

Integration der Patientensicht im Rahmen des Health Technology Assessments

Analyse des Berücksichtigungsstandes von Daten zur
gesundheitsbezogenen Lebensqualität (Health-Related
Quality of Life, HRQoL) im Verfahren der frühen
Nutzenbewertung in Deutschland

Isabelle Schatz, Thomas Hammerschmidt, Axel Olaf Kern

© COPYRIGHT 2016 SCHATZ, I; KERN, A. O.

CENTER FOR HEALTH AND SOCIAL POLICY (CHSP)
At Weingarten University

*Wir danken
Herrn Dr. Matthias Perleth, Leiter der Abteilung Fachberatung
Medizin in der Geschäftsstelle des G-BA, für seine wertvollen
Anregungen und Anmerkungen, die diese Analyse sehr bereichert
haben.*

Inhaltsverzeichnis

1	Einleitung	1
2	Historische und konzeptionelle Grundlagen der HRQoL	3
2.1	Patientenberichtete Studienendpunkte (Patient Reported Outcomes, PRO)	5
2.2	Lebensqualität (Quality of Life, QoL).....	9
2.3	Gesundheitsbezogene Lebensqualität (Health Related Quality of Life, HRQoL)	13
3	Messung und Validierung von HRQoL	17
3.1	Nutzentheoretische und psychometrische Ansätze	17
3.2	Profil- und Indexinstrumente	19
3.3	Generische und spezifische Messinstrumente	21
3.4	Erhebungsverfahren, -zeitpunkt und Recall-Periode.....	23
3.5	Gütekriterien.....	24
4	HRQoL als patientenrelevanter Endpunkt in der frühen Nutzenbewertung gemäß §35a SGB V	27
5	Fragestellung	32
6	Methodik	34
6.1	Informationsbeschaffung und -aufbereitung	34
6.2	Auswertung der Daten	37
7	Ergebnisdarstellung	40
7.1	Studiencharakteristika und Messinstrumentarium.....	41
7.2	Berücksichtigungsstand von HRQoL-Daten in der Nutzenbewertung.....	43
7.3	Das Kategoriensystem der Gründe für die Nichtberücksichtigung und das Versagen eines HRQoL-spezifischen ZN.....	45
8	Diskussion	47
8.1	Limitationen der Analyse	47
8.2	Erhebungs-, Einreichungsverhalten und Berücksichtigungsstand von HRQoL-Daten.....	48
8.3	Best Practice der Erhebung und Einreichung von HRQoL-Daten	51
8.4	Mängel in der Erhebung und Einreichung von HRQoL-Daten.....	53
8.4.1	Mängel im Studiendesign	53
8.4.2	Mängel des Erhebungsinstruments	55
8.4.3	Mängel in der Auswertung.....	57
8.4.4	Mangelnde Signifikanz/Relevanz der Daten.....	59
8.4.5	Formale Mängel.....	62

9	Zusammenfassung und Ausblick.....	64
	Anhang A	66
	Anhang B	67
	Anhang C	68
	Anhang D.....	70
	Literaturverzeichnis	71

Abbildungsverzeichnis

Abbildung 1: Modell nach Wilson und Cleary (1995)	7
Abbildung 2: PRO-Klassifikationsschema	8
Abbildung 3: Dimensionen der QoL auf Bevölkerungsebene	11
Abbildung 4: Konzeptionelles Modell der HRQoL	15
Abbildung 5: Operationalisierung der HRQoL.....	18
Abbildung 6: Ablaufmodell der zusammenfassenden Inhaltsanalyse	39
Abbildung 7: Studiencharakteristika - Studienkategorie (n = 60).....	41
Abbildung 8: Studiencharakteristika - Endpunktebene (n = 60).....	42
Abbildung 9: Prozentualer Anteil der verwendeten HRQoL-Messinstrumente in den Verfahren (n = 28)	42
Abbildung 10: Berücksichtigungsstand von HRQoL-Daten in der Nutzenbewertung (n=28)	44
Abbildung 11: Kategoriensystem der Gründe für die Nichtberücksichtigung und das Versagen eines HRQoL-spezifischen ZN (n = 28)	45

Tabellenverzeichnis

Tabelle 1: Domänen und Konzepte der HRQoL.....	16
Tabelle 2: IQWiG-Operationalisierung patientenrelevanter Endpunkte gemäß AM- NutzenV	30

Abkürzungsverzeichnis

Abkürzung	Bedeutung
AMNOG	Arzneimittelmarktneuordnungsgesetz
AM-NutzenV	Arzneimittel-Nutzenbewertungsverordnung
AVWG	Arzneimittelversorgungs-Wirtschaftlichkeitsgesetz
AWG	Anwendungsgebiet
BAnz	Bundesanzeiger
BMG	Bundesministerium für Gesundheit
BSC	Best Supportive Care
CFQ-R	Cystic Fibrosis Questionnaire-Revised
CushingQoL	Cushing's Quality of Life Questionnaire
DFS	Disease Free Survival
EMA	European Medicines Agency
EORTC- QLQ-C-30	European Organization for the Research and Treatment of Cancer Quality of Life Questionnaire
EQ-5D	European Quality of Life - 5 Dimensions
FACIT-D	Functional Assessment of Chronic Illness Therapy - Diarrhea
FACT-G	Functional Assessment of Cancer Therapy - General Scale
FACT-Lym	Functional Assessment of Cancer Therapy - Lymphoma
FACT-M	Functional Assessment of Cancer Therapy - Melanoma
FACT-P	Functional Assessment of Cancer Therapy - Prostate
FDA	Food and Drug Administration
FKSI-15	Functional Assessment of Cancer Therapy Kidney Symptom Index
G-BA	Gemeinsamer Bundesausschuss
HRQoL	Health Related Quality of Life
HTA	Health Technology Assessment
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ICD-10	Internat. Statistical Classification of Diseases and Related Health Problems - Version 10
ISOQOL	Internat. Society for Quality of Life Research
ISPOR	Internat. Society for Pharmacoeconomics and Outcomes Research
ITT	Intention to Treat
MID	Minimal Important Difference
MSQOL-54	Multiple Sclerosis Quality of Life-54 Instrument
NEI VFQ-25	National Eye Institute 25-item Visual Function Questionnaire
Norfolk QOL-DN	Norfolk Quality of Life - Diabetic Neuropathy Score
n.n.s.	nicht näher spezifiziert
OD	Orphan Drug

OS	Overall Survival
PRO	Patient Reported Outcome
PFS	Progression Free Survival
pU	pharmazeutischer Unternehmer
QALY	Quality-Adjusted Life Year
QoL	Quality of Life
QOLIE-31-P	Quality of Life in Epilepsy Inventory
RCT	Randomized Controlled Trial
SF-36	Short Form Health Survey
SGB	Sozialgesetzbuch
SGRQ	Saint George Respiratory Questionnaire
TrG	Tragende Gründe
VAS	Visuelle Analogskala
VD	Value Dossier
vfa	Verband forschender Arzneimittelhersteller e.V.
WHO	World Health Organization
WHO-QoL	World Health Organization Quality of Life Assessment
ZD	Zusammenfassende Dokumentation
ZN	Zusatznutzen
ZVT	zweckmäßige Vergleichstherapie

Abstract

With the 2011 introduction of an early benefit assessment (EBA) in Germany, health-related quality of life (HRQoL) was defined as one of three patient-relevant endpoints for showing the added benefits of new medicines. Pharmaceutical manufacturers have to submit a value dossier to the Joint Federal Committee (G-BA) stating the added benefit on at least one of these three endpoints.

This analysis depicts the status-quo of the submission and assessment of HRQoL data in the context of EBA during the period from January 1th, 2011 to August 31th, 2013. Moreover, it offers a framework to systemize the discussion around the assessment of HRQoL data within EBA. Although available in most value dossiers HRQoL data did not formally enter into the decision-making process in EBA to a large extent. Only in two cases was an additional benefit based on HRQoL data granted by the G-BA. Five main categories of structural deficiencies of submitted HRQoL data can be distinguished: Deficiencies in the study design, deficiencies of the assessment tool, deficiencies in the evaluation, lack of significance / relevance of data and formal deficiencies. In the face of those deficiencies, the development of a common set of mandatory requirements regarding HRQoL assessment seems highly desirable within the context of German EBA.

Keywords: Health-Related Quality of Life (HRQoL), Patient-Reported Outcomes (PRO), Early Benefit Assessment (EBA)

1 Einleitung

Steigende Gesundheitsausgaben haben in den vergangenen Jahren zur wachsenden Bedeutung des *Health Technology Assessment* (HTA) beigetragen, dessen elementarer Bestandteil es ist, Kosten und Nutzen medizinischer Technologien zu bestimmen, ins Verhältnis zu setzen und somit eine wirtschaftliche Ressourcenallokation zu gewährleisten.¹ Während sich die Kosten medizinischer Technologien vergleichsweise leicht bestimmen lassen, stellt sich im Kontext des HTA primär die Frage: Was ist Nutzen? Und wer soll bzw. kann darüber Auskunft geben?

Neben klassischen klinischen Studienendpunkten wie Morbidität und Mortalität sind zunehmend Daten gefragt, die den Mehrwert von Innovationen aus Patienten- und gesellschaftlicher Perspektive abbilden.² Die alleinige Berücksichtigung von somatischen Indikatoren des Therapieergebnisses und Überlebensraten ist demzufolge nicht ausreichend. Vielmehr gilt es, die Befindlichkeit und Funktionsfähigkeit von Patienten in verschiedenen Lebensbereichen im Rahmen des HTA zu berücksichtigen.³ Patientenberichtete Studienendpunkte (engl. *Patient Reported Outcomes*, PRO) etablierten sich vor diesem Hintergrund zu einer zentralen Nutzendimension des HTA.⁴ Das Konzept der gesundheitsbezogenen Lebensqualität (engl. *Health Related Quality of Life*, HRQoL) zählt zu den bekanntesten und am häufigsten erhobenen PRO.⁵ Anstatt rein ergänzende Daten zu Begleiterscheinungen der Behandlung zu liefern, sehen Autoren wie Kohlmann (2010) in der integrierten Betrachtung klinischer wie patientenberichteter Studienendpunkte ein erhebliches Potenzial, um ein vollständigeres Bild vom Nutzen und Schaden medizinischer Interventionen zu gewinnen.⁶ Laut Bullinger (2006) eröffnet die Erfassung der HRQoL die Chance, die Ressourcenallokation gezielter – im Sinne des Patienten – zu erbringen.⁷

Innerhalb des deutschen HTA-Verfahrens stellt HRQoL seit 2011 ein eigenständiges Nutzenkriterium dar.⁸ An Bedeutung gewinnt dabei die normative Komponente des Diskurses, die Nutzenbewertung deutlicher am Patienten auszurichten und die Vorlage von HRQoL-Daten stärker als bisher einzufordern.⁹ Herstellerseitig bedeutet dies gleichzeitig eine Herausforderung wie auch eine Chance. Einerseits gilt es, entsprechende Evidenz zu generieren. Andererseits eröffnen PRO neue Möglichkeiten, den Nutzen eines Arzneimittels abzubilden und zu vermarkten.¹⁰ Traditionelle Produktbeschreibungen (engl. *Claims*) bezüglich Sicherheit und

¹ Vgl. Perleth (2008), S. 14.

² Vgl. Klag et al. (2008), S.1.

³ Vgl. Bullinger/Schmidt (2006), S.41.

⁴ Vgl. Deutschinoff et al. (2005), S. 164; Doward/Gnanasakthy/Baker (2010), S.1.

⁵ Vgl. Klose et al. (2013), S. 14.e1.

⁶ Vgl. Kohlmann (2010), S. 261.

⁷ Vgl. Bullinger (2006), S.7.

⁸ Vgl. Klose et al. (2013), S. 14.e1.

⁹ Vgl. ÄZ (2013).

¹⁰ Vgl. Doward/Gnanasakthy/Baker (2010), S.1.

Wirksamkeit reichen alleine nicht mehr aus. Unter strategischen Gesichtspunkten gewinnt die Einbeziehung der Patientensicht im Rahmen klinischer Studien zunehmend an Bedeutung.¹¹

Dass PRO bereits auf Ebene der Arzneimittelzulassung Fuß gefasst haben, bestätigt eine Analyse von Caron et al. (2008).¹² Seit 1995 wurden von pharmazeutischen Unternehmen (pU) kontinuierlich entsprechende Daten bei der amerikanischen Arzneimittelzulassungsbehörde *Food and Drug Administration* (FDA) und der europäischen Arzneimittel-Agentur (engl. *European Medicines Agency*, EMA) eingereicht. Im Jahr 2008 wurden in 21 Prozent der EMA- und in 20 Prozent der FDA-Zulassungen PRO-Daten vorgelegt. Mit Fokus auf den Bereich der Onkologie identifizierten Kitchen et al. (2011) im Zeitraum von 1995 bis 2009 22 von 125 Substanzen (17,6 Prozent), bei denen die Zulassung durch die FDA oder die EMA zumindest teilweise auf PRO-Daten basierte.¹³ Erfolgsfaktoren der Einreichung waren u.a. die systematische Datensammlung, die klare Formulierung von Studienhypothese und Rationale der Erhebung, die Verankerung von PRO im statistischen Analyseplan (inkl. Subgruppenanalysen) und eine ausreichende Fallzahlplanung.

Während auf Ebene der Arzneimittelzulassung somit bereits eine Auseinandersetzung mit PRO-Daten stattfindet, ist ihre Berücksichtigung im Rahmen des nationalen HTA bisher kaum untersucht worden. Ziel der vorliegenden Arbeit ist die Bestimmung des Stellenwerts eines der zentralsten PRO-Konzepte – der HRQoL – im Kontext der frühen Nutzenbewertung in Deutschland. Abgebildet werden sollen einerseits das Einreichungsverhalten der pU sowie andererseits die Berücksichtigung der Daten innerhalb der Bewertungspraxis deutscher HTA-Institutionen.

In einem ersten, theoretischen Teil werden hierzu in Kapitel 2 die historischen und konzeptionellen Grundlagen der HRQoL vorgestellt. Daran anknüpfend wird in Kapitel 3 ein Überblick über die Ansätze, Instrumente und Güterkriterien der HRQoL-Erhebung gegeben. Kapitel 4 versteht sich als Überleitung zum empirischen Teil der vorliegenden Arbeit und geht auf die Besonderheiten der Etablierung von HRQoL als patientenrelevanter Endpunkt innerhalb der frühen Nutzenbewertung ein.

Der zweite, empirische Teil besteht in einer Analyse des bisherigen Standes der Berücksichtigung von Daten zur HRQoL im Rahmen der frühen Nutzenbewertung in Deutschland. Kapitel 5 präsentiert die forschungsleitende Fragestellung des empirischen Teils. In Kapitel 6 erfolgt eine Erläuterung der angewandten Methodik. Darauf folgt in Kapitel 7 die Präsentation der Ergebnisse. In Kapitel 8 werden diese diskutiert und schließlich in Kapitel 9 zusammengefasst und weiterer Forschungsbedarf aufgezeigt.

¹¹ Vgl. Ebd.

¹² Vgl. Caron et al. (2008).

¹³ Vgl. Kitchen et al. (2011).

2 Historische und konzeptionelle Grundlagen der HRQoL

Obwohl die Steigerung oder zumindest die Erhaltung der Lebensqualität (engl. *Quality of Life*, QoL) des Patienten immer schon eine maßgebliche Zielgröße des ärztlichen Handelns darstellte, etablierte sich erst Mitte der 70er Jahre im angloamerikanischen Raum und Anfang der 80er Jahre im deutschsprachigen Raum der Begriff der QoL als eigenständiger Terminus in der Medizin.¹⁴

Die explizite Beschäftigung mit der QoL des Patienten fußte auf der Einsicht, dass es neben den klassischen biomedizinischen Messgrößen noch andere Effekte von Behandlungen gibt, die die Gesundheit beeinflussen. Ihr zugrunde lag eine Erweiterung des traditionell biologischen Verständnisses von Gesundheit und Krankheit hin zu einem biopsychosozialen Modell.¹⁵ Das Konzept der QoL reicht über die objektive, körperliche Manifestation von Therapie und Behandlung hinaus und repräsentiert die subjektiv „erlebte Gesundheit“, bzw. die subjektive Repräsentation der Gesundheit aus Sicht der Betroffenen.¹⁶ Daher ist vielfach auch treffender die Rede von HRQoL. Anstatt den Blick auf einzelne Defizite zu richten, wird eine ganzheitliche Perspektive auf sämtliche Faktoren eingenommen, die die Wahrnehmung von Gesundheit bestimmen. Der Patient rückt vor dem Hintergrund seines erlebten Wissens über die Erkrankung in den Fokus des Forschungsinteresses.¹⁷ Der Begriff der (HR)QoL weist somit auch eine humanistische Konnotation auf, da der Patient zum informierten und handlungsleitenden Partner in der Interaktion mit dem Arzt wird.¹⁸ Klassische Zielkriterien (engl. *Outcomes*) der medizinischen Forschung, wie sie in Form von Mortalität und Morbidität bestehen, wurden im Zuge dessen auf ihre Relevanz für den Patienten hinterfragt.¹⁹

Laut Bullinger (2000a) markierte „*die explizite Anerkennung der Bedeutung von Patientenäußerungen zur eigenen Gesundheit [...] einen Wandel in der Therapieforschung*“.²⁰ Seitdem zeichnet sich eine wachsende Bedeutung des Konzepts der (HR)QoL in der Versorgungspraxis, der Sozialgesetzgebung und vor allem in der medizinischen Forschung ab. Indikatoren hierfür finden sich im Anwachsen einschlägiger Einträge in medizinischen Datenbanken. Zwischen 1975 und 2008 stieg die Anzahl der unter dem Stichwort QoL erfassten Veröffentlichungen in der Datenbank *MEDLINE* von ca. 3.000 auf über 8.000 Publikationen an, in der Datenbank *ClinicalTrials.gov* waren 2010 ca. 3.500 Studien verzeichnet.²¹ Auch das medizinische Literatursuchsystem *PubMed* konnte von 1975 bis 2013 einen kontinuierlichen Zuwachs verbuchen und listet heute über 100.000 Publikationen zum Thema QoL.²²

¹⁴ Vgl. Schölmerich/Thews (1990), S. 9; Bullinger (2000a), S.190.

¹⁵ Vgl. Bullinger (2000a), S.190.

¹⁶ Ebd., S.191.

¹⁷ Vgl. Ebd., S.190-192.

¹⁸ Vgl. Ebd., S.191.

¹⁹ Vgl. Bullinger (2014), S. 98.

²⁰ Bullinger (2000a), S.191.

²¹ Vgl. Kohlmann (2010), S. 260.

²² Vgl. Bullinger (2014), S. 98.

Die Gründe für diesen Paradigmenwechsel sind vielfältiger Natur. Laut Brenner (1995) sind für den Vormarsch der (HR)QoL-Forschung zwei Trends ursächlich: Einerseits steige die Lebenserwartung der Bevölkerung in sämtlichen industrialisierten Ländern, wodurch sich eine Verschiebung des Krankheitsspektrums von akuten zu chronischen Erkrankungen abzeichne.²³ Im Zuge dieser Entwicklung musste die Medizin erkennen, dass sie nicht nur heilen kann. Weniger das bloße Überleben sondern eher die Minimierung von Leiden und Einschränkungen des Patienten wurde damit zum Ziel neuer Arzneimitteltherapien – eine Einsicht, die (HR)QoL-Forscher bereits in den 90er Jahren zu der Forderung einer Neudefinition klinischer *Outcomes* veranlasste.²⁴ Andererseits bedinge der medizinisch-technische Fortschritt den Einsatz zunehmend hoch spezialisierter diagnostischer und therapeutischer Verfahren.²⁵ Infolge dessen ist eine weitaus feinere Betrachtung von Behandlungseffekten auf das Leben des Patienten möglich geworden, insbesondere dann, wenn nicht das reine Überleben im Fokus von Therapien steht, wie dies bei chronischen Erkrankungen oder in palliativen Situationen der Fall ist.²⁶ Zugleich bewirken ebenjene Innovationen in allen industrialisierten Ländern einen deutlich beobachtbaren Anstieg der Gesundheitsausgaben und damit verbunden eine wachsende Bedeutung der ökonomischen Evaluation von Gesundheitsleistungen.²⁷ Zur Legitimation von Allokationsentscheidungen und zur Beschreibung der Effekt- bzw. Nutzenseite wird zunehmend auf (präferenzbasierte) Bewertungen von Gesundheitszuständen zurückgegriffen.²⁸ Das Potenzial eines derartigen Paradigmenwechsels zeigt Graf-Baumann (1990) auf und spricht von (HR)QoL als einem „*kollektivem Ziel gesundheitsystemischer Entwicklung*“.²⁹ Durch die Einbeziehung der (HR)QoL unter individuellen und gesellschaftlichen Gesichtspunkten ließen sich effizientere und gerechtere, nicht nur gerechnete Verteilungsmodelle entwickeln.³⁰ All diese Argumente finden sich heute – rund 20 Jahre später – in der aktuellen Diskussion über den Stellenwert und das Potenzial des Konzepts der (HR)QoL wieder. Durch den Wandel der Patientenrolle wird dieser Trend zusätzlich befeuert.³¹ Schlagworte wie Patientensouveränität, -information und -sicherheit prägen die moderne Arzt-Patienten-Beziehung.³² Anstatt passiver Befehlsempfänger, ist der Patient heute aktiver Partner der Behandlung, dessen Wünsche und Bedürfnisse es zu erfragen und zu berücksichtigen gilt. Vor diesem Hintergrund steigt die Bedeutung der HRQoL als Ziel und Entscheidungsgrundlage der medizinischen Behandlung.

²³ Vgl. Brenner (1995), S. 41.

²⁴ Vgl. Schölmerich/Thews (1990), S. 10.

²⁵ Vgl. Brenner (1995), S. 41.

²⁶ Vgl. Ebd.

²⁷ Vgl. Brenner (1995), S. 41; Stratmann-Schöne (2001), S. 21.

²⁸ Vgl. Stratmann-Schöne (2001), S. 21f.

²⁹ Graf-Baumann (1990), S. 278.

³⁰ Vgl. Graf-Baumann (1990), S. 278.

³¹ Vgl. Klakow-Franck (2014), S. 152.

³² Vgl. Ebd.

