielhaft im Anhang zur EMA-Guideline betreffend die Bewertung von Arzneimitteln in der Onkologie [28] benannt werden: „*bias, missing data, quality of data, timing of assessments, only single-dimensional PRO measure reporting, [...] lack of post-progression data*“. Daneben beschreiben Schandelmaier et al. [33] in ihrer retrospektiven Untersuchung zu RCTs in der Onkologie eine bemerkenswerte Diskrepanz hinsichtlich der Erwähnung von Erhebungen der gesundheitsbezogenen Lebensqualität in den Studienprotokollen einerseits und den

¹ Im Wesentlichen die Entsprechung der Fachinformation eines Arzneimittels.

korrespondieren Publikationen andererseits. Dieser liegen laut Ausführungen der Autoren in der Mehrzahl fehlende (Prä-)Spezifikationen für den Report zugrunde.

Gleichwohl die prinzipielle Gleichwertigkeit der Zielgrößen schwerwiegende (bzw. schwere) Symptome, schwerwiegende (bzw. schwere) Nebenwirkungen und gesundheitsbezogene Lebensqualität in den Nutzenbewertungsverfahren postuliert wird [7], offenbart die Untersuchung von Bender und Dintsios [34] einen in den seltensten Fällen auf erhobenen Daten zur gesundheitsbezogenen Lebensqualität gründenden Zusatznutzen. Die dort herausgearbeiteten Ablehnungsgründe des IQWiG zu den betreffenden Endpunkten dokumentieren diverse methodische Unzulänglichkeiten auf den Ebenen der jeweilig einbezogenen Studien-/Analysepopulation, den Erhebungsbögen sowie den (mess-)methodischen Näherungen. Schlussfolgernd mahnen die Verfasser insbesondere die konsensuelle und regulatorisch harmonisierte Etablierung (mess-)methodischer Standards an. Dass Endpunkte der gesundheitsbezogenen Lebensqualität im Nutzenbewertungsprozess selten Zusatznutzenstiftend sind und überdies zu einem signifikanten Anteil nicht berichtet werden, legen bereits Lohrberg et al. dar [35]. Die Autoren kritisieren dabei eine fehlende Legaldefinition und fordern bindende methodische Vorgaben zur Minderung der Varianz in Aufbereitung und Auslegung von Zahlen zur (gesundheitsbezogenen) Lebensqualität. Die Arbeit von Blome et al. [36] greift die methodischen Ansprüche und Prämissen von IQWiG und G-BA in der Bewertungshistorie auf und konstatiert mit Blick auf das Validitätsgebot, dass eine gewisse Übereinstimmung hinsichtlich der Patientenpopulationen von Referenzstudien und solchen Studien, welche in die frühe Nutzenbewertung eingebracht werden, verlangt ist – ohne jedoch vorzugeben, unterhalb welcher Grenze vorliegende Evidenzen notwendigerweise verworfen werden. Ferner geht aus der Untersuchung hervor, dass Limitationen von Referenzstudien betreffend die Generalisierbarkeit der Ergebnisse aus Sicht der Bewertungsinstanzen erkannt und benannt werden. Was fehlt, sind konsistent aus solchen Feststellungen abgeleitete Wertungen für die Validität jener Informationen. Auch bleibt fraglich, ob und inwieweit anderweitige Patientencharakteristika differieren und welche Implikationen auf die Übertragbarkeit etwaig angenommen werden. Bullinger et al. [37] heben in diesem Zusammenhang die Obliegenheit der psychometrischen Testung für die betrachtete Untersuchungspopulation im Besonderen beim Einsatz (krankheits-)spezifischer Verfahren hervor.

Ansätze und Herausforderungen einer suffizienten MID-Schätzung

Einschlägige Forschungsarbeiten zu *PROs* setzen sich zunehmend mit der Frage auseinander, welche Anforderungen an die Herleitung belastbarer *MIDs* zu stellen sind, um eine sachgerechte Interpretation gewonnener Daten zur gesundheitsbezogenen Lebensqualität im Sinne einer klinischen Relevanzbeurteilung sicherzustellen [38–40]. Hinsichtlich der Nutzenbewertungsverfahren ist ein Vorzug ankerbasierter (vs. verteilungsbasierter) Methoden unter Verwendung nachweisbar patientenrelevanter Ankervariablen beobachtbar [36]. Wie die Übersichtsarbeit von Jayadevappa et al. [41] zeigt, gründet die Mehrzahl der *MID*-Bestimmungen sinnentsprechend auf einem ankerbasierten Ansatz beziehungsweise einem – in der Forschung präferierten – Konzept der gleichzeitigen Berücksichtigung anker- und verteilungsbasierter Approximationen. Diskutiert wird ferner eine ergänzende Berücksichtigung von aus klinischen Studien gewonnenen Erfahrungen im Wege einer Triangulierung und den Erhalt einer auf diese Weise abgeleiteten *MID* respektive *MID*-Wertespanne [39]. Gegenwärtig ist jedenfalls eine infolge der inkonsistenten Methodiken und divergenter Kontextfaktoren erzeugte Variabilität in Bezug auf die *MID*-Schätzung erkennbar [41]. Auch identifiziert die in der Onkologie verortete Publikation von Ousmen et al. [42] eine hohe numerische Varianz hinsichtlich der rekrutierten Studienteilnehmer und merkt zugleich einen Mangel an Untersuchungen zur Stabilität der *MID* im zeitlichen Verlauf an. Forschergruppen bemühen sich in der Folge um die Inventarisierung und Systematisierung von Schätzern und deren methodischen Fundierungen sowie um die Etablierung geeigneter Bewertungsstandards [43–45]. Beispielgebend in diesem Zusammenhang ist die Arbeit von Devji et al. [46], welche mit Verweis auf den fehlenden methodologischen Standard und die Multiplizität der *MID*-Schätzwerte für ein gegebenes *PRO*-Messverfahren ein Instrument zur Wertung der Zuverlässigkeit solcher Schätzungen offeriert. Hinsichtlich der kardinalen Beurteilungskriterien als günstig führt die Studie an: (i) die Verwendung patientenberichteter Anker (anstelle analoger Einschätzungen der Behandler oder klinischer Messwerte), (ii) für die Patienten einfach begreifliche und in hohem Maße relevante Anker, (iii) eng in Zusammenhang stehende Konzepte von Anker und eigentlichem Erhebungsinstrument, (iv) präzise *MID*-Schätzungen unter Einbezug der Konfidenzintervalle zum Niveau 95 Prozent (alternativ unter Beachtung der Anzahl inkludierter Patienten) sowie (v) ankerbezogene Schwellenwerte oder Gruppenunterschiede, die geeignet sind, geringe und zugleich relevante Abweichungen zu detektieren. Zugleich wird die Validierung von Instrumenten in neuen Studien-/Therapiesituationen im Sinne einer fortlaufenden Re-Evaluation als zwingend angesehen [47].

King formuliert diesbezüglich in Ihrer Kritik:

„While it may be tempting to adopt a single published MID *prima facie*, we need to be mindful that it is just an estimate, as prone to sampling variation as any other, and influenced by the method used, the patient population, the clinical context and perspective“ [48].

Schließlich fasst die Autorin die Charakteristik der gesundheitsbezogenen Lebensqualität und damit das implizite Dilemma der Interpretierbarkeit solcher Daten zusammen: (i) subjektiv, (ii) intra-/interindividuell variabel, (iii) arbiträr (Skalierung), (iv) nicht metrisch (Erfassung) sowie (v) abstrahiert [48].

Entwicklung von (Prüf-)Kriterien für einen standardisierten und validen Gebrauch patientenberichteter Erhebungsmethoden

Im Bewusstsein der diffizilen Beschaffenheit der gesundheitsbezogenen Lebensqualität erarbeiten Forschergruppen Handlungsempfehlungen, welche die robuste und werthaltige Einbettung der Zielgröße beabsichtigen, um schließlich Akzeptanz und substanzielle Wirkung in der klinischen Versorgung zu erlangen. Der Beitrag von Coon und Cook [49] anlässlich des Symposiums *Measuring What Matters* der *International Society for Quality of Life Research (ISOQOL)* [50] stellt etwa Forderungen an Referenzstudien zur Ableitung von aus Responderanalysen gewonnenen Schwellenwerten auf: (i) hinreichende konstruktive Ähnlichkeit, (ii) gleichzeitige Messung und (iii) Untersuchung im Längsschnitt in Bezug auf Anker und Messinstrument sowie Verwendung (iv) selbstberichteter und (v) multipler Ankervariablen. Ferner soll (vi) der Rückgriff auf verteilungsbasierte Analysen einzig der Absicherung reliabler ankerbasierter Schwellenwerte dienen und (vii) die Bestimmung eines Intervalls hinsichtlich der Responderdefinition erwogen werden, um etwaige Unsicherheiten zu illustrieren. Responder meint in diesem Zusammenhang Patienten, die während des Studienverlaufs – je nach Festlegung – eine Verbesserung oder Verschlechterung ihres gesundheitlichen Befindens erfahren, was sich in einem Überschreiten definierter (und die tatsächliche Spürbarkeit aussagender) Punktwertänderungen für einen Fragebogen ausdrückt. Für die Sparte der alltagsnahen Forschung per *Comparative Effectiveness Research (CER)* und *Patient-Centered Outcomes Research (PCOR)* formuliert die *ISOQOL*-Gruppe [51] ein Set von Minimalstandards für den Einbezug von *PROs*, welches umfasst: (i) die Erläuterung des (*PRO*-)Konstruktes, (ii) dessen Anwendbarkeit auf die untersuchte Population, (iii) eine Beschreibung der Operationalisierung, (iv) die Abschätzung der Reliabilität via interner

Konsistenz (Multi-Item-Skalen) respektive Test-Retest-Methode (Ein-Item-Skalen), (v) Informationen zur Validität anhand von Angaben zur Inhaltsvalidität, Konstruktvalidität und Änderungssensitivität, (vi) einfach zu interpretierende Punktwerte (einschließlich Belegmaterial zur numerischen Einordnung), (vii) die Verwendung geeigneter Übersetzungsstrategien, um ein einheitliches Verständnis im internationalen und multikulturellen Kontext sicherzustellen sowie (viii) ein Austarieren von zeitlichen Erfordernissen im Sinne insbesondere der Rücksichtnahme auf die mit dem Ausfüllen eines Fragebogens einhergehenden Lasten für die Patienten. Empfehlungen zum Umgang mit *PROs* in (klinischen) Studienprotokollen fasst ferner die *SPIRIT-PRO Extension Guideline* [52] zusammen – insbesondere im Falle einer primären oder sekundären Endpunkt-Konzeptualisierung. Demgemäß sind unter anderem Ausführungen betreffend die Rationale des *PRO*-Einbezugs sowie einfach zu implementierende Messzeitpunkte gefordert, um die Akzeptanz bei den Beteiligten zu gewährleisten und konsekutiv hohe Rücklaufquoten zu erlangen. Die Verfasser raten weiterhin an, die Verwendung jeweiliger Erhebungsmethoden nebst den messmethodischen Merkmalen zu begründen – idealtypisch mit Verweis auf einschlägige Evidenz in Bezug auf die Zielpopulation. Auf Ebene des Berichtswesens existiert mit der *CONSORT PRO Extension* [53] eine additive Prüfliste für einen adäquaten Report in randomisierten Studien. Die Handlungsempfehlungen schließen im Besonderen ein: (i) die Nutzung eines validierten und reliablen Instrumentes (idealerweise unter Bezugnahme auf die Zielpopulation), (ii) den Bericht von Daten zu Baseline zur Abschätzung von Relevanz und Generalisierbarkeit etwaiger Unterschiede, (iii) das Darlegen sämtlicher Befunde zu Domänen und Erhebungszeitpunkten im Falle multidimensionaler *PROs* und (iv) die Angabe *PRO*-spezifischer Limitationen, die einer Übertragbarkeit von Ergebnissen entgegenstehen können.

Forschungslücken und folgernde Bedarfe

In der Zusammenschau zeigt sich, dass die Beschreibung und Auslegung der Evidenzbasierung zur gesundheitsbezogenen Lebensqualität vielschichtig ist. Folglich bedarf es einer breiten Informationsgrundlage und -synthese, um die Schlüssigkeit diesbezüglicher Aussagen in ihrer Gesamtheit zu erfassen.

Erkennbar ist, dass zur gesundheitsbezogenen Lebensqualität im Kontext des AMNOG per se wenige Forschungsarbeiten vorzufinden sind. Überdies fokussieren vorliegende Publikationen vielfach Teilaspekte im Umgang mit dieser Zielgröße. Es fehlt insoweit an einer kohärenten

Analyse der Evidenzbasierung unter Einbezug verschiedenartiger, den praktischen Implikationen Rechnung tragender, Betrachtungsebenen.

Insbesondere fehlt es an einer systematischen Auswertung zu Unterschieden hinsichtlich der Behandlungskontexte und der Merkmale von Patienten in Studien, welche zum Nachweis der Validität eines jeweiligen Messinstrumentes, mithin der *MID*-Schätzung, in den Nutzenbewertungsverfahren angeführt werden. Eine weitere Insuffizienz betrifft die (strukturierte) Beurteilung von Aussagen zur Zuverlässigkeit einer Erhebung hinsichtlich einzelner, in der Sicherheit solcher Aussagen etwaig differierender, Skalen.

Nicht zuletzt bedarf es einer geordneten Bestandsaufnahme von Verweisen zum Beleg der Validität von Messinstrumenten der gesundheitsbezogenen Lebensqualität, um zum einen die Zulässigkeit des Referenzierens zu beurteilen und zum anderen, um zu weiterer Forschung hinsichtlich passgenauer Validierungsstudien anzuregen.

3 Ziele der Arbeit

Vor dem Hintergrund einer zugleich lückenhaften Erhebung und eingeschränkten Verwertung vorliegender Evidenzen zur gesundheitsbezogenen Lebensqualität [34] soll systematisch untersucht werden, ob und inwieweit in die frühe Nutzenbewertung eingebrachte Endpunktaussagen auf robusten Erkenntnissen gründen. Neben der Erfassung an sich sind hinsichtlich der Frage der Belastbarkeit gewonnener Evidenzen Kenntnisse bezüglich der methodischen Güte zugrundeliegender Studien einerseits sowie des konzeptionellen Stellenwertes und des Informationsumfangs zum Endpunkt gesundheitsbezogene Lebensqualität andererseits entscheidend und werden nachfolgend analysiert. Weiterhin stellt sich die Frage der Übertragbarkeit, mithin der Interpretierbarkeit, vorliegender Ergebnisse zur gesundheitsbezogenen Lebensqualität angesichts etwaiger Unterschiede im Hinblick auf die Patientencharakteristika von in die Nutzenbewertung eingebrachter Studien (im Folgenden: zu bewertende Studien) und Referenzstudie. Daneben soll eruiert werden, in welchem Maße die psychometrischen Gütekriterien einzelner Skalen von Erhebungsinstrumenten der gesundheitsbezogenen Lebensqualität differieren und die Relevanzangaben empirisch stützen.

Die Eingrenzung der zu sichtenden Literatur sowie der zu verarbeitenden Datenmenge auf ein der Art und dem Umfang der Arbeit angepasstes Maß macht eine zeitliche und/oder thematische Fokussierung notwendig. Die Maßgabe sowohl eine möglichst lange Zeitdauer im AMNOG-

Prozess zu überblicken als auch für die frühe Nutzenbewertung gewichtige und für die Versorgung relevante Verfahren zu selektieren, erklärt die gesonderte Auswahl von Nutzenbewertungsverfahren in der Onkologie. Vor dem Hintergrund der Verwendung multipler Erhebungsmethoden bewirkt der therapeutische Fokus zudem eine Verdichtung der Informationen und ermöglicht in der Folge eine fundiertere Darlegung der Evidenzbasierung einzelner Instrumente.

Die vorgenannten Teilaspekte zur Ableitung eines Evidenzprofils werden in der Arbeit wie folgt strukturiert:

- I. Ebenen-Betrachtung
- II. Kongruenz-Betrachtung
- III. Güte-Betrachtung
- IV. Katalogisierung

Reflexion von quantitativer und qualitativer Stellung der gesundheitsbezogenen Lebensqualität in den Nutzenbewertungsverfahren

Die Ebenen-Betrachtung befasst sich zunächst mit den Fragen:

- I.1 in welcher Häufigkeit die gesundheitsbezogene Lebensqualität in den eingeschlossenen Nutzenbewertungsverfahren adressiert wird (vgl. Verfahrensebene),
- I.2 wie die Anzahl und methodische Güte der zugrundeliegenden Studien zu beurteilen ist (vgl. Studienebene),
- I.3 welcher konzeptionelle Stellenwert der gesundheitsbezogenen Lebensqualität als Zielgröße im Rahmen der Studienplanung zuteilwird und wie der Informationsumfang der Endpunkte zu qualifizieren ist. Außerdem: zu welchem Grad statistische Signifikanzen und klinische Relevanzen erreicht werden auf deren Grundlage der pU einen Zusatznutzen tatsächlich beansprucht (vgl. Endpunktebene).

Analyse der Vergleichbarkeit von Behandlungskontexten und zentralen Patientenmerkmalen sowie etwaiger Restriktionen hinsichtlich der Übertragbarkeit von Evidenzen zur Beurteilung der Sicherheit von Ergebnissen zur gesundheitsbezogenen Lebensqualität

Im Verlauf befasst sich die Kongruenz-Betrachtung mit etwaigen Unterschieden betreffend die Patientencharakteristika Alter, Geschlecht, Grad der Erkrankungsschwere, Ethnizität und geographische Region. Dies bezüglich der Perspektiven:

- II.1 Interventionsgruppe versus Kontrollgruppe innerhalb der zu bewertenden Studien,
- II.2 zu bewertende Studie versus regelhaft benannter Referenzstudie innerhalb der Nutzenbewertungsverfahren und
- II.3 zu bewertende Studie versus Referenzstudie mit totaler oder partieller Übereinstimmung hinsichtlich der indizierten Tumorentität innerhalb der Nutzenbewertungsverfahren

– jeweilig gruppiert nach zugelassenem Anwendungsgebiet. Auf Basis der regelhaft benannten Referenzstudien wird im Kontext der Güte-Betrachtung sodann eruiert,

- III.1 inwieweit sich Disparitäten hinsichtlich der Validität und klinischen Relevanzbestimmung einzelner Skalen von Erhebungsinstrumenten der gesundheitsbezogenen Lebensqualität feststellen lassen,
- III.2 ob Skalen solcher Erhebungsinstrumente in der Beurteilung unberücksichtigt sind und
- III.3 zu welchem Grad etwaige Limitationen die Belastbarkeit von Zusatznutzenaussagen zur gesundheitsbezogenen Lebensqualität relativieren.

Dabei werden Referenzstudien, welche vordergründig die psychometrischen Gütekriterien (Validität, Reliabilität und Änderungssensitivität) untersuchen und solche, die einen klinisch relevanten Effekt im Sinne einer *MID* beschreiben, gesondert betrachtet.

Informationen zur Geeignetheit von Validierungsstudien für die untersuchten Zielpopulationen

Im Bestreben einer Ergebnissynthese werden die Befunde zur Kongruenz- und Gütebetrachtung abschließend katalogisiert (vgl. IV.1 bis IV.3) und damit praktisch nutzbar gemacht.

- IV.1 listet je Referenzstudie die zentralen Patientencharakteristika mit der Maßgabe – gegebenenfalls prospektiv – zu überblicken, inwieweit der Verweis auf eine Validierungsstudie Aussagen zur gesundheitsbezogenen Lebensqualität bei der zu bewertenden Studie stützen kann.

- IV.2** legt zunächst dar, ob Nutzenbewertungsverfahren, die den (geclusterten) Anwendungsgebieten zuzuordnen sind, die jeweilige Referenzstudie benennen. Überdies sind die ermittelten relativen Abweichungen zur Referenzstudie je Anwendungsgebiet nebst Rangfolge angegeben, welche auf die, den Referenzstudien zu entnehmenden, Patientencharakteristika Bezug nehmen. Ziel ist es, zu erkennen, inwieweit sich Unterschiede hinsichtlich der verschiedenen Kontexte ergeben, die etwaig eine Übertragbarkeit von Ergebnissen infrage stellen.
- IV.3** ordnet zuletzt die Skalen der Instrumente zur Erhebung der gesundheitsbezogenen Lebensqualität ihrer Aussagekraft entsprechend. Die Informationen basieren auf den Angaben und Erkenntnissen der Referenzstudien im Hinblick auf die Kriterien der psychometrischen Güte und klinischen Relevanz. Damit werden Unterschiede in der Wertigkeit einzelner Skalen in den Blick genommen, um auf solchen Skalen gründende Zusatznutzenaussagen gegebenenfalls zu relativieren.

4 Material und Methoden

Die informationelle Grundlage der Arbeit bildet das Informationsarchiv des G-BA zur Nutzenbewertung nach § 35a SGB V [54]. In die Untersuchung einbezogen sind sämtliche Nutzenbewertungsverfahren betreffend das Therapiegebiet onkologische Erkrankungen mit einem als abgeschlossen vermerkten Status (vgl. Abb. 1) und einem dokumentierten Bewertungsbeginn bis einschließlich Januar 2016. Das inkludierte Material umfasst damit 59 Verfahren zu antitumoralen Wirkstoffen mit einer letztmaligen Beschlussfassung am 16.06.2016 (vgl. Anhang A).

Verfahren der Nutzenbewertung nach § 35a SGB V

In der Übersicht finden Sie Informationen zu den laufenden und abgeschlossenen [Nutzenbewertungsverfahren nach § 35a SGB V](#). Veröffentlicht werden hier alle für das Verfahren relevanten Unterlagen: Dossier des pharmazeutischen Unternehmers, Informationen zur zweckmäßigen Vergleichstherapie, die Nutzenbewertung, das Wortprotokoll zur mündlichen Anhörung sowie Beschluss nebst Begründung.

Die maschinenlesbaren Fassungen der Beschlüsse (XML-Dateien) werden auf folgender Seite abrufbar sein: www.g-ba.de/ais

Wirkstoff A-Z
Bitte wählen

Therapiegebiet
onkologische Erkrankungen

Orphan Drug
Bitte wählen

Verfahrensstand
Verfahren abgeschlossen

Nur Verfahren mit englischer Übersetzung anzeigen/Show procedures with English translation only

Filter anwenden oder zurücksetzen

Hervorhebung durch den Verfasser.

Abb. 1: Rechercheoptionen im Informationsarchiv des G-BA

(Quelle: <https://www.g-ba.de/bewertungsverfahren/nutzenbewertung/> [Bildschirmausschnitt]. Letzter Zugriff: 14.02.2021 [54]).

In die systematische Bearbeitung von Literatur gehen jeweilig (i) der (Verfahrens-)Steckbrief, (ii) die Module 1 und 4 eines Dossiers, (iii) die Nutzenbewertung (zzgl. etwaiger Addenda), (iv) der Beschlusstext sowie (v) die tragenden Gründe zum Beschluss ein. Insoweit sind, was das Dossier anbelangt, neben den zusammenfassenden Aussagen (vgl. Modul 1) insbesondere die Beschreibungen des pU zu Studiendesign und Studienpopulation von in die Bewertung eingeschlossenen Studien, die Ergebnisse zu Endpunkten der gesundheitsbezogenen Lebensqualität und die Angaben zum medizinischen (Zusatz-)Nutzen einschließlich deren Wahrscheinlichkeit und Ausmaß im Vergleich zu einer zweckmäßigen Vergleichstherapie im zu bewertenden Anwendungsgebiet (vgl. Modul 4) relevant. Eine Übersicht zur Gliederung des Dossiers findet sich in der Dokumentvorlage des G-BA und ist nachstehend wiedergegeben (vgl. Abb. 2).

Modul 1

- Administrative Informationen (pU, Zulassungsinhaber)
- **Zusammenfassende Aussagen zu den Modulen 2, 3 und 4 – ggf. je zu bewertendem AWG (u. a. Feststellungen zu den zugelassenen AWGs des zu bewertenden Arzneimittels und zum medizinischen [Zusatz-]Nutzen je AWG)**

Modul 2

- Allgemeine Angaben über das zu bewertende Arzneimittel
- Beschreibung der AWGs, für die das zu bewertende Arzneimittel zugelassen wurde

Modul 3 (ggf. je zu bewertendem AWG)

- Bestimmung der zVT, der Anzahl an Patienten mit therapeutisch bedeutsamem Zusatznutzen und der Therapiekosten für die GKV
- Beschreibung der Anforderungen an eine qualitätsgesicherte Anwendung

Modul 4 (ggf. je zu bewertendem AWG)

- Angaben zur Methodik der im Dossier präsentierten Bewertung des medizinischen (Zusatz-)Nutzens
- **Ergebnisse zum medizinischen (Zusatz-)Nutzen (u. a. Charakterisierung der eingeschlossenen Studien, der Interventionen und der Studienpopulationen sowie spezifische Berichte zu Endpunkten der gesundheitsbezogenen Lebensqualität mit Angaben zur Operationalisierung, zur Bewertung des Verzerrungspotenzials und zu den Ergebnissen)**
- **Abschließende Bewertung der Unterlagen zum Nachweis des Zusatznutzens, einschließlich der Angabe von Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen angenommen wird (u. a. Beschreibung des Zusatznutzens einschließlich dessen Wahrscheinlichkeit und Ausmaß)**
- Ergänzende Informationen zur Begründung der vorgelegten Unterlagen
- **Darlegung der Methodik der eingeschlossenen Studien (u. a. Ausführungen zum Studiendesign, zur Studienorganisation und zum Ort der Studiendurchführung, zu den Zielkriterien [einschließlich der Validierung von Erhebungsinstrumenten] und zum Zeitraum der Studienaufnahme der Probanden/Patienten sowie der Nachbeobachtung)**

Modul 5

- Supplementäre Informationen (u. a. Dateien zur Dokumentation der Informationsbeschaffung, Volltexte der zitierten Quellen, Studienberichte, zulassungsrelevante Unterlagen und der Bewertungsbericht der Zulassungsbehörde)

AWG: Anwendungsgebiet, GKV: Gesetzliche Krankenversicherung, pU: pharmazeutischer Unternehmer, zVT: zweckmäßige Vergleichstherapie.

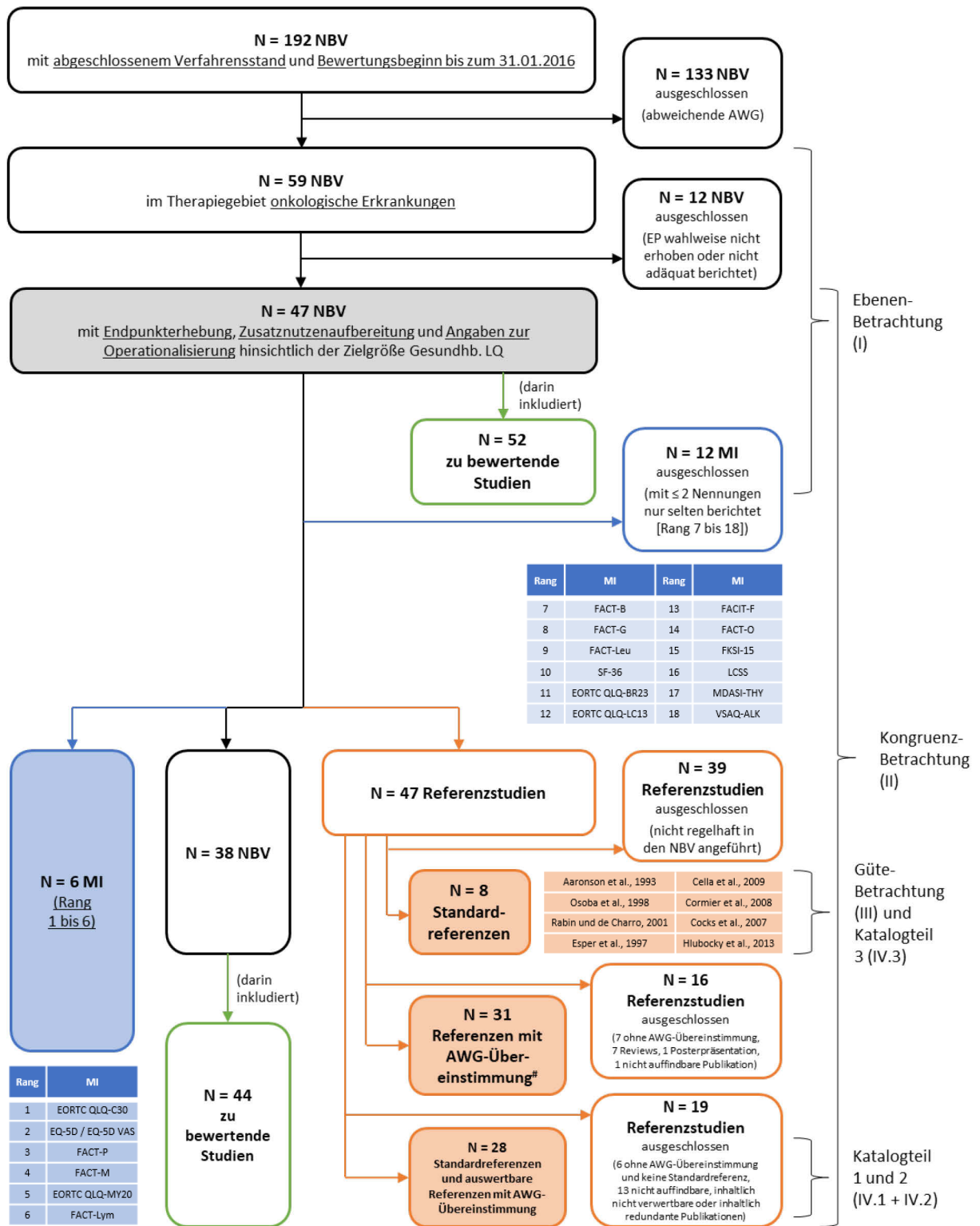
Anmerkung: Modul 5 unterliegt der Geheimhaltungspflicht und wird folglich nicht im Internet veröffentlicht.

Abb. 2: Modularer Aufbau des Dossiers zur Nutzenbewertung gemäß § 35a SGB V

(Quelle: Eigene Darstellung unter Bezugnahme auf die Anlagen zum 5. Kapitel der Verfahrensordnung des Gemeinsamen Bundesausschusses [55–59]).

Grundlage der Untersuchung sind weiterhin die in den Nutzenbewertungsverfahren zitierten Studien zur Beschreibung der Gütekriterien (Validität, Reliabilität, Änderungssensitivität bzw. abgeleiteter Responsekriterien) eingesetzter Instrumente zur Erhebung der gesundheitsbezogenen Lebensqualität. Ein Flussdiagramm dient der Veranschaulichung von Art und Umfang (i) der gesichteten Quellen, (ii) der Ein- und Ausschlusskriterien, (iii) der auswertungsgegenständlichen Literatur und (iv) seiner thematischen Verortung (vgl. Abb. 3).

Die Dokumente werden zunächst inhaltsanalytisch im Volltext entsprechend der nachstehenden Forschungsfragen gesichtet. Anschließend erfolgt die Sammlung von Auszügen im Wege inhaltlicher Kategorisierungen über Registerblätter des Tabellenkalkulationsprogramms Microsoft Excel, Version 2016 (Excel 2016). Dabei werden die Einheiten der Untersuchung (z. B. Nutzenbewertungsverfahren, zu bewertende Studien, Validierungsstudien) zeilenweise und die Merkmale der Untersuchungseinheiten (z. B. Anwendungsgebiet, Erhebung der Zielgröße gesundheitsbezogene Lebensqualität, Patientencharakteristika) spaltenweise erfasst. Erforderlichenfalls werden einzelnen Spalten Festlegungen zur Datenextraktion zugeordnet, um ein einheitliches Sammeln von Kennzahlen zu erreichen (am Beispiel Lageparameter: Median, ersatzweise Mittelwert) oder Abgrenzungen zu kennzeichnen (z. B. kein Einschluss von Fremdstudien, kein erneutes Erfassen bereits registrierter zu bewertender Studien). Weiterhin werden Informationen verdichtet beziehungsweise auf Kennzahlen heruntergebrochen, um eine spätere statistische Auswertung möglich zu machen. Anhang B zeigt eine Bewertungsmatrix zur Kongruenz-Betrachtung am Beispiel des zuletzt in das Datenset eingeschlossenen Nutzenbewertungsverfahrens zu Crizotinib (neues Anwendungsgebiet) [60] sowie eine weitere zur Güte-Betrachtung am Beispiel der dort vorzufindenden Validierungsstudie Osoba et al., 1998 [61], die sich aus den jeweils themenrelevanten Informationen zusammensetzen.



AWG: Anwendungsgebiet, EP: Endpunkt der gesundheitsbezogenen Lebensqualität, Gesundhb. LQ: Gesundheitsbezogene Lebensqualität, MI: Messinstrument der gesundheitsbezogenen Lebensqualität, NBV: Nutzenbewertungsverfahren, Referenzen mit AWG-Übereinstimmung: in Nutzenbewertungsverfahren zitierte und betreffend der Tumorentität wenigstens partiell übereinstimmende Referenzstudien, Standardreferenzen: in Nutzenbewertungsverfahren wiederkehrend zitierte Referenzstudien.

Ergebnisse zu den erfassten (7) Standardreferenzen werden infolge der separaten Betrachtung und Auswertung solcher Studien nicht (erneut) berichtet.

Abb. 3: Flussdiagramm zum allgemeinen Aufbau der Arbeit
(Quelle: Eigene Darstellung).

4.1 Ebenen-Betrachtung (I)

Um den Geltungsgrad von Endpunkten der gesundheitsbezogenen Lebensqualität bei onkologischen Erkrankungen zu beschreiben, wird Modul 4 von Dossiers zu den insgesamt 59 Nutzenbewertungsverfahren in die Auswertung einbezogen. Im Rahmen einer quantitativen Inhaltsanalyse werden in einer Bewertungsmatrix thematische Kategorisierungen zu Verfahren mit (i) Endpunkterhebung, (ii) Zusatznutzenaufbereitung und (iii) Angaben zur Operationalisierung gebildet und bezüglich der relativen Häufigkeiten kumulativ ausgewertet.

Zur Charakteristik der zu bewertenden Studien wird anhand von Modul 4 der jeweiligen Dossiers aus den Nutzenbewertungen zu den 47 in die Analyse eingeschlossenen Verfahren mit vorliegenden Endpunktangaben ausgeführt. Die thematischen Kategorisierungen fokussieren (i) die Methodik zum Vergleich von Interventionen, (ii) das Studiendesign und (iii) die klinische Studienphase. Die Ergebnisse werden in Form relativer Häufigkeiten präsentiert.

Die quantitative Inhaltsanalyse auf Endpunktebene trifft Aussagen zum konzeptionellen und tatsächlichen Stellenwert sowie zum Informationsumfang der Zielgröße gesundheitsbezogene Lebensqualität und schließt die Module 1 und 4 eines Dossiers, die Nutzenbewertung und die tragenden Gründe zum Beschluss hinsichtlich derselben 47 Nutzenbewertungsverfahren ein. Berichtet werden relative Häufigkeiten und Durchschnittsvergleiche je Tumorentität, fußend auf den thematischen Kategorisierungen (i) der Endpunkt-Konzeptualisierung, (ii) der Anzahl randomisierter Patienten je Nutzenbewertungsverfahren/zu bewertender Studie, (iii) der Behandlungsdauer je zu bewertender Studie, (iv) der Anzahl Messungen bezogen auf die (mediane) Behandlungsdauer je zu bewertender Studie sowie (v) der statistischen Signifikanzen, (vi) der klinischen Relevanzen und (vii) der Beanspruchung eines Zusatznutzens aus pU-Perspektive.

4.2 Kongruenz-Betrachtung (II)

Die Feststellung von Unterschieden hinsichtlich der Merkmale von Patienten zwischen Interventions- und Kontrollgruppe der zu bewertenden Studien basiert auf Modul 4 eines Dossiers und den Nutzenbewertungen zu wiederum 47 Verfahren. Im Wege einer quantitativen Inhaltsanalyse werden auf Basis der folgenden thematischen Kategorisierungen Durchschnittsvergleiche je Tumorentität präsentiert: (i) Alter, (ii) Anteil Frauen, (iii) häufigkeitsgewichtete Maßzahl zum *Eastern Cooperative Oncology Group Performance Status (ECOG PS)* (im Folgenden: *ECOG PS* gw.) und Anteil eines *ECOG PS* von 0 oder 1, (iv) Anteil Kaukasier/weiß und (v) Anteil (West-)Europa.

Unterschiede betreffend den Behandlungskontext zwischen den zu bewertenden Studien einerseits und den wiederkehrend zitierten Referenzstudien in solchen Verfahren (im Folgenden: Standardreferenzen) andererseits werden auf Grundlage von Modul 4 eines Dossiers und den Nutzenbewertungen quantitativ untersucht. Berücksichtigt sind insgesamt 38 Nutzenbewertungsverfahren, welche auf die führenden 6 Instrumente zur Erhebung der gesundheitsbezogenen Lebensqualität – im Einzelnen: *EORTC QLQ-C30*, *EQ-5D (VAS)*, *FACT-P*, *FACT-M*, *EORTC QLQ-MY20* und *FACT-Lym* – Bezug nehmen. Diese Einschränkung ist dadurch begründet, dass die weiteren 12 Messinstrumente nicht häufiger als 2 Mal in den Nutzenbewertungsverfahren benannt sind und folglich einer belastbaren Statistik entgegenstehen. Durch den Ausschluss der weniger exponierten Instrumente soll zusätzlich eine Multiplizität von Ergebnissen vermieden und die Anschaulichkeit gestärkt werden. Daneben gehen in die Analyse die Volltexte zu 8 erfassten Standardreferenzen (vgl. Anhang C) ein. Die Auswahl der Verweise basiert auf der Zählung sämtlicher Referenzstudien in den Nutzenbewertungsverfahren und berücksichtigt die nach Anzahl und Einflussnahme auf verschiedene Tumorentitäten dominierenden Publikationen. Kennzeichen solcher Studien ist insoweit die Interpretation als etablierter Standard für den Nachweis der Validität spezifischer Messinstrumente. In die Bewertungsmatrizen werden insbesondere folgende thematische Kategorisierungen eingefügt: (i) die Bezeichnung des Messinstrumentes, (ii) eine bibliographische Angabe nebst Kurztitel der angeführten Referenzstudie, (iii) die Zuordnung entsprechend der vordergründigen Zweckbestimmung (wahlweise Nachweis von Validität oder klinischer Relevanz), (iv) der Grad an Übereinstimmung in Bezug auf die Studienpopulation sowie Abweichungen zur Referenzstudie nebst statistischer Signifikanzaussage betreffend die Merkmale (v) Alter, (vi) Anteil Frauen, (vii) *ECOG PS* gw. und *ECOG PS* von 0 oder 1, (viii) Anteil Kaukasier/weiß und (ix) Anteil (West-)Europa. Die Ergebnisse werden deskriptiv mittels Angabe zu absoluten Häufigkeiten und relativen Abweichungen und in Teilen inferentiell berichtet. Letzteres schließt den ungepaarten t-Test im Falle der intervallskalierten beziehungsweise intervallskaliert operationalisierten Variablen (Alter, Grad der Erkrankungsschwere) und den exakten Test nach Fisher im Falle der nominalskalierten Variablen (Geschlecht, Ethnizität, geographische Region) bei einem gewählten Signifikanzniveau von 5 Prozent ein. Entgegen dem Verständnis der Entwickler und Anwender als ordinalskalierte Indizes [62], werden der *ECOG PS* und – ersatzweise – der *Karnofsky Performance Status (KPS)* vorliegend intervallskaliert interpretiert. Ziel ist es, funktionale Einschränkungen von Patientenkollektiven in den betrachteten Studien damit differenzierter abzubilden. Ein Umrechnungsmodus zwischen den Skalen *ECOG PS* und *KPS* ist beschrieben

[63] und findet ausnahmsweise Anwendung, sofern Angaben zum *KPS*, nicht aber solche zum *ECOG PS*, in den Unterlagen berichtet sind. Für die Kalkulation der p-Werte wird mit *GraphPad QuickCalcs* (<https://www.graphpad.com/quickcalcs/>) von einer kommerziell verfügbaren, browserbasierten Statistiksoftware Gebrauch gemacht. Im Zuge der Anwendung des t-Tests ermöglicht diese eine Berechnung von p-Werten anhand von Angaben (je Vergleichsgruppe) zu Mittelwert, Standardabweichung und Umfang der Stichprobe. Ein solches Dateneingabeformat ist nötig, da in Nutzenbewertungsverfahren und in diesen zitierten Publikationen keine Rohdaten – im Sinne des Berichts von Werten einzelner Patienten – verfügbar sind. Sofern Angaben zu Interventions- und Kontrollgruppen hinsichtlich der zu bewertenden Studien getrennt vorliegen, werden hilfsweise die gepoolte Standardabweichung

*(Berechnungsformel, Excel 2016: =WURZEL(((n_I-1)*s_I²+(n_K-1)*s_K²)/(n_I+n_K-2))*
mit n = Zahl der randomisierten Patienten je Therapiearm, s = Standardabweichung eines Merkmals [z. B. Alter] je Therapiearm)

und das gewichtete arithmetische Mittel

(Berechnungsformel, Excel 2016: =MW_I(n_I/SUMME(n_I;n_K))+*
MW_K(n_K/SUMME(n_I;n_K)) mit MW = arithmetisches Mittel eines Merkmals*
[z. B. Alter] je Therapiearm, n = Zahl der randomisierten Patienten je Therapiearm)

berechnet.

Für die weitere Gegenüberstellung der Charakteristik von Studienpopulationen werden die zu bewertenden Studien und die in den betreffenden Nutzenbewertungsverfahren zitierten Referenzstudien, welche sich ganz oder teilweise hinsichtlich der onkologischen Erkrankung gleichen (im Folgenden: Referenzen mit AWG-Übereinstimmung), einbezogen. Die Inhaltsanalysen fußen zum einen auf Modul 4 eines Dossiers und den Nutzenbewertungen zu wiederum 38 Verfahren, für die eine Bezugnahme auf wenigstens eines der führenden 6 Messinstrumente dokumentiert ist. Zum anderen werden die insgesamt 47 Referenzstudien gesichtet und nach betrachteter Tumorerkrankung geordnet. In der Folge gehen 31 Referenzen mit AWG-Übereinstimmung² (vgl. Anhang C) in die Analyse ein. Dabei werden die Referenzen mit AWG-Übereinstimmung qualitativ und die weiteren thematischen Kategorisierungen analog obiger Betrachtung quantitativ untersucht.

² Aufgrund der generischen Konzeption des *EQ-5D* werden für diesen Fragebogen auch Referenzstudien in die Betrachtung einbezogen, die sich in Bezug auf die onkologische Erkrankung von den zu bewertenden Studien unterscheiden.

4.3 Güte-Betrachtung (III)

Um Disparitäten hinsichtlich der psychometrischen Gütestruktur einzelner Skalen von Erhebungsinstrumenten zu erkennen und deren Auswirkungen auf die Belastbarkeit von Zusatznutzenaussagen zu beurteilen, werden die Inhalte der 8 Standardreferenzen qualitativ ausgewertet. Die ableitbaren Interpretationen zum Grad der Belastbarkeit werden den Ergebnisprofilen zu Nutzenbewertungsverfahren, die einen Zusatznutzen beanspruchen, gegenübergestellt. In den Bewertungsmatrizen werden als thematische Kategorisierungen aufgenommen: (i) das Studiendesign, (ii) die Anzahl einbezogener Patienten und (iii) Aussagen zu den Messzeitpunkten. Ferner: (iv) die psychometrischen Gütekriterien oder klinischen Relevanzkriterien und deren Bestimmungsmethoden, (v) gegebenenfalls Angaben zum Interpretationsansatz (z. B. ankerbasiert) und dem Referenzmodell (z. B. *Subjective Significance Questionnaire, SSQ*), (vi) die Maßzahlen inklusive etwaiger Anforderungen an diese, (vii) eine Übersicht der Ergebnisse und (viii) Auflistungen zu Art und Umfang der Bezugnahme in den Dossiers.

Exkurs: Referenzmodelle dienen insbesondere der Absicherung einer klinischen Relevanz in Bezug auf die Interpretation von Änderungen im Antwortverhalten bei Fragebögen zur Erhebung der gesundheitsbezogenen Lebensqualität. Dabei finden zum einen vom eigentlichen Messinstrument unabhängige Erhebungsmethoden Anwendung, welche darauf ausgelegt sind, den vom Patienten empfundenen Grad an Veränderung – beispielsweise in physischen, emotionalen, sozialen und allgemeinen Belangen (vgl. *SSQ* [4 Items, 7-stufige Likert-Skala] [64]) – zu erfassen. Dieser wiederum wird den Punktwertunterschieden der korrespondierenden Domänen eines zu bewertenden Fragebogens gegenübergestellt. Zum anderen lassen sich externe Kriterien (z. B. Laborparameter, Aktivitätsstatus) heranziehen, um ein Patientenkollektiv anhand zu definierender Schwellenwerte in disjunkte Gruppen zu unterteilen. Für die einzelnen Gruppen werden sodann die Unterschiede in den Scores eines Fragebogens im Querschnitt betrachtet. Schließlich sind Längsschnittuntersuchungen zur Änderungssensitivität nutzbar, um zu prüfen, ob Veränderungen einer Messgröße (z. B. Morbiditätsendpunkte, Nebenwirkungen) mit Veränderungen im Scorewert eines Fragebogens einhergehen. Auch sind in dieser Beziehung verteilungsbasierte Verfahren etabliert, welche die Veränderung im Antwortverhalten mit einem Maß der Variabilität verbinden – dabei jedoch den Bezug zur tatsächlichen Spürbarkeit einer Veränderung außer Acht lassen [15].

4.4 Katalogisierung (IV)

In **Teil 1** werden zur Beschreibung von Behandlungskontext und Charakteristik der Studienpopulationen folgende Spezifika zusammengetragen: (i) Studienpopulation/-indikation, (ii) Alter, (iii) Geschlecht, (iv) Grad der Erkrankungsschwere, (v) Ethnizität und (vi) geographische Region. Die Liste umfasst 28 Referenzstudien und schließt sowohl die Standardreferenzen als auch die auswertbaren Referenzen mit AWG-Übereinstimmung ein.

Teil 2 enthält eine Übersicht zu Abweichungen und Rangfolgen auf Ebene der onkologischen Entitäten – insofern Referenzstudien dort erwähnt sind. Um zu prüfen, welche Studie in welchem Anwendungsgebiet als Referenz benannt wird, werden auf Basis der Informationen aus Modul 4 eines Dossiers Verknüpfungen gebildet aus dem Nutzenbewertungsverfahren, dem Anwendungsgebiet in welchem das Verfahren verortet ist und den Referenzstudien, welche in den Nutzenbewertungsverfahren zitiert werden. Zur Berechnung einer relativen Abweichung wird je Anwendungsgebiet die Patientencharakteristik der zu bewertenden Studien in Beziehung gesetzt zu den verfügbaren Angaben der Referenzstudien. Die berücksichtigten und zu gleichen Teilen gewichteten Kriterien sind: (i) Studienpopulation/-indikation, (ii) Alter, (iii) Geschlecht, (iv) Grad der Erkrankungsschwere, (v) Ethnizität und (vi) geographische Region. Zur Sicherstellung aussagekräftiger Angaben ist die Nennung von mindestens 3 jener Kriterien je Referenzstudie Voraussetzung für die Berechnung einer (gemittelten) Abweichung. Tabelle 1 ist die Berechnung der relativen Abweichung zur Referenz anhand der Studie von Aaronson et al. [64] für das Anwendungsgebiet Leukämie, Lymphom beispielhaft zu entnehmen.

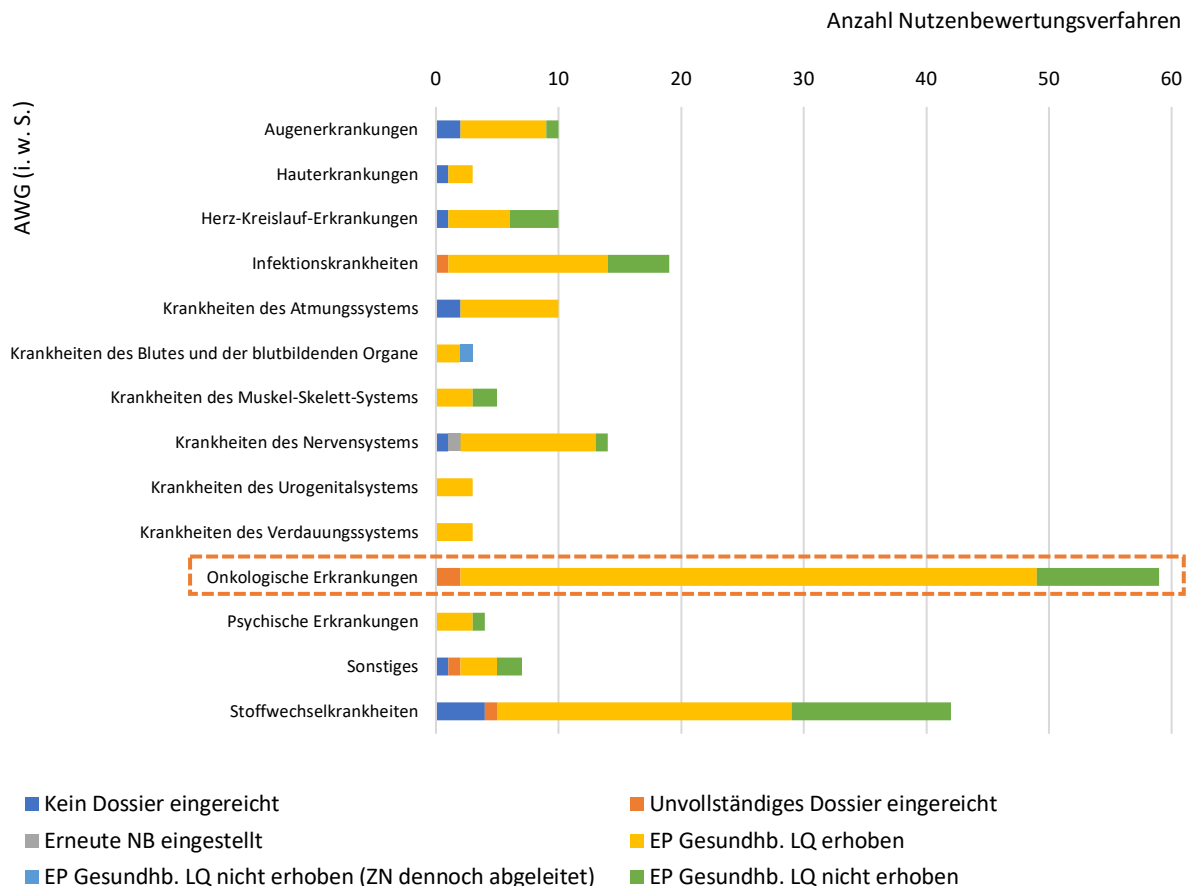
Aus den Ergebnissen zur Güte-Betrachtung erlangte Informationen betreffend die Belastbarkeit einzelner Skalen von Erhebungsinstrumenten werden in **Teil 3** zusammengetragen. Inkludiert sind die eingangs bestimmten führenden 6 Instrumente (vgl. *EORTC QLQ-C30*, *EQ-5D (VAS)*, *FACT-P*, *FACT-M*, *EORTC QLQ-MY20* und *FACT-Lym*). Zur Veranschaulichung werden je Skala farbliche Codierungen im Sinne (i) belastbar, (ii) eingeschränkt oder ungleich belastbar und (iii) fehlend/unberücksichtigt vergeben, um Unterschiede mit Blick auf die Aussagekraft zu illustrieren.

Kriterium	Referenzstudie (Aaronson et al., 1993 [64])	AWG-Cluster mit Verweis auf die Referenzstudie ([I]: Interventionsgruppe, [K]: Kontrollgruppe, zb. St.: zu bewertende Studie)		Berechnungs- formel, Excel 2016 (ABS: Absolutwert, MW: Mittelwert)	Abweichung zur Referenzstudie
Studienpopulation /-indikation	nicht-resezierbares Lungenkarzinom	Leukämie, Lymphom			100 %
Alter	63 Jahre	zb. St. (lfd. Nr.) 1 2 3 4 5 6 7	73 J. [I ₁]; 73 J. [K ₁] 52,8 J. [I ₂] 74 J. [I ₃]; 73 J. [K ₃] 71 J. [I ₄]; 70 J. [K ₄] 61 J. [I ₅] 67 J. [I ₆]; 67 J. [K ₆] 68 J. [I ₇]	=ABS((MW (MW([I ₁];[I ₂];[I ₃]; [I ₄];[I ₅];[I ₆];[I ₇]); MW([K ₁];[K ₃];[K ₄]; [K ₆]))-63)/63)	9,1 %
Anteil Frauen	24 %	zb. St. (lfd. Nr.) 1 2 3 4 5 6 7	43,4 % [I ₁]; 37,9 % [K ₁] 47,2 % [I ₂] 39 % [I ₃]; 38 % [K ₃] 30,9 % [I ₄]; 38,2 % [K ₄] 45,8 % [I ₅] 33,8 % [I ₆]; 30,1 % [K ₆] 23 % [I ₇]	=ABS((MW (MW([I ₁];[I ₂];[I ₃]; [I ₄];[I ₅];[I ₆];[I ₇]); MW([K ₁];[K ₃];[K ₄]; [K ₆]))-24)/100)	12,8 % (absolut)
ECOG PS gw. (häufigkeitsgewichtete Maßzahl zum <i>Eastern Cooperative Oncology Group Performance Status</i>)	1,143	zb. St. (lfd. Nr.) 1 2 3 4 5 6 7	1,075 [I ₁]; 1,073 [K ₁] 0,352 [I ₂] nicht berichtet 0,755 [I ₄]; 0,725 [K ₄] 0,291 [I ₅] 0,595 [I ₆]; 0,592 [K ₆] 0,630 [I ₇]	=ABS((MW (MW([I ₁];[I ₂];[I ₄]; [I ₅];[I ₆];[I ₇]);MW ([K ₁];[K ₄];[K ₆]))- 1,143)/1,143)	38,2 %
Anteil Kaukasier/weiß	nicht berichtet	zb. St. (lfd. Nr.) 1 2 3 4 5 6 7	86,4 % [I ₁]; 87,7 % [K ₁] 65,3 % [I ₂] 95,0 % [I ₃]; 95,0 [K ₃] 90,9 % [I ₄]; 89,1 % [K ₄] 88,9 % [I ₅] 89,2 % [I ₆]; 90,3 % [K ₆] nicht berichtet	-	- (nicht berücksichtigt)
Anteil (West-)Europa	69,5 %	zb. St. (lfd. Nr.) 1 2 3 4 5 6 7	66,1 % [I ₁]; 60,5 % [K ₁] nicht berichtet nicht berichtet 27,3 % [I ₄]; 24,1 % [K ₄] 38,9 % [I ₅] 45,1 % [I ₆]; 41,8 % [K ₆] nicht berichtet	=ABS((MW (MW([I ₁];[I ₄];[I ₅]; [I ₆]);MW([K ₁];[K ₄]; [K ₆]))-69,5)/100)	26,3 % (absolut)
Ergebnis:				=MW(1;0,091; 0,128;0,382;0,263)	37,3 %

Tabelle 1: Beispielrechnung zu Katalogteil 2 ([gemittelte] Abweichungen und Rangfolgen)
(Quelle: Eigene Darstellung).

5 Ergebnisse

Die Vollerhebung abgeschlossener Nutzenbewertungsverfahren mit einem Bewertungsbeginn bis zum 31.01.2016 zeigt, dass mit 59 Verfahren die Mehrzahl im Bereich der onkologischen Erkrankungen verortet ist (vgl. Abb. 4).



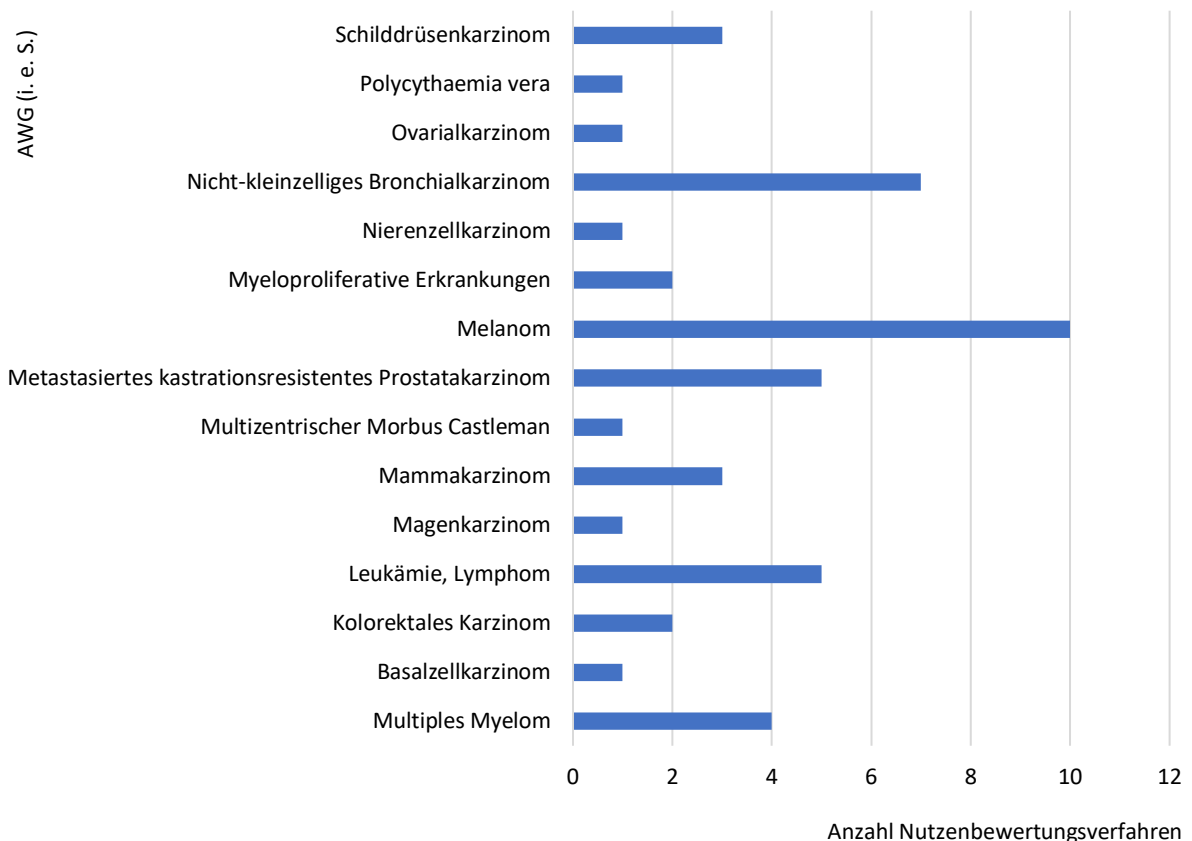
AWG: Anwendungsgebiet, EP: Endpunkt, Gesundhb. LQ: Gesundheitsbezogene Lebensqualität, NB: Nutzenbewertung, ZN: Zusatznutzen. N = 192 Nutzenbewertungsverfahren.

Abb. 4: Berücksichtigung der gesundheitsbezogenen Lebensqualität in den Therapiegebieten
(Quelle: Eigene Darstellung).

Der in den vorangestellten Kapiteln der Arbeit folgenden Struktur werden die Ergebnisse in thematischen Clustern aufbereitet. Das Evidenzprofil zur Nutzendimension gesundheitsbezogene Lebensqualität entwickelt sich in der Folge sukzessive. Anhang D informiert über die spezifischen Ein- und Ausschlüsse an Literatur und setzt jene Nutzenbewertungsverfahren und (Referenz-)Studien in Beziehung, welche in die nachfolgend präsentierten Auswertungen eingehen.

5.1 Verfahrensebene (I.1)

Im Hinblick auf onkologische Erkrankungen wird die Zielgröße gesundheitsbezogene Lebensqualität in rund 4 von 5 Nutzenbewertungsverfahren (80 %) erhoben, operationalisiert und zur Ableitung eines Zusatznutzens in den Dossiers berichtet. Die Verteilung dieser 47 Verfahren nach onkologischer Entität ist der nachstehenden Abbildung 5 zu entnehmen.

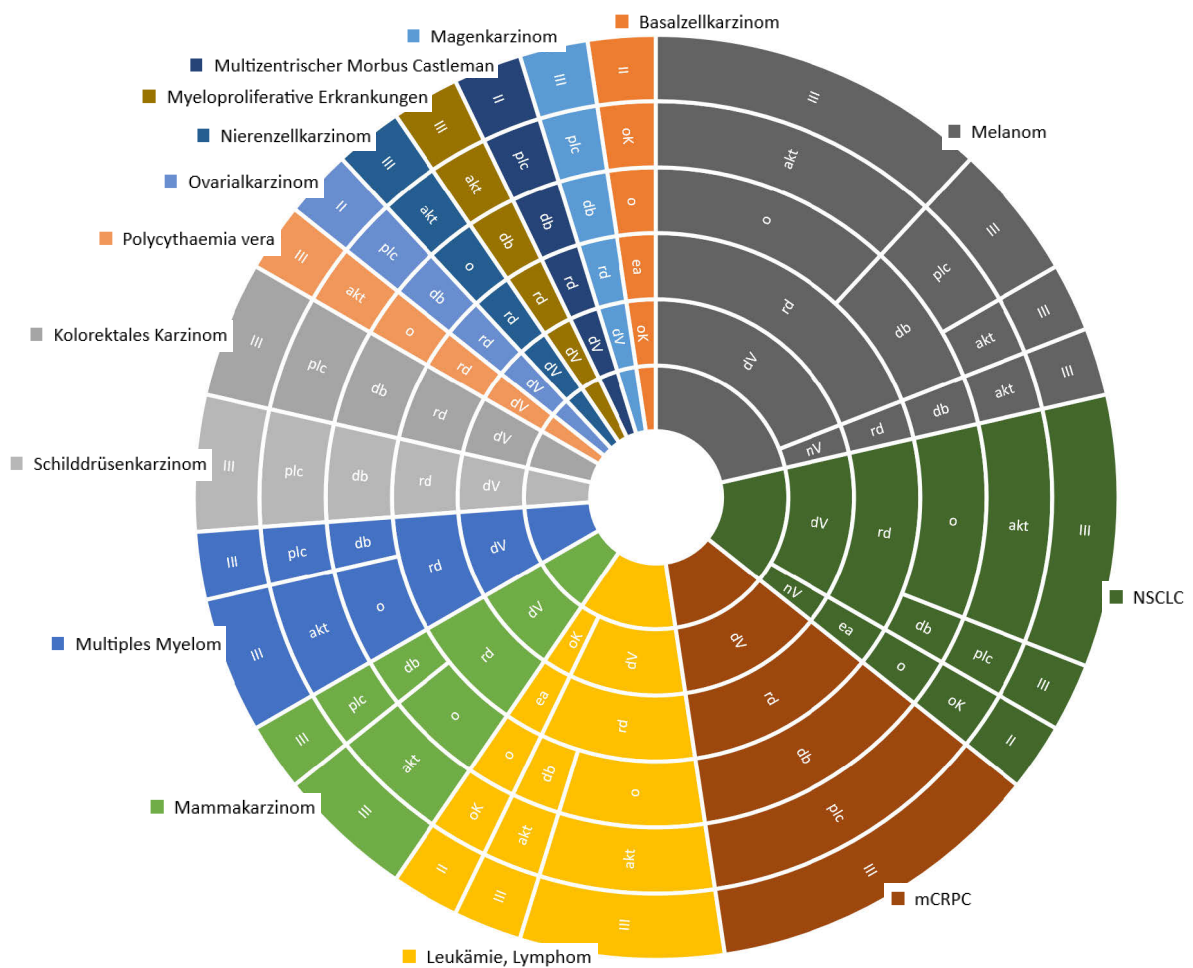


AWG: Anwendungsgebiet. N = 47 Nutzenbewertungsverfahren.

Abb. 5: Onkologische Entitäten mit Endpunktbedeutung für die gesundheitsbezogene Lebensqualität
(Quelle: Eigene Darstellung).

5.2 Studienebene (I.2)

Insgesamt 52 von den pUs eingebrachte Studien führen Nutzenbelege zur gesundheitsbezogenen Lebensqualität an. Dabei fußt die Bewertung von Behandlungsverfahren zu 85 Prozent auf direkten Vergleichen. Was den Typus der zu bewertenden Studien anbelangt, finden zu 88 Prozent randomisierte kontrollierte Studien Verwendung – davon 52 Prozent mit doppelblindem Design. Wiederum rund jede zweite dieser Studien (57 %) untersucht den Interventionseffekt gegenüber einem aktiven Komparator. Abbildung 6 legt die Zuordnung der Attribute nach onkologischer Entität dar.



dV: direkter Vergleich, nV: nicht-adjustierter indirekter Vergleich, oK: ohne Kontrollgruppe; ea: einarmig, rd: randomisiert; db: doppelblind, o: offen; akt: aktivkontrolliert, plc: placebokontrolliert; II: Phase-II-Studie, III: Phase-III-Studie.

mCRPC: Metastasiertes kastrationsresistentes Prostatakarzinom, NSCLC: Nicht-kleinzelliges Bronchialkarzinom.

Simplifizierte Darstellung: Im Falle des Einbringens mehrerer zu bewertender Studien gilt je Nutzenbewertungsverfahren die gemäß Evidenzgrad höherwertige Ausprägung.

Lesart: Dargestellt sind die nach Tumorentität bestimmten relativen Anteile für die Abfolge (i) direkter Vergleich vs. nicht-adjustierter indirekter Vergleich vs. ohne Kontrollgruppe, (ii) randomisiert vs. einarmig, (iii) doppelblind vs. offen, (iv) aktivkontrolliert vs. placebokontrolliert vs. ohne Kontrollgruppe und (v) Phase-II- vs. Phase-III-Studie. Im Falle des Clusters Leukämie, Lymphom beispielsweise sind die zu bewertenden Studien überwiegend randomisierte, direkte Vergleichsstudien der Phase III bei in der Mehrzahl offenem Design und aktiver Kontrolle.

N = 42 Nutzenbewertungsverfahren.

Abb. 6: Design klinischer Studien mit Endpunktbedeutung für die gesundheitsbezogene Lebensqualität (Quelle: Eigene Darstellung).

5.3 Endpunktebene (I.3)

Die Endpunkte der gesundheitsbezogenen Lebensqualität sind in den zu bewertenden Studien überwiegend als sekundäre Zielkriterien definiert. Eine konfirmatorische Konzeptualisierung – im Sinne einer Festlegung als primärer Endpunkt – fehlt hingegen vollständig (vgl. Abb. 7).



sekundär: sekundärer Studienendpunkt, explorativ: explorativer Studienendpunkt, abweichend: tertiärer Studienendpunkt (3-mal) bzw. als „weiterer“ Endpunkt der Studie nicht näher spezifiziert (6-mal).

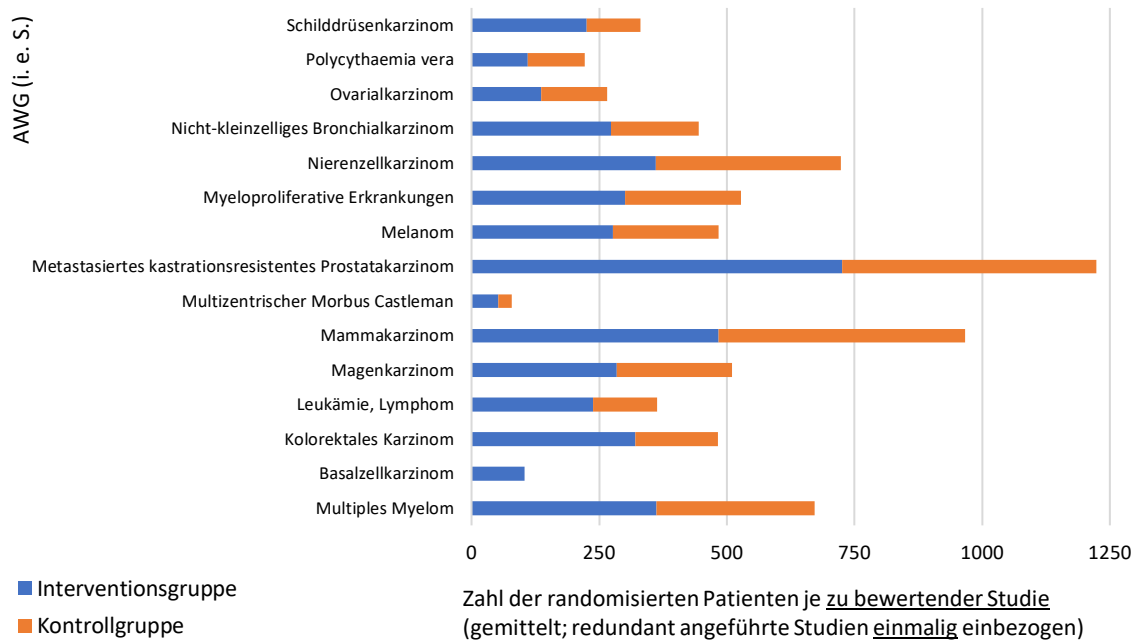
MCD: Multizentrischer Morbus Castleman, MPE: Myeloproliferative Erkrankungen.

Simplifizierte Darstellung: Im Falle des Einbringens mehrerer zu bewertender Studien gilt je Nutzenbewertungsverfahren die gemäß Evidenzgrad höherwertige Ausprägung.

N = 42 Nutzenbewertungsverfahren.

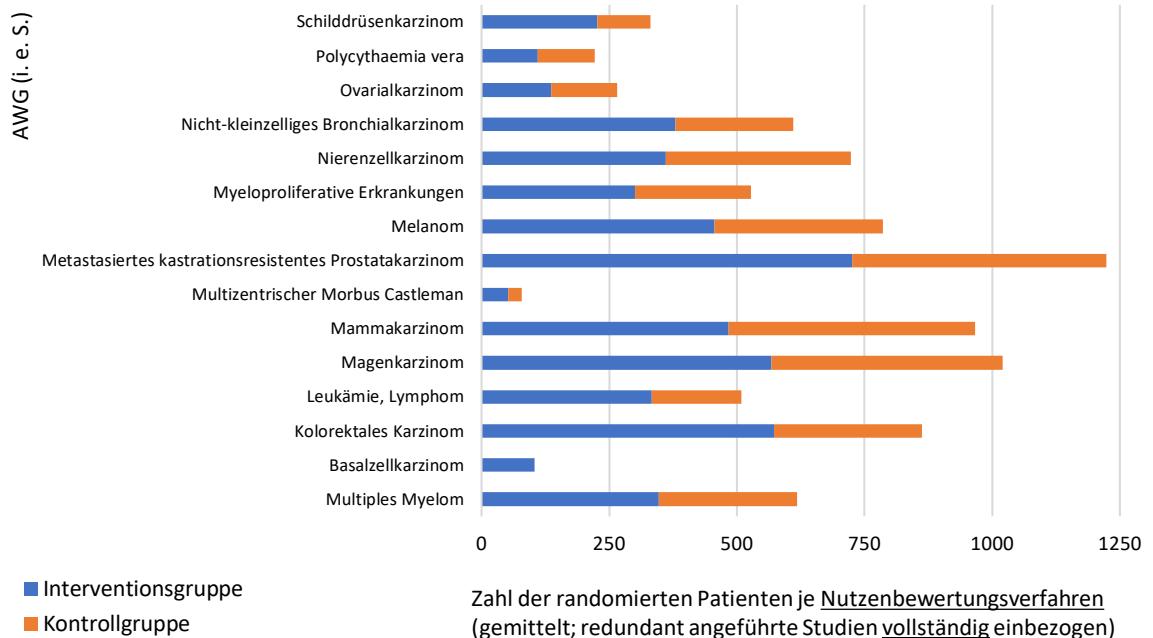
Abb. 7: Konzeptioneller Stellenwert von Endpunkten der gesundheitsbezogenen Lebensqualität (Quelle: Eigene Darstellung).