2.1 Patientenberichtete Studienendpunkte (Patient Reported Outcomes, PRO)

Eine Verortung des Konzepts der HRQoL im Kontext klinischer Studien kommt nicht ohne den Rückgriff auf den Begriff der PRO aus, der um 1999 erstmals in der englischsprachigen Literatur auftauchte.³³ Synonym werden auch die Begriffe der *Self Reported Health Outcomes* bzw. der *Patient Generated Outcomes* verwendet.³⁴ PRO stellen einen Oberbegriff für klinische *Outcomes* dar, die direkt vom Patienten erhoben werden und dessen Sicht auf die Erkrankung, mögliche Begleiterscheinungen und die Behandlung widerspiegeln.³⁵

Während sich das Konzept der HRQoL bereits wesentlich früher im klinischen Kontext etabliert hatte, bedurfte es Ende der 90er Jahre einer begrifflichen Öffnung.³⁶ Anlass gab ein neu entstandener Forschungszweig, der den Patienten im Rahmen klinischer Studien zunehmend selbst zu Wort kommend ließ. Die gewachsene Bedeutung von PRO spiegelt sich insbesondere auf der Ebene der Arzneimittelzulassung wider: Gleichzeitig mit dem Höhepunkt der Einreichung von PRO-Daten bei der FDA wurde im Jahr 2000 die *PRO Harmonization Group* ins Leben gerufen, der Vertreter von Zulassungsbehörden, wissenschaftlichen Fachgesellschaften, Verbänden der pharmazeutischen Industrie, der *International Society for Quality of Life Research* (ISOQOL) sowie der *International Society for Pharmacoeconomics and Outcomes Research* (ISPOR) angehörten.³⁷ Im Zuge der Konferenz wurden PRO zum offiziellen Oberbegriff aller direkt vom Patienten erhobenen klinischen *Outcomes* erklärt, unter den fortan auch der vorherige Oberbegriff der QoL subsumiert werden sollte.³⁸

Gemäß der Definition der FDA gilt als PRO:

*“[...] any report of the status of a patient’s health condition that comes **directly from the patient**, without interpretation of the patient’s response by a clinician or anyone else.”³⁹*

Ähnlich definiert auch die EMA PRO als:

*“Any outcome evaluated **directly by the patient** himself and based on patient’s perception of a disease and its treatment(s) [...]”⁴⁰*

Entscheidend für die Begriffsdefinitionen der FDA wie auch der EMA ist das Kriterium der Unmittelbarkeit. Als Konsequenz dieser stringenten Sichtweise sind Beurteilungen durch sog. Proxiberichterstatter wie Angehörige oder Pflegepersonal, die also stellvertretend für den Patienten Auskunft geben, kritisch zu bewerten oder ganz unzulässig.⁴¹ Die Definition der EMA bestimmt darüber hinaus den subjektiv-evaluativen Charakter als zentrales Wesensmerkmal von

³³ Vgl. Sloan et al. (2007), S. 60.

³⁴ Vgl. Ebd.

³⁵ Vgl. Ebd.

³⁶ Vgl. Ebd.

³⁷ Vgl. DIMDI (2011), S. 33.

³⁸ Vgl. Sloan et al. (2007), S. 60.

³⁹ FDA (2009), S. 2.

⁴⁰ EMA (2005), S. 3.

⁴¹ Vgl. DIMDI (2011), S. 33.

PRO. Beide Definitionen beziehen sich jedoch primär auf die Informationsquelle und lassen den eigentlichen Inhalt bzw. das zu evaluierende *Outcome* zunächst offen.

Das Urteil des Patienten kann zu vielfältigen unterschiedlichen Aspekten seines Gesundheitszustandes eingeholt werden. Die Erhebung von PRO erscheint immer dann geboten, wenn für den Patienten zentrale, messbare *Outcomes* der Behandlung nicht verfügbar, potenziell verzerrt oder äußerst selten sind.⁴² Dies betrifft bspw. Begleiterscheinungen der Behandlung wie Schmerzen, sexuelle Dysfunktion oder Schlaflosigkeit.⁴³ Im Kontext klinischer Studien gilt es zunächst, das jeweilige PRO-Konzept zu spezifizieren, d.h. den jeweiligen Gegenstand festzulegen, der gemessen werden soll. Laut FDA (2009) repräsentiert ein PRO-Konzept Aspekte der Funktionsfähigkeit oder des Befindens vor dem Hintergrund der jeweiligen Erkrankung und ihrer Behandlung.⁴⁴ Erhoben werden können direkt erfragbare Auswirkungen, wie beispielsweise Symptome oder Nebenwirkungen.⁴⁵ Darüber hinaus eröffnen Konzepte wie das der Funktionsfähigkeit, des Wohlbefindens, der HRQoL sowie der Behandlungszufriedenheit einen komplexeren Zugang zum Erleben der Behandlung aus Sicht des Patienten.⁴⁶ All diesen Konzepten ist ein subjektiver Charakter gemein; unterschiedlich ist hingegen der Grad, zu welchem eine Evaluation durch den Befragten stattfindet. Brenner (1995) verwendet hierfür den Begriff des „proximal-distalen Kontinuums“, das sich zwischen objektiv zugänglichen biologischen, hin zu allein subjektiv-evaluierbaren, psychologischen PRO-Konzepten aufspannt.⁴⁷ Kohlmann (2010) unterscheidet zwischen Einzelsymptomen (z.B. Schmerz, Übelkeit), Konzepten (z.B. körperliche Funktionsfähigkeit, Depressivität) und mehrdimensionalen Konstrukten (z.B. HRQoL).⁴⁸

Einen Ansatzpunkt zur Klassifikation von PRO-Konzepten stellt das biophysiologische Erklärungsmodell nach Wilson und Cleary (1995) dar, das fünf aufeinander aufbauende Ebenen von Gesundheitszuständen unterscheidet.⁴⁹

⁴² Vgl. Johnston et al. (2013), S. 3.

⁴³ Vgl. Ebd.

⁴⁴ Vgl. FDA (2009), S. 2.

⁴⁵ Vgl. Rothman et al. (2007), S. 67; Kohlmann (2010), S. 260f.

⁴⁶ Vgl. Rothman et al. (2007), S. 66f.; Kohlmann (2010), S. 260.

⁴⁷ Brenner (1995), S. 49.

⁴⁸ Vgl. Kohlmann (2010), S. 260f.

⁴⁹ Vgl. Wilson/Cleary (1995), zitiert nach DIMDI (2011), S. 40.

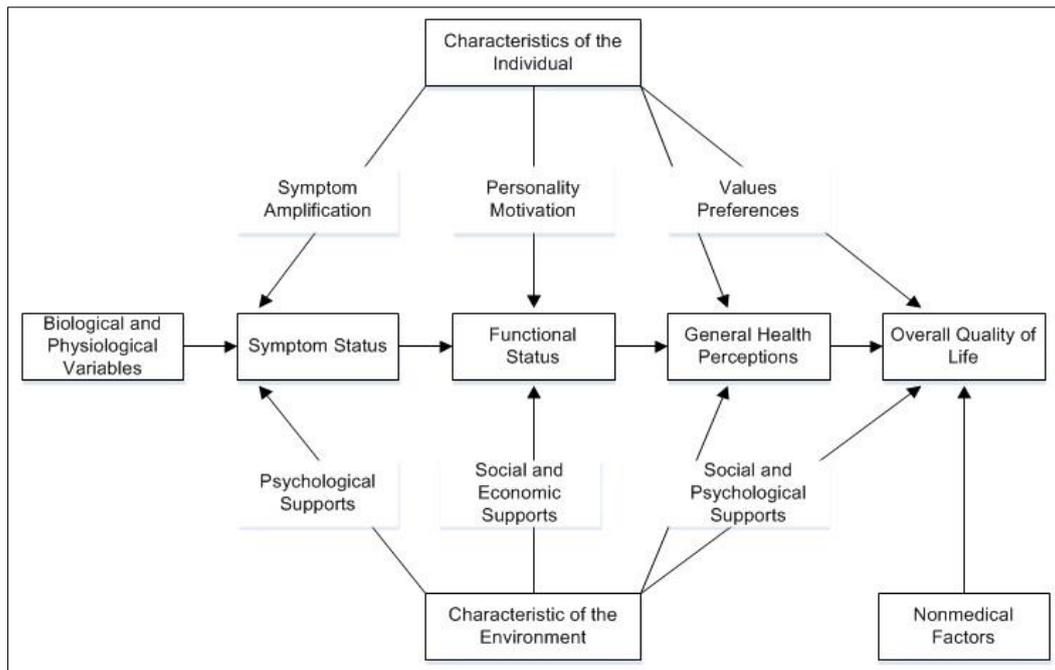


Abbildung 1: Modell nach Wilson und Cleary (1995), zitiert nach DIMDI (2011), S. 40.

Abbildung 1 zeigt das Modell von Wilson und Cleary, das eine Brücke zwischen dem biomedizinischen und dem sozialwissenschaftlichen Paradigma zu schlagen versucht.⁵⁰ Ausgehend von klassischen biophysiological Variablen (engl. *Biological and Psychological Variables*) bildet der Symptomstatus (engl. *Symptom Status*), der funktionelle Status (engl. *Functional Status*) und die allgemeine Gesundheitswahrnehmung (engl. *General Health Perceptions*) das Bindeglied zum Konzept der QoL.

Biologische und physiologische Daten sind als tendenziell objektive Indikatoren des Zustandes von Zellen, Organen und Organsystemen zu verstehen.⁵¹ Ihre Manifestation in Form von Symptomen, verstanden als durch den Patienten wahrgenommene, anomale körperliche oder kognitive Zustände, unterliegt jedoch bereits dem subjektiven Urteil des Befragten.⁵² Entscheidend ist, was der Patient tatsächlich erlebt. Eine exemplarische Frage zur Erhebung des Symptomstatus könnte lauten: „*Hatten Sie innerhalb der letzten zwei Wochen körperliche Schmerzen? Falls ja, in welcher Ausprägung?*“

Der Funktionsstatus bildet die Fähigkeit des Individuums ab, physische, soziale und psychologische Aufgaben vor dem Hintergrund des jeweiligen Symptomstatus zu erfüllen.⁵³ Eine exemplarische Frage zur Erhebung der Funktionsfähigkeit könnte lauten: „*Hindert Sie Ihr gesundheitlicher Zustand an folgenden Aktivitäten: Tragen von Einkäufen, etc.? Falls ja, wie sehr?*“

⁵⁰ Vgl. DIMDI (2011), S. 39-41.

⁵¹ Vgl. DIMDI (2011), S. 39-41.

⁵² Vgl. Ebd.

⁵³ Vgl. Ebd.

Die allgemeine Gesundheitswahrnehmung repräsentiert das subjektive Gesamturteil der drei vorgelagerten Stufen.⁵⁴ Eine exemplarische Frage zur Erhebung der Gesundheitswahrnehmung könnte lauten: „Haben Sie sich kürzlich krank gefühlt?“

Unter Berücksichtigung nicht-medizinischer Einflussfaktoren konstituiert sich schließlich die QoL als finaler Endpunkt innerhalb des Wilson-Cleary-Modells.⁵⁵ Eine exemplarische Frage zur Erhebung der QoL könnte lauten: „Wie hoch würden Sie Ihre Lebensqualität innerhalb der vergangenen Woche auf einer Skala von 1-7 einschätzen?“

Die Abfolge der Gesundheitszustände im Wilson-Cleary-Modell ist unidirektional, d.h. das Erreichen einer höheren Stufe wird erst durch Informationen der vorangegangenen Stufen ermöglicht.⁵⁶ Jede der Stufen stellt zugleich einen potenziellen Ansatzpunkt zur Erhebung von PRO-Konzepten dar. Entsprechend nehmen Valderas und Alonso (2008) das Wilson-Cleary-Modell als Ausgangspunkt ihres PRO-Klassifikationsansatzes und erweitern es einerseits durch die Ergebnisse einer systematischen Literaturrecherche sowie durch die Integration der zentralen Bestandteile der *International Classification of Functioning, Disability and Health (ICF)*.⁵⁷

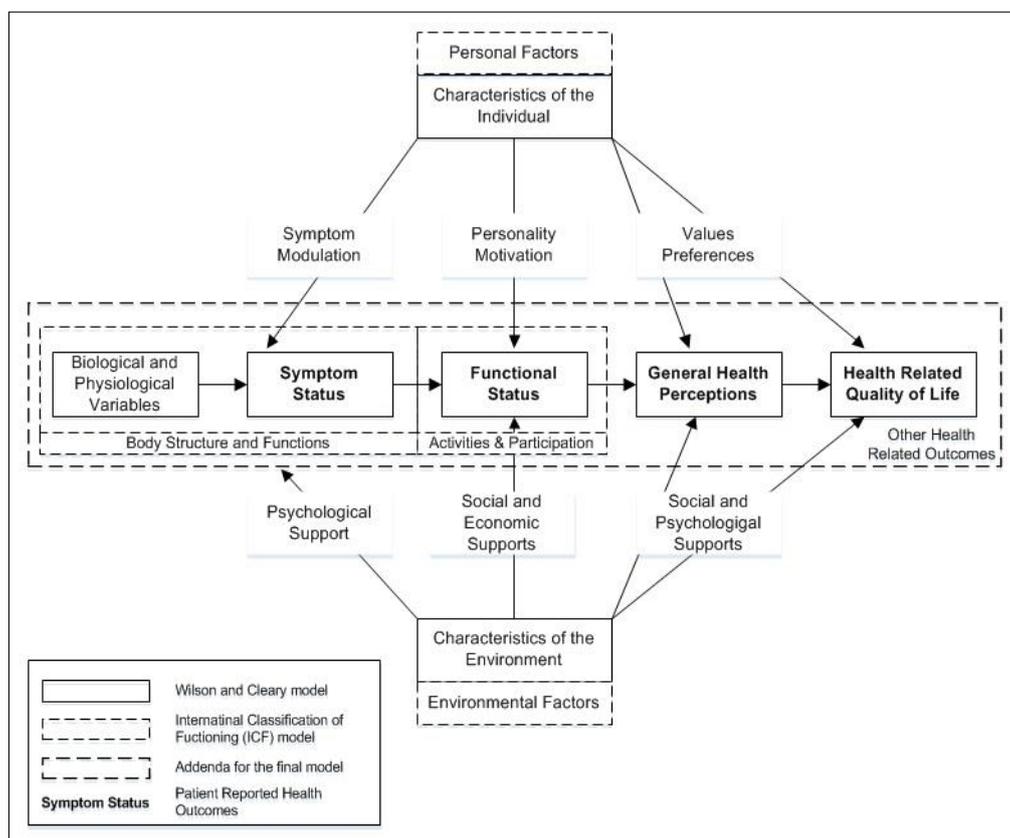


Abbildung 2: PRO-Klassifikationsschema nach Valderas und Alonso (2008), S. 1129.

⁵⁴ Vgl. Ebd.

⁵⁵ Vgl. Ebd.

⁵⁶ Vgl. Ebd.

⁵⁷ Vgl. Valderas/Alonso (2008).

Wie aus Abbildung 2 hervorgeht, lagern Valderas und Alonso die biologischen und psychologischen Variablen (engl. *Biological and Psychological Variables*) ihrem Klassifikationsschema vor, da diese in der Regel nicht vom Patienten direkt erfragt würden und somit nicht als PRO zu werten seien.⁵⁸ Ebenso plädieren sie vor dem Hintergrund der Ergebnisse ihrer Literaturrecherche dafür, das Konzept der *Overall QoL* durch das enger gefasste Konzept der HRQoL auszutauschen.⁵⁹ Darüber hinaus ergänzen sie das Modell um weitere gesundheitsbezogene Konstrukte (engl. *Other Health Related Constructs*), worunter sie gängige PRO-Konzepte fassen, die jedoch weder im Wilson-Cleary-Modell noch in der ICF enthalten sind. Zu diesen zählen Behandlungszufriedenheit sowie Resilienz, d.h. die Fähigkeit, mit Stress und Krankheit umzugehen.⁶⁰

Sowohl das Wilson-Cleary als auch das Valderas-Alonso-Modell eint der Gedanke einer Rangfolge von Gesundheitszuständen bzw. PRO. Entscheidend ist laut Wintner et al. (2012) die Differenzierung zwischen Symptom und erlebter Folge bzw. Konsequenz; erst ab dieser Stufe tragen PRO den innerpsychischen und sozialen Prozessen im Zuge der Krankheitsverarbeitung Rechnung.⁶¹ Die Relevanz dieser Schwelle formulieren Doward und McKenna (2004) wie folgt:

„It should be noted that “patient-reported” implies that the instrument is completed by the patient. It should not be inferred from this term that the information gained is necessarily of direct concern to the patient.”⁶²

Obleich weniger komplexe PRO einen Zugang zum Erleben des Patienten darstellen und somit die Einflüsse von Erkrankung und Behandlung realistischer abbilden, sagen sie ebenso wenig über deren Relevanz für den Patienten aus wie traditionelle biomedizinische Messgrößen. Um die für den Patienten relevanten Behandlungseffekte abzubilden, bedarf es komplexerer multidimensionaler PRO-Erhebungsverfahren, wie sie im Rahmen der (HR)QoL-Forschung entwickelt wurden.

2.2 Lebensqualität (Quality of Life, QoL)

Eine der frühen Erwähnungen des Begriffs der QoL findet sich in der Nikomachischen Ethik von Aristoteles (384-322 BC), welche sich als Leitfaden eines glücklichen Lebens versteht:

„Both the multitude and persons of refinement ... conceive “the good life” or “doing well” to be the same thing as “being happy”. But what constitutes happiness is a matter of dispute ... some say one thing and some another, indeed very often the same man says different things at different times: when he falls sick he thinks health is happiness, when he is poor, wealth.”⁶³

⁵⁸ Vgl. Valderas/Alonso (2008). S. 1127.

⁵⁹ Vgl. Ebd.

⁶⁰ Vgl. Ebd., S. 1128.

⁶¹ Vgl. Wintner et al. (2012), S. 294.

⁶² Doward/McKenna (2004), S. 4.

⁶³ Zitiert nach Fayers/Machin (2007), S. 5.

Bereits in dieser frühen Definition der QoL sind zwei wesentliche Aspekte enthalten: Einerseits der Aspekt der Subjektivität und damit die Feststellung, dass verschiedene Menschen ganz unterschiedliche Auffassungen davon haben, was unter QoL zu verstehen ist.⁶⁴ Andererseits der Aspekt intraindividuelle Schwankungen der QoL, anhängig von der jeweiligen Situation des Individuums.⁶⁵

Die Ursprünge der modernen QoL-Forschung sind in der sozialwissenschaftlichen Wohlfahrts- und Sozialindikatorenforschung zu verorten.⁶⁶ Im Zusammenhang mit dem *Social-Indicators Movement* der 1960er Jahre in Amerika wurde das allgegenwärtige Mantra des volkswirtschaftlichen Wachstums als zentrales Ziel nationalstaatlicher Politik und Abbild gesellschaftlichen Fortschritts zunehmend in Frage gestellt.⁶⁷ Der Begriff der QoL als Maß für die Wohlfahrt einer Nation entwickelte sich zu einer alternativen Zieldimension staatlichen Handelns und gesellschaftlicher Entwicklung und fand Eingang in die gängigen Instrumente der Sozialberichterstattung (bspw. Herausgabe des *Towards a Social Report* des *US Departments of Health, Education and Welfare*, 1969).⁶⁸ Aufgrund eines korrespondierenden gesellschaftlichen und politischen Klimas fanden ähnliche Forschungsansätze innerhalb der nachfolgenden zehn Jahre auch international rasch Verbreitung. Organisationen wie die Weltbank (*World Development Indicators*) und die Vereinten Nationen (*Human Development Index*) entwickelten eigene Indikatoren zur Erhebung des sozialen Wohlbefindens bzw. des Wohlfahrts- und Entwicklungsniveaus von Nationen.⁶⁹

Zu unterscheiden sind dabei zwei unterschiedliche Herangehensweise zur Erfassung der QoL auf Bevölkerungsebene: Während Vertreter des skandinavischen Ansatzes QoL als Maß für die allgemeine Wohlfahrt verstehen und primär über den Zugang zu Ressourcen, d.h. über objektive Indikatoren wie Geld, Besitz, Wissen und Sicherheit definieren, fokussiert der amerikanische Ansatz auf die Erhebung subjektiver Indikatoren wie Zufriedenheit und Freude des individuellen Bürgers.⁷⁰ Die Erhebung der QoL über objektive Indikatoren basiert auf der normativen Annahme eines Idealzustands, gegen den ein Vergleich der tatsächlich vorgefundenen Lebensumstände vorgenommen werden kann.⁷¹ Die Verwendung subjektiver Indikatoren zur

⁶⁴ Vgl. Ebd.

⁶⁵ Vgl. Ebd.

⁶⁶ Vgl. Schumacher/Klaiberg/Brähler (2003), S. 9.

⁶⁷ Vgl. Rapley (2003), S. 5.

⁶⁸ Vgl. Ebd., S. 5-6.

⁶⁹ Vgl. Ebd., S. 7.

⁷⁰ Vgl. Rapley (2003), S. 4-6.

⁷¹ Vgl. Ebd., S. 11f.

Messung der QoL geht davon aus, dass sich Wohlfahrt bzw. QoL nicht direkt messen lässt, sondern sich in der Wahrnehmung des Individuums konstituiert.⁷²

Darüber hinaus existieren Ansätze einer Synthese beider Positionen. Zapf (1984) versteht QoL als aggregiertes Maß der Übereinstimmung von objektiven Lebensbedingungen und deren subjektiver Bewertung.⁷³ Entsprechend bilden subjektive und objektive Variablen die zwei Achsen des Konzepts der QoL auf Bevölkerungsebene.

	Subjektives Wohlbefinden		
Objektive Lebensbedingungen		gut	schlecht
	gut	<i>WELL-BEING</i>	<i>DISSONANZ</i>
	schlecht	<i>ADAPTION</i>	<i>DEPRIVATION</i>

Abbildung 3: Dimensionen der QoL auf Bevölkerungsebene nach Zapf (1984), S. 25.

Abbildung 3 verdeutlicht den Zusammenhang zwischen objektiven Lebensbedingungen und subjektivem Wohlbefinden anhand der Dichotomisierung gut bzw. schlecht. Laut Zapf (1984) lassen sich vier unterschiedliche Wohlfahrtspositionen bzw. Dimensionen der QoL auf Bevölkerungsebene unterscheiden.⁷⁴ Die Dimension des Wohlbefindens – wofür Zapf den englischen Begriff *Well-Being* vorzieht – ist ein Zustand, in dem die objektiv als gut zu bewertenden Lebensbedingungen auch mit einem hohen Wohlbefinden des Individuums korrespondieren. Den Konterpart dazu stellt der Zustand der Deprivation dar, der aus der Kombination schlechter Lebensbedingungen mit einem korrespondierenden, schlechten Wohlbefinden resultiert. Dissonanz beschreibt einen inkonsistenten Zustand guter objektiver Lebensbedingungen bei gleichzeitigem niedrigem Wohlbefinden des Individuums, auch als sog.

⁷² Vgl. Ebd.