Gemittelt über die onkologischen Entitäten berücksichtigt die Erhebung der gesundheitsbezogenen Lebensqualität je zu bewertender Studie 284 Patienten der Interventionsgruppe und 210 Patienten der Kontrollgruppe (vgl. Abb. 8). Auf Ebene der Nutzenbewertungsverfahren unter Zählung sämtlicher – in Teilen identischer – zu bewertender Studien umfasst die Interventionsgruppe im Durchschnitt 344 Patienten und die Kontrollgruppe im Durchschnitt 246 Patienten (vgl. Abb. 9).



AWG: Anwendungsgebiet. N = 47 Nutzenbewertungsverfahren.

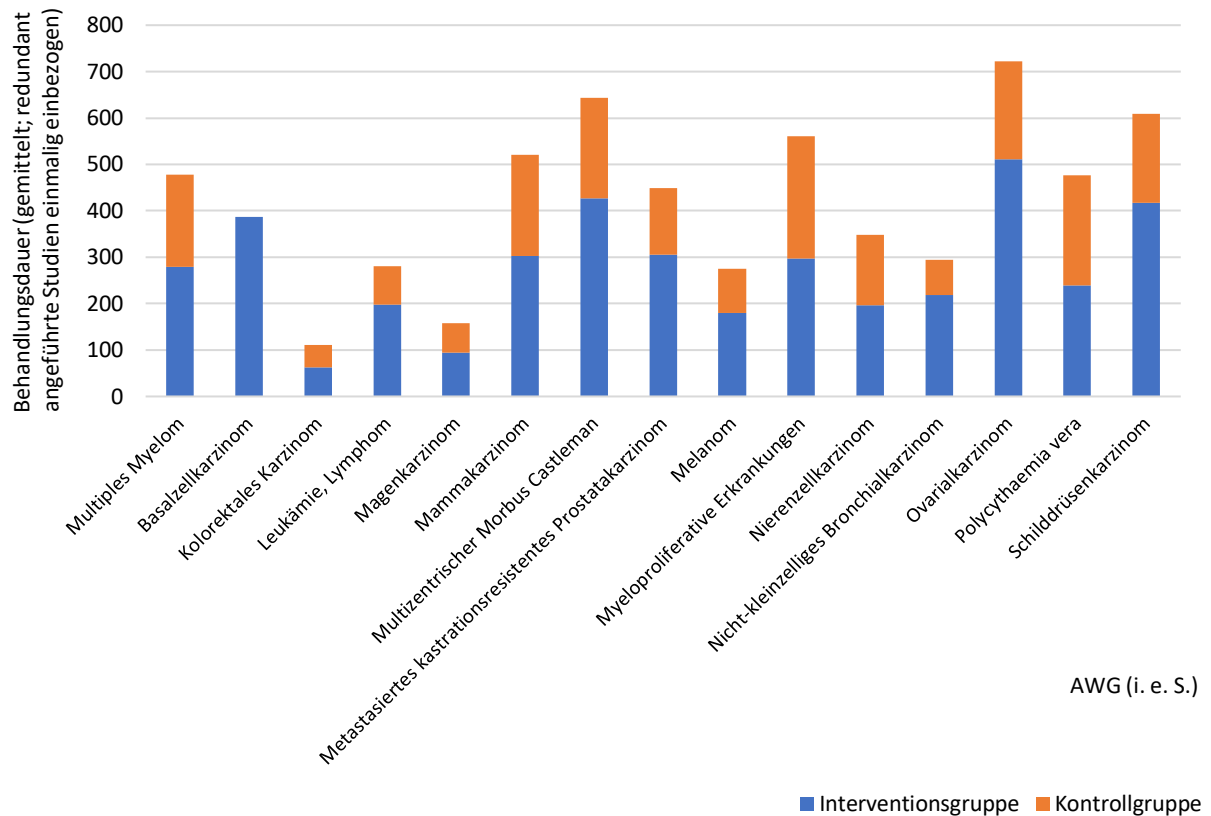
Abb. 8: Anzahl randomisierter Patienten je zu bewertender Studie
(Quelle: Eigene Darstellung).



AWG: Anwendungsgebiet. N = 47 Nutzenbewertungsverfahren.

Abb. 9: Anzahl randomisierter Patienten je Nutzenbewertungsverfahren
(Quelle: Eigene Darstellung).

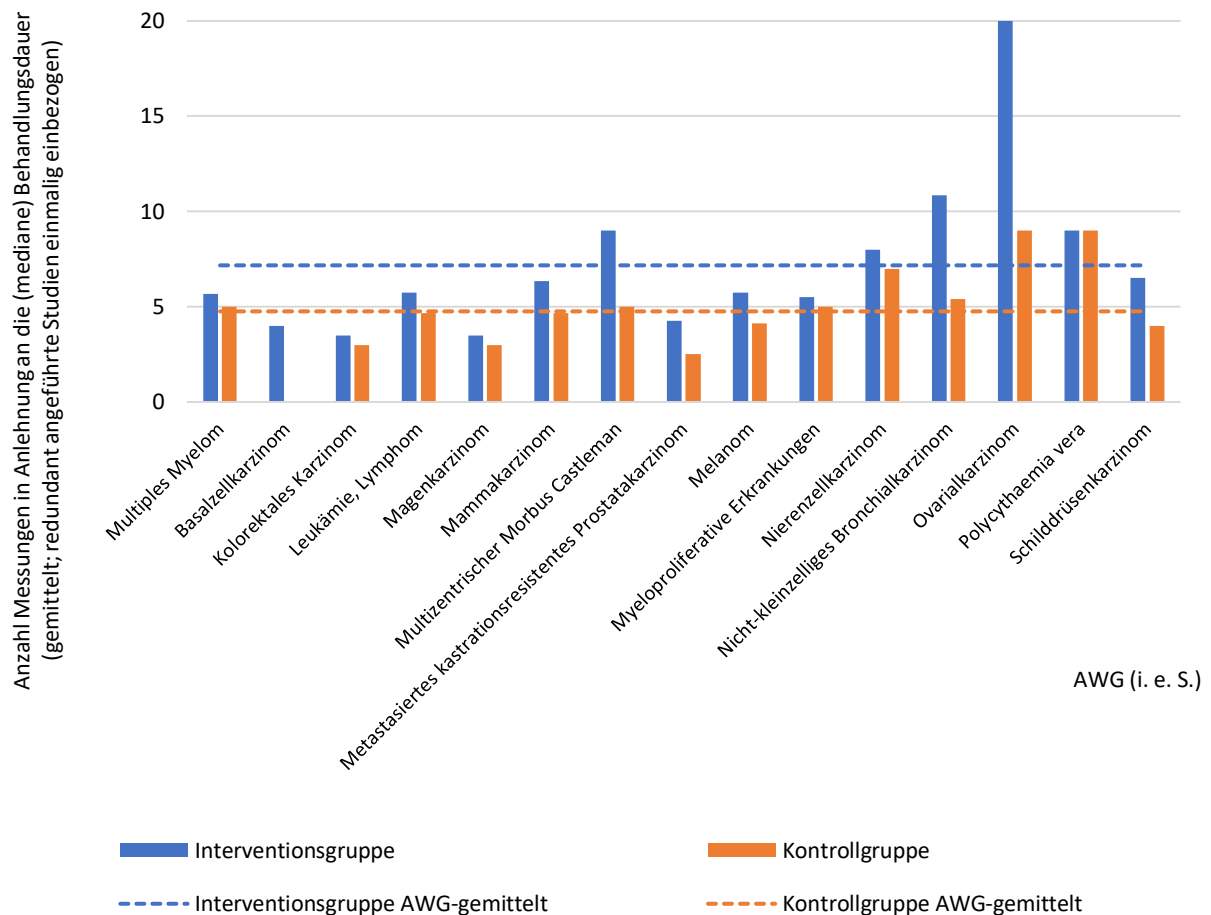
Bereinigt um redundant angeführte Studien in den Nutzenbewertungsverfahren beträgt die Behandlungsdauer im Falle der Interventionsgruppe im Mittel 274 Tage und im Falle der Kontrollgruppe im Mittel 147 Tage (vgl. Abb. 10).



Anmerkung: Die Rohdaten berücksichtigen Angaben zum Median (ersatzweise solche zum Mittelwert). AWG: Anwendungsgebiet. N = 37 Nutzenbewertungsverfahren, 42 zu bewertende Studien (ferner: keine Angabe).

Abb. 10: Behandlungsdauer je zu bewertender Studie
(Quelle: Eigene Darstellung).

Die ergänzende Verwertung des jeweiligen Modus der Zielgrößenerhebung liefert die in Abbildung 11 illustrierten Angaben zur mittleren tatsächlichen Anzahl durchgeführter Messungen je onkologischer Entität. Gemittelt über die (geclusterten) Anwendungsgebiete ergeben sich 7,2 Messungen für die Interventionsgruppe und 4,8 Messungen für die Kontrollgruppe.



Beispiel (vgl. Nutzenbewertungsverfahren zu Decitabin [65] aus dem AWG-Cluster Leukämie, Lymphom):
 Definierte Erhebungszeitpunkte: Studienbeginn, Tag 1 von Zyklus 3, 4 und 6 sowie am Ende der Studienbehandlung (Wiederholung des Behandlungszyklus alle 4 Wochen). Mittlere Behandlungsdauer: 132 Tage (Interventionsgruppe) bzw. 66 Tage (Kontrollgruppe). Rechnerische Anzahl Messungen im Behandlungszeitraum: 4 (Interventionsgruppe) bzw. 3 (Kontrollgruppe).

AWG: Anwendungsgebiet. N = 36 Nutzenbewertungsverfahren, 41 zu bewertende Studien (ferner: keine Angabe).

Abb. 11: Anzahl Messungen je zu bewertender Studie
 (Quelle: Eigene Darstellung).

Ungeachtet der Verfahren, in welchen eine formale statistische Analyse fehlt – respektive Angaben als solche fehlen – erreichen die Ergebnisse zur gesundheitsbezogenen Lebensqualität auf Grundlage der aufbereiteten Evidenzen der pUs in 36 Nutzenbewertungsverfahren (77 %) statistische Signifikanz. In 30 dieser Verfahren (64 %) wird der Unterschied gänzlich oder partiell als klinisch relevant beschrieben. Dabei meint partiell, dass sich die klinische Relevanz entweder auf einzelne Komponenten von Fragebögen, bestimmte Operationalisierungen, explizite Zeitpunkte der Erhebung oder separate Teilpopulationen erstreckt. In 28 jener Verfahren (60 %) beanspruchen pUs schließlich einen die Endpunktkategorie betreffenden Zusatznutzen – qualitativ oder in Anlehnung an die Kategorisierungsoptionen gemäß AM-NutzenV.

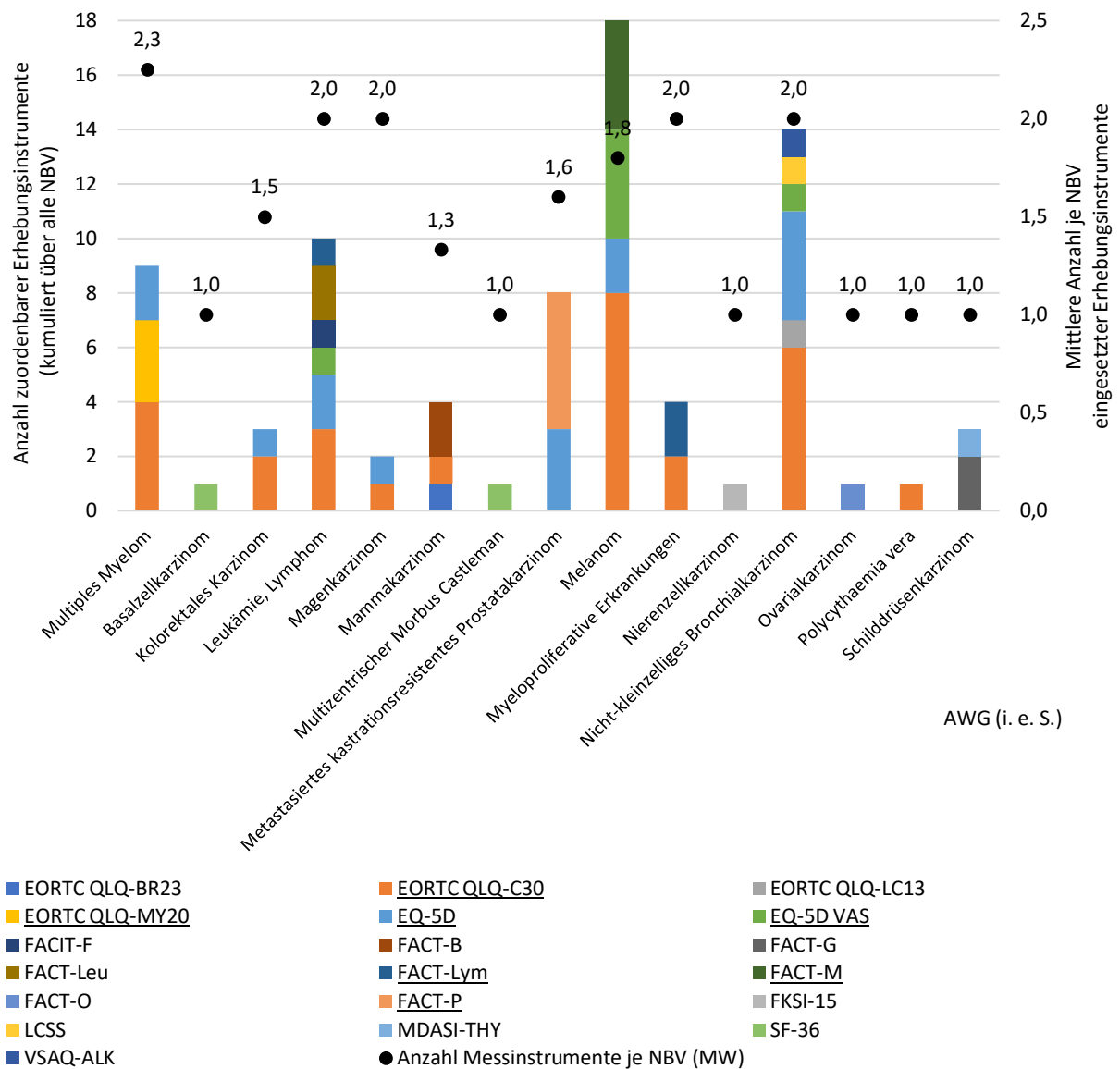
5.4 Kongruenz innerhalb der zu bewertenden Studien (II.1)

Hinsichtlich der zu bewertenden Studien sind auf Ebene der (geclusterten) onkologischen Entitäten folgende Unterschiede in den Patientencharakteristika zwischen den Interventionsgruppen einerseits und den Kontrollgruppen andererseits festzustellen: (i) Alter: 1,2 Jahre (Mittelwert), maximal 4,1 Jahre (Cluster: Leukämie, Lymphom), (ii) Anteil Frauen (exklusive der Cluster: Metastasiertes kastrationsresistentes Prostatakarzinom, Ovarialkarzinom): 5,5 Prozent (Mittelwert), maximal 28,0 Prozent (Cluster: Multizentrischer Morbus Castleman), (iii) *ECOG PS* gw.: 0,060 (Mittelwert), maximal 0,180 (Cluster: Leukämie, Lymphom) beziehungsweise Anteil eines *ECOG PS* von 0 oder 1: 2,9 Prozent (Mittelwert), maximal 13,2 Prozent (Cluster: Multizentrischer Morbus Castleman), (iv) Anteil Kaukasier/weiß: 2,8 Prozent (Mittelwert), maximal 10,4 Prozent (Cluster: Multizentrischer Morbus Castleman) und (v) Anteil (West-)Europa: 1,8 Prozent (Mittelwert), maximal 6,3 Prozent (Cluster: Multizentrischer Morbus Castleman).

Resümierend sind die numerisch größten Abweichungen bei den onkologischen Erkrankungen Leukämie, Lymphom und Multizentrischer Morbus Castleman zu beobachten. Die Verteilungen je Merkmal und onkologischem Cluster sind den Abbildungen E1 bis E6 des Anhangs E zu entnehmen.

5.5 Kongruenz hinsichtlich der regelhaft angeführten Validierungsstudien (II.2)

Die nachstehende Betrachtung schließt Rang 1 bis 6 der in den Nutzenbewertungsverfahren meistverwendeten Instrumente zur Erhebung der gesundheitsbezogenen Lebensqualität ein. Grundend auf den 47 berücksichtigungsfähigen Verfahren und seinen insgesamt 80 separaten Endpunktbetrachtungen ergibt sich folgende Rangordnung (Zählung jener Endpunktbetrachtungen nebst prozentualem Anteil je Messinstrument in absteigender Reihung): (i) *EORTC QLQ-C30*: 28 (35 %), (ii) *EQ-5D (VAS)*: 21 (26 %), (iii) *FACT-P*: 5 (6 %), (iv) *FACT-M*: 4 (5 %), (v) *EORTC QLQ-MY20*: 3 (4 %) und (vi) *FACT-Lym*: 3 (4 %). Weitere Messinstrumente sind mit einer Häufigkeit von jeweilig ≤ 2 in dem untersuchten Datenset zu finden. Abbildung 12 fasst den Grad der Berücksichtigung und die Zuordnung der einschlägigen Erhebungsverfahren gruppiert nach onkologischen Indikationen zusammen.



Anmerkung: Ausgeschlossen sind nicht berichtete sowie nicht als originäre Erhebungsinstrumente der gesundheitsbezogenen Lebensqualität klassifizierte Messverfahren (vgl. Schmerz [visuelle Analogskala], ECOG PS, FACIT-F, MCD SS, KPS, PGIC).

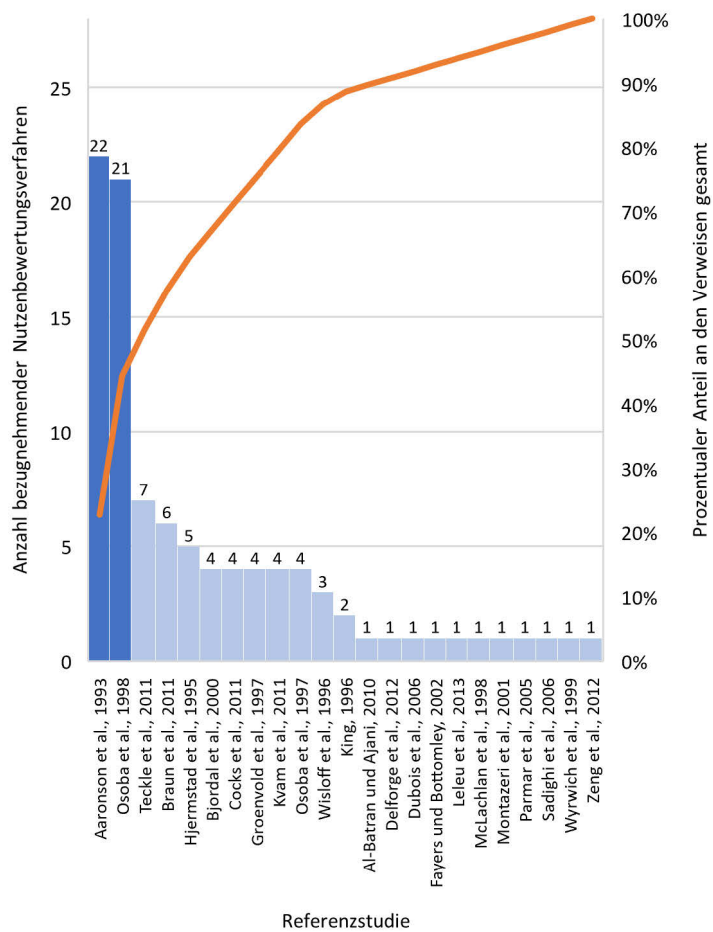
AWG: Anwendungsgebiet, MW: Mittelwert, NBV: Nutzenbewertungsverfahren. N = 47 Nutzenbewertungsverfahren.

Abb. 12: Bedeutungsgrad eingesetzter Messinstrumente
(Quelle: Eigene Darstellung).

Zu den Unterschieden hinsichtlich der zentralen Patientencharakteristika – Alter, Geschlecht, Grad der Erkrankungsschwere, Ethnizität und geographische Region – informieren die nachstehenden Subkapitel. Dabei wird der Betrachtung je Messinstrument die Selektion der Standardreferenzen vorangestellt. Die Tabellen 2 bis 7 fassen die nachfolgenden Ergebnisse am Ende dieses Kapitels zusammen.

Kongruenzlevel von Standardreferenzen: EORTC QLQ-C30

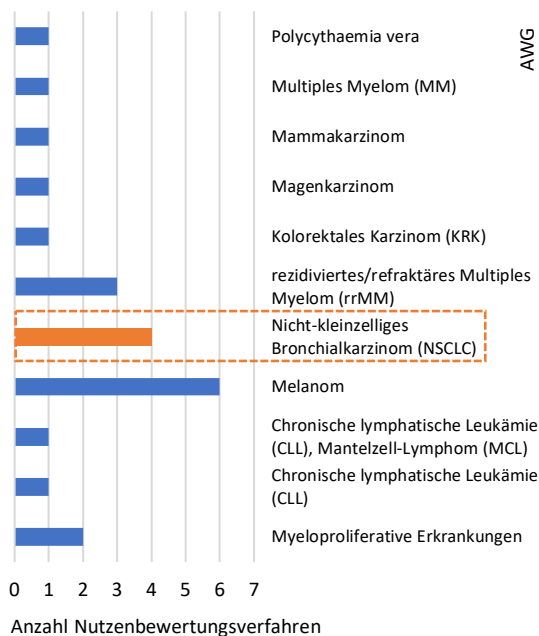
Mit Blick auf den *EORTC QLQ-C30* wird in 22 Nutzenbewertungsverfahren (79 %) auf die im Jahr 1993 publizierte Arbeit von Aaronson et al. [64] mit dem Titel „*The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology*“ (im Folgenden: **Aaronson et al., 1993**) verwiesen. Weiterhin wird in 21 jener Verfahren (75 %) die Studie von Osoba et al. [61] aus dem Jahr 1998 mit dem Titel „*Interpreting the significance of changes in health-related quality-of-life scores*“ (im Folgenden: **Osoba et al., 1998**) zitiert. Auf weitere Referenzen wird mit einer Häufigkeit von jeweilig ≤ 7 in den Nutzenbewertungsverfahren Bezug genommen (nicht weiter betrachtet). Die Validierungsstudien betreffend Aaronson et al., 1993 und Osoba et al., 1998 haben einen Anteil an den gesamten Verweisen, welche mit dem *EORTC QLQ-C30* in Beziehung stehen, von 44 Prozent. Abseits der Häufigkeitsverteilung ist Abbildung 13 zu entnehmen, dass diese Studien in sämtlichen onkologischen Clustern zum Validitätsnachweis angeführt werden, was der generisch-onkologischen Konzeption des *EORTC QLQ-C30* zuzurechnen ist.



Anmerkung: Verweise bereinigt um systematische Reviews (u. dgl.).
N = 28 Nutzenbewertungsverfahren.

Abb. 13: Ermittlung der Standardreferenzen; hier: EORTC QLQ-C30
(Quelle: Eigene Darstellung).

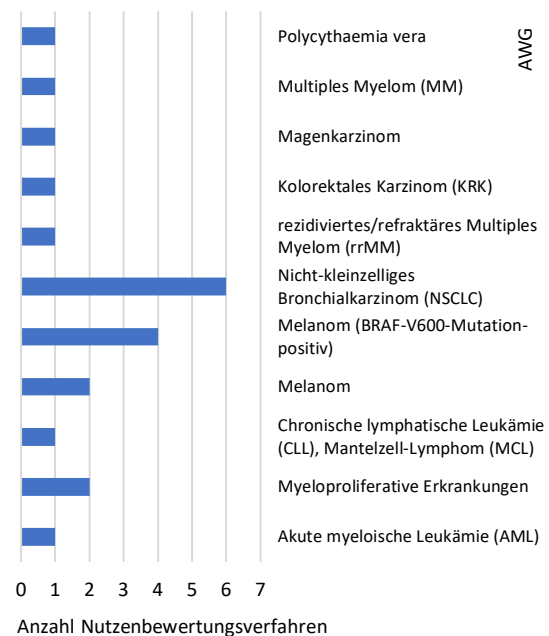
Während die Studienpopulationen in den Nutzenbewertungsverfahren gewichtet zu etwa 11 Prozent mit jener der vorgebrachten Validierungsstudie Aaronson et al., 1993 übereinstimmen³, was die zugrunde liegenden Erkrankungen anbelangt (vgl. Abb. 14), ist im Falle Osoba et al., 1998 keine Entsprechung auszumachen (vgl. Abb. 15).



Aaronson et al., 1993: *Non-small-cell lung cancer* 63,1 %, *small-cell lung cancer* 36,9 % [64].

AWG: Anwendungsgebiet. N = 22 Nutzenbewertungsverfahren.

Abb. 14: Entsprechung nach onkologischer Entität; hier: Aaronson et al., 1993
(Quelle: Eigene Darstellung).



Osoba et al., 1998: *recurrent or metastatic breast cancer, extensive-stage small-cell lung cancer* [61].

AWG: Anwendungsgebiet. N = 21 Nutzenbewertungsverfahren.

Abb. 15: Entsprechung nach onkologischer Entität; hier: Osoba et al., 1998
(Quelle: Eigene Darstellung).

Die relative Abweichung einer zu bewertenden Studie zur Referenz misst im Falle Aaronson et al., 1993 in der Spitze 19 Prozent für das Alter, 318 Prozent für den Anteil Frauen, 75 Prozent für den *ECOG PS* gw. sowie 44 Prozent⁴ für den Anteil (West-)Europa. Dabei ergeben die merkmalsbezogenen Testungen beim Alter in 15 Fällen (79 % [N = 25 zu bewertende Studien, davon 6 nicht auswertbar]), beim Anteil Frauen in 22 Fällen (88 %), beim *ECOG PS* gw. in 24 Fällen (100 %) und beim Anteil (West-)Europa in 8 Fällen (53 %) einen signifikanten Gruppenunterschied. Zusammengefasst weisen die zu bewertenden Studien nahezu geschlossen einen im Vergleich höheren Frauenanteil auf. Des Weiteren sind die Patienten in den zu

³ Berechnung: $4 \times 0,631 / 22$ (mit 4 = Nutzenbewertungsverfahren im Bereich des nicht-kleinzelligen Bronchialkarzinoms, 0,631 = relativer Anteil jener Histologie in der Referenzstudie und 22 = Nutzenbewertungsverfahren gesamt).

⁴ Unberücksichtigt der (2) Studien, in welchen gänzlich nicht in (West-)Europa rekrutiert wurde.

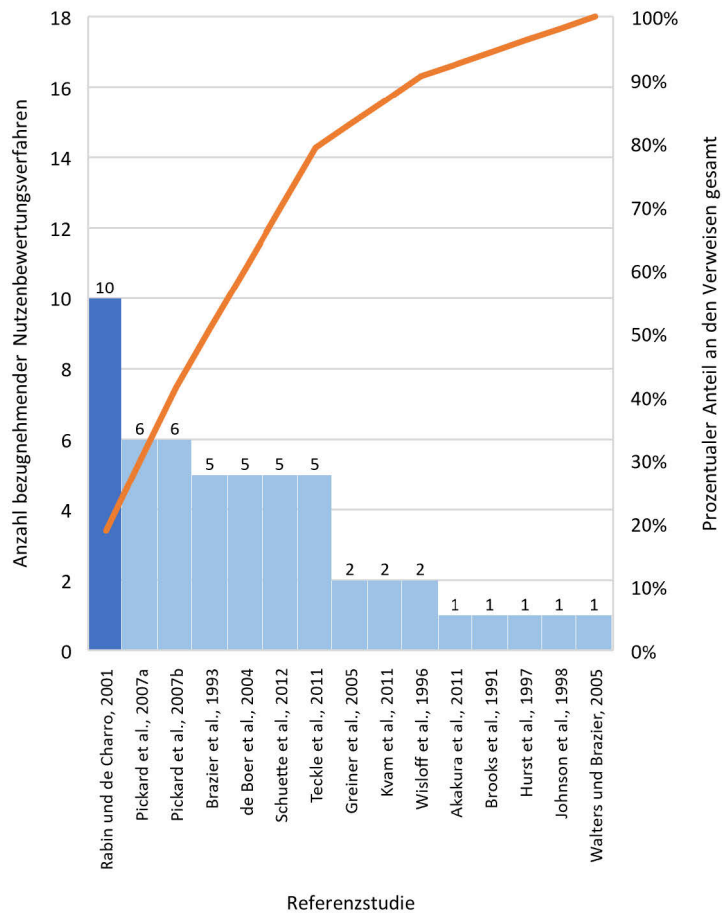
bewertenden Studien ausnahmslos gesünder und die Rekrutierung in (West-)Europa ist größtenteils geringer (vgl. Anhang F-I, Abb. F1- F4).

Bezüglich Osoba et al., 1998 beträgt die relative Abweichung einer zu bewertenden Studie maximal 32 Prozent für das Alter, 77 Prozent für den Anteil Frauen und 72 Prozent für den *ECOG PS* gw. Statistisch signifikante Unterschiede in den Gruppen sind beim Alter 11 Mal (65 % [N = 26 zu bewertende Studien, davon 9 nicht auswertbar]), beim Anteil Frauen 26 Mal (100 %) und beim *ECOG PS* gw. 24 Mal (92 %) in der Indikation Brustkrebs beziehungsweise 18 Mal (69 %) in der Indikation kleinzelliges Lungenkarzinom (*Small Cell Lung Cancer, SCLC*) errechnet. Im Ergebnis sind die Patienten in den zu bewertenden Studien häufig älter und nahezu alle sind gesünder. Zudem werden ausnahmslos weniger Frauen in die zu bewertenden Studien eingeschlossen⁵ (vgl. Anhang F-I, Abb. F5-F7).

Kongruenzlevel von Standardreferenzen: EQ-5D (VAS)

Der Beleg der Validierung von Komponenten des *EQ-5D* wird in 10 Nutzenbewertungsverfahren (56 %) anhand der Publikation „*EQ-5D: a measure of health status from the EuroQol Group*“ von Rabin und de Charro [66] aus dem Jahr 2001 (im Folgenden: **Rabin und de Charro, 2001**) erbracht. Dabei erfolgt ein Verweis auf jene Arbeit in 4 der insgesamt 6 onkologischen Entitäten, welche den Nutzenbewertungsverfahren mit diesem Messinstrument zuzuordnen sind. Auf weitere Validierungsstudien wird in jeweilig 6 oder weniger Fällen verwiesen (nicht weiter betrachtet). Die vollständige Segmentierung ist in Abbildung 16 wiedergegeben.

⁵ Entsprechende Angaben liegen lediglich für die Erkrankung Brustkrebs vor.



Prozentualer Anteil an den Verweisen gesamt

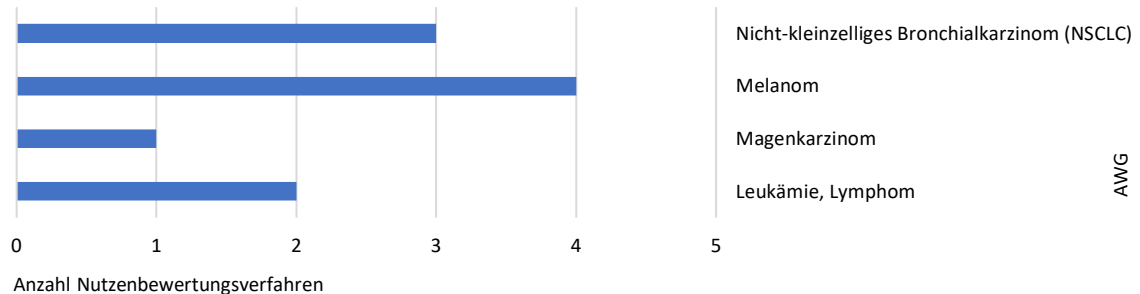


Anmerkung: Verweise bereinigt um systematische Reviews (u. dgl.).

mCRPC: Metastasiertes kastrationsresistentes Prostatakarzinom, NSCLC: Nicht-kleinzelliges Bronchialkarzinom. N = 18 Nutzenbewertungsverfahren.

Abb. 16: Ermittlung der Standardreferenzen; hier: EQ-5D (VAS)
(Quelle: Eigene Darstellung).

Die Teilnehmer in den zu bewertenden Studien unterscheiden sich hinsichtlich der onkologischen Indikationen vollständig von jenen, die zum Nachweis der Validität des Erhebungsverfahrens in Rabin und de Charro, 2001 angeführt sind (vgl. Abb. 17).



Rabin und de Charro, 2001: *rheumatoid arthritis (RA)* [66].

Anmerkung: Rabin und de Charro, 2001 [66] führt im Wesentlichen die Ergebnisse aus Hurst et al., 1997 [67] an.

AWG: Anwendungsgebiet. N = 10 Nutzenbewertungsverfahren.

Abb. 17: Entsprechung nach onkologischer Entität; hier: Rabin und de Charro, 2001
(Quelle: Eigene Darstellung).

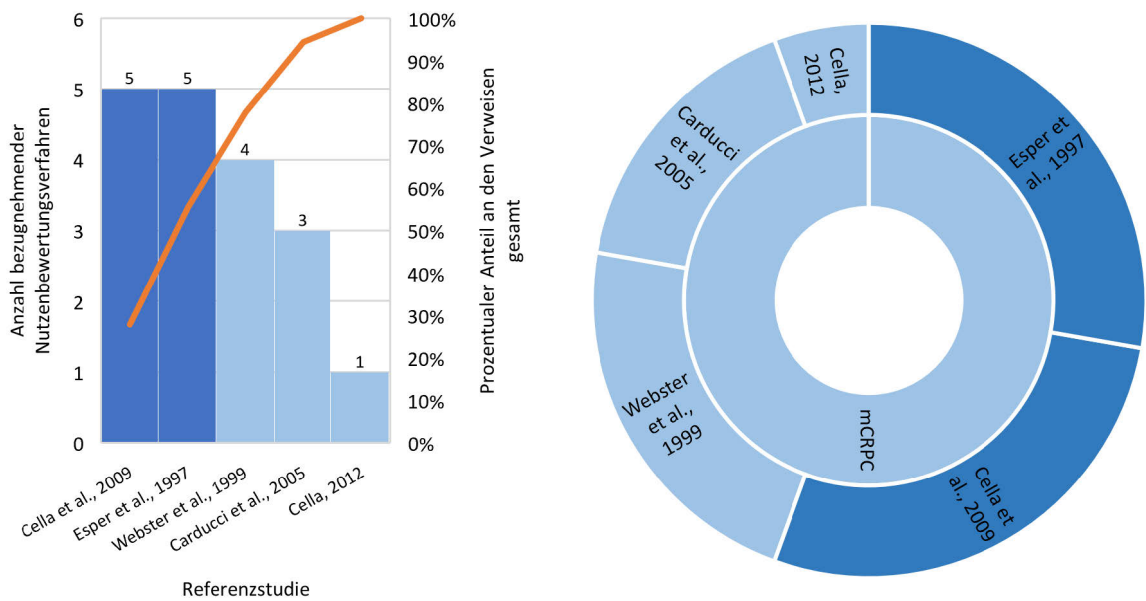
Für Rabin und de Charro, 2001 ergibt sich eine relative Abweichung einer zu bewertenden Studie zu dieser Referenz von höchstens 26 Prozent beim Alter, 64 Prozent beim Anteil Frauen und 74 Prozent⁶ beim Anteil (West-)Europa. Die Unterschiede sind statistisch signifikant in 9 Fällen für das Merkmal Alter (82 % [N = 13 zu bewertende Studien, davon 2 nicht auswertbar]), in 13 Fällen für das Merkmal Geschlecht (100 %) und in 8 Fällen für das Merkmal geographische Region (100 %). Tendenziell sind die Patienten in den zu bewertenden Studien älter. Zudem wurden durchgehend weniger Frauen in jene Studien eingeschlossen und es wurde ausnahmslos zu einem geringeren Anteil in (West-)Europa rekrutiert (vgl. Anhang F-II, Abb. F8-F10).

Kongruenzlevel von Standardreferenzen: FACT-P

Auf die Publikation „*Measuring quality of life in men with prostate cancer using the functional assessment of cancer therapy-prostate instrument*“ aus dem Jahr 1997 von Esper et al. [68] (im Folgenden: **Esper et al., 1997**) sowie auf die Arbeit „*Estimating clinically meaningful changes for the Functional Assessment of Cancer Therapy--Prostate: results from a clinical trial of patients with metastatic hormone-refractory prostate cancer*“ aus dem Jahr 2009 von Cella et al. [69] (im Folgenden: **Cella et al., 2009**) wird in je 5 Nutzenbewertungsverfahren (je 100 %) im Anwendungsfeld des fortgeschrittenen Prostatakarzinoms Bezug genommen, um eine

⁶ Unberücksichtigt der Studie, in welcher gänzlich nicht in (West-)Europa rekrutiert wurde.

hinreichende Güte der psychometrischen Eigenschaften der Instrumente zu belegen (vgl. Abb. 18).

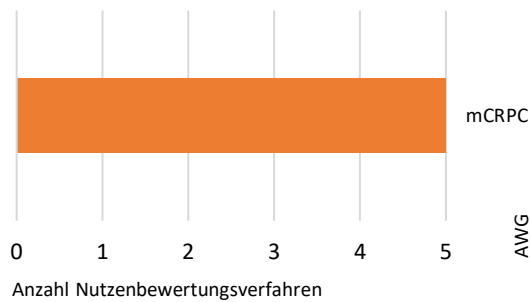


Anmerkung: Verweise bereinigt um systematische Reviews (u. dgl.).

mCRPC: Metastasiertes kastrationsresistentes Prostatakarzinom. N = 5 Nutzenbewertungsverfahren.

Abb. 18: Ermittlung der Standardreferenzen; hier: *FACT-P*
(Quelle: Eigene Darstellung).

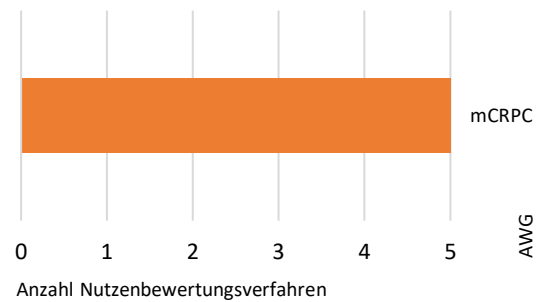
Ungeachtet des spezifischen Erkrankungsstadiums stimmen die Populationen der zu bewertenden Studien mit jenen der beiden Referenzstudien überein, was die onkologische Entität anbetrifft (vgl. Abb. 19-20).



Esper et al., 1997: (i) *prostate cancer (at various stages)*, (ii) *patients who had undergone radical prostatectomy (scale development sample)*, *advanced hormone refractory prostate cancer (validity sample 1)*, *patients who (i) were managing their disease conservatively*, (ii) *had evidence of metastasis or reoccurrence*, (iii) *had clinically localized disease*, (iv) *had presented with prostate problems, without yet having gotten a positive biopsy result (validity sample 2)* [68].

AWG: Anwendungsgebiet, mCRPC: Metastasiertes kastrationsresistentes Prostatakarzinom. N = 5 Nutzenbewertungsverfahren.

Abb. 19: Entsprechung nach onkologischer Entität; hier: Esper et al., 1997
(Quelle: Eigene Darstellung).



Cella et al., 2009: *metastatic hormone-refractory prostate cancer (HRPC)* [69].

AWG: Anwendungsgebiet, mCRPC: Metastasiertes kastrationsresistentes Prostatakarzinom. N = 5 Nutzenbewertungsverfahren.

Abb. 20: Entsprechung nach onkologischer Entität; hier: Cella et al., 2009
(Quelle: Eigene Darstellung).

Hinsichtlich Esper et al., 1997 betragen die relativen Abweichungen einer zu bewertenden Studie zu dieser Referenz bis zu 8 Prozent für das Alter und bis zu 26 Prozent für den Anteil Kaukasier/weiß. Für den Anteil (West-)Europa misst die Differenz der Verhältnisse 53 Prozent⁷. Statistische Signifikanzen erreichen die Gruppenunterschiede 4 Mal bezogen auf das Merkmal Alter (100 % [N = 5 zu bewertende Studien, davon 1 nicht auswertbar]), 4 Mal bezogen auf das Merkmal Ethnizität (80 %) und 1 Mal bezogen auf das Merkmal geographische Region (100 %). In der Gesamtschau sind die Patienten in den zu bewertenden Studien durchweg älter. Weiterhin wurden weit überwiegend Kaukasier in diese Studien eingeschlossen (vgl. Anhang F-III, Abb. F11-F13).

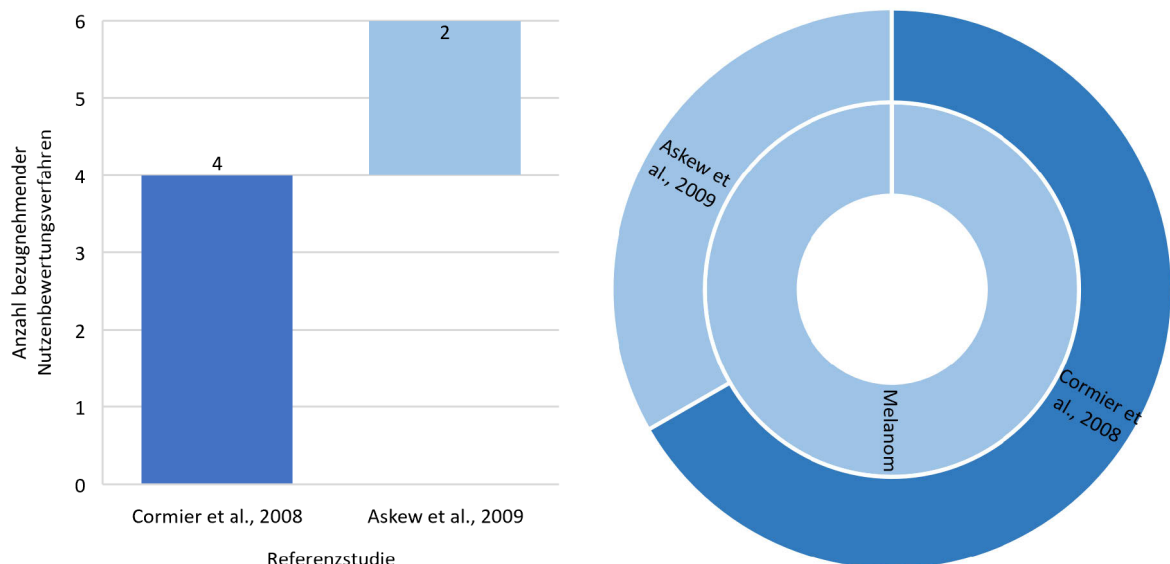
Die relativen Abweichungen der zu bewertenden Studien zu Cella et al., 2009 betragen in der Spitze 4 Prozent beim Alter sowie 19 Prozent beim Anteil Kaukasier/weiß. Die Differenz zur Validierungsstudie mit Blick auf den *ECOG PS gw.* beläuft sich auf maximal 0,871. Statistisch signifikant sind die Unterschiede im Hinblick auf das Alter bei 3 Vergleichen (75 % [N = 5 zu bewertende Studien, davon 1 nicht auswertbar]) und im Hinblick auf den Anteil Kaukasier/weiß bei 4 Vergleichen (80 %). Lediglich deskriptiv auswertbar waren indes die Befunde zum *ECOG PS gw.* (N = 4 zu bewertende Studien). Im Ergebnis sind die Patienten in den zu

⁷ Angaben hierzu liegen lediglich für eine zu bewertende Studie vor.

bewertenden Studien ohne Ausnahme jünger und durchweg weniger leistungsfähig. Zudem wurden in der Tendenz weniger Kaukasier in die zu bewertenden Studien aufgenommen (vgl. Anhang F-III, Abb. F14-F16).

Kongruenzlevel von Standardreferenzen: *FACT-M*

In 4 Nutzenbewertungsverfahren (100 %) wird auf die im Jahr 2008 veröffentlichte Studie von Cormier et al. [70] mit dem Titel „*Prospective Assessment of the Reliability, Validity, and Sensitivity to Change of the Functional Assessment of Cancer Therapy-Melanoma Questionnaire*“ (im Folgenden: **Cormier et al., 2008**) rekurriert. Cormier et al., 2008 wird damit im Bereich des Melanoms konstant als Referenz benannt, wenn es darum geht, einen Beleg der Validierung des *FACT-M* zu erbringen (vgl. Abb. 21).



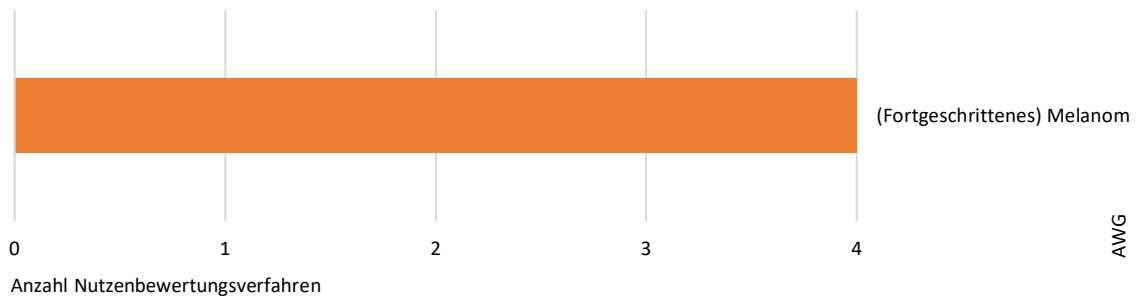
Anmerkung: Verweise bereinigt um systematische Reviews (u. dgl.).

N = 4 Nutzenbewertungsverfahren.

Abb. 21: Ermittlung der Standardreferenzen; hier: *FACT-M*

(Quelle: Eigene Darstellung).

Sowohl die in den Nutzenbewertungsprozess eingebrachten (zu bewertenden) Studien als auch die bezugnehmende Studie Cormier et al., 2008 sind im Bereich des (fortgeschrittenen) Melanoms verortet, was die Behandlungskontexte anbelangt (vgl. Abb. 22).



Cormier et al., 2008: *melanoma* [70].

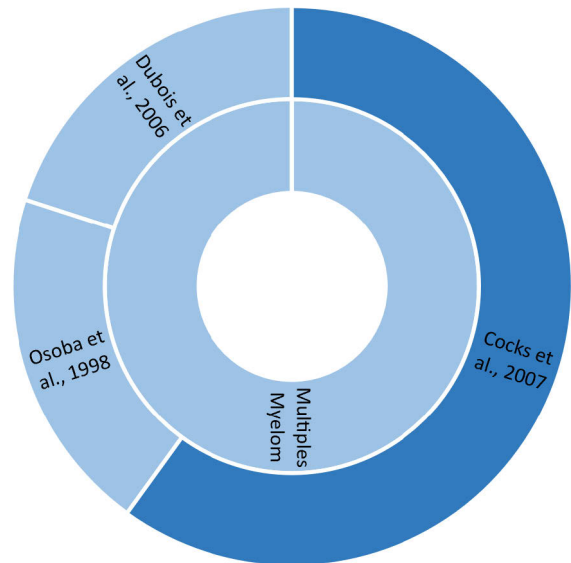
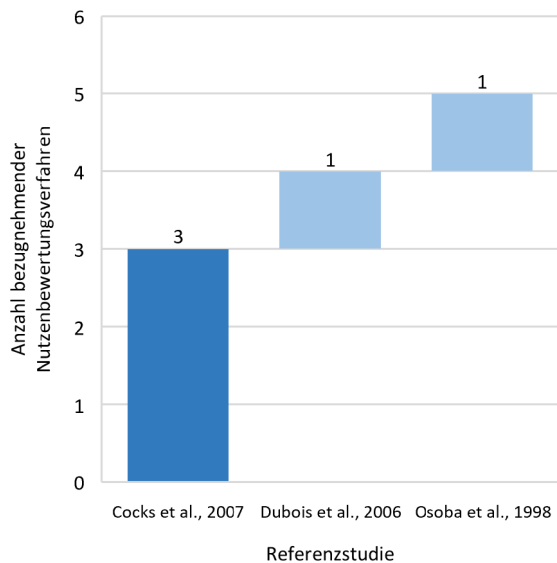
AWG: Anwendungsgebiet. N = 4 Nutzenbewertungsverfahren.

Abb. 22: Entsprechung nach onkologischer Entität; hier: Cormier et al., 2008
(Quelle: Eigene Darstellung).

Die relativen Abweichungen im Verhältnis von zu bewertender Studie und Cormier et al., 2008 als Referenzpublikation belaufen sich im Extremfall auf 5 Prozent beim Alter, jeweils 11 Prozent beim Anteil Frauen und beim *ECOG PS* gw. sowie 2 Prozent beim Anteil Kaukasier/weiß. Daneben beträgt die Differenz zu Cormier et al., 2008 für den Anteil (West-)Europa maximal 68 Prozent. Die Unterschiede in der Patientencharakteristik erreichen hinsichtlich der Kriterien Geschlecht (N = 3 zu bewertende Studien), Grad der Erkrankungsschwere (N = 3 zu bewertende Studien) und Ethnizität (N = 2 zu bewertende Studien) in keiner Konstellation statistische Signifikanz. Bezüglich des Kriteriums geographische Region ist die Differenz hingegen in 3 Fällen (100 %) statistisch signifikant. Keine Aussage kann indes für das Kriterium Alter getroffen werden, da relevante Angaben zur Verteilung nicht berichtet wurden. Zusammenfassend sind die Patienten in den zu bewertenden Studien ausnahmslos älter und der Anteil Frauen ist sämtlich höher. Betreffend den *ECOG PS* gw. und den Anteil Kaukasier/weiß ist die Verteilung im Mittel ausgeglichen. Schließlich wurde in einem weit überwiegenden Maße in (West-)Europa rekrutiert (vgl. Anhang F-IV, Abb. F17-F21).

Kongruenzlevel von Standardreferenzen: EORTC QLQ-MY20

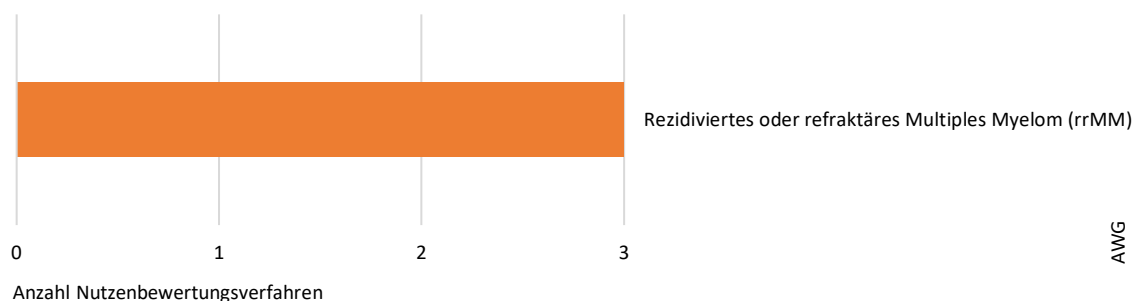
Im Bereich des Multiplen Myeloms wird die Validität des *EORTC QLQ-MY20* in 3 Nutzenbewertungsverfahren (100 %) mit Verweis auf die im Jahr 2007 erschienene Arbeit „*An international field study of the reliability and validity of a disease-specific questionnaire module (the QLQ-MY20) in assessing the quality of life of patients with multiple myeloma*“ von Cocks et al. [71] (im Folgenden: **Cocks et al., 2007**) belegt. Zwei weitere Referenzen werden zudem einmalig benannt (nicht weiter betrachtet) (vgl. Abb. 23).



Anmerkung: Verweise bereinigt um systematische Reviews (u. dgl.).
N = 3 Nutzenbewertungsverfahren.

Abb. 23: Ermittlung der Standardreferenzen; hier: EORTC QLQ-MY20
(Quelle: Eigene Darstellung).

Sowohl die zu bewertenden Studien als auch Cocks et al., 2007 als referenzielle Arbeit berücksichtigen Myelom-Patienten – gleichwohl in zum Teil differentieller Therapiesituation (vgl. Abb. 24).



Cocks et al., 2007: *myeloma patients (newly diagnosed or relapsed/refractory)* [71].

AWG: Anwendungsgebiet. N = 3 Nutzenbewertungsverfahren.

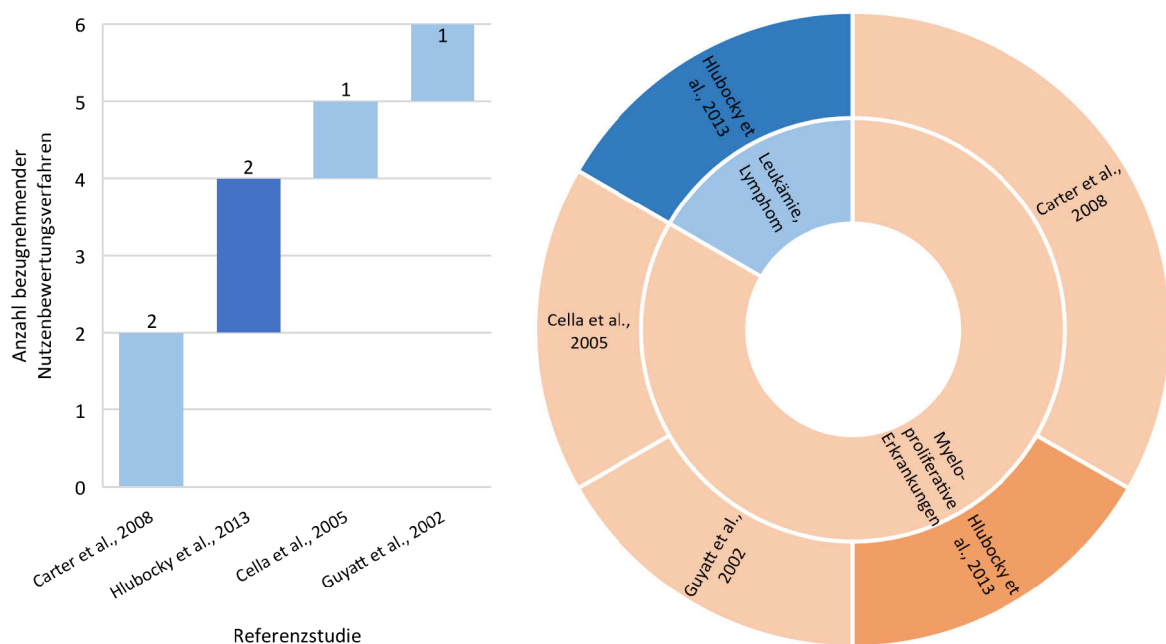
Abb. 24: Entsprechung nach onkologischer Entität; hier: Cocks et al., 2007
(Quelle: Eigene Darstellung).

Die relativen Abweichungen von zu bewertender Studie zu Cocks et al., 2007 ergeben bis zu 3 Prozent für das Spezifikum Alter, bis zu 14 Prozent für das Spezifikum Geschlecht und bis zu 37 Prozent für das Spezifikum Erkrankungsschwere. Die Unterschiede in den Vergleichsgruppen sind bezüglich des Geschlechts in keinem Fall (N = 2 zu bewertende Studien) und bezüglich der Erkrankungsschwere in einem Fall (50 %) statistisch signifikant. Für das Alter (N = 2 zu bewertende Studien) sind die Diskrepanzen infolge des Fehlens

notwendiger Angaben nicht analog auswertbar. In der Zusammenschau sind die Patienten in den zu bewertenden Studien zu einem geringen Grade älter, der Anteil Frauen ist tendenziell höher und die Leistungsfähigkeit ist ohne Ausnahme günstiger (vgl. Anhang F-V, Abb. F22-F24).

Kongruenzlevel von Standardreferenzen: *FACT-Lym*

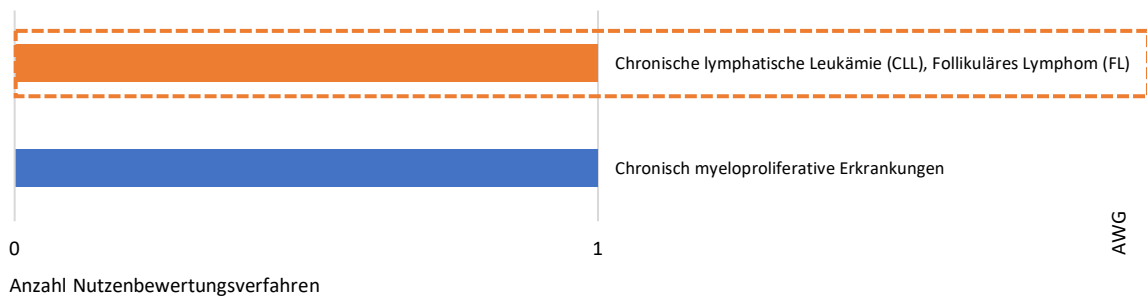
Bei Verwendung des *FACT-Lym* wird in 2 Nutzenbewertungsverfahren (67 %) auf die Validierungsstudie „*The Development and Validation of a Measure of Health-Related Quality of Life for Non-Hodgkin's Lymphoma: The Functional Assessment of Cancer Therapy-Lymphoma (FACTLym)*“ von Hlubocky et al. [72] aus dem Jahr 2013 (im Folgenden: **Hlubocky et al., 2013**) referenziert. Anders als die weiterhin benannten Referenzen (nicht weiter betrachtet) wird Hlubocky et al., 2013 sowohl im Anwendungsgebiet Leukämie, Lymphom als auch bei den myeloproliferativen Erkrankungen zum Nachweis der Validität des *FACT-Lym* angeführt (vgl. Abb. 25).



Anmerkung: Verweise bereinigt um systematische Reviews (u. dgl.).
N = 3 Nutzenbewertungsverfahren.

Abb. 25: Ermittlung der Standardreferenzen; hier: *FACT-Lym*
(Quelle: Eigene Darstellung).

Was die Behandlungskontexte anbelangt, gibt es mit dem Non-Hodgkin-Lymphom (NHL) eine Übereinstimmung (50 %) zwischen zu bewertender Studie und Validierungsstudie. Dabei bezieht Hlubocky et al., 2013 in die Entwicklung und Validierung nicht nur einzelne Formen von malignen Lymphomen ein (vgl. Abb. 26).



Hlubocky et al., 2013: *non-Hodgkin's lymphoma (NHL)* [72].

AWG: Anwendungsgebiet. N = 2 Nutzenbewertungsverfahren.

Abb. 26: Entsprechung nach onkologischer Entität; hier: Hlubocky et al., 2013
(Quelle: Eigene Darstellung).

Die relativen Abweichungen in der Gegenüberstellung von zu bewertender Studie und Hlubocky et al., 2013 liegen in der Spitze bei 7 Prozent für das Alter, bei 23 Prozent für den Anteil Frauen, bei 46 Prozent für den *ECOG PS* gw. und bei 18 Prozent für den Anteil Kaukasier/weiß. Die Differenz zur Validierungsstudie mit Blick auf den Anteil (West-)Europa beläuft sich überdies auf 39 Prozent⁸. Hinsichtlich der als Variablen einbezogenen Patientencharakteristika zeigt sich in keinem Fall beim Geschlecht (N = 2 zu bewertende Studien), in je 2 Fällen (je 100 %) beim Grad der Erkrankungsschwere und bei der geographischen Region sowie in einem Fall (50 %) bei der Ethnizität ein statistisch signifikanter Gruppenunterschied. Nicht auf diesem Wege auswertbar waren schließlich die Unterschiede beim Alter (N = 2 zu bewertende Studien). Insgesamt sind die Patienten in den zu bewertenden Studien entweder jünger oder älter, zu einem geringeren Anteil Frauen und entweder deutlich gesünder oder deutlich weniger leistungsfähig. Des Weiteren ist der Anteil Kaukasier/weiß einerseits und der Anteil (West-)Europa andererseits in den zu bewertenden Studien durchweg höher (vgl. Anhang F-VI, Abb. F25-F29).

⁸ Unberücksichtigt der Studie, in welcher gänzlich nicht in (West-)Europa rekrutiert wurde.

Erhebungsinstrument	EORTC QLQ-C30							
Kategoriale Einordnung	generisch	-						
	krankheitsspezifisch	zutreffend (onkologiespezifisch-generisch)						
Art und Umfang der Bezugnahme			Nutzenbewertungsverfahren					
	Referenzstudien	gesamt	N = 28					
		unberücksichtigt	N ≤ 7					
	Klassifikation als Standardreferenz	Aaronson et al., 1993		N = 22 (79 %)				
Osoba et al., 1998		N = 21 (75 %)						
Grad an Übereinstimmung im Verhältnis von Nutzenbewertungsverfahren zu Standardreferenz	Tumorentität	partiell	Aaronson et al., 1993					
		fehlend	Osoba et al., 1998					
			<u>Logik:</u> - Metrische bzw. metrisch interpretierte Variablen (Alter, Grad der Erkrankungsschwere): Nutzenbewertungsverfahren (ggf. zu bewertende Studie): <i>Mean, SD, N; p-Wert</i> (ggf. Subgruppe) - Nominalskalierte Variablen (Geschlecht, geographische Region): Nutzenbewertungsverfahren (ggf. zu bewertende Studie): N weiblich bzw. (West-)Europa, N männlich bzw. abweichend; p-Wert (ggf. Subgruppe)					
			Aaronson et al., 1993 [64]			Osoba et al., 1998 [61]		
	Patientencharakteristika	Alter (J.: Jahre)	Merkmalsbeschreibung (Standardreferenzen)	<i>Mean</i>	63 J.	55,2 J.	<i>Breast Cancer</i>	57,6 J.
<u>Einschluss:</u> Zu bewertende Studien, deren zugehöriges Nutzenbewertungsverfahren eine der Standardreferenzen für einen Validitätsbeleg anführt.	<i>SD</i>			10 J.	9,9 J.	7,7 J.		
	<i>N</i>			305	246	111		
		statistisch signifikanter Gruppenunterschied	N = 15 (79 %)	- Dabrafenib: 53,0 J. (gw.), 13,9 J. (gepoolt), 250; p < 0,0001 - Regorafenib: 60,7 J. (gw.), 10,0 J. (gepoolt), 760; p = 0,0007	N = 11 (65 %)	- Dabrafenib: 53,0 J. (gw.), 13,9 J. (gepoolt), 250; p = 0,0432 (<i>Breast Cancer</i>), p = 0,0012 (<i>SCLC</i>) - Regorafenib: 60,7 J. (gw.), 10,0 J. (gepoolt), 760; p < 0,0001 (<i>Breast Cancer</i>), p = 0,0018 (<i>SCLC</i>)		

<p>Grad an Übereinstimmung im Verhältnis von Nutzenbewertungsverfahren zu Standardreferenz</p>	<p>Patientencharakteristika Einschluss: Zu bewertende Studien, deren zugehöriges Nutzenbewertungsverfahren eine der Standardreferenzen für einen Validitätsbeleg anführt.</p>			<ul style="list-style-type: none"> - Afatinib (LUX-Lung 3): 60,2 J. (gw.), 10,1 J. (gepoolt), 345; p = 0,0004 - Afatinib (LUX-Lung 6): 56,3 J. (gw.), 10,8 J. (gepoolt), 364; p < 0,0001 - Obinutuzumab: 71,7 J. (gw.), 8,7 J. (gepoolt), 663; p < 0,0001 - Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) (PCYC-1112-CA): 66,5 J. (gw.), 9,5 J. (gepoolt), 391; p < 0,0001 - Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) (PCYC-1104-CA): 67,1 J., 8,6 J., 111; p < 0,0001 - Nintedanib (nicht-kleinzelliges Lungenkarzinom): 58,6 J. (gw.), 9,8 J. (gepoolt), 658; p < 0,0001 - Ruxolitinib (Polycythaemia vera): 60,1 J. (gw.), 10,4 J. (gepoolt), 222; p = 0,0013 - Ceritinib: 51,2 J., 11,6 J., 140; p < 0,0001 - Pembrolizumab (KEYNOTE 002): 60,0 J. (gw.), 13,8 J. (gepoolt), 359; p = 0,0017 - Pembrolizumab (KEYNOTE 006): 60,5 J. (gw.), 13,9 J. (gepoolt), 555; p = 0,0057 - Dabrafenib (neues Anwendungsgebiet): 54,2 J. (gw.), 13,9 J. (gepoolt), 704; p < 0,0001 - Trametinib: 53,7 J. (gw.), 13,3 J. (gepoolt), 322; p < 0,0001 - Cobimetinib: 55,1 J. (gw.), 13,9 J. (gepoolt), 495; p < 0,0001 	<ul style="list-style-type: none"> - Afatinib (LUX-Lung 3): 60,2 J. (gw.), 10,1 J. (gepoolt), 345; p < 0,0001 (<i>Breast Cancer</i>), p = 0,0132 (<i>SCLC</i>) - Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) (PCYC-1112-CA): 66,5 J. (gw.), 9,5 J. (gepoolt), 391; p < 0,0001 (<i>Breast Cancer</i>, <i>SCLC</i>) - Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) (PCYC-1104-CA): 67,1 J., 8,6 J., 111; p < 0,0001 (<i>Breast Cancer</i>, <i>SCLC</i>) - Nintedanib (nicht-kleinzelliges Lungenkarzinom): 58,6 J. (gw.), 9,8 J. (gepoolt), 658; p < 0,0001 (<i>Breast Cancer</i>) - Ruxolitinib (Polycythaemia vera): 60,1 J. (gw.), 10,4 J. (gepoolt), 222; p < 0,0001 (<i>Breast Cancer</i>), p = 0,0256 (<i>SCLC</i>) - Ceritinib: 51,2 J., 11,6 J., 140; p = 0,0004 (<i>Breast Cancer</i>), p < 0,0001 (<i>SCLC</i>) - Pembrolizumab (KEYNOTE 002): 60,0 J. (gw.), 13,8 J. (gepoolt), 359; p < 0,0001 (<i>Breast Cancer</i>) - Pembrolizumab (KEYNOTE 006): 60,5 J. (gw.), 13,9 J. (gepoolt), 555; p < 0,0001 (<i>Breast Cancer</i>), p = 0,0333 (<i>SCLC</i>) - Dabrafenib (neues Anwendungsgebiet): 54,2 J. (gw.), 13,9 J. (gepoolt), 704; p = 0,0121 (<i>SCLC</i>) - Panobinostat: 62,1 J. (gw.), 9,3 J. (gepoolt), 768; p < 0,0001 (<i>Breast Cancer</i>, <i>SCLC</i>) - Trametinib: 53,7 J. (gw.), 13,3 J. (gepoolt), 322; p = 0,0036 (<i>SCLC</i>)
				<p>Gesamtanzahl zu bewertender Studien</p>	<p>N = 25 (davon 19 inferentiell auswertbar)</p>

Grad an Übereinstimmung im Verhältnis von Nutzenbewertungsverfahren zu Standardreferenz	Patientencharakteristika <u>Einschluss:</u> Zu bewertende Studien, deren zugehöriges Nutzenbewertungsverfahren eine der Standardreferenzen für einen Validitätsbeleg anführt.	Geschlecht	Merkmalsbeschreibung (Standardreferenzen)	N weiblich 73	246	<i>Breast Cancer</i>	nicht berichtet	<i>SCLC</i>
				N männlich 232	0			
			statistisch signifikanter Gruppenunterschied	N = 22 (88 %)	N = 26 (100 %)	<ul style="list-style-type: none"> - Ruxolitinib (Chronisch myeloproliferative Erkrankungen) (COMFORT I): 141, 167; p < 0,0001 - Ruxolitinib (Chronisch myeloproliferative Erkrankungen) (COMFORT II): 94, 125; p < 0,0001 - Pomalidomid: 187, 268; p < 0,0001 - Dabrafenib: 101, 149; p < 0,0001 - Regorafenib: 296, 464; p < 0,0001 - Afatinib (LUX-Lung 3): 224, 121; p < 0,0001 - Afatinib (LUX-Lung 6): 238, 126; p < 0,0001 - Eribulin (neues Anwendungsgebiet, Neubewertung nach Fristablauf): 882, 0; p < 0,0001 - Obinutuzumab: 256, 407; p < 0,0001 - Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) (PCYC-1112-CA): 125, 266; p = 0,0222 - Nintedanib (nicht-kleinzelliges Lungenkarzinom): 247, 411; p < 0,0001 - Ruxolitinib (Polycythaemia vera): 76, 146; p = 0,0109 - Ceritinib: 70, 70; p < 0,0001 - Nivolumab (Melanom) (CA209-066): 172, 246; p < 0,0001 - Nivolumab (Melanom) (CA209-067): 149, 285; p = 0,0026 - Pembrolizumab (KEYNOTE 002): 141, 218; p < 0,0001 	<ul style="list-style-type: none"> - Ruxolitinib (Chronisch myeloproliferative Erkrankungen) (COMFORT I): 141, 167; p < 0,0001 - Ruxolitinib (Chronisch myeloproliferative Erkrankungen) (COMFORT II): 94, 125; p < 0,0001 - Decitabin: 197, 288; p < 0,0001 - Crizotinib (PROFILE 1007): 194, 153; p < 0,0001 - Crizotinib (PROFILE 1005): 233, 206; p < 0,0001 - Dabrafenib: 101, 149; p < 0,0001 - Regorafenib: 296, 464; p < 0,0001 - Afatinib (LUX-Lung 3): 224, 121; p < 0,0001 - Afatinib (LUX-Lung 6): 238, 126; p < 0,0001 - Ipilimumab (neues Anwendungsgebiet) (CA184-022): 73, 144; p < 0,0001 - Ipilimumab (neues Anwendungsgebiet) (CA184-024): 201, 301; p < 0,0001 - Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) (PCYC-1112-CA): 125, 266; p < 0,0001 - Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) (PCYC-1104-CA): 26, 85; p < 0,0001 - Nintedanib (nicht-kleinzelliges Lungenkarzinom): 247, 411; p < 0,0001 	

Grad an Übereinstimmung im Verhältnis von Nutzenbewertungsverfahren zu Standardreferenz	Patientencharakteristika <u>Einschluss:</u> Zu bewertende Studien, deren zugehöriges Nutzenbewertungsverfahren eine der Standardreferenzen für einen Validitätsbeleg anführt.				<ul style="list-style-type: none"> - Pembrolizumab (KEYNOTE 006): 219, 336; p < 0,0001 - Dabrafenib (neues Anwendungsgebiet): 316, 388; p < 0,0001 - Panobinostat: 361, 407; p < 0,0001 - Trametinib: 149, 173; p < 0,0001 - Carfilzomib: 345, 447; p < 0,0001 - Cobimetinib: 209, 286; p < 0,0001 	<ul style="list-style-type: none"> - Ramucirumab (RAINBOW): 193, 472; p < 0,0001 - Ramucirumab (REGARD): 107, 248; p < 0,0001 - Ruxolitinib (Polycythaemia vera): 76, 146; p < 0,0001 - Ceritinib: 70, 70; p < 0,0001 - Pembrolizumab (KEYNOTE 002): 141, 218; p < 0,0001 - Pembrolizumab (KEYNOTE 006): 219, 336; p < 0,0001 - Dabrafenib (neues Anwendungsgebiet): 316, 388; p < 0,0001 - Panobinostat: 361, 407; p < 0,0001 - Trametinib: 149, 173; p < 0,0001 - Carfilzomib: 345, 447; p < 0,0001 - Cobimetinib: 209, 286; p < 0,0001 - Crizotinib (neues Anwendungsgebiet): 212, 131; p < 0,0001 														
					Gesamtanzahl zu bewertender Studien	N = 25	N = 26													
					Merkmalsbeschreibung (Standardreferenzen)	<table border="1"> <tr> <td><i>Mean</i></td> <td>1,143</td> <td>1,012</td> <td rowspan="3"> <i>Breast Cancer</i> </td> <td>0,809</td> <td rowspan="3"> <i>SCLC</i> </td> </tr> <tr> <td><i>SD</i></td> <td>0,825</td> <td>0,695</td> <td>0,595</td> </tr> <tr> <td><i>N</i></td> <td>305</td> <td>246</td> <td>111</td> </tr> </table>	<i>Mean</i>	1,143	1,012	<i>Breast Cancer</i>	0,809	<i>SCLC</i>	<i>SD</i>	0,825	0,695	0,595	<i>N</i>	305	246	111
					<i>Mean</i>	1,143	1,012	<i>Breast Cancer</i>	0,809		<i>SCLC</i>									
<i>SD</i>	0,825	0,695	0,595																	
<i>N</i>	305	246	111																	
Grad der Erkrankungsschwere (Operationalisierung: ECOG PS gw.)	statistisch signifikanter Gruppenunterschied	N = 24 (100 %)	<ul style="list-style-type: none"> - Ruxolitinib (Chronisch myeloproliferative Erkrankungen) (COMFORT I): 0,894 (gw.), 0,700 (gepoolt), 309; p < 0,0001 - Ruxolitinib (Chronisch myeloproliferative Erkrankungen) (COMFORT II): 0,721 (gw.), 0,653 (gepoolt), 219; p < 0,0001 - Pomalidomid: 0,850 (gw.), 0,705 (gepoolt), 455; p < 0,0001 - Dabrafenib: 0,310 (gw.), 0,464 (gepoolt), 250; p < 0,0001 	<ul style="list-style-type: none"> - Ruxolitinib (Chronisch myeloproliferative Erkrankungen) (COMFORT I): 0,894 (gw.), 0,700 (gepoolt), 309; p = 0,0483 (<i>Breast Cancer</i>) - Ruxolitinib (Chronisch myeloproliferative Erkrankungen) (COMFORT II): 0,721 (gw.), 0,653 (gepoolt), 219; p < 0,0001 (<i>Breast Cancer</i>) - Decitabin: 1,074 (gw.), 0,660 (gepoolt), 485; p < 0,0001 (<i>SCLC</i>) 																

<p>Grad an Übereinstimmung im Verhältnis von Nutzenbewertungsverfahren zu Standardreferenz</p>	<p>Patientencharakteristika <u>Einschluss:</u> Zu bewertende Studien, deren zugehöriges Nutzenbewertungsverfahren eine der Standardreferenzen für einen Validitätsbeleg anführt</p>			<ul style="list-style-type: none"> - Regorafenib: 0,459 (gw.), 0,498 (gepoolt), 760; p < 0,0001 - Afatinib (LUX-Lung 3): 0,618 (gw.), 0,491 (gepoolt), 345; p < 0,0001 - Afatinib (LUX-Lung 6): 0,756 (gw.), 0,425 (gepoolt), 364; p < 0,0001 - Eribulin (neues Anwendungsgebiet, Neubewertung nach Fristablauf): 0,626 (gw.), 0,539 (gepoolt), 882; p < 0,0001 - Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) (PCYC-1112-CA): 0,593 (gw.), 0,491 (gepoolt), 391; p < 0,0001 - Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) (PCYC-1104-CA): 0,630, 0,691, 111; p < 0,0001 - Nintedanib (nicht-kleinzelliges Lungenkarzinom): 0,704 (gw.), 0,457 (gepoolt), 658; p < 0,0001 - Ramucirumab (RAINBOW): 0,607 (gw.), 0,487 (gepoolt), 665; p < 0,0001 - Ramucirumab (REGARD): 0,727 (gw.), 0,452 (gepoolt), 355; p < 0,0001 - Ruxolitinib (Polycythaemia vera): 0,329 (gw.), 0,507 (gepoolt), 222; p < 0,0001 - Ceritinib: 0,843, 0,647, 140; p = 0,0002 - Nivolumab (Melanom) (CA209-066): 0,364 (gw.), 0,496 (gepoolt), 418; p < 0,0001 - Nivolumab (Melanom) (CA209-067): 0,302 (gw.), 0,464 (gepoolt), 434; p < 0,0001 - Pembrolizumab (KEYNOTE 002): 0,445 (gw.), 0,497 (gepoolt), 359; p < 0,0001 	<ul style="list-style-type: none"> - Crizotinib (PROFILE 1007): 0,688 (gw.), 0,622 (gepoolt), 347; p < 0,0001 (<i>Breast Cancer</i>) - Dabrafenib: 0,310 (gw.), 0,464 (gepoolt), 250; p < 0,0001 (<i>Breast Cancer, SCLC</i>) - Regorafenib: 0,459 (gw.), 0,498 (gepoolt), 760; p < 0,0001 (<i>Breast Cancer, SCLC</i>) - Afatinib (LUX-Lung 3): 0,618 (gw.), 0,491 (gepoolt), 345; p < 0,0001 (<i>Breast Cancer</i>), p = 0,0008 (<i>SCLC</i>) - Afatinib (LUX-Lung 6): 0,756 (gw.), 0,425 (gepoolt), 364; p < 0,0001 (<i>Breast Cancer</i>) - Ipilimumab (neues Anwendungsgebiet) (CA184-022): 0,392, 0,499, 217; p < 0,0001 (<i>Breast Cancer, SCLC</i>) - Ipilimumab (neues Anwendungsgebiet) (CA184-024): 0,291 (gw.), 0,454 (gepoolt), 502; p < 0,0001 (<i>Breast Cancer, SCLC</i>) - Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) (PCYC-1112-CA): 0,593 (gw.), 0,491 (gepoolt), 391; p < 0,0001 (<i>Breast Cancer, SCLC</i>) - Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) (PCYC-1104-CA): 0,630, 0,691, 111; p < 0,0001 (<i>Breast Cancer</i>), p = 0,0398 (<i>SCLC</i>) - Nintedanib (nicht-kleinzelliges Lungenkarzinom): 0,704 (gw.), 0,457 (gepoolt), 658; p < 0,0001 (<i>Breast Cancer</i>), p = 0,0331 (<i>SCLC</i>) - Ramucirumab (RAINBOW): 0,607 (gw.), 0,487 (gepoolt), 665; p < 0,0001 (<i>Breast Cancer, SCLC</i>)
---	---	--	--	--	---

Grad an Übereinstimmung im Verhältnis von Nutzenbewertungsverfahren zu Standardreferenz	Patientencharakteristika <u>Einschluss:</u> Zu bewertende Studien, deren zugehöriges Nutzenbewertungsverfahren eine der Standardreferenzen für einen Validitätsbeleg anführt			<ul style="list-style-type: none"> - Pembrolizumab (KEYNOTE 006): 0,321 (gw.), 0,467 (gepoolt), 555; p < 0,0001 - Dabrafenib (neues Anwendungsgebiet): 0,295 (gw.), 0,455 (gepoolt), 704; p < 0,0001 - Panobinostat: 0,616 (gw.), 0,606 (gepoolt), 768; p < 0,0001 - Trametinib: 0,360 (gw.), 0,481 (gepoolt), 322; p < 0,0001 - Carfilzomib: 0,665 (gw.), 0,642 (gepoolt), 792; p < 0,0001 - Cobimetinib: 0,288 (gw.), 0,455 (gepoolt), 495; p < 0,0001 	<ul style="list-style-type: none"> - Ramucirumab (REGARD): 0,727 (gw.), 0,452 (gepoolt), 355; p < 0,0001 (<i>Breast Cancer</i>) - Ruxolitinib (Polycythaemia vera): 0,329 (gw.), 0,507 (gepoolt), 222; p < 0,0001 (<i>Breast Cancer, SCLC</i>) - Ceritinib: 0,843, 0,647, 140; p = 0,0191 (<i>Breast Cancer</i>) - Pembrolizumab (KEYNOTE 002): 0,445 (gw.), 0,497 (gepoolt), 359; p < 0,0001 (<i>Breast Cancer, SCLC</i>) - Pembrolizumab (KEYNOTE 006): 0,321 (gw.), 0,467 (gepoolt), 555; p < 0,0001 (<i>Breast Cancer, SCLC</i>) - Dabrafenib (neues Anwendungsgebiet): 0,295 (gw.), 0,455 (gepoolt), 704; p < 0,0001 (<i>Breast Cancer, SCLC</i>) - Panobinostat: 0,616 (gw.), 0,606 (gepoolt), 768; p < 0,0001 (<i>Breast Cancer</i>), p = 0,0017 (<i>SCLC</i>) - Trametinib: 0,360 (gw.), 0,481 (gepoolt), 322; p < 0,0001 (<i>Breast Cancer, SCLC</i>) - Carfilzomib: 0,665 (gw.), 0,642 (gepoolt), 792; p < 0,0001 (<i>Breast Cancer</i>), p = 0,0258 (<i>SCLC</i>) - Cobimetinib: 0,288 (gw.), 0,455 (gepoolt), 495; p < 0,0001 (<i>Breast Cancer, SCLC</i>) - Crizotinib (neues Anwendungsgebiet): 0,740 (gw.), 0,534 (gepoolt), 343; p < 0,0001 (<i>Breast Cancer</i>)
				Gesamtanzahl zu bewertender Studien	N = 24

Grad an Übereinstimmung im Verhältnis von Nutzenbewertungsverfahren zu Standardreferenz	Patientencharakteristika <u>Einschluss:</u> Zu bewertende Studien, deren zugehöriges Nutzenbewertungsverfahren eine der Standardreferenzen für einen Validitätsbeleg anführt.	Ethnizität	Merkmalsbeschreibung (Standardreferenzen)	N Kaukasier/weiß N abweichend	nicht auswertbar (relevante Angaben dazu fehlen in der Referenzstudie)	nicht auswertbar (relevante Angaben dazu fehlen in der Referenzstudie)
		geographische Region	Merkmalsbeschreibung (Standardreferenzen)	N (West-) Europa N abweichend	212 93	nicht auswertbar (relevante Angaben dazu fehlen in der Referenzstudie)
			statistisch signifikanter Gruppenunterschied	N = 8 (53 %)	<ul style="list-style-type: none"> - Ruxolitinib (Chronisch myeloproliferative Erkrankungen) (COMFORT I): 0, 309; p < 0,0001 - Ruxolitinib (Chronisch myeloproliferative Erkrankungen) (COMFORT II): 219, 0; p < 0,0001 - Afatinib (LUX-Lung 6): 0, 364; p < 0,0001 - Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) (PCYC-1112-CA): 170, 221; p < 0,0001 - Ramucirumab (RAINBOW): 303, 362; p < 0,0001 - Ramucirumab (REGARD): 172, 183; p < 0,0001 - Ceritinib: 57, 83; p < 0,0001 - Nivolumab (Melanom) (CA209-067): 221, 213; p < 0,0001 	
			Gesamtanzahl zu bewertender Studien	N = 15		

Erläuterungen: gw.: gewichtetes arithmetisches Mittel, gepoolt: gepoolte Standardabweichung. Berechnung unter Einbezug der Zahl an (randomisierten) Patienten je Therapiearm (Interventions-/Kontrollgruppe) und der jeweiligen Mittelwerte bzw. der jeweiligen Standardabweichungen einer betreffenden Merkmalsbeschreibung. *ECOG PS* gw.: Häufigkeitsgewichtete Maßzahl zum *Eastern Cooperative Oncology Group Performance Status*, *Mean*: arithmetisches Mittel, *SCLC*: Kleinzelliges Lungenkarzinom, *SD*: Standardabweichung, *N*: wahlweise Zahl der in die Auswertung einbezogenen Patienten bzw. Anzahl Nutzenbewertungsverfahren, Anzahl zu bewertender Studien oder Anzahl paarweiser Vergleiche.

Tabelle 2: Zusammenschau bezüglich der Standardreferenzen; hier: EORTC QLQ-C30
(Quelle: Eigene Darstellung).

Erhebungsinstrument	<i>EQ-5D (VAS)</i>				
Kategoriale Einordnung	generisch	zutreffend			
	krankheitsspezifisch	-			
Art und Umfang der Bezugnahme				Nutzenbewertungsverfahren	
	Referenzstudien	gesamt		N = 18	
		unberücksichtigt		N ≤ 6	
Klassifikation als Standardreferenz	Rabin und de Charro, 2001 [66]		N = 10 (56 %)		
Grad an Übereinstimmung im Verhältnis von Nutzenbewertungsverfahren zu Standardreferenz	Tumorentität	fehlend			
				<u>Logik:</u> - Metrische Variable (Alter): Nutzenbewertungsverfahren (ggf. zu bewertende Studie): <i>Mean, SD, N; p-Wert</i> - Nominalskalierte Variablen (Geschlecht, geographische Region): Nutzenbewertungsverfahren (ggf. zu bewertende Studie): N weiblich bzw. (West-)Europa, N männlich bzw. abweichend; p-Wert	
	Patientencharakteristika <u>Einschluss:</u> Zu bewertende Studien, deren zugehöriges Nutzenbewertungsverfahren die Standardreferenz für einen Validitätsbeleg anführt.	Alter (J.: Jahre)	Merkmalsbeschreibung (Standardreferenz)	<i>Mean</i>	56 J.
				<i>SD</i>	14 J.
		statistisch signifikanter Gruppenunterschied	N = 9 (82 %)	- Bosutinib: 51,3 J. (gw.), 14,6 J. (gepoolt), 546; p < 0,0001 - Dabrafenib: 53,0 J. (gw.), 13,9 J. (gepoolt), 250; p = 0,0186 - Afatinib (LUX-Lung 3): 60,2 J. (gw.), 10,1 J. (gepoolt), 345; p < 0,0001 - Idelalisib (GS-US-312-0116): 70,5 J. (gw.), 7,9 J. (gepoolt), 220; p < 0,0001 - Idelalisib (101-09 [Subgruppe: Follikuläres Lymphom-Patienten]): 61 J., 11,9 J., 72; p = 0,0065	

Grad an Übereinstimmung im Verhältnis von Nutzenbewertungsverfahren zu Standardreferenz	Patientencharakteristika <u>Einschluss:</u> Zu bewertende Studien, deren zugehöriges Nutzenbewertungsverfahren die Standardreferenz für einen Validitätsbeleg anführt.				<ul style="list-style-type: none"> - Nintedanib (nicht-kleinzelliges Lungenkarzinom): 58,6 J. (gw.), 9,8 J. (gepoolt), 658; p = 0,0021 - Pembrolizumab (KEYNOTE 002): 60,0 J. (gw.), 13,8 J. (gepoolt), 359; p = 0,0007 - Pembrolizumab (KEYNOTE 006): 60,5 J. (gw.), 13,9 J. (gepoolt), 555; p < 0,0001 - Trametinib: 53,7 J. (gw.), 13,3 J. (gepoolt), 322; p = 0,0497 		
		Gesamtanzahl zu bewertender Studien		N = 13 (davon 11 inferentiell auswertbar)			
		Merkmalsbeschreibung (Standardreferenz)	N weiblich	188			
			N männlich	45			
		statistisch signifikanter Gruppenunterschied		N = 13 (100 %)	<ul style="list-style-type: none"> - Bosutinib: 135, 153; p < 0,0001 - Dabrafenib: 101, 149; p < 0,0001 - Afatinib (LUX-Lung 3): 224, 121; p < 0,0001 - Afatinib (LUX-Lung 6): 238, 126; p < 0,0001 - Idelalisib (GS-US-312-0116): 76, 144; p < 0,0001 - Idelalisib (101-09 [Subgruppe: Follikuläres Lymphom-Patienten]): 33, 39; p < 0,0001 - Nintedanib (nicht-kleinzelliges Lungenkarzinom): 247, 411; p < 0,0001 - Ramucirumab (RAINBOW): 193, 472; p < 0,0001 - Ramucirumab (REGARD): 107, 248; p < 0,0001 - Pembrolizumab (KEYNOTE 002): 141, 218; p < 0,0001 - Pembrolizumab (KEYNOTE 006): 219, 336; p < 0,0001 - Dabrafenib (neues Anwendungsgebiet): 316, 388; p < 0,0001 - Trametinib: 149, 173; p < 0,0001 		
		Gesamtanzahl zu bewertender Studien		N = 13			
Geschlecht							

Grad an Übereinstimmung im Verhältnis von Nutzenbewertungsverfahren zu Standardreferenz	Patientencharakteristika <u>Einschluss:</u> Zu bewertende Studien, deren zugehöriges Nutzenbewertungsverfahren die Standardreferenz für einen Validitätsbeleg anführt.	Grad der Erkrankungsschwere	Merkmalsbeschreibung (Standardreferenz)	<i>Mean</i>	nicht auswertbar (relevante Angaben dazu fehlen in der Referenzstudie)
				<i>SD</i>	
				N	
		Ethnizität	Merkmalsbeschreibung (Standardreferenz)	N Kaukasier/weiß	nicht auswertbar (relevante Angaben dazu fehlen in der Referenzstudie)
				N abweichend	
		geographische Region	Merkmalsbeschreibung (Standardreferenz)	N (West-) Europa	233
				N abweichend	0
statistisch signifikanter Gruppenunterschied	N = 8 (100 %)		<ul style="list-style-type: none"> - Afatinib (LUX-Lung 6): 0, 364; p < 0,0001 - Idelalisib (GS-US-312-0116): 57, 163; p < 0,0001 - Idelalisib (101-09 [Subgruppe: Follikuläres Lymphom-Patienten]): 28, 44; p < 0,0001 - Nintedanib (nicht-kleinzelliges Lungenkarzinom): 463, 195; p < 0,0001 - Ramucirumab (RAINBOW): 303, 362; p < 0,0001 - Ramucirumab (REGARD): 172, 183; p < 0,0001 - Dabrafenib (neues Anwendungsgebiet): 464, 240; p < 0,0001 - Trametinib: 219, 103; p < 0,0001 		
Gesamtanzahl zu bewertender Studien	N = 8				

Erläuterungen: gw.: gewichtetes arithmetisches Mittel, gepoolt: gepoolte Standardabweichung. Berechnung unter Einbezug der Zahl an (randomisierten) Patienten je Therapiearm (Interventions-/Kontrollgruppe) und der jeweiligen Mittelwerte bzw. der jeweiligen Standardabweichungen einer betreffenden Merkmalsbeschreibung.

Mean: arithmetisches Mittel, *SD:* Standardabweichung, N: wahlweise Zahl der in die Auswertung einbezogenen Patienten bzw. Anzahl Nutzenbewertungsverfahren, Anzahl zu bewertender Studien oder Anzahl paarweiser Vergleiche. Anmerkung: Rabin und de Charro, 2001 [66] führt im Wesentlichen die Ergebnisse aus Hurst et al., 1997 [67] an.

Tabelle 3: Zusammenschau bezüglich der Standardreferenzen; hier: EQ-5D (VAS)
 (Quelle: Eigene Darstellung).

Erhebungsinstrument	FACT-P						
Kategoriale Einordnung	generisch	-					
	krankheitsspezifisch	zutreffend					
Art und Umfang der Bezugnahme				Nutzenbewertungsverfahren			
	Referenzstudien	gesamt		N = 5			
		unberücksichtigt		N ≤ 4			
	Klassifikation als Standardreferenz	Esper et al., 1997		N = 5 (100 %)			
Cella et al., 2009		N = 5 (100 %)					
Grad an Übereinstimmung im Verhältnis von Nutzenbewertungsverfahren zu Standardreferenz	Tumorentität	umfassend	Esper et al., 1997				
			Cella et al., 2009				
				<u>Logik:</u> - Metrische Variable (Alter): Nutzenbewertungsverfahren: Mean, SD, N; p-Wert - Nominalskalierte Variablen (Ethnizität, geographische Region): Nutzenbewertungsverfahren: N Kaukasier/weiß bzw. (West-)Europa, N abweichend; p-Wert			
				Esper et al., 1997 [68]		Cella et al., 2009 [69]	
	Patientencharakteristika	Alter (J.: Jahre)	Merkmalsbeschreibung (Standardreferenzen)	Mean	66,3 J. (gw.)		71,8 J.
				SD	8,7 J. (gepoolt)		8,2 J.
N				130		809	
statistisch signifikanter Gruppenunterschied		N = 4 (100 %)	- Abirateronacetat (neues Anwendungsgebiet): 70,5 J. (gw.), 9,0 J. (gepoolt), 1088; p < 0,0001 - Enzalutamid: 68,7 J. (gw.), 8,1 J. (gepoolt), 1199, p = 0,0015 - Radium-223-dichlorid: 70,3 J. (gw.), 8,0 J. (gepoolt), 921; p < 0,0001	N = 3 (75 %)	- Abirateronacetat (neues Anwendungsgebiet): 70,5 J. (gw.), 9,0 J. (gepoolt), 1088; p = 0,0013 - Enzalutamid: 68,7 J. (gw.), 8,1 J. (gepoolt), 1199, p < 0,0001 - Radium-223-dichlorid: 70,3 J. (gw.), 8,0 J. (gepoolt), 921; p = 0,0001		
<u>Einschluss:</u> Zu bewertende Studien, deren zugehöriges Nutzenbewertungsverfahren eine der Standardreferenzen für einen Validitätsbeleg anführt.							

Grad an Übereinstimmung im Verhältnis von Nutzenbewertungsverfahren zu Standardreferenz	Patientencharakteristika <u>Einschluss:</u> Zu bewertende Studien, deren zugehöriges Nutzenbewertungsverfahren eine der Standardreferenzen für einen Validitätsbeleg anführt.				- Enzalutamid (neues Anwendungsgebiet): 71,3 J. (gw.), 8,5 J. (gepoolt), 1717; p < 0,0001		
			Gesamtanzahl zu bewertender Studien	N = 5 (davon 4 inferentiell auswertbar)		N = 5 (davon 4 inferentiell auswertbar)	
		Geschlecht	Merkmalsbeschreibung (Standardreferenzen)	N weiblich	n/a	n/a	
				N männlich			
		Grad der Erkrankungsschwere	Merkmalsbeschreibung (Standardreferenzen)	Mean	nicht auswertbar (relevante Angaben dazu fehlen in der Referenzstudie)	nicht inferentiell auswertbar (relevante Angaben dazu fehlen in der Referenzstudie)	
				SD			
				N			
		Ethnizität	Merkmalsbeschreibung (Standardreferenzen)	N Kaukasier/weiß	58	770	
				N abweichend	15	39	
			statistisch signifikanter Gruppenunterschied		N = 4 (80 %)	- Abirateronacetat: 1195, 0; p < 0,0001 - Abirateronacetat (neues Anwendungsgebiet): 1088, 0; p < 0,0001 - Enzalutamid: 1111, 88, p = 0,0004 - Radium-223-dichlorid: 865, 56; p < 0,0001	N = 4 (80 %) - Abirateronacetat: 1195, 0; p < 0,0001 - Abirateronacetat (neues Anwendungsgebiet): 1088, 0; p < 0,0001 - Enzalutamid: 1111, 88, p = 0,0247 - Enzalutamid (neues Anwendungsgebiet): 1322, 395; p < 0,0001
Gesamtanzahl zu bewertender Studien			N = 5		N = 5		

Grad an Übereinstimmung im Verhältnis von Nutzenbewertungsverfahren zu Standardreferenz	Patientencharakteristika <u>Einschluss:</u> Zu bewertende Studien, deren zugehöriges Nutzenbewertungsverfahren eine der Standardreferenzen für einen Validitätsbeleg anführt.	geographische Region	Merkmalsbeschreibung (Standardreferenzen)	N (West-Europa)	0		nicht auswertbar (relevante Angaben dazu fehlen in der Referenzstudie)
				N abweichend	173		
			statistisch signifikanter Gruppenunterschied	N = 1 (100 %)	- Enzalutamid (neues Anwendungsgebiet): 911, 806; p < 0,0001		
			Gesamtanzahl zu bewertender Studien	N = 1			

Erläuterungen: gw.: gewichtetes arithmetisches Mittel, gepoolt: gepoolte Standardabweichung. Berechnung unter Einbezug der Zahl an (randomisierten) Patienten je Therapiearm (Interventions-/Kontrollgruppe) und der jeweiligen Mittelwerte bzw. der jeweiligen Standardabweichungen einer betreffenden Merkmalsbeschreibung.

Mean: arithmetisches Mittel, *n/a:* nicht anwendbar, *SD:* Standardabweichung, *N:* wahlweise Zahl der in die Auswertung einbezogenen Patienten bzw. Anzahl Nutzenbewertungsverfahren, Anzahl zu bewertender Studien oder Anzahl paarweiser Vergleiche.

Tabelle 4: Zusammenschau bezüglich der Standardreferenzen; hier: *FACT-P*
(Quelle: Eigene Darstellung).

Erhebungsinstrument	FACT-M					
Kategoriale Einordnung	generisch	-				
	krankheitsspezifisch	zutreffend				
Art und Umfang der Bezugnahme				Nutzenbewertungsverfahren		
	Referenzstudien	gesamt		N = 4		
		unberücksichtigt		N = 2		
Klassifikation als Standardreferenz	Cormier et al., 2008 [70]		N = 4 (100 %)			
Grad an Übereinstimmung im Verhältnis von Nutzenbewertungsverfahren zu Standardreferenz	Tumorentität	umfassend				
				<u>Logik:</u> - Nominalskalierte Variable (geographische Region): Nutzenbewertungsverfahren: N (West-)Europa, N abweichend; p-Wert		
	Patientencharakteristika <u>Einschluss:</u> Zu bewertende Studien, deren zugehöriges Nutzenbewertungsverfahren die Standardreferenz für einen Validitätsbeleg anführt.	Alter (J.: Jahre)	Merkmalsbeschreibung (Standardreferenz)	Mean	nicht berichtet	
				SD	(ersatzweise: Median 52 J., Range 20-79 J.)	
				N	273	
		Geschlecht	Merkmalsbeschreibung (Standardreferenz)	N weiblich	114	
N männlich				159		
statistisch signifikanter Gruppenunterschied			in keinem Fall (0 %)			
Gesamtanzahl zu bewertender Studien		N = 3				
	Merkmalsbeschreibung	Mean	0,325			
		SD	0,521			

Grad an Übereinstimmung im Verhältnis von Nutzenbewertungsverfahren zu Standardreferenz	Patientencharakteristika <u>Einschluss:</u> Zu bewertende Studien, deren zugehöriges Nutzenbewertungsverfahren die Standardreferenz für einen Validitätsbeleg anführt.	Grad der Erkrankungsschwere (Operationalisierung: <i>ECOG PS gw.</i>)	(Standardreferenz)	N	273		
			statistisch signifikanter Gruppenunterschied		in keinem Fall (0 %)		
			Gesamtanzahl zu bewertender Studien		N = 3		
		Ethnizität	Merkmalsbeschreibung (Standardreferenz)	N Kaukasier /weiß	268		
				N abweichend	5		
			statistisch signifikanter Gruppenunterschied		in keinem Fall (0 %)		
			Gesamtanzahl zu bewertender Studien		N = 2		
		geographische Region	Merkmalsbeschreibung (Standardreferenz)	N (West-) Europa	0		
				N abweichend	273		
			statistisch signifikanter Gruppenunterschied		N = 3 (100 %)	- Vemurafenib: 408, 267; p < 0,0001 - Dabrafenib (neues Anwendungsgebiet): 464, 240; p < 0,0001 - Trametinib: 219, 103; p < 0,0001	
			Gesamtanzahl zu bewertender Studien		N = 3		

ECOG PS gw.: Häufigkeitsgewichtete Maßzahl zum *Eastern Cooperative Oncology Group Performance Status*, N: wahlweise Zahl der in die Auswertung einbezogenen Patienten bzw. Anzahl Nutzenbewertungsverfahren, Anzahl zu bewertender Studien oder Anzahl paarweiser Vergleiche.

Tabelle 5: Zusammenschau bezüglich der Standardreferenzen; hier: *FACT-M*
(Quelle: Eigene Darstellung).

Erhebungsinstrument	EORTC QLQ-MY20					
Kategoriale Einordnung	generisch	-				
	krankheitsspezifisch	zutreffend				
Art und Umfang der Bezugnahme		Nutzenbewertungsverfahren				
	Referenzstudien	gesamt	N = 3			
		unberücksichtigt	N = 1			
	Klassifikation als Standardreferenz	Cocks et al., 2007 [71]	N = 3 (100 %)			
Grad an Übereinstimmung im Verhältnis von Nutzenbewertungsverfahren zu Standardreferenz	Tumorentität	umfassend				
		<u>Logik:</u> - Metrisch interpretierte Variable (Grad der Erkrankungsschwere): Nutzenbewertungsverfahren: <i>Mean, SD, N</i> ; p-Wert				
	Patientencharakteristika <u>Einschluss:</u> Zu bewertende Studien, deren zugehöriges Nutzenbewertungsverfahren die Standardreferenz für einen Validitätsbeleg anführt.	Alter (J.: Jahre)	Merkmalsbeschreibung (Standardreferenz)	<i>Mean</i>	62,3 J.	nicht inferentiell auswertbar (relevante Angaben dazu fehlen in der Referenzstudie)
				<i>SD</i>	nicht berichtet (ersatzweise: Range 27-89 J.)	
				N	240	
	Geschlecht		Merkmalsbeschreibung (Standardreferenz)	N weiblich	99	
				N männlich	141	
				statistisch signifikanter Gruppenunterschied		in keinem Fall (0 %)
				Gesamtanzahl zu bewertender Studien		N = 2
			Merkmalsbeschreibung	<i>Mean</i>	0,971	
<i>SD</i>				0,995		

Grad an Übereinstimmung im Verhältnis von Nutzenbewertungsverfahren zu Standardreferenz	Patientencharakteristika <u>Einschluss:</u> Zu bewertende Studien, deren zugehöriges Nutzenbewertungsverfahren die Standardreferenz für einen Validitätsbeleg anführt.	Grad der Erkrankungsschwere (Operationalisierung: <i>ECOG PS</i> gw.)	(Standardreferenz)	N	240		
			statistisch signifikanter Gruppenunterschied		N = 1 (50 %)	- Panobinostat: 0,616 (gw.), 0,606 (gepoolt), 768; p < 0,0001	
			Gesamtanzahl zu bewertender Studien		N = 2		
		Ethnizität	Merkmalsbeschreibung (Standardreferenz)	N Kaukasier /weiß	nicht auswertbar (relevante Angaben dazu fehlen in der Referenzstudie)		
				N abweichend			
		geographische Region	Merkmalsbeschreibung (Standardreferenz)	N (West-) Europa	218	nicht auswertbar (relevante Angaben dazu fehlen in den Nutzenbewertungsverfahren)	
N abweichend	22						

Erläuterungen: gw.: gewichtetes arithmetisches Mittel, gepoolt: gepoolte Standardabweichung. Berechnung unter Einbezug der Zahl an (randomisierten) Patienten je Therapiearm (Interventions-/Kontrollgruppe) und der jeweiligen Mittelwerte bzw. der jeweiligen Standardabweichungen einer betreffenden Merkmalsbeschreibung.

ECOG PS gw.: Häufigkeitsgewichtete Maßzahl zum *Eastern Cooperative Oncology Group Performance Status*, *Mean*: arithmetisches Mittel, *SD*: Standardabweichung, N: wahlweise Zahl der in die Auswertung einbezogenen Patienten bzw. Anzahl Nutzenbewertungsverfahren, Anzahl zu bewertender Studien oder Anzahl paarweiser Vergleiche.

Tabelle 6: Zusammenschau bezüglich der Standardreferenzen; hier: *EORTC QLQ-MY20*
(Quelle: Eigene Darstellung).

Erhebungsinstrument	FACT-Lym						
Kategoriale Einordnung	generisch	-					
	krankheitsspezifisch	zutreffend					
Art und Umfang der Bezugnahme						Nutzenbewertungsverfahren	
	Referenzstudien	gesamt			N = 3		
		unberücksichtigt			N ≤ 2		
Klassifikation als Standardreferenz	Hlubocky et al., 2013 [72]			N = 2 (67 %)			
Grad an Übereinstimmung im Verhältnis von Nutzenbewertungsverfahren zu Standardreferenz	Tumorentität	partiell					
						<p><u>Logik:</u></p> <ul style="list-style-type: none"> - Metrisch interpretierte Variable (Grad der Erkrankungsschwere): Nutzenbewertungsverfahren: <i>Mean, SD, N</i>; p-Wert - Nominalskalierte Variablen (Ethnizität, geographische Region): Nutzenbewertungsverfahren: N Kaukasier/weiß bzw. (West-)Europa, N abweichend; p-Wert 	
	Patientencharakteristika <u>Einschluss:</u> Zu bewertende Studien, deren zugehöriges Nutzenbewertungsverfahren die Standardreferenz für einen Validitätsbeleg anführt.	Alter (J.: Jahre)	Merkmalsbeschreibung (Standardreferenz)	<i>Mean</i>	nicht berichtet		nicht inferentiell auswertbar (relevante Angaben dazu fehlen in der Referenzstudie)
				<i>SD</i>	(ersatzweise: Median 62 J., Range 23-82 J.)		
				N	84		
Geschlecht		Merkmalsbeschreibung (Standardreferenz)	N weiblich	47			
	N männlich		37				
	statistisch signifikanter Gruppenunterschied			in keinem Fall (0 %)			
Gesamtanzahl zu bewertender Studien			N = 2				

Grad an Übereinstimmung im Verhältnis von Nutzenbewertungsverfahren zu Standardreferenz	Patientencharakteristika <u>Einschluss:</u> Zu bewertende Studien, deren zugehöriges Nutzenbewertungsverfahren die Standardreferenz für einen Validitätsbeleg anführt.	Grad der Erkrankungsschwere (Operationalisierung: <i>ECOG PS</i> gw.)	Merkmalsbeschreibung (Standardreferenz)	<i>Mean</i>	0,535			
				<i>SD</i>	0,626			
				N	84			
			statistisch signifikanter Gruppenunterschied			N = 2 (100 %)	- Ruxolitinib (Chronisch myeloproliferative Erkrankungen) (COMFORT II): 0,721 (gw.), 0,653 (gepoolt), 219; p = 0,0255 - Idelalisib (101-09 [Subgruppe: Follikuläres Lymphom-Patienten]): 0,291, 0,484, 72; p = 0,0080	
			Gesamtanzahl zu bewertender Studien			N = 2		
		Ethnizität	Merkmalsbeschreibung (Standardreferenz)	N Kaukasier/weiß	64			
				N abweichend	20			
			statistisch signifikanter Gruppenunterschied			N = 1 (50 %)	- Idelalisib (101-09 [Subgruppe: Follikuläres Lymphom-Patienten]): 64, 7; p = 0,0324	
			Gesamtanzahl zu bewertender Studien			N = 2		
		geographische Region	Merkmalsbeschreibung (Standardreferenz)	N (West-) Europa	0			
				N abweichend	84			
			statistisch signifikanter Gruppenunterschied			N = 2 (100 %)	- Ruxolitinib (Chronisch myeloproliferative Erkrankungen) (COMFORT II): 219, 0; p < 0,0001 - Idelalisib (101-09 [Subgruppe: Follikuläres Lymphom-Patienten]): 28, 44; p < 0,0001	
Gesamtanzahl zu bewertender Studien			N = 2					

Erläuterungen: gw.: gewichtetes arithmetisches Mittel, gepoolt: gepoolte Standardabweichung. Berechnung unter Einbezug der Zahl an (randomisierten) Patienten je Therapiearm (Interventions-/Kontrollgruppe) und der jeweiligen Mittelwerte bzw. der jeweiligen Standardabweichungen einer betreffenden Merkmalsbeschreibung.

ECOG PS gw.: Häufigkeitsgewichtete Maßzahl zum *Eastern Cooperative Oncology Group Performance Status*, *Mean*: arithmetisches Mittel, *SD*: Standardabweichung, N: wahlweise Zahl der in die Auswertung einbezogenen Patienten bzw. Anzahl Nutzenbewertungsverfahren, Anzahl zu bewertender Studien oder Anzahl paarweiser Vergleiche.

Tabelle 7: Zusammenschau bezüglich der Standardreferenzen; hier: *FACT-Lym*
(Quelle: Eigene Darstellung).

5.6 Kongruenz hinsichtlich der Validierungsstudien mit Übereinstimmung in den Anwendungsgebieten (II.3)

In der Absicht, die relevanten Validierungsstudien vollumfänglich zu erfassen, ist – abseits der Betrachtung von Standardreferenzen – der Bezugsrahmen im Folgenden neu gesetzt. Ungeachtet der Häufigkeit, mit welcher eine Studie zur Rechtfertigung der Validität eines Messinstrumentes in den Nutzenbewertungsverfahren genannt wird, schließt die nachstehende Betrachtung sämtliche Verweise ein, die eine wenigstens partielle Übereinstimmung zu der zu bewertenden Studie im Hinblick auf die maligne Erkrankung aufweisen. Analog der Logik in Kapitel 5.5 fokussieren die Auswertungen mit dem Alter, dem Geschlecht, dem Grad der Erkrankungsschwere, der Ethnizität und der geographischen Region zentrale Merkmale der Studienpopulationen und berücksichtigen dabei mit dem *EORTC QLQ-C30*, dem *EQ-5D (VAS)*, dem *FACT-P*, dem *FACT-M*, dem *EORTC QLQ-MY20* und dem *FACT-Lym* wiederholt die führenden 6 Instrumente zur Erhebung der gesundheitsbezogenen Lebensqualität. Die Tabellen 8 bis 13 fassen die nachfolgenden Ergebnisse am Ende dieses Kapitels zusammen.

Kongruenzlevel von Referenzen mit AWG-Übereinstimmung: EORTC QLO-C30

Ausgehend von 28 Nutzenbewertungsverfahren, welche auf den *EORTC QLQ-C30* Bezug nehmen, konnten 18 Verweise zur Validierung des Instrumentes identifiziert werden. Diese stimmen hinsichtlich der Behandlungskontexte in 6 Fällen (33 %) nicht, in 4 Fällen (22 %) in Teilen und in 8 Fällen (44 %) gänzlich mit jenen der zu bewertenden Studien überein (vgl. Anhang G-I, Tab. G1-G18).

Reduziert auf die Validierungsstudien mit hinreichender Passung bezüglich der onkologischen Anwendungsgebiete zeigen sich statistisch signifikante Unterschiede für die paarweisen Vergleiche von zu bewertender Studie und bezugnehmender Referenz (ungeachtet der Standardreferenz) beim Alter in 3 Fällen (100 % [N = 25 paarweise Vergleiche, davon 22 nicht auswertbar]), beim Geschlecht in 12 Fällen (48 %), beim Grad der Erkrankungsschwere in 7 Fällen (88 % [N = 25 paarweise Vergleiche, davon 17 nicht auswertbar]), bei der Ethnizität in 3 Fällen (50 % [N = 20 paarweise Vergleiche, davon 14 nicht auswertbar]) sowie bei der geographischen Region in 10 Fällen (77 %) (vgl. Anhang G-I, Tab. G19-G23).

Kongruenzlevel von Referenzen jedweder Art: EQ-5D (VAS)

Der *EQ-5D (VAS)* dient in 18 Nutzenbewertungsverfahren der Erhebung der gesundheitsbezogenen Lebensqualität. Dabei stützen sich die Validitätsannahmen des Fragebogens auf zusammen 13 Publikationen. Die Konformität hinsichtlich der onkologischen Erkrankungen ist bei 6 dieser Arbeiten (46 %) nicht gegeben, bei weiteren 4 Arbeiten (31 %) zum Teil gegeben und bei 3 Arbeiten (23 %) vollumfänglich gegeben (vgl. Anhang G-II, Tab. G24-G36).

Der generischen Eignung des *EQ-5D (VAS)* folgend, bringt die Analyse sämtlicher Vergleiche von zu bewertender Studie und angeführter Validierungsstudie (ungeachtet der Standardreferenz) einen statistisch signifikanten Unterschied in 12 Fällen (100 % [N = 36 paarweise Vergleiche, davon 24 nicht auswertbar]) für das Alter, in 25 Fällen (74 % [N = 36 paarweise Vergleiche, davon 2 nicht auswertbar]) für das Geschlecht, in 15 Fällen (79 % [N = 36 paarweise Vergleiche, davon 17 nicht auswertbar]) für den Grad der Erkrankungsschwere, in 11 Fällen (73 % [N = 28 paarweise Vergleiche, davon 13 nicht auswertbar]) für die Ethnizität und in 21 Fällen (95 %) für die geographische Region hervor (vgl. Anhang G-II, Tab. G37-G41).

Kongruenzlevel von Referenzen mit AWG-Übereinstimmung: FACT-P

Insgesamt 5 Nutzenbewertungsverfahren verwerten für Befunde zur gesundheitsbezogenen Lebensqualität den *FACT-P*. Dabei werden 3 Referenzstudien (60 %) mit umfassender Passung hinsichtlich der onkologischen Entität, eine nicht auffindbare (20 %) sowie eine nicht hinreichend spezifische Referenz (20 %) zwecks Dokumentation der Validität angeführt (vgl. Anhang G-III, Tab. G42-G46).

Für das Merkmal Geschlecht zeigen die Auswertungen in keinem Fall (0 % [N = 7 paarweise Vergleiche, davon 4 nicht auswertbar]) einen statistisch signifikanten Gruppenunterschied, was den Vergleich von zu bewertender Studie und Referenzstudie (ungeachtet den Standardreferenzen) anbelangt (vgl. Anhang G-III, Tab. G47). Weitere Berechnungen mit Blick auf die Kriterien Alter, Grad der Erkrankungsschwere, Ethnizität und geographische Region konnten infolge fehlender Angaben in den Referenzen nicht durchgeführt werden.

Kongruenzlevel von Referenzen mit AWG-Übereinstimmung: FACT-M

Der *FACT-M* findet in 4 Nutzenbewertungsverfahren Berücksichtigung. Diese ziehen wiederum 2 Referenzstudien als Quellen heran, um die Validität des Fragebogens zu belegen.

Die Anwendungsgebiete stimmen jeweilig (100 %) mit denen der zu bewertenden Studien überein (vgl. Anhang G-IV, Tab. G48-G49).

Mit Blick auf die verbliebene Referenz mit AWG-Übereinstimmung (ungeachtet der Standardreferenz) ergeben sich sowohl für das Geschlecht (N = 2 paarweise Vergleiche) als auch für den Grad der Erkrankungsschwere (N = 2 paarweise Vergleiche) und die Ethnizität (N = 1 paarweiser Vergleich) in keinem Fall (0 %) statistisch signifikante Unterschiede betreffend den Vergleich von zu bewertender Studie und Referenzstudie (vgl. Anhang G-IV, Tab. G50-G52). Hinsichtlich des Alters (N = 2 paarweise Vergleiche) und der geographischen Region (N = 2 paarweise Vergleiche) sind die Ergebnisse nicht inferentiell auswertbar (vgl. Anhang G-IV, Tab. G53-G54).

Kongruenzlevel von Referenzen mit AWG-Übereinstimmung: EORTC QLQ-MY20

Für die Untersuchung des *EORTC QLQ-MY20* können 3 Nutzenbewertungsverfahren und 3 Verweise zur Rechtfertigung der Validität herangezogen werden. Zwei dieser Referenzstudien (67 %) gleichen sich jeweils in Bezug auf die zugrundeliegende onkologische Erkrankung der zu bewertenden Studien. Der Behandlungskontext der dritten Referenzstudie (33 %) ist hingegen ein abweichender (vgl. Anhang G-V, Tab. G55-G57).

Reduziert auf die Referenzstudie, welche eine Entsprechung im Anwendungsgebiet aufweist (ungeachtet der Standardreferenz), zeigt sich beim Geschlecht ein Gruppenunterschied ohne statistische Signifikanz und bei der Ethnizität ein solcher mit statistischer Signifikanz (vgl. Anhang G-V, Tab. G58-G59). Überdies ist die Abweichung beim Alter der inferentiellen Auswertung nicht zugänglich (vgl. Anhang G-V, Tab. G60). Nicht beziffert werden konnten infolge fehlender Angaben in den Referenzstudien etwaige Differenzen mit Blick auf die Kriterien Grad der Erkrankungsschwere und geographische Region.

Kongruenzlevel von Referenzen mit AWG-Übereinstimmung: FACT-Lym

Auf den *FACT-Lym* wird in 3 Nutzenbewertungsverfahren Bezug genommen. Bemessen an den dort zitierten Referenzstudien ist in einem Fall (33 %) eine partielle Übereinstimmung auszumachen, was die beschriebenen onkologischen Erkrankungen der zu bewertenden Studien anbelangt (vgl. Anhang G-VI, Tab. G61). Eine Referenzstudie (33 %) stimmt diesbezüglich nicht überein und eine weitere Arbeit (33 %) ist nicht auffindbar und in der Folge nicht bewertbar (vgl. Anhang G-VI, Tab. G62-G63).

Abseits der in Kapitel 5.5 dargelegten Standardreferenz ergeben sich keine weiteren Ansätze für Vergleiche mit Blick auf die Kriterien Alter, Geschlecht, Grad der Erkrankungsschwere, Ethnizität und geographische Region.

Erhebungsinstrument	EORTC QLQ-C30			
Kategoriale Einordnung	generisch	-		
	krankheitsspezifisch	zutreffend (onkologiespezifisch-generisch)		
Art und Umfang der Bezugnahme	Nutzenbewertungsverfahren	N = 28		
	Referenzstudien	N = 18		
Grad an Übereinstimmung im Verhältnis von zu bewertender Studie zu Referenzstudie	Tumorentität	umfassend	N = 8 Referenzstudien (44 %)	
		partiell	N = 4 Referenzstudien (22 %)	
		fehlend	N = 6 Referenzstudien (33 %)	
			<u>Logik:</u> - Metrische bzw. metrisch interpretierte Variablen (Alter, Grad der Erkrankungsschwere): Nutzenbewertungsverfahren (ggf. zu bewertende Studie): <i>Mean, SD, N</i> ; Referenzstudie (ggf. Subgruppe): <i>Mean, SD, N</i> ; p-Wert - Nominalskalierte Variablen (Geschlecht, Ethnizität, geographische Region): Nutzenbewertungsverfahren (ggf. zu bewertende Studie): N weiblich bzw. Kaukasier/weiß bzw. (West-)Europa, N männlich bzw. abweichend; Referenzstudie (ggf. Subgruppe): N weiblich bzw. Kaukasier/weiß bzw. (West-)Europa, N männlich bzw. abweichend; p-Wert	
Patientencharakteristika <u>Einschluss:</u> Referenzstudien mit umfassender oder partieller Übereinstimmung hinsichtlich der onkologischen Erkrankung. <u>Ausschluss:</u> Standardreferenzen gemäß Kapitel 5.5 aufgrund redundanter Betrachtung.	Alter (J.: Jahre)	statistisch signifikanter Gruppenunterschied	N = 3 (100 %)	- Afatinib (LUX-Lung 3): 60,2 J. (gw.), 10,1 J. (gepoolt), 345; Teckle et al., 2011 (<i>Lung cancer</i>): 63,0 J., 9,8 J., 61; p = 0,0457 - Afatinib (LUX-Lung 6): 56,3 J. (gw.), 10,8 J. (gepoolt), 364; Teckle et al., 2011 (<i>Lung cancer</i>): 63,0 J., 9,8 J., 61; p < 0,0001 - Nintedanib (nicht-kleinzelliges Lungenkarzinom): 58,6 J. (gw.), 9,8 J. (gepoolt), 658; Teckle et al., 2011 (<i>Lung cancer</i>): 63,0 J., 9,8 J., 61; p = 0,0008
		Gesamtanzahl paarweiser Vergleiche	N = 25 (davon 3 inferentiell auswertbar)	

Grad an Übereinstimmung im Verhältnis von zu bewertender Studie zu Referenzstudie	Patientencharakteristika <u>Einschluss:</u> Referenzstudien mit umfassender oder partieller Übereinstimmung hinsichtlich der onkologischen Erkrankung. <u>Ausschluss:</u> Standardreferenzen gemäß Kapitel 5.5 aufgrund redundanter Betrachtung.	Geschlecht	statistisch signifikanter Gruppenunterschied N = 12 (48 %)	<ul style="list-style-type: none"> - Crizotinib (PROFILE 1007): 194, 153; Braun et al., 2011: 593, 601; p = 0,0441 - Regorafenib: 296, 464; Osoba et al., 1997: 280, 205; p < 0,0001 - Afatinib (LUX-Lung 3): 224, 121; Braun et al., 2011: 593, 601; p < 0,0001 - Afatinib (LUX-Lung 3): 224, 121; Teckle et al., 2011 (<i>Lung cancer</i>): 29, 32; p = 0,0143 - Afatinib (LUX-Lung 6): 238, 126; Braun et al., 2011: 593, 601; p < 0,0001 - Afatinib (LUX-Lung 6): 238, 126; Teckle et al., 2011 (<i>Lung cancer</i>): 29, 32; p = 0,0098 - Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) (PCYC-1104-CA): 26, 85; Hjermstad et al., 1995: 72, 118; p = 0,0108 - Nintedanib (nicht-kleinzelliges Lungenkarzinom): 247, 411; Braun et al., 2011: 593, 601; p < 0,0001 - Regorafenib (Neubewertung nach Fristablauf): 86, 118; Osoba et al., 1997: 280, 205; p = 0,0002 - Carfilzomib: 345, 447; Delforge et al., 2012: 329, 320; p = 0,0080 - Crizotinib (neues Anwendungsgebiet): 212, 131; Braun et al., 2011: 593, 601; p < 0,0001 - Crizotinib (neues Anwendungsgebiet): 212, 131; Teckle et al., 2011 (<i>Lung cancer</i>): 29, 32; p = 0,0468
			Gesamtanzahl paarweiser Vergleiche N = 25	
		Grad der Erkrankungsschwere (Operationalisierung: <i>ECOG PS gw.</i>)	statistisch signifikanter Gruppenunterschied N = 7 (88 %)	<ul style="list-style-type: none"> - Crizotinib (PROFILE 1007): 0,688 (gw.), 0,622 (gepoolt), 347; Teckle et al., 2011 (<i>Lung cancer</i>): 1,017, 0,713, 61; p = 0,0002 - Afatinib (LUX-Lung 3): 0,618 (gw.), 0,491 (gepoolt), 345; Teckle et al., 2011 (<i>Lung cancer</i>): 1,017, 0,713, 61; p < 0,0001 - Afatinib (LUX-Lung 6): 0,756 (gw.), 0,425 (gepoolt), 364; Teckle et al., 2011 (<i>Lung cancer</i>): 1,017, 0,713, 61; p < 0,0001 - Eribulin (neues Anwendungsgebiet, Neubewertung nach Fristablauf): 0,626 (gw.), 0,539 (gepoolt), 882; McLachlan et al., 1998: 0,730, 0,709, 150; p = 0,0380

Grad an Übereinstimmung im Verhältnis von zu bewertender Studie zu Referenzstudie	Patientencharakteristika <u>Einschluss:</u> Referenzstudien mit umfassender oder partieller Übereinstimmung hinsichtlich der onkologischen Erkrankung. <u>Ausschluss:</u> Standardreferenzen gemäß Kapitel 5.5 aufgrund redundanter Betrachtung.				<ul style="list-style-type: none"> - Nintedanib (nicht-kleinzelliges Lungenkarzinom): 0,704 (gw.), 0,457 (gepoolt), 658; Teckle et al., 2011 (<i>Lung cancer</i>): 1,017, 0,713, 61; p < 0,0001 - Carfilzomib: 0,665 (gw.), 0,642 (gepoolt), 792; Delforge et al., 2012: 0,760, 0,664, 649; p = 0,0060 - Crizotinib (neues Anwendungsgebiet): 0,740 (gw.), 0,534 (gepoolt), 343; Teckle et al., 2011 (<i>Lung cancer</i>): 1,017, 0,713, 61; p = 0,0005
			Gesamtanzahl paarweiser Vergleiche	N = 25 (davon 8 inferentiell auswertbar)	
		Ethnizität	statistisch signifikanter Gruppenunterschied	N = 3 (50 %)	<ul style="list-style-type: none"> - Crizotinib (PROFILE 1005): 268, 171; Teckle et al., 2011 (<i>Lung cancer</i>): 24, 37; p = 0,0021 - Afatinib (LUX-Lung 6): 0, 364; Teckle et al., 2011 (<i>Lung cancer</i>): 24, 37; p < 0,0001 - Panobinostat: 499, 269; Dubois et al., 2006: 164, 38; p < 0,0001
			Gesamtanzahl paarweiser Vergleiche	N = 20 (davon 6 inferentiell auswertbar)	
		geographische Region	statistisch signifikanter Gruppenunterschied	N = 10 (77 %)	<ul style="list-style-type: none"> - Eribulin (neues Anwendungsgebiet, Neubewertung nach Fristablauf): 636, 246; McLachlan et al., 1998: 0, 150; p < 0,0001 - Eribulin (neues Anwendungsgebiet, Neubewertung nach Fristablauf): 636, 246; Parmar et al., 2005: 0, 299; p < 0,0001 - Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom): 170, 221; Hjermstad et al., 1995: 190, 0; p < 0,0001 - Nintedanib (nicht-kleinzelliges Lungenkarzinom): 463, 195; Braun et al., 2011: 0, 1194; p < 0,0001 - Nintedanib (nicht-kleinzelliges Lungenkarzinom): 463, 195; Teckle et al., 2011 (<i>Lung cancer</i>): 0, 61; p < 0,0001 - Regorafenib (Neubewertung nach Fristablauf): 0, 204; Osoba et al., 1997: 485, 0; p < 0,0001 - Carfilzomib: 590, 202; Kvam et al., 2011: 239, 0; p < 0,0001 - Carfilzomib: 590, 202; Wisloff et al., 1996: 524, 0; p < 0,0001

Grad an Übereinstimmung im Verhältnis von zu bewertender Studie zu Referenzstudie	Patientencharakteristika <u>Einschluss:</u> Referenzstudien mit umfassender oder partieller Übereinstimmung hinsichtlich der onkologischen Erkrankung. <u>Ausschluss:</u> Standardreferenzen gemäß Kapitel 5.5 aufgrund redundanter Betrachtung.				- Crizotinib (neues Anwendungsgebiet): 127, 213; Braun et al., 2011: 0, 1194; p < 0,0001 - Crizotinib (neues Anwendungsgebiet): 127, 213; Teckle et al., 2011 (<i>Lung cancer</i>): 0, 61; p < 0,0001
		Gesamtanzahl paarweiser Vergleiche	N = 13		

Erläuterungen: gw.: gewichtetes arithmetisches Mittel, gepoolt: gepoolte Standardabweichung. Berechnung unter Einbezug der Zahl an (randomisierten) Patienten je Therapiearm (Interventions-/Kontrollgruppe) und der jeweiligen Mittelwerte bzw. der jeweiligen Standardabweichungen einer betreffenden Merkmalsbeschreibung.

ECOG PS gw.: Häufigkeitsgewichtete Maßzahl zum *Eastern Cooperative Oncology Group Performance Status*, *Mean:* arithmetisches Mittel, *SD:* Standardabweichung, *N:* wahlweise Zahl der in die Auswertung einbezogenen Patienten bzw. Anzahl Nutzenbewertungsverfahren, Anzahl Referenzstudien oder Anzahl paarweiser Vergleiche.

Tabelle 8: Zusammenschau hinsichtlich der Referenzen mit AWG-Übereinstimmung; hier: *EORTC QLQ-C30*

(Quelle: Eigene Darstellung).

Erhebungsinstrument	EQ-5D (VAS)			
Kategoriale Einordnung	generisch	zutreffend		
	krankheitsspezifisch	-		
Art und Umfang der Bezugnahme	Nutzenbewertungsverfahren	N = 18		
	Referenzstudien	N = 13		
Grad an Übereinstimmung im Verhältnis von zu bewertender Studie zu Referenzstudie	Tumorentität	umfassend	N = 3 Referenzstudien (23 %)	
		partiell	N = 4 Referenzstudien (31 %)	
		fehlend	N = 6 Referenzstudien (46 %)	
			<u>Logik:</u> - Metrische bzw. metrisch interpretierte Variablen (Alter, Grad der Erkrankungsschwere): Nutzenbewertungsverfahren (ggf. zu bewertende Studie): <i>Mean, SD, N</i> ; Referenzstudie (ggf. Subgruppe): <i>Mean, SD, N</i> ; p-Wert - Nominalskalierte Variablen (Geschlecht, Ethnizität, geographische Region): Nutzenbewertungsverfahren (ggf. zu bewertende Studie): N weiblich bzw. Kaukasier/weiß bzw. (West-)Europa, N männlich bzw. abweichend; Referenzstudie (ggf. Subgruppe): N weiblich bzw. Kaukasier/weiß bzw. (West-)Europa, N männlich bzw. abweichend; p-Wert	
Patientencharakteristika	Alter (J.: Jahre)	statistisch signifikanter Gruppenunterschied	N = 12 (100 %)	- Bosutinib: 51,3 J. (gw.), 14,6 J. (gepoolt), 546; Pickard et al., 2007a: 59 J., 12 J., 534; p < 0,0001 - Afatinib (LUX-Lung 3): 60,2 J. (gw.), 10,1 J. (gepoolt), 345; Teckle et al., 2011 (<i>Lung cancer</i>): 63,0 J., 9,8 J., 61; p = 0,0457 - Afatinib (LUX-Lung 6): 56,3 J. (gw.), 10,8 J. (gepoolt), 364; Teckle et al., 2011 (<i>Lung cancer</i>): 63,0 J., 9,8 J., 61; p < 0,0001 - Radium-223-dichlorid: 70,3 J. (gw.), 8,0 J. (gepoolt), 921; Johnson et al., 1998: 52,4 J., 17,3 J., 905; p < 0,0001 - Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) (PCYC-1112-CA): 66,5 J. (gw.), 9,5 J. (gepoolt), 391; Pickard et al., 2007a: 59 J., 12 J., 534; p < 0,0001
Einschluss: Sämtliche Referenzstudien, ungeachtet der Übereinstimmung hinsichtlich der onkologischen Erkrankung. Ausschluss: Standardreferenz gemäß Kapitel 5.5 aufgrund redundanter Betrachtung.				

Grad an Übereinstimmung im Verhältnis von zu bewertender Studie zu Referenzstudie	Patientencharakteristika <u>Einschluss:</u> Sämtliche Referenzstudien, ungeachtet der Übereinstimmung hinsichtlich der onkologischen Erkrankung. <u>Ausschluss:</u> Standardreferenz gemäß Kapitel 5.5 aufgrund redundanter Betrachtung.			<ul style="list-style-type: none"> - Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) (PCYC-1104-CA): 67,1 J., 8,6 J., 111; Pickard et al., 2007a: 59 J., 12 J., 534; p < 0,0001 - Nintedanib (nicht-kleinzelliges Lungenkarzinom): 58,6 J. (gw.), 9,8 J. (gepoolt), 658; Teckle et al., 2011 (<i>Lung cancer</i>): 63,0 J., 9,8 J., 61; p = 0,0008 - Nivolumab (nicht-kleinzelliges Lungenkarzinom): 63,3 J. (gw.), 8,3 J. (gepoolt), 272; Pickard et al., 2007a: 59 J., 12 J., 534; p < 0,0001 - Dabrafenib (neues Anwendungsgebiet): 54,2 J. (gw.), 13,9 J. (gepoolt), 704; Pickard et al., 2007a: 59 J., 12 J., 534; p < 0,0001 - Trametinib: 53,7 J. (gw.), 13,3 J. (gepoolt), 322; Pickard et al., 2007a: 59 J., 12 J., 534; p < 0,0001 - Cobimetinib: 55,1 J. (gw.), 13,9 J. (gepoolt), 495; Pickard et al., 2007a: 59 J., 12 J., 534; p < 0,0001 - Cobimetinib: 55,1 J. (gw.), 13,9 J. (gepoolt), 495; Teckle et al., 2011: 58,5 J., 11,5 J., 184; p = 0,0032
			Gesamtanzahl paarweiser Vergleiche N = 36 (davon 12 inferentiell auswertbar)	
		Geschlecht	statistisch signifikanter Gruppenunterschied N = 25 (74 %)	<ul style="list-style-type: none"> - Crizotinib (PROFILE 1007): 194, 153; Schuette et al., 2012: 158, 363; p < 0,0001 - Crizotinib (PROFILE 1005): 233, 206; Schuette et al., 2012: 158, 363; p < 0,0001 - Afatinib (LUX-Lung 3): 224, 121; Brazier et al., 1993: 798, 655; p = 0,0007 - Afatinib (LUX-Lung 3): 224, 121; de Boer et al., 2004: 12, 71; p < 0,0001 - Afatinib (LUX-Lung 3): 224, 121; Schuette et al., 2012: 158, 363; p < 0,0001 - Afatinib (LUX-Lung 3): 224, 121; Teckle et al., 2011 (<i>Lung cancer</i>): 29, 32; p = 0,0143 - Afatinib (LUX-Lung 6): 238, 126; Brazier et al., 1993: 798, 655; p = 0,0003 - Afatinib (LUX-Lung 6): 238, 126; de Boer et al., 2004: 12, 71; p < 0,0001 - Afatinib (LUX-Lung 6): 238, 126; Schuette et al., 2012: 158, 363; p < 0,0001

Grad an Übereinstimmung im Verhältnis von zu bewertender Studie zu Referenzstudie	Patientencharakteristika <u>Einschluss:</u> Sämtliche Referenzstudien, ungeachtet der Übereinstimmung hinsichtlich der onkologischen Erkrankung. <u>Ausschluss:</u> Standardreferenz gemäß Kapitel 5.5 aufgrund redundanter Betrachtung.				<ul style="list-style-type: none"> - Afatinib (LUX-Lung 6): 238, 126; Teckle et al., 2011 (<i>Lung cancer</i>): 29, 32; p = 0,0098 - Radium-223-dichlorid: 0, 921; Brooks et al., 1991: 185, 164; p < 0,0001 - Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) (PCYC-1112-CA): 125, 266; Pickard et al., 2007a: 258, 276; p < 0,0001 - Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) (PCYC-1104-CA): 26, 85; Pickard et al., 2007a: 258, 276; p < 0,0001 - Nintedanib (nicht-kleinzelliges Lungenkarzinom): 247, 411; Brazier et al., 1993: 798, 655; p < 0,0001 - Nintedanib (nicht-kleinzelliges Lungenkarzinom): 247, 411; de Boer et al., 2004: 12, 71; p < 0,0001 - Nintedanib (nicht-kleinzelliges Lungenkarzinom): 247, 411; Schuette et al., 2012: 158, 363; p = 0,0113 - Ramucirumab (RAINBOW): 193, 472; Brazier et al., 1993: 798, 655; p < 0,0001 - Ramucirumab (RAINBOW): 193, 472; de Boer et al., 2004: 12, 71; p = 0,0040 - Ramucirumab (REGARD): 107, 248; Brazier et al., 1993: 798, 655; p < 0,0001 - Ramucirumab (REGARD): 107, 248; de Boer et al., 2004: 12, 71; p = 0,0038 - Nivolumab (nicht-kleinzelliges Lungenkarzinom): 64, 208; Pickard et al., 2007a: 258, 276; p < 0,0001 - Cobimetinib: 209, 286; Brazier et al., 1993: 798, 655; p < 0,0001 - Cobimetinib: 209, 286; de Boer et al., 2004: 12, 71; p < 0,0001 - Cobimetinib: 209, 286; Schuette et al., 2012: 158, 363; p < 0,0001 - Cobimetinib: 209, 286; Teckle et al., 2011: 119, 65; p < 0,0001
					Gesamtanzahl paarweiser Vergleiche

<p>Grad an Übereinstimmung im Verhältnis von zu bewertender Studie zu Referenzstudie</p>	<p>Patientencharakteristika <u>Einschluss:</u> Sämtliche Referenzstudien, ungeachtet der Übereinstimmung hinsichtlich der onkologischen Erkrankung. <u>Ausschluss:</u> Standardreferenz gemäß Kapitel 5.5 aufgrund redundanter Betrachtung.</p>	<p>Grad der Erkrankungsschwere (Operationalisierung: ECOG PS gw.)</p>	<p>statistisch signifikanter Gruppenunterschied</p>	<p>N = 15 (79 %)</p>	<ul style="list-style-type: none"> - Crizotinib (PROFILE 1007): 0,688 (gw.), 0,622 (gepoolt), 347; Teckle et al., 2011 (<i>Lung cancer</i>): 1,017, 0,713, 61; p = 0,0002 - Crizotinib (PROFILE 1005): 0,926, 0,707, 439; Schuette et al., 2012: 0,706, 0,558, 521; p < 0,0001 - Bosutinib: 0,352, 0,539, 546; Pickard et al., 2007a: 0,860, 0,741, 534; p < 0,0001 - Afatinib (LUX-Lung 3): 0,618 (gw.), 0,491 (gepoolt), 345; Schuette et al., 2012: 0,706, 0,558, 521; p = 0,0175 - Afatinib (LUX-Lung 3): 0,618 (gw.), 0,491 (gepoolt), 345; Teckle et al., 2011 (<i>Lung cancer</i>): 1,017, 0,713, 61; p < 0,0001 - Afatinib (LUX-Lung 6): 0,756 (gw.), 0,425 (gepoolt), 364; Teckle et al., 2011 (<i>Lung cancer</i>): 1,017, 0,713, 61; p < 0,0001 - Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) (PCYC-1112-CA): 0,593 (gw.), 0,491 (gepoolt), 391; Pickard et al., 2007a: 0,860, 0,741, 534; p < 0,0001 - Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) (PCYC-1104-CA): 0,630, 0,691, 111; Pickard et al., 2007a: 0,860, 0,741, 534; p = 0,0027 - Nintedanib (nicht-kleinzelliges Lungenkarzinom): 0,704 (gw.), 0,457 (gepoolt), 658; Teckle et al., 2011 (<i>Lung cancer</i>): 1,017, 0,713, 61; p < 0,0001 - Nivolumab (nicht-kleinzelliges Lungenkarzinom): 0,757 (gw.), 0,424 (gepoolt), 272; Pickard et al., 2007a: 0,860, 0,741, 534; p = 0,0341 - Dabrafenib (neues Anwendungsgebiet): 0,295 (gw.), 0,455 (gepoolt), 704; Pickard et al., 2007a: 0,860, 0,741, 534; p < 0,0001 - Trametinib: 0,360 (gw.), 0,481 (gepoolt), 322; Pickard et al., 2007a: 0,860, 0,741, 534; p < 0,0001 - Cobimetinib: 0,288 (gw.), 0,455 (gepoolt), 495; Pickard et al., 2007a: 0,860, 0,741, 534; p < 0,0001 - Cobimetinib: 0,288 (gw.), 0,455 (gepoolt), 495; Schuette et al., 2012: 0,706, 0,558, 521; p < 0,0001 - Cobimetinib: 0,288 (gw.), 0,455 (gepoolt), 495; Teckle et al., 2011: 0,767, 0,709, 184; p < 0,0001
---	--	---	---	----------------------	---

Grad an Übereinstimmung im Verhältnis von zu bewertender Studie zu Referenzstudie	Patientencharakteristika <u>Einschluss:</u> Sämtliche Referenzstudien, ungeachtet der Übereinstimmung hinsichtlich der onkologischen Erkrankung. <u>Ausschluss:</u> Standardreferenz gemäß Kapitel 5.5 aufgrund redundanter Betrachtung.		Gesamtanzahl paarweiser Vergleiche N = 36 (davon 19 inferentiell auswertbar)	
		Ethnizität	statistisch signifikanter Gruppenunterschied N = 11 (73 %)	<ul style="list-style-type: none"> - Crizotinib (PROFILE 1007): 181, 166; Schuette et al., 2012: 514, 7; p < 0,0001 - Crizotinib (PROFILE 1005): 268, 171; Schuette et al., 2012: 514, 7; p < 0,0001 - Crizotinib (PROFILE 1005): 268, 171; Teckle et al., 2011 (<i>Lung cancer</i>): 24, 37; p = 0,0021 - Bosutinib: 356, 190; Pickard et al., 2007a: 474, 60; p < 0,0001 - Afatinib (LUX-Lung 3): 96, 249; Schuette et al., 2012: 514, 7; p < 0,0001 - Afatinib (LUX-Lung 6): 0, 364; Schuette et al., 2012: 514, 7; p < 0,0001 - Afatinib (LUX-Lung 6): 0, 364; Teckle et al., 2011 (<i>Lung cancer</i>): 24, 37; p < 0,0001 - Dabrafenib (neues Anwendungsgebiet): 678, 26; Pickard et al., 2007a: 474, 60; p < 0,0001 - Cobimetinib: 462, 33; Pickard et al., 2007a: 474, 60; p = 0,0121 - Cobimetinib: 462, 33; Schuette et al., 2012: 514, 7; p < 0,0001 - Cobimetinib: 462, 33; Teckle et al., 2011: 85, 99; p < 0,0001
			Gesamtanzahl paarweiser Vergleiche N = 28 (davon 15 inferentiell auswertbar)	
		geographische Region	statistisch signifikanter Gruppenunterschied N = 21 (95 %)	<ul style="list-style-type: none"> - Afatinib (LUX-Lung 6): 0, 364; Brazier et al., 1993: 1980, 0; p < 0,0001 - Afatinib (LUX-Lung 6): 0, 364; de Boer et al., 2004: 83, 0; p < 0,0001 - Afatinib (LUX-Lung 6): 0, 364; Schuette et al., 2012: 521, 0; p < 0,0001 - Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) (PCYC-1112-CA): 170, 221; Pickard et al., 2007a: 0, 534; p < 0,0001 - Nintedanib (nicht-kleinzelliges Lungenkarzinom): 463, 195; Brazier et al., 1993: 1980, 0; p < 0,0001 - Nintedanib (nicht-kleinzelliges Lungenkarzinom): 463, 195; de Boer et al., 2004: 83, 0; p < 0,0001

Grad an Übereinstimmung im Verhältnis von zu bewertender Studie zu Referenzstudie	Patientencharakteristika <u>Einschluss:</u> Sämtliche Referenzstudien, ungeachtet der Übereinstimmung hinsichtlich der onkologischen Erkrankung. <u>Ausschluss:</u> Standardreferenz gemäß Kapitel 5.5 aufgrund redundanter Betrachtung.				<ul style="list-style-type: none"> - Nintedanib (nicht-kleinzelliges Lungenkarzinom): 463, 195; Schuette et al., 2012: 521, 0; p < 0,0001 - Nintedanib (nicht-kleinzelliges Lungenkarzinom): 463, 195; Teckle et al., 2011 (<i>Lung cancer</i>): 0, 61; p < 0,0001 - Ramucirumab (RAINBOW): 303, 362; Brazier et al., 1993: 1980, 0; p < 0,0001 - Ramucirumab (RAINBOW): 303, 362; de Boer et al., 2004: 83, 0; p < 0,0001 - Ramucirumab (REGARD): 172, 183; Brazier et al., 1993: 1980, 0; p < 0,0001 - Ramucirumab (REGARD): 172, 183; de Boer et al., 2004: 83, 0; p < 0,0001 - Nivolumab (nicht-kleinzelliges Lungenkarzinom): 155, 117; Greiner et al., 2005: 339, 0; p < 0,0001 - Nivolumab (nicht-kleinzelliges Lungenkarzinom): 155, 117; Pickard et al., 2007a: 0, 534; p < 0,0001 - Dabrafenib (neues Anwendungsgebiet): 464, 240; Pickard et al., 2007a: 0, 534; p < 0,0001 - Trametinib: 219, 103; Pickard et al., 2007a: 0, 534; p < 0,0001 - Cobimetinib: 366, 129; Brazier et al., 1993: 1980, 0; p < 0,0001 - Cobimetinib: 366, 129; de Boer et al., 2004: 83, 0; p < 0,0001 - Cobimetinib: 366, 129; Pickard et al., 2007a: 0, 534; p < 0,0001 - Cobimetinib: 366, 129; Schuette et al., 2012: 521, 0; p < 0,0001 - Cobimetinib: 366, 129; Teckle et al., 2011: 0, 184; p < 0,0001
					Gesamtanzahl paarweiser Vergleiche

Erläuterungen: gw.: gewichtetes arithmetisches Mittel, gepoolt: gepoolte Standardabweichung. Berechnung unter Einbezug der Zahl an (randomisierten) Patienten je Therapiearm (Interventions-/Kontrollgruppe) und der jeweiligen Mittelwerte bzw. der jeweiligen Standardabweichungen einer betreffenden Merkmalsbeschreibung.