⁷³ Vgl. Zapf (1984), S. 24f.

⁷⁴ Vgl. Zapf (1984), S. 25f.

Unzufriedenheitsdilemma bezeichnet.⁷⁵ Der Zustand der Adaption bildet das Phänomen eines hohen subjektiven Wohlbefindens trotz objektiv als schlecht zu bewertender Lebensbedingungen an, welches auch als sog. Zufriedenheitsparadoxon bezeichnet wird.⁷⁶ Aus den beiden letztgenannten Dimensionen, der Dissonanz und der Adaption, wird deutlich, dass die Beziehung zwischen objektiven und subjektiven Faktoren eine lose ist. Objektive Faktoren können nur einen Teil der Zufriedenheit des Individuums mit seinen jeweiligen Lebensumständen erklären. Zufriedenheit hinsichtlich des Einkommens, Bildungsstands und Lebensstandards ergibt sich jedoch zu einem großen Teil aus dem subjektiven Vergleich, den das Individuum kontinuierlich mit seinen bisherigen Erfahrungen, Erwartungen und seinem Umfeld (Verwandte, Freunde, Gesellschaft) vornimmt.⁷⁷

Einen unterschiedlichen Zugang wählen QoL-Konzepte auf Ebene des Individuums.⁷⁸ Zu den zentralsten konzeptionellen Ansätzen zählt Bullinger (2014) subjektive Wohlbefindens-, Zufriedenheits-, Bedürfnis-, Rollenfunktions- und Vergleichsmodelle.⁷⁹ Zu unterscheiden sind hierbei Ansätze, die sich auf die Gesamtheit aller Lebensumstände einer Person beziehen (globale QoL) und solche, die sich bestimmten Teilbereichen der menschlichen Existenz widmen, wie beispielsweise Gesundheit (HRQoL) oder Krankheit.⁸⁰ Dem globalen Begriff der QoL auf Ebene des Individuums ist die die Begriffsdefinition der Weltgesundheitsorganisation (engl. *World Health Organization*, WHO) aus dem Jahre 1993 zuzuordnen. Sie definiert QoL wie folgt:

„Lebensqualität ist die subjektive Wahrnehmung einer Person über ihre Stellung im Leben in Relation zur Kultur und den Wertesystemen, in denen sie lebt und in Bezug auf ihre Ziele, Erwartungen, Maßstäbe und Anliegen. Es handelt sich um ein breites Konzept, das in komplexer Weise beeinflusst wird durch die körperliche Gesundheit einer Person, den psychischen Zustand, die sozialen Beziehungen, die persönlichen Überzeugungen und ihre Stellung zu den hervorstechenden Eigenschaften der Umwelt.“⁸¹

Indem die WHO den Einfluss von Kultur und Werten auf die subjektive Einschätzung der Lebensumstände durch das Individuum hervorhebt, eröffnet sie einen universellen Zugang zur QoL.⁸² Gleichzeitig integriert sie das breite Spektrum konzeptioneller Ansätze. Das körperliche, seelische, soziale und spirituelle Wohlbefinden werden als zentrale Dimensionen oder Domänen der QoL identifiziert.

⁷⁵ Vgl. Ebd.

⁷⁶ Vgl. Rapley (2003), S. 31.

⁷⁷ Vgl. Levine (1995), S. 9.

⁷⁸ Vgl. Rapley (2003), S. 49.

⁷⁹ Vgl. Bullinger (2014), S. 99.

⁸⁰ Vgl. Rapley (2003), S. 49.

⁸¹ WHO (1993), S. 153.

⁸² Vgl. Rapley (2003), S. 50.

2.3 Gesundheitsbezogene Lebensqualität (Health Related Quality of Life, HRQoL)

Im Kontext klinischer Studien ist weniger der globale Begriff der QoL Gegenstand des Forschungsinteresses. Stattdessen wird mit dem Konzept der HRQoL ein spezieller Aspekt der QoL in den Blick genommen, der sich mit den Einflussfaktoren von Erkrankung und Behandlung auf den Patienten befasst.⁸³ Laut Bullinger (2000b) ist der Begriff der HRQoL überall dort zu verwenden, „*wo bedeutsame Aspekte des Befindens und Handlungsvermögens von Personen Beachtung finden, welche unter gesundheitlichen Einschränkungen leiden bzw. chronisch krank sind.*“⁸⁴ Im Kern der HRQoL-Forschung geht es darum, das „*Befinden und Handlungsvermögen kranker Menschen in Beziehung [...] zu Therapie und Krankheitsverlauf*“ zu setzen.⁸⁵ Der Fokus liegt dabei auf der Abbildung der Veränderung der HRQoL bei Individuen oder Patientengruppen im Sinne eines subjektiven Evaluationsindizes im Zeitverlauf.⁸⁶

Ebenso wie QoL stellt HRQoL einen nicht eindeutig zu definierenden Begriff dar.⁸⁷ Umstritten ist, welche Domänen der QoL darunter zu fassen sind, ob also beispielsweise auch die indirekten Auswirkungen, wie sie sich in Form von Arbeitslosigkeit oder finanzieller Schwierigkeiten darstellen können, damit abzubilden sind, weil sich Krankheit und Behandlung darauf auswirken.⁸⁸ Laut Raspe (1990) bedarf es innerhalb des medizinischen Kontextes eines gewissen Grades an Reduktionismus.⁸⁹ Zwar blende das Konzept der HRQoL politische, kulturelle, spirituelle, ökologische und finanzielle Aspekte aus der Betrachtung aus, dadurch würde es jedoch bestimmter und der wissenschaftlichen Erhebung zugänglich.⁹⁰ Ähnlich pragmatisch plädiert auch Schöffski (2007a) für eine Einschränkung des Begriffs der HRQoL im Rahmen gesundheitsökonomischer Evaluationen, da die Steuerungsmöglichkeiten des Adressaten, d.h. primär der Gesundheitspolitik, auf den medizinischen Kontext begrenzt seien.⁹¹

Für die Praxis der HRQoL-Erhebung spielt die Auseinandersetzung mit den konzeptionellen Modellen der QoL kaum eine Rolle.⁹² Ausgangspunkt vieler begrifflicher Bestimmungen der HRQoL stellt die im Jahr 1948 von der WHO vorgeschlagene Gesundheitsdefinition dar.

⁸³ Vgl. Wood-Dauphinee (1999), S. 356; Fayers/Machin (2007), S. 4.

⁸⁴ Bullinger (2000b), S. 11.

⁸⁵ Siegrist (1990), S. 59.

⁸⁶ Vgl. King (2001), S. 60.

⁸⁷ Vgl. Fayers/Machin (2007), S. 4.

⁸⁸ Vgl. Ebd.

⁸⁹ Vgl. Raspe (1990), S. 28.

⁹⁰ Vgl. Ebd.

⁹¹ Vgl. Schöffski (2007a), S. 323.

⁹² Vgl. Bullinger (2014), S. 99.

Demzufolge ist Gesundheit mehr als die reine Abwesenheit von Krankheit und stellt einen Zustand des „*umfassenden physischen, psychischen und sozialen Wohlbefindens*“ dar.⁹³ Wenngleich diese Definition hinsichtlich ihrer Abdeckung des Konzepts der HRQoL nicht unumstritten ist, gilt sie dennoch als Minimalkonsens und Referenzpunkt der Entwicklung von HRQoL-Messinstrumenten. In Anlehnung daran definiert Bullinger (2000) HRQoL als:

„[...] multidimensionales Konstrukt, das sich auf körperliche, emotionale, mentale, soziale und verhaltensbezogene Komponenten von Wohlbefinden und Funktionsfähigkeit aus der Sicht von Patienten und/oder Beobachtern bezieht.“⁹⁴

Nicht nur Krankheit und Behandlung, sondern auch individuelle Charakteristika der Person (Persönlichkeit, Verarbeitungsstrategien) und strukturelle Merkmale (Lebensbedingungen, Schichtzugehörigkeit) determinieren die HRQoL.⁹⁵ Ähnlich konstatiert Radoschewski (2000), das Konzept der HRQoL würde:

„[...] von den strukturellen und funktionalen Auswirkungen des Gesundheitszustandes auf Lebenssituation und Lebensgestaltung, seine Wahrnehmung und Wertigkeit für die Menschen bestimmt. Im Konzept der gesundheitsbezogenen Lebensqualität erhalten die psychische und soziale Dimension von Gesundheit eine Leitfunktion.“⁹⁶

Enthalten in dieser Definition ist die Attribuierung eines Wertes in Bezug auf die HRQoL, wodurch in definatorischer Hinsicht eine Brücke zu den nutzentheoretischen Verfahren der HRQoL-Messung geschlagen wird. Die EMA (2005) definiert HRQoL als:

“[...] a specific type/subset of PROs, distinguished by its multi-dimensionality. Indeed, HRQL is a broad concept which can be defined as the patient’s subjective perception of the impact of his disease and its treatment(s) on his daily life, physical, psychological and social functioning and well-being.”⁹⁷

Deutlich wird daran nochmals die Abgrenzung zum Oberbegriff der PRO: Anders als z.B. der Symptomstatus ist die HRQoL das „*Resultat individueller komplexer Bewertungs- und Beurteilungsprozesse*“ und entsprechend multidimensional zu erfassen.⁹⁸

Testa und Simonson (1996) zufolge lässt sich jede Domäne der HRQoL aus zwei Dimensionen messen: als objektive Erhebung des Gesundheitszustandes (engl. *Objective Health Status*) einer Person, basierend auf einer Bewertung ihrer physischen, psychischen und sozialen

⁹³ WHO (1948).

⁹⁴ Bullinger (2006), S. 5.

⁹⁵ Vgl. Bullinger (2014), S. 99.

⁹⁶ Radoschewski (2000), S. 186.

⁹⁷ EMA (2005), S. 3.

⁹⁸ Rennenberg/Hammelstein (2006), S. 31.

Funktionsfähigkeit sowie als subjektive Wahrnehmung (engl. *Subjective Perception*) des jeweiligen Zustandes, im Sinne eines physischen, psychischen und sozialen Wohlbefindens.⁹⁹

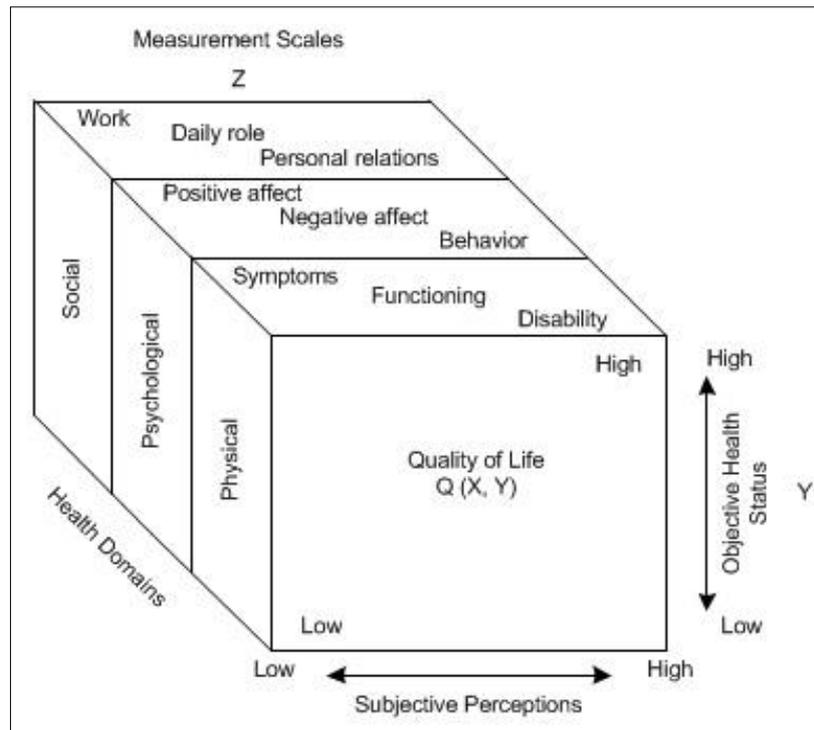


Abbildung 4: Konzeptionelles Modell der HRQoL nach Testa und Simonson (1996), S. 835.

Abbildung 4 zeigt das konzeptionelle Modell der HRQoL nach Testa und Simonson (1996). Die x-Achse bildet dabei die subjektiven, die y-Achse die objektiven Dimensionen der drei zentralen Domänen des Konzepts der HRQoL ab.¹⁰⁰ Zwischen beiden Achsen spannt sich das Konzept der HRQoL auf. Die objektive Dimension der HRQoL ist als Gradmesser des Gesundheitszustandes in physischer, psychischer und sozialer Hinsicht zu verstehen, der jedoch erst durch das subjektive Urteil des Patienten in das Konzept der HRQoL transformiert wird.¹⁰¹ Am Beispiel der physischen Domäne verdeutlicht bedeutet das, nicht nur die Häufigkeit und Schwere von Symptomen sondern zugleich deren Relevanz aus Sicht des Patienten zu erfragen.¹⁰²

⁹⁹ Vgl. Testa/Simonson (1996), S. 835.

¹⁰⁰ Vgl. Testa/Simonson (1996), S. 835f.

¹⁰¹ Vgl. Ebd., S. 835.

¹⁰² Vgl. Fairclough (2010), S. 5.

Physisches Wohlbefinden/Funktionsfähigkeit	Psychisches Wohlbefinden/Funktionsfähigkeit	Soziales Wohlbefinden/Funktionsfähigkeit
<ul style="list-style-type: none"> • Beschwerden • Symptome • Nebenwirkungen (Schmerz, Übelkeit, Erbrechen) • Kraft • Energie • Etc. 	<ul style="list-style-type: none"> • Stimmung • Isolation • Niedergeschlagenheit • Angst • Depression • Furcht • Etc. 	<ul style="list-style-type: none"> • Rollenfunktion • Individuelle Beziehungen • Kontakte zu Familie, Freunden, Arbeitskollegen und der Gemeinde • Fähigkeiten, Aktivitäten des täglichen Lebens (ATL) durchzuführen • Berufstätigkeit • Abhängigkeit von anderen • Etc.
<ul style="list-style-type: none"> • Individuelle Charakteristika der Person • Strukturelle Merkmale 		

Tabelle 1: Domänen und Konzepte der HRQoL, eigene Abbildung in Anlehnung an King (2001), S. 58; EMA (2005), S. 3; Walters (2009), S. 2; Schöffski (2007a), S. 324; Bullinger (2014), S. 99.

Tabelle 1 gibt einen Überblick über die drei zentralen Domänen der HRQoL und eine Auswahl der darunter zu subsumierenden Konzepte. Die Mehrheit der HRQoL-Instrumente deckt zumindest ein Konzept der jeweiligen Domäne ab.¹⁰³ In Anbetracht der enormen Bandbreite beinhaltet die Instrumentenentwicklung immer auch eine Selektion der für die jeweilige Fragestellung relevanten Konzepte.¹⁰⁴

Eine abschließende Definition scheint vor dem Hintergrund der Komplexität und der Multidimensionalität des Konzepts der HRQoL schwer.¹⁰⁵ Weitestgehender Konsens besteht hinsichtlich der Dynamik, der Subjektivität und der Multidimensionalität des Konzepts der HRQoL.¹⁰⁶ Im Kontext der vorliegenden Arbeit soll HRQoL sowohl als Beschreibung des Gesundheitszustandes aus Sicht des Betroffenen (Bullinger 2000, EMA 2005) als auch als

¹⁰³ Vgl. Fayers/Machin (2007), S. 5.

¹⁰⁴ Vgl. Ebd.

¹⁰⁵ Vgl. Wood-Dauphinee (1999), S. 356.

¹⁰⁶ Vgl. King (2001), S.57; Walters (2009), S. 1; Fayers/Machin (2007), S. 5; DIMDI (2011), S. 22.

präferenzbasierte Einschätzung (Radoschewski 2000) desselben verstanden werden. Zur Abgrenzung von anderen Konzepten wird das Kriterium der Mehr- oder Multidimensionalität verwendet, basierend auf den aus dem WHO-Gesundheitsbegriff abgeleiteten Domänen des physischen, psychischen und sozialen Wohlbefindens und der Funktionsfähigkeit.

3 Messung und Validierung von HRQoL

Die begriffliche Vielfalt des Konzepts der HRQoL schlägt sich in den unterschiedlichen Modellen nieder, die die Beziehung der enthaltenen Domänen und Konzepte abzubilden und zu erklären versuchen.¹⁰⁷ In Ermangelung einer allgemeingültigen Begriffsdefinition wird HRQoL oftmals vor dem jeweiligen Untersuchungsgegenstand oder überhaupt nicht näher definiert. In Folge dessen wird der HRQoL-Forschung eine Kluft zwischen Theorie und Praxis attestiert.¹⁰⁸ Laut Bullinger (2000) hat sich jedoch in der Zwischenzeit – individuell für die jeweils betrachtete Patientenpopulation – ein Konsens über die operationale Definition des Konzepts sowie über das Messinstrumentarium herausgebildet.¹⁰⁹ Einen Überblick präsentieren Schumacher/Klaiberg/Brähler (2003).¹¹⁰ Das zur Verfügung stehende Messinstrumentarium lässt sich anhand verschiedener Kriterien unterscheiden.

3.1 Nutzentheoretische und psychometrische Ansätze

Hinsichtlich der Zugänge zu dem latenten Konzept HRQoL kann zwischen nutzentheoretischen und psychometrischen Ansätzen unterschieden werden.¹¹¹

Nutzentheoretische oder auch als präferenzbasiert bezeichnete Ansätze der HRQoL-Messung zielen auf die Bewertung eines Gesundheitszustands.¹¹² Solche Präferenzen können entweder direkt vom Patienten erhoben, oder indirekt durch eine Befragung der Öffentlichkeit ermittelt werden.¹¹³ Ausgehend von diesen Präferenzen werden spezifische Nutzwerte (engl. *Utilities*) einzelner Gesundheitszustände abgeleitet, wie sie im Rahmen von Kosteneffektivitätsanalysen zur Berechnung sog. *Quality-Adjusted Life Years* (QALY) benötigt werden.¹¹⁴ Aufgrund des

¹⁰⁷ Vgl. Hazard Vallerand/Breckenridge/Hodgson (2001), S. 77-94.

¹⁰⁸ Vgl. Bullinger (2014), S. 99.

¹⁰⁹ Vgl. Bullinger (2000a), S. 191f; Bullinger (2014), S. 99.

¹¹⁰ Vgl. Schumacher/Klaiberg/Brähler (2003).

¹¹¹ Vgl. Fairclough (2010), S. 3-5; Schöffski (2007a), S. 327.

¹¹² Vgl. DIMDI (2011), S. 48.

¹¹³ Vgl. Fairclough (2010), S. 4.

¹¹⁴ Vgl. Ebd.

multidimensionalen Charakters der HRQoL ist dieses Vorgehen nicht unumstritten.¹¹⁵ Vertreter der präferenzbasierten Messung der HRQoL argumentieren, dass das Abwägen zwischen verschiedenen Gesundheitszuständen ein integraler Bestandteil der Therapieentscheidung des Patienten ist.¹¹⁶ Für eine tiefergehende Auseinandersetzung mit präferenzbasierten Ansätzen der HRQoL-Erhebung sei an dieser Stelle auf Greiner/Klose (2014) verwiesen.¹¹⁷

Psychometrische Ansätze dienen der Beschreibung der externen, der internen und der antizipierten Erfahrungen des Individuums.¹¹⁸ Die Mehrheit der zur Erfassung von PRO und der HRQoL zur Verfügung stehenden Messinstrumente basiert auf dem psychometrischen Ansatz.¹¹⁹ Ausgehend von der Annahme, dass das latente Konzept der HRQoL sich nicht direkt messen lässt, wird mittels einzelnen Domänen, Komponenten und letztlich einzelnen Fragen (engl. *Items*) der Versuch einer Operationalisierung unternommen.¹²⁰

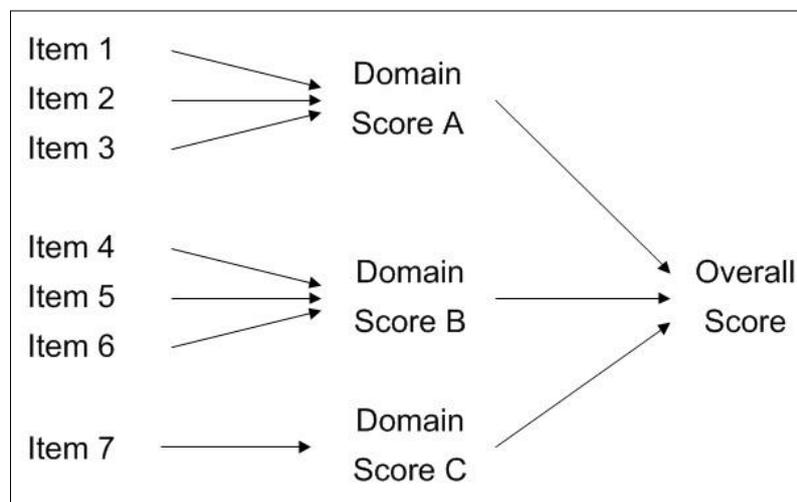


Abbildung 5: Operationalisierung der HRQoL, eigene Darstellung in Anlehnung an Walters (2009), S. 14.

Abbildung 5 verdeutlicht die Operationalisierung des Konzepts der HRQoL entsprechend des psychometrischen Ansatzes. Dem Patienten wird eine Auswahl geordneter *Items* präsentiert, zu denen in der Regel ordinalskalierte Antwortmöglichkeiten vorgegeben werden.¹²¹ Der Patient wird daraufhin gebeten, über das Vorhandensein, die Häufigkeit und die Intensität von Symptomen oder Verhaltensweisen Auskunft zu geben; erfragt werden können aber auch

¹¹⁵ Vgl. Stratmann-Schöne (2001), S. 22f.

¹¹⁶ Vgl. Stratmann-Schöne (2001), S. 22f.

¹¹⁷ Vgl. Greiner/Klose (2014).

¹¹⁸ Vgl. Walters (2009), S. 17-18; DIMDI (2011), S. 47.

¹¹⁹ Vgl. Marquis et al. (2006), S. 195; DIMDI (2011), S. 47.

¹²⁰ Vgl. Walter (2009), S. 13f.

¹²¹ Vgl. Ebd., S. 17f.