ECOG PS gw.: Häufigkeitsgewichtete Maßzahl zum *Eastern Cooperative Oncology Group Performance Status*, *Mean*: arithmetisches Mittel, *SD*: Standardabweichung, N: wahlweise Zahl der in die Auswertung einbezogenen Patienten bzw. Anzahl Nutzenbewertungsverfahren, Anzahl Referenzstudien oder Anzahl paarweiser Vergleiche.

Tabelle 9: Zusammenschau hinsichtlich der Referenzen mit AWG-Übereinstimmung; hier: EQ-5D (VAS)
 (Quelle: Eigene Darstellung).

Erhebungsinstrument	FACT-P				
Kategoriale Einordnung	generisch	-			
	krankheitsspezifisch	zutreffend			
Art und Umfang der Bezugnahme	Nutzenbewertungsverfahren	N = 5			
	Referenzstudien	N = 5			
Grad an Übereinstimmung im Verhältnis von zu bewertender Studie zu Referenzstudie	Tumorentität	umfassend	N = 3 Referenzstudien (60 %)		
		unspezifisch	N = 1 Referenzstudie (20 %) („mixed sample of cancer patients“)		
		nicht auffindbar	N = 1 Referenzstudie (20 %)		
	Patientencharakteristika <u>Einschluss:</u> Referenzstudien mit umfassender Übereinstimmung hinsichtlich der onkologischen Erkrankung („unspezifisch“ ergänzend einbezogen). <u>Ausschluss:</u> Standardreferenzen gemäß Kapitel 5.5 aufgrund redundanter Betrachtung.	Geschlecht	statistisch signifikanter Gruppenunterschied	in keinem Fall (0 %)	
			Gesamtanzahl paarweiser Vergleiche	N = 7 (davon 3 inferentiell auswertbar)	
		Alter, Grad der Erkrankungsschwere, Ethnizität, geographische Region	nicht auswertbar (relevante Angaben dazu fehlen in den Referenzstudien)		

N: wahlweise Anzahl Nutzenbewertungsverfahren, Anzahl Referenzstudien oder Anzahl paarweiser Vergleiche.

Tabelle 10: Zusammenschau hinsichtlich der Referenzen mit AWG-Übereinstimmung; hier: FACT-P
(Quelle: Eigene Darstellung).

Erhebungsinstrument	FACT-M				
Kategoriale Einordnung	generisch	-			
	krankheitsspezifisch	zutreffend			
Art und Umfang der Bezugnahme	Nutzenbewertungsverfahren	N = 4			
	Referenzstudien	N = 2			
Grad an Übereinstimmung im Verhältnis von zu bewertender Studie zu Referenzstudie	Tumorentität	umfassend	N = 2 Referenzstudien (100 %)		
	Patientencharakteristika <u>Einschluss:</u> Referenzstudien mit umfassender Übereinstimmung hinsichtlich der onkologischen Erkrankung. <u>Ausschluss:</u> Standardreferenz gemäß Kapitel 5.5 aufgrund redundanter Betrachtung.	Geschlecht, Grad der Erkrankungsschwere, Ethnizität	statistisch signifikanter Gruppenunterschied	in keinem Fall (je 0 %)	
			Gesamtanzahl paarweiser Vergleiche	je N = 2 (Geschlecht, Grad der Erkrankungsschwere) N = 1 (Ethnizität)	
		Alter, geographische Region	statistisch signifikanter Gruppenunterschied	n/a	
			Gesamtanzahl paarweiser Vergleiche	je N = 2 (davon 0 inferentiell auswertbar)	

N: wahlweise Anzahl Nutzenbewertungsverfahren, Anzahl Referenzstudien oder Anzahl paarweiser Vergleiche. n/a: nicht anwendbar.

Tabelle 11: Zusammenschau hinsichtlich der Referenzen mit AWG-Übereinstimmung; hier: FACT-M
(Quelle: Eigene Darstellung).

Erhebungsinstrument	EORTC QLQ-MY20					
Kategoriale Einordnung	generisch	-				
	krankheitsspezifisch	zutreffend				
Art und Umfang der Bezugnahme	Nutzenbewertungsverfahren	N = 3				
	Referenzstudien	N = 3				
Grad an Übereinstimmung im Verhältnis von zu bewertender Studie zu Referenzstudie	Tumorentität	umfassend	N = 2 Referenzstudien (67 %)			
		fehlend	N = 1 Referenzstudie (33 %)			
				Logik: Nutzenbewertungsverfahren: N Kaukasier/weiß, N abweichend; Referenzstudie: N Kaukasier/weiß, N abweichend; p-Wert		
	Patientencharakteristika <u>Einschluss:</u> Referenzstudien mit umfassender Übereinstimmung hinsichtlich der onkologischen Erkrankung. <u>Ausschluss:</u> Standardreferenz gemäß Kapitel 5.5 aufgrund redundanter Betrachtung.	Geschlecht	statistisch signifikanter Gruppenunterschied	nicht gegeben (0 %)		
			Gesamtanzahl paarweiser Vergleiche	N = 1		
		Ethnizität	statistisch signifikanter Gruppenunterschied	N = 1 (100 %)	- Panobinostat: 499, 269; Dubois et al., 2006: 164, 38; p < 0,0001	
			Gesamtanzahl paarweiser Vergleiche	N = 1		
		Alter	statistisch signifikanter Gruppenunterschied	n/a		
			Gesamtanzahl paarweiser Vergleiche	N = 1 (davon 0 inferentiell auswertbar)		
	Grad der Erkrankungsschwere, geographische Region	nicht auswertbar (relevante Angaben dazu fehlen in der Referenzstudie)				

N: wahlweise Zahl der in die Auswertung einbezogenen Patienten bzw. Anzahl Nutzenbewertungsverfahren, Anzahl Referenzstudien oder Anzahl paarweiser Vergleiche.
n/a: nicht anwendbar.

Tabelle 12: Zusammenschau hinsichtlich der Referenzen mit AWG-Übereinstimmung; hier: EORTC QLQ-MY20
(Quelle: Eigene Darstellung).

Erhebungsinstrument	FACT-Lym		
Kategoriale Einordnung	generisch	-	
	krankheitsspezifisch	zutreffend	
Art und Umfang der Bezugnahme	Nutzenbewertungsverfahren	N = 3	
	Referenzstudien	N = 3	
Grad an Übereinstimmung im Verhältnis von zu bewertender Studie zu Referenzstudie	Tumorentität	partiell	N = 1 Referenzstudie (33 %)
		fehlend	N = 1 Referenzstudie (33 %)
		nicht auffindbar	N = 1 Referenzstudie (33 %)
	Patientencharakteristika <u>Ausschluss:</u> Standardreferenz gemäß Kapitel 5.5 aufgrund redundanter Betrachtung (hebt vorliegend den Einschluss der Referenzstudie mit partieller Übereinstimmung hinsichtlich der onkologischen Erkrankung auf).	Alter, Geschlecht, Grad der Erkrankungsschwere, Ethnizität, geographische Region	kein Ansatz für Vergleiche gegeben

N: wahlweise Anzahl Nutzenbewertungsverfahren oder Anzahl Referenzstudien.

Tabelle 13: Zusammenschau hinsichtlich der Referenzen mit AWG-Übereinstimmung; hier: FACT-Lym
(Quelle: Eigene Darstellung).

5.7 Psychometrische Güte und klinische Relevanzkriterien (III.1-III.3)

Die Analyse zu Disparitäten oder gar unberücksichtigten Skalen von Instrumenten zur Erhebung der gesundheitsbezogenen Lebensqualität berücksichtigt die in Kapitel 5.5 selektieren 8 Standardreferenzen (vgl. Aaronson et al., 1993, Osoba et al., 1998, Rabin und de Charro, 2001, Esper et al., 1997, Cella et al., 2009, Cormier et al., 2008, Cocks et al., 2007 und Hlubocky et al., 2013). Die Darlegung und Einordnung relevanter Kennzahlen sowie die Betrachtung etwaiger Implikationen hinsichtlich der Sicherheit von Feststellungen zum endpunktbezogenen Zusatznutzen in den Nutzenbewertungsverfahren erfolgen gruppiert für (i) Referenzstudien, welche den primären Fokus auf die Testung psychometrischer Gütekriterien (v. a. Validität, Reliabilität und Änderungssensitivität) legen und (ii) für solche, die vorrangig auf eine Bestimmung der klinischen Relevanz im Sinne einer *MID*-Schätzung abzielen.

Standardreferenzen mit vorrangiger Testung psychometrischer Gütekriterien

Die internationale Feldstudie zu **Aaronson et al., 1993** rekrutiert 354 Patienten, wobei 305 dieser Patienten schließlich in die Auswertung eingehen. Die Studie wurde im Längsschnitt mit den Messzeitpunkten „*pre-treatment*“ und „*on-treatment*“ design und sieht die Bestimmung (i) der (Item-)Konvergenz- und Diskriminanzvalidität, (ii) der Reliabilität (via interner Konsistenz [Forderung: Cronbach's alpha $\geq 0,70$]) sowie der Validität via (iii) Interskalenkorrelation, (iv) „Technik der bekannten Gruppen“ (*known-group comparisons*) (Variablen: *ECOG PS* [fremdbeurteilt], *weight loss*, *toxicity*) und (v) Änderungssensitivität (Variable: *ECOG PS* [fremdbeurteilt]) vor.

Die Studienautoren legen dar, dass hinsichtlich der Skala Rollenfunktion des *EORTC QLQ-C30* Einschränkungen gelten, was die psychometrische Güte angeht. Zum Hintergrund: Trotz guter Daten unter anderem zur Änderungssensitivität sind die Ergebnisse insgesamt unzureichend (vgl. Unzulänglichkeiten betreffend die Item-Skala-Korrelation und die interne Konsistenz mit einem Cronbach's alpha $< 0,70$). Diese Erkenntnis mindert im Ergebnis insbesondere die Belastbarkeit jener Zusatznutzenaussagen, die einschließlich der Skala Rollenfunktion des *EORTC QLQ-C30* auf wenigen weiteren Skalen des Fragebogens beruhen (vgl. Tab. 14; Nutzenbewertungsverfahren zu Pomalidomid [erneute Nutzenbewertung] [73], Ruxolitinib [Chronisch myeloproliferative Erkrankungen] [74] und Ruxolitinib [Chronisch myeloproliferative Erkrankungen, erneute Nutzenbewertung] [75]).

Nutzenbewertungsverfahren	<i>QL</i>	<i>PF</i>	<i>RF</i>	<i>CF</i>	<i>EF</i>	<i>SF</i>
Afatinib	X	X	X	X	X	X
Afatinib (Neubewertung nach Fristablauf)	X	X	X	X	X	X
Carfilzomib	X					
Ceritinib	X	X				
Cobimetinib		X				X
Dabrafenib					X	
Dabrafenib (neues Anwendungsgebiet)	X	X	X	X	X	X
Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom)				X		
Nintedanib (nicht-kleinzelliges Lungenkarzinom)	X			X		
Pembrolizumab						X
Pomalidomid (erneute Nutzenbewertung)		X	X		X	
Ramucirumab					X	
Ruxolitinib (Chronisch myeloproliferative Erkrankungen)	X		X			
Ruxolitinib (Chronisch myeloproliferative Erkrankungen, erneute Nutzenbewertung)	X	X	X			X
Ruxolitinib (Polycythaemia vera)	X	X		X	X	X
Trametinib	X	X	X	X	X	X

Anmerkung: Eingeschränkt auf Nutzenbewertungsverfahren mit statistischen Signifikanzen, (zusätzlich) klinischen Relevanzen und (zusätzlicher) Beanspruchung eines Zusatznutzens hinsichtlich der Endpunktkategorie gesundheitsbezogene Lebensqualität seitens der pharmazeutischen Unternehmer (pU). X = statistisch signifikanter Unterschied (lt. Dossier des pU). Die orange Einfärbung kennzeichnet eine eingeschränkte Belastbarkeit von Skalen des *EORTC QLQ-C30*.

QL: Globaler Gesundheitsstatus, *PF:* Körperliche Funktion, *RF:* Rollenfunktion, *CF:* Kognitive Funktion, *EF:* Emotionale Funktion, *SF:* Soziale Funktion (Skalen des Fragebogens *EORTC QLQ-C30*).

Tabelle 14: Zusatznutzenbeanspruchung und Bezugnahme in den Dossiers; hier: Aaronson et al., 1993 (Quelle: Eigene Darstellung).

Rabin und de Charro, 2001 bringt größtenteils die Ergebnisse aus Hurst et al., 1997 [67] ein. Die Studie rekrutiert 233 Patienten und gibt insgesamt 3 Messzeitpunkte an. Indessen bleibt der Ergebnisbericht auf die Messungen zu Baseline und Monat 3 beschränkt. Die Wertung der

psychometrischen Güte des *EQ-5D (VAS)* erfolgt anhand von (i) Angaben zur diskriminanten Validität als Element der Konstruktvalidität (Variable: *functional class [I-IV]*) sowie Berechnungen zur Änderungssensitivität via (ii) *Standardized Response Mean (SRM)* und (iii) Regressionsanalysen (Variablen: u. a. *Health Assessment Questionnaire (HAQ) score*, *Hospital Anxiety and Depression (HAD)-mood*).

Die Befunde lassen erkennen, dass der *EQ-5D-Index* in einem höheren Maße in der Lage ist, zwischen Patientengruppen zu unterscheiden, verglichen mit der visuellen Analogskala (VAS) (vgl. Ergebnisse zur diskriminanten Validität im Falle eines verminderten Leistungsstatus). Hinsichtlich der Änderungssensitivität – bemessen an der *SRM* – zeigt sich eine für Index und VAS vergleichbare Güte, welche nach Cohen [76] mit moderat zu quantifizieren ist.

Esper et al.(, 1997) beziehen in ihre Untersuchung insgesamt 3 Kollektive ein („*subscale development sample*“, „*validity sample 1*“ und „*validity sample 2*“). Während das Entwickler-Kollektiv 43 Patienten umfasst, schließen die Validierungskollektive 34 Patienten (vgl. 1.) respektive 96 Patienten (vgl. 2.) ein. Zum Zweck der tatsächlichen Validierung des *FACT-P* sind zwei konzeptionelle Ansätze beschrieben: eine klinische (Längsschnitt-)Studie mit Messzeitpunkten zu Baseline, Monat 1 und Monat 2 (1. Validierungskollektiv) sowie die Betrachtung im Querschnitt von Prostatakarzinom-Patienten, die eine Zweitmeinung hinsichtlich ihrer Erkrankung nachfragen (2. Validierungskollektiv). Zur Erfassung der psychometrischen Eigenschaften des Fragebogens werden (i) die Reliabilität (via interner Konsistenz [Forderung: Cronbach's alpha > 0,60]), (ii) die konkurrente Validität (Variable: *disease stage*) und (iii) die Änderungssensitivität (Variablen: *Performance Status Rating, PSR* [fremdbeurteilt], *Prostate-specific Antigen, PSA*) ausgewertet.

Zu bemerken ist, dass die Ergebnisse zu Esper et al., 1997 auf einer veralteten Version (2) des *FACT-P* fußen. Zur Einordnung: Die zum Zeitpunkt der angeführten Nutzenbewertungsverfahren aktuelle und verwendete Version ist die vierte. Was die Validität und Änderungssensitivität anbelangt, zeigt sich ein differenziertes Ansprechen der einzelnen Subskalen hinsichtlich der verwendeten Anker (vgl. *PSR, PSA*). Während die physischen Dimensionen sich hinreichend unterscheiden beziehungsweise änderungssensitiv sind, fehlt die Fähigkeit einer solchen Diskriminierung bei den Aspekten des sozialen und emotionalen Wohlbefindens. Die Bedeutsamkeit einzelner (Sub-)Skalen des *FACT-P* für einen beanspruchten Zusatznutzen in den Nutzenbewertungsverfahren ist in Tabelle 15 illustriert.

Bei gegebener statistischer Signifikanz ist mit Blick auf die Bestimmung der Änderungssensitivität für das Kriterium *PSR* ferner festzustellen, dass sich die Skalenwerte

zwischen den Ausprägungen *improved* und *unchanged* gering und zwischen den Ausprägungen *unchanged* und *worsened* erheblich unterscheiden, was nahelegt, dass lediglich eine „Verschlechterung“ valide erfasst werden kann. Allgemein ist eine geringe und in den Auswertungen variable Studienpopulation zu konstatieren (vgl. Reliabilität [N = 173] > Validität [N = 130] > Änderungssensitivität [N = 34, davon 28 {PSR} resp. 15 {PSA} in die Analyse eingehend]). Zugleich differieren die Validierungskollektive in Bezug auf das Erkrankungsstadium.

Nutzenbewertungsverfahren	FACT-P total score	PWB	FWB	SWB	EWB	FACT-P PCS	FACT-P TOI
Abirateronacetat	X	X	X		X	X	n/a
Abirateronacetat (neues Anwendungsgebiet) [§]	X	X	X	X	X	X	X
Enzalutamid	X	X	X		X	X	n/a
Radium-223-dichlorid	X	n/a	n/a	n/a	n/a	n/a	n/a

Anmerkung: Eingeschränkt auf Nutzenbewertungsverfahren mit statistischen Signifikanzen, (zusätzlich) klinischen Relevanzen und (zusätzlicher) Beanspruchung eines Zusatznutzens hinsichtlich der Endpunktkategorie gesundheitsbezogene Lebensqualität seitens der pharmazeutischen Unternehmer (pU).

X = statistisch signifikanter Unterschied (lt. Dossier des pU). Die orange Einfärbung kennzeichnet eine eingeschränkte Belastbarkeit von Skalen des FACT-P. n/a = nicht übertragbar (entweder nicht erhoben oder nicht berichtet). § Mit Verweis auf „Günstige-Prognose-Gruppe“.

FACT-P: Functional Assessment of Cancer Therapy – Prostate, PWB: Physisches Wohlbefinden, FWB: Funktionales Wohlbefinden, SWB: Soziales Wohlbefinden, EWB: Emotionales Wohlbefinden, PCS: Prostatakarzinom-spezifische Subskala, TOI: Trial Outcome Index (PWB + FWB + PCS).

Tabelle 15: Zusatznutzenbeanspruchung und Bezugnahme in den Dossiers; hier: Esper et al., 1997 (Quelle: Eigene Darstellung).

Als prospektive Studie mit 273 Patienten und Messzeitpunkten zu Baseline, Woche 1, Monat 3, Monat 6 und Monat 12 nebst Follow-up-Visiten ist die Arbeit zu **Cormier et al., 2008** konzipiert. Darin wird die Eignung des FACT-M mittels Bestimmung der formalen Zuverlässigkeit via (i) interner Konsistenz (Forderung: Cronbach’s alpha > 0,70) und (ii) Test-Retest-Reliabilität sowie anhand von Analysen zur (iii) Konvergenz- und Divergenzvalidität (vs. *Visual Analog Scale*, *EORTC QLQ Melanoma Module*, *KPS* [selbstbeurteilt], *Profile of Mood States* [*POMS*], *Marlowe-Crowne Social Desirability Scale* [*MC-SDS*]), (iv) kriteriumsbezogener Validität (Variablen: *disease stage*, *ECOG PS* [fremdbeurteilt], *KPS*

[selbstbeurteilt], *treatment status*) und (v) Änderungssensitivität (Variablen: *KPS* [selbstbeurteilt], *ECOG PS* [fremdbeurteilt]) erfasst.

Die Ergebnisberichte deuten darauf hin, dass die Subskalen soziales Wohlbefinden und emotionales Wohlbefinden in Bezug auf ihre psychometrischen Eigenschaften den weiteren unterlegen sind (vgl. Insuffizienzen zur Konvergenz- und Divergenzvalidität [hier: Korrelation des *MC-SDS*, Subskala: emotionales Wohlbefinden] und zur kriteriumsbezogenen Validität [hier: Differenzierung nach Erkrankungsstadium, Subskala: soziales Wohlbefinden] sowie eine allgemein fehlende Änderungssensitivität). Eine Relativierung von Aussagen zum Zusatznutzen entfällt jedoch, da der pU in keinem der zugeordneten Nutzenbewertungsverfahren einen solchen auf Grundlage der vermeintlich schwächeren Subskalen beansprucht. Mit Blick auf die verwendeten Anker zur Beurteilung der Erkrankungsschwere stellen die Autoren der Studie fest, dass die Patientenbeurteilung – via *KPS* – eine bessere Korrelation liefert als die Fremdeinschätzung – via *ECOG PS*. Bezüglich der Melanom-Subskala ist beschrieben, dass die Mehrzahl der Items eine (positiv gerichtete) schiefe Verteilung hinsichtlich der Beantwortung der Fragen aufwies. Dies – so konstatieren die Studienautoren – regt eine Überarbeitung der in die Skala einzubeziehenden Items an. Schließlich ist festzustellen, dass die Angaben zur Änderungssensitivität auf Daten von 67 Prozent der in die Studie eingeschlossenen Patienten basieren. Damit bleibt ein substantieller Anteil an Patienten in der Analyse unberücksichtigt.

Cocks et al., 2007 fasst den Ausgang einer internationalen Feldstudie mit 477 rekrutierten und 240 in die Analyse einbezogenen Teilnehmern zusammen. Erhebungen mittels des *EORTC QLQ-MY20* erfolgen zu Baseline und zu einem weiteren Zeitpunkt („*during or on completion of trial treatment*“). Zur Beschreibung der psychometrischen Struktur des Fragebogens wird von folgenden Gütekriterien Gebrauch gemacht: (i) der Reliabilität (via interner Konsistenz [Forderung: Cronbach's alpha $\geq 0,70$]), der Konstruktvalidität via (ii) konvergenter Validität, (iii) diskriminanter Validität und (iv) Korrelationsanalyse, (v) der klinischen Validität via *known-group comparisons* (Variablen: *ECOG PS*, *fractures*, *beta-2 microglobulin level*, *disease*, *chemotherapy intensity*) sowie (vi) der Änderungssensitivität (Variable: *response level*).

Die Autoren beschränken sich in ihrer Analyse zur psychometrischen Eignung auf die Myelom-spezifische Skala und regen an, diese gemeinsam mit dem (Kern-)Fragebogen *EORTC QLQ-C30* in klinischen Studienprogrammen zu verwenden. Bei insgesamt gegebener Validität legen die Detailergebnisse zum Modul nahe, dass die Nebenwirkungen der Behandlung – neben den

Krankheitssymptomen, der Körperwahrnehmung und der Zukunftsperspektive eine der Subskalen der 20 Items umfassenden Myelom-spezifischen Skala – am wenigsten treffsicher wiedergegeben werden können (vgl. nicht erreichte Korrelationen betreffend die Konstruktvalidität). Dabei ist hervorzuheben, dass solche Ergebnisse beispielsweise im Nutzenbewertungsverfahren zu Pomalidomid (erneute Nutzenbewertung) [73] abseits der gesundheitsbezogenen Lebensqualität unter der Zielkategorie Sicherheit subsummiert werden. In Bezug auf die Ergebnisse zur Änderungssensitivität wird überdies angemerkt, dass der Nachweis einer „Verschlechterung“ aufgrund einer zu geringen Studiengröße nicht erbracht werden kann. Angesichts der Tatsache, dass in den bezugnehmenden Nutzenbewertungsverfahren regelhaft die Zeit bis zur Verschlechterung als Endpunkt erfasst wird, ist dieser Umstand von besonderem Belang. Auch sind Patienten in fortgeschrittenen Erkrankungsstadien in der Validierungsstudie unterrepräsentiert, wohingegen die pivotalen Studien durchgängig Patienten mit rezidiviertem oder refraktärem Multiplen Myelom (rrMM) einschließen.

Standardreferenzen mit vorrangiger Bestimmung der klinischen Relevanz (MID-Schätzung)

Osoba et al., 1998 führt zur Bestimmung einer klinischen Relevanzschwelle für Änderungen im Scorewert des *EORTC QLQ-C30* zwei randomisierte Studien an. Eine dieser Studien ist als solche der Phase III im Bereich des rezidivierten oder metastasierten Mammakarzinoms mit 300 „qualifizierten“ Patienten und einer Analysepopulation von rund 246 Patienten beschrieben. Die weitere – nicht näher spezifizierte – Studie ist im Feld des kleinzelligen Lungenkarzinoms verortet und benennt 111 „qualifizierte“ Patienten und eine Analysepopulation von rund 80 Patienten. Die Erhebungen erfolgen im Längsschnitt zu jeweilig drei Messzeitpunkten (Baseline, Tag 1 des 3. Zyklus und unmittelbar vor dem 6. Zyklus bzw. Baseline, Woche 4 und Woche 7). Als Korrespondenzmodell findet der *SSQ* Verwendung, welcher (selbstberichtete) Zustandsänderungen mit „*a little*“, „*moderately*“ und „*very much*“ einordnet.

Von den Studienautoren unerwähnt, bleiben die Subskalen kognitive Funktion und Rollenfunktion des *EORTC QLQ-C30* in den Auswertungen zur klinischen Relevanzbestimmung unberücksichtigt. Daneben weist die zur Ableitung von Responsekriterien herangezogene Regressionsanalyse hinsichtlich des (kleinzelligen) Lungenkarzinom-Kollektivs eine teilweise Überschreitung der in den Nutzenbewertungsverfahren regelhaft angenommenen Spürbarkeitsschwelle in Höhe von

10 Skalenswerten auf (vgl. soziale Funktion: 12,2 Punkte, globaler Gesundheitsstatus: 10,7 Punkte). Zur Abschätzung der Relevanz dieser Erkenntnis enthält Tabelle 16 eine Übersicht von Nutzenbewertungsverfahren mit einem auf den (Teil-)Ergebnissen des *EORTC QLQ-C30* gründenden und von Seiten der pUs beanspruchten Zusatznutzen.

Nutzenbewertungsverfahren	QL	PF	RF	CF	EF	SF
Afatinib	X	X	X	X	X	X
Afatinib (Neubewertung nach Fristablauf)	X	X	X	X	X	X
Carfilzomib	X					
Cobimetinib		X				X
Crizotinib	X	X	X		X	X
Crizotinib (neues Anwendungsgebiet)	X	X	X	X	X	X
Dabrafenib					X	
Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom)				X		
Nintedanib (nicht-kleinzelliges Lungenkarzinom)	X			X		
Pembrolizumab						X
Ramucirumab					X	
Ruxolitinib (Chronisch myeloproliferative Erkrankungen)	X		X			
Ruxolitinib (Chronisch myeloproliferative Erkrankungen, erneute Nutzenbewertung)	X	X	X			X
Trametinib	X	X	X	X	X	X

Anmerkung: Eingeschränkt auf Nutzenbewertungsverfahren mit statistischen Signifikanzen, (zusätzlich) klinischen Relevanzen und (zusätzlicher) Beanspruchung eines Zusatznutzens hinsichtlich der Endpunktkategorie gesundheitsbezogene Lebensqualität seitens der pharmazeutischen Unternehmer (pU). X = statistisch signifikanter Unterschied (lt. Dossier des pU). Die orange Einfärbung kennzeichnet eine eingeschränkte Belastbarkeit von Skalen des *EORTC QLQ-C30*.

QL: Globaler Gesundheitsstatus, PF: Körperliche Funktion, RF: Rollenfunktion, CF: Kognitive Funktion, EF: Emotionale Funktion, SF: Soziale Funktion (Skalen des Fragebogens *EORTC QLQ-C30*).

Tabelle 16: Zusatznutzenbeanspruchung und Bezugnahme in den Dossiers; hier: Osoba et al., 1998 (Quelle: Eigene Darstellung).

Cella et al., 2009 präsentiert eine *RCT* der Phase III mit 809 Patienten und einem Einsatz des *FACT-P* zu diversen Messzeitpunkten (Baseline, Woche 4, Woche 12, alle 12 Wochen bis

Woche 72 bzw. Krankheitsprogress). Der Interpretation einer *MID* dienen sowohl Kennzahlen zur Verteilung (*one-third and one-half Standard Deviation [SD]*, *Standard Error of Measurement [SEM]*) wie auch ankerbasierte Ansätze. Zu Letzteren zählt die Auslegung der kriteriumsbezogenen Validität unter Einsatz der Ankervariablen *KPS* [fremdbeurteilt], *Bone Alkaline Phosphatase (BAP)* und *hemoglobin* sowie die Dimensionierung der Sensitivität anhand der Ankervariablen *Time to Disease Progression (TTP)*, *Adverse Events (AE)* und *survival*.

Jene in der Arbeit als relevante Änderungen beschriebenen *MIDs* werden infolge einer kombinierten Betrachtung der Analyseansätze als Wertebereiche angegeben (z. B. *FACT-P* Gesamtscore: 6 bis 10 Punkte). In den Nutzenbewertungsverfahren wird dabei der jeweils obere Grenzwert für eine konservative Schätzung herangezogen (z. B. Änderung im *FACT-P* Gesamtscore um ≥ 10 Punkte). Zu Cella et al., 2009 ist ferner zu bemerken, dass der Bericht einzelner Subskalen (vgl. physisches Wohlbefinden, funktionales Wohlbefinden, soziales Wohlbefinden und emotionales Wohlbefinden) schlicht ausbleibt. Inwieweit dies für die Beanspruchung eines Zusatznutzens in den Nutzenbewertungsverfahren von Belang ist, erfasst Tabelle 17.

Nutzenbewertungsverfahren	<i>FACT-P total score</i>	<i>PWB</i>	<i>FWB</i>	<i>SWB</i>	<i>EWB</i>	<i>FACT-P PCS</i>	<i>FACT-P TOI</i>
Abirateronacetat	X	X	X		X	X	n/a
Abirateronacetat (neues Anwendungsgebiet) [§]	X	X	X	X	X	X	X
Enzalutamid	X	X	X		X	X	n/a
Radium-223-dichlorid	X	n/a	n/a	n/a	n/a	n/a	n/a

Anmerkung: Eingeschränkt auf Nutzenbewertungsverfahren mit statistischen Signifikanzen, (zusätzlich) klinischen Relevanzen und (zusätzlicher) Beanspruchung eines Zusatznutzens hinsichtlich der Endpunktkategorie gesundheitsbezogene Lebensqualität seitens der pharmazeutischen Unternehmer (pU).

X = statistisch signifikanter Unterschied (lt. Dossier des pU). Die orange Einfärbung kennzeichnet eine eingeschränkte Belastbarkeit von Skalen des *FACT-P*. n/a = nicht übertragbar (entweder nicht erhoben oder nicht berichtet). [§] Mit Verweis auf „Günstige-Prognose-Gruppe“.

FACT-P: Functional Assessment of Cancer Therapy – Prostate, *PWB*: Physisches Wohlbefinden, *FWB*: Funktionales Wohlbefinden, *SWB*: Soziales Wohlbefinden, *EWB*: Emotionales Wohlbefinden, *PCS*: Prostatakarzinom-spezifische Subskala, *TOI*: Trial Outcome Index (*PWB* + *FWB* + *PCS*).

Tabelle 17: Zusatznutzenbeanspruchung und Bezugnahme in den Dossiers; hier: Cella et al., 2009 (Quelle: Eigene Darstellung).

Neben einer umfangreichen Analyse zur psychometrischen Struktur des *FACT-Lym* ist **Hlubocky et al., 2013** die Herleitung einer *MID*-Wertespanne zu entnehmen. Die Urteile basieren auf einer Validierungsstudie mit 84 Patienten und Erhebungszeitpunkten zu Baseline, Tag 3-7 und Woche 8-12. Im Einzelnen werden die formale Zuverlässigkeit via (i) interner Konsistenz (Forderung: Cronbach's alpha $\geq 0,70$) und (ii) Test-Retest-Reliabilität, (iii) die Konvergenz- und Divergenzvalidität (Variablen: *POMS-Short Form* [*BriefPOMS*], *Short Form* [*SF*]-36 *Health Survey Questionnaire*, *MC-SDS-Short Form* [*MC-10*]) sowie (iv) die kriteriumsbezogene Validität via *known-group comparisons* ([Anker-]Variablen: *NHL grade*, *ECOG PS* [selbstbeurteilt], *active treatment*), (v) die Änderungssensitivität ([Anker-]Variablen: *Global Rating of Change Scale* [selbstbeurteilt], *ECOG PS* [selbstbeurteilt]) und (vi) die *MID* im Zuge einer verteilungsbasierten Näherung (*0.33 to 0.50 SD range*) bestimmt.

Die Schätzung der *MID* beruht gleicherweise auf den Ergebnissen zum verteilungs- und ankerbasierten Ansatz und wird für die Domäne *LymS* in einer Range von 3 bis 5 Punkten angegeben. Am Beispiel des Nutzenbewertungsverfahrens zu Idelalisib [77] wird auf den oberen Wert (5 Punkte) referenziert, um von einer Relevanz der Änderung auf der *LymS*-Skala auszugehen. Anzumerken ist, dass neben der lymphomspezifischen Subskala hinsichtlich der Bestimmungen zur kriteriumsbezogenen Validität, der Änderungssensitivität und der *MID* keine weiteren Skalen in die Betrachtung einbezogen werden. Hinsichtlich der Übertragbarkeit der Befunde ist die Tatsache bedeutend, dass die Validierungsstudie vielfach Patienten mit einem geringen Grad der Erkrankung einschließt, wohingegen die bezugnehmenden pivotalen Studien der pUs abweichende Patientenkollektive berücksichtigen (vgl. Nutzenbewertungsverfahren zu Idelalisib [77]: vorbehandeltes indolentes B-Zell-Non-Hodgkin-Lymphom, Nutzenbewertungsverfahren zu Ruxolitinib [Chronisch myeloproliferative Erkrankungen, erneute Nutzenbewertung] [75]: primäre oder sekundäre Myelofibrose mit Intermediär-2- oder Hochrisikoprofil).

5.8 Charakteristik der Studienpopulationen bezüglich der Verweise (IV.1)

Insgesamt 28 Referenzstudien sind maßgebend für die Validitätsgewähr beim Einsatz der führenden 6 Instrumente zur Erhebung der gesundheitsbezogenen Lebensqualität. Bei 20 dieser Arbeiten (71 %) steht die Bestimmung der psychometrischen Gütekriterien (v. a. Validität, Reliabilität und Änderungssensitivität) im Vordergrund. Die weiteren 8 Arbeiten (29 %) haben vorrangig die klinische Relevanzbestimmung (via *MID*-Schätzung) zum Inhalt. Tabelle 18 listet je Referenzstudie die zentralen Patientencharakteristika.

Charakteristika

Referenzstudie	Studienpopulation/-indikation	Charakteristika					Geographische Region
		Alter	Geschlecht	Grad der Erkrankungsschwere		Ethnizität	
		[Jahre]	[Anteil Frauen]	[<i>ECOG PS</i> gw.]	[<i>ECOG PS</i> von 0 oder 1]	[Anteil Kaukasier/weiß]	[Anteil (West-) Europa]
Aaronson et al., 1993	nonresectable lung cancer (histologic type: NSCLC [63,1 %], SCLC [36,9 %]; disease stage: local/loco-regional [70,3 %], metastatic [29,7 %])	63	24 %	1,143	71,5 %	n. b.	69,5 %
Askew et al., 2009	melanoma (stage I/II [37,4 %], stage III [36,6 %], stage IV [26,0 %])	52	41,8 %	0,325	97,5 %	98,1 %	0 %
Braun et al., 2011	non-small cell lung cancer (NSCLC) (stage I [8,4 %], stage II [5,3 %], stage III [29,1 %], stage IV [54,9 %], indeterminate [2,3 %])	58,5	49,7 %	n. b.	n. b.	n. b.	0 %
Brazier et al., 1993	general population in Sheffield, UK	41,5	54,9 %	n. b.	n. b.	n. b.	100 %
Brooks et al., 1991	representative sample of the Swedish population	46,4	53 %	n. b.	n. b.	n. b.	100 %
Carducci et al., 2005	metastatic hormone-refractory prostate cancer (mHRPC)	n. b.	0 %	n. b.	n. b.	n. b.	n. b.
Cella et al., 2009	metastatic hormone-refractory prostate cancer (mHRPC)	72	0 %	0 (vgl. [<i>KPS</i>] (MW) 93,8, (Median) 100)	100 %	95 %	n. b.
Cocks et al., 2007	multiple myeloma (MM) (newly diagnosed [93,8 %], relapsed/refractory [6,3 %])	62	41,3 %	0,971	76,1 %	n. b.	90,8 %
Cormier et al., 2008	melanoma (stage I [22 %], stage II [15,4 %], stage III [36,6 %], stage IV [26 %])	52	41,8 %	0,325	97,5 %	98,1 %	0 %
de Boer et al., 2004	adenocarcinoma of the oesophagus or oesophagogastric junction	64	14 %	n. b.	n. b.	n. b.	100 %
Delforge et al., 2012	transplant-ineligible multiple myeloma (MM) (stage I [19,3 %], stage II [47,5 %], stage III [33,3 %])	71	50,5 %	0,760	87,1 %	n. b.	77 %
Dubois et al., 2006	relapsed, refractory multiple myeloma (MM) (stage III [72 %])	60	40 %	n. b.	n. b.	81 %	n. b.
Esper et al., 1997	prostate cancer (primarily men without hormone refractory disease [subscale development sample]; advanced hormone refractory prostate cancer [validity sample 1]; clinically localized disease (n = 67), evidence of metastasis or recurrence (n = 17), patients who were managing their disease conservatively (n = 10), patients with prostate problems, without having gotten a positive biopsy result (n = 2) [validity sample 2])	66 (65 [subscale development sample], 67 [validity sample 1], 66 [validity sample 2])	0 %	n. b.	n. b.	80,2 % (81 % [subscale development sample], 79,4 % [validity sample 1], n. b. [validity sample 2])	0 %
Greiner et al., 2005	representative sample of the German population in the northern German states of Lower Saxony and Bremen	n. b.	n. b.	n. b.	n. b.	n. b.	100 %

Hjermstad et al., 1995	breast cancer (30 %); testicular cancer (11 %); malignant lymphomas (9 %); cervical cancer (9 %); other cancer (41 %)	54	38 %	n. b.	n. b.	n. b.	100 %
Hlubocky et al., 2013	non-Hodgkin´s-lymphoma (NHL) (NHL grade: low [59,5 %], intermediate [26,2 %], high [14,3 %])	62	56,0 %	0,535	92,9 %	76,2 %	0 %
Johnson et al., 1998	national sample of the continental US population	52,4	"predominantly male"	n. b.	n. b.	"predominantly Caucasian"	0 %
Kvam et al., 2011	multiple myeloma (MM) (newly diagnosed [36 %], Plateau [34 %], relapse/progression [29 %])	66	46 %	n. b.	n. b.	n. b.	100 %
McLachlan et al., 1998	metastatic breast cancer	49	100 %	0,730	85 %	n. b.	0 %
Osoba et al., 1997	breast cancer (39,7 %); lung cancer (29,4 %); colorectal cancer (9,9 %); gynaecologic cancer (6,4 %); other cancer (14,6 %) (local/regional [61 %], distant metastases [39 %])	57	57,7 %	n. b.	n. b.	n. b.	100 %
Osoba et al., 1998	recurrent or metastatic breast cancer (68,9 %); extensive-stage small-cell lung cancer (SCLC) (31,1 %)	55,2 (breast cancer); 57,6 (SCLC)	100 % (breast cancer); n. b. (SCLC)	1,012 (breast cancer); 0,809 (SCLC)	75,2 % (breast cancer); 89,9 % (SCLC)	n. b.	n. b.
Parmar et al., 2005	operable breast cancer	44,2	100 %	n. b.	n. b.	n. b.	0 %
Pickard et al., 2007a	bladder cancer (5,8 %); brain cancer (approx. 9,6 %); breast cancer (approx. 9,6 %); colorectal cancer (approx. 9,6 %); head and neck cancer (approx. 9,6 %); liver cancer (approx. 4,8 %); pancreatic cancer (approx. 4,8 %); kidney cancer (approx. 9,6 %); lung cancer (9,4 %); lymphoma (approx. 9,6 %); ovarian cancer (approx. 9,6 %); prostate cancer (approx. 9,6 %)	59	48 %	0,860	74 %	89 %	0 %
Rabin und de Charro, 2001	rheumatoid arthritis (RA) (stage I [26 %], stage II [27 %], stage III [21,4 %], stage IV [19,3 %])	56	81 %	n. b.	n. b.	n. b.	100 %
Schuette et al., 2012	advanced non-small-cell lung cancer (NSCLC) (stage IIIa [6,7 %], stage IIIb [19,8 %], stage IV [73,5 %])	66,3	30,3 %	0,706	87,7 %	98,7 %	100 %
Teckle et al., 2011	breast cancer (35,9 %); colorectal cancer (31,0 %); lung cancer (33,2 %) (stage I [8 %], stage II [17 %], stage III [25 %], stage IV [50 %])	63	48 %	1,017	80,3 %	61 %	0 %
Webster et al., 1999	mixed sample of cancer patients	n. b.	n. b.	n. b.	n. b.	n. b.	n. b.
Wisloff et al., 1996	multiple myeloma (MM)	66	44 %	n. b.	39,6 %	n. b.	100 %

Anmerkung: Die blaue Einfärbung kennzeichnet eine vorrangige Testung psychometrischer Gütekriterien bei Referenzstudien. Die orange Einfärbung drückt hingegen eine vorrangige Bestimmung der klinischen Relevanz (MID-Schätzung) bezüglich der Referenzstudien aus. Standardreferenzen sind in der Schriftart hervorgehoben.

ECOG PS: Eastern Cooperative Oncology Group Performance Status, gw.: häufigkeitsgewichtet, KPS: Karnofsky Performance Status, MW: Mittelwert, n. b.: nicht berichtet.

Tabelle 18: Katalogteil 1 zur Charakteristik von Studienpopulationen
(Quelle: Eigene Darstellung).

5.9 (Gemittelte) Abweichungen und Rangfolgen bei Referenzstudien (IV.2)

Unter Einsatz der in Kapitel 5.8 angeführten Literatur und der systematischen Identifikation und Aufbereitung darin beschriebener therapeutischer Situationen und Merkmalsausprägungen fügt sich die in Tabelle 19 dargelegte Übersicht zur Verortung und Konformität von Referenzstudien auf Ebene der Anwendungsgebiete (Cluster zu bewertender Studien je onkologischer Entität) zusammen. Die Durchsetzung – im Sinne der Angabe gleicher Referenzstudien in verschiedenen onkologischen Clustern – ist dabei different und variiert zwischen 1 von 15 Clustern (diverse Referenzstudien) und 9 von 15 Clustern (vgl. Aaronson et al., 1993). Was die Abweichungen zur Referenzstudie für Anwendungsgebiete betrifft, die Nutzenbewertungsverfahren mit wenigstens einem bezugnehmenden Verweis einschließen, ergibt sich eine relative Wertespanne von 1 % (minimale Abweichung) bis 61 % (maximale Abweichung) und ein arithmetisches Mittel von 37 % (mittlere Abweichung).

(1-7 / 28)

	Aaronson et al., 1993			Askew et al., 2009			Braun et al., 2011			Brazier et al., 1993			Brooks et al., 1991			Carducci et al., 2005			Cella et al., 2009		
	nonresectable lung cancer			melanoma			non-small-cell lung cancer			general population in Sheffield, UK			representative sample of Swedish population			metastatic hormone-refractory prostate cancer			metastatic hormone-refractory prostate cancer		
AWG (geclustert)	Match ¹	Rel. Abw. z. Ref. ²	Rang	Match ¹	Rel. Abw. z. Ref. ²	Rang	Match ¹	Rel. Abw. z. Ref. ²	Rang	Match ¹	Rel. Abw. z. Ref. ²	Rang	Match ¹	Rel. Abw. z. Ref. ²	Rang	Match ¹	Rel. Abw. z. Ref. ²	Rang	Match ¹	Rel. Abw. z. Ref. ²	Rang
Basalzellkarzinom	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
mCRPC	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	ja	50,3 %	1	ja	keine valide Angabe möglich		ja	1,3 %	1
KRK	ja	48,2 %	8	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
Leukämie, Lymphom	ja	37,3 %	5	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
Magenkarzinom	ja	34,8 %	3	nein	n/a	n/a	nein	n/a	n/a	ja	41,3 %	3	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
Mammakarzinom	ja	54,6 %	9	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
MCD	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
Melanom	ja	40,3 %	6	ja	13,5 %	1	nein	n/a	n/a	ja	29,8 %	1	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
(rr)MM	ja	32,7 %	2	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
Myeloproliferative Erkrankungen	ja	35,0 %	4	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
Nierenzellkarzinom	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
NSCLC	ja	27,9 %	1	nein	n/a	n/a	ja	11,6 %	1	ja	32,9 %	2	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
Ovarialkarzinom	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
Polycythaemia vera	ja	46,2 %	7	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
Schilddrüsenkarzinom	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a

(8-14 / 28)

	Cocks et al., 2007			Cormier et al., 2008			de Boer et al., 2004			Delforge et al., 2012			Dubois et al., 2006			Esper et al., 1997			Greiner et al., 2005		
	Multiple myeloma			melanoma			Adenocarcinoma of the oesophagus or oesophagogastric junction			transplant-ineligible multiple myeloma			relapsed, refractory multiple myeloma			prostate cancer			representative sample of the German population in the northern German states of Lower Saxony and Bremen		
AWG (geclustert)	Match ¹	Rel. Abw. z. Ref. ²	Rang	Match ¹	Rel. Abw. z. Ref. ²	Rang	Match ¹	Rel. Abw. z. Ref. ²	Rang	Match ¹	Rel. Abw. z. Ref. ²	Rang	Match ¹	Rel. Abw. z. Ref. ²	Rang	Match ¹	Rel. Abw. z. Ref. ²	Rang	Match ¹	Rel. Abw. z. Ref. ²	Rang
Basalzellkarzinom	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
mCRPC	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	ja	14,3 %	1	nein	n/a	n/a
KRK	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
Leukämie, Lymphom	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
Magenkarzinom	nein	n/a	n/a	nein	n/a	n/a	ja	43,7 %	2	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
Mammakarzinom	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
MCD	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
Melanom	nein	n/a	n/a	ja	13,5 %	1	ja	43,3 %	1	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
(rr)MM	ja	9,6 %	1	nein	n/a	n/a	nein	n/a	n/a	ja	4,9 %	1	ja	2,8 %	1	nein	n/a	n/a	ja	keine valide Angabe möglich	
Myeloproliferative Erkrankungen	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
Nierenzellkarzinom	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
NSCLC	nein	n/a	n/a	nein	n/a	n/a	ja	56,0 %	3	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	ja	keine valide Angabe möglich	
Ovarialkarzinom	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
Polycythaemia vera	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
Schilddrüsenkarzinom	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a

(15-21 / 28)

	Hjermstad et al., 1995			Hlubocky et al., 2013			Johnson et al., 1998			Kvam et al., 2011			McLachlan et al., 1998			Osoba et al., 1997			Osoba et al., 1998					
	breast cancer, testicular cancer, malignant lymphomas, cervical cancer, other cancer			non-Hodgkin`s-lymphoma			national sample of the continental US population			Multiple myeloma			metastatic breast cancer			breast cancer, lung cancer, colorectal cancer, gynaecologic cancer, other cancer			recurrent or metastatic breast cancer			extensive-stage small-cell lung cancer		
AWG (geclustert)	Match ¹	Rel. Abw. z. Ref. ²	Rang	Match ¹	Rel. Abw. z. Ref. ²	Rang	Match ¹	Rel. Abw. z. Ref. ²	Rang	Match ¹	Rel. Abw. z. Ref. ²	Rang	Match ¹	Rel. Abw. z. Ref. ²	Rang	Match ¹	Rel. Abw. z. Ref. ²	Rang	Match ¹	Rel. Abw. z. Ref. ²	Rang	Match ¹	Rel. Abw. z. Ref. ²	Rang
Basalzellkarzinom	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	n/a	n/a	n/a
mCRPC	nein	n/a	n/a	nein	n/a	n/a	ja	keine valide Angabe möglich		nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	n/a	n/a	n/a
KRK	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	ja	52,3 %	2	ja	51,2 %	4	42,4 %	5	
Leukämie, Lymphom	ja	44,0 %	3	ja	19,6 %	1	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	ja	54,5 %	6	44,0 %	6	
Magenkarzinom	ja	43,3 %	2	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	ja	53,4 %	5	40,8 %	4	
Mammakarzinom	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	ja	19,0 %	1	nein	n/a	n/a	nein	n/a	n/a	n/a	n/a	n/a
MCD	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	n/a	n/a	n/a
Melanom	ja	36,6 %	1	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	ja	38,6 %	1	ja	57,5 %	7	52,7 %	7	
(rr)MM	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	ja	7,7 %	1	nein	n/a	n/a	nein	n/a	n/a	ja	50,1 %	3	40,6 %	3	
Myeloproliferative Erkrankungen	nein	n/a	n/a	ja	39,1 %	2	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	ja	49,2 %	2	39,2 %	2	
Nierenzellkarzinom	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	n/a	n/a	n/a
NSCLC	ja	44,5 %	4	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	ja	44,7 %	1	37,1 %	1	
Ovarialkarzinom	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	n/a	n/a	n/a
Polycythaemia vera	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	ja	60,9 %	8	55,1 %	8	
Schilddrüsenkarzinom	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	n/a	n/a	n/a

(22-28 / 28)

	Parmar et al., 2005			Pickard et al., 2007a			Rabin und de Charro, 2001			Schuette et al., 2012			Teckle et al., 2011			Webster et al., 1999			Wisloff et al., 1996		
	operable breast cancer			bladder cancer, brain cancer, breast cancer, colorectal cancer, head and neck cancer, liver cancer, pancreatic cancer, kidney cancer, lung cancer, lymphoma, ovarian cancer, prostate cancer			Rheumatoid arthritis			advanced non-small-cell lung cancer			breast cancer, colorectal cancer, lung cancer			mixed sample of cancer patients			Multiple myeloma		
AWG (geclustert)	Match ¹	Rel. Abw. z. Ref. ²	Rang	Match ¹	Rel. Abw. z. Ref. ²	Rang	Match ¹	Rel. Abw. z. Ref. ²	Rang	Match ¹	Rel. Abw. z. Ref. ²	Rang	Match ¹	Rel. Abw. z. Ref. ²	Rang	Match ¹	Rel. Abw. z. Ref. ²	Rang	Match ¹	Rel. Abw. z. Ref. ²	Rang
Basalzellkarzinom	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
mCRPC	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	ja	keine valide Angabe möglich	nein	n/a	n/a	
KRK	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
Leukämie, Lymphom	nein	n/a	n/a	ja	30,0 %	1	ja	55,9 %	4	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
Magenkarzinom	nein	n/a	n/a	nein	n/a	n/a	ja	53,1 %	3	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
Mammakarzinom	ja	17,3 %	1	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
MCD	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
Melanom	nein	n/a	n/a	ja	40,2 %	3	ja	44,9 %	1	ja	35,6 %	2	ja	46,9 %	2	nein	n/a	n/a	nein	n/a	n/a
(rr)MM	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	ja	7,3 %	1
Myeloproliferative Erkrankungen	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
Nierenzellkarzinom	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
NSCLC	nein	n/a	n/a	ja	32,7 %	2	ja	47,4 %	2	ja	25,0 %	1	ja	27,3 %	1	nein	n/a	n/a	nein	n/a	n/a
Ovarialkarzinom	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
Polycythaemia vera	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a
Schilddrüsenkarzinom	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a	nein	n/a	n/a

Erläuterungen: ¹ (vgl. Match) Wurde die Studie in einem Nutzenbewertungsverfahren als Referenz benannt, das dem jeweiligen Anwendungsgebiet zuzuordnen ist? ² (vgl. Relative Abweichung zur Referenz) Zur Berechnung des Wertes wird die Patientencharakteristik der zu bewertenden Studien in Beziehung gesetzt zu den verfügbaren Angaben der Referenzstudie. Die berücksichtigten und zu gleichen Teilen gewichteten Kriterien sind: (i) Studienpopulation/-indikation, (ii) Alter, (iii) Anteil Frauen, (iv) ECOG PS gw., (v) Anteil Kaukasier/weiß und (vi) Anteil (West-)Europa. **Konventionen:** (i) Um der Variabilität einzelner Kenngrößen Rechnung zu tragen, werden bei prozentualen Angaben absolute Abweichungen gerechnet (vgl. Geschlecht, Ethnizität, geographische Region). (ii) Zur Sicherstellung aussagekräftiger Angaben ist der Bericht von ≥ 3 Kriterien je Referenzstudie Voraussetzung für die Berechnung einer (gemittelten) Abweichung. AWG: Anwendungsgebiet, KRK: Kolorektalkarzinom, MCD: Multizentrischer Morbus Castleman, mCRPC: Metastasiertes kastrationsresistentes Prostatakarzinom, n/a: nicht anwendbar, NSCLC: Nicht-kleinzelliges Bronchialkarzinom, (rr)MM: Rezidiertes oder refraktäres Multiples Myelom.

Tabelle 19: Katalogteil 2 zu (gemittelten) Abweichungen und Rangfolgen
(Quelle: Eigene Darstellung).

5.10 Kategorisierung nach Belastbarkeit einzelner Skalen von Erhebungsinstrumenten (IV.3)

Die Einordnung der Belastbarkeit einzelner Skalen gründet auf den 8 Standardreferenzen und zieht Bilanz aus den (Teil-)Ergebnissen zur psychometrischen Güte und klinischen Relevanzbestimmung (vgl. Kapitel 5.7).

Bemessen an den 6 Skalen des *EORTC QLQ-C30* (Globaler Gesundheitsstatus [*QL*], Körperliche Funktion [*PF*], Rollenfunktion [*RF*], Kognitive Funktion [*CF*], Emotionale Funktion [*EF*], Soziale Funktion [*SF*]) ist bei Aaronson et al., 1993 einzig im Falle der Skala Rollenfunktion von einer eingeschränkten Belastbarkeit auszugehen. Bei Osoba et al., 1998 sind die Skalen Rollenfunktion und kognitive Funktion in den Analysen unberücksichtigt. Zudem lassen die Ergebnisse darauf schließen, dass die Skalen globaler Gesundheitsstatus und soziale Funktion nur eingeschränkt belastbar sind (vgl. Anhang H, Tab. H1).

Die Ergebnisse zu Rabin und de Charro, 2001 legen nahe, dass die Sicherheit der Aussagen zum Indexwert (*EQ-5D_{utility}*) jene der Aussagen zur visuellen Analogskala (*EQ-5D_{VAS}*) übersteigt (vgl. Anhang H, Tab. H2).

Hinsichtlich der 7 Komponenten des *FACT-P* (*FACT-P* Gesamtscore, Physisches Wohlbefinden [*PWB*], Funktionales Wohlbefinden [*FWB*], Soziales Wohlbefinden [*SWB*], Emotionales Wohlbefinden [*EWB*], Prostatakarzinomspezifische Subskala [*PCS*], *Trial Outcome Index* [*TOI = PWB + FWB + PCS*]) lässt sich für Esper et al., 1997 eine geringere Belastbarkeit der Skalen soziales Wohlbefinden und emotionales Wohlbefinden in Relation zu den Skalen physisches Wohlbefinden und funktionales Wohlbefinden ableiten. Für Cella et al., 2009 gilt, dass 4 der 5 Subklassen (*PWB*, *FWB*, *SWB* und *EWB*) in den Auswertungen fehlend sind (vgl. Anhang H, Tab. H3).

In Bezug auf den *FACT-M* ist zu Cormier et al., 2008 zu bemerken, dass die Aussagesicherheit im Falle der Skalen soziales Wohlbefinden und emotionales Wohlbefinden vermeintlich eine geringere ist, verglichen mit den weiteren Komponenten des Fragebogens (*FACT-M* Gesamtscore, Physisches Wohlbefinden [*PWB*], Funktionales Wohlbefinden [*FWB*], Melanomspezifische Subskala [*MS*]) (vgl. Anhang H, Tab. H4).

Untersuchungen zu dem 4 Subskalen umfassenden *EORTC QLQ-MY20* (Krankheitssymptome [*DS*], Nebenwirkungen [*SE*], Körperwahrnehmung [*BI*], Zukunftsperspektive [*FP*]) lassen mit Blick auf Cocks et al., 2007 vermuten, dass die Subskala Nebenwirkungen die am wenigsten belastbare Domäne im Gesamtgefüge darstellt (vgl. Anhang H, Tab. H5).

Mit Verweis auf Hlubocky et al., 2013 sind für den **FACT-Lym** keine Vorbehalte hinsichtlich der Aussagesicherheit der zusammen 7 Komponenten (*FACT-Lym* Gesamtscore, Physisches Wohlbefinden [*PWB*], Funktionales Wohlbefinden [*FWB*], Soziales Wohlbefinden [*SWB*], Emotionales Wohlbefinden [*EWB*], Lymphomspezifische Subskala [*LymS*], *Trial Outcome Index* [$TOI = PWB + FWB + LymS$]) zu vermerken (vgl. Anhang H, Tab. H6).

6 Diskussion

Mit einem Berichtsanteil von 80 Prozent kommt den Endpunkten der gesundheitsbezogenen Lebensqualität bei onkologischen Erkrankungen eine hohe und im Vergleich zu den Therapiegebieten insgesamt überdurchschnittliche Bedeutung zu [31]. Vor dem Hintergrund eines in den Anfängen der frühen Nutzenbewertung ausgeprägten Defizits in der Erhebung und Verwertung dieser Zielgröße [78] ist dies in besonderer Weise bedeutsam. Grundsätzlich ist die Aufbereitung der Evidenz in den Nutzenbewertungsverfahren dienlich, um die beschriebene Publikationslücke [33] jener Outcome-Daten zu verringern.

Allgemein gründen die in den AMNOG-Prozess eingebrachten Ergebnisse zur gesundheitsbezogenen Lebensqualität auf klinischen Studien – insbesondere randomisierten, direkten Vergleichsstudien – und damit auf solchen der höchsten Evidenzstufe (Ib) [5]. Studien minderer Qualität spielen bei der Endpunktbetrachtung insoweit keine Rolle. Infolge der subjektiven Erfassung dieser Zielgröße ist jedoch bei den oftmals unverblindeten Designs von einem relevanten Verzerrungspotenzial auszugehen, auf welches auch das IQWiG in seinen Allgemeinen Methoden [7] hinweist.

Mit Blick auf die Onkologie ist ersichtlich, dass Endpunkte der gesundheitsbezogenen Lebensqualität in keinem Fall als primäre Zielvariablen in den klinischen Studien berücksichtigt sind. Obgleich sich solche Endpunkte als konfirmatorische Ziele einer Studie eignen können [79], steht dies mit der Haltung der Zulassungsbehörden in Einklang, die den *PROs* eine im Allgemeinen supportive Wertigkeit [80] zuschreiben respektive eine Erhebung der gesundheitsbezogenen Lebensqualität als optional [81] oder additiven Wert [27] klassifizieren. Konsequenz der ungleichen Anforderungen ist, dass Endpunkte der gesundheitsbezogenen Lebensqualität seltener im Zulassungsprozess als in der frühen Nutzenbewertung die Entscheidungsfindung prägen [82]. Der Umstand, dass solche Studienendpunkte – abseits einer sekundären (Ziel-)Festlegung – vielfach explorativer Art oder nicht näher bestimmt sind, lässt eine bisweilen reduzierte Aussagekraft der Nachweise

annehmen. Überdies ist von Bedeutung, dass evidenz- und konsensbasierte Leitlinien wie die *SPIRIT-PRO Extension* [52] und die *CONSORT PRO Extension* [53] den Fokus explizit auf solche *PROs* legen, die als primäre oder sekundäre Zielparameter in klinischen Studien ihren Nutzen haben. Obwohl in dieser Arbeit nicht näher untersucht, dürfte die Art der Konzeptualisierung des Endpunktes nicht zuletzt mit dem Sensibilisierungsgrad der an einer Studie Beteiligten und in der Folge mit der Rücklaufquote der Fragebögen und der Vollständigkeit der Berichte in Zusammenhang stehen. Mängel dieser Art wiederum sind mitunter entscheidend in Bezug auf den Ausschluss entsprechender Daten im Rahmen der Dossierbewertung [34].

Die Befunde zur Anzahl der randomisierten Patienten, zur Dauer der Behandlung und zur Rate der Endpunkterhebung lassen eine heterogene Verteilung erkennen, was die Krebslokalisationen einerseits und die Therapiearme der zu bewertenden Studien andererseits anbelangt. Anzunehmen ist, dass die Differenzen in der Patientenzahl gemeinhin auf dem Wesen der (randomisierten) Gruppenteilung (vgl. Randomisierungsverhältnis 1:n) und Unterschieden in der Erkrankungsprävalenz beruhen. Ein Crossover-Design, eine insgesamt überlegene Wirksamkeit der Intervention und sich verändernde Dosisschemata – neuere Arzneimittel wie Proteinkinaseinhibitoren oder Immunonkologika werden meist über einen längeren Zeitraum eingesetzt als klassische Chemotherapien mit prädefinierten Zyklen – führen zu längeren Behandlungsdauern und in der Folge zu zusätzlichen Messungen in der Interventionsgruppe. Dies bedingt, dass unter anderem die Forderung nach einer in den Behandlungsgruppen gleichlangen Beobachtungszeit [83] nicht durchgängig erfüllt werden kann. Im Ergebnis zeigt sich, dass Evidenzen – im Sinne der tatsächlichen Erhebung von Daten zur gesundheitsbezogenen Lebensqualität – aufgrund tendenziell länger andauernder Behandlungen und einem privilegierten Zuteilungsmodus, was den numerischen Einschluss von Patienten betrifft, zugunsten der Interventionsgruppe verschoben sind. Jedoch steht die Zunahme an Information nicht notwendigerweise mit einer höheren Aussagesicherheit in Zusammenhang, was Untersuchungen zu Crossover-Effekten und deren Einfluss auf den Evidenzgehalt bei onkologischen Studien im Kontext des AMNOG belegen [84].

Der Umfang verfügbarer Informationen zur gesundheitsbezogenen Lebensqualität gewinnt an Bedeutung, sobald die klinische Relevanz einer Änderung im Score abseits der typischerweise in den Nutzenbewertungsverfahren vorzufindenden prä-post-Vergleiche (z. B. Verbesserung von 10 Punkten in der globalen *FACT-P*-Auswertung) beurteilt wird. Von einem hohen Informationsgehalt profitieren insbesondere Operationalisierungen, welche die dauerhafte Aufrechterhaltung einer definierten Response verlangen, den Zeitpunkt der Änderung

zusätzlich berücksichtigende Ereigniszeitanalysen sowie (zeitadjustierte) Berechnungen der Fläche unter der Kurve (*Area Under the Curve, AUC*) [83]. Beispielhaft sei auf das Nutzenbewertungsverfahren zu Regorafenib [85] verwiesen, in welchem die *AUC*-Analyse als Operationalisierungsoption ausnahmsweise Anwendung findet.

Der Erhebung der gesundheitsbezogenen Lebensqualität liegen multiple Messverfahren zugrunde. Dabei zeigt sich eine ausgesprochene Variabilität hinsichtlich der onkologischen Indikationen. Die Effektschätzungen betreffend die Endpunktkategorie fußen überwiegend auf mehr als einem Fragebogen und es dominiert die gleichzeitige generische und krankheitsspezifische Konzeption. Insoweit ist die Empfehlung des G-BA [86] bezüglich einer Kombination beider Methoden in den betrachteten Nutzenbewertungsverfahren als mehrheitlich umgesetzt zu betrachten.

Hinsichtlich der zu bewertenden Studien sind die Unterschiede zwischen den Interventions- und Kontrollgruppen, was die Merkmale Alter, Geschlecht, Grad der Erkrankungsschwere, Ethnizität und geographische Region anbelangt, gering – eine Folge der vorherrschenden randomisierten Zuteilung, die Strukturunterschiede egalisiert. Das Delta kann insoweit als indirekter Güteindikator für die Evidenzbasierung gelten, als der Einbezug indirekter Vergleiche oder Studien ohne Kontrollgruppe das Ungleichgewicht zwischen den Gruppen tendenziell vergrößert. Ein in der Gesamtschau feststellbarer Überhang männlicher Studienteilnehmer lässt sich mit den Krebsregisterdaten in Deutschland größtenteils in Einklang bringen [87]. Manifeste Abweichungen bezüglich der Geschlechterverteilung beschränken sich in diesem Vergleich auf die Lokalisationen Lunge und Schilddrüse. Ferner ist die gemeinhin wenig eingeschränkte Leistungsfähigkeit der Teilnehmer in den zu bewertenden Studien Ausdruck der vielfach restriktiven Einschlusskriterien. Der tendenziell hohe Anteil an Kaukasiern deckt sich mit dem Petition des G-BA, wonach Studien repräsentativ für die Population in Deutschland sein sollen [22]. In dieser Hinsicht ist aus Sicht des Verfassers auch die im Mittel hälftige Rekrutierung in (West-)Europa ein akzeptabler Wert, zumal pivotale Studien üblicherweise international und multizentrisch konzipiert und durchgeführt werden.

Der isolierte Vergleich von zu bewertender Studie und regelhaft benannter Referenzstudie – Letzteres im Sinne einer Leitpublikation, was den Nachweis der Validität des verwandten Messinstrumentes anbelangt – offenbart in der Mehrzahl statistisch signifikante Unterschiede hinsichtlich der betrachteten Merkmalsgruppen. Am ausgeprägtesten sind die Unähnlichkeiten beim Geschlecht und beim Grad der Erkrankungsschwere; am wenigsten auffallend bei der Ethnizität. Abseits der originär krankheitsspezifischen Messinstrumente (*FACT-P, FACT-M,*

EORTC QLQ-MY20 und *FACT-Lym*) sind darüber hinaus weitgehend fehlende Übereinstimmungen hinsichtlich der untersuchten Erkrankungssituationen festzustellen. Damit ist die Übertragbarkeit der Erkenntnisse aus Validierungsstudien auf die spezifischen Behandlungskontexte von Therapiestudien insgesamt in Frage zu stellen. Erkennbar ist zudem, dass die Tendenzen je Merkmal selten ausgeglichen sind. Das heißt, die Patienten in den zu bewertenden Studien sind (i) entweder häufig bis durchweg älter oder häufig bis durchweg jünger, (ii) entweder sämtlich zu einem höheren Anteil männlich oder tendenziell bis sämtlich zu einem höheren Anteil weiblich, (iii) entweder größtenteils bis ausnahmslos gesünder oder ausnahmslos weniger leistungsfähig, (iv) weit überwiegend bis ausschließlich Kaukasier beziehungsweise wurden die Patienten (v) entweder häufig bis immer zu einem geringeren Grade oder immer zu einem höheren Grade in (West-)Europa rekrutiert. Im Ergebnis spricht dies für eine weitreichende Homogenität der Populationen in den zu bewertenden Studien. Schließlich fällt auf, dass drei der acht Standardreferenzen bereits in den 1990er-Jahren veröffentlicht wurden und folglich weit zurückreichende Therapiesituationen repräsentieren. Eine einzige dieser Referenzen wurde hingegen ab dem Jahr 2011 (post-AMNOG) publiziert.

Die Gegenüberstellung von zu bewertender Studie und Referenzstudie mit wenigstens partieller Entsprechung hinsichtlich der onkologischen Erkrankung stellt eine in gewisser Weise ergänzende Art der Betrachtung dar – zumal in den Nutzenbewertungsverfahren meist mehrere Validierungsstudien angeführt werden. Bei abweichender Merkmalsreihung – Alter > geographische Region > Grad der Erkrankungsschwere > Ethnizität > Geschlecht – weisen die Befunde auch hier eine Dominanz statistisch signifikanter Unterschiede hinsichtlich der erschlossenen Charakteristika der Studienpopulationen aus. Und genauso zeigt eine Klassifizierung in Publikationen (i) der Jahre bis 1999, (ii) der Jahre von 2000 bis 2010 und (iii) der Jahre ab 2011, dass mehrheitlich ältere Studien als Referenzen zum Nachweis der Güte eines Instrumentes herangezogen werden. Insoweit erhärtet diese Sichtweise die Zweifelhaftigkeit der externen Validität und legt zudem ein Fehlen zeitgemäßer beziehungsweise überlegener Publikationen zur Validitätsannahme nahe.

Generell zeigt sich, dass die in Forschungsarbeiten [38, 39, 47] formulierte Forderung nach Validierungsstudien mit gleichartigen Behandlungskontexten unzureichend umgesetzt ist. Zudem irritiert die Tatsache, dass einerseits Ergebnisse auf Basis indirekter Vergleiche im Nutzenbewertungsprozess verworfen werden, sofern die zu vergleichenden Studien(-arme) als nicht hinreichend ähnlich beurteilt werden [88]. Andererseits wird die psychometrische Güte eines Erhebungsverfahrens mit Verweis auf Studien als gegeben erachtet, welche eine solche

Ähnlichkeit vermissen lassen. Das überarbeitete Methodenpapier des IQWiG [7] lässt nun das Bewusstsein über die Probleme der unter anderem durch abweichende Kontextfaktoren hervorgerufenen Variabilität bei empirisch ermittelten *MIDs* erkennen.

Die Detailbetrachtung der als Standardreferenzen in dieser Arbeit untersuchten Publikationen offenbart eine substantielle Heterogenität in Bezug auf (i) den Studientypus, (ii) der spezifischen Therapiesituationen, (iii) der Anzahl rekrutierter und (u. a. je Ankerkategorie) analysierter Patienten sowie (iv) der Methoden zur Erfassung der psychometrischen Gütekriterien. Erkennbar ist unter anderem, dass gerade nicht das im Nutzenbewertungsprozess etablierte *RCT*-Design vorherrscht. Vielmehr koexistiert ein Gemenge aus klinischen Studien, Feldstudien und solchen ohne nähere Angabe mit den wesensgemäßen Unterschieden bezüglich der Aussagekraft der Nachweise. Kennzeichnend ist zudem ein vielfach inhomogenes Patientenkollektiv als Folge breit angelegter Einschlusskriterien (u. a. Konglomerat verschiedener Erkrankungsstadien, Therapielinien und Behandlungsmodalitäten). Mit in der Mehrzahl zwei bis drei Erhebungszeitpunkten wird das in der Wissenschaft vorgetragene Verlangen nach einer Längsschnitterhebung [49] zuverlässig umgesetzt. Gleichzeitig ist auszumachen, dass sich die Datenerhebung selten über einen längeren Zeitraum erstreckt. Hinzukommend werden Erhebungen zu späteren Zeitpunkten teilweise nicht berichtet oder nicht näher spezifiziert. In diesem Punkt bestätigt sich der Bericht von Ousmen et al. [42], wonach die Stabilität einer *MID* im Zeitverlauf nur selten untersucht wird. Schließlich werden abseits des Spektrums an Bestimmungsmethoden selbst gleiche Maßzahlen wie das Cronbach's alpha hinsichtlich der Werte verschieden interpretiert (vgl. Akzeptanzschwellen von $> 0,60$, $\geq 0,70$ oder $> 0,70$), was einen Kontrast schafft zum allgemein gesetzten Minimalstandard [24, 51]. Auch ist Achtsamkeit geboten, was die Zulässigkeit der Richtung einer Response betrifft, wie die Angaben in Esper et al., 1997 und Cocks et al., 2007 belegen. Coon und Cappelleri [40] haben dazu den Hinweis gegeben, dass Schwellenwerte für eine Verbesserung nicht zwangsläufig für eine Verschlechterung nutzbar sind.

Allgemein bestätigt sich der in Forschungsarbeiten [40–42] dokumentierte Mangel an Standardisierung, was sowohl die Planung und Durchführung als auch das Berichtswesen von Validierungsstudien angeht. Dabei ist das Verstehen von *PROs* einschließlich ihrer Messeigenschaften nicht nur in der Wissenschaft von Belang, sondern auch für die Umsetzung in der klinischen Praxis [89]. Angesichts der Befunde ist überdies zu fordern, dass in Validierungsstudien unberücksichtigte Skalen oder festgestellte Unterschiede in der Belastbarkeit von Skalen thematisiert werden, sobald die Validität von Instrumenten zur

Erfassung der gesundheitsbezogenen Lebensqualität in den Nutzenbewertungsverfahren proklamiert wird. Dies gilt umso mehr, da jene Endpunktergebnisse in den Verfahren vielfach skalenbezogen präsentiert und interpretiert werden. In dieser Hinsicht sollte sich also die Forderung des IQWiG [83] nach Bereitstellung der Gesamtheit relevanter Subskalen widerspiegeln. Auch ist zu diskutieren, inwieweit es sachgerecht ist, einerseits die Relevanz des *EQ-5D*-Indexwertes mit Verweis auf den Unterschied von Selbsteinschätzung der Gesundheit durch Patienten im relevanten Anwendungsgebiet und Fremdeinschätzung derselben durch eine Bevölkerungsstichprobe in Frage zu stellen (vgl. Ausführungen des G-BA zum *EQ-5D* im Nutzenbewertungsverfahren zu Ibrutinib [chronische lymphatische Leukämie, Mantelzell-Lymphom] [90] und jene des IQWiG im Nutzenbewertungsverfahren zu Nivolumab [nicht-kleinzelliges Lungenkarzinom] [91]) oder nicht vom Patienten selbst beurteilte Nutzenendpunkte abzuwerten (vgl. u. a. Kritik des G-BA am *ECOG PS* im Nutzenbewertungsverfahren zu Ramucirumab [92]) und andererseits Referenzstudien zu akzeptieren, welche den Nachweis der Güte psychometrischer Instrumente anhand fremdbeurteilter Validitäts- und Ankerkriterien (mit-)erbringen.⁹ Schließlich ist zu bemerken, dass die Validierungsstudien insgesamt wenig Aufschluss darüber geben, inwieweit verwendete Anker ihrerseits geeignet sind, die Güte der Messinstrumente valide zu erfassen. Dabei ist anerkannt, dass die Validität und Reliabilität des Ankers entscheidend sind für die Bestimmung einer validen *MID* [94].

Die Unsicherheiten bei der Relevanzbewertung von Gruppenunterschieden thematisiert das IQWiG in der Neuauflage seines Methodenpapiers [7]. Dabei geht das Institut im Besonderen auf die Herausforderungen ein, die beim Einsatz empirisch abgeleiteter *MIDs* im Rahmen von Responderanalysen bei psychometrischen Skalen entstehen und offeriert ein Verfahren, welches relevante Veränderungen sicher detektieren und einer ergebnisgesteuerten Berichterstattung vorbeugen soll. Trotz des Anerkenntnisses der Variabilität von Responsekriterien berücksichtigt der Ansatz eine pauschale Schwellenwertfestlegung (mindestens 15 % der Skalenspannweite) und verkennt dabei die von Kontextfaktoren ausgelösten Abhängigkeiten. Die fehlende Spezifität und die Loslösung vom Prinzip der *MID* wird auch in den Stellungnahmen zum Entwurf der Allgemeinen Methoden 6.0 [95] bemerkt. Auch befasst sich die Kritik mit den praktischen Implikationen, sollten bisher akzeptierte *MIDs*

⁹ An dieser Stelle sei insbesondere auf die folgenden Validierungsstudien verwiesen: Aaronson et al., 1993, Esper et al., 1997, Cormier et al., 2008 und Cocks et al., 2007. Hinsichtlich Cella et al., 2009 kritisiert das IQWiG im Nutzenbewertungsverfahren zu Abirateronacetat (neues Anwendungsgebiet) [93] die Verwendung von Laborparametern oder allein auf der Einschätzung des behandelnden Arztes beruhender Anker und zieht die Belastbarkeit der *MID* in Folge fehlender Patientenrelevanz in Zweifel.

nicht länger berücksichtigt werden (z. B. mit Blick auf den *EORTC QLQ-C30*). Anstatt fiktive Evidenzen mit unklarer Verlässlichkeit zusätzlich in den Bewertungsprozess einzubringen, sollte auch hier gelten, sich der bestverfügbaren Evidenz zu bedienen, die in Form der Validierungsstudien existent ist. Gleichzeitig sind Initiativen notwendig, um an den tatsächlichen Bedarfen ausgerichtete Validierungsstudien neu aufzulegen. Neben den Sponsoren klinischer Studien stehen mit der *European Organisation for Research and Treatment of Cancer (EORTC)*, der *Functional Assessment of Chronic Illness Therapy (FACIT) group* und der *EuroQol Group* insbesondere jene Forschergruppen und akademischen Netzwerke in der Verantwortung, die weit verbreitete Messinstrumente entwickeln und ihre Eignung verkünden.

Die Katalogisierung der Befunde greift die gegenwärtigen Inventarisierungsansätze auf [43–45] und folgt den Zielen, die Charakteristik von Validierungsstudien offenzulegen und zur Reflexion anzuregen, soweit Verweise für die Eignung von Messinstrumenten verlangt sind. Zweck der berechneten Abweichungen und klassifizierten Ungleichheiten ist nicht die präzise Verortung von Referenzstudien oder eine profunde Qualifizierung von Skalen der Erhebungsinstrumente – zumal es an Informationen zur adäquaten Gewichtung der Studien- und Patientenmerkmale oder den für die Einordnung der Skalen in Belastbarkeitsgrade maßgebenden Gütekriterien mangelt. Beabsichtigt ist vielmehr, den Leser zu sensibilisieren, indem in komprimierter Form auf Art und Umfang von Diskrepanzen hingewiesen wird, die der Übertragbarkeit von Erkenntnissen aus Validierungsstudien auf einen gegebenen Behandlungskontext entgegenstehen können.

Zweifellos kann eine derartige Systematik nur ein erster Aufschlag sein und zur Fortentwicklung anregen. Dies gilt umso mehr, da kontinuierlich und dynamisch ansteigend neue Nutzenbewertungsverfahren initiiert werden – getriggert nicht zuletzt durch Befristungen von G-BA-Beschlüssen und Zulassungserweiterungen. Insoweit ist auch die vorliegende Arbeit mit der Limitation behaftet, die Charakteristik der Evidenzbasierung von Instrumenten zur Erhebung der gesundheitsbezogenen Lebensqualität lediglich fragmentarisch untersuchen und diskutieren zu können. Dazu gehören die Eingrenzungen von Betrachtungszeitraum (Januar 2011 bis Januar 2016) und Messverfahren (6 numerisch vorherrschende Erhebungsinstrumente) sowie die Fokussierung auf das Therapiefeld Onkologie, vor dem Hintergrund pragmatischer Erwägungen im Sinne einer Reduzierung einzuschließender Nutzenbewertungsverfahren einerseits und der Bedeutsamkeit der Zielgröße in der onkologischen Versorgung – speziell dem palliativen Setting – andererseits. Insoweit bleibt offen, ob die Befunde dieser Arbeit auf

spätere Zeiträume oder abweichende Untersuchungskontexte übertragbar sind. Der Einsatz gleicher Messinstrumente und eine sich wiederholende Bezugnahme auf bereits erfasste Referenzstudien lässt indes eine gewisse Saturierung erkennen, weshalb weitere erkenntnisstiftende Informationen durch den Einbezug späterer Nutzenbewertungsverfahren nicht zu erwarten sind. Die Aktualität und Wichtigkeit der Thematik insgesamt belegen nicht zuletzt die kritischen Einlassungen des IQWiG zur Beurteilung der klinischen Relevanz – im Besonderen bei (komplexen) Skalen – in seinem neu aufgelegten Methodenpapier [7]. Auch ist anzumerken, dass mit Stand 01.11.2020 Beschlüsse des G-BA zu einer Vielzahl der in dieser Arbeit berücksichtigten Nutzenbewertungsverfahren noch immer gültig sind (31 von 47) beziehungsweise in Folgeverfahren auf Ergebnisse derselben zu bewertenden Studien ganz (12 von 47) oder in Teilen (2 von 47) Bezug genommen wird.

Es sei darauf hingewiesen, dass das alleinige Sichten und Exzerpieren der mannigfaltigen Literatur durch den Verfasser sowie die weitgehend manuelle Datenverarbeitung – trotz der gewissenhaften Ausführung – mit einer Fehleranfälligkeit behaftet ist. Gleichwohl sprechen der Einbezug eines umfangreichen Datensets und die in sich konsistenten Ergebnisse für einen validen Bericht.

7 Fazit

Die gesundheitsbezogene Lebensqualität nimmt eine zentrale Rolle im Urteilsgefüge der frühen Nutzenbewertung ein. Speziell in der Onkologie sind die Ergebnisse mitunter essenziell, was die Entscheidungsfindung anbelangt. Neben einer inzwischen regelhaften Erhebung dieser Zielgröße auf Basis hochwertiger klinischer Studien zeigt sich, dass die Endpunkte einzig als zweit- oder nachrangige Ziele in solchen Studien festgeschrieben sind – einhergehend mit einer potenziell limitierten Aussagekraft. Hinsichtlich der Behandlungskontexte und zentraler Patientencharakteristika sind vielfach signifikante Unterschiede erkennbar, was eine Übertragbarkeit von Aussagen zur Güte eines Erhebungsverfahrens in Validierungsstudien auf die, in den zu bewertenden Studien erlangten, Befunde zur gesundheitsbezogenen Lebensqualität einschränkt. Hinzukommend berichten Autoren von Validierungsstudien bisweilen nicht alle Skalen von Messinstrumenten oder beschreiben unterschiedliche Güteprofile solcher Skalen, was eine ungleiche Belastbarkeit der Ergebnisse einzelner Domänen nahelegt. Von eminenter Bedeutung ist folglich, Restriktionen dieser Art im Zuge entweder der Validitätsannahme eines Fragebogens oder der Verwendung von

Responsekriterien bei einem Fragebogen zu kennen, zu benennen und zu diskutieren. Perspektivisch sollte ein Ordnungsrahmen zur Durchführung von Validierungsstudien geschaffen werden, welcher in der Forschung konsentierete Qualitätsindikatoren berücksichtigt und in der Folge einen minimalen Gütestandard solcher Studien sichert. Zusätzlich sind insbesondere jene Forschungsverbände und akademischen Netzwerke, die weit verbreitete Fragebögen zur Erhebung der gesundheitsbezogenen Lebensqualität in der Onkologie (fort-) entwickeln und eine Verwendung dieser Instrumente in klinischen Studien möglich machen, aufgefordert, mit der Auflage methodisch verfeinerter und auf die Behandlungskontexte gegenwärtiger Therapiestudien ausgerichteter Validierungsstudien, Unsicherheiten in der Interpretation von Ergebnissen zu mindern und damit die Akzeptanz der Messungen zu fördern. Damit würde zugleich Konsistenz geschaffen, was die, den AMNOG-Prozess prägende, Orientierung am Prinzip der evidenzbasierten Medizin betrifft.

Anhang

Anhang A: Eingeschlossene Nutzenbewertungsverfahren	111
Anhang B: Bewertungsmatrizen (exemplarisch).....	115
Anhang C: Eingeschlossene Referenzstudien	117
Anhang D: Flussdiagramm zur spezifischen Literaturlauswahl.....	122
Anhang E: Entsprechungsgrade von Patientencharakteristika innerhalb der zu bewertenden Studien.....	123
Anhang F: Entsprechungsgrade von Patientencharakteristika bei Standardreferenzen	126
F-I: Abweichungen zur Standardreferenz; hier: <i>EORTC QLQ-C30</i>	126
F-II: Abweichungen zur Standardreferenz; hier: <i>EQ-5D (VAS)</i>	133
F-III: Abweichungen zur Standardreferenz; hier: <i>FACT-P</i>	135
F-IV: Abweichungen zur Standardreferenz; hier: <i>FACT-M</i>	138
F-V: Abweichungen zur Standardreferenz; hier: <i>EORTC QLQ-MY20</i>	141
F-VI: Abweichungen zur Standardreferenz; hier: <i>FACT-Lym</i>	143
Anhang G: Entsprechungsgrade von Patientencharakteristika bei Referenzen mit AWG-Übereinstimmung	146
G-I: Abweichungen zur Referenz mit AWG-Übereinstimmung; hier: <i>EORTC QLQ-C30</i>	146
G-II: Abweichungen zur Referenz mit AWG-Übereinstimmung; hier: <i>EQ-5D (VAS)</i>	157
G-III: Abweichungen zur Referenz mit AWG-Übereinstimmung; hier: <i>FACT-P</i>	167
G-IV: Abweichungen zur Referenz mit AWG-Übereinstimmung; hier: <i>FACT-M</i>	169
G-V: Abweichungen zur Referenz mit AWG-Übereinstimmung; hier: <i>EORTC QLQ-MY20</i>	171
G-VI: Abweichungen zur Referenz mit AWG-Übereinstimmung; hier: <i>FACT-Lym</i>	172
Anhang H: Ordnung der Skalen von Erhebungsinstrumenten ihrer Aussagekraft entsprechend	173

Anhang A: Eingeschlossene Nutzenbewertungsverfahren

Wirkstoff (Bezeichnungen zum Zeitpunkt des Datenabzugs am 23.08.2016)	Beginn des Bewertungs- verfahren	Status	Anwendungs- gebiet (i. w. S.)	Anwendungsgebiet (i. e. S.)
Cabazitaxel	15.04.2011	Verfahren abgeschlossen	Onkologische Erkrankungen	Metastasiertes hormonrefraktäres Prostatakarzinom (mCRPC)
Eribulin	01.05.2011	Verfahren abgeschlossen	Onkologische Erkrankungen	Mammakarzinom
Ipilimumab	01.08.2011	Verfahren abgeschlossen	Onkologische Erkrankungen	Fortgeschrittenes Melanom
Abirateronacetat	01.10.2011	Verfahren abgeschlossen	Onkologische Erkrankungen	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)
Vandetanib	15.03.2012	Verfahren abgeschlossen	Onkologische Erkrankungen	Medulläres Schilddrüsenkarzinom (MTC)
Vemurafenib	15.03.2012	Verfahren abgeschlossen	Onkologische Erkrankungen	Fortgeschrittenes Melanom (BRAF- V600 Mutation-positiv)
Tegafur, Gimeracil, Oteracil	01.07.2012	Verfahren abgeschlossen	Onkologische Erkrankungen	Magenkarzinom
Ruxolitinib (Chronisch myeloproliferative Erkrankungen)	15.09.2012	Verfahren abgeschlossen	Onkologische Erkrankungen	Chronisch Myeloproliferative Erkrankungen
Axitinib	01.10.2012	Verfahren abgeschlossen	Onkologische Erkrankungen	Nierenzellkarzinom
Decitabin	01.11.2012	Verfahren abgeschlossen	Onkologische Erkrankungen	Akute myeloische Leukämie (AML)
Crizotinib	15.11.2012	Verfahren abgeschlossen	Onkologische Erkrankungen	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)
Brentuximab Vedotin	01.12.2012	Verfahren abgeschlossen	Onkologische Erkrankungen	Hodgkin-, anaplastische großzellige Lymphome (HL, sALCL)
Pixantron	01.12.2012	Verfahren abgeschlossen	Onkologische Erkrankungen	Non-Hodgkin Lymphome (NHL)

Abirateronacetat (neues Anwendungsgebiet)	15.01.2013	Verfahren abgeschlossen	Onkologische Erkrankungen	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)
Aflibercept (metastasiertes kolorektales Karzinom)	01.03.2013	Verfahren abgeschlossen	Onkologische Erkrankungen	Metastasiertes kolorektales Karzinom (mCRC)
Vandetanib (Absatz 5b)	15.03.2013	Verfahren abgeschlossen	Onkologische Erkrankungen	Medulläres Schilddrüsenkarzinom (MTC)
Pertuzumab	01.04.2013	Verfahren abgeschlossen	Onkologische Erkrankungen	(HER2-positives) metastasiertes Mammakarzinom
Bosutinib	01.05.2013	Verfahren abgeschlossen	Onkologische Erkrankungen	Chronische myeloische Leukämie (CML)
Ponatinib	01.08.2013	Verfahren abgeschlossen	Onkologische Erkrankungen	Chronische myeloische Leukämie (CML); (Philadelphia-chromosom positive) akute Lymphoblastenleukämie (Ph+ ALL)
Vismodegib	15.08.2013	Verfahren abgeschlossen	Onkologische Erkrankungen	Basalzellkarzinom
Enzalutamid	01.09.2013	Verfahren abgeschlossen	Onkologische Erkrankungen	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)
Pomalidomid	01.09.2013	Verfahren abgeschlossen	Onkologische Erkrankungen	Rezidiertes oder refraktäres Multiples Myelom (rrMM)
Vemurafenib (Neubewertung nach Fristablauf)	15.09.2013	Verfahren abgeschlossen	Onkologische Erkrankungen	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)
Dabrafenib	01.10.2013	Verfahren abgeschlossen	Onkologische Erkrankungen	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)
Regorafenib	01.10.2013	Verfahren abgeschlossen	Onkologische Erkrankungen	Kolorektalkarzinom (KRK)
Afatinib	15.11.2013	Verfahren abgeschlossen	Onkologische Erkrankungen	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)
Ipilimumab (neues Anwendungsgebiet)	15.12.2013	Verfahren abgeschlossen	Onkologische Erkrankungen	Fortgeschrittenes Melanom
Radium-223-dichlorid	01.01.2014	Verfahren abgeschlossen	Onkologische Erkrankungen	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)

Trastuzumab Emtansin	01.01.2014	Verfahren abgeschlossen	Onkologische Erkrankungen	(HER2-positives) metastasiertes Mammakarzinom
Ruxolitinib (Chronisch myeloproliferative Erkrankungen, erneute Nutzenbewertung)	15.05.2014	Verfahren abgeschlossen	Onkologische Erkrankungen	Chronisch Myeloproliferative Erkrankungen
Siltuximab	15.06.2014	Verfahren abgeschlossen	Onkologische Erkrankungen	Multizentrischer Morbus Castleman (MCD)
Cabozantinib	01.08.2014	Verfahren abgeschlossen	Onkologische Erkrankungen	Medulläres Schilddrüsenkarzinom (MTC)
Eribulin (neues Anwendungsgebiet, Neubewertung nach Fristablauf)	01.08.2014	Verfahren abgeschlossen	Onkologische Erkrankungen	Mammakarzinom
Obinutuzumab	15.08.2014	Verfahren abgeschlossen	Onkologische Erkrankungen	Chronische lymphatische Leukämie (CLL)
Regorafenib (neues Anwendungsgebiet)	01.09.2014	Verfahren abgeschlossen	Onkologische Erkrankungen	Gastrointestinale Stromatumore (GIST)
Idelalisib	01.10.2014	Verfahren abgeschlossen	Onkologische Erkrankungen	Chronische lymphatische Leukämie (CLL); Follikuläres Lymphom (FL)
Sipuleucel-T	01.10.2014	Verfahren abgeschlossen	Onkologische Erkrankungen	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)
Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom)	01.11.2014	Verfahren abgeschlossen	Onkologische Erkrankungen	Chronische lymphatische Leukämie (CLL); Rezidiertes oder refraktäres Mantelzell-Lymphom (MCL)
Enzalutamid (neues Anwendungsgebiet)	01.01.2015	Verfahren abgeschlossen	Onkologische Erkrankungen	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)
Nintedanib (nicht-kleinzelliges Lungenkarzinom)	01.01.2015	Verfahren abgeschlossen	Onkologische Erkrankungen	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)
Ramucirumab	01.02.2015	Verfahren abgeschlossen	Onkologische Erkrankungen	Magenkarzinom
Ruxolitinib (Polycythaemia vera)	15.04.2015	Verfahren abgeschlossen	Onkologische Erkrankungen	Polycythaemia vera
Afatinib (Neubewertung nach Fristablauf)	15.05.2015	Verfahren abgeschlossen	Onkologische Erkrankungen	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)

Olaparib	01.06.2015	Verfahren abgeschlossen	Onkologische Erkrankungen	Ovarialkarzinom
Ceritinib	01.07.2015	Verfahren abgeschlossen	Onkologische Erkrankungen	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)
Lenvatinib	01.07.2015	Verfahren abgeschlossen	Onkologische Erkrankungen	Differenziertes Schilddrüsenkarzinom (DTC)
Nivolumab (Melanom)	15.07.2015	Verfahren abgeschlossen	Onkologische Erkrankungen	Fortgeschrittenes Melanom
Nivolumab (nicht-kleinzelliges Lungenkarzinom)	15.08.2015	Verfahren abgeschlossen	Onkologische Erkrankungen	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)
Pembrolizumab	15.08.2015	Verfahren abgeschlossen	Onkologische Erkrankungen	Fortgeschrittenes Melanom
Pertuzumab (neues Anwendungsgebiet)	01.09.2015	Verfahren abgeschlossen	Onkologische Erkrankungen	(HER2-positives) Mammakarzinom (neoadjuvante Therapie)
Dabrafenib (neues Anwendungsgebiet)	01.10.2015	Verfahren abgeschlossen	Onkologische Erkrankungen	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)
Panobinostat	01.10.2015	Verfahren abgeschlossen	Onkologische Erkrankungen	Rezidiertes oder refraktäres Multiples Myelom (rrMM)
Pomalidomid (erneute Nutzenbewertung)	01.10.2015	Verfahren abgeschlossen	Onkologische Erkrankungen	Rezidiertes oder refraktäres Multiples Myelom (rrMM)
Regorafenib (Neubewertung nach Fristablauf)	01.10.2015	Verfahren abgeschlossen	Onkologische Erkrankungen	Kolorektalkarzinom (KRK)
Trametinib	01.10.2015	Verfahren abgeschlossen	Onkologische Erkrankungen	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)
Blinatumomab	15.12.2015	Verfahren abgeschlossen	Onkologische Erkrankungen	Akute lymphatische Leukämie (ALL)
Carfilzomib	15.12.2015	Verfahren abgeschlossen	Onkologische Erkrankungen	Multiples Myelom (MM)
Cobimetinib	15.12.2015	Verfahren abgeschlossen	Onkologische Erkrankungen	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)
Crizotinib (neues Anwendungsgebiet)	01.01.2016	Verfahren abgeschlossen	Onkologische Erkrankungen	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)

Anhang B: Bewertungsmatrizen (exemplarisch)

Angabe zum NBV	AWG (i. w. S.)	Bezeichnung des MI	Bibliogr. Angabe der angeführten Validierungsstudie	Zuordnung entspr. der vordergründigen Zweckbestimmung	Angabe der Studienpopulation	Grad an Übereinstimmung in Bezug auf die Studienpopulation					
						Tumor-entität	Patientencharakteristika				
							Alter	Anteil Frauen	ECOG PS gw.	Anteil Kaukasier/ weiß	Anteil (West-) Europa
Crizotinib (neues AWG)	NSCLC	EORTC QLQ-C30			nicht vorbehandelte erwachsene Patienten mit ALK-positivem <u>NSCLC</u> in lokal fortgeschrittenem oder metastasiertem Stadium	Fehlend	(Median) 52 J. [I]; 54 J. [K]	60,5 %, N = 104 [I] (abweichend: N = 68 [I]; 63,2 %, N = 108 [K] (abweichend: N = 63 [K])	0,714, N = 172 [I]; 0,766, N = 171 [K] (Status 0: 33,7 % [I]; 27,5 % [K], Status 1: 61,0 % [I]; 68,4 % [K], Status 2: 5,2 % [I]; 4,1 % [K])	Weiß: 52,9 % [I]; 49,7 % [K] (Schwarze: 0,0 % [I]; 2,3 % [K], Asiaten: 44,8 % [I]; 46,8 % [K], Andere: 2,3 % [I]; 1,2 % [K])	Europa: 35,5 % [I]; 38,6 % [K] (Nordamerika: 11,0 % [I]; 7,6 % [K], Asien: 41,9 % [I]; 42,7 % [K], Andere: 11,0 % [I]; 9,9 % [K])
			Osoba et al., 1998	Nachweis der klinischen Relevanz via Responsekriterium	Rezidivierender oder metastasierter <u>Brustkrebs</u> ; kleinzelliges Lungenkarzinom (<u>SCLC</u>)		(MW) 55,2 J. (Breast Cancer, N = 246) bzw. 57,6 J. (SCLC, N = 111)	100 % (Breast Cancer, N = 246) bzw. n. b. (SCLC, N = 111)	1,012, N = 246 (Breast Cancer, Status 0: 23,6 %, Status 1: 51,6 %, Status 2: 24,8 %) bzw. 0,809, N = 111 (SCLC, Status 0: 29,2 %, Status 1: 60,7 %, Status 2: 10,1 %)	Ohne Angaben	Ohne Angaben
							Rel. Abw. zur Val. Studie: - 4,0 % (Breast Cancer) bzw. - 8,0 % (SCLC)	Rel. Abw. zur Val. Studie: - 38,2 % (Breast Cancer), p < 0,0001	Rel. Abw. zur Val. Studie: - 26,9 % (Breast Cancer), p < 0,0001 bzw. - 8,5 % (SCLC), p = 0,2508		

AWG: Anwendungsgebiet, *ECOG PS gw.*: Häufigkeitsgewichtete Maßzahl zum *Eastern Cooperative Oncology Group Performance Status*, MI: Messinstrument der gesundheitsbezogenen Lebensqualität, MW: Mittelwert, n. b.: nicht berichtet, NBV: Nutzenbewertungsverfahren, (N)SCLC: (Nicht-)kleinzelliges Bronchialkarzinom, Rel. Abw. zur Val.Studie: Relative Abweichung zur Validierungsstudie, SD: Standardabweichung. [I]: Interventionsgruppe, [K] Kontrollgruppe. Anmerkung: Getrennte Angaben zu den Studienarmen gehen als Mittelwert in die Berechnung der relativen Abweichung zur Validierungsstudie ein.

Bibliographische Angabe der angeführten Validierungsstudie	Studiendesign	Anzahl einbezogener Patienten	Aussagen zu den Messzeitpunkten	Klinische Relevanzkriterien und deren Bestimmungsmethoden	Angaben zur anker- oder verteilungsbasierten Interpretation klinisch relevanter Änderungen und dem Referenzmodell	Auflistungen zu Art und Umfang der Bezugnahme in den Dossiers
Osoba et al., 1998	2 randomisierte Studien a) Studie der Phase III im Bereich des rezidierten oder metastasierten Mammakarzinoms b) (nicht näher spezifizierte) Studie im Feld des kleinzelligen Lungenkarzinoms	a) „qualifiziert“: 300; Fragebogen ausgefüllt an T1 + T2 (Analysepopulation): 246 b) „qualifiziert“: 111; Fragebogen ausgefüllt an T1 + T2 (Analysepopulation): 80	a) "before the first cycle of chemotherapy" (T1; EORTC QLQ-C30), "day 1 of the third cycle of chemotherapy" (T2; EORTC QLQ-C30 + SSQ), "just before the sixth cycle of chemotherapy" (T3; EORTC QLQ-C30 + SSQ) b) "before the first cycle of chemotherapy" (T1; EORTC QLQ C30), "week 4 of chemotherapy" (T2; EORTC QLQ-C30 + SSQ), "week 7 of chemotherapy" (T3; EORTC QLQ-C30 + SSQ)	a) Rangkorrelationsanalyse nach Spearman b) Assoziationen hinsichtlich Richtung und Größe von Änderungen in den Score-Werten (u. a. relativer Anteil am Zielwert ["a little" 5 - 10 Punkte, "moderate" 10 - 20 Punkte, "very much" > 20 Punkte]) c) Regressionsanalyse (mittlere EORTC QLQ-C30-Score-Änderung je Unterschied zur benachbarten Kategorie)	ankerbasiert (vgl. SSQ [self-report] mit den berichteten Zustandsänderungen "a little", "moderately", "very much") <u>Erläuterung:</u> SSQ (Subjective Significance Questionnaire) ["The four items in the SSQ were intended to correspond to the physical functioning, emotional functioning, social functioning, and global quality of life (QL) domains of the QLQ-C30"]	Afatinib: QL, PF, RF, CF, EF, SF Afatinib (Neubewertung nach Fristablauf): QL, PF, RF, CF, EF, SF Carfilzomib: QL Cobimetinib: PF, SF Crizotinib: QL, PF, RF, EF, SF Crizotinib (neues Anwendungsgebiet): QL, PF, RF, CF, EF, SF Dabrafenib: EF Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom): CF Nintedanib (nicht-kleinzelliges Lungenkarzinom): QL, CF Pembrolizumab: SF Ramucirumab: EF Ruxolitinib (Chronisch myeloproliferative Erkrankungen): QL, RF Ruxolitinib (Chronisch myeloproliferative Erkrankungen, erneute Nutzenbewertung): QL, PF, RF, SF Trametinib: QL, PF, RF, CF, EF, SF

QL: Globaler Gesundheitsstatus, PF: Körperliche Funktion, RF: Rollenfunktion, CF: Kognitive Funktion, EF: Emotionale Funktion, SF: Soziale Funktion (Skalen des Fragebogens EORTC QLQ-C30); SSQ: Subjective Significance Questionnaire.

Anhang C: Eingeschlossene Referenzstudien

Kurztitel	Kategorisierung (CAVE: Betreffend den EQ-5D (VAS) werden sämtliche Referenzstudien betrachtet. Übereinstimmung im Anwendungsgebiet (AWG) ist insoweit im übertragenen Sinne auf die generische Konzeption des Fragebogens bezogen [vgl. mitunter (X)].)		Bibliographische Angabe (Quelle: Modul 4 der Dossiers, Dossier-/Nutzenbewertungen [ggf. wahlweise])
	Standardreferenz	Referenz mit AWG-Übereinstimmung	
Aaronson et al., 1993	X	X	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):365-376.
Akakura et al., 2011		X (Publikation in japanischer Sprache; nur Abstract in Englisch verfasst)	Akakura K, Matsuzaki K, Kobayashi T, Kitoh H, Mizoguchi K, Tomikawa G, et al. [Evaluation of utility index of quality of life (QOL) in prostate cancer patients: comparison of QOL utility index EuroQol-5D (EQ-5D) and visual analogue scale (VAS) with health-related QOL questionnaires SF-36 and EPIC]. Nihon Hinyokika Gakkai Zasshi. 2011;102(1):9-13. Epub 2011/04/28.
Al-Batran und Ajani, 2010			Al-Batran SE, Ajani JA. Impact of chemotherapy on quality of life in patients with metastatic esophagogastric cancer. Cancer. 2010;116(11):2511-2518.
Askew et al., 2009		X	Askew RL, Xing Y, Palmer JL, Cella D, Moye LA, Cormier JN. Evaluating minimal important differences for the FACT-Melanoma quality of life questionnaire. Value Health. 2009;12(8):1144-50.
Bjordal et al., 2000			Bjordal K, de Graeff A, Fayers PM, Hammerlid E, van Pottelsberghe C, Curran D, et al. A 12 country field study of the EORTC QLQ-C30 (version 3.0) and the head and neck cancer specific module (EORTC QLQ-H&N35) in head and neck patients. European Journal of Cancer. 2000;36(14):1796-807.
Braun et al., 2011		X	Braun DP, Gupta D, Staren ED. Quality of life assessment as a predictor of survival in non-small cell lung cancer. BMC Cancer. 2011;11:353.
Brazier et al., 1993		(X)	Brazier J, Jones N, Kind P. Testing the validity of the Euroqol and comparing it with the SF-36 health survey questionnaire. Quality of Life Research. 1993;2(3):169-80.

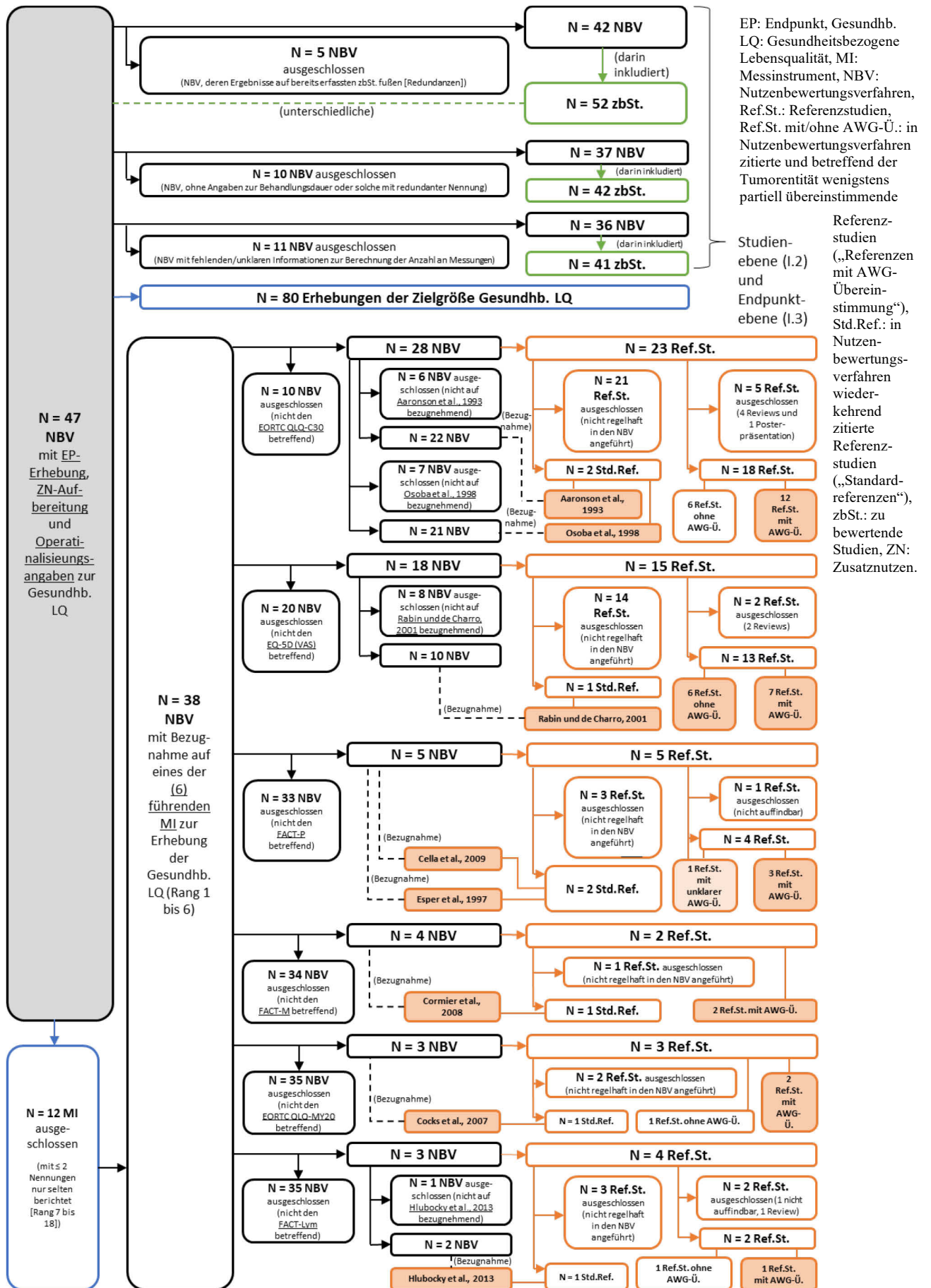
Brooks et al., 1991		(X)	Brooks RG, Jendteg S, Lindgren B, Persson U, Bjork S. EuroQol: health-related quality of life measurement. Results of the Swedish questionnaire exercise. Health Policy. 1991;18(1):37-48. Epub 1991/05/09.
Carducci et al., 2005		X	Carducci MA, Nelson JB, Vogelzang NJ, Mulani P. Determination of clinically meaningful change (CMC) for functional assessment of cancer therapy-prostate (FACT-P). J Clin Oncol 2005;23(16S(June 1 Supplement)):8077.
Carter et al., 2008			Carter GC, Liepa AM, Zimmermann AH, Morschhauser F. Validation of the Functional Assessment of Cancer Therapy - Lymphoma (FACT-Lym) in patients with relapsed/refractory mantle cell lymphoma. ASH Annual Meeting Abstracts; 2008 Dec 7.
Cella et al., 2005		X (nicht auffindbar; keine Detailanalysen möglich)	Cella D, Webster K, Cashy J, Kutikova L, Burgess MF, Lin BK, et.al. Development of a Measure of Health-Related Quality of Life for Non-Hodgkin's Lymphoma Clinical Research: The Functional Assessment of Cancer Therapy - Lymphoma (FACT-Lym). ASH Annual Meeting Abstracts 2005; 106: 750.
Cella et al., 2009	X	X	Cella D, Nichol MB, Eton D, Nelson JB, Mulani P. Estimating clinically meaningful changes for the Functional Assessment of Cancer Therapy--Prostate: results from a clinical trial of patients with metastatic hormone-refractory prostate cancer. Value Health 2009;12(1):124-9.
Cella, 2012		(nicht auswertbar)	Cella, D. 2012. The validity of version 4 of the Functional Assessment of Cancer Therapy - Prostate (FACT-P).
Cocks et al., 2007	X	X	Cocks K, Cohen D, Wisloff F, Sezer O, Lee S, Hippe E, Gimsing P, Turesson I, Hajek R, Smith A, Graham L, Phillips A, Stead M, Velikova G, Brown J, Group EQoL. An international field study of the reliability and validity of a disease-specific questionnaire module (the QLQ-MY20) in assessing the quality of life of patients with multiple myeloma. European journal of cancer. 2007;43(11):1670-8.
Cocks et al., 2011			Cocks K, King MT, Velikova G, Martyn St-James M, Fayers PM, Brown JM. Evidence-based guidelines for determination of sample size and interpretation of the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2011;29(1):89-96.
Cormier et al., 2008	X	X	Cormier JN, Ross MI, Gershenwald JE, Lee JE, Mansfield PF, Camacho LH, et al. Prospective Assessment of the Reliability, Validity, and Sensitivity to Change of the Functional Assessment of Cancer Therapy-Melanoma Questionnaire. Cancer. 2008;112(10):2249-57.
de Boer et al., 2004		X	de Boer AG, van Lanschot JJ, Stalmeier PF, van Sandick JW, Hulscher JB, de Haes JC, et al. Is a single-item visual analogue scale as valid, reliable and responsive as multi-item scales in measuring quality of life? Quality of Life Research. 2004;13(2):311-20.

Delforge et al., 2012		X	Delforge M., Dhawan R., et al. 2012. Health-related quality of life in elderly, newly diagnosed multiple myeloma patients treated with VMP vs. MP: results from the VISTA trial. Eur J Haematol, 89(1): 16-27.
Dubois et al., 2006		X	Dubois D, Dhawan R, van de Velde H, Esseltine D, Gupta S, Viala M, de la Loge C. Descriptive and prognostic value of patient-reported outcomes: the bortezomib experience in relapsed and refractory multiple myeloma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2006;24(6):976-82.
Esper et al., 1997	X	X	Esper P, Mo F, Chodak G, Sinner M, Cella D, Pienta KJ. Measuring quality of life in men with prostate cancer using the functional assessment of cancer therapy-prostate instrument. Urology 1997;50(6):920-8.
Fayers und Bottomley, 2002			Fayers P, Bottomley A. Quality of life research within the EORTC-the EORTC QLQ-C30. European Organisation for Research and Treatment of Cancer. Eur J Cancer. 2002;38 Suppl 4:S125-33. Epub 2002/02/23.
Greiner et al., 2005		(X)	Greiner W, Claes C, Busschbach JJ, von der Schulenburg JM. Validating the EQ-5D with time trade off for the German population. Eur J Health Econ. 2005;6(2):124-30. Epub 2005/06/01.
Groenvold et al., 1997			Groenvold M, Klee MC, Sprangers MA, Aaronson NK. Validation of the EORTC QLQ-C30 quality of life questionnaire through combined qualitative and quantitative assessment of patient-observer agreement. Journal of clinical epidemiology. 1997;50(4):441-450.
Guyatt et al., 2002			Guyatt, GH, Osboa, D, Wu A et al., Methods to Explain the Clinical Significance of Health Status Measures, Mayo Clin Proc. 2002;77:371-383.
Hjermstad et al., 1995		X	Hjermstad MJ, Fossa SD, Bjordal K, Kaasa S. Test/retest study of the European Organization for Research and Treatment of Cancer Core Quality-of-Life Questionnaire. J Clin Oncol. 1995;13(5):1249-54.
Hlubocky et al., 2013	X	X	Hlubocky FJ, Webster K, Cashy J, Beaumont J, Cella D. The Development and Validation of a Measure of Health-Related Quality of Life for Non-Hodgkin's Lymphoma: The Functional Assessment of Cancer Therapy-Lymphoma (FACTLym). Lymphoma. 2013;2013.
Hurst et al., 1997		(X) (im Wesentlichen redundant zu Rabin und de Charro, 2001)	NP Hurst, P Kind, D Ruta, M Hunter & A Stubbings 1997. Measuring the Health-related Quality of Life in Rheumatoid Arthritis: Validity, Responsivness and Reliability of Euroqol (ED-5D). British Journal of Rheumatology 36, 551-9.
Johnson et al., 1998		(X)	Johnson JA, Coons SJ, Ergo A, Szava-Kovats G. Valuation of EuroQOL (EQ-5D) health states in an adult US sample. Pharmacoeconomics. 1998;13(4):421-33. Epub 1998/03/08.
King, 1996			King MT. The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. Qual Life Res 1996; 5(6): 555-567.

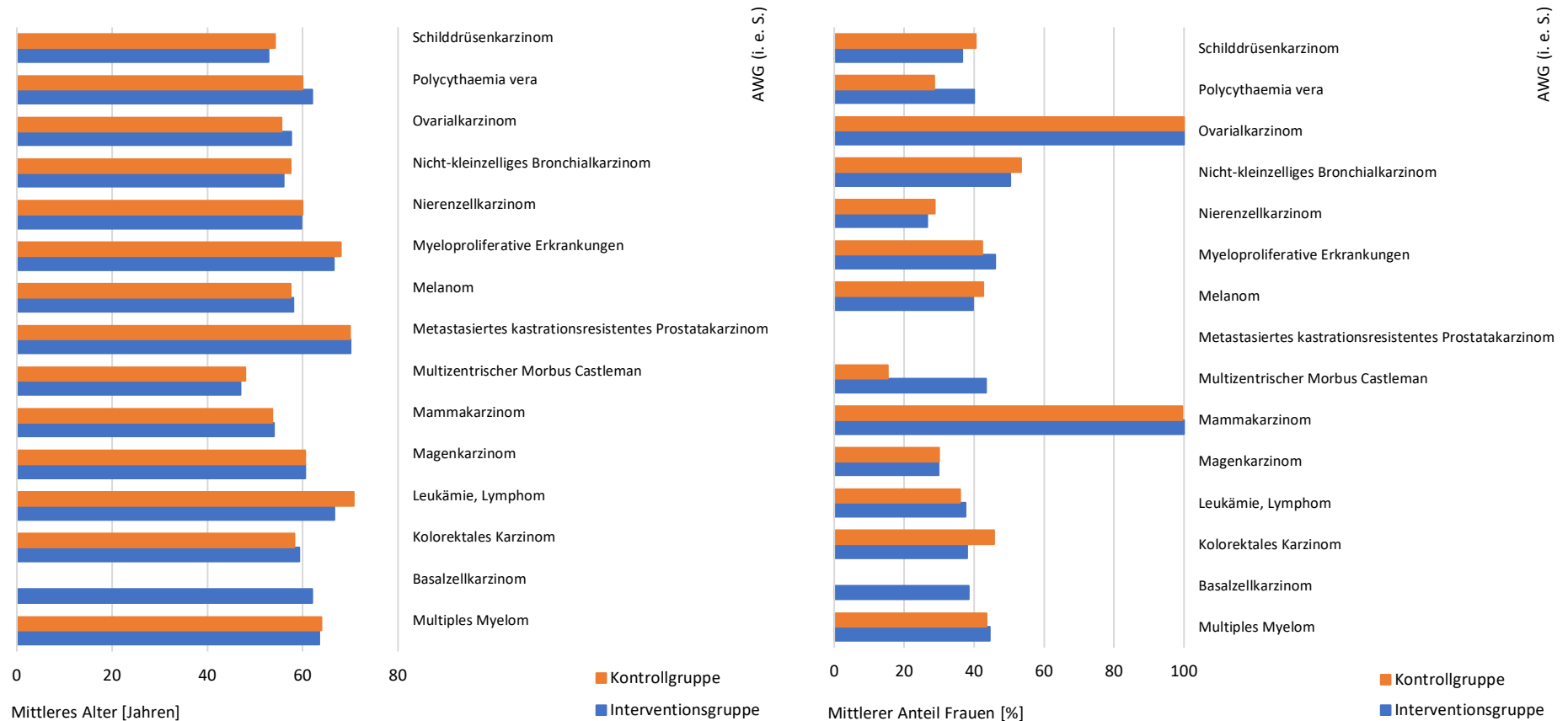
Kvam et al., 2011		X	Kvam AK, Fayers PM, Wisloff F. Responsiveness and minimal important score differences in quality-of-life questionnaires: a comparison of the EORTC QLQ-C30 cancer-specific questionnaire to the generic utility questionnaires EQ-5D and 15D in patients with multiple myeloma. <i>European journal of haematology</i> . 2011;87(4):330-7.
Leleu et al., 2013			Leleu X., Petrucci M.T., et al. 2013. Psychometric Performance Of The EORTC Quality-Of-Life Core Questionnaire (QLQ-C30) and QLQ-Multiple Myeloma (QLQ-MY20) In Relapsed/Refractory Multiple Myeloma (RRMM). Session: 901 - Health Services and Outcomes Research: Poster I. Abstract 1721, Conference: ASH Annual Meeting 2013 (New Orleans, USA).
McLachlan et al., 1998		X	McLachlan S-A, Devins G, Goodwin P. Validation of the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (QLQ-C30) as a measure of psychosocial function in breast cancer patients. <i>European Journal of Cancer</i> . 1998;34(4):510-7.
Montazeri et al., 2001			Montazeri A, Milroy R, Hole D, McEwen J, Gillis CR. Quality of life in lung cancer patients: as an important prognostic factor. <i>Lung cancer</i> . 2001;31(2):233-40.
Osoba et al., 1997		X	Osoba D, Aaronson N, Zee B, Sprangers M, te Velde A. Modification of the EORTC QLQ-C30 (version 2.0) based on content validity and reliability testing in large samples of patients with cancer. The Study Group on Quality of Life of the EORTC and the Symptom Control and Quality of Life Committees of the NCI of Canada Clinical Trials Group. <i>Qual Life Res</i> . 1997;6(2):103-8.
Osoba et al., 1998	X		Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. <i>J Clin Oncol</i> 1998; 16 (1): 139-44.
Parmar et al., 2005		X	Parmar V, Badwe R, Hawaldar R, Rayabhattachanavar S, Varghese A, Sharma R, et al. Validation of EORTC quality-of-life questionnaire in Indian women with operable breast cancer. <i>National Medical Journal of India</i> . 2005;18(4):172-7.
Pickard et al., 2007a		X	Pickard AS, Neary MP, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. <i>Health and quality of life outcomes</i> . 2007;5:70.
Pickard et al., 2007b			Pickard AS, Wilke CT, Lin HW, Lloyd A. Health utilities using the EQ-5D in studies of cancer. <i>Pharmacoeconomics</i> . 2007;25(5):365-84.
Rabin und de Charro, 2001	X	(X)	Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. <i>Ann Med</i> . 2001;33(5):337-43.
Sadighi et al., 2006		X (nach Sichtung nicht als originäre Validierungsstudie qualifiziert)	Sadighi S, Mohagheghi MA, Montazeri A, Sadighi Z. Quality of life in patients with advanced gastric cancer: a randomized trial comparing docetaxel, cisplatin, 5-FU (TCF) with epirubicin, cisplatin, 5-FU (ECF). <i>BMC cancer</i> . 2006;6:274.

Schuette et al., 2012		X	Schuette W, Tesch H, Büttner H, Krause T, Soldatenkova V, Stoffregen C. Second-line treatment of Stage III/IV Non-Small-Cell Lung Cancer (NSCLC) with pemetrexed in routine clinical practice: Evaluation of performance status and health-related quality of life. BMC Cancer. 2012;12:14.
Teckle et al., 2011		X	Teckle P, Peacock S, McTaggart-Cowan H, van der Hoek K, Chia S, Melosky B, et al. The ability of cancer-specific and generic preference-based instruments to discriminate across clinical and self-reported measures of cancer severities. Health Qual Life Outcomes. 2011;9:106.
Walters und Brazier, 2005			S. J. Walters & J. E. Brazier 2005. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation, 14, 1523-32.
Webster et al., 1999		X (wenngleich weitgehend unbestimmt)	Webster K, Odom L, Peterman A, Lent L, Cella D. The functional assessment of chronic illness therapy (FACIT) measurement system: validation of version 4 the core questionnaire. Qual Life Res 1999; 8(7): 604.
Wisloff et al., 1996		X	Wisloff F., Eika S., et al. 1996. Measurement of health-related quality of life in multiple myeloma. Nordic Myeloma Study Group. Br J Haematol, 92(3): 604-613.
Wyrwich et al., 1999			WYRWICH, K. W., NIENABER, N. A., TIERNEY, W. M. & WOLINSKY, F. D. 1999. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. Medical Care, 37, 469-478.
Zeng et al., 2012			L. Zeng, E. Chow, L. Zhang, L. M. Tseng, M. F. Hou, A. Fairchild, V. Vassiliou, R. Jesus-Garcia, M. A. Alm El-Din, A. Kumar, F. Forges, W. C. Chie, G. Bedard & A. Bottomley 2012. An international prospective study establishing minimal clinically important differences in the EORTC QLQ-BM22 and QLQ-C30 in cancer patients with bone metastases. Supportive care in cancer : official journal of the Multinational Association of Supportive Care in Cancer, 20, 3307-13.

Anhang D: Flussdiagramm zur spezifischen Literatursuche



Anhang E: Entsprechungsgrade von Patientencharakteristika innerhalb der zu bewertenden Studien

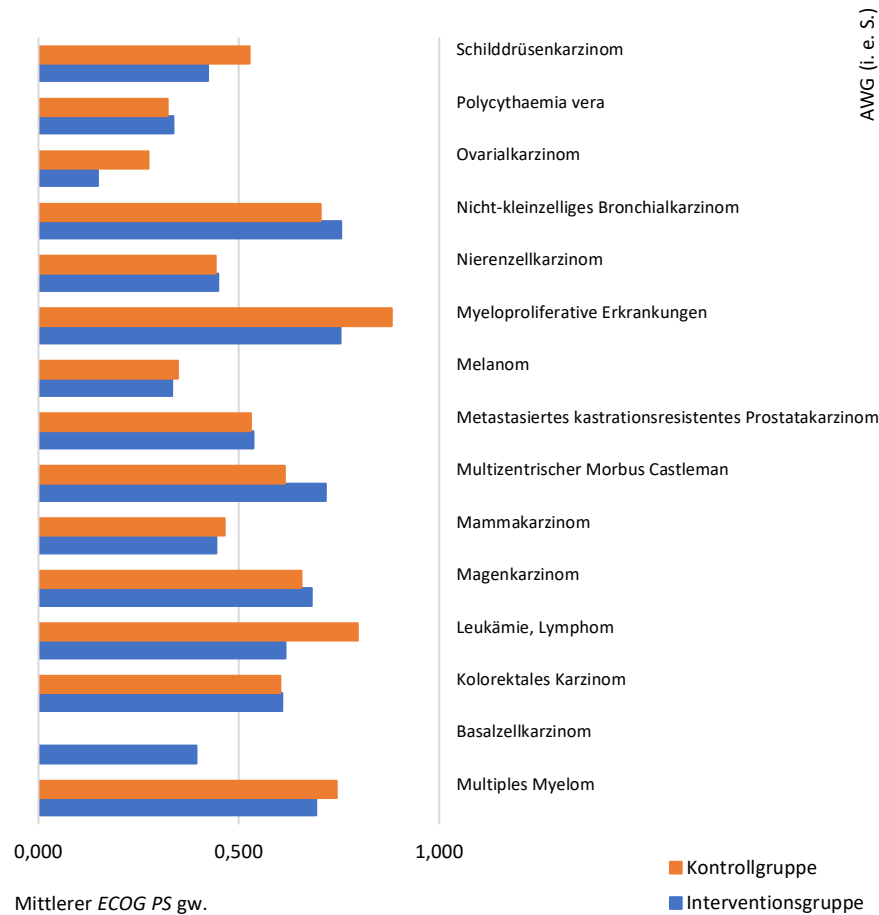


Anmerkung: Unberücksichtigt sind etwaige Fremdstudien; ohne Redundanzen. Der Verteilung zugrunde liegen entsprechende Angaben zum Median (ersatzweise Mittelwert).
 AWG: Anwendungsgebiet. N = 42 Nutzenbewertungsverfahren, 51 zu bewertende Studien.

Anmerkung: Unberücksichtigt sind etwaige Fremdstudien; ohne Redundanzen.
 AWG: Anwendungsgebiet. N = 42 Nutzenbewertungsverfahren, 51 zu bewertende Studien.

Abb. E1: Unterschiede innerhalb der zu bewertenden Studien; hier: Alter
 (Quelle: Eigene Darstellung).

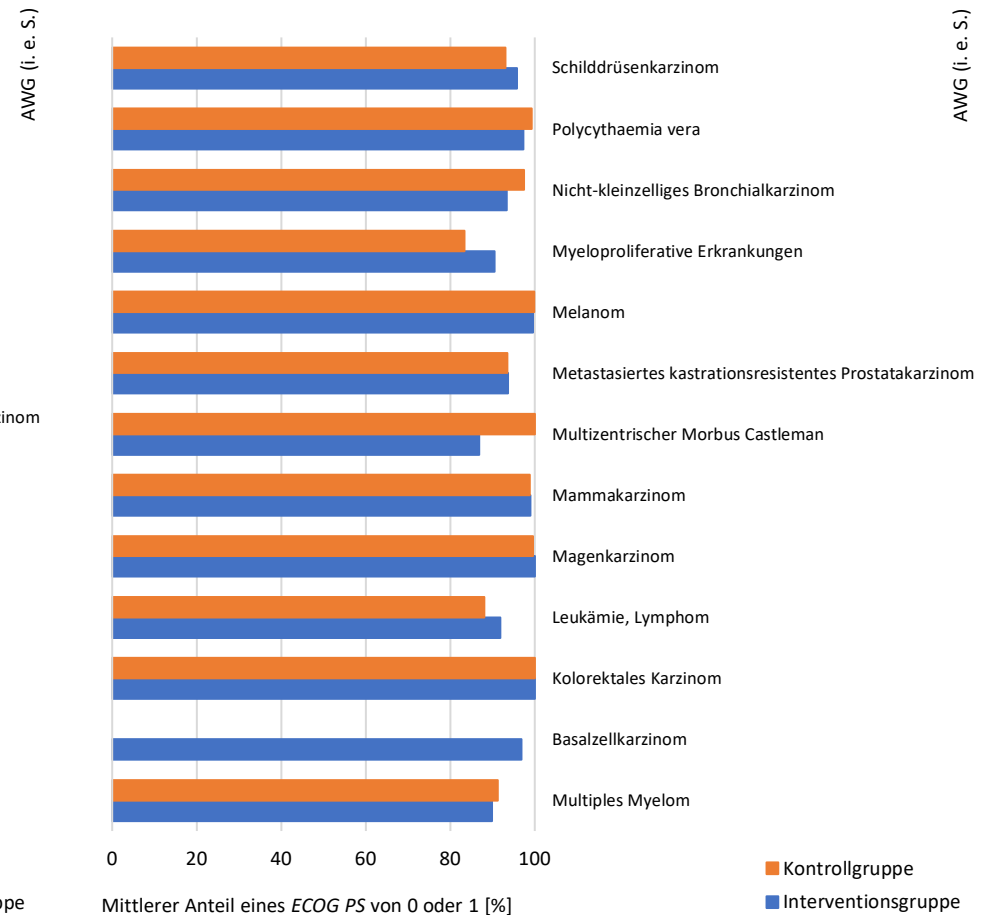
Abb. E2: Unterschiede innerhalb der zu bewertenden Studien; hier: Geschlecht
 (Quelle: Eigene Darstellung).



Anmerkung: Unberücksichtigt sind etwaige Fremdstudien; ohne Redundanzen. Konvention: Studie ausgeschlossen, sofern nicht zu $\geq 95\%$ diskriminante Angaben vorliegen. *Karnofsky Performance Status (KPS)* ggf. in *Eastern Cooperative Oncology Group Performance Status (ECOG PS)* umgerechnet.

AWG: Anwendungsgebiet. N = 40 Nutzenbewertungsverfahren, 49 zu bewertende Studien.

Abb. E3: Unterschiede innerhalb der zu bewertenden Studien; hier: Grad der Erkrankungsschwere (ECOG PS gw.) (Quelle: Eigene Darstellung).



Anmerkung: Unberücksichtigt sind etwaige Fremdstudien; ohne Redundanzen. Konvention: Studie ausgeschlossen, sofern nicht zu $\geq 99\%$ diskriminante Angaben vorliegen. *Karnofsky Performance Status (KPS)* ggf. in *Eastern Cooperative Oncology Group Performance Status (ECOG PS)* umgerechnet.

AWG: Anwendungsgebiet. N = 35 Nutzenbewertungsverfahren, 44 zu bewertende Studien.

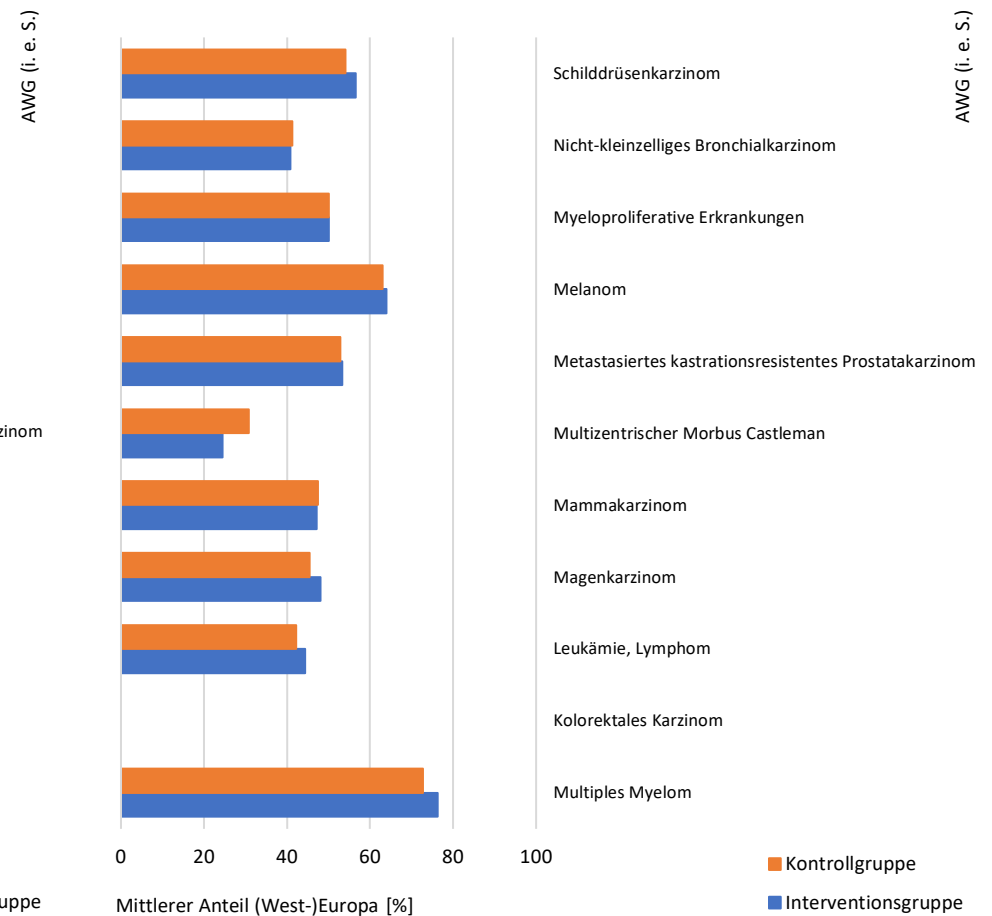
Abb. E4: Unterschiede innerhalb der zu bewertenden Studien; hier: Grad der Erkrankungsschwere (ECOG PS von 0 od. 1) (Quelle: Eigene Darstellung)



Anmerkung: Unberücksichtigt sind etwaige Fremdstudien; ohne Redundanzen. Konvention: Sofern ausschließlich für die Gesamtstudienpopulation berichtet wurde, werden entsprechende Angaben fiktiv auf die Interventions- und Kontrollgruppe übertragen.

AWG: Anwendungsgebiet. N = 39 Nutzenbewertungsverfahren, 47 zu bewertende Studien.

Abb. E5: Unterschiede innerhalb der zu bewertenden Studien; hier: Ethnizität (Quelle: Eigene Darstellung).



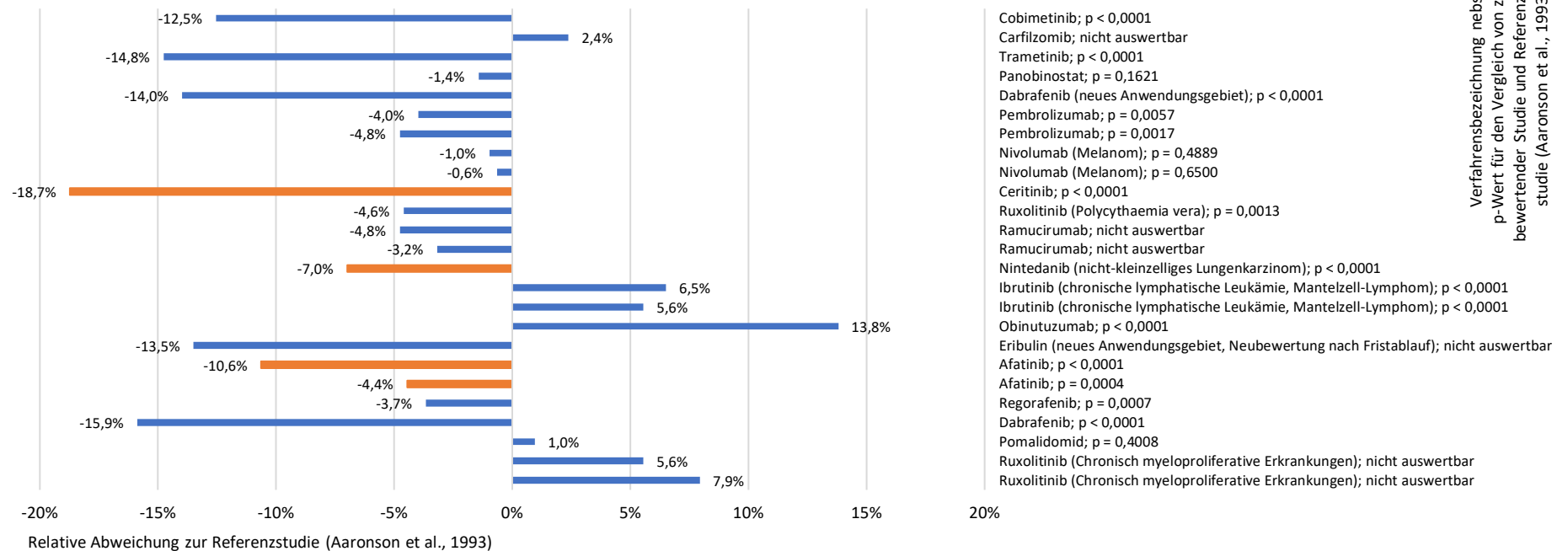
Anmerkung: Unberücksichtigt sind etwaige Fremdstudien; ohne Redundanzen. Konvention: Wurden West- und Osteuropa separat berichtet, werden die Regionen vorliegend zusammengefasst.

AWG: Anwendungsgebiet. N = 23 Nutzenbewertungsverfahren, 27 zu bewertende Studien.

Abb. E6: Unterschiede innerhalb der zu bewertenden Studien; hier: Geographische Region (Quelle: Eigene Darstellung).

Anhang F: Entsprechungsgrade von Patientencharakteristika bei Standardreferenzen

F-I: Abweichungen zur Standardreferenz; hier: *EORTC QLQ-C30*



Verfahrensbezeichnung nebst p-Wert für den Vergleich von zu bewertender Studie und Referenzstudie (Aaronson et al., 1993)

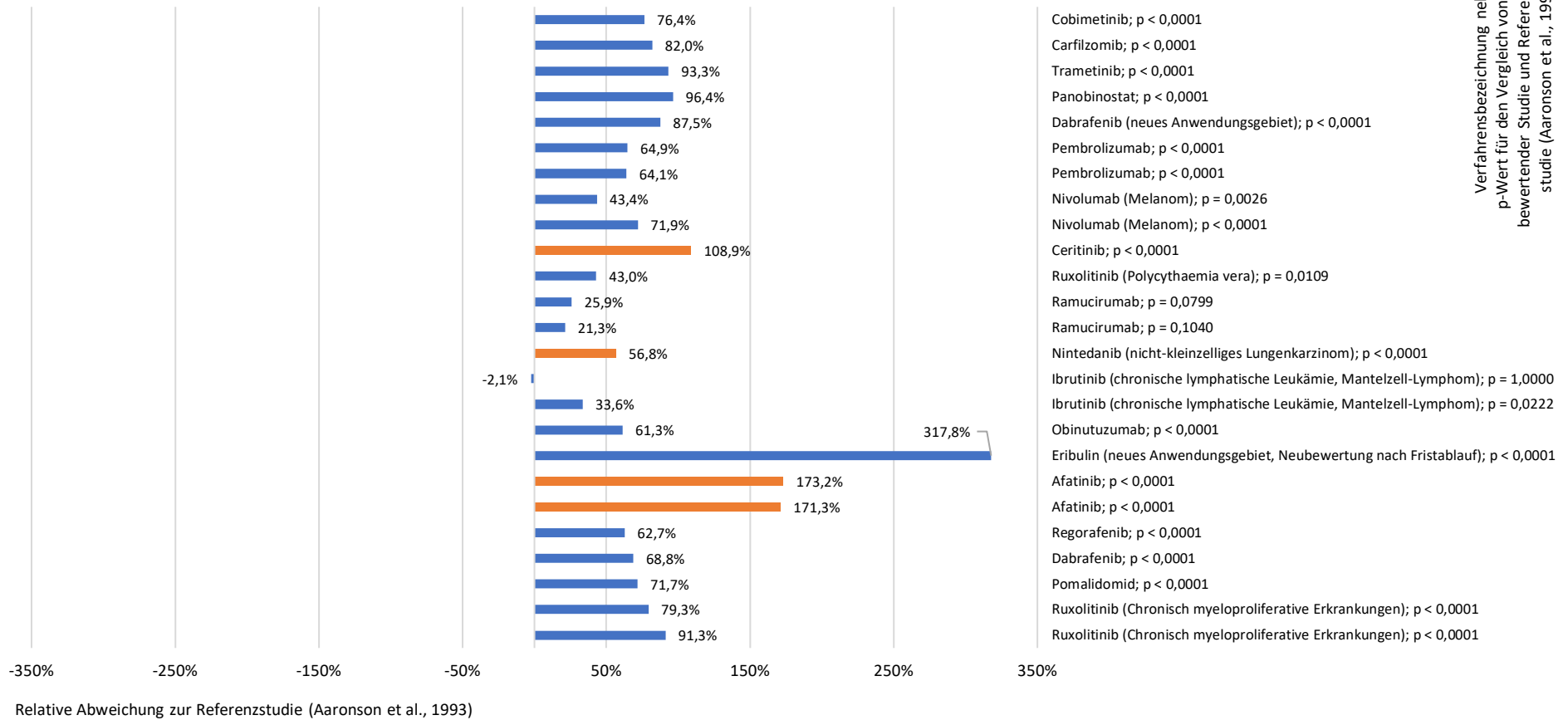
Aaronson et al., 1993: Mean age: 63 years (SD: 10 years) [64].