Fähigkeiten, Gefühle oder die generelle Wahrnehmungen des Gesundheitszustandes.¹²² Entsprechend des Zustimmungsgrades werden Punkte vergeben.¹²³ Der Punktwert (engl. *Score*) von Antworten, die der Beschreibung einer übergeordneten Domäne (engl. *Domain*) dienen, wird zusammengefasst.¹²⁴ Aggregiert oder unter Abbildung der einzelnen Domänen konstituiert sich schließlich das übergeordnete Konzept der HRQoL.¹²⁵

Die Erfassung der HRQoL kann sowohl mittels Profil- als auch Indexinstrumenten (s.u.) erfolgen, wobei einige Autoren dem psychometrischen Ansatz ausschließlich das Profilinstrument zuordnen.¹²⁶

3.2 Profil- und Indexinstrumente

Profilinstrumente tragen der Multidimensionalität der HRQoL Rechnung, da sie die einzelnen Domänen wie zuvor beschrieben getrennt erfassen und abbilden.¹²⁷ Entsprechend der Minimaldefinition der WHO ermitteln Profilinstrumente getrennte *Scores* für die Domänen des psychischen, physischen und sozialen Wohlbefindens, lassen sich jedoch abhängig von der zugrundeliegenden Begriffsdefinition der HRQoL um beliebig viele Domänen erweitern. Jedes *Item* geht dabei mit gleicher Gewichtung in den domänenspezifischen Gesamtscore ein.¹²⁸ Indem sie die einzelnen Domänen der HRQoL getrennt ausweisen, umgehen Profilinstrumente die Problematik der Gewichtung von *Items* entsprechend ihres Einflusses auf das Gesamtkonzept der HRQoL. Laut Bullinger (2000) liegt der Vorteil des Profilinstruments in der Reichhaltigkeit der Information.¹²⁹ Profile ermöglichen durch ihre Differenzierung zwischen den einzelnen Domänen eine genauere Abbildung eines Therapieeffektes. Hieraus ergibt sich jedoch die Problematik, die Manifestation gegensätzlicher Therapieeffekte zu bewerten.¹³⁰ Zeichnet sich bspw. in einzelnen Domänen eine leichte Verbesserung ab, in einer anderen jedoch eine deutliche Verschlechterung, stellt sich die Frage, wie die Ergebnisse in der Summe zu bewerten sind.¹³¹

¹²² Vgl. Marquis et al. (2006), S. 195.

¹²³ Vgl. Walters (2009), S. 49.

¹²⁴ Vgl. Marquis et al. (2006), S. 195.

¹²⁵ Vgl. Testa/Simonson (1996), S. 835f.

¹²⁶ Vgl. DIMDI (2011), S. 47.

¹²⁷ Vgl. Bullinger (2000a), S.195.

¹²⁸ Vgl. Walters (2009), S. 49.

¹²⁹ Vgl. Bullinger (2000a), S.195.

¹³⁰ Vgl. PCORI (2012), S. 15f.

¹³¹ Vgl. Ebd.

Aufgrund ihrer detaillierten Betrachtung lassen sich Profile zudem schlechter zwischen unterschiedlichen Indikationen oder Therapien vergleichen.¹³²

Indexinstrumente sind handlicher und werden vor allem im Rahmen gesundheitsökonomischer Analysen zur Generierung von Nutzwerten eingesetzt.¹³³ Gemeint sind damit Instrumente, die sich aus mehreren *Items* zusammensetzen und ein aggregiertes Ergebnis in Form einer Maßzahl generieren.¹³⁴ Sie basieren somit auf der Annahme eines zugrundeliegenden, umfassenderen theoretischen Konzepts der HRQoL, welches sich in Form gleich oder unterschiedlich gewichteter Indikatoren abbilden lässt.¹³⁵ Aufgrund der komprimierten Ergebnisdarstellung bieten sich Indizes für den Vergleich zwischen verschiedenen Therapien an.¹³⁶ In Anbetracht der Multidimensionalität der HRQoL ist die Zulässigkeit dieses hohen Abstraktionsgrades jedoch umstritten. Laut Raspe (1990) besteht die Grundproblematik der Indexbildung darin, dass gesundheitliche Störungen auf unterschiedlichen Ebenen nicht zwangsläufig miteinander kovariieren, sondern nur mehr oder weniger schwache statistische Zusammenhänge bestehen.¹³⁷ Im Rahmen einer Aggregation der *Item*-Werte ist daher eine Nivellierung möglicher Effekte nicht auszuschließen. Ein weiteres Problem der Indexbildung besteht in dem Erfordernis einer Gewichtung der *Items*.¹³⁸ Zur Ermittlung entsprechender Gewichtungsfaktoren (engl. *Tariffs*) stehen drei unterschiedliche Möglichkeiten zur Verfügung: die Verwendung von Referenzwerten aus der Literatur, Plausibilitätsüberlegungen (engl. *Judgments*) sowie die direkte Ermittlung der Präferenzen der Befragten.¹³⁹ Zur vertiefenden Auseinandersetzung mit der Präferenzmessung sei an dieser Stelle auf Schöffski (2007b) verwiesen.¹⁴⁰

In der Gesamtschau der jeweiligen Vor- und Nachteile konstatiert Schöffski (2007a), Indexinstrumente seien vor allem eine wichtige Grundlage für Allokationsentscheidungen, wohingegen er den Einsatz von Profilinstrumenten eher im Bereich klinischer Untersuchungen verortet.¹⁴¹ Der erforderliche Beweis von Wirksamkeit und Wirtschaftlichkeit im Rahmen der Zulassung sowie des nachgelagerten HTA scheint somit die kombinierte Anwendung beider Instrumententypen zu erfordern.

¹³² Vgl. Bullinger (2000a), S.195.

¹³³ Vgl. DIMDI (2011), S. 47.

¹³⁴ Vgl. Raspe (1990), S. 32.

¹³⁵ Vgl. Ebd.

¹³⁶ Vgl. Bullinger (2000a), S.195.

¹³⁷ Vgl. Raspe (1990), S. 36.

¹³⁸ Vgl. Schöffski (2007a), S. 328.

¹³⁹ Vgl. Schöffski (2007b), S. 341.

¹⁴⁰ Vgl. Ebd., S. 335-385.

¹⁴¹ Vgl. Schöffski (2007a), S. 329.

3.3 Generische und spezifische Messinstrumente

Hinsichtlich des Instrumententyps kann zwischen krankheitsübergreifenden (generischen), spezifischen und individuellen Messinstrumenten unterschieden werden.

Generische Instrumente sind darauf ausgelegt, sämtliche relevanten HRQoL-Domänen zu erfassen, losgelöst vom jeweiligen Erkrankungskontext.¹⁴² Daher sind sie mitunter auch sehr komplex und zeitaufwendig in der Durchführung, was insbesondere bei Schwerstkranken ein Problem darstellt.¹⁴³ Durch ihren weiten Fokus sind generische Instrumente in der Lage, auch unvorhersehbare Behandlungseffekte zu erfassen, weswegen sie sich besonders für klinische Studien eignen, in denen neue Wirkungsprofile erprobt werden.¹⁴⁴ Ihrem Ursprung nach sind generische Instrumente im Bereich der Bevölkerungsstudien zu verorten, wo sie zunächst der Erfassung der QoL an Gesunden dienen.¹⁴⁵ Obwohl in der Zwischenzeit auch generische Instrumente für klinische *Settings* entwickelt wurden, sind einige davon auch heute noch bei gesunden Befragten einsetzbar.¹⁴⁶ Generische Instrumente ermöglichen somit nicht nur den Vergleich der HRQoL zwischen verschiedenen Therapien und Indikationen, sondern auch gegenüber einem Normwert, wie er sich aus der Bevölkerung ableiten lässt.¹⁴⁷ Der Preis einer krankheitsübergreifenden HRQoL-Erfassung liegt in der geringen Sensitivität generischer Instrumente, also der Fähigkeit, die für den Patienten maßgeblichen Einflussflussfaktoren der Behandlung auf die HRQoL zu erfassen.¹⁴⁸ Seitens regulatorischer Behörden wie der EMA wird daher der Einsatz von krankheitsspezifischen Instrumenten im Kontext klinischer Studien gefordert.¹⁴⁹

Generische Fragebögen können sowohl Profil- als auch Indexinstrumente sein.¹⁵⁰ Spezifische Fragebögen sind hingegen in der Regel als Profilinstrumente konzipiert.¹⁵¹ Sie sind auf bestimmte Erkrankungen, Populationen, HRQoL-Domänen oder Zustände zugeschnitten und insofern die sensibleren Messinstrumente.¹⁵² Am häufigsten finden spezifische Instrumente Anwendung, die zur Erfassung der HRQoL bestimmter Erkrankungen (z.B. Diabetes, Asthma, Mukoviszidose)

¹⁴² Vgl. King (2001), S. 61.

¹⁴³ Vgl. Ebd.

¹⁴⁴ Vgl. DIMDI (2011), S. 48.

¹⁴⁵ Vgl. Fayers/Machin (2007), S. 19.

¹⁴⁶ Vgl. Ebd.

¹⁴⁷ Vgl. Ebd., S. 23.

¹⁴⁸ Vgl. Ebd.

¹⁴⁹ Vgl. EMA (2005), S. 3.

¹⁵⁰ Vgl. Schöffski (2007a), S. 330.

¹⁵¹ Vgl. Ebd.

¹⁵² Vgl. King (2001), S. 62.

konzipiert sind.¹⁵³ Ihr Einsatz verringert den Befragungsaufwand und ist in der Regel mit einer höheren Akzeptanz seitens der Befragten verbunden, da die Relevanz der Fragen für den Patienten klar erkennbar ist.¹⁵⁴ Daher ist auch davon auszugehen, dass der Fragebogenrücklauf spezifischer Instrumente höher ausfällt.¹⁵⁵ Jedoch bedingt die hohe Sensitivität im Umkehrschluss, dass diese Instrumente nicht für den Vergleich mit anderen Erkrankungen oder Zuständen eingesetzt werden können und in der Regel auch keine Normwerte zur Verfügung stehen.¹⁵⁶

Neben generischen und krankheitsspezifischen Instrumenten, die im klinischen Kontext als etabliert angesehen werden können, existieren auch individuelle HRQoL-Messinstrumente. Im Unterschied zu den konventionellen Befragungsinstrumenten nehmen individuelle Instrumente eine radikal-subjektive Perspektive auf die HRQoL des Patienten ein und sind daher weder standardisiert, noch werden Daten im Rahmen der Auswertung aggregiert.¹⁵⁷ Individuelle Instrumente ermöglichen es dem Patienten, die für ihn relevanten Dimensionen der HRQoL selbst festzulegen.¹⁵⁸ Die Ergebnisse werden auf einer Fall-zu-Fall Basis beurteilt, ausgehend von der jeweils unterschiedlichen, individuellen Reaktion und Gewichtung eines Lebensumstands durch den Patienten.¹⁵⁹

Die Online-Datenbank *PROQOLID* (www.proqolid.org) listet zwischenzeitlich 803 HRQoL-Messinstrumente. Insbesondere krankheitsspezifische Instrumente wurden oftmals nur im Rahmen einer einzelnen Studie eingesetzt und sind somit unzureichend validiert.¹⁶⁰ Dennoch existiert heute ein reichhaltiger Methodenköffer etablierter Messinstrumente, die die drei Validierungsstufen der Übersetzung, der psychometrischen Testung und der Normierung in unterschiedlichen Ländern erfolgreich durchlaufen haben.¹⁶¹ In der Praxis besteht daher oftmals die Problematik der Auswahl des richtigen Messinstruments. Dies gilt im Speziellen für die Erfassung der HRQoL, im Allgemeinen jedoch für alle PRO. Verschiedene Autoren haben daher Kriterienkataloge für die Auswahl von Instrumenten vorgestellt.¹⁶² Darüber hinaus liegt ein Schwerpunkt der aktuellen PRO-Forschung auf *Head-to-Head*-Studien, die ähnliche Instrumente

¹⁵³ Vgl. DIMDI (2011), S.48.

¹⁵⁴ Vgl. Ebd.; Fitzpatrick et al. (1998), S. 8.

¹⁵⁵ Vgl. Fitzpatrick et al. (1998), S. 8.

¹⁵⁶ Vgl. Ebd., S. 8f.

¹⁵⁷ Vgl. DIMDI (2011), S.48.

¹⁵⁸ Vgl. Fitzpatrick et al. (1998), S. 12.

¹⁵⁹ Vgl. DIMDI (2011), S. 48.

¹⁶⁰ Vgl. Bullinger (2000a), S. 192.

¹⁶¹ Vgl. Ebd.; Kohlmann (2014), S. 105.

¹⁶² Vgl. Fitzpatrick et al. (2001); Morkink et al. (2010).

miteinander vergleichen.¹⁶³ Ihr Ziel ist es, eine sachliche Grundlage für die Auswahl von Messinstrumenten im Kontext klinischer Studien zu schaffen.

3.4 Erhebungsverfahren, -zeitpunkt und Recall-Periode

Das HRQoL-Erhebungsverfahren der Wahl ist in der Regel der selbst auszufüllende Fragebogen.¹⁶⁴ Vor allem zur Entwicklung neuer Fragebögen wird jedoch auch auf das Interview als exploratives Erhebungsverfahren zurückgegriffen.¹⁶⁵

Fragebögen eignen sich besonders gut für den Einsatz bei großen Patientenzahlen, erfordern weniger Zeit als ein Interview und sind tendenziell einfacher in der Handhabung.¹⁶⁶ Aufgrund dieser Eigenschaften sind die Fragebogenentwicklung und ihr Einsatz als Hauptströmung der HRQoL-Erhebung zu betrachten.¹⁶⁷ Fragebögen können entweder in Papierform oder elektronisch vorliegen, wobei Untersuchungen zu dem Schluss kommen, dass beide Verfahren hinsichtlich der Ergebnisqualität als gleichwertig zu betrachten sind.¹⁶⁸ Im Vergleich zum Interview weist die HRQoL-Erhebung mittels Fragebögen tendenziell eine schlechtere Datenqualität auf, da mit unvollständigem und/oder geringem Rücklauf zu rechnen ist.¹⁶⁹ Eine weitere Option stellt die Erfassung der HRQoL in Form eines Patienten-Tagebuchs dar.¹⁷⁰ Neuere Ansätze bestehen in computerisiert vorgegebenen, auf Basis eines Algorithmus ausgewählten *Items* im Rahmen des sog. computeradaptiven Testens (CAT).¹⁷¹ Darüber hinaus existieren erste Projekte zum Aufbau von *Item*-Datenbanken, um die jeweils besten *Items* fragebogenübergreifend zu sammeln und in neue Instrumente zu überführen.¹⁷²

Vorgeschlagen wird, die HRQoL zu definierten Intervallen innerhalb der Behandlung zu erheben und die Patienten beispielsweise nach zwei, vier und sechs Wochen nach ihrem jeweils aktuellen Status zu befragen.¹⁷³ Auf diese Weise lässt sich ein möglicher Therapieeffekt besser im Behandlungsverlauf abbilden. Als wenig aussagekräftig werden hingegen aggregierte HRQoL-

¹⁶³ Vgl. Radoschewski (2000), S. 186; Kohlmann (2010), S. 262.

¹⁶⁴ Vgl. DIMDI (2011), S. 51.

¹⁶⁵ Vgl. Flechtner (2001), S. 24.

¹⁶⁶ Vgl. Ebd.

¹⁶⁷ Vgl. Ebd.; Kohlmann (2014), S. 105.

¹⁶⁸ Vgl. DIMDI (2011), S. 51.

¹⁶⁹ Vgl. Schöffski (2007a), S. 326.

¹⁷⁰ Vgl. Ebd., S. 326f.

¹⁷¹ Vgl. Bullinger (2014), S. 100.

¹⁷² Vgl. Ebd.

¹⁷³ Vgl. FDA (2009), S. 14.

Daten angesehen, basierend auf kurzen, stichprobenhaften Befragungszeiträumen.¹⁷⁴ Bei der Festlegung des Erhebungszeitpunktes ist zu berücksichtigen, inwiefern für den Patienten maßgebliche Intervalle gewählt werden.¹⁷⁵ Anstatt auf das voraussichtliche Auftreten eines Symptoms zu fokussieren, spiegeln längere Beobachtungszeiträume unter Umständen besser die Fähigkeit des Patienten wider, Aktivitäten des täglichen Lebens auszuüben.

Als *Recall*-Periode wird der Zeitraum bezeichnet, den der Patient zur Beantwortung der Frage vergegenwärtigen können muss.¹⁷⁶ Je kürzer die *Recall*-Periode ist, umso geringer ist die Wahrscheinlichkeit, dass die Ergebnisse systematisch verzerrt sind (sog. *Recall-Bias*).¹⁷⁷ Insofern wird die Festlegung der *Recall*-Periode auch seitens der FDA als wichtiger methodischer Faktor angesehen, der insbesondere im Zusammenhang mit komplexeren Erinnerungsvorgängen über die Güte der Erhebung entscheidet.¹⁷⁸ Zur Festlegung der *Recall*-Periode sollte der Zweck des Erhebungsinstruments reflektiert werden sowie die Variabilität, Dauer, Frequenz und Intensität des Konzepts, das gemessen werden soll.¹⁷⁹ Des Weiteren sollten die Charakteristika der Erkrankung sowie der Therapie berücksichtigt werden.¹⁸⁰ Üblicherweise beträgt der erfragte Zeitraum die vergangenen sieben Tage oder vier Wochen.¹⁸¹

3.5 Gütekriterien

Klassische Gütekriterien der PRO-Erhebung sind die Validität, Reliabilität und Responsivität (auch: Änderungssensitivität) eines Messinstruments.¹⁸² Einige Autoren führen darüber hinaus das Kriterium der Interpretierbarkeit von Messergebnissen als weiteres Gütekriterium der HRQoL-Erhebung an.¹⁸³

Das Kriterium der Validität gibt an, ob das Messinstrument misst, was es messen soll und ist somit als Maß für die Genauigkeit der Messung zu verstehen.¹⁸⁴ Unterschieden wird zwischen der Augenschein-, der Kriteriums-, der Inhalts- und der Konstruktvalidität.¹⁸⁵

¹⁷⁴ Vgl. FDA (2009), S. 14.

¹⁷⁵ Vgl. King (2001), S. 107.

¹⁷⁶ Vgl. DIMDI (2011), S. 51.

¹⁷⁷ Vgl. Ebd.

¹⁷⁸ Vgl. FDA (2009), S. 14.

¹⁷⁹ Vgl. Ebd.

¹⁸⁰ Vgl. Ebd.

¹⁸¹ Vgl. Fairclough (2010), S. 7.

¹⁸² Vgl. King (2001), S. 60f; FDA (2009), S. 15; PCORI (2012), S. 11-13.

¹⁸³ Vgl. Walters (2009), S. 29; Kohlmann (2014), S. 106.

¹⁸⁴ Vgl. King (2001), S. 60; DIMDI (2011), S. 51.

¹⁸⁵ Vgl. DIMDI (2011), S. 51f.

Die Augenscheinvalidität ist ein subjektives Maß und gilt dann als gewährleistet, wenn das Instrument im Rahmen einer oberflächlichen Betrachtung misst, was es messen soll.¹⁸⁶ Insofern ähnelt die Augenscheinvalidität der Inhaltsvalidität und wird mitunter auch als ein Bestandteil derselben bezeichnet.¹⁸⁷ Jedoch lässt sich die Augenscheinvalidität im Unterschied nur ex post, d.h. nach Einsatz des Messinstruments, einschätzen.¹⁸⁸

In Ermangelung eines „wahren“ Wertes erfolgt die Beurteilung der Kriteriumsvalidität anhand eines Goldstandardverfahrens, d.h. anhand eines Verfahrens, das nach dem aktuellen Kenntnisstand das zu messende Konzept der HRQoL mit höchstmöglicher Sicherheit erfasst.¹⁸⁹ Dies trifft beispielsweise im Falle der Entwicklung von Kurzversionen etablierter Fragebögen zu, deren Ergebnisse sich anhand der ursprünglichen Langversion beurteilen lassen. Die Kriteriumsvalidität gilt als erfüllt, wenn die Skalenwerte mit Ergebnissen des externen Vergleichs assoziiert sind.¹⁹⁰

Die Inhaltsvalidität bezieht sich auf die Angemessenheit der Zusammenstellung einzelner *Items* zur Erfassung der HRQoL.¹⁹¹ Die zugrundeliegende Frage ist, ob die ermittelten Werte mit dem tatsächlichen Zustand des Patienten korrespondieren, das Instrument also in der Lage ist, diesen in all seinen Facetten abzubilden. Perfekte Inhaltsvalidität gilt jedoch als unerreichbar.¹⁹² Im Gegensatz zur Augenscheinqualität setzt die Inhaltsvalidität vor dem eigentlichen Einsatz bei der Entwicklung des Instruments an. Bereits im Rahmen der Fragebogenentwicklung sollten Experten und Patienten zur Beurteilung der Vollständigkeit, des Umfangs und der Akzeptanz der Fragen und Antwortmöglichkeiten konsultiert werden.¹⁹³ Nur auf diese Weise kann sichergestellt werden, dass das Instrument sämtliche Einflussfaktoren auf das zu messende Konzept erfasst und dass diese aus Sicht des Betroffenen auch tatsächlich relevant sind.¹⁹⁴

Die Beurteilung der Konstruktvalidität erfolgt anhand des zugrundeliegenden konzeptionellen Modells der HRQoL.¹⁹⁵ Insbesondere dann, wenn für die Validierung eines Instruments keine vergleichbare Methodik bzw. kein Goldstandard existiert, wie dies bei der Entwicklung neuer

¹⁸⁶ Vgl. DIMDI (2011), S. 51.

¹⁸⁷ Vgl. Fayers/Machin (2007), S. 79.

¹⁸⁸ Vgl. Ebd.

¹⁸⁹ Vgl. DIMDI (2011), S. 52; Fayers/Machin (2007), S. 79.

¹⁹⁰ Vgl. Fayers/Machin (2007), S. 77.

¹⁹¹ Vgl. DIMDI (2011), S. 52.

¹⁹² Vgl. Fayers/Machin (2007), S. 78.

¹⁹³ Vgl. Ebd., S. 78f.

¹⁹⁴ Vgl. Ebd., S. 79.

¹⁹⁵ Vgl. Ebd.

Instrumente zutrifft, kommt der Konstruktvalidität eine zentrale Rolle zu.¹⁹⁶ Unterformen der Konstruktvalidität sind die konvergente und die divergente oder auch diskriminante Validität: Die konvergente Validität wird über den Vergleich mit ähnlichen Messinstrumenten ermittelt und bildet ab, ob das Instrument zu ähnlichen Ergebnissen gelangt.¹⁹⁷ Die diskriminante Validität hingegen beschreibt seine Fähigkeit, zwischen verschiedenen Patientengruppen zu unterscheiden.¹⁹⁸

Vorbehalte hinsichtlich der Validität erscheinen insbesondere dann geboten, wenn spezifische HRQoL-Instrumente in einer anderen Patientenpopulation oder in einem kulturell und sprachlich abweichenden Umfeld verwendet werden.¹⁹⁹

Das Kriterium der Reliabilität bezieht sich auf die Konsistenz, Stabilität und Reproduzierbarkeit der Messergebnisse.²⁰⁰ Um formal als möglichst zuverlässig (reliabel) angesehen zu werden, dürfen unterschiedliche Messergebnisse bei wiederholtem Einsatz desselben Instrumentes allein auf Veränderungen des Gesundheitszustandes zurückzuführen sein.²⁰¹ Zur Beurteilung der Reliabilität wird ein Instrument bei denselben Befragten mehrfach eingesetzt (sog. *Test-Retest-Reliabilität*).²⁰² Liegen Fremdbeurteilungen vor, wird das Ausmaß der Übereinstimmungen der Einschätzungsergebnisse bei unterschiedlichen Untersuchern (engl. *Rater*) abgebildet (sog. *Interrater-Reliabilität*).²⁰³ Darüber hinaus wird die Reproduzierbarkeit der Ergebnisse bei der Erhebung an verschiedenen Orten, z.B. im Krankenhaus oder zu Hause, bestimmt.²⁰⁴

Das Kriterium der Responsivität bzw. Änderungssensitivität beschreibt die Fähigkeit eines Messinstruments, Veränderungen der HRQoL abzubilden.²⁰⁵ Dies betrifft die Verbesserung, die Konstanz und die Verschlechterung der HRQoL. Zu unterscheiden sind dabei zwei Formen der Responsivität: Die interne Responsivität bezieht sich auf die Anwendung eines Messinstruments im Zeitverlauf und spiegelt wider, wie gut ein Instrument Veränderungen der HRQoL abbildet.²⁰⁶ Erlebt der Patient beispielsweise vermehrt Symptome, sollte sich dies entsprechend abbilden. Die

¹⁹⁶ Vgl. DIMDI (2011), S. 52.

¹⁹⁷ Vgl. Ebd.

¹⁹⁸ Vgl. Bullinger/Schmidt (2006), S. 2507.

¹⁹⁹ Vgl. Johnston et al. (2013), S. 6.

²⁰⁰ Vgl. DIMDI (2011), S. 52.

²⁰¹ Vgl. Schöffski (2007b), S. 340.

²⁰² Vgl. Ebd.

²⁰³ Vgl. DIMDI (2011), S. 53.

²⁰⁴ Vgl. Walters (2009), S. 27.

²⁰⁵ Vgl. DIMDI (2011), S. 53; Kohlmann (2014), S. 106.

²⁰⁶ Vgl. DIMDI (2011), S. 53.

HRQoL als patientenrelevanter Endpunkt in der frühen Nutzenbewertung gemäß §35a SGB V

externe Responsivität bezieht sich auf ein Standardmessinstrument und vergleicht die ermittelten Messergebnisse im Zeitverlauf auf ihre Korrelation hin.²⁰⁷

Die Problematik einer geringen Responsivität, bzw. falsch negativer Ergebnisse besteht vor allem bei der Verwendung generischer Instrumente, da diese nicht auf den jeweiligen Erkrankungskontext zugeschnitten sind (Vgl. 3.3. Generische und spezifische Messinstrumente).²⁰⁸

Das Kriterium der Interpretierbarkeit knüpft an die Fähigkeit des Messinstruments an, relevante Veränderungen der HRQoL zu identifizieren. Sind beispielsweise nur recht geringe Veränderungen der HRQoL feststellbar, stellt sich die Frage, inwiefern diese für den Patienten tatsächlich spürbar und letztendlich klinisch relevant sind.²⁰⁹ Für Individuen gilt es daher, die im Zeitverlauf kleinste relevante Differenz der HRQoL festzulegen. Dieser intraindividuelle Bezugswert (engl. *Benchmark*) wird auch als *Minimal Important Difference* (MID) bezeichnet.²¹⁰ Gleichmaßen gilt es, eine spezifische MID zur Bewertung relevanter Gruppenunterschiede zu definieren.²¹¹

4 HRQoL als patientenrelevanter Endpunkt in der frühen Nutzenbewertung gemäß §35a SGB V

Mit dem Gesetz zur Neuordnung des Arzneimittelmarktes (AMNOG) wurde im Jahr 2011 die Grundlage für das deutsche HTA in Form der frühen Nutzenbewertung geschaffen. Diese umfasst „die Evaluation medizinischer Interventionen hinsichtlich ihrer kausal begründeten positiven und negativen Effekte im Vergleich mit einer klar definierten anderen Therapie, einem Placebo (oder einer andersartigen Scheinbehandlung) oder keiner Behandlung“.²¹² Zentrale Akteure der Nutzenbewertung sind der Gemeinsame Bundesausschuss (G-BA) sowie das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG).²¹³

Der G-BA ist das oberste Beschlussgremium der Gemeinsamen Selbstverwaltung von Ärzten, Krankenhäusern und Krankenkassen.²¹⁴ Während der Gesetzgeber den gesundheitspolitischen Rahmen vorgibt, obliegt dem G-BA dessen Ausfüllung im Rahmen des fünften

²⁰⁷ Vgl. DIMDI (2011), S. 53.

²⁰⁸ Vgl. Johnston et al. (2013), S. 7.

²⁰⁹ Vgl. Walters (2009), S. 29.

²¹⁰ Vgl. Ebd.

²¹¹ Vgl. Lange et al. (2010).

²¹² IQWiG (2013a), S. 32.

²¹³ Vgl. Gibis (2008), S. 27-29.

²¹⁴ Vgl. Schwalm/Perleth/Matthias (2010), S. 324.

Sozialgesetzbuches (SGB V).²¹⁵ Hierzu verabschiedet der G-BA Richtlinien, die den Charakter untergesetzlicher Normen haben und für Kostenträger, Leistungserbringer und die gesetzlich Versicherten verbindlich sind.²¹⁶ Sämtliche Beschlüsse des G-BA unterliegen der Rechtsaufsicht des Bundesministeriums für Gesundheit (BMG) und treten erst im Falle dessen Nichtbeanstandung in Kraft.²¹⁷ Die zentralste Aufgabe des G-BA besteht in der Definition des Leistungskatalogs für die gesetzlich Versicherten in Deutschland.²¹⁸

Das IQWiG wurde mit der Gesundheitsreform 2004 gegründet und stellt ein fachlich unabhängiges wissenschaftliches Institut dar.²¹⁹ Seine Hauptaufgabe besteht in der Bewertung des Nutzens sowie der Qualität und Wirtschaftlichkeit von medizinischen Leistungen.²²⁰

Mit dem AMNOG erfolgte die Verpflichtung von pU, dem G-BA vor Markteinführung ein *Value Dossier* (VD) vorzulegen. Gemäß § 5 Abs. 4 Satz 1 der Arzneimittel-Nutzenbewertungsverordnung (AM-NutzenV) ist darin darzulegen, mit welcher Wahrscheinlichkeit und in welchem Ausmaß das neue Arzneimittel gegenüber der gegenwärtigen Standardtherapie einen Zusatznutzen (ZN) aufweist. Die Festlegung dieser sog. zweckmäßigen Vergleichstherapie (ZVT) erfolgt durch den G-BA.²²¹ Innerhalb von drei Monaten erfolgt eine Bewertung des VD, mit der der G-BA in der Regel das IQWiG beauftragt.²²² Weitere drei Monate später trifft der G-BA auf Basis des eingereichten VD, der Empfehlungen des IQWiG und einer öffentlichen Stellungnahme eine Entscheidung über die Art und das Ausmaß des ZN (oder Schadens) anhand sog. patientenrelevanter Endpunkte.²²³

Die Verankerung patientenrelevanter Endpunkte im SGB V geht zurück auf das Gesetz zur Verbesserung der Wirtschaftlichkeit in der Arzneimittelversorgung (AVWG) vom 26. April 2006.²²⁴ Synonym werden auch die Bezeichnungen patientenrelevante Zielgrößen sowie patientenbezogene Endpunkte verwendet.²²⁵ In Ermangelung einer Konkretisierung innerhalb des Gesetzestextes definiert das IQWiG als patientenrelevant „*wie ein Patient fühlt, seine Funktionen und Aktivitäten wahrnehmen kann oder ob er überlebt.*“²²⁶ Gemäß § 35b SGB V sollen als Patientennutzen die Verbesserung des Gesundheitszustandes, eine Verkürzung der Krankheitsdauer, eine Verlängerung der Lebensdauer, eine Verringerung der Nebenwirkungen sowie eine Verbesserung der Lebensqualität angemessen berücksichtigt werden.

²¹⁵ Vgl. Ebd.

²¹⁶ Vgl. Ebd.

²¹⁷ Vgl. Ebd.

²¹⁸ Vgl. Mühlbauer/Teupen (2014), S. 147.

²¹⁹ Vgl. Ebd., S. 149.

²²⁰ Vgl. Ebd.

²²¹ Vgl. Ruof et al. (2013), S. 2.

²²² Vgl. Ebd.

²²³ Vgl. Ebd.

²²⁴ Vgl. Matthias (2010), S. 273.

²²⁵ Vgl. Ebd., S. 274.

²²⁶ IQWiG (2013a), S. 30.

Hieraus ergibt sich zunächst ein scheinbarer Widerspruch zu § 34 Abs. 1 Satz 7 und 8 SGB V, der eine Kostenerstattung für Präparate, „*bei denen eine Erhöhung der Lebensqualität im Vordergrund steht*“ im Rahmen des Sozialversicherungssystems ausschließt. Gemeint sind hiermit jedoch sog. *Lifestyle-Präparate*, bspw. zur Behandlung der erektilen Dysfunktion, der Abmagerung oder der Raucherentwöhnung.²²⁷

Aus den Nutzendimensionen gemäß § 35b SGB V ergeben sich folgende patientenrelevante Zielgrößen:

1. Mortalität,
2. Morbidität (Beschwerden und Komplikationen),
3. HRQoL

Darüber hinaus können der interventions- und erkrankungsbezogene Aufwand und die Zufriedenheit der Patientinnen und Patienten mit der Behandlung berücksichtigt werden, ein ZN lässt sich aus diesen Zielgrößen alleine jedoch nicht ableiten.²²⁸

In seinem Methodenpapier weist das IQWiG darauf hin, dass zur Beurteilung des ZN eine Zusammenschau der Zielgrößen nötig ist.²²⁹ Vor dem jeweiligen indikationsspezifischen Hintergrund wird eine Hierarchisierung der Endpunkte vorgenommen.²³⁰ Der Nachweis eines Überlebenszeitgewinns wird vom IQWiG in der Regel als essentielle Voraussetzung eines wie auch immer gearteten ZN angesehen.²³¹ Eine gesetzliche Grundlage bietet das SGB V hierfür nicht, eine Rangfolge patientenrelevanter Endpunkte ergibt sich daraus allenfalls implizit. Auch einschlägige Rechtsprechung existiert hierzu bisher nicht.

Zur Quantifizierung des Ausmaßes des ZN definiert § 5 Abs. 7 der AM-NutzenV sechs Kategorien: (1) erheblicher ZN, (2) beträchtlicher ZN, (3) geringer ZN, (4) nicht quantifizierbarer ZN, (5) kein ZN belegt, (6) geringerer Nutzen. Um eine konsistente Grundlage für die Einstufung des ZN-Ausmaßes zu schaffen, erfolgte für die ersten drei Kategorien eine Operationalisierung durch das IQWiG.

²²⁷ Vgl. Klakow-Franck (2014), S. 152.

²²⁸ Vgl. IQWiG (2013a), S.31.

²²⁹ Vgl. Ebd.

²³⁰ Vgl. Ebd., S. 32.

²³¹ Vgl. Klakow-Frank (2014), S. 153.

		Zielgrößenkategorie			
		Gesamtmortalität	Schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen	HRQoL	Nicht schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen
Ausmaßkategorie	Erheblich	Erhebliche Verlängerung der Überlebensdauer	Langfristige Freiheit bzw. eitestgehende Vermeidung	<i>Erhebliche Verbesserung</i>	<i>Nicht besetzt</i>
	Beträchtlich	Moderate Verlängerung der Überlebensdauer	Abschwächung bzw. relevante Vermeidung	<i>Bedeutsame Verbesserung</i>	Bedeutsame Vermeidung
	Gering	Jegliche Verlängerung der Überlebensdauer	<i>Jegliche Verringerung</i>	<i>Relevante Verbesserung</i>	Relevante Vermeidung

Tabelle 2: IQWiG-Operationalisierung patientenrelevanter Endpunkte gemäß AM-NutzenV, eigene Darstellung in Anlehnung an IQWiG (2013a), S. 154.

Ergänzungen gegenüber der AM-NutzenV sind in Tabelle 2 kursiv dargestellt. Eine Schwierigkeit der Operationalisierung ergibt sich aus dem Umstand, dass die AM-NutzenV HRQOL nicht als Zielgrößenkategorie zur Bestimmung des Ausmaßes des ZN vorsieht. Das IQWiG ordnet sie gleichbedeutend wie schwerwiegende (bzw. schwere) Symptome, Folgekomplikationen und Nebenwirkungen ein.²³²

Grundsätzlich gelten für den Nachweis eines HRQoL-spezifischen ZN die gleichen Prinzipien wie für andere Endpunkte. So fordert das IQWiG den Nachweis entsprechender Effektstärken.²³³ Zudem sind randomisierte kontrollierte Studien (engl. *Randomized Controlled Trials*, RCT) am besten für den Nachweis eines Effekts geeignet.²³⁴ Als Goldstandard gilt die *double-blind* RCT, bei der die Zuteilung in Behandlungs- und Kontrollgruppe zufällig und für den Patienten sowie den Behandler verdeckt erfolgt.²³⁵ Bei adäquater Umsetzung sollten sich die Patienten in der Behandlungs- und Kontrollgruppe somit allein hinsichtlich der Therapie unterscheiden.

²³² Vgl. IQWiG (2013a), S. 152, 158.

²³³ Vgl. Ebd.

²³⁴ Vgl. Ebd., S. 46.

²³⁵ Vgl. Walters (2009), S. 181f.

Das IQWiG-Methodenpapier fordert darüber hinaus, dass zur Erfassung der HRQoL nur Instrumente eingesetzt werden sollen, die für klinische Studien geeignet und entsprechend validiert sind.²³⁶ Zudem sind die Vorgaben zur Vermeidung von Verzerrungen (engl. *Bias*) einzuhalten, wozu insbesondere ein statistischer Analyseplan, Angaben zur Durchführung der Erhebungen, zu den Methoden der Berechnung von Skalenwerten, zum Umgang mit fehlenden Daten und den statistischen Analyseverfahren a priori im Studienprotokoll festzulegen sind.²³⁷ Ergänzend hierzu fordert das IQWiG die Verwendung eines validierten Responsekriteriums für das jeweilige HRQoL-Messinstrument, was neben der statistischen Signifikanz als zusätzliche Hürde hinsichtlich der Feststellung eines endpunktspezifischen ZN zu werten ist.²³⁸

Eine weitere Herausforderung stellen die unterschiedlichen Anforderungen dar, die Zulassungs- und nationale HTA-Behörden an die einzureichenden Daten stellen. Im Rahmen der Zulassung liegt der Fokus auf der Feststellung der Wirksamkeit eines neuen Arzneimittels, die Beweisführung des pU erfolgt anhand der vorab definierten, primären Endpunkte.²³⁹ Laut FDA (2009) kann HRQoL prinzipiell sowohl als primärer als auch als sekundärer Endpunkt erhoben werden.²⁴⁰ Jedoch wird gefordert, der primäre Endpunkt solle „*the most clinically relevant and convincing evidence*“ darstellen.²⁴¹ Der Leitfaden der EMA (2005) zur Durchführung klinischer Studien im Bereich onkologischer Präparate definiert progressionsfreies Überleben (engl. *Progression Free Survival*, PFS), krankheitsfreies Überleben (engl. *Disease Free Survival*, DFS) und Gesamtüberleben (engl. *Overall Survival*, OS) als obligatorische primäre Endpunkte und gesteht HRQoL allenfalls einen zusätzlichen, informativen Mehrwert als sekundäre Zielgröße zu.²⁴²

Während auf Ebene der Zulassung folglich noch Vorbehalte zu bestehen scheinen, HRQoL als gleichwertigen Wirksamkeitsparameter anzuerkennen, spitzt sich im Rahmen des deutschen HTA-Diskurses die Lage zu einer Einforderung von HRQoL-Daten zu. Laut Klakow-Franck (2014) sind HRQoL-Daten insbesondere in der immer individueller gestalteten, onkologischen Behandlung zwingend erforderlich, um Patientenpräferenzen adäquat berücksichtigen zu können.²⁴³ Gefordert wird eine systematische Verankerung der HRQoL als primärer und sekundärer Endpunkt.²⁴⁴

²³⁶ Vgl. IQWiG (2013a), S. 31.

²³⁷ Vgl. Kohlmann (2010), S. 262.

²³⁸ Vgl. IQWiG (2013a), S. 53.

²³⁹ Vgl. Leverkus (2014), S. 113.

²⁴⁰ Vgl. FDA (2009), S. 3f.

²⁴¹ ICH (1998), S. 5.

²⁴² EMA (2013), S. 23.

²⁴³ Vgl. Klakow-Franck (2014), S. 155.

²⁴⁴ Vgl. Ebd., S. 156.

5 Fragestellung

Wie die vorangegangenen Ausführungen zeigen, sind die methodischen Anforderungen, die hinsichtlich der Erhebung von PRO wie der HRQoL im Rahmen der frühen Nutzenbewertung gelten, durchaus als hoch zu bewerten. Eine weitere Herausforderung ergibt sich durch die unterschiedlichen Vorgaben, die seitens der Zulassungs- und der nationalen HTA-Behörden an die einzureichenden Studiendaten bestehen. In einer Zwischenbilanz im November 2013 sprach der G-BA von einem „Flaschenhalsproblem“ in Bezug auf die Berücksichtigung von HRQoL-Daten in der frühen Nutzenbewertung.²⁴⁵ Trotz einer erkennbaren Steigerung des Forschungsinteresses und der Erhebung des Endpunktes HRQoL in klinischen Studien seien die bisher im Rahmen der Nutzenbewertung präsentierten Daten aufgrund methodischer Mängel nicht oder nur eingeschränkt verwertbar.²⁴⁶

Diese Einschätzung deckt sich mit den Ergebnissen von Ruof et al. (2012).²⁴⁷ In einer Analyse der ersten 18 Monate der frühen Nutzenbewertung stellten die Autoren fest, dass in keinem der 27 abgeschlossenen Verfahren ein HRQoL-spezifischer ZN ausgesprochen worden war.²⁴⁸ Offen blieb im Zusammenhang mit dieser Feststellung jedoch, in welchem Umfang HRQoL-Daten bereits durch pU eingereicht worden waren. Darüber hinaus stellten die Autoren erhebliche Differenzen bezüglich des durch den pU beanspruchten und im Rahmen der Nutzenbewertung attestierten ZN fest.²⁴⁹

Ein ähnliches Bild förderte eine Analyse von Klose et al. (2013) zutage, deren Beobachtungszeitraum die ersten 30 Monate der Nutzenbewertung umfasste.²⁵⁰ In den nunmehr 43 abgeschlossenen Verfahren ließen sich lediglich drei Fälle identifizieren, in denen ein HRQoL-spezifischer ZN ausgesprochen worden war.²⁵¹ Angaben zur HRQoL fanden sich jedoch in 25 Dossiers, wovon herstellerseitig in 12 Fällen ein ZN geltend gemacht worden war.²⁵²

Ziel der vorliegenden Arbeit ist die Bestimmung des aktuellen Stellenwerts eines der zentralsten PRO Konzepte – der HRQoL – im Kontext der frühen Nutzenbewertung in Deutschland. Die Arbeit konzentriert sich hierzu auf die Erhebung von HRQoL im Kontext von klinischen Interventionsstudien und geht folgenden Fragestellungen nach:

1. Inwiefern kommen pU bereits der Anforderung gemäß §35b SGB V nach, HRQoL-Daten zu erheben und im Rahmen der Nutzenbewertung einzureichen?
2. Wurden diese Daten in der Nutzenbewertung berücksichtigt und ließ sich aus ihnen ein endpunktspezifischer ZN ableiten?

²⁴⁵ Vgl. IQWiG (2013a).

²⁴⁶ Vgl. Klakow-Franck (2014), S. 155.

²⁴⁷ Vgl. Ruof et al. (2012).

²⁴⁸ Vgl. Ebd., S. 7.

²⁴⁹ Vgl. Ebd., S. 12.

²⁵⁰ Vgl. Klose et al. (2013).

²⁵¹ Vgl. Ebd., S. 14.e1.

²⁵² Vgl. Ebd.

Fragestellung

- a. Falls ja – welche Daten wurden berücksichtigt und lassen sich gegebenenfalls *Best-Practice*-Strategien ableiten?
- b. Falls nein – welche Ablehnungsgründe werden genannt? Lassen sich gegebenenfalls strukturelle Mängel der eingereichten HRQoL-Daten identifizieren?

6 Methodik

Im Rahmen der vorliegenden Arbeit wurden sämtliche abgeschlossenen Verfahren gemäß § 35a SGB V im Zeitraum vom 1. Januar 2011 bis 31. August 2013 ausgewertet. In die Analyse wurden diejenigen Verfahren eingeschlossen, in denen der pU im VD eigene Studien zu dem patientenrelevanten Endpunkt HRQoL einreichte. Basis für die Identifikation relevanter Verfahren bildete die unter Kapitel 2.3 - Gesundheitsbezogene Lebensqualität (*Health Related Quality of Life, HRQoL*) vorgestellte Begriffsdefinition. Berücksichtigt wurden demzufolge sowohl Profil- als auch Indexinstrumente. Aus der Betrachtung ausgeschlossen waren hingegen eindimensionale Symptomskalen, wie beispielsweise die *Functional Assessment of Chronic Illness Therapy-Fatigue-Scale*.

6.1 Informationsbeschaffung und -aufbereitung

Grundlage für die Analyse der Verfahren bildeten das Modul 4 des eingereichten VD des pU, die Dossierbewertung des IQWiG sowie der G-BA-Beschluss (Bundesanzeiger (BAnz), Tragende Gründe (TrG), Zusammenfassende Dokumentation (ZD)). Sämtliche Unterlagen sowie Steckbriefe und Fristen der Verfahren wurden über die Website des G-BA (<http://www.g-ba.de/>) bezogen und in eine Übersichtsmatrix überführt [siehe Anhang A].

Wirkstoff, Handelsname, Indikation, Datum der Beschlussfassung und pU wurden dem jeweiligen Verfahrenssteckbrief auf der Website des G-BA unter der Rubrik *Übersicht der Wirkstoffe* (<http://www.g-ba.de/informationen/nutzenbewertung/>) entnommen.