Operationalisierung: Alter [Jahren].

Anmerkung: „nicht auswertbar“ bei fehlenden Angaben zu Mittelwert und Standardabweichung. Eine orange Füllfarbe im Diagramm kennzeichnet zu bewertende Studien, deren Anwendungsgebiet dem der Standardreferenz ganz oder teilweise entspricht.

N = 22 Nutzenbewertungsverfahren, 25 zu bewertende Studien.

Abb. F1: Abweichung zur Standardreferenz Aaronson et al., 1993; hier: Alter (Quelle: Eigene Darstellung).



Verfahrensbezeichnung nebst p-Wert für den Vergleich von zu bewertender Studie und Referenzstudie (Aaronson et al., 1993)

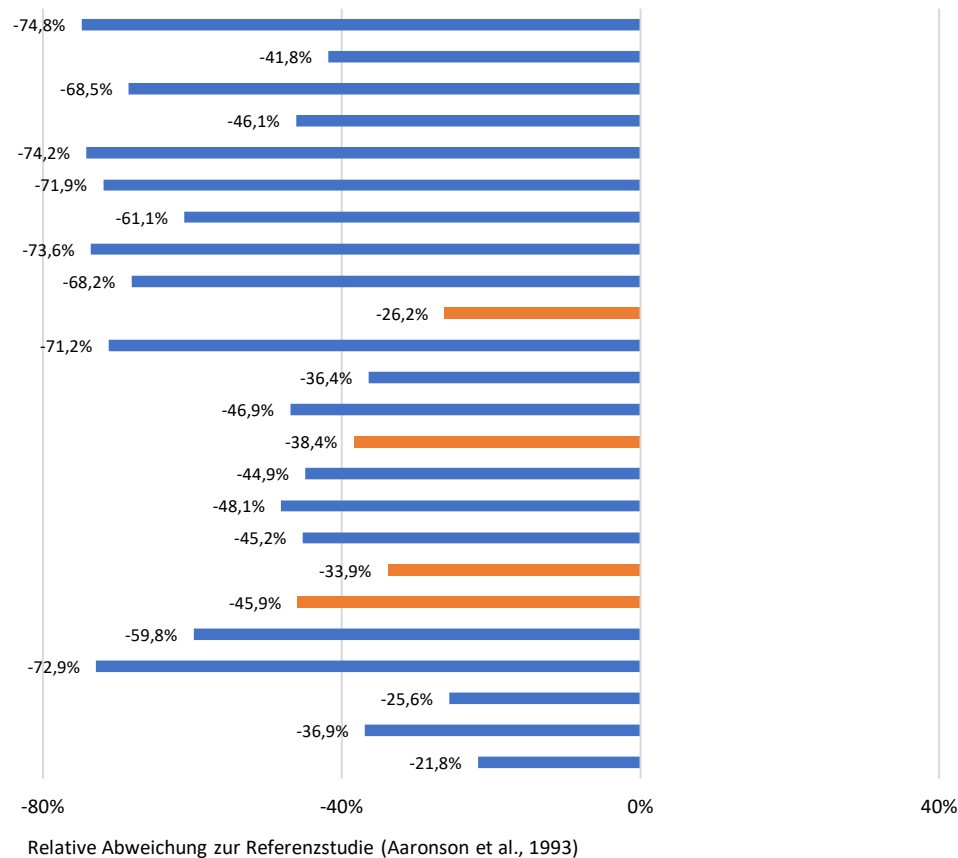
Aaronson et al., 1993: *Female* 24 % [64].

Operationalisierung: Anteil Frauen.

Anmerkung: Eine orange Füllfarbe im Diagramm kennzeichnet zu bewertende Studien, deren Anwendungsgebiet dem der Standardreferenz ganz oder teilweise entspricht.

N = 22 Nutzenbewertungsverfahren, 25 zu bewertende Studien.

Abb. F2: Abweichung zur Standardreferenz Aaronson et al., 1993; hier: Geschlecht (Quelle: Eigene Darstellung).



- Cobimetinib; p < 0,0001
- Carfilzomib; p < 0,0001
- Trametinib; p < 0,0001
- Panobinostat; p < 0,0001
- Dabrafenib (neues Anwendungsgebiet); p < 0,0001
- Pembrolizumab; p < 0,0001
- Pembrolizumab; p < 0,0001
- Nivolumab (Melanom); p < 0,0001
- Nivolumab (Melanom); p < 0,0001
- Ceritinib; p = 0,0002
- Ruxolitinib (Polycythaemia vera); p < 0,0001
- Ramucirumab; p < 0,0001
- Ramucirumab; p < 0,0001
- Nintedanib (nicht-kleinzelliges Lungenkarzinom); p < 0,0001
- Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom); p < 0,0001
- Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom); p < 0,0001
- Eribulin (neues Anwendungsgebiet, Neubewertung nach Fristablauf); p < 0,0001
- Afatinib; p < 0,0001
- Afatinib; p < 0,0001
- Regorafenib; p < 0,0001
- Dabrafenib; p < 0,0001
- Pomalidomid; p < 0,0001
- Ruxolitinib (Chronisch myeloproliferative Erkrankungen); p < 0,0001
- Ruxolitinib (Chronisch myeloproliferative Erkrankungen); p < 0,0001

Verfahrensbezeichnung nebst p-Wert für den Vergleich von zu bewertender Studie und Referenzstudie (Aaronson et al., 1993)

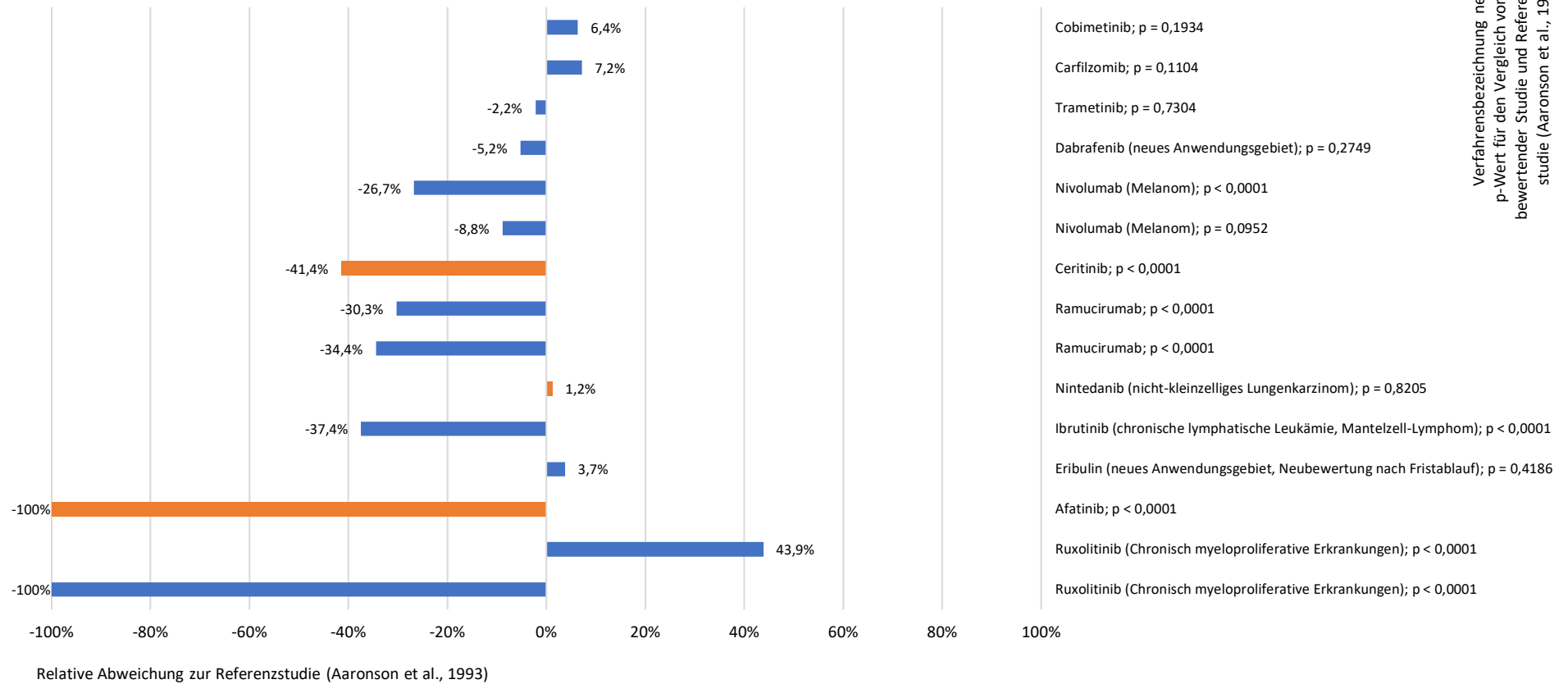
Aaronson et al., 1993: Performance status 0: 20,7 %, 1: 50,8 %, 2: 22,0 %, 3: 6,1 %, 4: 0,3 % [64].

Operationalisierung: Häufigkeitsgewichtete Maßzahl zum Eastern Cooperative Oncology Group Performance Status (ECOG PS gw.).

Anmerkung: Eine orange Füllfarbe im Diagramm kennzeichnet zu bewertende Studien, deren Anwendungsgebiet dem der Standardreferenz ganz oder teilweise entspricht.

N = 21 Nutzenbewertungsverfahren, 24 zu bewertende Studien.

Abb. F3: Abweichung zur Standardreferenz Aaronson et al., 1993; hier: Grad der Erkrankungsschwere (Quelle: Eigene Darstellung).



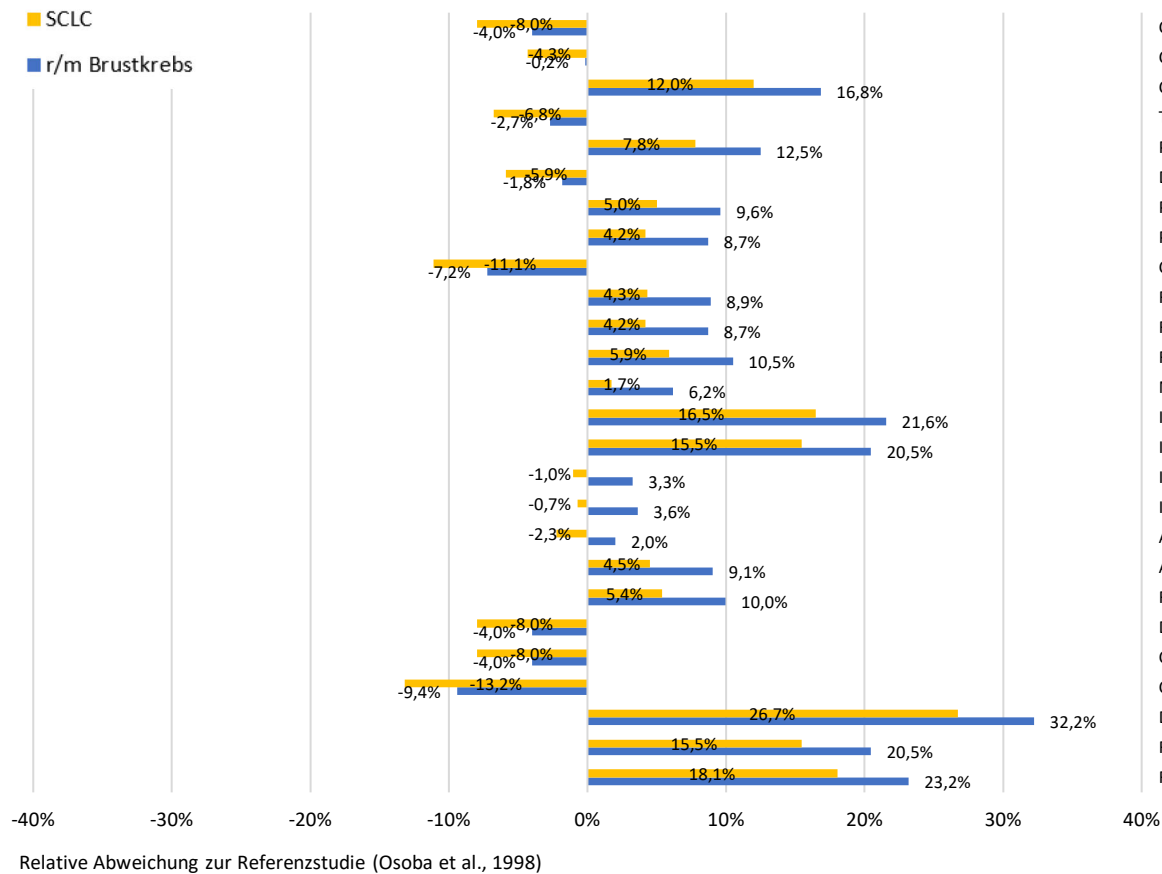
Aaronson et al., 1993: *No. of patients participating*: N = 313; *United Kingdom*: N = 40, *Federal Republic of Germany*: N = 32, *The Netherlands*: N = 28, *Denmark*: N = 4, *Norway*: N = 30, *Sweden*: N = 26, *Belgium*: N = 22, *France*: N = 17, *Italy*: N = 13 [64].

Operationalisierung: Anteil (West-)Europa.

Anmerkung: Eine orange Füllfarbe im Diagramm kennzeichnet zu bewertende Studien, deren Anwendungsgebiet dem der Standardreferenz ganz oder teilweise entspricht.

N = 15 Nutzenbewertungsverfahren, 15 zu bewertende Studien.

Abb. F4: Abweichung zur Standardreferenz Aaronson et al., 1993; hier: Geographische Region (Quelle: Eigene Darstellung).



Crizotinib (neues Anwendungsgebiet); nicht auswertbar
 Cobimetinib; p = 0,9197 (r/m Brustkrebs), p = 0,0674 (SCLC)
 Carfilizomib; nicht auswertbar
 Trametinib; p = 0,1387 (r/m Brustkrebs), p = 0,0036 (SCLC)
 Panobinostat; p < 0,0001
 Dabrafenib (neues Anwendungsgebiet); p = 0,2987 (r/m Brustkrebs), p = 0,0121 (SCLC)
 Pembrolizumab; p < 0,0001 (r/m Brustkrebs), p = 0,0333 (SCLC)
 Pembrolizumab; p < 0,0001 (r/m Brustkrebs), p = 0,0809 (SCLC)
 Ceritinib; p = 0,0004 (r/m Brustkrebs), p < 0,0001 (SCLC)
 Ruxolitinib (Polycythaemia vera); p < 0,0001 (r/m Brustkrebs), p = 0,0256 (SCLC)
 Ramucirumab; nicht auswertbar
 Ramucirumab; nicht auswertbar
 Nintedanib (nicht-kleinzelliges Lungenkarzinom); p < 0,0001 (r/m Brustkrebs), p = 0,3067 (SCLC)
 Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom); p < 0,0001
 Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom); p < 0,0001
 Ipilimumab (neues Anwendungsgebiet); p = 0,0648 (r/m Brustkrebs), p = 0,6536 (SCLC)
 Ipilimumab (neues Anwendungsgebiet); p = 0,0689 (r/m Brustkrebs), p = 0,7742 (SCLC)
 Afatinib; p = 0,2025 (r/m Brustkrebs), p = 0,2387 (SCLC)
 Afatinib; p < 0,0001 (r/m Brustkrebs), p = 0,0132 (SCLC)
 Regorafenib; p < 0,0001 (r/m Brustkrebs), p = 0,0018 (SCLC)
 Dabrafenib; p = 0,0432 (r/m Brustkrebs), p = 0,0012 (SCLC)
 Crizotinib; nicht auswertbar
 Crizotinib; nicht auswertbar
 Decitabin; nicht auswertbar
 Ruxolitinib (Chronisch myeloproliferative Erkrankungen); nicht auswertbar
 Ruxolitinib (Chronisch myeloproliferative Erkrankungen); nicht auswertbar

Verfahrensbezeichnung nebst
 p-Wert für den Vergleich von zu
 bewertender Studie und Referenz-
 studie (Osoba et al., 1998)

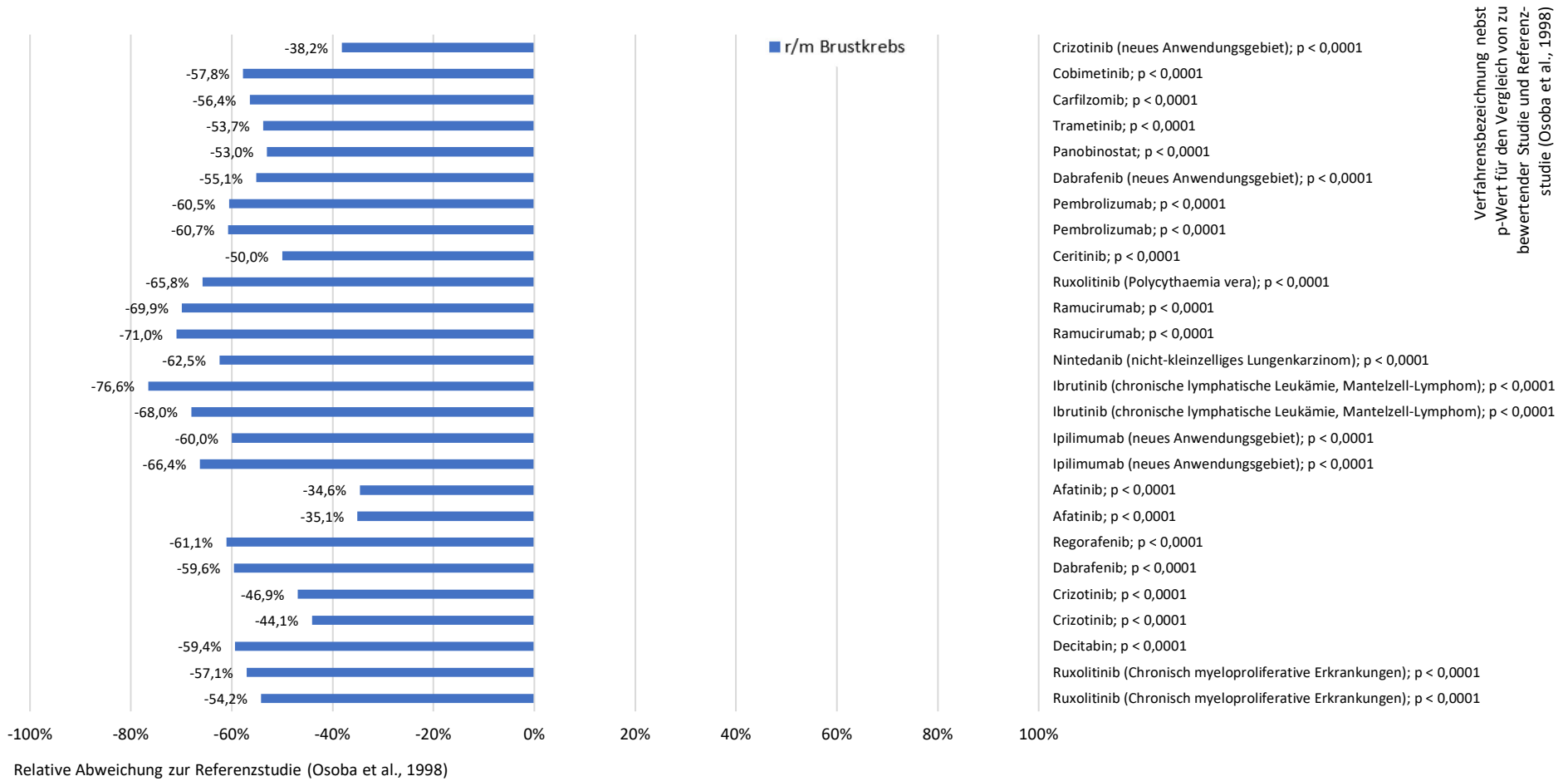
Osoba et al., 1998: Mean age: 55,2 years (SD: 9,9 years) (Breast Cancer); 57,6 years (SD: 7,7 years) (SCLC) [61].

Operationalisierung: Alter [Jahren].

Anmerkung: „nicht auswertbar“ bei fehlenden Angaben zu Mittelwert und Standardabweichung.

N = 21 Nutzenbewertungsverfahren, 26 zu bewertende Studien.

Abb. F5: Abweichung zur Standardreferenz Osoba et al., 1998; hier: Alter (Quelle: Eigene Darstellung).



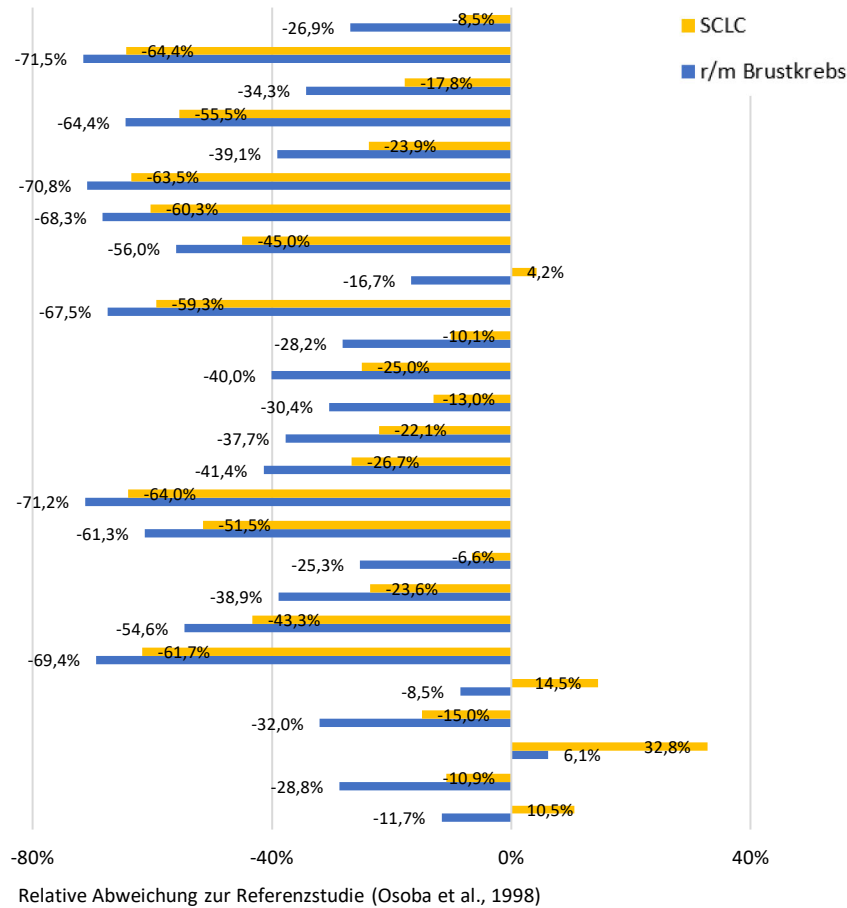
Verfahrensbezeichnung, nebst p-Wert für den Vergleich von zu bewertender Studie und Referenzstudie (Osoba et al., 1998)

Osoba et al., 1998: *Women (Breast Cancer)* [61].

Operationalisierung: Anteil Frauen.

N = 21 Nutzenbewertungsverfahren, 26 zu bewertende Studien.

Abb. F6: Abweichung zur Standardreferenz Osoba et al., 1998; hier: Geschlecht (Quelle: Eigene Darstellung).



- Crizotinib (neues Anwendungsgebiet); p < 0,0001 (r/m Brustkrebs), p = 0,2508 (SCLC)
- Cobimetinib; p < 0,0001 (r/m Brustkrebs), p < 0,0001 (SCLC)
- Carfilzomib; p < 0,0001 (r/m Brustkrebs), p = 0,0258 (SCLC)
- Trametinib; p < 0,0001 (r/m Brustkrebs), p < 0,0001 (SCLC)
- Panobinostat; p < 0,0001 (r/m Brustkrebs), p = 0,0017 (SCLC)
- Dabrafenib (neues Anwendungsgebiet); p < 0,0001 (r/m Brustkrebs), p < 0,0001 (SCLC)
- Pembrolizumab; p < 0,0001 (r/m Brustkrebs), p < 0,0001 (SCLC)
- Pembrolizumab; p < 0,0001 (r/m Brustkrebs), p < 0,0001 (SCLC)
- Ceritinib; p = 0,0191 (r/m Brustkrebs), p = 0,6688 (SCLC)
- Ruxolitinib (Polycythaemia vera); p < 0,0001 (r/m Brustkrebs), p < 0,0001 (SCLC)
- Ramucirumab; p < 0,0001 (r/m Brustkrebs), p = 0,1243 (SCLC)
- Ramucirumab; p < 0,0001 (r/m Brustkrebs), p < 0,0001 (SCLC)
- Nintedanib (nicht-kleinzelliges Lungenkarzinom); p < 0,0001 (r/m Brustkrebs), p = 0,0331 (SCLC)
- Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom); p < 0,0001 (r/m Brustkrebs), p = 0,0398 (SCLC)
- Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom); p < 0,0001 (r/m Brustkrebs), p < 0,0001 (SCLC)
- Ipilimumab (neues Anwendungsgebiet); p < 0,0001 (r/m Brustkrebs), p < 0,0001 (SCLC)
- Ipilimumab (neues Anwendungsgebiet); p < 0,0001 (r/m Brustkrebs), p < 0,0001 (SCLC)
- Afatinib; p < 0,0001 (r/m Brustkrebs), p = 0,2989 (SCLC)
- Afatinib; p < 0,0001 (r/m Brustkrebs), p = 0,0008 (SCLC)
- Regorafenib; p < 0,0001 (r/m Brustkrebs), p < 0,0001 (SCLC)
- Dabrafenib; p < 0,0001 (r/m Brustkrebs), p < 0,0001 (SCLC)
- Crizotinib; p = 0,1248 (r/m Brustkrebs), p = 0,1090 (SCLC)
- Crizotinib; p < 0,0001 (r/m Brustkrebs), p = 0,0721 (SCLC)
- Decitabin; p = 0,2389 (r/m Brustkrebs), p < 0,0001 (SCLC)
- Ruxolitinib (Chronisch myeloproliferative Erkrankungen); p < 0,0001 (r/m Brustkrebs), p = 0,2345 (SCLC)
- Ruxolitinib (Chronisch myeloproliferative Erkrankungen); p = 0,0483 (r/m Brustkrebs), p = 0,2551 (SCLC)

Verfahrensbezeichnung, nebst p-Wert für den Vergleich von zu bewertender Studie und Referenzstudie (Osoba et al., 1998)

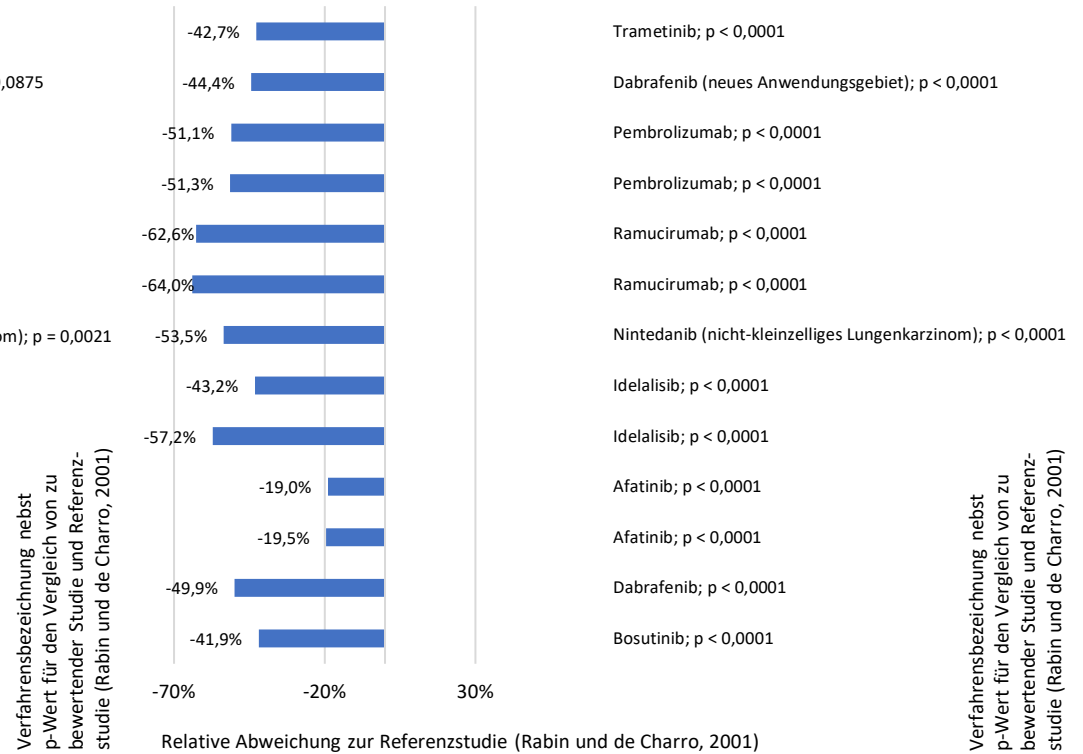
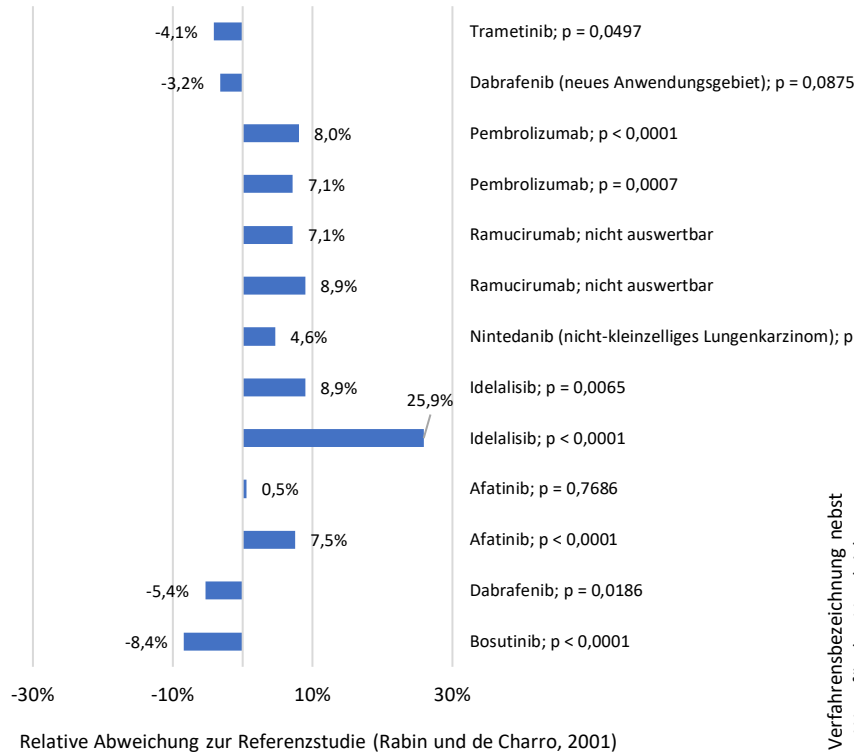
Osoba et al., 1998: Performance status 0: 23,6 %, 1: 51,6 %, 2: 24,8 % (Breast Cancer); 0: 29,2 %, 1: 60,7 %, 2: 10,1 % (SCLC) [61].

Operationalisierung: Häufigkeitsgewichtete Maßzahl zum Eastern Cooperative Oncology Group Performance Status (ECOG PS gw.).

N = 21 Nutzenbewertungsverfahren, 26 zu bewertende Studien.

Abb. F7: Abweichung zur Standardreferenz Osoba et al., 1998; hier: Grad der Erkrankungsschwere (Quelle: Eigene Darstellung).

F-II: Abweichungen zur Standardreferenz; hier: EQ-5D (VAS)



Rabin und de Charro, 2001: *Mean age: 56 years (SD: 14 years)* [66].

Operationalisierung: Alter [Jahren].

Anmerkung: Rabin und de Charro, 2001 [66] führt im Wesentlichen die Ergebnisse aus Hurst et al., 1997 [67] an. „nicht auswertbar“ bei fehlenden Angaben zu Mittelwert und Standardabweichung.

N = 10 Nutzenbewertungsverfahren, 13 zu bewertende Studien.

Abb. F8: Abweichung zur Standardreferenz Rabin und de Charro, 2001; hier: Alter (Quelle: Eigene Darstellung).

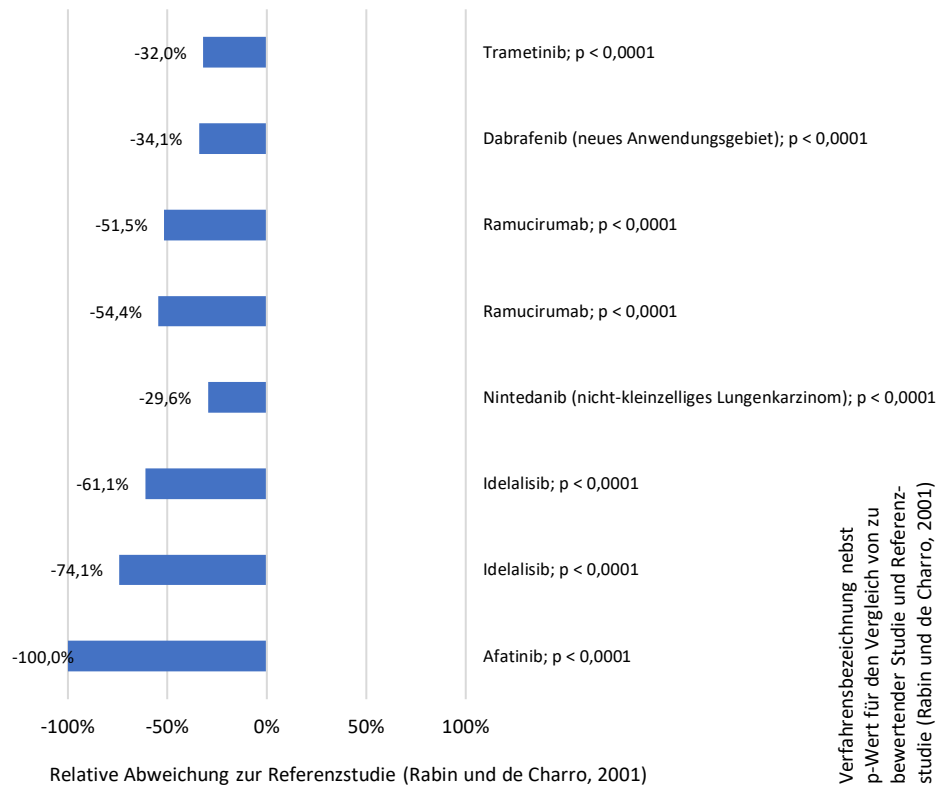
Rabin und de Charro, 2001: *Females: N = 188, Total: N = 233* [66].

Operationalisierung: Anteil Frauen.

Anmerkung: Rabin und de Charro, 2001 [66] führt im Wesentlichen die Ergebnisse aus Hurst et al., 1997 [67] an.

N = 10 Nutzenbewertungsverfahren, 13 zu bewertende Studien.

Abb. F9: Abweichung zur Standardreferenz Rabin und de Charro, 2001; hier: Geschlecht (Quelle: Eigene Darstellung).



Rabin und de Charro, 2001: „[...] rheumatology ward (Western General Hospital NHS Trust, Edinburgh) and by contacting GPs and nursing homes in the Lothian and Fife Regions.“ [66].

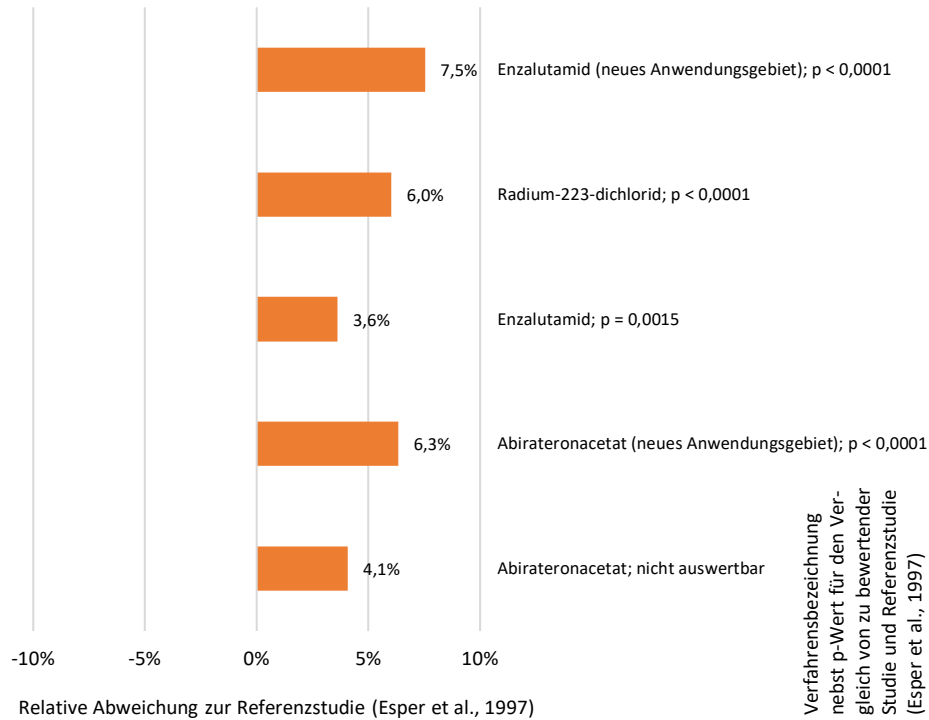
Operationalisierung: Anteil (West-)Europa.

Anmerkung: Rabin und de Charro, 2001 [66] führt im Wesentlichen die Ergebnisse aus Hurst et al., 1997 [67] an.

N = 7 Nutzenbewertungsverfahren, 8 zu bewertende Studien.

Abb. F10: Abweichung zur Standardreferenz Rabin und de Charro, 2001; hier: Geographische Region (Quelle: Eigene Darstellung).

F-III: Abweichungen zur Standardreferenz; hier: *FACT-P*



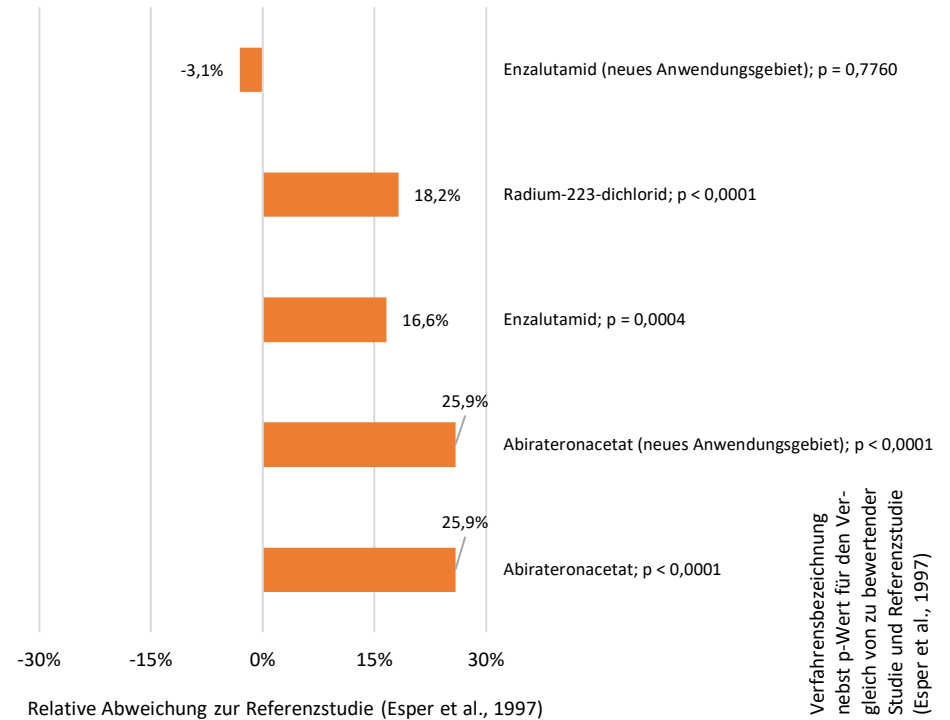
Esper et al., 1997: *Mean age: 65 years (Scale Development Sample), 67 years (SD: 8,6 years) (Validity Sample 1), 66 years (SD: 8,7 years) (Validity Sample 2)* [68].

Operationalisierung: Alter [Jahren].

Anmerkung: „nicht auswertbar“ bei fehlenden Angaben zu Mittelwert und Standardabweichung. Eine orange Füllfarbe im Diagramm kennzeichnet zu bewertende Studien mit im Vergleich zur Standardreferenz übereinstimmendem Anwendungsgebiet.

N = 5 Nutzenbewertungsverfahren, 5 zu bewertende Studien.

Abb. F11: Abweichung zur Standardreferenz Esper et al., 1997; hier: Alter (Quelle: Eigene Darstellung).



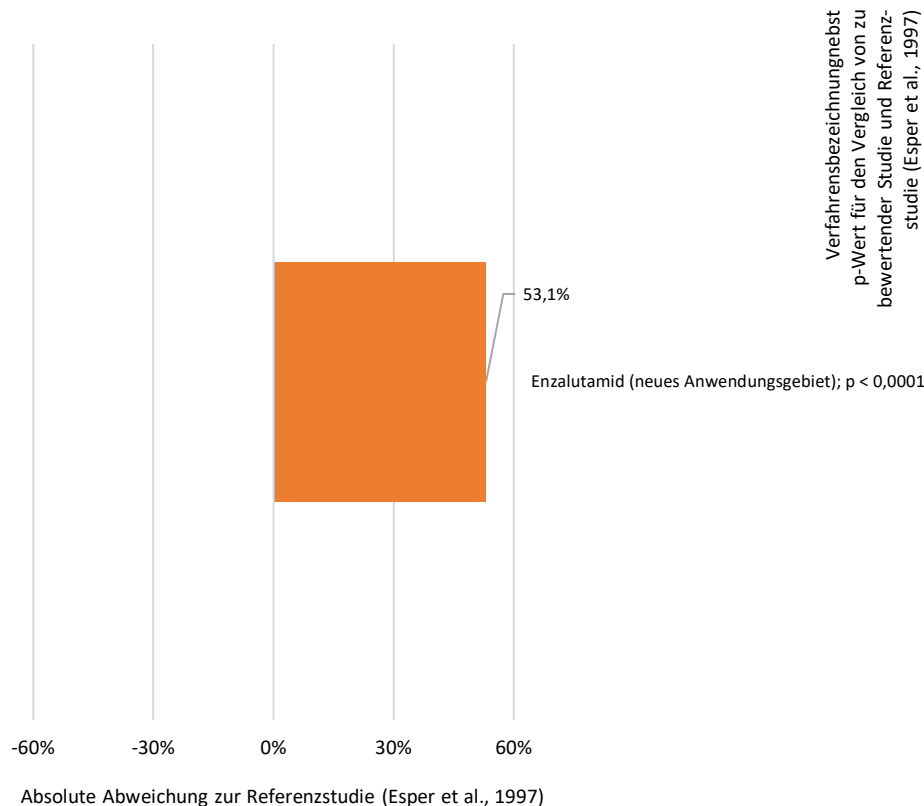
Esper et al., 1997: *Race, White: N = 35 (81 %) (Scale Development Sample); N = 23 (79,4 %) (Validity Sample 1)* [68].

Operationalisierung: Anteil Kaukasier/weiß.

Anmerkung: Eine orange Füllfarbe im Diagramm kennzeichnet zu bewertende Studien mit im Vergleich zur Standardreferenz übereinstimmendem Anwendungsgebiet.

N = 5 Nutzenbewertungsverfahren, 5 zu bewertende Studien.

Abb. F12: Abweichung zur Standardreferenz Esper et al., 1997; hier: Ethnizität (Quelle: Eigene Darstellung).



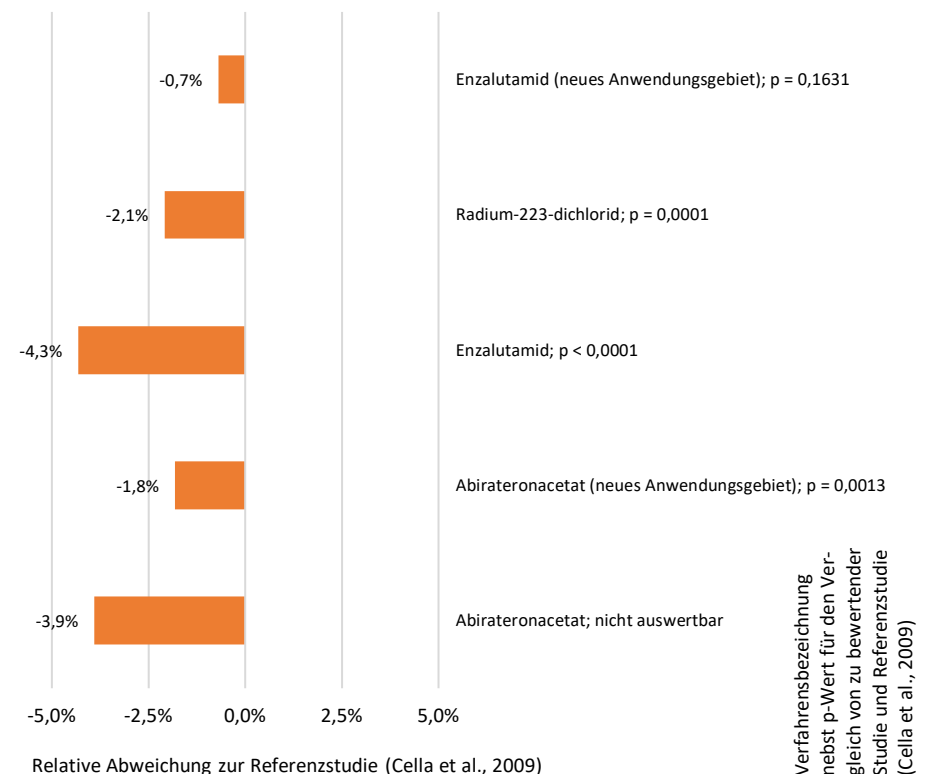
Esper et al., 1997: *Meyer L. Prentis Comprehensive Cancer Center (MLPCCC, Detroit, Mich) (Scale Development Sample), MLPCCC and the University of Michigan (Validity Sample 1), Prostate and Urology Center of the Louis A. Weiss Memorial Hospital, affiliated with the University of Chicago (Validity Sample 2)* [68].

Anmerkung: Eine orange Füllfarbe im Diagramm kennzeichnet zu bewertende Studien mit im Vergleich zur Standardreferenz übereinstimmendem Anwendungsgebiet.

Operationalisierung: Anteil (West-)Europa.

N = 1 Nutzenbewertungsverfahren, 1 zu bewertende Studie.

Abb. F13: Abweichung zur Standardreferenz Esper et al., 1997; hier: Geographische Region (Quelle: Eigene Darstellung).



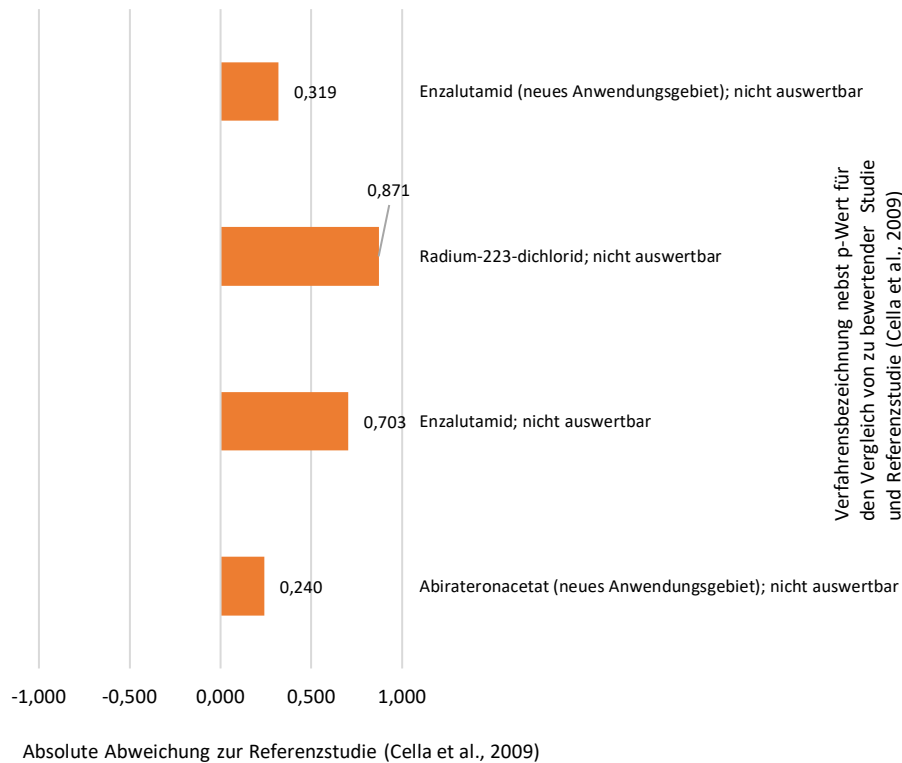
Cella et al., 2009: *Mean age: 71,8 years (SD: 0,29 years)* [69].

Operationalisierung: Alter [Jahren].

Anmerkung: „nicht auswertbar“ bei fehlenden Angaben zu Mittelwert und Standardabweichung. Eine orange Füllfarbe im Diagramm kennzeichnet zu bewertende Studien mit im Vergleich zur Standardreferenz übereinstimmendem Anwendungsgebiet.

N = 5 Nutzenbewertungsverfahren, 5 zu bewertende Studien.

Abb. F14: Abweichung zur Standardreferenz Cella et al., 2009; hier: Alter (Quelle: Eigene Darstellung).



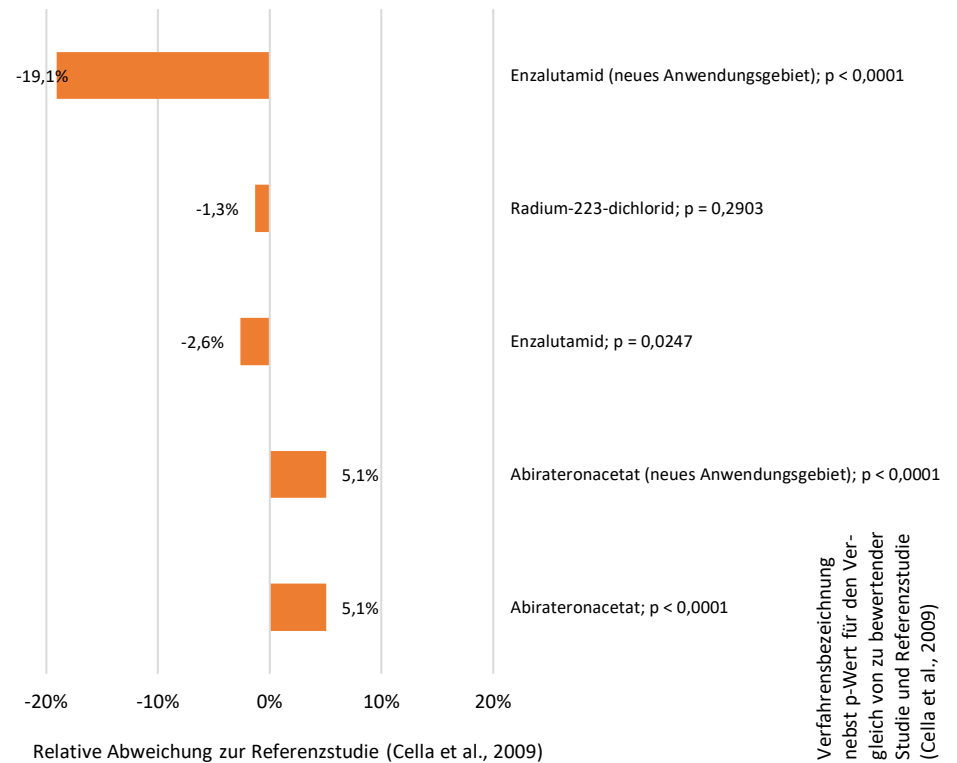
Cella et al., 2009: *Karnofsky score, Median: 100* [69].

Operationalisierung: Häufigkeitsgewichtete Maßzahl zum *Eastern Cooperative Oncology Group Performance Status (ECOG PS gw.)*.

Anmerkung: Der *ECOG PS* in Cella et al., 2009 [69] entspricht im Median 0 (Darstellung der jeweils absoluten Abweichung). Eine orange Füllfarbe im Diagramm kennzeichnet zu bewertende Studien mit im Vergleich zur Standardreferenz übereinstimmendem Anwendungsgebiet.

N = 4 Nutzenbewertungsverfahren, 4 zu bewertende Studien.

Abb. F15: Abweichung zur Standardreferenz Cella et al., 2009; hier: Grad der Erkrankungsschwere (Quelle: Eigene Darstellung).



Cella et al., 2009: *Race, Caucasian: N = 770 (95 %)* [69].

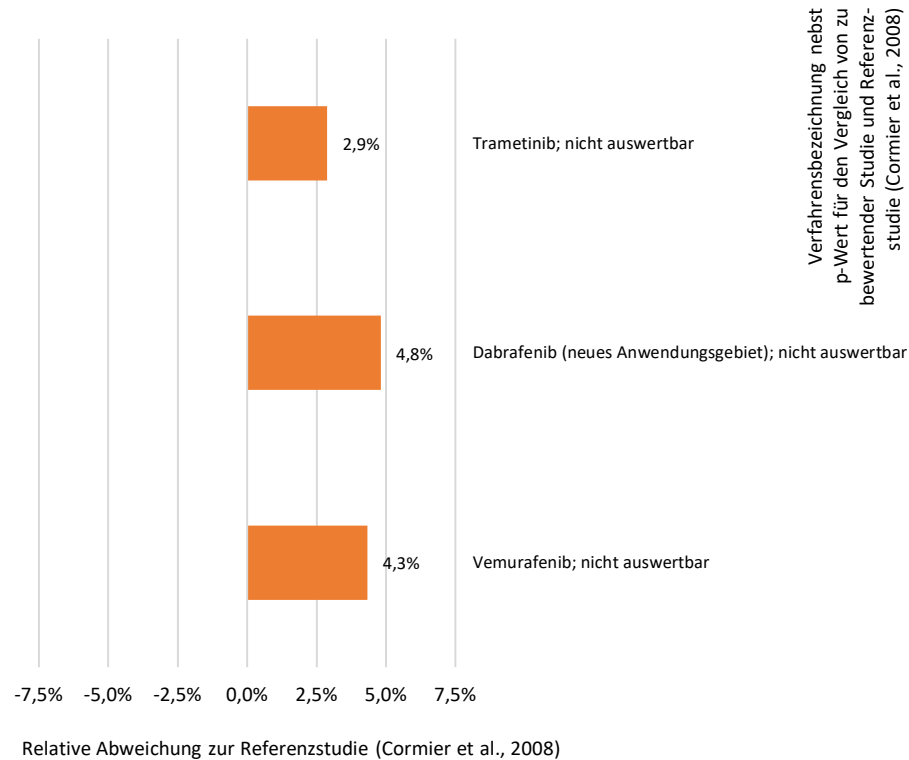
Operationalisierung: Anteil Kaukasier/weiß.

Anmerkung: Eine orange Füllfarbe im Diagramm kennzeichnet zu bewertende Studien mit im Vergleich zur Standardreferenz übereinstimmendem Anwendungsgebiet.

N = 5 Nutzenbewertungsverfahren, 5 zu bewertende Studien.

Abb. F16: Abweichung zur Standardreferenz Cella et al., 2009; hier: Ethnizität (Quelle: Eigene Darstellung).

F-IV: Abweichungen zur Standardreferenz; hier: *FACT-M*



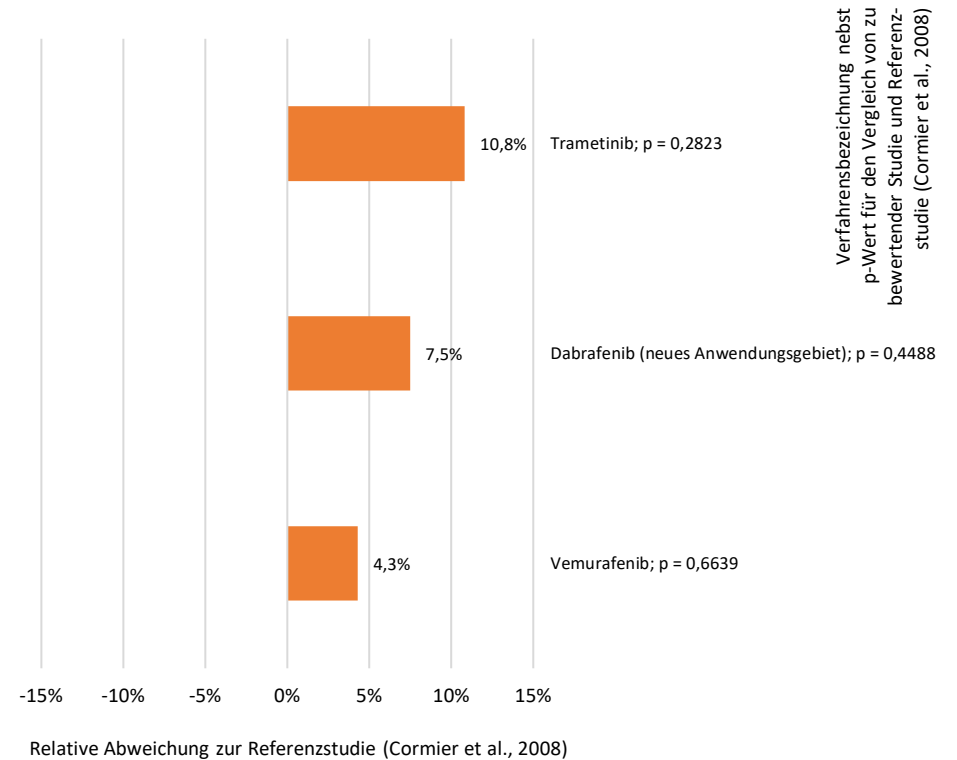
Cormier et al., 2008: *Median age: 52 years* [70].

Operationalisierung: Alter [Jahren].

Anmerkung: „nicht auswertbar“ bei fehlenden Angaben zu Mittelwert und Standardabweichung. Eine orange Füllfarbe im Diagramm kennzeichnet zu bewertende Studien mit im Vergleich zur Standardreferenz übereinstimmendem Anwendungsgebiet.

N = 3 Nutzenbewertungsverfahren, 3 zu bewertende Studien.

Abb. F17: Abweichung zur Standardreferenz Cormier et al., 2008; hier: Alter (Quelle: Eigene Darstellung).



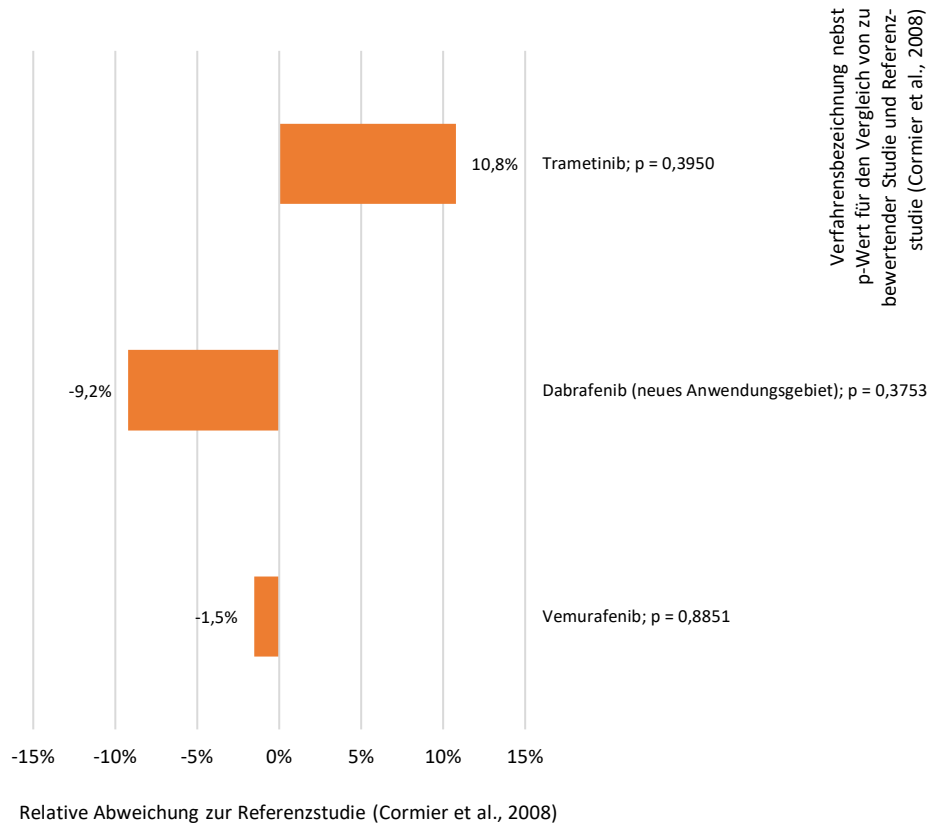
Cormier et al., 2008: *Women: N = 114 (41,8 %)* [70].

Operationalisierung: Anteil Frauen.

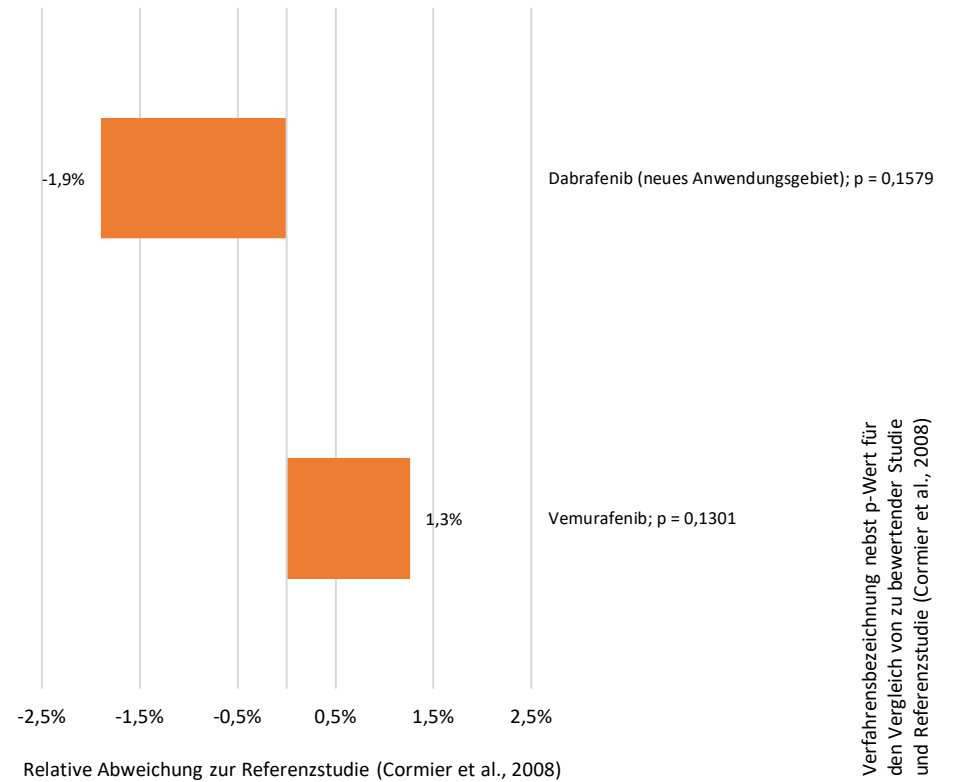
Anmerkung: Eine orange Füllfarbe im Diagramm kennzeichnet zu bewertende Studien mit im Vergleich zur Standardreferenz übereinstimmendem Anwendungsgebiet.

N = 3 Nutzenbewertungsverfahren, 3 zu bewertende Studien.

Abb. F18: Abweichung zur Standardreferenz Cormier et al., 2008; hier: Geschlecht (Quelle: Eigene Darstellung).



Verfahrensbezeichnung nebst p-Wert für den Vergleich von zu bewertender Studie und Referenzstudie (Cormier et al., 2008)



Verfahrensbezeichnung nebst p-Wert für den Vergleich von zu bewertender Studie und Referenzstudie (Cormier et al., 2008)

Cormier et al., 2008: *ECOG performance status criteria* 0: N = 191 (70 %), 1: 75 (27,5 %), 2: 7 (2,5 %) [70].

Cormier et al., 2008: *Race/ethnicity, White, non-Hispanic*: N = 268 (98,1 %) [70].

Operationalisierung: Häufigkeitsgewichtete Maßzahl zum *Eastern Cooperative Oncology Group Performance Status (ECOG PS gw.)*.

Operationalisierung: Anteil Kaukasier/weiß.

Anmerkung: Eine orange Füllfarbe im Diagramm kennzeichnet zu bewertende Studien mit im Vergleich zur Standardreferenz übereinstimmendem Anwendungsgebiet.

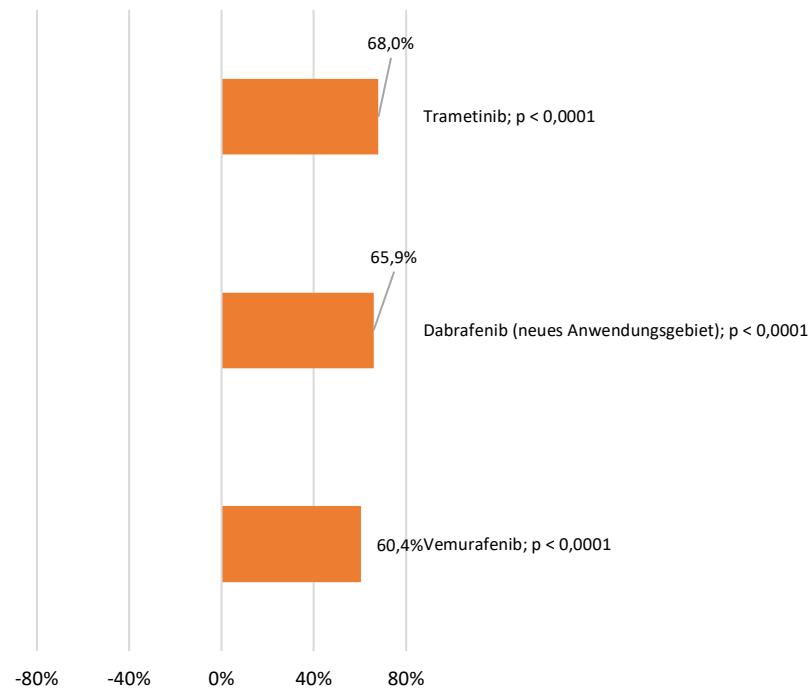
Anmerkung: Eine orange Füllfarbe im Diagramm kennzeichnet zu bewertende Studien mit im Vergleich zur Standardreferenz übereinstimmendem Anwendungsgebiet.

N = 3 Nutzenbewertungsverfahren, 3 zu bewertende Studien.

N = 2 Nutzenbewertungsverfahren, 2 zu bewertende Studien.

Abb. F19: Abweichung zur Standardreferenz Cormier et al., 2008; hier: Grad der Erkrankungsschwere (Quelle: Eigene Darstellung).

Abb. F20: Abweichung zur Standardreferenz Cormier et al., 2008; hier: Ethnizität (Quelle: Eigene Darstellung).



Absolute Abweichung zur Referenzstudie (Cormier et al., 2008)

Verfahrensbezeichnung nebst p-Wert für den Vergleich von zu bewertender Studie und Referenzstudie (Cormier et al., 2008)

Cormier et al., 2008: „*This study was approved by the University of Texas M. D. Anderson Cancer Center Institutional Review Board [...]. Patients were recruited for the study when they presented for their regularly scheduled appointments at the Melanoma and Skin Center.*“ [70].

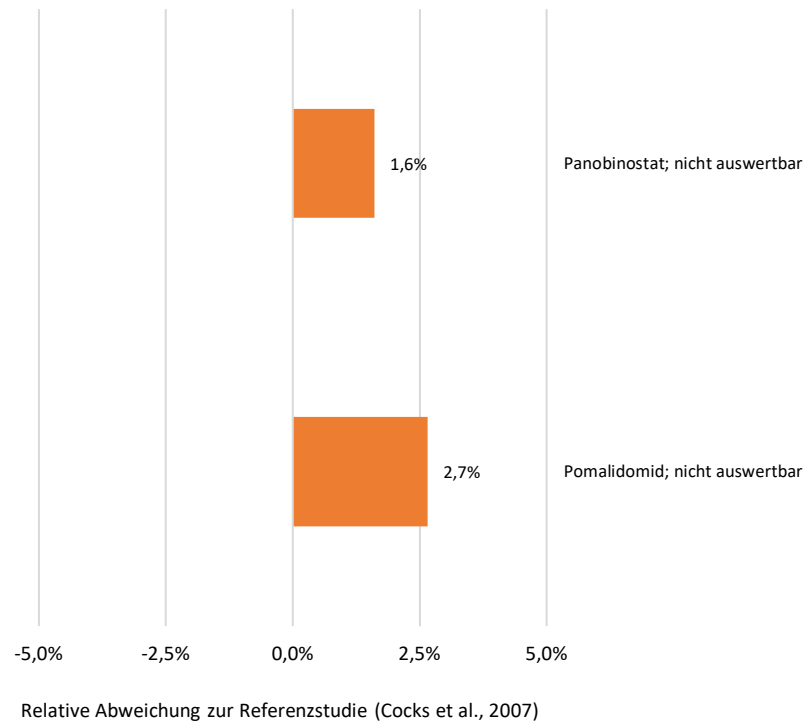
Operationalisierung: Anteil (West-)Europa.

Anmerkung: Eine orange Füllfarbe im Diagramm kennzeichnet zu bewertende Studien mit im Vergleich zur Standardreferenz übereinstimmendem Anwendungsgebiet.

N = 3 Nutzenbewertungsverfahren, 3 zu bewertende Studien.

Abb. F21: Abweichung zur Standardreferenz Cormier et al., 2008; hier: Geographische Region (Quelle: Eigene Darstellung).

F-V: Abweichungen zur Standardreferenz; hier: *EORTC QLQ-MY20*



Cocks et al., 2007: *Median age: 62,0 years* [71].