ZN-Wahrscheinlichkeit und Ausmaß wurden den Angaben des G-BA im BAnz entnommen.

Die Identifikation von Studien des pU, die HRQoL-Daten enthalten, stützte sich auf die Angaben des pU im VD unter Punkt

4.3.1.2 *Charakteristika der in die Bewertung eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel,*

4.3.2.1.2 *Charakteristika der Studien für indirekte Vergleiche,*

4.3.2.2.2 *Charakteristika der nicht randomisierten vergleichenden Studien,*

4.3.2.3.2 *Charakteristika der weiteren Untersuchungen.*

Die Feststellung der Studienkategorie zielte auf die Frage, ob es sich um eine Zulassungsstudie handelt. Sie erfolgte auf Basis der Angaben des pU im VD unter Punkt

4.3.1.1.4 *Resultierender Studienpool: RCT mit dem zu bewertenden Arzneimittel,*

4.3.2.1.1 *Ergebnis der Informationsbeschaffung – Studien für indirekte Vergleiche,*

4.3.2.2.1 *Ergebnis der Informationsbeschaffung – nicht randomisierte vergleichende Studien,*

4.3.2.3.1 *Ergebnis der Informationsbeschaffung – weitere Untersuchungen.* Folgende Merkmalsausprägungen waren für die Feststellung der Studienkategorie vorgesehen: *ja, nein*.

Die Feststellung der Studienendpunkte sowie der Endpunktebene erfolgte auf Basis der Angaben des pU im VD unter Punkt 4.3.1.3.1 *Endpunkte – RCT*, 4.3.2.1.3.1 *Endpunkte – indirekte Vergleiche*

aus RCT und 4.3.2.2.3.1 Endpunkte – nicht randomisierte vergleichende Studien, 4.3.2.3.3.1 Endpunkt – weitere Untersuchungen. Folgende Merkmalsausprägungen waren für die Feststellung der Endpunktebene vorgesehen: *primär/co-primär, sekundär, tertiär, explorativ, nicht näher spezifiziert*. Die Endpunktebene wurde als *nicht näher spezifiziert* eingestuft, sofern aus dem VD des pU die Einteilung bzw. Hierarchisierung des Endpunktes nicht hervorging.

Die Identifikation der in den Studien eingesetzten HRQoL-Messinstrumente stützte sich auf die Angaben des pU im VD unter Punkt 4.3.1.3.1 Endpunkte– RCT – Operationalisierung, 4.3.2.1.3.1 Endpunkte – indirekte Vergleiche aus RCT – Operationalisierung, 4.3.2.2.3.1 Endpunkte – nicht randomisierte vergleichende Studien – Operationalisierung, 4.3.2.3.3.1 Endpunkt – weitere Untersuchungen – Operationalisierung. Die Kategorisierung des Instrumententyps erfolgte unter Konsultation der PROQOLID Datenbank (<http://www.proqolid.org/>). Folgende Merkmalsausprägungen waren vorgesehen: generisch, spezifisch.

Eine abweichende Methodik wurde für Arzneimittel angewendet, die zur Behandlung eines seltenen Leidens nach der Verordnung (EG) Nr. 141/2000 des Europäischen Parlaments und des Rates vom 16. Dezember 1999 zugelassen sind. Laut 5. Kapitel § 12 Absatz 1 Satz 1 der Verfahrensordnung des G-BA gilt der ZN für diese sog. Orphan Drugs (OD) nach § 35a Absatz 1 Satz 3 Nummer 2 und 3 SGB V durch die Zulassung als belegt, sofern der Umsatz des Arzneimittels zu Apothekenverkaufspreisen einschließlich Umsatzsteuer in den letzten zwölf Kalendermonaten einen Betrag von 50 Millionen Euro nicht übersteigt.²⁵³ Nachweise gemäß 5. Kapitel § 5 Absatz 1 bis 6 müssen nicht vorgelegt werden, sondern das Ausmaß des ZN ist nur für die Anzahl der Patienten und Patientengruppen, für die ein therapeutisch bedeutsamer ZN besteht, nachzuweisen.²⁵⁴ Die Ermittlung der aufgeführten Bestandteile der Übersichtsmatrix stütze sich für OD daher auf die obligatorischen Angaben des pU unter Punkt 4.4.4 Angabe der Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht – Arzneimittel zur Behandlung eines seltenen Leidens sowie auf freiwillige Angaben des pU im VD.

Der Berücksichtigungsstand der eingereichten HRQoL-Daten in der Dossierbewertung wurde auf Basis der Angaben des IQWiG-Gutachtens unter Punkt 2.4 Ergebnisse zum Zusatznutzen – Matrix der Endpunkte & Datenverfügbarkeit ermittelt. Folgende Merkmalsausprägungen waren vorgesehen: ja, berücksichtigt, aber kein patientenrelevanter ZN, nein. Der Berücksichtigungsstand wurde mit ja angegeben, wenn die prinzipielle Datenverfügbarkeit seitens des IQWiG bestätigt und die HRQoL-Daten in der Bemessung des Ausmaßes des ZN berücksichtigt wurden. Der Berücksichtigungsstand wurde mit berücksichtigt, aber kein patientenrelevanter ZN angegeben, wenn die prinzipielle Datenverfügbarkeit seitens des IQWiG bestätigt, die HRQoL-Daten jedoch aufgrund der diskutierten Mängel nicht in der Bemessung des ZN berücksichtigt wurden. Der Berücksichtigungsstand wurde mit nein angegeben, wenn die prinzipielle Datenverfügbarkeit seitens des IQWiG als nicht gewährleistet eingestuft wurde.

²⁵³ Vgl. G-BA (2014), S. 114f.

²⁵⁴ Vgl. Ebd.

Die Gründe für die Nichtberücksichtigung von HRQoL-Daten sowie für das Versagen eines endpunktspezifischen ZN wurden den Angaben des IQWiG-Gutachtes unter Punkt 2.7 Kommentare zum Dossier des pU entnommen.

Eine abweichende Methodik wurde für OD angewendet. Gemäß Beschlussfassung des G-BA vom 15. März 2012 hat bei OD eine Beauftragung des IQWiG mit der Durchführung einer Nutzenbewertung erst dann zu erfolgen, wenn der Umsatz des betreffenden Arzneimittels die gesetzliche Grenze von 50 Millionen Euro überschritten hat und es somit einer uneingeschränkten Nutzenbewertung unterliegt.²⁵⁵ Andernfalls bestimmt der G-BA das ZN-Ausmaß auf Grundlage der Zulassung und der die Zulassung begründenden Studien.²⁵⁶ Die Ermittlung des Berücksichtigungsstandes der HRQoL-Daten stützt sich für OD somit auf die durch den G-BA durchgeführte Dossierbewertung und die Angaben unter 2.5. Endpunkte und Erhebungsinstrumente. Folgende Merkmalsausprägungen waren vorgesehen: ja, berücksichtigt, aber kein patientenrelevanter ZN, nein. Der Berücksichtigungsstand wurde mit ja angegeben, wenn die prinzipielle Datenverfügbarkeit seitens des G-BA bestätigt und die HRQoL-Daten in der Bemessung des ZN-Ausmaßes berücksichtigt wurden. Der Berücksichtigungsstand wurde mit berücksichtigt, aber kein patientenrelevanter ZN angegeben, wenn die prinzipielle Datenverfügbarkeit seitens des G-BA bestätigt, die HRQoL-Daten jedoch aufgrund der diskutierten Mängel nicht in der Bemessung des ZN berücksichtigt wurden. Der Berücksichtigungsstand wurde mit nein angegeben, wenn die prinzipielle Datenverfügbarkeit seitens des G-BA als nicht gewährleistet eingestuft wurde.

Der Berücksichtigungsstand der eingereichten HRQoL-Daten in der Beschlussfassung des G-BA wurde auf Basis der Angaben im BAnz ermittelt. Folgende Merkmalsausprägungen waren vorgesehen: ja, berücksichtigt, aber kein patientenrelevanter ZN, nein. Der Berücksichtigungsstand wurde mit ja angegeben, wenn der Studienendpunkt und -ergebnisse im BAnz gelistet waren und HRQoL-Daten in der Bemessung des Ausmaßes des ZN berücksichtigt wurden. Der Berücksichtigungsstand wurde mit berücksichtigt, aber kein patientenrelevanter ZN angegeben, wenn der Studienendpunkt sowie die entsprechenden -ergebnisse im BAnz gelistet waren, die HRQoL-Daten jedoch aufgrund der diskutierten Mängel nicht in der Bemessung des ZN berücksichtigt wurden. Der Berücksichtigungsstand wurde mit nein angegeben, wenn entweder gar keine Erwähnung im BAnz oder allein eine Auflistung mit der Angabe keine verwertbaren Daten verfügbar erfolgte.

Die Gründe für die Nichtberücksichtigung der HRQoL-Daten in der Beschlussfassung sowie für das Versagen eines endpunktspezifischen ZN wurden den Angaben des G-BA im BAnz, in den TrG und in der ZD entnommen.

²⁵⁵ Vgl. G-BA (2012a), S. 4.

²⁵⁶ Vgl. Ebd.

6.2 Auswertung der Daten

Die Auswertung besteht in einer deskriptiven Analyse der Studien und der eingesetzten HRQoL-Messinstrumente, zu deren Beschreibung absolute sowie relative Häufigkeitsverteilungen unter Angabe des jeweiligen Stichprobenumfangs n eingesetzt wurden.²⁵⁷

Darüber hinaus wird eine Kategorisierung der geäußerten Gründe für die Nichtberücksichtigung von HRQoL-Daten sowie das Versagen eines endpunktspezifischen ZN in der Dossierbewertung und der Beschlussfassung vorgenommen. Zur Entwicklung der Kategorien wurde ein induktives Vorgehen gewählt, dass sich an der qualitativen Inhaltsanalyse nach Mayring orientiert. Ziel der qualitativen Inhaltsanalyse ist es, „Texte systematisch [zu] analysieren indem sie das Material schrittweise und mit theoriegeleitet am Material entwickelten Kategoriensystemen bearbeitet“.²⁵⁸ Laut Mayring (2010) nimmt die qualitative Inhaltsanalyse hinsichtlich der beiden Gegenpole der quantitativen und der qualitativen Analyse eine Zwischenposition ein.²⁵⁹ Ergänzend zu dem qualitativen Interpretationsvorgang und der induktiven Kategorienbildung ermöglicht sie eine quantitative Weiterverarbeitung der Ergebnisse, bspw. in Form einer Auswertung von Kategorienhäufigkeiten.²⁶⁰ Das hierzu von Mayring vorgeschlagene inhaltsanalytische Ablaufmodell umfasst die nachfolgenden acht Schritte:

1. Festlegung des Materials
2. Analyse der Entstehungssituation
3. Formale Charakterisierung des Materials
4. Festlegung der Analyserichtung
5. Theoretische Differenzierung der Fragestellung
6. Bestimmung der Analysetechnik
7. Definition der Analyseeinheiten
8. Durchführung der Materialanalyse

Die ersten drei Schritte beziehen sich auf die Bestimmung des Ausgangsmaterials.²⁶¹ Die Festlegung des Materials umfasst in der vorliegenden Arbeit die unter *Kapitel 6.1 - Informationsbeschaffung und Aufbereitung* genannten Dokumente bzw. Dokumentenabschnitte. Zur Analyse der Entstehungssituation ist festzuhalten, dass sämtliche Dokumente im Kontext der Regelungen von § 35a SGB V entstanden sind und entsprechend der damit verbundenen prozessualen Standards verfasst wurden. Im Falle des VD handelt es sich bei dem Verfasser um den pU. Die Dossierbewertung wird in der Regel durch das IQWiG durchgeführt. Eine Ausnahme stellen OD dar, für die eine Nutzenbewertung durch den G-BA vorgenommen

²⁵⁷ Vgl. Bortz/Schuster (2010), S. 40.

²⁵⁸ Mayring (2002), S. 14; zitiert nach Ramsenthaler (2013), S. 23.

²⁵⁹ Vgl. Mayring (2010), S. 8.

²⁶⁰ Vgl. Mayring (2010), S. 8.

²⁶¹ Vgl. Ramsenthaler (2013), S. 27.

wird.²⁶² Im Falle der Beschlussdokumente (BA_{nz}, TrG, ZD) handelt es sich bei dem Verfasser um den G-BA. Zu den formalen Charakteristika des Materials lässt sich feststellen, dass es sich um verschriftlichte Bewertungsberichte handelt.

Schritt 4 und 5 beziehen sich auf die Herausarbeitung der Fragestellung.²⁶³ Die im Rahmen der vorliegenden Arbeit gewählte Analyserichtung zielt auf die Beschreibung der in den Texten behandelten Nutzenbewertungsverfahren sowie auf den Textverfasser ab. Die Dokumente sollen Auskunft über die in der Dossierbewertung und Beschlussfassung identifizierten, charakteristischen Mängel geben, die zu einer Nichtberücksichtigung von HRQoL-Daten oder dem Versagen eines endpunktspezifischen ZN geführt haben. Für die theoretische Differenzierung der Fragestellung ist auf *Kapitel 2 - Historische Grundlagen und konzeptionelle Einordnung gesundheitsbezogener Lebensqualität*, *Kapitel 3 - Messung und Validierung von HRQoL* und *Kapitel 4 - HRQoL als patientenrelevanter Endpunkt in der frühen Nutzenbewertung gemäß § 35b SGB V* zu verweisen. Diese Kapitel verstehen sich als theoretische Grundlage für die in Kapitel 5 vorgestellte Fragestellung.

Die Schritte 6, 7 und 8 beziehen sich auf die Entwicklung und Durchführung des Ablaufmodells.²⁶⁴ Die Bestimmung der Analysetechnik besteht in einer Festlegung auf eine Form des Interpretierens. Mayring (2010) unterscheidet zwischen der Zusammenfassung, der Explikation und der Strukturierung.²⁶⁵ Im Rahmen der vorliegenden Arbeit wurde die Analysetechnik der Zusammenfassung gewählt. Das Grundprinzip der zusammenfassenden Inhaltsanalyse ist es, sämtliches Material zu berücksichtigen und so zu reduzieren, dass die wesentlichen Inhalte erhalten bleiben.²⁶⁶ Hierzu werden anhand des Materials gewonnene, textgeleitete Definitionskriterien verwendet, um die relevanten Textbestandteile herauszufiltern und anschließend zu Kategorien zu verdichten.²⁶⁷ Die Kontexteinheit legt den größten Textbestandteil fest, der unter eine Kategorie fallen darf.²⁶⁸ Die Kodiereinheit legt den kleinsten Materialbestandteil fest, der ausgewertet bzw. unter eine Kategorie gefasst werden darf.²⁶⁹ Die Auswertungseinheit legt die Reihenfolge fest, in der die einzelnen Textteile ausgewertet werden.²⁷⁰

²⁶² Anmerkung: Da diese Regelung erst mit Beschlussfassung zum 15. März 2012 in Kraft trat, fand für die OD Pirfenidon (Esbriet®) und Tafamidis Meglumine (Vyndaqel®) noch eine Dossierbewertung durch das IQWiG statt.

²⁶³ Vgl. Ramsenthaler (2013), S. 27f.

²⁶⁴ Vgl. Ramsenthaler (2013), S. 28f.

²⁶⁵ Vgl. Mayring (2010), S. 65.

²⁶⁶ Vgl. Ebd., S. 65f.

²⁶⁷ Vgl. Ebd., S. 66.

²⁶⁸ Vgl. Ebd., S. 59.

²⁶⁹ Vgl. Ebd.

²⁷⁰ Vgl. Ebd.

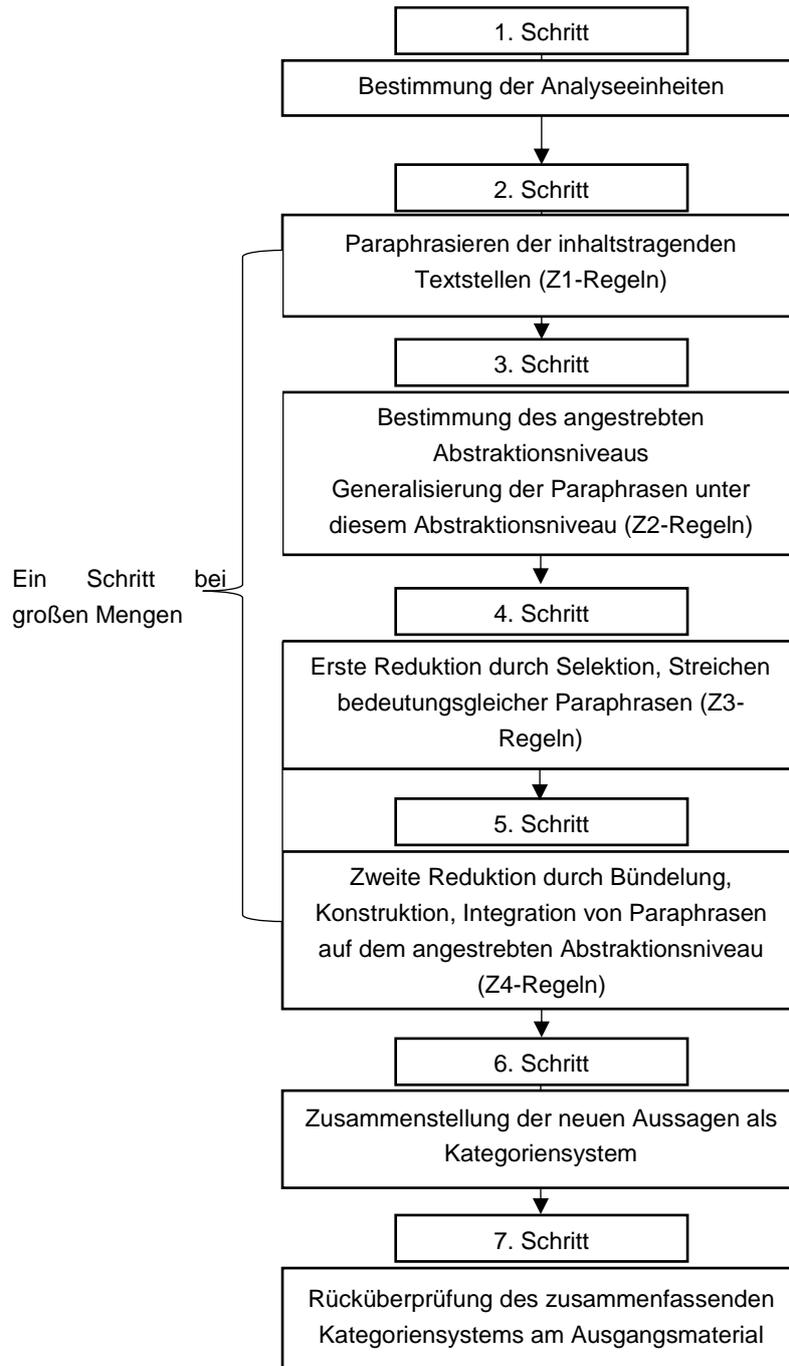


Abbildung 6: Ablaufmodell der zusammenfassenden Inhaltsanalyse nach Mayring (2010), S.68.

Relevante Textstellen waren sämtliche Stellen, an denen Gründe für die Nichtberücksichtigung von HRQoL-Daten sowie das Versagen eines endpunktspezifischen ZN durch das IQWiG oder den G-BA diskutiert wurden. Die Identifikation der Textstellen erfolgte anhand der Suchbegriffe gesundheitsbezogene Lebensqualität bzw. HRQoL sowie per Handsuche in den oben beschriebenen Datenquellen. Kodiereinheiten bildeten in diesem Zusammenhang explizit benannte Mängel der Daten und somit einzelne Sätze. Die Auswertung erfolgte entlang der durch die Gliederung bzw. Kapitel vorgegebene Textstruktur. Zunächst wurde die Dossierbewertung

und anschließend die Dokumente der Beschlussfassung in der Reihenfolge BAnz, TrG und ZD ausgewertet.

Einzelne Kodiereinheiten wurden in einer auf den Inhalt beschränkten, beschreibenden Form paraphrasiert, wobei ausschmückende Textbestandteile entsprechend der Z1-Regeln [siehe Anhang B] getilgt wurden. Anschließend wurde das Abstraktionsniveau dieser ersten Reduktion bestimmt. Sämtliche Paraphrasen, die unter diesem Niveau lagen, wurden generalisiert; Paraphrasen, die über dem Abstraktionsniveau lagen, wurden entsprechend der Z2-Regeln zunächst belassen. Anschließend wurden inhaltsgleiche, unwichtige und nichtssagende Paraphrasen entsprechend der Z3-Regeln gestrichen. In einer Zusammenschau wurden sich aufeinander beziehende Paraphrasen entsprechend der Z4-Regeln zusammengefasst und zu übergeordneten Kategorien generalisiert. Es erfolgte eine Rückkopplung an das Material, um zu prüfen, ob die derart gewonnen Kategorien das Ausgangsmaterial noch repräsentieren. Textstellen, die sich keiner Kategorie zuordnen ließen, lösten eine erneute Paraphrasierung, Bestimmung des Abstraktionsniveaus und Kategorienbildung aus. Laut Mayring (2010) wird dieser Verdichtungsprozess so lange durchlaufen, bis das Ergebnis der angestrebten Reduzierung des Materials entspricht.²⁷¹

Problematisch hierfür erwies sich die innerhalb der ersten zehn Verfahren feststellbare, geringe Begründungspraxis des G-BA. So fanden sich im Falle der Nichtberücksichtigung oder des Versagens eines endpunktspezifischen ZN innerhalb des Beschlussfassungszeitraums vom 01.01.2011-29.03.2012 in den TrG sämtlicher Verfahren entweder gar keine Angaben oder allein der Hinweis, zum Endpunkt HRQoL seien „*keine verwertbaren Daten verfügbar*“.²⁷² Um die Anwendung eines einheitlichen Kategoriensystems zu ermöglichen, wurde für die vorliegende Arbeit angenommen, dass der G-BA in diesen Fällen der Bewertung des IQWiG folgte. Die vom IQWiG angeführten Gründe für die Nichtberücksichtigung von HRQoL-Daten sowie das Versagen eines endpunktspezifischen ZN wurden daraufhin identisch für den G-BA übernommen.

Die Auswertung des Kategoriensystems umfasst die Anwendung quantitativer Analyseverfahren auf die qualitative Inhaltsanalyse in Form einer Darstellung von absoluten sowie relativen Häufigkeitsverteilungen unter Angabe des jeweiligen Stichprobenumfangs n.²⁷³

7 Ergebnisdarstellung

Es konnten 29 Verfahren von insgesamt 46 Nutzenbewertungsverfahren (63 Prozent) identifiziert werden, in denen durch den pU innerhalb des Verfahrens der frühen Nutzenbewertung gemäß § 35a SGB V im Zeitraum vom 1. Januar 2011 bis 31. August 2013 Daten zum Endpunkt HRQoL erhoben wurden [siehe Anhang C]. Im Zuge der Auswertung

²⁷¹ Vgl. Mayring (2010), S. 69.