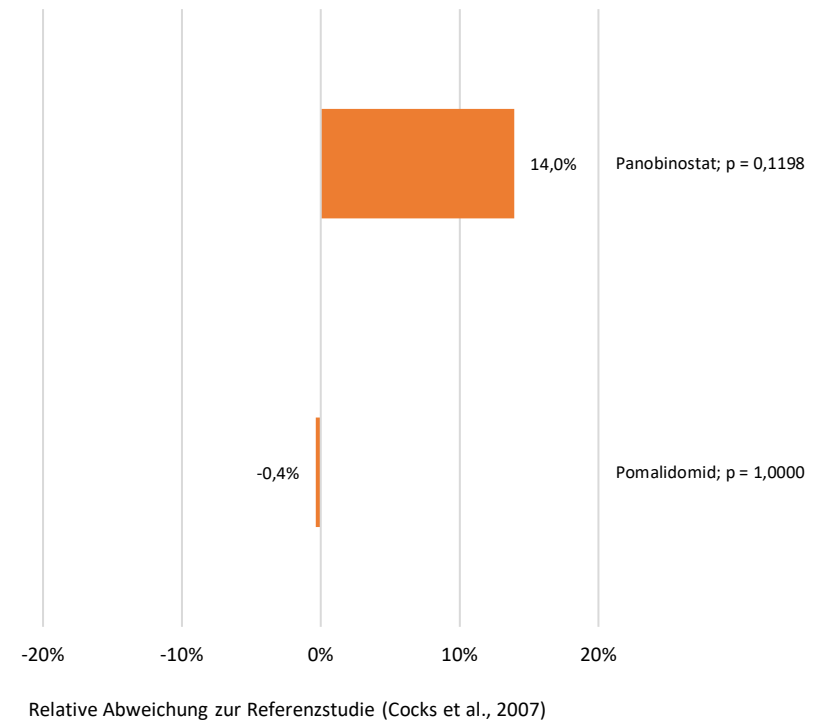
Operationalisierung: Alter [Jahren].

Anmerkung: „nicht auswertbar“ bei fehlenden Angaben zu Mittelwert und Standardabweichung. Eine orange Füllfarbe im Diagramm kennzeichnet zu bewertende Studien mit im Vergleich zur Standardreferenz übereinstimmendem Anwendungsgebiet.

N = 2 Nutzenbewertungsverfahren, 2 zu bewertende Studien.

Abb. F22: Abweichung zur Standardreferenz Cocks et al., 2007; hier: Alter (Quelle: Eigene Darstellung).

Verfahrensbezeichnung nebst p-Wert für den Vergleich von zu bewertender Studie und Referenzstudie (Cocks et al., 2007)



Cocks et al., 2007: *Females: N = 99 (41,3 %)* [71].

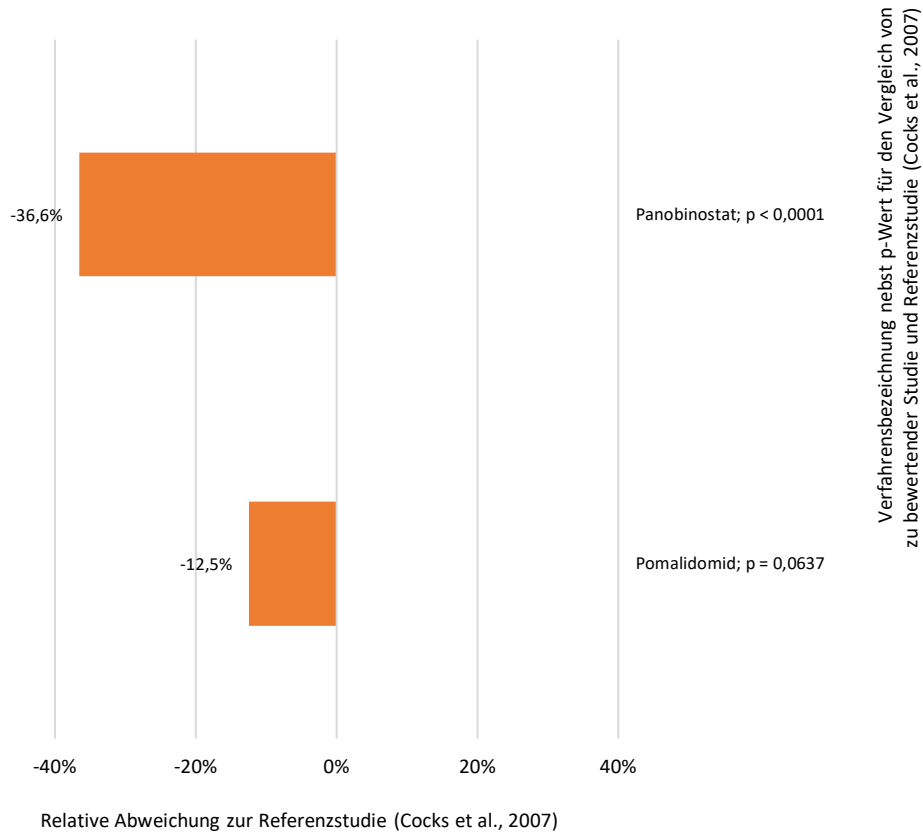
Operationalisierung: Anteil Frauen.

Anmerkung: Eine orange Füllfarbe im Diagramm kennzeichnet zu bewertende Studien mit im Vergleich zur Standardreferenz übereinstimmendem Anwendungsgebiet.

N = 2 Nutzenbewertungsverfahren, 2 zu bewertende Studien.

Abb. F23: Abweichung zur Standardreferenz Cocks et al., 2007; hier: Geschlecht (Quelle: Eigene Darstellung).

Verfahrensbezeichnung nebst p-Wert für den Vergleich von zu bewertender Studie und Referenzstudie (Cocks et al., 2007)



Cocks et al., 2007: WHO performance status 0: N = 87 (36,3 %), 1: N = 84 (35,0 %), 2: N = 29 (12,0 %), 3: N = 23 (9,6 %), 4: N = 2 (0,8 %), Missing: N = 15 (6,3 %) [71].

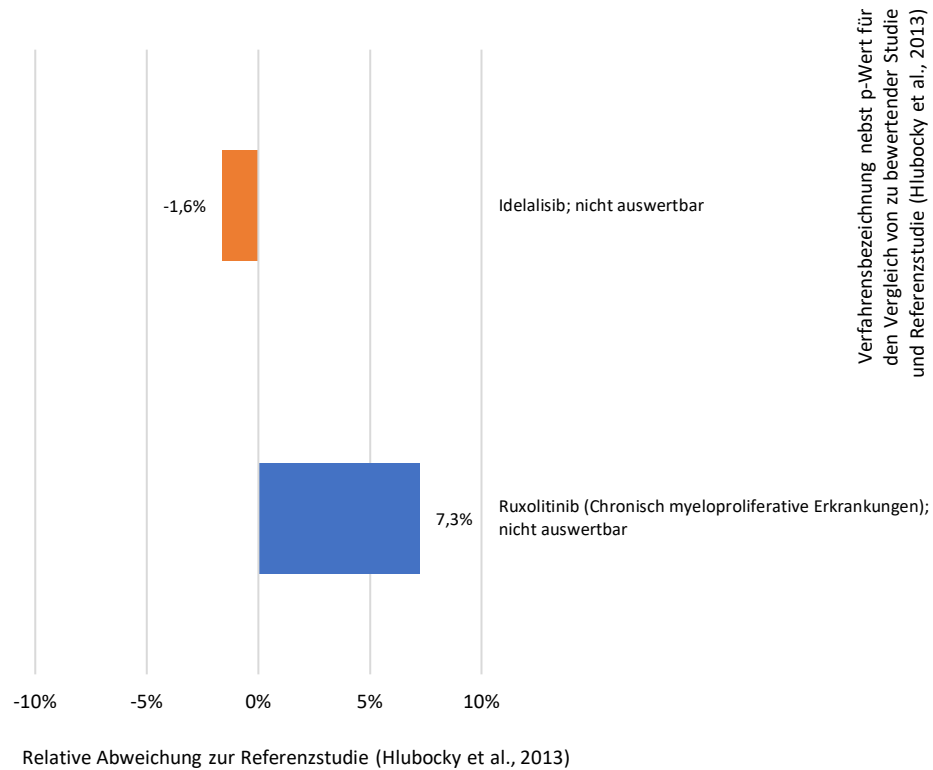
Operationalisierung: Häufigkeitsgewichtete Maßzahl zum Eastern Cooperative Oncology Group Performance Status (ECOG PS gw.).

Anmerkung: Eine orange Füllfarbe im Diagramm kennzeichnet zu bewertende Studien mit im Vergleich zur Standardreferenz übereinstimmendem Anwendungsgebiet.

N = 2 Nutzenbewertungsverfahren, 2 zu bewertende Studien.

Abb. F24: Abweichung zur Standardreferenz Cocks et al., 2007; hier: Grad der Erkrankungsschwere (Quelle: Eigene Darstellung).

F-VI: Abweichungen zur Standardreferenz; hier: *FACT-Lym*



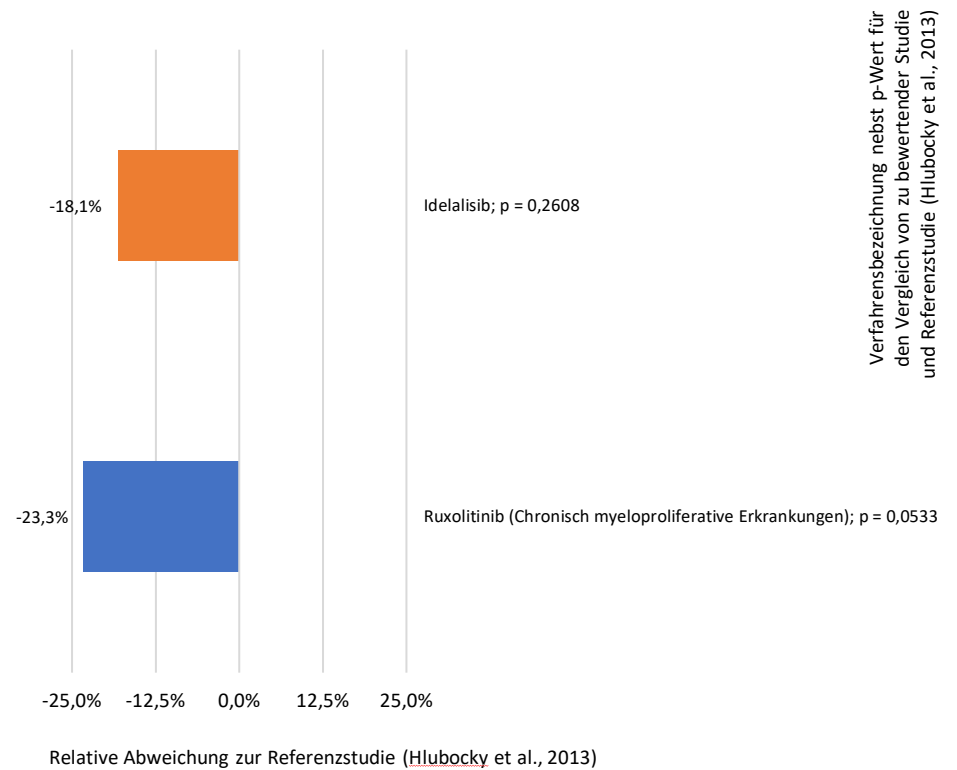
Hlubocky et al., 2013: *Median age: 62 years* [72].

Operationalisierung: Alter [Jahren].

Anmerkung: „nicht auswertbar“ bei fehlenden Angaben zu Mittelwert und Standardabweichung. Eine orange Füllfarbe im Diagramm kennzeichnet zu bewertende Studien mit im Vergleich zur Standardreferenz übereinstimmendem Anwendungsgebiet.

N = 2 Nutzenbewertungsverfahren, 2 zu bewertende Studien.

Abb. F25: Abweichung zur Standardreferenz Hlubocky et al., 2013; hier: Alter (Quelle: Eigene Darstellung).



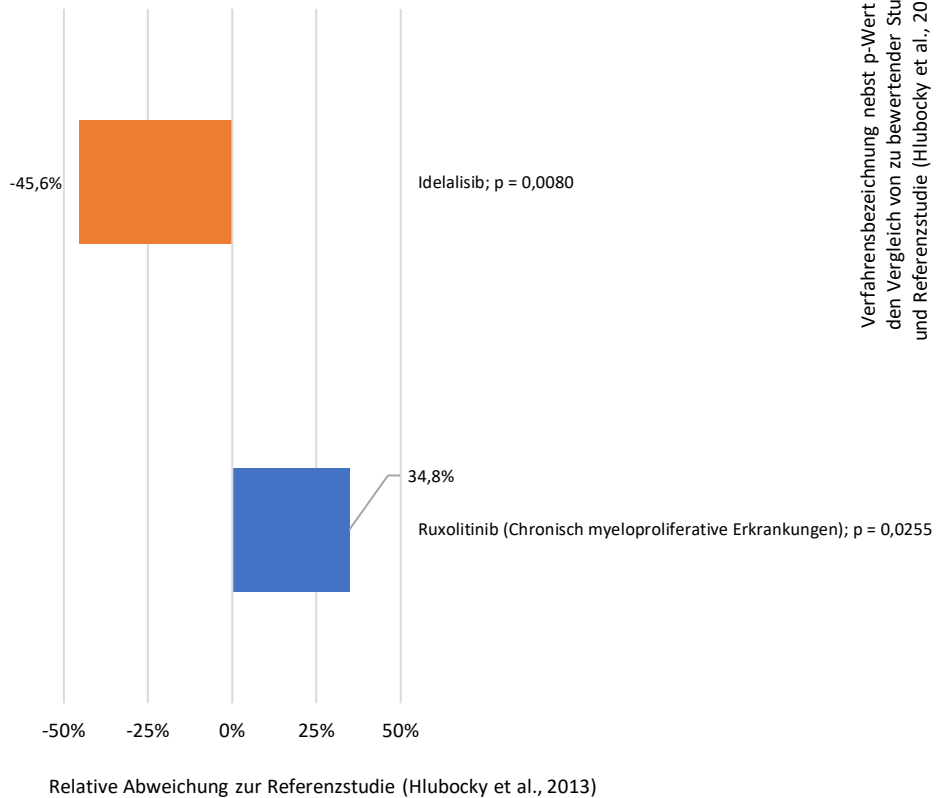
Hlubocky et al., 2013: *Females: N = 47 (56,0 %)* [72].

Operationalisierung: Anteil Frauen.

Anmerkung: Eine orange Füllfarbe im Diagramm kennzeichnet zu bewertende Studien mit im Vergleich zur Standardreferenz übereinstimmendem Anwendungsgebiet.

N = 2 Nutzenbewertungsverfahren, 2 zu bewertende Studien.

Abb. F26: Abweichung zur Standardreferenz Hlubocky et al., 2013; hier: Geschlecht (Quelle: Eigene Darstellung).



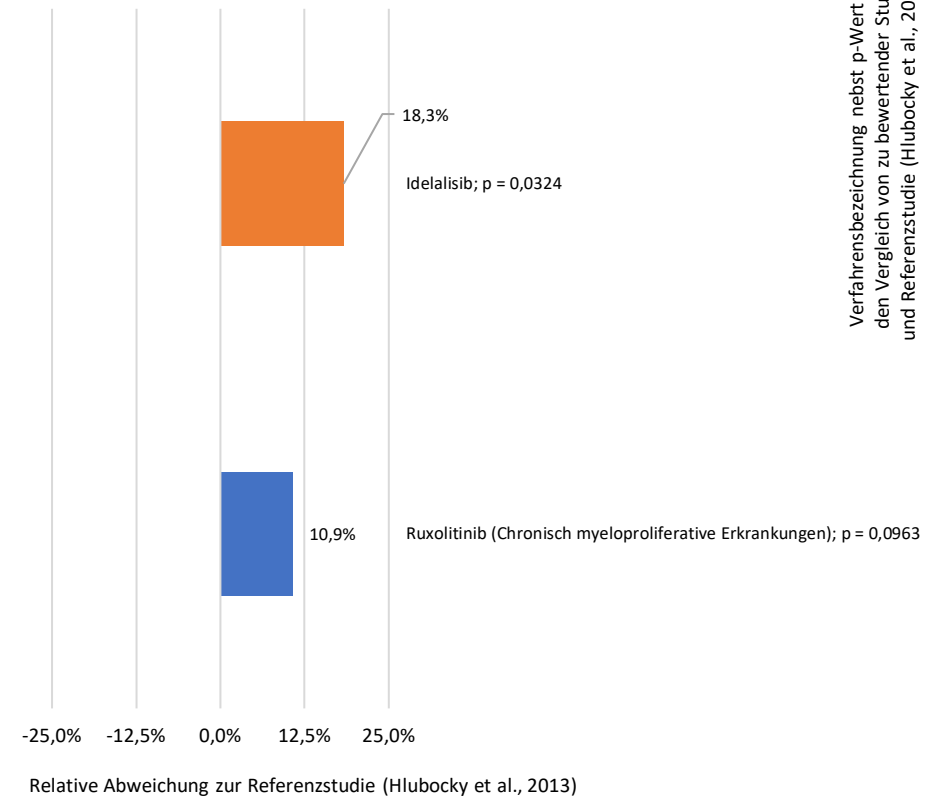
Hlubocky et al., 2013: *ECOG PSR 0*: $N = 45$ (53,6 %), 1 : $N = 33$ (39,3 %), 2 : $N = 6$ (7,1 %) [72].

Operationalisierung: Häufigkeitsgewichtete Maßzahl zum *Eastern Cooperative Oncology Group Performance Status (ECOG PS gw.)*.

Anmerkung: Eine orange Füllfarbe im Diagramm kennzeichnet zu bewertende Studien mit im Vergleich zur Standardreferenz übereinstimmendem Anwendungsgebiet.

$N = 2$ Nutzenbewertungsverfahren, 2 zu bewertende Studien.

Abb. F27: Abweichung zur Standardreferenz Hlubocky et al., 2013; hier: Grad der Erkrankungsschwere (Quelle: Eigene Darstellung).



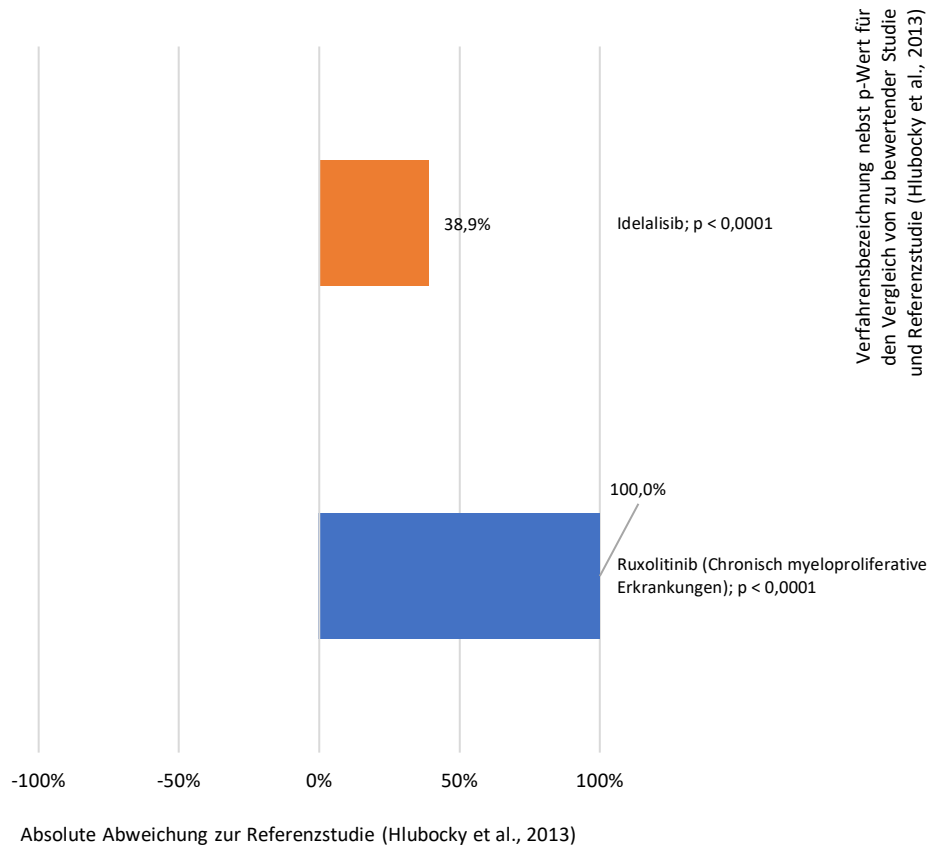
Hlubocky et al., 2013: *Race/ethnicity, White*: $N = 64$ (76,2 %) [72].

Operationalisierung: Anteil Kaukasier/weiß.

Anmerkung: Eine orange Füllfarbe im Diagramm kennzeichnet zu bewertende Studien mit im Vergleich zur Standardreferenz übereinstimmendem Anwendungsgebiet.

$N = 2$ Nutzenbewertungsverfahren, 2 zu bewertende Studien.

Abb. F28: Abweichung zur Standardreferenz Hlubocky et al., 2013; hier: Ethnizität (Quelle: Eigene Darstellung).



Hlubocky et al., 2013: *urban Midwestern academic medical centers* [72].

Operationalisierung: Anteil (West-)Europa.

Anmerkung: Eine orange Füllfarbe im Diagramm kennzeichnet zu bewertende Studien mit im Vergleich zur Standardreferenz übereinstimmendem Anwendungsgebiet.

N = 2 Nutzenbewertungsverfahren, 2 zu bewertende Studien.

Abb. F29: Abweichung zur Standardreferenz Hlubocky et al., 2013; hier: Geographische Region (Quelle: Eigene Darstellung).

Anhang G: Entsprechungsgrade von Patientencharakteristika bei Referenzen mit AWG-Übereinstimmung

G-I: Abweichungen zur Referenz mit AWG-Übereinstimmung; hier: *EORTC QLQ-C30*

		Aaronson et al., 1993 [64]				Aaronson et al., 1993 [64]	
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	non-small-cell lung cancer (NSCLC)	small-cell lung cancer (SCLC)	Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	non-small-cell lung cancer (NSCLC)	small-cell lung cancer (SCLC)
Relativer Anteil:		63,1 %	36,9 %	Relativer Anteil:		63,1 %	36,9 %
Ruxolitinib (Chronisch myeloproliferative Erkrankungen)	Chronisch Myeloproliferative Erkrankungen	-	-	Ruxolitinib (Polycythaemia vera)	Polycythaemia vera	-	-
Pomalidomid	Rezidiertes oder refraktäres Multiples Myelom (rrMM)	-	-	Afatinib (Neubewertung nach Fristablauf)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	X	-
Dabrafenib	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)	-	-	Ceritinib	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	X	-
Regorafenib	Kolorektalkarzinom (KRK)	-	-	Nivolumab (Melanom)	Fortgeschrittenes Melanom	-	-
Afatinib	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	X	-	Pembrolizumab	Fortgeschrittenes Melanom	-	-
Ruxolitinib (Chronisch myeloproliferative Erkrankungen, erneute Nutzenbewertung)	Chronisch Myeloproliferative Erkrankungen	-	-	Dabrafenib (neues Anwendungsgebiet)	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)	-	-
Eribulin (neues Anwendungsgebiet, Neubewertung nach Fristablauf)	Mammakarzinom	-	-	Panobinostat	Rezidiertes oder refraktäres Multiples Myelom (rrMM)	-	-
Obinutuzumab	Chronische lymphatische Leukämie (CLL)	-	-	Pomalidomid (erneute Nutzenbewertung)	Rezidiertes oder refraktäres Multiples Myelom (rrMM)	-	-
Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom)	Chronische lymphatische Leukämie (CLL); Rezidiertes oder refraktäres Mantelzell-Lymphom (MCL)	-	-	Trametinib	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)	-	-
Nintedanib (nicht-kleinzelliges Lungenkarzinom)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	X	-	Carfilzomib	Multiples Myelom (MM)	-	-
Ramucirumab	Magenkarzinom	-	-	Cobimetinib	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)	-	-

Tabelle G1: Entsprechung nach AWG beim *EORTC QLQ-C30*; hier: Aaronson et al., 1993 (Quelle: Eigene Darstellung).

Bjordal et al., 2000 [96]			
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	laryngeal cancer	cancer of the oral cavity or pharynx
Relativer Anteil:		47,1 %	52,9 %
Dabrafenib	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)	-	-
Obinutuzumab	Chronische lymphatische Leukämie (CLL)	-	-
Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom)	Chronische lymphatische Leukämie (CLL); Rezidiertes oder refraktäres Mantelzell-Lymphom (MCL)	-	-
Cobimetinib	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)	-	-

Tabelle G2: Entsprechung nach AWG beim EORTC QLQ-C30; hier: Bjordal et al., 2000 (Quelle: Eigene Darstellung).

Braun et al., 2011 [97]		
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	non-small-cell lung cancer (NSCLC)
Relativer Anteil:		100 %
Crizotinib	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	X
Afatinib	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	X
Nintedanib (nicht-kleinzelliges Lungenkarzinom)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	X
Afatinib (Neubewertung nach Fristablauf)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	X
Crizotinib (neues Anwendungsgebiet)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	X

Tabelle G3: Entsprechung nach AWG beim EORTC QLQ-C30; hier: Braun et al., 2011 (Quelle: Eigene Darstellung).

Delforge et al., 2012 [98]		
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	transplant-ineligible multiple myeloma (MM)
Relativer Anteil:		100 %
Carfilzomib	Multiples Myelom (MM)	X

Tabelle G4: Entsprechung nach AWG beim EORTC QLQ-C30; hier: Delforge et al., 2012 (Quelle: Eigene Darstellung).

Dubois et al., 2006 [99]		
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	multiple myeloma (MM)
Relativer Anteil:		100 %
Panobinostat	Rezidiertes oder refraktäres Multiples Myelom (rrMM)	X

Tabelle G5: Entsprechung nach AWG beim EORTC QLQ-C30; hier: Dubois et al., 2006 (Quelle: Eigene Darstellung).

Groenvold et al., 1997 [100]			
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	breast cancer	gynecological cancer
Relativer Anteil:		48,4 %	51,6 %
Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom)	Chronische lymphatische Leukämie (CLL); Rezidiertes oder refraktäres Mantelzell-Lymphom (MCL)	-	-
Nintedanib (nicht-kleinzelliges Lungenkarzinom)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	-	-
Ramucirumab	Magenkarzinom	-	-
Afatinib (Neubewertung nach Fristablauf)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	-	-

Tabelle G6: Entsprechung nach AWG beim EORTC QLQ-C30; hier: Groenvold et al., 1997 (Quelle: Eigene Darstellung).

n/a: nicht übertragbar

Hjermstad et al., 1995 [101]

Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	breast cancer	testicular cancer	malignant lymphomas	cervical cancer	other cancer
Relativer Anteil:		30 %	11 %	9 %	9 %	41 %
Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom)	Chronische lymphatische Leukämie (CLL); Rezidiertes oder refraktäres Mantelzell-Lymphom (MCL)	-	-	X	-	n/a
Nintedanib (nicht-kleinzelliges Lungenkarzinom)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	-	-	-	-	n/a
Ramucirumab	Magenkarzinom	-	-	-	-	n/a
Afatinib (Neubewertung nach Fristablauf)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	-	-	-	-	n/a
Pembrolizumab	Fortgeschrittenes Melanom	-	-	-	-	n/a

Tabelle G7: Entsprechung nach AWG beim EORTC QLQ-C30; hier: Hjermstad et al., 1995 (Quelle: Eigene Darstellung).

Kvam et al., 2011 [102]

Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	multiple myeloma (MM)
Relativer Anteil:		100 %
Pomalidomid	Rezidiertes oder refraktäres Multiples Myelom (rrMM)	X
Panobinostat	Rezidiertes oder refraktäres Multiples Myelom (rrMM)	X
Pomalidomid (erneute Nutzenbewertung)	Rezidiertes oder refraktäres Multiples Myelom (rrMM)	X
Carfilzomib	Multiples Myelom (MM)	X

Tabelle G8: Entsprechung nach AWG beim EORTC QLQ-C30; hier: Kvam et al., 2011 (Quelle: Eigene Darstellung).

McLachlan et al., 1998 [103]

Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	metastatic breast cancer
Relativer Anteil:		100 %
Eribulin (neues Anwendungsgebiet, Neubewertung nach Fristablauf)	Mammakarzinom	X

Tabelle G9: Entsprechung nach AWG beim EORTC QLQ-C30; hier: McLachlan et al., 1998 (Quelle: Eigene Darstellung).

Montazeri et al., 2001 [104]

Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	non-small-cell lung cancer (NSCLC)	small-cell lung cancer (SCLC)	unspecified lung cancer
Relativer Anteil:		51,9 %	20,9 %	27,1 %
Eribulin (neues Anwendungsgebiet, Neubewertung nach Fristablauf)	Mammakarzinom	-	-	-

Tabelle G10: Entsprechung nach AWG beim EORTC QLQ-C30; hier: Montazeri et al., 2001 (Quelle: Eigene Darstellung).

n/a: nicht übertragbar

Osoba et al., 1997 [105]

Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	breast cancer	lung cancer	colorectal cancer	gynaecologic cancer	other cancer
Relativer Anteil:		39,7 %	29,4 %	9,9 %	6,4 %	14,6 %
Ipilimumab	Fortgeschrittenes Melanom	-	-	-	-	n/a
Regorafenib	Kolorektalkarzinom (KRK)	-	-	X	-	-
Pembrolizumab	Fortgeschrittenes Melanom	-	-	-	-	n/a
Regorafenib (Neubewertung nach Fristablauf)	Kolorektalkarzinom (KRK)	-	-	X	-	-

Tabelle G11: Entsprechung nach AWG beim EORTC QLQ-C30; hier: Osoba et al., 1997 (Quelle: Eigene Darstellung).

Osoba et al., 1998 [61]				Osoba et al., 1998 [61]			
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	recurrent or metastatic breast cancer	extensive-stage small-cell lung cancer (SCLC)	Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	recurrent or metastatic breast cancer	extensive-stage small-cell lung cancer (SCLC)
Relativer Anteil:		68,9 %	31,1 %	Relativer Anteil:		68,9 %	31,1 %
Ruxolitinib (Chronisch myeloproliferative Erkrankungen)	Chronisch Myeloproliferative Erkrankungen	-	-	Dabrafenib (neues Anwendungsgebiet)	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)	-	-
Decitabin	Akute myeloische Leukämie (AML)	-	-	Panobinostat	Rezidiertes oder refraktäres Multiples Myelom (rrMM)	-	-
Crizotinib	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	-	-	Trametinib	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)	-	-
Dabrafenib	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)	-	-	Carfilzomib	Multiples Myelom (MM)	-	-
Regorafenib	Kolorektalkarzinom (KRK)	-	-	Cobimetinib	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)	-	-
Afatinib	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	-	-	Crizotinib (neues Anwendungsgebiet)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	-	-
Ipilimumab (neues Anwendungsgebiet)	Fortgeschrittenes Melanom	-	-				
Ruxolitinib (Chronisch myeloproliferative Erkrankungen, erneute Nutzenbewertung)	Chronisch Myeloproliferative Erkrankungen	-	-				
Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom)	Chronische lymphatische Leukämie (CLL); Rezidiertes oder refraktäres Mantelzell-Lymphom (MCL)	-	-				
Nintedanib (nicht-kleinzelliges Lungenkarzinom)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	-	-				
Ramucirumab	Magenkarzinom	-	-				
Ruxolitinib (Polycythaemia vera)	Polycythaemia vera	-	-				
Afatinib (Neubewertung nach Fristablauf)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	-	-				
Ceritinib	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	-	-				
Pembrolizumab	Fortgeschrittenes Melanom	-	-				

Tabelle G12: Entsprechung nach AWG beim EORTC QLQ-C30; hier: Osoba et al., 1998 (Quelle: Eigene Darstellung).

Parmar et al., 2005 [106]		
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	operable breast cancer
Relativer Anteil:		100 %
Eribulin (neues Anwendungsgebiet, Neubewertung nach Fristablauf)	Mammakarzinom	X

Tabelle G13: Entsprechung nach AWG beim EORTC QLQ-C30; hier: Parmar et al., 2005 (Quelle: Eigene Darstellung).

Sadighi et al., 2006 [107]		
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	gastric cancer
Relativer Anteil:		100 %
Ramucirumab	Magenkarzinom	X

Tabelle G14: Entsprechung nach AWG beim EORTC QLQ-C30; hier: Sadighi et al., 2006 (Quelle: Eigene Darstellung).

Teckle et al., 2011 [108]				
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	breast cancer	colorectal cancer	lung cancer
Relativer Anteil:		35,9 %	31,0 %	33,2 %
Crizotinib	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	-	-	(X)
Afatinib	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	-	-	(X)
Nintedanib (nicht-kleinzelliges Lungenkarzinom)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	-	-	(X)
Afatinib (Neubewertung nach Fristablauf)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	-	-	(X)
Cobimetinib	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)	-	-	-
Crizotinib (neues Anwendungsgebiet)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	-	-	(X)

Tabelle G15: Entsprechung nach AWG beim EORTC QLQ-C30; hier: Teckle et al., 2011 (Quelle: Eigene Darstellung).

Wisloff et al., 1996 [109]		
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	multiple myeloma (MM)
Relativer Anteil:		100 %
Pomalidomid	Rezidiertes oder refraktäres Multiples Myelom (rrMM)	X
Pomalidomid (erneute Nutzenbewertung)	Rezidiertes oder refraktäres Multiples Myelom (rrMM)	X
Carfilzomib	Multiples Myelom (MM)	X

Tabelle G16: Entsprechung nach AWG beim EORTC QLQ-C30; hier: Wisloff et al., 1996 (Quelle: Eigene Darstellung).

Wyrwich et al., 1999 [110]				
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	coronary artery disease (CAD) only	congestive heart failure (CHF) only	CAD + CHF
Relativer Anteil:		45 %	18 %	37 %
Pomalidomid (erneute Nutzenbewertung)	Rezidiertes oder refraktäres Multiples Myelom (rrMM)	-	-	-

Tabelle G17: Entsprechung nach AWG beim EORTC QLQ-C30; hier: Wyrwich et al., 1999 (Quelle: Eigene Darstellung).

n/a: nicht übertragbar

		Zeng et al., 2012 [111]													
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	breast cancer	prostate cancer	lung cancer	colorectal cancer	multiple myeloma	renal cell / kidney cancer	liver cancer	ovarian cancer	stomach cancer	oesophagus cancer	pancreatic cancer	other cancer	unknown	
Relativer Anteil:		30,1 %	22,6 %	16,1 %	5,4 %	4,3 %	3,2 %	2,2 %	2,2 %	2,2 %	1,1 %	1,1 %	8,6 %	1,1 %	
Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom)	Chronische lymphatische Leukämie (CLL); Rezidiertes oder refraktäres Mantelzell-Lymphom (MCL)	-	-	-	-	-	-	-	-	-	-	-	n/a	n/a	

Tabelle G18: Entsprechung nach AWG beim EORTC QLQ-C30; hier: Zeng et al., 2012 (Quelle: Eigene Darstellung).

Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie		Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie	
			absolut	relativ				absolut	relativ
Crizotinib [PROFILE 1007]; nicht auswertbar	NSCLC	Braun et al., 2011	- 8,3	- 14,2 %	Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) [PCYC-1112-CA]; nicht auswertbar	Leukämie, Lymphom	Hjermstad et al., 1995	12,5	23,1 %
Crizotinib [PROFILE 1007]; nicht auswertbar	NSCLC	Teckle et al., 2011	- 13,0	- 20,6 %	Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) [PCYC-1104-CA]; nicht auswertbar	Leukämie, Lymphom	Hjermstad et al., 1995	13,1	24,3 %
Crizotinib [PROFILE 1005]; nicht auswertbar	NSCLC	Braun et al., 2011	- 5,3	- 9,1 %	Nintedanib (nicht-kleinzelliges Lungenkarzinom); nicht auswertbar	NSCLC	Braun et al., 2011	0,3	0,5 %
Crizotinib [PROFILE 1005]; nicht auswertbar	NSCLC	Teckle et al., 2011	- 10,0	- 15,9 %	Nintedanib (nicht-kleinzelliges Lungenkarzinom); p = 0,0008	NSCLC	Teckle et al., 2011	- 4,4	- 7,0 %
Pomalidomid; nicht auswertbar	(rr)MM	Kvam et al., 2011	- 2,4	- 3,6 %	Panobinostat; nicht auswertbar	(rr)MM	Dubois et al., 2006	2,1	3,5 %
Pomalidomid; nicht auswertbar	(rr)MM	Wisloff et al., 1996	- 2,4	- 3,6 %	Panobinostat; nicht auswertbar	(rr)MM	Kvam et al., 2011	- 3,9	- 5,9 %
Regorafenib; nicht auswertbar	KRK	Osoba et al., 1997	3,7	6,5 %	Regorafenib (Neubewertung nach Fristablauf); nicht auswertbar	KRK	Osoba et al., 1997	- 0,5	- 0,9 %
Afatinib [LUX-Lung 3]; nicht auswertbar	NSCLC	Braun et al., 2011	1,9	3,3 %	Carfilzomib; nicht auswertbar	(rr)MM	Delforge et al., 2012	- 6,5	- 9,2 %
Afatinib [LUX-Lung 3]; p = 0,0457	NSCLC	Teckle et al., 2011	- 2,8	- 4,4 %	Carfilzomib; nicht auswertbar	(rr)MM	Kvam et al., 2011	- 1,5	- 2,3 %
Afatinib [LUX-Lung 6]; nicht auswertbar	NSCLC	Braun et al., 2011	- 2,0	- 3,4 %	Carfilzomib; nicht auswertbar	(rr)MM	Wisloff et al., 1996	- 1,5	- 2,3 %
Afatinib [LUX-Lung 6]; p < 0,0001	NSCLC	Teckle et al., 2011	- 6,7	- 10,6 %	Crizotinib (neues Anwendungsgebiet); nicht auswertbar	NSCLC	Braun et al., 2011	- 5,3	- 9,1 %
Eribulin (neues Anwendungsgebiet, Neubewertung nach Fristablauf); nicht auswertbar	Mammakarzinom	McLachlan et al., 1998	5,5	11,2 %	Crizotinib (neues Anwendungsgebiet); nicht auswertbar	NSCLC	Teckle et al., 2011	- 10,0	- 15,9 %
Eribulin (neues Anwendungsgebiet, Neubewertung nach Fristablauf); nicht auswertbar	Mammakarzinom	Parmar et al., 2005	10,3	23,3 %					

Anmerkung: Exklusive Aaronson et al., 1993 [64] (vgl. detaillierte Betrachtung in Kapitel 5.5, Komplex „Standardreferenzen“) und Sadighi et al., 2006 [107] (keine originäre Validierungsstudie). Präzisierungen: Osoba et al., 1997 [105] (Subgruppe: Dutch sample), Teckle et al., 2011 [108] (Subgruppe: Lung cancer).

KRK: Kolorektalkarzinom, NSCLC: Nicht-kleinzelliges Bronchialkarzinom, (rr)MM: (Rezidiertes oder refraktäres) Multiples Myelom.

Tabelle G19: Abweichung zur Referenz mit AWG-Übereinstimmung beim EORTC QLQ-C30; hier: Alter (Quelle: Eigene Darstellung).

Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie		Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie	
			absolut	relativ				absolut	relativ
Crizotinib [PROFILE 1007]; p = 0,0441	NSCLC	Braun et al., 2011	6,2 %	12,6 %	Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) [PCYC-1112-CA]; p = 0,1622	Leukämie, Lymphom	Hjermstad et al., 1995	- 5,9 %	- 15,6 %
Crizotinib [PROFILE 1007]; p = 0,2649	NSCLC	Teckle et al., 2011	8,4 %	17,6 %	Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) [PCYC-1104-CA]; p = 0,0108	Leukämie, Lymphom	Hjermstad et al., 1995	- 14,5 %	- 38,2 %
Crizotinib [PROFILE 1005]; p = 0,2411	NSCLC	Braun et al., 2011	3,4 %	6,9 %	Nintedanib (nicht-kleinzelliges Lungenkarzinom); p < 0,0001	NSCLC	Braun et al., 2011	- 12,1 %	- 24,4 %
Crizotinib [PROFILE 1005]; p = 0,4942	NSCLC	Teckle et al., 2011	5,5 %	11,6 %	Nintedanib (nicht-kleinzelliges Lungenkarzinom); p = 0,1317	NSCLC	Teckle et al., 2011	- 10,0 %	- 21,0 %
Pomalidomid; p = 0,1967	(rr)MM	Kvam et al., 2011	- 5,3 %	- 11,5 %	Panobinostat; p = 0,0813	(rr)MM	Dubois et al., 2006	6,9 %	17,2 %
Pomalidomid; p = 0,3647	(rr)MM	Wisloff et al., 1996	- 2,9 %	- 6,6 %	Panobinostat; p = 0,8823	(rr)MM	Kvam et al., 2011	0,6 %	1,2 %
Regorafenib; p < 0,0001	KRK	Osoba et al., 1997	- 18,8 %	- 32,5 %	Regorafenib (Neubewertung nach Fristablauf); p = 0,0002	KRK	Osoba et al., 1997	- 15,6 %	- 27,0 %
Afatinib [LUX-Lung 3]; p < 0,0001	NSCLC	Braun et al., 2011	15,3 %	30,7 %	Carfilzomib; p = 0,0080	(rr)MM	Delforge et al., 2012	- 7,1 %	- 14,1 %
Afatinib [LUX-Lung 3]; p = 0,0143	NSCLC	Teckle et al., 2011	17,4 %	36,6 %	Carfilzomib; p = 0,4577	(rr)MM	Kvam et al., 2011	- 2,9 %	- 6,2 %
Afatinib [LUX-Lung 6]; p < 0,0001	NSCLC	Braun et al., 2011	15,7 %	31,7 %	Carfilzomib; p = 0,8649	(rr)MM	Wisloff et al., 1996	- 0,4 %	- 1,0 %
Afatinib [LUX-Lung 6]; p = 0,0098	NSCLC	Teckle et al., 2011	17,8 %	37,5 %	Crizotinib (neues Anwendungsgebiet); p < 0,0001	NSCLC	Braun et al., 2011	12,1 %	24,4 %
Eribulin (neues Anwendungsgebiet, Neubewertung nach Fristablauf); p = 1,0000	Mammakarzinom	McLachlan et al., 1998	0,0 %	n/a	Crizotinib (neues Anwendungsgebiet); p = 0,0468	NSCLC	Teckle et al., 2011	14,3 %	30,0 %
Eribulin (neues Anwendungsgebiet, Neubewertung nach Fristablauf); p = 1,0000	Mammakarzinom	Parmar et al., 2005	0,0 %	n/a					

Anmerkung: Exklusive Aaronson et al., 1993 [64] (vgl. detaillierte Betrachtung in Kapitel 5.5, Komplex „Standardreferenzen“) und Sadighi et al., 2006 [107] (keine originäre Validierungsstudie). Präzisierungen: Osoba et al., 1997 [105] (Subgruppe: Dutch sample), Teckle et al., 2011 [108] (Subgruppe: Lung cancer).

KRK: Kolorektalkarzinom, n/a: nicht anwendbar, NSCLC: Nicht-kleinzelliges Bronchialkarzinom, (rr)MM: (Rezidiertes oder refraktäres) Multiples Myelom.

Tabelle G20: Abweichung zur Referenz mit AWG-Übereinstimmung beim EORTC QLQ-C30; hier: Geschlecht (Quelle: Eigene Darstellung).

Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie		Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie	
			absolut	relativ				absolut	relativ
Crizotinib [PROFILE 1007]; nicht auswertbar	NSCLC	Braun et al., 2011	nicht berichtet	n/a	Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) [PCYC-1112-CA]; nicht auswertbar	Leukämie, Lymphom	Hjermstad et al., 1995	nicht berichtet	n/a
Crizotinib [PROFILE 1007]; p = 0,0002	NSCLC	Teckle et al., 2011	- 0,329	- 32,4 %	Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) [PCYC-1104-CA]; nicht auswertbar	Leukämie, Lymphom	Hjermstad et al., 1995	nicht berichtet	n/a
Crizotinib [PROFILE 1005]; nicht auswertbar	NSCLC	Braun et al., 2011	nicht berichtet	n/a	Nintedanib (nicht-kleinzelliges Lungenkarzinom); nicht auswertbar	NSCLC	Braun et al., 2011	nicht berichtet	n/a
Crizotinib [PROFILE 1005]; p = 0,3472	NSCLC	Teckle et al., 2011	- 0,091	- 8,9 %	Nintedanib (nicht-kleinzelliges Lungenkarzinom); p < 0,0001	NSCLC	Teckle et al., 2011	- 0,313	- 30,8 %
Pomalidomid; nicht auswertbar	(rr)MM	Kvam et al., 2011	nicht berichtet	n/a	Panobinostat; nicht auswertbar	(rr)MM	Dubois et al., 2006	nicht berichtet	n/a
Pomalidomid; nicht auswertbar	(rr)MM	Wisloff et al., 1996	nicht berichtet	n/a	Panobinostat; nicht auswertbar	(rr)MM	Kvam et al., 2011	nicht berichtet	n/a
Regorafenib; nicht auswertbar	KRK	Osoba et al., 1997	nicht berichtet	n/a	Regorafenib (Neubewertung nach Fristablauf); nicht auswertbar	KRK	Osoba et al., 1997	nicht berichtet	n/a
Afatinib [LUX-Lung 3]; nicht auswertbar	NSCLC	Braun et al., 2011	nicht berichtet	n/a	Carfilzomib; p = 0,0060	(rr)MM	Delforge et al., 2012	- 0,095	- 12,5 %
Afatinib [LUX-Lung 3]; p < 0,0001	NSCLC	Teckle et al., 2011	- 0,399	- 39,2 %	Carfilzomib; nicht auswertbar	(rr)MM	Kvam et al., 2011	nicht berichtet	n/a
Afatinib [LUX-Lung 6]; nicht auswertbar	NSCLC	Braun et al., 2011	nicht berichtet	n/a	Carfilzomib; nicht auswertbar	(rr)MM	Wisloff et al., 1996	nicht berichtet	n/a
Afatinib [LUX-Lung 6]; p < 0,0001	NSCLC	Teckle et al., 2011	- 0,261	- 25,7 %	Crizotinib (neues Anwendungsgebiet); nicht auswertbar	NSCLC	Braun et al., 2011	nicht berichtet	n/a
Eribulin (neues Anwendungsgebiet, Neubewertung nach Fristablauf); p = 0,0380	Mammakarzinom	McLachlan et al., 1998	- 0,104	- 14,2 %	Crizotinib (neues Anwendungsgebiet); p = 0,0005	NSCLC	Teckle et al., 2011	- 0,277	- 27,2 %
Eribulin (neues Anwendungsgebiet, Neubewertung nach Fristablauf); nicht auswertbar	Mammakarzinom	Parmar et al., 2005	nicht berichtet	n/a					

Operationalisierung: Häufigkeitsgewichtete Maßzahl zum *Eastern Cooperative Oncology Group Performance Status (ECOG PS gw.)*.

Anmerkung: Exklusive Aaronson et al., 1993 [64] (vgl. detaillierte Betrachtung in Kapitel 5.5, Komplex „Standardreferenzen“) und Sadighi et al., 2006 [107] (keine originäre Validierungsstudie). Präzisierungen: Osoba et al., 1997 [105] (Subgruppe: Dutch sample), Teckle et al., 2011 [108] (Subgruppe: Lung cancer).

KRK: Kolorektalkarzinom, n/a: nicht anwendbar, NSCLC: Nicht-kleinzelliges Bronchialkarzinom, (rr)MM: (Rezidiertes oder refraktäres) Multiples Myelom.

Tabelle G21: Abweichung zur Referenz mit AWG-Übereinstimmung beim *EORTC QLQ-C30*; hier: Grad der Erkrankungsschwere (Quelle: Eigene Darstellung).

Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie		Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie	
			absolut	relativ				absolut	relativ
Crizotinib [PROFILE 1007]; nicht auswertbar	NSCLC	Braun et al., 2011	nicht berichtet	n/a	Panobinostat; nicht auswertbar	(rr)MM	Kvam et al., 2011	nicht berichtet	n/a
Crizotinib [PROFILE 1007]; p = 0,0718	NSCLC	Teckle et al., 2011	12,8 %	32,6 %	Regorafenib (Neubewertung nach Fristablauf); nicht auswertbar	KRK	Osoba et al., 1997	nicht berichtet	n/a
Crizotinib [PROFILE 1005]; nicht auswertbar	NSCLC	Braun et al., 2011	nicht berichtet	n/a	Carfilzomib; nicht auswertbar	(rr)MM	Delforge et al., 2012	nicht berichtet	n/a
Crizotinib [PROFILE 1005]; p = 0,0021	NSCLC	Teckle et al., 2011	21,7 %	55,2 %	Carfilzomib; nicht auswertbar	(rr)MM	Kvam et al., 2011	nicht berichtet	n/a
Regorafenib; nicht auswertbar	KRK	Osoba et al., 1997	nicht berichtet	n/a	Carfilzomib; nicht auswertbar	(rr)MM	Wisloff et al., 1996	nicht berichtet	n/a
Afatinib [LUX-Lung 3]; nicht auswertbar	NSCLC	Braun et al., 2011	nicht berichtet	n/a	Crizotinib (neues Anwendungsgebiet); nicht auswertbar	NSCLC	Braun et al., 2011	nicht berichtet	n/a
Afatinib [LUX-Lung 3]; p = 0,0931	NSCLC	Teckle et al., 2011	- 11,5 %	- 29,3 %	Crizotinib (neues Anwendungsgebiet); p = 0,0960	NSCLC	Teckle et al., 2011	12,0 %	30,4 %
Afatinib [LUX-Lung 6]; nicht auswertbar	NSCLC	Braun et al., 2011	nicht berichtet	n/a					
Afatinib [LUX-Lung 6]; p < 0,0001	NSCLC	Teckle et al., 2011	- 39,3 %	n/a (vs. 0 %)					
Eribulin (neues Anwendungsgebiet, Neubewertung nach Fristablauf); nicht auswertbar	Mammakarzinom	McLachlan et al., 1998	nicht berichtet	n/a					
Eribulin (neues Anwendungsgebiet, Neubewertung nach Fristablauf); nicht auswertbar	Mammakarzinom	Parmar et al., 2005	nicht berichtet	n/a					
Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) [PCYC-1112-CA]; nicht auswertbar	Leukämie, Lymphom	Hjermstad et al., 1995	nicht berichtet	n/a					
Panobinostat; p < 0,0001	(rr)MM	Dubois et al., 2006	- 16,2 %	- 20,0 %					

Anmerkung: Exklusive Aaronson et al., 1993 [64] (vgl. detaillierte Betrachtung in Kapitel 5.5, Komplex „Standardreferenzen“) und Sadighi et al., 2006 [107] (keine originäre Validierungsstudie). Präzisierungen: Osoba et al., 1997 [105] (Subgruppe: Dutch sample), Teckle et al., 2011 [108] (Subgruppe: Lung cancer).

KRK: Kolorektalkarzinom, n/a: nicht anwendbar, NSCLC: Nicht-kleinzelliges Bronchialkarzinom, (rr)MM: (Rezidiertes oder refraktäres) Multiples Myelom.

Tabelle G22: Abweichung zur Referenz mit AWG-Übereinstimmung beim EORTC QLQ-C30; hier: Ethnizität (Quelle: Eigene Darstellung).

Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie		Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie	
			absolut	relativ				absolut	relativ
Afatinib [LUX-Lung 6]; p = 1,0000	NSCLC	Braun et al., 2011	0 %	n/a	Crizotinib (neues Anwendungsgebiet); p < 0,0001	NSCLC	Braun et al., 2011	37,4 %	n/a (vs. 0 %)
Afatinib [LUX-Lung 6]; p = 1,0000	NSCLC	Teckle et al., 2011	0 %	n/a	Crizotinib (neues Anwendungsgebiet); p < 0,0001	NSCLC	Teckle et al., 2011	37,4 %	n/a (vs. 0 %)
Eribulin (neues Anwendungsgebiet, Neubewertung nach Fristablauf); p < 0,0001	Mammakarzinom	McLachlan et al., 1998	72,1 %	n/a (vs. 0 %)					
Eribulin (neues Anwendungsgebiet, Neubewertung nach Fristablauf); p < 0,0001	Mammakarzinom	Parmar et al., 2005	72,1 %	n/a (vs. 0 %)					
Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) [PCYC-1112-CA]; p < 0,0001	Leukämie, Lymphom	Hjermstad et al., 1995	- 56,5 %	n/a (vs. 100 %)					
Nintedanib (nicht-kleinzelliges Lungenkarzinom); p < 0,0001	NSCLC	Braun et al., 2011	70,4 %	n/a (vs. 0 %)					
Nintedanib (nicht-kleinzelliges Lungenkarzinom); p < 0,0001	NSCLC	Teckle et al., 2011	70,4 %	n/a (vs. 0 %)					
Regorafenib (Neubewertung nach Fristablauf); p < 0,0001	KRK	Osoba et al., 1997	- 100 %	n/a					
Carfilzomib; p = 0,1942	(rr)MM	Delforge et al., 2012	- 3,0 %	- 3,9 %					
Carfilzomib; p < 0,0001	(rr)MM	Kvam et al., 2011	- 25,5 %	n/a (vs. 100 %)					
Carfilzomib; p < 0,0001	(rr)MM	Wisloff et al., 1996	- 25,5 %	n/a (vs. 100 %)					

Anmerkung: Exklusive Aaronson et al., 1993 [64] (vgl. detaillierte Betrachtung in Kapitel 5.5, Komplex „Standardreferenzen“) und Sadighi et al., 2006 [107] (keine originäre Validierungsstudie). Präzisierungen: Osoba et al., 1997 [105] (Subgruppe: Dutch sample), Teckle et al., 2011 [108] (Subgruppe: Lung cancer).

KRK: Kolorektalkarzinom, n/a: nicht anwendbar, NSCLC: Nicht-kleinzelliges Bronchialkarzinom, (rr)MM: (Rezidiertes oder refraktäres) Multiples Myelom.

Tabelle G23: Abweichung zur Referenz mit AWG-Übereinstimmung beim EORTC QLQ-C30; hier: Geographische Region (Quelle: Eigene Darstellung).

G-II: Abweichungen zur Referenz mit AWG-Übereinstimmung; hier: *EQ-5D (VAS)*

Akakura et al., 2011 [112]		
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	prostate cancer
Relativer Anteil:		100 %
Radium-223-dichlorid	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)	X

Tabelle G24: Entsprechung nach AWG beim *EQ-5D (VAS)*; hier: Akakura et al., 2011
(Quelle: Eigene Darstellung).

Brazier et al., 1993 [113]		
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	general population in Sheffield, UK
Relativer Anteil:		100 %
Afatinib	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	n/a
Nintedanib (nicht-kleinzelliges Lungenkarzinom)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	n/a
Ramucirumab	Magenkarzinom	n/a
Afatinib (Neubewertung nach Fristablauf)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	n/a
Cobimetinib	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)	n/a
Radium-223-dichlorid	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)	n/a

Tabelle G25: Entsprechung nach AWG beim *EQ-5D (VAS)*; hier: Brazier et al., 1993
(Quelle: Eigene Darstellung).

Brooks et al., 1991 [114]		
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	representative sample of the Swedish population
Relativer Anteil:		100 %
Radium-223-dichlorid	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)	n/a

Tabelle G26: Entsprechung nach AWG beim *EQ-5D (VAS)*; hier: Brooks et al., 1991
(Quelle: Eigene Darstellung).

de Boer et al., 2004 [115]		
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	adenocarcinoma of the oesophagus or oesophagogastric junction
Relativer Anteil:		100 %
Afatinib	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	-
Nintedanib (nicht-kleinzelliges Lungenkarzinom)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	-
Ramucirumab	Magenkarzinom	(X)
Afatinib (Neubewertung nach Fristablauf)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	-
Cobimetinib	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)	-

Tabelle G27: Entsprechung nach AWG beim *EQ-5D (VAS)*; hier: de Boer et al., 2004
(Quelle: Eigene Darstellung).

n/a: nicht übertragbar			Greiner et al., 2005 [116]
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	representative sample of the German population in the northern German states of Lower Saxony and Bremen	
Relativer Anteil:		100 %	
Nivolumab (nicht-kleinzelliges Lungenkarzinom)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	n/a	
Pomalidomid (erneute Nutzenbewertung)	Rezidiertes oder refraktäres Multiples Myelom (rrMM)	n/a	

Tabelle G28: Entsprechung nach AWG beim EQ-5D (VAS); hier: Greiner et al., 2005
(Quelle: Eigene Darstellung).

Hurst et al., 1997 [67]		
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	rheumatoid arthritis (RA)
Relativer Anteil:		100 %
Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom)	Chronische lymphatische Leukämie (CLL); Rezidiertes oder refraktäres Mantelzell-Lymphom (MCL)	-

Tabelle G29: Entsprechung nach AWG beim EQ-5D (VAS); hier: Hurst et al., 1997
(Quelle: Eigene Darstellung).

n/a: nicht übertragbar			Johnson et al., 1998 [117]
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	national sample of the continental US population	
Relativer Anteil:		100 %	
Radium-223-dichlorid	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)	n/a	

Tabelle G30: Entsprechung nach AWG beim EQ-5D (VAS); hier: Johnson et al., 1998
(Quelle: Eigene Darstellung).

Kvam et al., 2011 [102]		
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	multiple myeloma (MM)
Relativer Anteil:		100 %
Pomalidomid	Rezidiertes oder refraktäres Multiples Myelom (rrMM)	X
Pomalidomid (erneute Nutzenbewertung)	Rezidiertes oder refraktäres Multiples Myelom (rrMM)	X

Tabelle G31: Entsprechung nach AWG beim EQ-5D (VAS); hier: Kvam et al., 2011
(Quelle: Eigene Darstellung).

Pickard et al., 2007a [118]

Nutzenbewertungs- verfahren	Anwendungsgebiet (i. e. S.)	bladder cancer	brain cancer	breast cancer	colorectal cancer	head and neck cancer	liver cancer	pancreatic cancer	kidney cancer	lung cancer	lymphoma	ovarian cancer	prostate cancer
Relativer Anteil:		5,8 %	ca. 9,6 %	ca. 9,6 %	ca. 9,6 %	ca. 9,6 %	ca. 4,8 %	ca. 4,8 %	ca. 9,6 %	9,4 %	ca. 9,6 %	ca. 9,6 %	ca. 9,6 %
Bosutinib	Chronische myeloische Leukämie (CML)	-	-	-	-	-	-	-	-	-	-	-	-
Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom)	Chronische lymphatische Leukämie (CLL); Rezidiertes oder refraktäres Mantelzell-Lymphom (MCL)	-	-	-	-	-	-	-	-	-	X	-	-
Nivolumab (nicht-kleinzelliges Lungenkarzinom)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	-	-	-	-	-	-	-	-	(X)	-	-	-
Dabrafenib (neues Anwendungsgebiet)	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)	-	-	-	-	-	-	-	-	-	-	-	-
Trametinib	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)	-	-	-	-	-	-	-	-	-	-	-	-
Cobimetinib	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)	-	-	-	-	-	-	-	-	-	-	-	-

Tabelle G32: Entsprechung nach AWG beim EQ-5D (VAS); hier: Pickard et al., 2007a (Quelle: Eigene Darstellung).

Rabin und de Charro, 2001 [66]		
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	rheumatoid arthritis (RA) [#]
Relativer Anteil:		100 %
Bosutinib	Chronische myeloische Leukämie (CML)	-
Dabrafenib	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)	-
Afatinib	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	-
Idelalisib	Chronische lymphatische Leukämie (CLL); Follikuläres Lymphom (FL)	-
Nintedanib (nicht-kleinzelliges Lungenkarzinom)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	-
Ramucirumab	Magenkarzinom	-
Afatinib (Neubewertung nach Fristablauf)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	-
Pembrolizumab	Fortgeschrittenes Melanom	-
Dabrafenib (neues Anwendungsgebiet)	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)	-
Trametinib	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)	-

[#] Verweis auf Hurst et al., 1997 [67].

Tabelle G33: Entsprechung nach AWG beim EQ-5D (VAS); hier: Rabin und de Charro, 2001 (Quelle: Eigene Darstellung).

Schuette et al., 2012 [119]		
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	(advanced) non-small-cell lung cancer (NSCLC)
Relativer Anteil:		100 %
Crizotinib	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	X
Afatinib	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	X
Nintedanib (nicht-kleinzelliges Lungenkarzinom)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	X
Afatinib (Neubewertung nach Fristablauf)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	X
Cobimetinib	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)	-

Tabelle G34: Entsprechung nach AWG beim EQ-5D (VAS); hier: Schuette et al., 2012 (Quelle: Eigene Darstellung).

Teckle et al., 2011 [108]				
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	breast cancer	colorectal cancer	lung cancer
Relativer Anteil:		35,9 %	31,0 %	33,2 %
Crizotinib	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	-	-	(X)
Afatinib	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	-	-	(X)
Nintedanib (nicht-kleinzelliges Lungenkarzinom)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	-	-	(X)
Afatinib (Neubewertung nach Fristablauf)	Nicht-kleinzelliges Bronchialkarzinom (NSCLC)	-	-	(X)
Cobimetinib	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)	-	-	-

Tabelle G35: Entsprechung nach AWG beim EQ-5D (VAS); hier: Teckle et al., 2011 (Quelle: Eigene Darstellung).

Wisloff et al., 1996 [109]		
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	multiple myeloma (MM)
Relativer Anteil:		100 %
Pomalidomid	Rezidiertes oder refraktäres Multiples Myelom (rrMM)	X
Pomalidomid (erneute Nutzenbewertung)	Rezidiertes oder refraktäres Multiples Myelom (rrMM)	X

Tabelle G36: Entsprechung nach AWG beim EQ-5D (VAS); hier: Wisloff et al., 1996
(Quelle: Eigene Darstellung).

Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie		Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie	
			absolut	relativ				absolut	relativ
Crizotinib [PROFILE 1007]; nicht auswertbar	NSCLC	Schuette et al., 2012	- 16,3	- 24,6 %	Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) [PCYC-1104-CA]; p < 0,0001	Leukämie, Lymphom	Pickard et al., 2007a	8,1	13,7 %
Crizotinib [PROFILE 1007]; nicht auswertbar	NSCLC	Teckle et al., 2011	- 13,0	- 20,6 %	Nintedanib (nicht-kleinzelliges Lungenkarzinom); nicht auswertbar	NSCLC	Brazier et al., 1993	17,1	41,2 %
Crizotinib [PROFILE 1005]; nicht auswertbar	NSCLC	Schuette et al., 2012	- 13,3	- 20,1 %	Nintedanib (nicht-kleinzelliges Lungenkarzinom); nicht auswertbar	NSCLC	de Boer et al., 2004	- 5,4	- 8,4 %
Crizotinib [PROFILE 1005]; nicht auswertbar	NSCLC	Teckle et al., 2011	- 10,0	- 15,9 %	Nintedanib (nicht-kleinzelliges Lungenkarzinom); nicht auswertbar	NSCLC	Schuette et al., 2012	- 7,7	- 11,6 %
Bosutinib; p < 0,0001	Leukämie, Lymphom	Pickard et al., 2007a	- 7,7	- 13,1 %	Nintedanib (nicht-kleinzelliges Lungenkarzinom); p = 0,0008	NSCLC	Teckle et al., 2011	- 4,4	- 7,0 %
Pomalidomid; nicht auswertbar	(rr)MM	Kvam et al., 2011	- 2,4	- 3,6 %	Ramucirumab [RAINBOW]; nicht auswertbar	Magenkarzinom	Brazier et al., 1993	19,5	47,0 %
Pomalidomid; nicht auswertbar	(rr)MM	Wisloff et al., 1996	- 2,4	- 3,6 %	Ramucirumab [RAINBOW]; nicht auswertbar	Magenkarzinom	de Boer et al., 2004	- 3,0	- 4,7 %
Afatinib [LUX-Lung 3]; nicht auswertbar	NSCLC	Brazier et al., 1993	18,7	45,1 %	Ramucirumab [REGARD]; nicht auswertbar	Magenkarzinom	Brazier et al., 1993	18,5	44,6 %
Afatinib [LUX-Lung 3]; nicht auswertbar	NSCLC	de Boer et al., 2004	- 3,8	- 5,9 %	Ramucirumab [REGARD]; nicht auswertbar	Magenkarzinom	de Boer et al., 2004	- 4,0	- 6,3 %
Afatinib [LUX-Lung 3]; nicht auswertbar	NSCLC	Schuette et al., 2012	- 6,1	- 9,2 %	Nivolumab (nicht-kleinzelliges Lungenkarzinom); nicht auswertbar	NSCLC	Greiner et al., 2005	nicht berichtet	n/a
Afatinib [LUX-Lung 3]; p = 0,0457	NSCLC	Teckle et al., 2011	- 2,8	- 4,4 %	Nivolumab (nicht-kleinzelliges Lungenkarzinom); p < 0,0001	NSCLC	Pickard et al., 2007a	4,3	7,3 %
Afatinib [LUX-Lung 6]; nicht auswertbar	NSCLC	Brazier et al., 1993	14,8	35,7 %	Dabrafenib (neues Anwendungsgebiet); p < 0,0001	Melanom	Pickard et al., 2007a	- 4,8	- 8,1 %
Afatinib [LUX-Lung 6]; nicht auswertbar	NSCLC	de Boer et al., 2004	- 7,7	- 12,0 %	Trametinib; p < 0,0001	Melanom	Pickard et al., 2007a	- 5,3	- 9,0 %
Afatinib [LUX-Lung 6]; nicht auswertbar	NSCLC	Schuette et al., 2012	- 10,0	- 15,1 %	Cobimetinib; nicht auswertbar	Melanom	Brazier et al., 1993	13,6	32,8 %
Afatinib [LUX-Lung 6]; p < 0,0001	NSCLC	Teckle et al., 2011	- 6,7	- 10,6 %	Cobimetinib; nicht auswertbar	Melanom	de Boer et al., 2004	- 8,9	- 13,9 %
Radium-223-dichlorid; nicht auswertbar	mCRPC	Brooks et al., 1991	23,9	51,4 %	Cobimetinib; p < 0,0001	Melanom	Pickard et al., 2007a	- 3,9	- 6,6 %
Radium-223-dichlorid; p < 0,0001	mCRPC	Johnson et al., 1998	17,9	34,2 %	Cobimetinib; nicht auswertbar	Melanom	Schuette et al., 2012	- 11,2	- 16,9 %
Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) [PCYC-1112-CA]; p < 0,0001	Leukämie, Lymphom	Pickard et al., 2007a	7,5	12,7 %	Cobimetinib; p = 0,0032	Melanom	Teckle et al., 2011	- 3,4	- 5,8 %

Anmerkung: Exklusive Rabin und de Charro, 2001 [66] sowie Hurst et al., 1997 [67] (vgl. jeweils detaillierte Betrachtung in Kapitel 5.5, Komplex „Standardreferenzen“). Weiterer Ausschluss: Akakura et al., 2011 [112] (Originalartikel in japanischer Sprache verfasst). Präzisierungen/Bemerkungen: Teckle et al., 2011 [108] (Crizotinib, Afatinib, Nintedanib: Subgruppe Lung cancer; Cobimetinib: All cancers); Brazier et al., 1993 [113] (approximiert anhand vorliegender Angaben zu Altersklassen).

mCRPC: Metastasiertes kastrationsresistentes Prostatakarzinom, n/a: nicht anwendbar, NSCLC: Nicht-kleinzelliges Bronchialkarzinom, (rr)MM: (Rezidiviertes oder refraktäres) Multiples Myelom.

Tabelle G37: Abweichung zur Referenz mit AWG-Übereinstimmung beim EQ-5D (VAS); hier: Alter (Quelle: Eigene Darstellung).

Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie	
			absolut	relativ
Crizotinib [PROFILE 1007]; p < 0,0001	NSCLC	Schuette et al., 2012	25,6 %	84,4 %
Crizotinib [PROFILE 1007]; p = 0,2649	NSCLC	Teckle et al., 2011	8,4 %	17,6 %
Crizotinib [PROFILE 1005]; p < 0,0001	NSCLC	Schuette et al., 2012	22,7 %	75,0 %
Crizotinib [PROFILE 1005]; p = 0,4942	NSCLC	Teckle et al., 2011	5,5 %	11,6 %
Bosutinib; p = 0,7148	Leukämie, Lymphom	Pickard et al., 2007a	- 1,4 %	- 3,0 %
Pomalidomid; p = 0,1967	(rr)MM	Kvam et al., 2011	- 5,3 %	- 11,5 %
Pomalidomid; p = 0,3647	(rr)MM	Wisloff et al., 1996	- 2,9 %	- 6,6 %
Afatinib [LUX-Lung 3]; p = 0,0007	NSCLC	Brazier et al., 1993	10,0 %	18,2 %
Afatinib [LUX-Lung 3]; p < 0,0001	NSCLC	de Boer et al., 2004	50,5 %	349,1 %
Afatinib [LUX-Lung 3]; p < 0,0001	NSCLC	Schuette et al., 2012	34,6 %	114,1 %
Afatinib [LUX-Lung 3]; p = 0,0143	NSCLC	Teckle et al., 2011	17,4 %	36,6 %
Afatinib [LUX-Lung 6]; p = 0,0003	NSCLC	Brazier et al., 1993	10,5 %	19,1 %
Afatinib [LUX-Lung 6]; p < 0,0001	NSCLC	de Boer et al., 2004	50,9 %	352,2 %
Afatinib [LUX-Lung 6]; p < 0,0001	NSCLC	Schuette et al., 2012	35,1 %	115,6 %
Afatinib [LUX-Lung 6]; p = 0,0098	NSCLC	Teckle et al., 2011	17,8 %	37,5 %
Radium-223-dichlorid; p < 0,0001	mCRPC	Brooks et al., 1991	- 53,0 %	n/a (vs. 0 %)
Radium-223-dichlorid; nicht auswertbar	mCRPC	Johnson et al., 1998	„predominantly male“	
Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) [PCYC-1112-CA]; p < 0,0001	Leukämie, Lymphom	Pickard et al., 2007a	- 16,3 %	- 33,8 %

Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie	
			absolut	relativ
Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) [PCYC-1104-CA]; p < 0,0001	Leukämie, Lymphom	Pickard et al., 2007a	- 24,9 %	- 51,5 %
Nintedanib (nicht-kleinzelliges Lungenkarzinom); p < 0,0001	NSCLC	Brazier et al., 1993	- 17,4 %	- 31,7 %
Nintedanib (nicht-kleinzelliges Lungenkarzinom); p < 0,0001	NSCLC	de Boer et al., 2004	23,1 %	159,6 %
Nintedanib (nicht-kleinzelliges Lungenkarzinom); p = 0,0113	NSCLC	Schuette et al., 2012	7,2 %	23,8 %
Nintedanib (nicht-kleinzelliges Lungenkarzinom); p = 0,1317	NSCLC	Teckle et al., 2011	- 10,0 %	- 21,0 %
Ramucirumab [RAINBOW]; p < 0,0001	Magenkarzinom	Brazier et al., 1993	- 25,9 %	- 47,2 %
Ramucirumab [RAINBOW]; p = 0,0040	Magenkarzinom	de Boer et al., 2004	14,6 %	100,7 %
Ramucirumab [REGARD]; p < 0,0001	Magenkarzinom	Brazier et al., 1993	- 24,8 %	- 45,1 %
Ramucirumab [REGARD]; p = 0,0038	Magenkarzinom	de Boer et al., 2004	15,7 %	108,5 %
Nivolumab (nicht-kleinzelliges Lungenkarzinom); nicht auswertbar	NSCLC	Greiner et al., 2005	nicht berichtet	n/a
Nivolumab (nicht-kleinzelliges Lungenkarzinom); p < 0,0001	NSCLC	Pickard et al., 2007a	- 24,8 %	- 51,3 %
Dabrafenib (neues Anwendungsgebiet); p = 0,2498	Melanom	Pickard et al., 2007a	- 3,4 %	- 7,1 %
Trametinib; p = 0,5726	Melanom	Pickard et al., 2007a	- 2,0 %	- 4,2 %
Cobimetinib; p < 0,0001	Melanom	Brazier et al., 1993	- 12,7 %	- 23,1 %
Cobimetinib; p < 0,0001	Melanom	de Boer et al., 2004	27,8 %	192,0 %
Cobimetinib; p = 0,0523	Melanom	Pickard et al., 2007a	- 6,1 %	- 12,6 %
Cobimetinib; p < 0,0001	Melanom	Schuette et al., 2012	11,9 %	39,2 %
Cobimetinib; p < 0,0001	Melanom	Teckle et al., 2011	- 22,5 %	- 34,7 %

Anmerkung: Exklusive Rabin und de Charro, 2001 [66] sowie Hurst et al., 1997 [67] (vgl. jeweils detaillierte Betrachtung in Kapitel 5.5, Komplex „Standardreferenzen“). Weiterer Ausschluss: Akakura et al., 2011 [112] (Originalartikel in japanischer Sprache verfasst). Präzisierungen/Bemerkungen: Teckle et al., 2011 [108] (Crizotinib, Afatinib, Nintedanib: Subgruppe Lung cancer; Cobimetinib: All cancers). mCRPC: Metastasiertes kastrationsresistentes Prostatakarzinom, n/a: nicht anwendbar, NSCLC: Nicht-kleinzelliges Bronchialkarzinom, (rr)MM: (Rezidiertes oder refraktäres) Multiples Myelom.