²⁷² Vgl. G-BA (2011a), S. 11; G-BA (2012a), S. 4; G-BA (2012b), S. 8.

²⁷³ Vgl. Bortz/Schuster (2010), S. 40.

ergab sich der Sonderfall, dass in einigen Studien zwar HRQoL-Messinstrumente eingesetzt worden waren, im Rahmen des VD jedoch keine Präsentation der Ergebnisse erfolgte.

In den Verfahren Ticagrelor (Brilique[®]), Axitinib (Inlyta[®]) und Acridiniumbromid (Eklira[®], Genuair[®] / Bretaris[®], Genuair[®]) verzichtete der pU jeweils auf die Präsentation der Ergebnisse des EQ-5D. Im Falle von Ipilimumab (Yervoy[®]) wurden seitens des pU keine Ergebnisse des SF-36 präsentiert. Im Falle von Fingolimod (Gilenya[®]) legte der pU keine Ergebnisse des Primus QoL vor. Da diese HRQoL-Messinstrumente einer Bewertung nicht zugänglich waren, wurden sie der Vollständigkeit halber in die Verfahrensübersicht aufgenommen, jedoch von der weiteren Auswertung ausgeschlossen. Im Falle von Ticagrelor (Brilique[®]) führte die mangelnde Datenverfügbarkeit zum völligen Ausschluss des Verfahrens aus der Auswertung, da der pU neben dem Verzicht auf die Präsentation der Ergebnisse des EQ-5D auch keine alternativen HRQoL-Daten im VD präsentierte. Somit ging in die Auswertung eine Gesamtzahl von 28 Verfahren ein.

7.1 Studiencharakteristika und Messinstrumentarium

In den 28 identifizierten Verfahren wurden insgesamt 60 Studien eingereicht, die HRQoL-Daten enthielten, d.h. durchschnittlich wurden pro Verfahren 2,2 Studien eingereicht, die den Endpunkt HRQoL einschlossen. Hiervon waren 58 Zulassungsstudien (96,7 Prozent), lediglich zwei Studien (3,3 Prozent) stellten keine Zulassungsstudien dar.

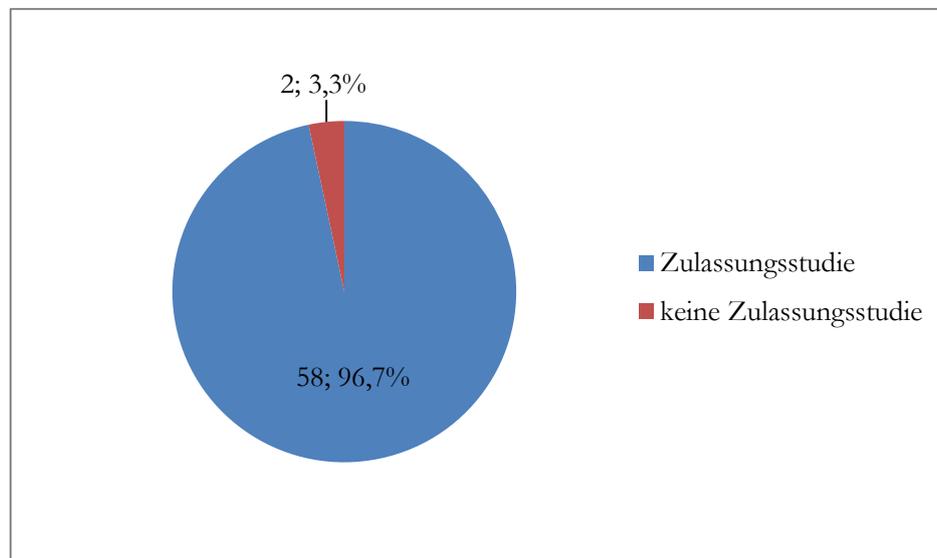


Abbildung 7: Studiencharakteristika - Studienkategorie (n = 60), eigene Darstellung.

In einer Studie (1,7 Prozent) wurde der Endpunkt HRQoL co-primär erhoben, in 40 Studien (66,7 Prozent) wurde der Endpunkt HRQoL sekundär erhoben, in zwei Studien (3,3 Prozent) wurde der Endpunkt HRQoL tertiär erhoben sowie in vier Studien (6,7 Prozent) explorativ. In 13 Studien (21,7 Prozent) wurde die Endpunktebene nicht näher spezifiziert (n.n.s.).

Ergebnisdarstellung

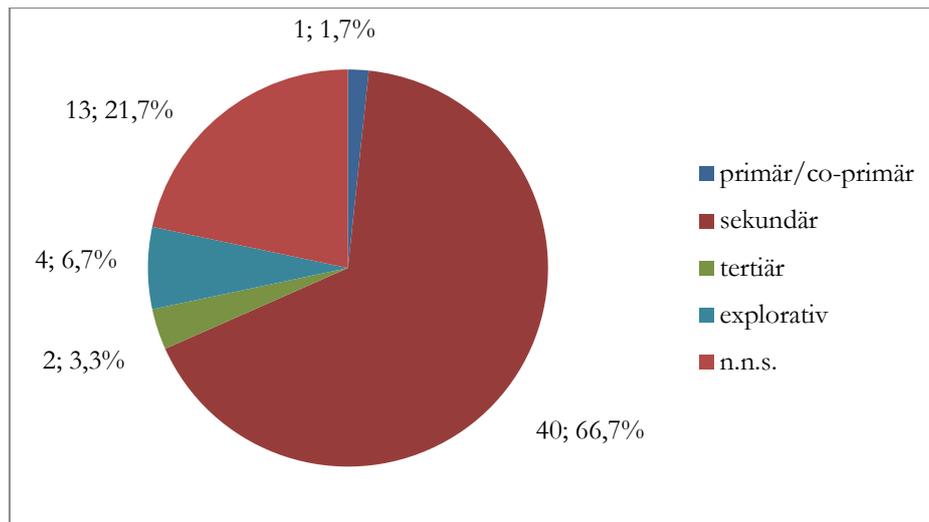


Abbildung 8: Studiencharakteristika - Endpunktebene (n = 60), eigene Darstellung.

In den 28 identifizierten Verfahren kamen 40 Instrumente zum Einsatz, d.h. durchschnittlich wurden pro Verfahren 1,4 HRQoL-Messinstrumente eingesetzt. Nach Abzug der Dopplungen ließen sich 20 unterschiedliche HRQoL-Instrumente identifizieren.

In neun Verfahren (32,1 Prozent) kamen ausschließlich generische Instrumente zum Einsatz, in 14 Verfahren (46,4 Prozent) kamen spezifische Instrumente zum Einsatz. In fünf Verfahren (21,4 Prozent) wurde eine Kombination aus beiden Instrumententypen eingesetzt.

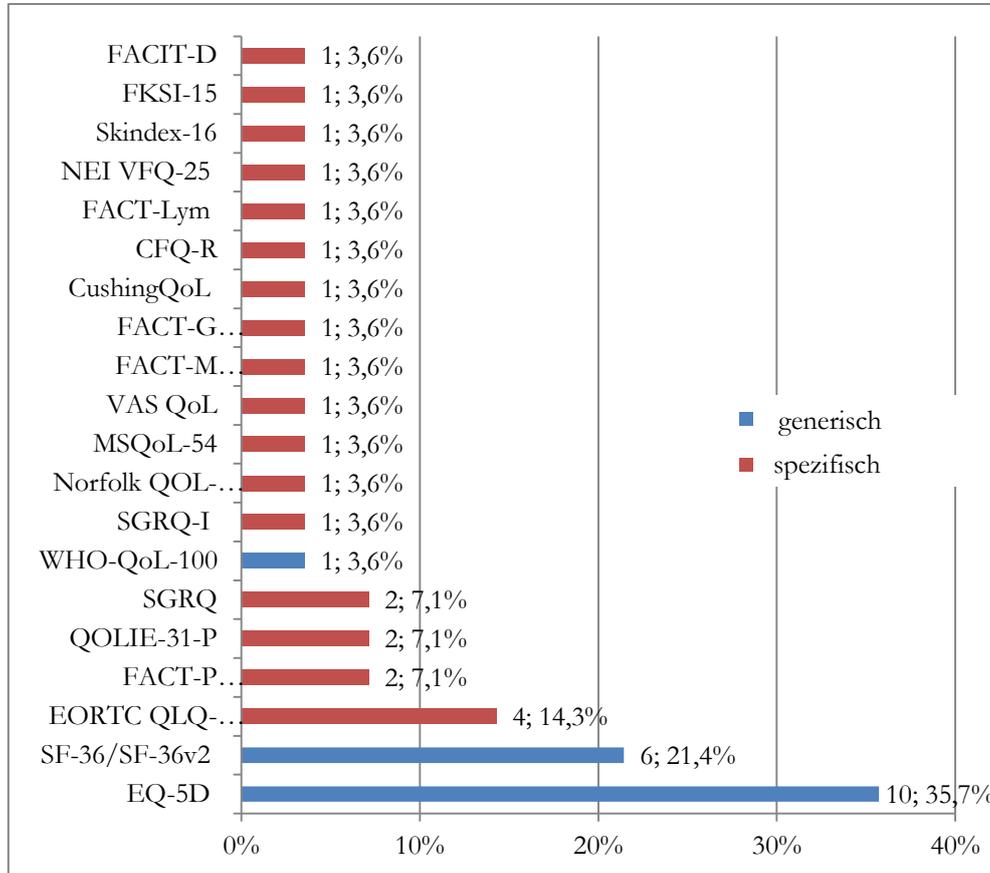


Abbildung 9: Prozentualer Anteil der verwendeten HRQoL-Messinstrumente in den Verfahren (n = 28), eigene Darstellung.

Abbildung 9 zeigt den prozentualen Anteil der verwendeten HRQoL-Messinstrumente in den ausgewerteten Verfahren. Mit Abstand am häufigsten eingesetzt wurde der EQ-5D, zu dem in zehn Verfahren (35,7 Prozent) Ergebnisse präsentiert wurden. Darauf folgt der SF-36/SF-36v2, zu dem in sechs Verfahren (21,4 Prozent) Ergebnisse präsentiert wurden. Der EORTC QLQ-C30 wurde in vier Verfahren (14,3 Prozent) eingesetzt. In jeweils zwei Verfahren (7,1 Prozent) kamen der FACT-P, der QOLIE-31-P sowie der SGRQ zum Einsatz. Jeweils nur in einem Verfahren (3,6 Prozent) wurden der WHO-QoL-100, der SGRQ-I, der Norfolk QoL-DN, der MSQoL-54, die VAS QoL, der FACT-M, der FACT-G, der ChushingQoL, der CFQ-R, der FACT-Lym, der NEI VFQ-25, der Skindex-16, der FKSI-15 sowie die FACIT-D eingesetzt.

Ergebnis:

- Die Nutzenbewertung basiert hauptsächlich auf Zulassungsstudien.
- Der Endpunkt HRQoL wird mehrheitlich sekundär erhoben.
- Es lässt sich eine breite Streuung der verwendeten HRQoL-Messinstrumente feststellen.
- Am häufigsten wurden spezifische Instrumente zur Erhebung der HRQoL eingesetzt.
- Das generische Instrument EQ-5D wurde am häufigsten eingesetzt.

7.2 Berücksichtigungsstand von HRQoL-Daten in der Nutzenbewertung

Den 28 im Beobachtungszeitraum identifizierten Verfahren, die Ergebnisse zum Endpunkt HRQoL präsentierten, steht eine Gesamtzahl von 46 innerhalb des Beobachtungszeitraums abgeschlossenen Verfahren gegenüber [siehe Anhang B]. Hieraus ergibt sich eine HRQoL-Einreichungsquote von 61 Prozent.

Eine Nichtberücksichtigung von HRQoL-Daten in der Dossierbewertung ließ sich in 13 Verfahren (46,4 Prozent) feststellen. Berücksichtigt in der Dossierbewertung, jedoch ohne Feststellung eines endpunktspezifischen ZN, wurden HRQoL-Daten in 13 Verfahren (46,4 Prozent). Mit einem endpunktspezifischen ZN berücksichtigt wurden HRQoL-Daten in der Dossierbewertung der Verfahren Ivacaftor (Kalydeco Kalydeco™) und Crizotinib (Xalkori®) (7,1 Prozent).

Eine Nichtberücksichtigung von HRQoL-Daten in der Beschlussfassung ließ sich in zwölf Verfahren (42,9 Prozent) feststellen. Berücksichtigt in der Beschlussfassung, jedoch ohne Feststellung eines endpunktspezifischen ZN, wurden HRQoL-Daten in 14 Verfahren (50 Prozent). Mit einem endpunktspezifischen ZN berücksichtigt wurden HRQoL-Daten in der Beschlussfassung der Verfahren Ivacaftor (Kalydeco Kalydeco™) und Crizotinib (Xalkori®) (7,1 Prozent).

Ergebnisdarstellung

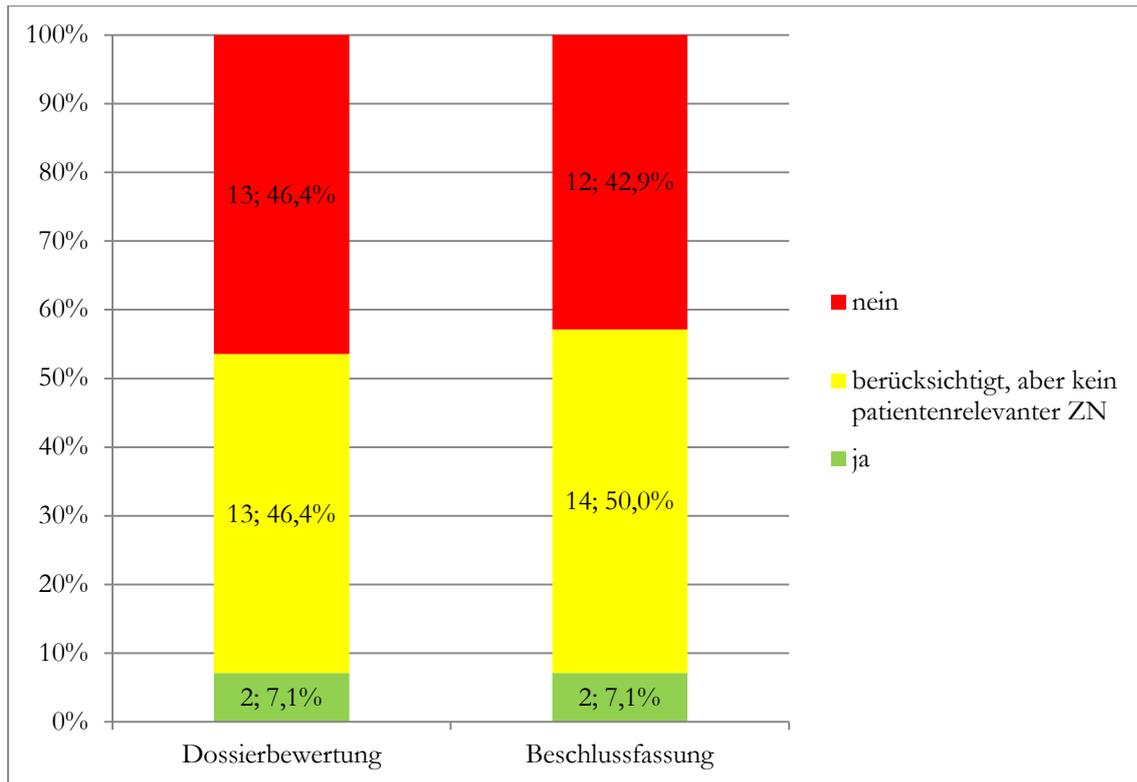


Abbildung 10: Berücksichtigungsstand von HRQoL-Daten in der Nutzenbewertung (n=28), eigene Darstellung.

Ergebnis:

- Seitens pU werden bereits mehrheitlich Daten zu dem patientenrelevanten Endpunkt HRQoL erhoben und im Rahmen der Nutzenbewertung eingereicht.
- Hinsichtlich des Berücksichtigungsstandes von HRQoL-Daten in der Dossierbewertung und der Beschlussfassung zeigt sich tendenziell ein homogenes Bild.
- In knapp der Hälfte aller Verfahren wurden HRQoL-Daten nicht berücksichtigt.
- In der Beschlussfassung wurden HRQoL-Daten geringfügig häufiger berücksichtigt, jedoch leitet sich hieraus kein zusätzlicher ZN ab.
- Im Verhältnis zur Einreichungsquote ist der in der Dossierbewertung und der Beschlussfassung zuerkannte ZN auf Endpunktebene als extrem niedrig zu bewerten.

7.3 Das Kategoriensystem der Gründe für die Nichtberücksichtigung und das Versagen eines HRQoL-spezifischen ZN

Das Kategoriensystem ermöglicht einen Überblick über die in der Dossierbewertung und der Beschlussfassung geäußerten Gründe, weshalb die vorgelegten HRQoL-Daten keine Berücksichtigung fanden, oder daraus kein patientenrelevanter ZN ableitbar war.

Diese Gründe ließen sich den Oberkategorien *Mängel im Studiendesign*, *Mängel des Erhebungsinstrumentes*, *Mängel in der Auswertung*, *mangelnde Signifikanz/Relevanz der Daten* und *formale Mängel* zuordnen. Zur verfeinerten Betrachtung erfolgt für die einzelnen Kategorien eine weitere Aufschlüsselung der Gründe [siehe Anhang D].

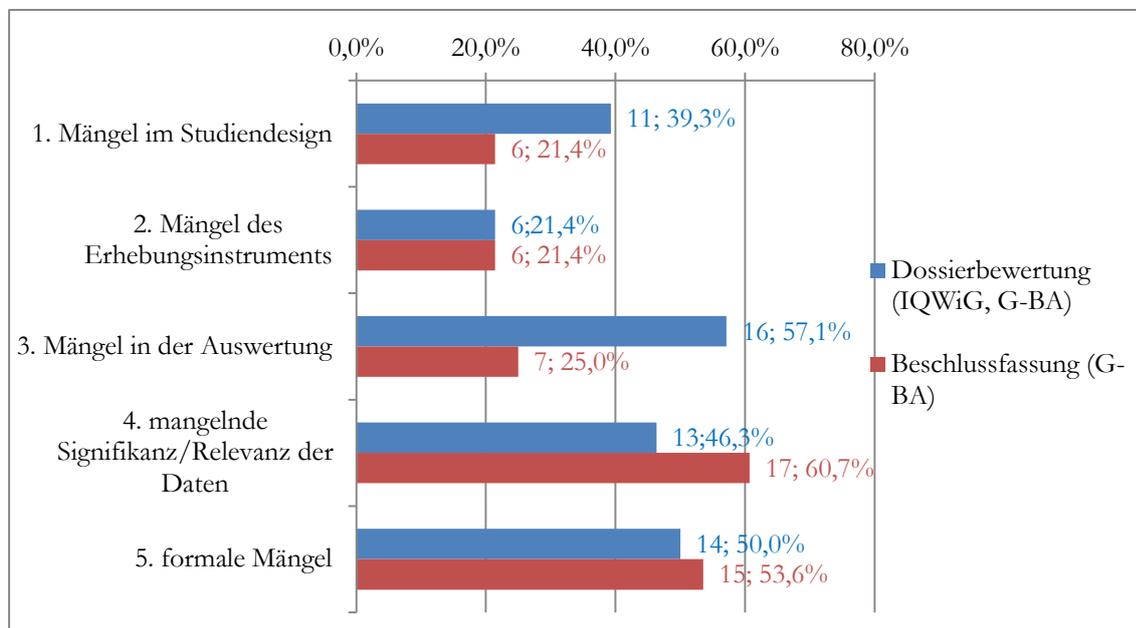


Abbildung 11: Kategoriensystem der Gründe für die Nichtberücksichtigung und das Versagen eines HRQoL-spezifischen ZN (n = 28), eigene Darstellung.

Mängel im Studiendesign wurden in der Dossierbewertung in elf Verfahren (39,3 Prozent) sowie der Beschlussfassung in 6 Verfahren (21,4 Prozent) festgestellt. Mängel im Studiendesign ließen sich im Falle der Dossierbewertung in *unvollständige Dokumentation und Präsentation der Daten* (7,1 Prozent), *systematische Unterschiede zwischen den Behandlungsgruppen* (3,6 Prozent) sowie *hohes Verzerrungspotenzial* (28,6 Prozent) unterscheiden. Die Unterkategorie *hohes Verzerrungspotenzial* ließ sich weiter differenzieren in *hohes Verzerrungspotenzial aufgrund mangelnder Verblindung* (17,9 Prozent), *hohes Verzerrungspotenzial aufgrund hoher Rate an Studienabbrechern* (3,6 Prozent), *hohes Verzerrungspotenzial aufgrund unzureichender Darstellung des Umgangs mit Lost-to-follow-up Patienten und Protokollverletzungen* (3,6 Prozent) und *hohes Verzerrungspotenzial aufgrund informativer Zensurierung* (3,6 Prozent). Im Falle der Beschlussfassung ließ sich eine weitere Differenzierung in die Unterkategorien *unvollständige Dokumentation und Präsentation der Daten* (7,1 Prozent), *systematische Unterschiede zwischen den Behandlungsgruppen* (3,6 Prozent) sowie *hohes Verzerrungspotenzial* (10,7

Prozent) unterscheiden. Die Unterkategorie *hohes Verzerrungspotenzial* ließ sich weiter differenzieren in *hohes Verzerrungspotenzial aufgrund mangelnder Verblindung* (3,6 Prozent), *hohes Verzerrungspotenzial aufgrund hoher Rate an Studienabbruchern* (3,6 Prozent), *hohes Verzerrungspotenzial aufgrund unzureichender Darstellung des Umgangs mit Lost-to-follow-up Patienten und Protokollverletzern* (3,6 Prozent).

Mängel hinsichtlich der verwendeten Erhebungsinstrumente wurden in jeweils sechs Verfahren (21,4 Prozent) in der Dossierbewertung und der Beschlussfassung festgestellt. Die verfeinerte Betrachtung dieser Kategorie zeigt, dass in der Dossierbewertung in zwei Verfahren (7,1 Prozent) sowie in der Beschlussfassung in vier Verfahren (10,7 Prozent) eine *zu geringe Rücklaufquote des Erhebungsinstruments* bemängelt wurde. Eine *mangelnde Validierung des Erhebungsinstruments* wurde in der Dossierbewertung in vier Verfahren (14,3 Prozent) sowie in der Beschlussfassung in drei Verfahren (10,7 Prozent) bemängelt.