Tabelle G38: Abweichung zur Referenz mit AWG-Übereinstimmung beim EQ-5D (VAS); hier: Geschlecht (Quelle: Eigene Darstellung).

Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie		Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie	
			absolut	relativ				absolut	relativ
Crizotinib [PROFILE 1007]; p = 0,6568	NSCLC	Schuette et al., 2012	- 0,018	- 2,5 %	Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) [PCYC-1104-CA]; p = 0,0027	Leukämie, Lymphom	Pickard et al., 2007a	- 0,230	- 26,7 %
Crizotinib [PROFILE 1007]; p = 0,0002	NSCLC	Teckle et al., 2011	- 0,329	- 32,4 %	Nintedanib (nicht-kleinzelliges Lungenkarzinom); nicht auswertbar	NSCLC	Brazier et al., 1993	nicht berichtet	n/a
Crizotinib [PROFILE 1005]; p < 0,0001	NSCLC	Schuette et al., 2012	0,220	31,2 %	Nintedanib (nicht-kleinzelliges Lungenkarzinom); nicht auswertbar	NSCLC	de Boer et al., 2004	nicht berichtet	n/a
Crizotinib [PROFILE 1005]; p = 0,3472	NSCLC	Teckle et al., 2011	- 0,091	- 8,9 %	Nintedanib (nicht-kleinzelliges Lungenkarzinom); p = 0,9461	NSCLC	Schuette et al., 2012	- 0,002	- 0,2 %
Bosutinib; p < 0,0001	Leukämie, Lymphom	Pickard et al., 2007a	- 0,508	- 59,1 %	Nintedanib (nicht-kleinzelliges Lungenkarzinom); p < 0,0001	NSCLC	Teckle et al., 2011	- 0,313	- 30,8 %
Pomalidomid; nicht auswertbar	(rr)MM	Kvam et al., 2011	nicht berichtet	n/a	Ramucirumab [RAINBOW]; nicht auswertbar	Magenkarzinom	Brazier et al., 1993	nicht berichtet	n/a
Pomalidomid; nicht auswertbar	(rr)MM	Wisloff et al., 1996	nicht berichtet	n/a	Ramucirumab [RAINBOW]; nicht auswertbar	Magenkarzinom	de Boer et al., 2004	nicht berichtet	n/a
Afatinib [LUX-Lung 3]; nicht auswertbar	NSCLC	Brazier et al., 1993	nicht berichtet	n/a	Ramucirumab [REGARD]; nicht auswertbar	Magenkarzinom	Brazier et al., 1993	nicht berichtet	n/a
Afatinib [LUX-Lung 3]; nicht auswertbar	NSCLC	de Boer et al., 2004	nicht berichtet	n/a	Ramucirumab [REGARD]; nicht auswertbar	Magenkarzinom	de Boer et al., 2004	nicht berichtet	n/a
Afatinib [LUX-Lung 3]; p = 0,0175	NSCLC	Schuette et al., 2012	- 0,088	- 12,4 %	Nivolumab (nicht-kleinzelliges Lungenkarzinom); nicht auswertbar	NSCLC	Greiner et al., 2005	nicht berichtet	n/a
Afatinib [LUX-Lung 3]; p < 0,0001	NSCLC	Teckle et al., 2011	- 0,399	- 39,2 %	Nivolumab (nicht-kleinzelliges Lungenkarzinom); p = 0,0341	NSCLC	Pickard et al., 2007a	- 0,103	- 12,0 %
Afatinib [LUX-Lung 6]; nicht auswertbar	NSCLC	Brazier et al., 1993	nicht berichtet	n/a	Dabrafenib (neues Anwendungsgebiet); p < 0,0001	Melanom	Pickard et al., 2007a	- 0,565	- 65,7 %
Afatinib [LUX-Lung 6]; nicht auswertbar	NSCLC	de Boer et al., 2004	nicht berichtet	n/a	Trametinib; p < 0,0001	Melanom	Pickard et al., 2007a	- 0,500	- 58,1 %
Afatinib [LUX-Lung 6]; p = 0,1496	NSCLC	Schuette et al., 2012	0,050	7,1 %	Cobimetinib; nicht auswertbar	Melanom	Brazier et al., 1993	nicht berichtet	n/a
Afatinib [LUX-Lung 6]; p < 0,0001	NSCLC	Teckle et al., 2011	- 0,261	- 25,7 %	Cobimetinib; nicht auswertbar	Melanom	de Boer et al., 2004	nicht berichtet	n/a
Radium-223-dichlorid; nicht auswertbar	mCRPC	Brooks et al., 1991	nicht berichtet	n/a	Cobimetinib; p < 0,0001	Melanom	Pickard et al., 2007a	- 0,572	- 66,5 %
Radium-223-dichlorid; nicht auswertbar	mCRPC	Johnson et al., 1998	nicht berichtet	n/a	Cobimetinib; p < 0,0001	Melanom	Schuette et al., 2012	- 0,418	- 59,2 %
Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) [PCYC-1112-CA]; p < 0,0001	Leukämie, Lymphom	Pickard et al., 2007a	- 0,267	- 31,0 %	Cobimetinib; p < 0,0001	Melanom	Teckle et al., 2011	- 0,479	- 62,5 %

Operationalisierung: Häufigkeitsgewichtete Maßzahl zum Eastern Cooperative Oncology Group Performance Status (ECOG PS gw.).

Anmerkung: Exklusive Rabin und de Charro, 2001 [66] sowie Hurst et al., 1997 [67] (vgl. jeweils detaillierte Betrachtung in Kapitel 5.5, Komplex „Standardreferenzen“). Weiterer Ausschluss: Akakura et al., 2011 [112] (Originalartikel in japanischer Sprache verfasst). Präzisierungen/Bemerkungen: Teckle et al., 2011 [108] (Crizotinib, Afatinib, Nintedanib: Subgruppe Lung cancer; Cobimetinib: All cancers). mCRPC: Metastasiertes kastrationsresistentes Prostatakarzinom, n/a: nicht anwendbar, NSCLC: Nicht-kleinzelliges Bronchialkarzinom, (rr)MM: (Rezidiertes oder refraktäres) Multiples Myelom.

Tabelle G39: Abweichung zur Referenz mit AWG-Übereinstimmung beim EQ-5D (VAS); hier: Grad der Erkrankungsschwere (Quelle: Eigene Darstellung).

Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie		Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie	
			absolut	relativ				absolut	relativ
Crizotinib [PROFILE 1007]; p < 0,0001	NSCLC	Schuette et al., 2012	- 46,5 %	- 47,1 %	Ramucirumab [RAINBOW]; nicht auswertbar	Magenkarzinom	Brazier et al., 1993	nicht berichtet	n/a
Crizotinib [PROFILE 1007]; p = 0,0718	NSCLC	Teckle et al., 2011	12,8 %	32,6 %	Ramucirumab [RAINBOW]; nicht auswertbar	Magenkarzinom	de Boer et al., 2004	nicht berichtet	n/a
Crizotinib [PROFILE 1005]; p < 0,0001	NSCLC	Schuette et al., 2012	- 37,6 %	- 38,1 %	Ramucirumab [REGARD]; nicht auswertbar	Magenkarzinom	Brazier et al., 1993	nicht berichtet	n/a
Crizotinib [PROFILE 1005]; p = 0,0021	NSCLC	Teckle et al., 2011	21,7 %	55,2 %	Ramucirumab [REGARD]; nicht auswertbar	Magenkarzinom	de Boer et al., 2004	nicht berichtet	n/a
Bosutinib; p < 0,0001	Leukämie, Lymphom	Pickard et al., 2007a	- 23,6 %	- 26,5 %	Nivolumab (nicht-kleinzelliges Lungenkarzinom); nicht auswertbar	NSCLC	Greiner et al., 2005	nicht berichtet	n/a
Afatinib [LUX-Lung 3]; nicht auswertbar	NSCLC	Brazier et al., 1993	nicht berichtet	n/a	Nivolumab (nicht-kleinzelliges Lungenkarzinom); p = 0,1047	NSCLC	Pickard et al., 2007a	3,9 %	4,4 %
Afatinib [LUX-Lung 3]; nicht auswertbar	NSCLC	de Boer et al., 2004	nicht berichtet	n/a	Dabrafenib (neues Anwendungsgebiet); p < 0,0001	Melanom	Pickard et al., 2007a	7,5 %	8,5 %
Afatinib [LUX-Lung 3]; p < 0,0001	NSCLC	Schuette et al., 2012	- 70,8 %	- 71,8 %	Cobimetinib; nicht auswertbar	Melanom	Brazier et al., 1993	nicht berichtet	n/a
Afatinib [LUX-Lung 3]; p = 0,0931	NSCLC	Teckle et al., 2011	- 11,5 %	- 29,3 %	Cobimetinib; nicht auswertbar	Melanom	de Boer et al., 2004	nicht berichtet	n/a
Afatinib [LUX-Lung 6]; nicht auswertbar	NSCLC	Brazier et al., 1993	nicht berichtet	n/a	Cobimetinib; p = 0,0121	Melanom	Pickard et al., 2007a	4,6 %	5,1 %
Afatinib [LUX-Lung 6]; nicht auswertbar	NSCLC	de Boer et al., 2004	nicht berichtet	n/a	Cobimetinib; p < 0,0001	Melanom	Schuette et al., 2012	- 5,3 %	- 5,4 %
Afatinib [LUX-Lung 6]; p < 0,0001	NSCLC	Schuette et al., 2012	- 98,7 %	n/a (vs. 0 %)	Cobimetinib; p < 0,0001	Melanom	Teckle et al., 2011	47,1 %	102,0 %
Afatinib [LUX-Lung 6]; p < 0,0001	NSCLC	Teckle et al., 2011	- 39,3 %	n/a (vs. 0 %)					
Radium-223-dichlorid; nicht auswertbar	mCRPC	Brooks et al., 1991	nicht berichtet	n/a					
Radium-223-dichlorid; nicht auswertbar	mCRPC	Johnson et al., 1998	„predominantly Caucasian“						
Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) [PCYC-1112-CA]; p = 0,6689	Leukämie, Lymphom	Pickard et al., 2007a	1,0 %	1,1 %					

Anmerkung: Exklusive Rabin und de Charro, 2001 [66] sowie Hurst et al., 1997 [67] (vgl. jeweils detaillierte Betrachtung in Kapitel 5.5, Komplex „Standardreferenzen“). Weiterer Ausschluss: Akakura et al., 2011 [112] (Originalartikel in japanischer Sprache verfasst). Präzisierungen/Bemerkungen: Teckle et al., 2011 [108] (Crizotinib, Afatinib, Nintedanib: Subgruppe Lung cancer; Cobimetinib: All cancers). mCRPC: Metastasiertes kastrationsresistentes Prostatakarzinom, n/a: nicht anwendbar, NSCLC: Nicht-kleinzelliges Bronchialkarzinom.

Tabelle G40: Abweichung zur Referenz mit AWG-Übereinstimmung beim EQ-5D (VAS); hier: Ethnizität (Quelle: Eigene Darstellung).

Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie		Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie	
			absolut	relativ				absolut	relativ
Afatinib [LUX-Lung 6]; p < 0,0001	NSCLC	Brazier et al., 1993	- 100 %	n/a	Nivolumab (nicht-kleinzelliges Lungenkarzinom); p < 0,0001	NSCLC	Greiner et al., 2005	- 43,0 %	n/a (vs. 100 %)
Afatinib [LUX-Lung 6]; p < 0,0001	NSCLC	de Boer et al., 2004	- 100 %	n/a	Nivolumab (nicht-kleinzelliges Lungenkarzinom); p < 0,0001	NSCLC	Pickard et al., 2007a	57,0 %	n/a (vs. 0 %)
Afatinib [LUX-Lung 6]; p < 0,0001	NSCLC	Schuette et al., 2012	- 100 %	n/a	Dabrafenib (neues Anwendungsgebiet); p < 0,0001	Melanom	Pickard et al., 2007a	65,9 %	n/a (vs. 0 %)
Afatinib [LUX-Lung 6]; p = 1,0000	NSCLC	Teckle et al., 2011	0 %	n/a	Trametinib; p < 0,0001	Melanom	Pickard et al., 2007a	68,0 %	n/a (vs. 0 %)
Ibrutinib (chronische lymphatische Leukämie, Mantelzell-Lymphom) [PCYC-1112-CA]; p < 0,0001	Leukämie, Lymphom	Pickard et al., 2007a	43,5 %	n/a (vs. 0 %)	Cobimetinib; p < 0,0001	Melanom	Brazier et al., 1993	- 26,1 %	n/a (vs. 100 %)
Nintedanib (nicht-kleinzelliges Lungenkarzinom); p < 0,0001	NSCLC	Brazier et al., 1993	- 29,6 %	n/a (vs. 100 %)	Cobimetinib; p < 0,0001	Melanom	de Boer et al., 2004	- 26,1 %	n/a (vs. 100 %)
Nintedanib (nicht-kleinzelliges Lungenkarzinom); p < 0,0001	NSCLC	de Boer et al., 2004	- 29,6 %	n/a (vs. 100 %)	Cobimetinib; p < 0,0001	Melanom	Pickard et al., 2007a	73,9 %	n/a (vs. 0 %)
Nintedanib (nicht-kleinzelliges Lungenkarzinom); p < 0,0001	NSCLC	Schuette et al., 2012	- 29,6 %	n/a (vs. 100 %)	Cobimetinib; p < 0,0001	Melanom	Schuette et al., 2012	- 26,1 %	n/a (vs. 100 %)
Nintedanib (nicht-kleinzelliges Lungenkarzinom); p < 0,0001	NSCLC	Teckle et al., 2011	70,4 %	n/a (vs. 0 %)	Cobimetinib; p < 0,0001	Melanom	Teckle et al., 2011	73,9 %	n/a (vs. 0 %)
Ramucirumab [RAINBOW]; p < 0,0001	Magenkarzinom	Brazier et al., 1993	- 54,4 %	n/a (vs. 100 %)	Detailangaben:				
Ramucirumab [RAINBOW]; p < 0,0001	Magenkarzinom	de Boer et al., 2004	- 54,4 %	n/a (vs. 100 %)	(i) Brazier et al., 1993 [113]: UK (100 %)				
Ramucirumab [REGARD]; p < 0,0001	Magenkarzinom	Brazier et al., 1993	- 51,5 %	n/a (vs. 100 %)	(ii) Brooks et al., 1991 [114]: Schweden (100 %)				
Ramucirumab [REGARD]; p < 0,0001	Magenkarzinom	de Boer et al., 2004	- 51,5 %	n/a (vs. 100 %)	(iii) de Boer et al., 2004 [115]: Niederlande (100 %)				
					(iv) Greiner et al., 2005 [116]: Deutschland (100 %)				
					(v) Johnson et al., 1998 [117]: US (100 %)				
					(vi) Kvam et al., 2011 [102]: Norwegen (100 %)				
					(vii) Pickard et al., 2007a [118]: US (100 %)				
					(viii) Schuette et al., 2012 [119] Deutschland und Österreich (100 %)				

Anmerkung: Exklusive Rabin und de Charro, 2001 [66] sowie Hurst et al., 1997 [67] (vgl. jeweils detaillierte Betrachtung in Kapitel 5.5, Komplex „Standardreferenzen“). Weiterer Ausschluss: Akakura et al., 2011 [112] (Originalartikel in japanischer Sprache verfasst). Präzisierungen/Bemerkungen: Teckle et al., 2011 [108] (Afatinib, Nintedanib: Subgruppe Lung cancer; Cobimetinib: All cancers). n/a: nicht anwendbar, NSCLC: Nicht-kleinzelliges Bronchialkarzinom.

Tabelle G41: Abweichung zur Referenz mit AWG-Übereinstimmung beim EQ-5D (VAS); hier: Geographische Region (Quelle: Eigene Darstellung).

G-III: Abweichungen zur Referenz mit AWG-Übereinstimmung; hier: *FACT-P*

Carducci et al., 2005 [120]		
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	metastatic hormone-refractory prostate cancer
Relativer Anteil:		100 %
Abirateronacetat	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)	X
Abirateronacetat (neues Anwendungsgebiet)	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)	X
Radium-223-dichlorid	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)	X

Tabelle G42: Entsprechung nach AWG beim *FACT-P*; hier: Carducci et al., 2005
(Quelle: Eigene Darstellung).

Cella et al., 2009 [69]		
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	metastatic hormone-refractory prostate cancer
Relativer Anteil:		100 %
Abirateronacetat	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)	X
Abirateronacetat (neues Anwendungsgebiet)	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)	X
Enzalutamid	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)	X
Radium-223-dichlorid	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)	X
Enzalutamid (neues Anwendungsgebiet)	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)	X

Tabelle G43: Entsprechung nach AWG beim *FACT-P*; hier: Cella et al., 2009
(Quelle: Eigene Darstellung).

Cella, 2012		
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	keine Angabe (nicht auffindbar)
Relativer Anteil:		-
Abirateronacetat (neues Anwendungsgebiet)	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)	n/a

Tabelle G44: Entsprechung nach AWG beim *FACT-P*; hier: Cella, 2012
(Quelle: Eigene Darstellung).

Esper et al., 1997 [68]		
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	prostate cancer
Relativer Anteil:		100 %
Abirateronacetat	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)	X
Abirateronacetat (neues Anwendungsgebiet)	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)	X
Enzalutamid	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)	X
Radium-223-dichlorid	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)	X
Enzalutamid (neues Anwendungsgebiet)	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)	X

Tabelle G45: Entsprechung nach AWG beim *FACT-P*; hier: Esper et al., 1997
(Quelle: Eigene Darstellung).

n/a: nicht übertragbar		Webster et al., 1999 [121]
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	mixed sample of cancer patients
Relativer Anteil:		100 %
Abirateronacetat (neues Anwendungsgebiet)	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)	n/a
Enzalutamid	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)	n/a
Radium-223-dichlorid	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)	n/a
Enzalutamid (neues Anwendungsgebiet)	Metastasiertes kastrationsresistentes Prostatakarzinom (mCRPC)	n/a

Tabelle G46: Entsprechung nach AWG beim *FACT-P*; hier: Webster et al., 1999
(Quelle: Eigene Darstellung).

Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie	
			absolut	relativ
Abirateronacetat; p = 1,0000	mCRPC	Carducci et al., 2005	0 %	n/a
Abirateronacetat (neues Anwendungsgebiet); p = 1,0000	mCRPC	Carducci et al., 2005	0 %	n/a
Abirateronacetat (neues Anwendungsgebiet); nicht auswertbar	mCRPC	Webster et al., 1999	nicht berichtet	n/a
Enzalutamid; nicht auswertbar	mCRPC	Webster et al., 1999	nicht berichtet	n/a
Radium-223-dichlorid; p = 1,0000	mCRPC	Carducci et al., 2005	0 %	n/a
Radium-223-dichlorid; nicht auswertbar	mCRPC	Webster et al., 1999	nicht berichtet	n/a
Enzalutamid (neues Anwendungsgebiet); nicht auswertbar	mCRPC	Webster et al., 1999	nicht berichtet	n/a

Anmerkung: Exklusive Cella et al., 2009 [69], Esper et al., 1997 [68] (vgl. jeweils detaillierte Betrachtung in Kapitel 5.5, Komplex „Standardreferenzen“) und Cella, 2012 (nicht auffindbar). Ergänzende Darstellung von Webster et al., 1999 [121] („mixed sample of cancer patients“; einzig als Abstract vorliegend).

mCRPC: Metastasiertes kastrationsresistentes Prostatakarzinom, n/a: nicht anwendbar.

Tabelle G47: Abweichung zur Referenz mit AWG-Übereinstimmung beim *FACT-P*; hier: Geschlecht (Quelle: Eigene Darstellung).

G-IV: Abweichungen zur Referenz mit AWG-Übereinstimmung; hier: *FACT-M*

Askew et al., 2009 [122]		
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	melanoma
Relativer Anteil:		100 %
Dabrafenib (neues Anwendungsgebiet)	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)	X
Trametinib	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)	X

Tabelle G48: Entsprechung nach AWG beim *FACT-M*; hier: Askew et al., 2009
(Quelle: Eigene Darstellung).

Cormier et al., 2008 [70]		
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	melanoma
Relativer Anteil:		100 %
Vemurafenib	Fortgeschrittenes Melanom (BRAF-V600 Mutation-positiv)	X
Vemurafenib (Neubewertung nach Fristablauf)	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)	X
Dabrafenib (neues Anwendungsgebiet)	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)	X
Trametinib	Fortgeschrittenes Melanom (BRAF-V600-Mutation-positiv)	X

Tabelle G49: Entsprechung nach AWG beim *FACT-M*; hier: Cormier et al., 2008
(Quelle: Eigene Darstellung).

Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie	
			absolut	relativ
Dabrafenib (neues Anwendungsgebiet); p = 0,3895	Melanom	Askew et al., 2009	3,1 %	7,5 %
Trametinib; p = 0,2823	Melanom	Askew et al., 2009	4,5 %	10,8 %

Anmerkung: Exklusive Cormier et al., 2008 [70] (vgl. detaillierte Betrachtung in Kapitel 5.5, Komplex „Standardreferenzen“).

Tabelle G50: Abweichung zur Referenz mit AWG-Übereinstimmung beim *FACT-M*; hier: Geschlecht (Quelle: Eigene Darstellung).

Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie	
			absolut	relativ
Dabrafenib (neues Anwendungsgebiet); p = 0,3753	Melanom	Askew et al., 2009	- 0,030	- 9,2 %
Trametinib; p = 0,3950	Melanom	Askew et al., 2009	0,035	10,8 %

Operationalisierung: Häufigkeitsgewichtete Maßzahl zum *Eastern Cooperative Oncology Group Performance Status (ECOG PS gw.)*.

Anmerkung: Exklusive Cormier et al., 2008 [70] (vgl. detaillierte Betrachtung in Kapitel 5.5, Komplex „Standardreferenzen“).

Tabelle G51: Abweichung zur Referenz mit AWG-Übereinstimmung beim *FACT-M*; hier: Grad der Erkrankungsschwere (Quelle: Eigene Darstellung).

Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie	
			absolut	relativ
Dabrafenib (neues Anwendungsgebiet); p = 0,1579	Melanom	Askew et al., 2009	- 1,9 %	- 1,9 %

Anmerkung: Exklusive Cormier et al., 2008 [70] (vgl. detaillierte Betrachtung in Kapitel 5.5, Komplex „Standardreferenzen“).

Tabelle G52: Abweichung zur Referenz mit AWG-Übereinstimmung beim *FACT-M*;
hier: Ethnizität (Quelle: Eigene Darstellung).

Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie	
			absolut	relativ
Dabrafenib (neues Anwendungsgebiet); nicht auswertbar	Melanom	Askew et al., 2009	2,2	4,2 %
Trametinib; nicht auswertbar	Melanom	Askew et al., 2009	1,7	3,3 %

Anmerkung: Exklusive Cormier et al., 2008 [70] (vgl. detaillierte Betrachtung in Kapitel 5.5, Komplex „Standardreferenzen“).

Tabelle G53: Abweichung zur Referenz mit AWG-Übereinstimmung beim *FACT-M*;
hier: Alter (Quelle: Eigene Darstellung).

Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie	
			absolut	relativ
Dabrafenib (neues Anwendungsgebiet); nicht auswertbar	Melanom	Askew et al., 2009	65,9 %	n/a (vs. 0 %)
Trametinib; nicht auswertbar	Melanom	Askew et al., 2009	66,8 %	n/a (vs. 0 %)

Detailangabe: Askew et al., 2009 [122]: US (100 %).

Anmerkung: Exklusive Cormier et al., 2008 [70] (vgl. detaillierte Betrachtung in Kapitel 5.5, Komplex „Standardreferenzen“). n/a: nicht anwendbar.

Tabelle G54: Abweichung zur Referenz mit AWG-Übereinstimmung beim *FACT-M*;
hier: Geographische Region (Quelle: Eigene Darstellung).

G-V: Abweichungen zur Referenz mit AWG-Übereinstimmung; hier: *EORTC QLQ-MY20*

Cocks et al., 2007 [71]		
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	multiple myeloma (MM)
Relativer Anteil:		100 %
Pomalidomid	Rezidiertes oder refraktäres Multiples Myelom (rrMM)	X
Panobinostat	Rezidiertes oder refraktäres Multiples Myelom (rrMM)	X
Pomalidomid (erneute Nutzenbewertung)	Rezidiertes oder refraktäres Multiples Myelom (rrMM)	X

Tabelle G55: Entsprechung nach AWG beim *EORTC QLQ-MY20*;
hier: Cocks et al., 2007 (Quelle: Eigene Darstellung).

Dubois et al., 2006 [99]		
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	multiple myeloma (MM)
Relativer Anteil:		100 %
Panobinostat	Rezidiertes oder refraktäres Multiples Myelom (rrMM)	X

Tabelle G56: Entsprechung nach AWG beim *EORTC QLQ-MY20*;
hier: Dubois et al., 2006 (Quelle: Eigene Darstellung).

Osoba et al., 1998 [61]			
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	recurrent or metastatic breast cancer	extensive-stage small-cell lung cancer (SCLC)
Relativer Anteil:		68,9 %	31,1 %
Panobinostat	Rezidiertes oder refraktäres Multiples Myelom (rrMM)	-	-

Tabelle G57: Entsprechung nach AWG beim *EORTC QLQ-MY20*;
hier: Osoba et al., 1998 (Quelle: Eigene Darstellung).

Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie	
			absolut	relativ
Panobinostat; p = 0,0813	(rr)MM	Dubois et al., 2006	6,9 %	17,2 %

Anmerkung: Exklusive Cocks et al., 2007 [71] (vgl. detaillierte Betrachtung in Kapitel 5.5, Komplex „Standardreferenzen“). Beachte: Dubois et al., 2006 [99] untersucht mit dem *EORTC QLQ-MY24* eine abweichende Version des betreffenden Messinstrumentes. (rr)MM: (Rezidiertes oder refraktäres) Multiples Myelom.

Tabelle G58: Abweichung zur Referenz mit AWG-Übereinstimmung beim *EORTC QLQ-MY20*; hier: Geschlecht (Quelle: Eigene Darstellung).

Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie	
			absolut	relativ
Panobinostat; p < 0,0001	(rr)MM	Dubois et al., 2006	- 16,2 %	- 20,0 %

Anmerkung: Exklusive Cocks et al., 2007 [71] (vgl. detaillierte Betrachtung in Kapitel 5.5, Komplex „Standardreferenzen“). Beachte: Dubois et al., 2006 [99] untersucht mit dem *EORTC QLQ-MY24* eine abweichende Version des betreffenden Messinstrumentes. (rr)MM: (Rezidiertes oder refraktäres) Multiples Myelom.

Tabelle G59: Abweichung zur Referenz mit AWG-Übereinstimmung beim *EORTC QLQ-MY20*; hier: Ethnizität (Quelle: Eigene Darstellung).

Nutzenbewertungsverfahren [ggf. zu bewertende Studie] und p-Wert	Anwendungsgebiet (geclustert)	Referenzstudie	Abweichung zur Referenzstudie	
			absolut	relativ
Panobinostat; nicht auswertbar	(rr)MM	Dubois et al., 2006	2,1	3,5 %

Anmerkung: Exklusive Cocks et al., 2007 [71] (vgl. detaillierte Betrachtung in Kapitel 5.5, Komplex „Standardreferenzen“). Beachte: Dubois et al., 2006 [99] untersucht mit dem *EORTC QLQ-MY24* eine abweichende Version des betreffenden Messinstrumentes. (rr)MM: (Rezidiertes oder refraktäres) Multiples Myelom.

Tabelle G60: Abweichung zur Referenz mit AWG-Übereinstimmung beim *EORTC QLQ-MY20*; hier: Alter (Quelle: Eigene Darstellung).

G-VI: Abweichungen zur Referenz mit AWG-Übereinstimmung; hier: *FACT-Lym*

Hlubocky et al., 2013 [72]		
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	non-Hodgkin's lymphoma (NHL)
Relativer Anteil:		100 %
Ruxolitinib (Chronisch myeloproliferative Erkrankungen, erneute Nutzenbewertung)	Chronisch Myeloproliferative Erkrankungen	-
Idelalisib	Chronische lymphatische Leukämie (CLL); Follikuläres Lymphom (FL)	X

Tabelle G61: Entsprechung nach AWG beim *FACT-Lym*; hier: Hlubocky et al., 2013
(Quelle: Eigene Darstellung).

Carter et al., 2008 [#] [123]		
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	relapsed/refractory mantle cell lymphoma (MCL)
Relativer Anteil:		100 %
Ruxolitinib (Chronisch myeloproliferative Erkrankungen)	Chronisch Myeloproliferative Erkrankungen	-

[#] *ASH Annual Meeting Abstract* (Kongressbeitrag).

Tabelle G62: Entsprechung nach AWG beim *FACT-Lym*; hier: Carter et al., 2008
(Quelle: Eigene Darstellung).

n/a: nicht übertragbar		
Cella et al., 2005 [§]		
Nutzenbewertungsverfahren	Anwendungsgebiet (i. e. S.)	keine Angabe (nicht auffindbar)
Relativer Anteil:		100 %
Ruxolitinib (Chronisch myeloproliferative Erkrankungen)	Chronisch Myeloproliferative Erkrankungen	n/a

[§] *Abstract*.

Tabelle G63: Entsprechung nach AWG beim *FACT-Lym*; hier: Cella et al., 2005
(Quelle: Eigene Darstellung).

Anhang H: Ordnung der Skalen von Erhebungsinstrumenten ihrer Aussagekraft entsprechend

(Sub-)Skala	Referenzstudie		Komponente	Referenzstudie	(Sub-)Skala	Referenzstudie	
	Aaronson et al., 1993	Osoba et al., 1998				Rabin und de Charro, 2001	Esper et al., 1997
<i>QL</i>	belastbar	eingeschränkt belastbar	<i>EQ-5D_{utility}</i>	$EQ-5D_{utility} > EQ-5D_{VAS}$ (ungleich belastbar)	<i>FACT-P total score</i>	belastbar	belastbar
<i>PF</i>	belastbar	belastbar	<i>EQ-5D_{VAS}</i>		<i>PWB</i>	belastbar	fehlend/ unberücksichtigt
<i>RF</i>	eingeschränkt belastbar	fehlend/ unberücksichtigt			<i>FWB</i>	belastbar	fehlend/ unberücksichtigt
<i>CF</i>	belastbar	fehlend/ unberücksichtigt			<i>SWB</i>	<i>PWB, FWB > SWB, EWB</i>	fehlend/ unberücksichtigt
<i>EF</i>	belastbar	belastbar			<i>EWB</i>	(ungleich belastbar)	fehlend/ unberücksichtigt
<i>SF</i>	belastbar	eingeschränkt belastbar			<i>PCS</i>	belastbar	belastbar
					<i>TOI</i>	belastbar	belastbar

Anmerkung: Die blaue Einfärbung kennzeichnet eine vorrangige Testung psychometrischer Gütekriterien (u. a. Validität, Reliabilität). Die orange Einfärbung kennzeichnet eine vorrangige Bestimmung der klinischen Relevanz (*MID*-Schätzung).

QL: Globaler Gesundheitsstatus, *PF*: Körperliche Funktion, *RF*: Rollenfunktion, *CF*: Kognitive Funktion, *EF*: Emotionale Funktion, *SF*: Soziale Funktion (Skalen des Fragebogens *EORTC QLQ-C30*).

Tabelle H1: Katalogteil 3 zur Kategorisierung einzelner Skalen nach Belastbarkeit; hier: *EORTC QLQ-C30*
(Quelle: Eigene Darstellung).

Anmerkung: Die blaue Einfärbung kennzeichnet eine vorrangige Testung psychometrischer Gütekriterien (u. a. Validität, Reliabilität).

EQ-5D: *European Quality of Life Questionnaire 5 Dimensions, utility*: Index(-wert), *VAS*: Visuelle Analogskala.

Tabelle H2: Katalogteil 3 zur Kategorisierung einzelner Skalen nach Belastbarkeit; hier: *EQ-5D (VAS)*
(Quelle: Eigene Darstellung).

Anmerkung: Die blaue Einfärbung kennzeichnet eine vorrangige Testung psychometrischer Gütekriterien (u. a. Validität, Reliabilität). Die orange Einfärbung kennzeichnet eine vorrangige Bestimmung der klinischen Relevanz (*MID*-Schätzung).

FACT-P: *Functional Assessment of Cancer Therapy – Prostate*, *PWB*: Physisches Wohlbefinden, *FWB*: Funktionales Wohlbefinden, *SWB*: Soziales Wohlbefinden, *EWB*: Emotionales Wohlbefinden, *PCS*: Prostatakarzinom-spezifische Subskala, *TOI*: *Trial Outcome Index (PWB + FWB + PCS)*.

Tabelle H3: Katalogteil 3 zur Kategorisierung einzelner Skalen nach Belastbarkeit; hier: *FACT-P*
(Quelle: Eigene Darstellung).

(Sub-)Skala	Referenzstudie		(Sub-)Skala	Referenzstudie		(Sub-)Skala	Referenzstudie	
	Cormier et al., 2008			Cocks et al., 2007			Hlubocky et al., 2013	
<i>FACT-M total score</i>	belastbar		<i>DS</i>	belastbar		<i>FACT-Lym total score</i>	belastbar	
<i>PWB</i>	belastbar		<i>SE</i>	DS, BI, FP > SE (ungleich belastbar)		<i>PWB</i>	belastbar	
<i>FWB</i>	belastbar		<i>BI</i>	belastbar		<i>FWB</i>	belastbar	
<i>SWB</i>	PWB, FWB > SWB, EWB (ungleich belastbar)		<i>FP</i>	belastbar		<i>SWB</i>	belastbar	
<i>EWB</i>						<i>EWB</i>	belastbar	
<i>MS</i>	belastbar					<i>LymS</i>	belastbar	
						<i>TOI</i>	belastbar	

Anmerkung: Die blaue Einfärbung kennzeichnet eine vorrangige Testung psychometrischer Gütekriterien (u. a. Validität, Reliabilität).

FACT-M: Functional Assessment of Cancer Therapy – Melanoma, PWB: Physisches Wohlbefinden, FWB: Funktionales Wohlbefinden, SWB: Soziales Wohlbefinden, EWB: Emotionales Wohlbefinden, MS: Melanom-spezifische Subskala.

Tabelle H4: Katalogteil 3 zur Kategorisierung einzelner Skalen nach Belastbarkeit; hier: *FACT-M*
(Quelle: Eigene Darstellung).

Anmerkung: Die blaue Einfärbung kennzeichnet eine vorrangige Testung psychometrischer Gütekriterien (u. a. Validität, Reliabilität).

DS: Krankheitssymptome, SE: Nebenwirkungen, BI: Körperwahrnehmung, FP: Zukunftsperspektive (Subskalen der Myelom-spezifischen Skala).

Tabelle H5: Katalogteil 3 zur Kategorisierung einzelner Skalen nach Belastbarkeit; hier: *EORTC QLQ-MY20*
(Quelle: Eigene Darstellung).

Anmerkung: Die orange Einfärbung kennzeichnet eine vorrangige Bestimmung der klinischen Relevanz (*MID*-Schätzung).

FACT-Lym: Functional Assessment of Cancer Therapy – Lymphoma, PWB: Physisches Wohlbefinden, FWB: Funktionales Wohlbefinden, SWB: Soziales Wohlbefinden, EWB: Emotionales Wohlbefinden, LymS: Lymphom-spezifische Subskala, TOI: Trial Outcome Index (PWB + FWB + LymS).

Tabelle H6: Katalogteil 3 zur Kategorisierung einzelner Skalen nach Belastbarkeit; hier: *FACT-Lym*
(Quelle: Eigene Darstellung).

Literatur

1. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Cabozantinib (Nierenzellkarzinom): Addendum zum Auftrag A17-56. https://www.g-ba.de/downloads/92-975-2249/2018-04-05_AM-RL-XII_Cabozantinib_D-317_Addendum-1.pdf (letzter Zugriff am: 14.02.2021)
2. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Allgemeine Methoden: Entwurf für Version 6.0 vom 05.12.2019. https://www.iqwig.de/download/Allgemeine-Methoden_Entwurf-fuer-Version-6-0.pdf (letzter Zugriff am: 14.02.2021)
3. Gesetz zur Neuordnung des Arzneimittelmarktes in der gesetzlichen Krankenversicherung (Arzneimittelmarktneuordnungsgesetz - AMNOG). Bundesgesetzblatt, Jahrgang 2010, Teil I Nr. 67, 2262-2277
4. Gesetzentwurf der Bundesregierung. Entwurf eines Gesetzes zur Neuordnung des Arzneimittelmarktes in der gesetzlichen Krankenversicherung (Arzneimittelmarktneuordnungsgesetz – AMNOG). Drucksache 17/3116, 17. Wahlperiode, 01.10.2010
5. Verordnung über die Nutzenbewertung von Arzneimitteln nach § 35a Absatz 1 SGB V für Erstattungsvereinbarungen nach § 130b SGB V. Arzneimittel-Nutzenbewertungsverordnung - AM-NutzenV in der Fassung vom 28. Dezember 2010 (BGBl. I S. 2324), zuletzt geändert am 9. August 2019 (BGBl. I S. 1202)
6. Gemeinsamer Bundesausschuss (G-BA). Beschluss des Gemeinsamen Bundesausschusses über die Änderung der Verfahrensordnung: Änderung im 5. Kapitel – Verfahren zur Forderung einer anwendungsbegleitenden Datenerhebung nach § 35a Absatz 3b SGB V vom 16. Juli 2020. https://www.g-ba.de/downloads/39-261-4402/2020-07-16_VerfO_Kapitel-5.pdf (letzter Zugriff am: 14.02.2021)
7. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Allgemeine Methoden: Version 6.0 vom 05.11.2020. https://www.iqwig.de/methoden/allgemeine-methoden_version-6-0.pdf?rev=144030 (letzter Zugriff am: 14.02.2021)
8. Bullinger M. Lebensqualität: Ein neues Thema in der Medizin? Zentralblatt für Gynäkologie 2002; 124: 153–156
9. Howard-Jones N. The World Health Organization in Historical Perspective. Perspectives in Biology and Medicine 1981; 24: 467–482
10. Walters SJ. Quality of Life Outcomes in Clinical Trials and Health-Care Evaluation. A Practical Guide to Analysis and Interpretation. Chichester: Wiley, 2009
11. Schumacher, J, Klaiberg, A, Brähler, E (Hrsg.). Diagnostische Verfahren zu Lebensqualität und Wohlbefinden. Göttingen, Bern, Toronto, Seattle: Hogrefe Verlag für Psychologie, 2003

12. Valderas JM, Alonso J. Patient reported outcome measures: a model-based classification system for research and clinical practice. *Quality of Life Research* 2008; 17: 1125–1135
13. Bullinger M. Das Konzept der Lebensqualität in der Medizin – Entwicklung und heutiger Stellenwert. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 2014; 108: 97–103
14. The WHOQOL Group. The World Health Organization quality of life assessment (WHOQOL): Position paper from the World Health Organization. *Social Science & Medicine* 1995; 41: 1403–1409
15. Brettschneider C, Lühmann D, Raspe H. Der Stellenwert von Patient Reported Outcomes (PRO) im Kontext von Health Technology Assessment (HTA). Köln: Deutsches Institut für Medizinische Dokumentation und Information (DIMDI), 2011
16. Greiner W, Klose K. Lebensqualitätsbewertung und Utilities in der Gesundheitsökonomie. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 2014; 108: 120–125
17. Schöffski O. Lebensqualität als Ergebnisparameter in gesundheitsökonomischen Studien. In: Schöffski O, Graf von der Schulenburg J-M (Hrsg.). *Gesundheitsökonomische Evaluationen*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012: 327–340
18. Fayers PM, Machin D. *Quality of Life: The assessment, analysis and reporting of patient-reported outcomes*. Chichester, West Sussex, UK, Hoboken, NJ: John Wiley & Sons Inc, 2016
19. Patrick DL, Deyo RA. Generic and Disease-Specific Measures in Assessing Health Status and Quality of Life. *Medical Care* 1989; 27: 217-32
20. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: Ascertaining the minimal clinically important difference. *Controlled Clinical Trials* 1989; 10: 407–415
21. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to Explain the Clinical Significance of Health Status Measures. *Mayo Clinic Proceedings* 2002; 77: 371–383
22. Gemeinsamer Bundesausschuss (G-BA). Verfahrensordnung des Gemeinsamen Bundesausschusses in der Fassung vom 18. Dezember 2008, zuletzt geändert am 16. Juli 2020. https://www.g-ba.de/downloads/62-492-2323/VerfO_2020-07-16_iK-2020-12-16.pdf (letzter Zugriff am: 14.02.2021)
23. Kohlmann T. Messung von Lebensqualität: So einfach wie möglich, so differenziert wie nötig. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 2014; 108: 104–110
24. Bullinger M, Petersen C, Mehnert A. Erfassung der Lebensqualität in der Onkologie. In: Siewert JR (Hrsg.). *Onkologische Chirurgie*. Heidelberg: Springer Medizin Verl., 2006: 323–332

25. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *Journal of Clinical Epidemiology* 2003; 56: 395–407
26. Klakow-Franck R. Die Bedeutung von Lebensqualität für die Arbeit des Gemeinsamen Bundesausschusses. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 2014; 108: 151–156
27. European Medicines Agency (EMA). Reflection Paper on the use of patient reported outcome (PRO) measures in oncology studies (EMA/CHMP/292464/2014, 17 June 2014). https://www.ema.europa.eu/en/documents/scientific-guideline/draft-reflection-paper-use-patient-reported-outcome-pro-measures-oncology-studies_en.pdf (letzter Zugriff am: 14.02.2021)
28. European Medicines Agency (EMA). Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man: The use of patient-reported outcome (PRO) measures in oncology studies (EMA/CHMP/292464/2014, 1 April 2016). https://www.ema.europa.eu/en/documents/other/appendix-2-guideline-evaluation-anticancer-medicinal-products-man_en.pdf (letzter Zugriff am: 14.02.2021)
29. Ludwig W-D, Schwabe U. Onkologika. In: Schwabe U, Ludwig W-D (Hrsg.). *Arzneiverordnungs-Report 2020*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2020: 671–732
30. Lohmüller J, Bauckmann J, Schröder M, Niepraschk-Von Dollen K, Telschow C, Sollmann S, Zawinell A, Hoberg M. *Der GKV-Arzneimittelmarkt*. Berlin: Wissenschaftliches Institut der AOK (WiDO), 2020
31. Witte J, Greiner W, Gensorowsky D. *AMNOG-Report 2019*. Heidelberg, Neckar: medhochzwei Verlag, 2019
32. Ruof J, Knoerzer D, Dünne A-A, Dintsios C-M, Staab T, Schwartz FW. Analysis of endpoints used in marketing authorisations versus value assessments of oncology medicines in Germany. *Health Policy* 2014; 118: 242–254
33. Schandelmaier S, Conen K, Elm E von, You JJ, Blümle A, Tomonaga Y, Amstutz A, Briel M, Kasenda B. Planning and reporting of quality-of-life outcomes in cancer trials. *Annals of Oncology* 2015; 26: 1966–1973
34. Bender H, Dintsios C-M. Gesundheitsbezogene Lebensqualität im Rahmen der frühen Nutzenbewertung von Arzneimitteln nach § 35a SGB V: Ein Endpunkt mit vielen Herausforderungen für alle beteiligten Akteure. *Gesundheitswesen* 2018; 80: 132–143
35. Lohrberg D, Augustin M, Blome C. The definition and role of quality of life in Germany's early assessment of drug benefit: a qualitative approach. *Quality of Life Research* 2016; 25: 447–455
36. Blome C, Augustin M, Metin H, Lohrberg D. Four years of early benefit assessment of new drugs in Germany: a qualitative study on methodological requirements for quality of life data. *The European Journal of Health Economics* 2017; 18: 181–193

37. Bullinger M, Blome C, Sommer R, Lohrberg D, Augustin M. Gesundheitsbezogene Lebensqualität – ein zentraler patientenrelevanter Endpunkt in der Nutzenbewertung medizinischer Maßnahmen. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* 2015; 58: 283–290
38. Wyrwich KW, Bullinger M, Aaronson N, Hays RD, Patrick DL, Symonds T. Estimating clinically significant differences in quality of life outcomes. *Quality of Life Research* 2005; 14: 285–295
39. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology* 2008; 61: 102–109
40. Coon CD, Cappelleri JC. Interpreting Change in Scores on Patient-Reported Outcome Instruments. *Therapeutic Innovation & Regulatory Science* 2016; 50: 22–29
41. Jayadevappa R, Cook R, Chhatre S. Minimal important difference to infer changes in health-related quality of life – a systematic review. *Journal of Clinical Epidemiology* 2017; 89: 188–198
42. Ousmen A, Touraine C, Deliu N, Cottone F, Bonnetain F, Efficace F, Brédart A, Mollevi C, Anota A. Distribution- and anchor-based methods to determine the minimally important difference on patient-reported outcome questionnaires in oncology: a structured review. *Health and Quality of Life Outcomes* 2018; 16: 228
43. Nordin Å, Taft C, Lundgren-Nilsson Å, Dencker A. Minimal important differences for fatigue patient reported outcome measures – a systematic review. *BMC Medical Research Methodology* 2016; 16: 62
44. Johnston BC, Ebrahim S, Carrasco-Labra A, Furukawa TA, Patrick DL, Crawford MW, Hemmelgarn BR, Schunemann HJ, Guyatt GH, Nesrallah G. Minimally important difference estimates and methods: a protocol. *BMJ Open* 2015; 5: e007953
45. Carrasco-Labra A, Devji T, Qasim A, Phillips M, Devasenapathy N, Zeraatkar D, Bhatt M, Jin X, Brignardello-Petersen R, Urquhart O, Faroutan F, Schandelmaier S, Pardo-Hernandez H, Vernooij RW, Huang W, Rizwan Y, Lytvyn L, Siemieniuk R, Johnston BC, Ebrahim S, Furukawa TA, Patrick DL, Schünemann HJ, Nesrallah G, Guyatt G. Interpretation of patient-reported outcome measures: an inventory of over 3000 minimally important difference estimates and an assessment of their credibility. Abstracts of the 25th Cochrane Colloquium, Edinburgh, UK. *Cochrane Database of Systematic Reviews* 2018; (9 Suppl 1)
46. Devji T, Carrasco-Labra A, Qasim A, Phillips M, Johnston BC, Devasenapathy N, Zeraatkar D, Bhatt M, Jin X, Brignardello-Petersen R, Urquhart O, Faroutan F, Schandelmaier S, Pardo-Hernandez H, Vernooij RW, Huang H, Rizwan Y, Siemieniuk R, Lytvyn L, Patrick DL, Ebrahim S, Furukawa T, Nesrallah G, Schünemann HJ, Bhandari M, Thabane L, Guyatt GH. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. *BMJ* 2020; 369: m1714

47. Chang EM, Gillespie EF, Shaverdian N. Truthfulness in patient-reported outcomes: factors affecting patients' responses and impact on data quality. *Patient Related Outcome Measures* 2019; 10: 171–186
48. King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Review of Pharmacoeconomics & Outcomes Research* 2011; 11: 171–184
49. Coon CD, Cook KF. Moving from significance to real-world meaning: methods for interpreting change in clinical outcome assessment scores. *Quality of Life Research* 2018; 27: 33–40
50. Blum SI, Ahmed S, Flood E, Oort FJ, Schwartz CE. Introduction to special section: measuring what matters. *Quality of Life Research* 2018; 27: 1–3
51. Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, Schwartz C, Revicki DA, Moinpour CM, McLeod LD, Lyons JC, Lenderking WR, Hinds PS, Hays RD, Greenhalgh J, Gershon R, Feeny D, Fayers PM, Cella D, Brundage M, Ahmed S, Aaronson NK, Butt Z. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Quality of Life Research* 2013; 22: 1889–1905
52. Calvert M, Kyte D, Mercieca-Bebber R, Slade A, Chan A-W, King MT, Hunn A, Bottomley A, Regnault A, Ells C, O'Connor D, Revicki D, Patrick D, Altman D, Basch E, Velikova G, Price G, Draper H, Blazeby J, Scott J, Coast J, Norquist J, Brown J, Haywood K, Johnson LL, Campbell L, Frank L, Hildebrand M von, Brundage M, Palmer M, Klutz P, Stephens R, Golub RM, Mitchell S, Groves T. Guidelines for Inclusion of Patient-Reported Outcomes in Clinical Trial Protocols: The SPIRIT-PRO Extension. *JAMA* 2018; 319: 483–494
53. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD. Reporting of Patient-Reported Outcomes in Randomized Trials: The CONSORT PRO Extension. *JAMA* 2013; 309: 814–822
54. Gemeinsamer Bundesausschuss (G-BA). Verfahren der Nutzenbewertung nach § 35a SGB V: Übersicht. <https://www.g-ba.de/bewertungsverfahren/nutzenbewertung/> (letzter Zugriff am: 14.02.2021)
55. Gemeinsamer Bundesausschuss (G-BA). Dokumentvorlage: Modul 1 - Zusammenfassung der Aussagen im Dossier (Anlage II.2 zum 5. Kapitel der Verfahrensordnung des Gemeinsamen Bundesausschusses, Version vom 20.02.2020). https://www.g-ba.de/downloads/17-98-4822/2020-02-20_An12_2_Modul1.pdf (letzter Zugriff am: 14.02.2021)
56. Gemeinsamer Bundesausschuss (G-BA). Dokumentvorlage: Modul 2 - Allgemeine Angaben zum Arzneimittel, zugelassene Anwendungsgebiete (Anlage II.4 zum 5. Kapitel der Verfahrensordnung des Gemeinsamen Bundesausschusses, Version vom 16.03.2018). https://www.g-ba.de/downloads/17-98-4823/2018-03-16_An12_4_Modul2.pdf (letzter Zugriff am: 14.02.2021)

57. Gemeinsamer Bundesausschuss (G-BA). Dokumentvorlage: Modul 3 - Zweckmäßige Vergleichstherapie, Anzahl der Patienten mit therapeutisch bedeutsamem Zusatznutzen, Kosten der Therapie für die GKV, Anforderungen an eine qualitätsgesicherte Anwendung (Anlage II.5 zum 5. Kapitel der Verfahrensordnung des Gemeinsamen Bundesausschusses, Version vom 16.03.2018/16.08.2018). https://www.g-ba.de/downloads/17-98-4824/2018-03-16_2018-08-16_An12_5_Modul3.pdf (letzter Zugriff am: 14.02.2021)
58. Gemeinsamer Bundesausschuss (G-BA). Dokumentvorlage: Modul 4 - Medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen (Anlage II.6 zum 5. Kapitel der Verfahrensordnung des Gemeinsamen Bundesausschusses, Version vom 21.02.2019). https://www.g-ba.de/downloads/17-98-4825/2019-02-21_An12_6_Modul4.pdf (letzter Zugriff am: 14.02.2021)
59. Gemeinsamer Bundesausschuss (G-BA). Dokumentvorlage: Erstellung und Einreichung eines Dossiers zur Nutzenbewertung gemäß § 35a SGB V (Anlage II.1 zum 5. Kapitel der Verfahrensordnung des Gemeinsamen Bundesausschusses, Version vom 20.02.2020). https://www.g-ba.de/downloads/17-98-4827/2020-02-20_An12_1_Erstellung-Einreichung-Dossier.pdf (letzter Zugriff am: 14.02.2021)
60. Gemeinsamer Bundesausschuss (G-BA). Nutzenbewertungsverfahren zum Wirkstoff Crizotinib (neues Anwendungsgebiet: nicht-kleinzelliges Lungenkarzinom, ALK+, Erstlinie). <https://www.g-ba.de/bewertungsverfahren/nutzenbewertung/209/> (letzter Zugriff am: 14.02.2021)
61. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. *Journal of Clinical Oncology* 1998; 16: 139–144
62. Broderick JE, May M, Schwartz JE, Li M, Mejia A, Nocera L, Kolatkar A, Ueno NT, Yennu S, Lee JSH, Hanlon SE, Cozzens Philips FA, Shahabi C, Kuhn P, Nieva J. Patient reported outcomes can improve performance status assessment: a pilot study. *Journal of Patient-Reported Outcomes* 2019; 3: 41
63. Center for International Blood and Marrow Transplant Research (CIBMTR). Forms Manual: Appendix L - Karnofsky/Lansky Performance Status (A00428 revision 1). <https://www.cibmtr.org/DataManagement/TrainingReference/Manuals/DataManagement/Documents/appendix-l.pdf> (letzter Zugriff am: 14.02.2021)
64. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, Filiberti A, Flechtner H, Fleishman SB, Haes JC de. The European Organization for Research and Treatment of Cancer QLQ-C30: A Quality-of-Life Instrument for Use in International Clinical Trials in Oncology. *Journal of the National Cancer Institute* 1993; 85: 365–376
65. Gemeinsamer Bundesausschuss (G-BA). Nutzenbewertungsverfahren zum Wirkstoff Decitabin (Akute myeloische Leukämie). <https://www.g-ba.de/bewertungsverfahren/nutzenbewertung/42/> (letzter Zugriff am: 14.02.2021)

66. Rabin R, Charro F de. EQ-5D: a measure of health status from the EuroQol Group. *Annals of Medicine* 2001; 33: 337–343
67. Hurst NP, Kind P, Ruta D, Hunter M, Stubbings A. Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). *British Journal of Rheumatology* 1997; 36: 551–559
68. Esper P, Mo F, Chodak G, Sinner M, Cella D, Pienta KJ. Measuring quality of life in men with prostate cancer using the Functional Assessment of Cancer Therapy-prostate instrument. *Urology* 1997; 50: 920–928
69. Cella D, Nichol MB, Eton D, Nelson JB, Mulani P. Estimating Clinically Meaningful Changes for the Functional Assessment of Cancer Therapy-Prostate: Results from a Clinical Trial of Patients with Metastatic Hormone-Refractory Prostate Cancer. *Value in Health* 2009; 12: 124–129
70. Cormier JN, Ross MI, Gershengwald JE, Lee JE, Mansfield PF, Camacho LH, Kim K, Webster K, Cella D, Palmer JL. Prospective Assessment of the Reliability, Validity, and Sensitivity to Change of the Functional Assessment of Cancer Therapy-Melanoma Questionnaire. *Cancer* 2008; 112: 2249–2257
71. Cocks K, Cohen D, Wisløff F, Sezer O, Lee S, Hippe E, Gimsing P, Turesson I, Hajek R, Smith A, Graham L, Phillips A, Stead M, Velikova G, Brown J. An international field study of the reliability and validity of a disease-specific questionnaire module (the QLQ-MY20) in assessing the quality of life of patients with multiple myeloma. *European Journal of Cancer* 2007; 43: 1670–1678
72. Hlubocky FJ, Webster K, Cashy J, Beaumont J, Cella D. The Development and Validation of a Measure of Health-Related Quality of Life for Non-Hodgkin's Lymphoma: The Functional Assessment of Cancer Therapy-Lymphoma (FACT-Lym). *Lymphoma* 2013; 2013: 1–9
73. Gemeinsamer Bundesausschuss (G-BA). Nutzenbewertungsverfahren zum Wirkstoff Pomalidomid (Überschreitung 50 Mio € Grenze: Multiples Myelom, mind. 2 Vortherapien, Kombination mit Dexamethason). <https://www.g-ba.de/bewertungsverfahren/nutzenbewertung/194/> (letzter Zugriff am: 14.02.2021)
74. Gemeinsamer Bundesausschuss (G-BA). Nutzenbewertungsverfahren zum Wirkstoff Ruxolitinib. <https://www.g-ba.de/bewertungsverfahren/nutzenbewertung/38/> (letzter Zugriff am: 14.02.2021)
75. Gemeinsamer Bundesausschuss (G-BA). Nutzenbewertungsverfahren zum Wirkstoff Ruxolitinib (Überschreitung 50 Mio € Grenze: Myelofibrose). <https://www.g-ba.de/bewertungsverfahren/nutzenbewertung/116/> (letzter Zugriff am: 14.02.2021)
76. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Hoboken: Taylor and Francis, 2013
77. Gemeinsamer Bundesausschuss (G-BA). Nutzenbewertungsverfahren zum Wirkstoff Idelalisib (Chronische lymphatische Leukämie, Follikuläres Lymphom). <https://www.g-ba.de/bewertungsverfahren/nutzenbewertung/142/> (letzter Zugriff am: 14.02.2021)

78. Greiner W, Witte J. AMNOG-Report 2016. Heidelberg: Medhochzwei Verlag GmbH, 2016
79. European Medicines Agency (EMA). ICH Topic E 9. Statistical Principles for Clinical Trials (CPMP/ICH/363/96, September 1998). https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf (letzter Zugriff am: 14.02.2021)
80. Food and Drug Administration (FDA). Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims (FDA-2006-D-0362, December 2009). <https://www.fda.gov/media/77832/download> (letzter Zugriff am: 14.02.2021)
81. European Medicines Agency (EMA). Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products (EMA/CHMP/EWP/139391/2004, 27 July 2005). https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-regulatory-guidance-use-healthrelated-quality-life-hrql-measures-evaluation_en.pdf (letzter Zugriff am: 14.02.2021)
82. Niehaus I, Dintsios C-M. Confirmatory versus explorative endpoint analysis: Decision-making on the basis of evidence available from market authorization and early benefit assessment for oncology drugs. *Health Policy* 2018; 122: 599–606
83. Thomas S, Schürmann C, Abbas S, Schwalm A. Lebensqualität in der Nutzenbewertung: Die Kunst des Lösbaren. In: Andreas S (Hrsg.). AMNOG-Report 2019. Heidelberg, Neckar: medhochzwei Verlag, 2019: 95–115
84. Isbary G, Staab TR, Amelung VE, Dintsios C-M, Iking-Konert C, Nesurini SM, Walter M, Ruof J. Effect of Crossover in Oncology Clinical Trials on Evidence Levels in Early Benefit Assessment in Germany. *Value in Health* 2018; 21: 698–706
85. Gemeinsamer Bundesausschuss (G-BA). Nutzenbewertungsverfahren zum Wirkstoff Regorafenib. <https://www.g-ba.de/bewertungsverfahren/nutzenbewertung/82/> (letzter Zugriff am: 14.02.2021)
86. Hecken J. Der Stellenwert von Daten zur Lebensqualität bei der Bemessung des Zusatznutzens durch den G-BA. In: Andreas S (Hrsg.). AMNOG-Report 2019. Heidelberg, Neckar: medhochzwei Verlag, 2019: 84–94
87. Gemeinsame Publikation des Zentrums für Krebsregisterdaten und der Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. Krebs in Deutschland für 2015/2016. Berlin, 2019
88. Werner S, Lechterbeck L, Rasch A, Merkesdal S, Ruof J. Untersuchung der Akzeptanz und der Ablehnungsgründe indirekter Vergleiche in IQWiG-Nutzenbewertungen. *Gesundheitsökonomie & Qualitätsmanagement* 2020; 25: 24–36
89. Chan EKH, Edwards TC, Haywood K, Mikles SP, Newton L. Implementing patient-reported outcome measures in clinical practice: a companion guide to the ISOQOL user's guide. *Quality of Life Research* 2019; 28: 621–627

90. Gemeinsamer Bundesausschuss (G-BA). Nutzenbewertungsverfahren zum Wirkstoff Ibrutinib. <https://www.g-ba.de/bewertungsverfahren/nutzenbewertung/145/> (letzter Zugriff am: 14.02.2021)
91. Gemeinsamer Bundesausschuss (G-BA). Nutzenbewertungsverfahren zum Wirkstoff Nivolumab (neues Anwendungsgebiet: nicht-kleinzelliges Lungenkarzinom, Plattenepithelhistologie, nach vorheriger Chemotherapie). <https://www.g-ba.de/bewertungsverfahren/nutzenbewertung/186/> (letzter Zugriff am: 14.02.2021)
92. Gemeinsamer Bundesausschuss (G-BA). Nutzenbewertungsverfahren zum Wirkstoff Ramucirumab. <https://www.g-ba.de/bewertungsverfahren/nutzenbewertung/158/> (letzter Zugriff am: 14.02.2021)
93. Gemeinsamer Bundesausschuss (G-BA). Nutzenbewertungsverfahren zum Wirkstoff Abirateronacetat (neues Anwendungsgebiet: Prostatakarzinom, nach Androgenentzugstherapie, keine Indikation für Chemotherapie). <https://www.g-ba.de/bewertungsverfahren/nutzenbewertung/60/> (letzter Zugriff am: 14.02.2021)
94. McGlothlin AE, Lewis RJ. Minimal Clinically Important Difference: Defining What Really Matters to Patients. *JAMA* 2014; 312: 1342–1343
95. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden 6.0 vom 05.11.2020. https://www.iqwig.de/methoden/allgemeine-methoden_dwa-entwurf-fuer-version-6-0_v1-0.pdf?rev=144030 (letzter Zugriff am: 14.02.2021)
96. Bjordal K, Graeff A de, Fayers PM, Hammerlid E, van Pottelsberghe C, Curran D, Ahlner-Elmqvist M, Maher EJ, Meyza JW, Brédart A, Söderholm AL, Arraras JJ, Feine JS, Abendstein H, Morton RP, Pignon T, Huguenin P, Bottomly A, Kaasa S. A 12 country field study of the EORTC QLQ-C30 (version 3.0) and the head and neck cancer specific module (EORTC QLQ-H&N35) in head and neck patients. *European Journal of Cancer* 2000; 36: 1796–1807
97. Braun DP, Gupta D, Staren ED. Quality of life assessment as a predictor of survival in non-small cell lung cancer. *BMC Cancer* 2011; 11: 353
98. Delforge M, Dhawan R, Robinson D, Meunier J, Regnault A, Esseltine D-L, Cakana A, van de Velde H, Richardson PG, San Miguel JF. Health-related quality of life in elderly, newly diagnosed multiple myeloma patients treated with VMP vs. MP: results from the VISTA trial. *European Journal of Haematology* 2012; 89: 16–27
99. Dubois D, Dhawan R, van de Velde H, Esseltine D, Gupta S, Viala M, La Loge C de. Descriptive and Prognostic Value of Patient-Reported Outcomes: The Bortezomib Experience in Relapsed and Refractory Multiple Myeloma. *Journal of Clinical Oncology* 2006; 24: 976–982
100. Groenvold M, Klee MC, Sprangers MAG, Aaronson NK. Validation of the EORTC QLQ-C30 quality of life questionnaire through combined qualitative and quantitative assessment of patient-observer agreement. *Journal of Clinical Epidemiology* 1997; 50: 441–450

101. Hjermstad MJ, Fossa SD, Bjordal K, Kaasa S. Test/retest study of the European Organization for Research and Treatment of Cancer Core Quality-of-Life Questionnaire. *Journal of Clinical Oncology* 1995; 13: 1249–1254
102. Kvam AK, Fayers PM, Wisloff F. Responsiveness and minimal important score differences in quality-of-life questionnaires: a comparison of the EORTC QLQ-C30 cancer-specific questionnaire to the generic utility questionnaires EQ-5D and 15D in patients with multiple myeloma. *European Journal of Haematology* 2011; 87: 330–337
103. McLachlan S-A, Devins GM, Goodwin PJ. Validation of the european organization for research and treatment of cancer quality of life questionnaire (QLQ-C30) as a measure of psychosocial function in breast cancer patients. *European Journal of Cancer* 1998; 34: 510–517
104. Montazeri A, Milroy R, Hole D, McEwen J, Gillis CR. Quality of life in lung cancer patients: As an important prognostic factor. *Lung Cancer* 2001; 31: 233–240
105. Osoba D, Aaronson N, Zee B, Sprangers M, te Velde A. Modification of the EORTC QLQ-C30 (version 2.0) based on content validity and reliability testing in large samples of patients with cancer. *Quality of Life Research* 1997; 6: 103–108
106. Parmar V, Badwe RA, Hawaldar R, Rayabhattachanavar S, Varghese A, Sharma R, Mitra I. Validation of EORTC quality-of-life questionnaire in Indian women with operable breast cancer. *The National Medical Journal of India* 2005; 18: 172–177
107. Sadighi S, Mohagheghi MA, Montazeri A, Sadighi Z. Quality of life in patients with advanced gastric cancer: a randomized trial comparing docetaxel, cisplatin, 5-FU (TCF) with epirubicin, cisplatin, 5-FU (ECF). *BMC Cancer* 2006; 6: 274
108. Teckle P, Peacock S, McTaggart-Cowan H, van der Hoek K, Chia S, Melosky B, Gelmon K. The ability of cancer-specific and generic preference-based instruments to discriminate across clinical and self-reported measures of cancer severities. *Health and Quality of Life Outcomes* 2011; 9: 106
109. Wisløff F, Eika S, Hippe E, Hjorth M, Holmberg E, Kaasa S, Palva I, Westin J. Measurement of health-related quality of life in multiple myeloma. *British Journal of Haematology* 1996; 92: 604–613
110. Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking Clinical Relevance and Statistical Significance in Evaluating Intra-Individual Changes in Health-Related Quality of Life. *Medical Care* 1999; 37: 469–478
111. Zeng L, Chow E, Zhang L, Tseng L-M, Hou M-F, Fairchild A, Vassiliou V, Jesus-Garcia R, Alm El-Din MA, Kumar A, Forges F, Chie W-C, Bedard G, Bottomley A. An international prospective study establishing minimal clinically important differences in the EORTC QLQ-BM22 and QLQ-C30 in cancer patients with bone metastases. *Supportive Care in Cancer* 2012; 20: 3307–3313

112. Akakura K, Matsuzaki K, Kobayashi T, Kitoh H, Mizoguchi K, Tomikawa G, Takura T, Kawabuchi K. Evaluation of utility index of quality of life (QOL) in prostate cancer patients: comparison of QOL utility index EuroQol-5D (EQ-5D) and visual analogue scale (VAS) with health-related QOL questionnaires SF-36 and EPIC. *Nihon Hinyokika Gakkai zasshi. The Japanese Journal of Urology* 2011; 102: 9–13
113. Brazier J, Jones N, Kind P. Testing the validity of the Euroqol and comparing it with the SF-36 health survey questionnaire. *Quality of Life Research* 1993; 2: 169–180
114. Brooks RG, Jendteg S, Lindgren B, Persson U, Björk S. EuroQol©: health-related quality of life measurement. Results of the Swedish questionnaire exercise. *Health Policy* 1991; 18: 37–48
115. Boer AGEM de, van Lanschot JJB, Stalmeier PFM, van Sandick JW, Hulscher JBF, Haes JCJM de, Sprangers MAG. Is a single-item visual analogue scale as valid, reliable and responsive as multi-item scales in measuring quality of life? *Quality of Life Research* 2004; 13: 311–320
116. Greiner W, Claes C, Busschbach JJV, Schulenburg J-MG von der. Validating the EQ-5D with time trade off for the German population. *The European Journal of Health Economics* 2005; 6: 124–130
117. Johnson JA, Coons SJ, Ergo A, Szava-Kovats G. Valuation of EuroQOL (EQ-5D) Health States in an Adult US Sample. *PharmacoEconomics* 1998; 13: 421–433
118. Pickard AS, Neary MP, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. *Health and Quality of Life Outcomes* 2007; 5: 70
119. Schuette W, Tesch H, Büttner H, Krause T, Soldatenkova V, Stoffregen C. Second-line Treatment of Stage III/IV Non-Small-Cell Lung Cancer (NSCLC) with pemetrexed in routine clinical practice: Evaluation of performance status and health-related quality of life. *BMC Cancer* 2012; 12: 14
120. Carducci MA, Nelson JB, Vogelzang NJ, Mulani P. Determination of clinically meaningful change (CMC) for functional assessment of cancer therapy-prostate (FACT-P). *Journal of Clinical Oncology* 2005; 23: 8077
121. Webster K, Odom L, Peterman A, Lent L CD. The Functional Assessment of Chronic Illness Therapy (FACIT) measurement system: Validation of version 4 of the core questionnaire. *Quality of Life Research* 1999; 8: 604
122. Askew RL, Xing Y, Palmer JL, Cella D, Moye LA, Cormier JN. Evaluating Minimal Important Differences for the FACT-Melanoma Quality of Life Questionnaire. *Value in Health* 2009; 12: 1144–1150
123. Carter GC, Liepa AM, Zimmermann AH, Morschhauser F. Validation of the Functional Assessment of Cancer Therapy–Lymphoma (FACT-LYM) in Patients with Relapsed/Refractory Mantle Cell Lymphoma. *Blood* 2008; 112: 2376

Danksagung

Gelegenheit, Inspiration und Motivation waren die Grundvoraussetzungen für das Anfertigen dieser Arbeit. All jenen, die Anteil daran haben, möchte ich Danke sagen!

Im Besonderen gilt mein Dank Frau Prof. Andrea Icks und Herrn Dr. Charalabos-Markos Dintsios, die gemeinsam meine Arbeit betreuten und mir kritisch-konstruktiv und fachlich fundiert zur Seite standen.

Auch möchte ich mich bei meiner Familie bedanken, ohne deren fortwährende Unterstützung und gelegentliche Nachsicht das Abfassen der Dissertation nicht möglich gewesen wäre.

Holger Bender

Nieheim, 22. Februar 2021