Mängel in der Auswertung der Daten wurden in der Dossierbewertung in 16 Verfahren (57,1 Prozent) sowie in der Beschlussfassung in sieben Verfahren (25 Prozent) kritisiert. Die verfeinerte Betrachtung dieser Kategorie zeigt, dass in der Dossierbewertung in neun Verfahren (32,1 Prozent) sowie in der Beschlussfassung in vier Verfahren (14,3 Prozent) eine *Verletzung des Intention-to-Treat (ITT)-Prinzips* bemängelt wurde. Eine *unzureichende Responderanalyse/MID* wurde in der Dossierbewertung in sieben Verfahren (25 Prozent) sowie der Beschlussfassung in drei Verfahren (10,7 Prozent) beanstandet.

Mangelnde Signifikanz bzw. Relevanz wurde in der Dossierbewertung in 13 Verfahren (46,4 Prozent) sowie in der Beschlussfassung in 17 Verfahren (60,7 Prozent) festgestellt. *Mangelnde statistische Signifikanz* wurde in der Dossierbewertung in zwölf Verfahren (42,9 Prozent) sowie in der Beschlussfassung in 15 Verfahren (53,6 Prozent) bemängelt. *Mangelnde klinische Relevanz* wurde in einem Verfahren (3,6 Prozent) in der Dossierbewertung und in zwei Verfahren (7,1 Prozent) in der Beschlussfassung beanstandet.

Die stärkste Ausprägung auf aggregierter Ebene weist die Kategorie *formale Mängel* auf. Insgesamt ließen sich dieser Kategorie 14 Verfahren der Dossierbewertung (50 Prozent) sowie 15 Verfahren der Beschlussfassung zuordnen (53,6 Prozent). Die verfeinerte Betrachtung dieser Kategorie zeigt, dass insbesondere das *Nichterfolgen der Nutzenbewertung* ausschlaggebend für die starke Ausprägung dieser Kategorie ist. So fand in zehn Verfahren (35,7 Prozent) keine Dossierbewertung statt. In sieben Verfahren (25 Prozent) fand ebenfalls keine nachträgliche Bewertung innerhalb der Beschlussfassung statt. Des Weiteren wurden einzelne Studien, die HRQoL-Daten enthielten, aus formalen Gründen aus der Nutzenbewertung ausgeschlossen. Innerhalb der Dossierbewertung wurden in vier Verfahren (14,3 Prozent), innerhalb der Beschlussfassung in acht Verfahren (28,6 Prozent) einzelne Studien *aus formalen Gründen ausgeschlossen*.

Ergebnis:

- Die Gründe für die Nichtberücksichtigung von HRQoL-Daten sowie das Versagen eines endpunktspezifischen ZN ließen sich fünf verschiedenen Kategorien zuordnen.
- Weitestgehende Übereinstimmung zeigte sich in den Kategorien *Mängel des Erhebungsinstruments* sowie *formale Mängel*.
- Abweichungen zeigten sich in den Kategorien *Mängel im Studiendesign*, *Mängel in der Auswertung* und *mangelnde Signifikanz/Relevanz*.
- In der Dossierbewertung weist die Kategorie *Mängel in der Auswertung* die stärkste Ausprägung auf.
- In der Beschlussfassung weist die Kategorie *mangelnde Signifikanz/Relevanz* die stärkste Ausprägung auf.

8 Diskussion

8.1 Limitationen der Analyse

Die vorliegende Arbeit besteht in einer deskriptiven Analyse sämtlicher abgeschlossener Verfahren gemäß § 35a SGB V im Zeitraum vom 1. Januar 2011 bis 31. August 2013. Der zur Entwicklung und Auswertung des Kategoriensystems gewählte Ansatz der qualitativen Inhaltsanalyse nach Mayring entspricht der aktuellen sozialwissenschaftlichen Diskussion über eine *Mixed Methodology* zur Erzielung von gegenstandsadäquaten, relevanten und reichhaltigen Forschungsergebnissen.²⁷⁴ Die gewählte Methodik ermöglicht einen Brückenschlag zwischen einer qualitativen und quantitativen Analyse der im Rahmen dieser Arbeit eingeschlossenen Nutzenbewertungsverfahren. Durch die induktive Entwicklung von Kategorien aus dem Material heraus ist sie dafür besser geeignet als andere inhaltsanalytischen Verfahren wie z.B. die Frequenz-, die Valenz- und Intensitäts- oder die Kontingenzanalyse, da diese die Aufstellung eines Kategoriensystems ex ante beinhalten.²⁷⁵ Dies setzt jedoch die Existenz eines entsprechenden Suchrasters voraus und würde somit einen unzulässigen Vorgriff in Anbetracht der offen formulierten Fragestellung bedeuten. In der qualitativen Inhaltsanalyse wird hingegen ein besonderes Augenmerk auf die regelgeleitete Konstruktion von Kategorien gelegt.²⁷⁶

Problematisch hierfür erwies sich die innerhalb der ersten zehn Verfahren feststellbare, geringe Begründungspraxis des G-BA. Um die Anwendung eines einheitlichen Kategoriensystems zu ermöglichen, wurde für die vorliegende Arbeit angenommen, dass der G-BA in diesen Fällen der Bewertung und Begründung des IQWiG folgte.

²⁷⁴ Vgl. Mayring (2010), S. 8.

²⁷⁵ Vgl. Ebd.

²⁷⁶ Vgl. Ebd., S. 86-85.

Eine weitere Limitation der Analyse besteht in dem Umstand, dass das entwickelte Kategoriensystem nicht hinsichtlich des Gütekriteriums der *Intercoder*-Reliabilität getestet wurde, wonach mehrere Inhaltsanalytiker unabhängig voneinander das gleiche Material bewerten und die Ergebnisse ihrer Analysen anschließend vergleichen.²⁷⁷ Vor dem Hintergrund außerhalb des Beobachtungszeitraums liegender Verfahren scheint zudem eine fortlaufende Weiterentwicklung und Anpassung des Kategoriensystems geboten.

Zusammenfassend ist davon auszugehen, dass die Analyse die für die Verfahren maßgeblichen Mängel zwar gegenstandsadäquat, aufgrund der genannten Limitationen jedoch nur eingeschränkt detailgetreu abbildet.

8.2 Erhebungs-, Einreichungsverhalten und Berücksichtigungsstand von HRQoL-Daten

Die Analyse sämtlicher abgeschlossener Nutzenbewertungsverfahren zeigt, dass pU bereits mehrheitlich Daten zu dem patientenrelevanten Endpunkt HRQoL erheben und im Rahmen der Nutzenbewertung einreichen. In 28 der 46 Nutzenbewertungsverfahren wurden HRQoL-Daten im VD präsentiert. Somit betrug die HRQoL-Einreichungsquote 61 Prozent. Hinsichtlich der Rolle der Daten in der Nutzenbewertung lässt sich allenfalls ein sehr geringer Einfluss feststellen. Nur in den beiden Verfahren Ivacaftor (Kalydeco™) und Crizotinib (Xalkori®) wurde ein HRQoL-spezifischer ZN ausgesprochen.

Dieses Ergebnis deckt sich in der Tendenz mit der Analyse von Klose et al. (2013), weicht im Detail jedoch von ihr ab: In 43 betrachteten Verfahren identifizierten die Autoren 25 VD, die HRQoL-Daten enthielten.²⁷⁸ Hieraus ergibt sich eine HRQoL-Einreichungsquote von 58 Prozent. Darüber hinaus konstatieren sie in drei der 25 Verfahren einen HRQoL-spezifischen ZN.²⁷⁹ Die Analyse bildet den Stand der Einreichung bis zum 1. Juli 2013 ab.²⁸⁰ Aufgrund des um einen Monat abweichenden Beobachtungszeitraums ist davon auszugehen, dass die Verfahren Ingenolmebutat (Picato®) und die Erweiterung des Anwendungsgebiets (AWG) von Abirateronacetat (Zytiga®) nicht in die Analyse eingegangen sind. Mögliche Gründe für die darüber hinausgehende Abweichung der Einreichungsquote um ein Verfahren könnten in einer unterschiedlichen Definition von HRQoL bestehen, wodurch bestimmte Messinstrumente aus der Analyse ausgeschlossen wurden. Nicht nachvollzogen werden kann hingegen die Identifikation eines dritten Verfahrens mit HRQoL-spezifischem ZN auf Basis der veröffentlichten G-BA Beschlüsse.

²⁷⁷ Vgl. Ebd., S. 51.

²⁷⁸ Vgl. Klose et al. (2013), S. 14.e1.

²⁷⁹ Vgl. Ebd.; Anmerkung: Eine Kontaktaufnahme mit den Autoren führte nur bedingt zur Klärung des Sachverhalts. So wurde zwar der genaue Beobachtungszeitraum mitgeteilt, der alle vom G-BA veröffentlichten Nutzenbewertungen bis 01. Juli 2013 umfasst. Eine Aufschlüsselung der verarbeiteten Studien erfolgte jedoch nicht mit der Begründung, dies sei derzeit aus organisatorischen Gründen nicht möglich.

²⁸⁰ Vgl. Klose et al. (2013), S. 14.e1.

Die Auswertung der eingereichten Studien zeigt, dass die betrachteten Verfahren fast ausschließlich auf Zulassungsstudien beruhen. Darüber hinaus konnte festgestellt werden, dass HRQoL in der Mehrheit der Verfahren als sekundärer Endpunkt erhoben wurde. Nur in dem OD-Verfahren Tafamidis Meglumin (Vyndaqel[®]) wurde HRQoL als (co-)primärer Endpunkt definiert. Zu einem ähnlichen Ergebnis gelangt das IQWiG (2014) in einer systematischen Übersichtsarbeit zu RCT an Patienten mit fortgeschrittenen Krebserkrankungen.²⁸¹ Das Gesamtüberleben sowie das PFS waren die am häufigsten erhobenen primären Studienendpunkte; PRO waren hingegen in keiner der insgesamt 100 betrachteten Studien als primärer Endpunkt definiert.²⁸² In 36 Publikationen war die Erhebung von PRO als sekundärer oder tertiärer Endpunkt vorgesehen, jedoch wurden nur in 31 RCT Ergebnisse in der Primärpublikation präsentiert.²⁸³ Davon wurden in 21 Studien HRQoL-Daten erhoben.²⁸⁴ Auch Leverkus (2014) konstatiert, dass HRQoL in der Praxis eher selten als primärer Endpunkt definiert wird.²⁸⁵ Ursächlich hierfür könnte vor allem der unterschiedliche Stellenwert des Konzepts der HRQoL aus Sicht von Zulassungs- und HTA-Behörden sein. Während HRQoL-Daten im Rahmen der Zulassung eher ein supportiver Charakter zugestanden wird, stellen diese innerhalb des AMNOG ein zentrales Nutzenkriterium dar. Daher fände die Etablierung von HRQoL als primärem Endpunkt bisher auch oftmals ihre „Grenzen in der Multinationalität der Studienprogramme“, so Leverkus.²⁸⁶ Darüber hinaus sei die *Medical Community* primär an klinischen (nicht an patientenberichteten) Studiendaten interessiert. Zudem sei die HRQoL-Messung methodisch z.B. hinsichtlich der Imputation fehlender Werte, der Interpretation der Ergebnisse, der Fallzahlplanung und der Verblindung problematisch. Mit am stärksten wiege jedoch, dass die publizierte Evidenz zu PFS um ein Vielfaches größer sei als zu HRQoL, woraus gegenläufige Anreize zu deren Definition als primäre Zielgröße resultierten.

Die Auswertung des Messinstrumentariums zeigt, dass zur Erhebung des Endpunktes HRQoL fast ausschließlich etablierte und validierte Messinstrumente eingesetzt wurden. Zudem wurden mehrheitlich spezifische Instrumente verwendet. Auch diese Feststellung deckt sich mit den Ergebnissen der zuvor angeführten Übersicht des IQWiG (2014).²⁸⁷ Nur im Falle der Verfahren Fingolimod (Gilenya[®]) und Pasireotid (Signifor[®]) erfolgte der erstmalige Einsatz von Messinstrumenten. Im Zusammenhang mit einer mangelnden Validierung von HRQoL-Instrumenten waren die diskutierten Mängel in der Regel eher formaler Natur und bezogen sich auf Fehler in der Zitation, die unsachgemäße Anwendung von Messinstrumenten und die mangelnde Spezifität von Instrumenten für das betrachtete Patientenkollektiv (Vgl. 8.5.3 Mängel

²⁸¹ Vgl. IQWiG (2014).

²⁸² Vgl. IQWiG (2014), S. 29-31.

²⁸³ Vgl. Ebd., S. 31.

²⁸⁴ Vgl. Ebd.

²⁸⁵ Vgl. Leverkus (2014), S. 114.

²⁸⁶ Ebd., S. 118.

²⁸⁷ Vgl. IQWiG (2014), S. 32f.

in der Auswertung). So wurde in der Nutzenbewertung von Pasireotid gar nicht explizit die mangelnde Validierung des Instruments sondern der verwendeten MID diskutiert.²⁸⁸

Im Zuge der Auswertung ergab sich der Sonderfall, dass in einigen Studien des pU zwar HRQoL-Messinstrumente eingesetzt worden waren, im Rahmen des VD dazu jedoch keine Präsentation der Ergebnisse erfolgte (Vgl. 7. Ergebnisdarstellung). Die nähere Betrachtung dieser Verfahren zeigt, dass seitens pU eher auf die Darstellung generischer Instrumente, speziell des EQ-5D, verzichtet wurde. Ein möglicher Grund hierfür könnte sein, dass generische Instrumente als zu wenig sensitiv erachtet werden, um die für den Patienten relevanten Einflussflussfaktoren der Behandlung zu erfassen. Diese Ansicht vertritt bspw. die EMA (2005) und fordert daher der Einsatz von krankheitsspezifischen HRQoL-Instrumenten.²⁸⁹ Bei erfolgtem Einsatz stellt ein Verzicht auf die Präsentation der Ergebnisse einen Vorgriff des pU gegenüber der Nutzenbewertung dar, der darüber hinaus den Verdacht einer ergebnisgesteuerten Berichterstattung erweckt. Johnston et al. (2013) verweisen darauf, dass speziell der Einsatz mehrerer PRO-Instrumente die Gefahr einer selektiven Ergebnisdarstellung (engl. *Reporting Bias*) birgt.²⁹⁰ Im Falle von Axitinib (Inlyta[®]) und Ipilimumab (Yervoy[®]) gab der pU an, aufgrund konsistenter bzw. gleichartiger Ergebnisse auf die Darstellung der generischen Instrumente zu verzichten, um Redundanzen zu vermeiden. Gerade die kombinierte Anwendung von HRQoL-Instrumenten ist jedoch als *Best-Practice* der HRQoL-Erhebung zu bewerten (Vgl. 8.4 *Best Practice* der Erhebung und Einreichung von HRQoL-Daten).

Erstmalig anhand des Addendums zum Verfahren Crizotinib zeigte sich eine neue Bewertungspraxis des IQWiG wonach der Summenscore der EQ-5D allenfalls als Nutzwert, nicht jedoch als patientenrelevanter Nutzenparameter interpretiert werden kann.²⁹¹ Auch die visuelle Analogskala (VAS) des EQ-5D, die einen globalen Gesundheitszustand abbildet, wurde in späteren Verfahren nicht mehr als patientenrelevant anerkannt. Ursächlich für den Ausschluss des EQ-5D-Index ist die Verwendung bevölkerungsbezogener Tarife, d.h. einer aus der allgemeinen Bevölkerung abgeleiteten Präferenzfunktion. Dieses Vorgehen steht jedoch im Widerspruch zur bisherigen Bewertungspraxis des IQWiG und des G-BA und stellt insofern eine Verfahrensinkonsistenz dar. Noch in der Dossierbewertung von Axitinib (Inlyta[®]) bemängelte das IQWiG, dass der pU auf die Präsentation der Ergebnisse des EQ-5D verzichtete.²⁹² Im Falle des OD Ivacaftor (Kalydeco[™]) bezeichnete der G-BA den Einsatz des EQ-5D ausdrücklich als wünschenswert.²⁹³ Darüber hinaus findet der EQ-5D auf Ebene der Arzneimittelzulassung weiterhin Berücksichtigung. So erkennt die EMA sowohl den Index als auch die VAS des EQ-5D nach wie vor im *European Public Assessment* an.²⁹⁴

²⁸⁸ Vgl. G-BA (2012d), S. 18f.

²⁸⁹ Vgl. EMA (2005), S. 3.

²⁹⁰ Vgl. Johnston et al. (2013), S. 7.

²⁹¹ Vgl. IQWiG (2013b), S. 3.

²⁹² Vgl. IQWiG (2012a), S. 41f.

²⁹³ Vgl. IQWiG (2012b), S. 10.

²⁹⁴ Vgl. EMA (2013), S. 44.

Problematisch erscheint der Ausschluss des EQ-5D in Anbetracht seines breiten Einsatzes und seiner generellen Akzeptanz in der Lebensqualitätsforschung und der Gesundheitsökonomie. So ergab die Auswertung des Messinstrumentariums, dass der EQ-5D in 46 Prozent aller Verfahren mit Angaben zur HRQoL verwendet wurde. In 36 Prozent der Verfahren präsentierte der pU entsprechende Ergebnisse im VD. Die Fortführung des Ausschlusses des EQ-5D bedeutet somit perspektivisch einen erheblichen Ausschluss von HRQoL-spezifischer Evidenz. Da der EQ-5D international im Kontext des HTA als Basis für die Berechnung von QALYs etabliert ist, ist der formale Ausschluss im Rahmen der frühen Nutzenbewertung als deutscher „Sonderweg“ zu bewerten. Dies erschwert eine weitere Harmonisierung – insbesondere des europäischen HTA – und steht im Kontrast zu Initiativen wie EUnetHTA.²⁹⁵

8.3 Best Practice der Erhebung und Einreichung von HRQoL-Daten

Es ließen sich zwei Verfahren identifizieren, in denen pU HRQoL-Daten erfolgreich einreichten und ein endpunktspezifischer ZN ausgesprochen wurde.

Im Falle des OD Ivacaftor (Kalydeco™) zur Behandlung der Mukoviszidose mit G551D-Mutation präsentierte der pU im VD Daten zur Quantifizierung des ZN im Vergleich zu *Best Supportive Care* (BSC). A priori definierte patientenrelevante Endpunkte waren Morbidität und HRQoL. In zwei Zulassungsstudien erhob der pU einerseits für Patienten im Alter von sechs bis elf Jahren (Studie VX08-770-103) sowie andererseits für Patienten im Alter von mindestens zwölf Jahren (Studie VX08-770-102) HRQoL-Daten mittels des krankheitsspezifischen Instruments CFQ-R (Domäne Atmungssystem) und des generischen Instruments EQ-5D.

Positiv hervorzuheben ist der Einsatz eines in deutschen wie auch in internationalen Studien validierten HRQoL-Erhebungsinstruments, was der G-BA entsprechend bestätigte.²⁹⁶ Zudem erfolgte die Verwendung eines Profilinstruments, das der Multidimensionalität des Konstrukts HRQoL Rechnung trägt. Dieses Potenzial schöpfte der pU jedoch nicht aus, da er allein Ergebnisse zu der Domäne Atmungssystem präsentierte. Der pU riskierte dadurch bereits auf Ebene der Zulassung die Nichtberücksichtigung der HRQoL-Daten, da die EMA ausdrücklich die Präsentation der Ergebnisse sämtlicher Domänen empfiehlt.²⁹⁷ Vorbildcharakter für die Erhebung und Präsentation von HRQoL-Daten im Rahmen der Nutzenbewertung erlangt das Verfahren Ivacaftor insbesondere aufgrund des kombinierten Einsatzes von spezifischen und generischen HRQoL-Instrumenten. Dieses Vorgehen entspricht den Empfehlungen der *Scientific Community*, wonach der kombinierte Einsatz beider Instrumententypen das Potenzial birgt, komplementäre Evidenz zu generieren, auch als sog. „belt-and-braces-approach“ bezeichnet.²⁹⁸

²⁹⁵ Vgl. Zentner/Busse/Velasco Garrido (2011).

²⁹⁶ Vgl. G-BA (2012c), S. 10.

²⁹⁷ Vgl. EMA (2005), S.3f.

²⁹⁸ Vgl. Fitzpatrick et al. (1998), S. 16; Marquis et al. (2006), S. 195; Osoba et al. (2007), S. 140; Walters (2009), S. 31.

Auch der G-BA empfiehlt zwischenzeitlich explizit die kombinierte Anwendung von HRQoL-Instrumenten.²⁹⁹

Im Falle des Wirkstoffs Crizotinib (Xalkori[®]) zur Behandlung des nicht kleinzelligen Bronchialkarzinoms präsentierte der pU im VD Daten zur Quantifizierung des ZN im Vergleich zu Pemetrexed/Docetaxel bzw. BSC. A priori definierte patientenrelevante Endpunkte waren Mortalität, Morbidität und HRQoL.³⁰⁰ In der Studie PROFILE 1007 erhob der pU HRQoL-Daten mittels des spezifischen Instruments EORTC QLQ-C30 sowie des generischen Instruments EQ-5D.³⁰¹ In der Studie PROFILE 1005 erhob der pU HRQoL-Daten mittels des generischen Instruments EQ-5D.³⁰² Eine Besonderheit des Verfahrens besteht in dem Umstand, dass für Crizotinib gemäß der Verordnung (EG) Nr. 507/2006 in Europa eine bedingte Zulassung ausgesprochen wurde, wonach der Wirkstoff noch vor Abschluss der vollständigen klinischen Prüfung zur Behandlung von Patienten eingesetzt werden durfte.³⁰³ Entsprechend erfolgt die Argumentation des pU anhand der zur Verfügung stehenden eingeschränkten Daten- bzw. Evidenzlage.

Crizotinib kann als Paradebeispiel einer onkologischen Substanz gelten, die vor allem aufgrund von überzeugenden HRQoL-Daten positiv bewertet wurde.³⁰⁴ Vorbildcharakter für die Erhebung und Präsentation von HRQoL-Daten im Rahmen der Nutzenbewertung erlangt das Verfahren aufgrund des Einsatzes eines Profilinstruments, bzw. der getrennten Erfassung und Abbildung der einzelnen Domänen der HRQoL. Dieses Potenzial schöpfte der pU auch vollumfänglich aus, indem er Ergebnisse zu sämtlichen Domänen präsentierte. Für fünf der sechs Domänen bestätigte der G-BA ein statistisch signifikantes Ergebnis zugunsten der Behandlung mit Crizotinib.³⁰⁵ Angesichts der einheitlichen Effektrichtung fiel die mangelnde statistische Signifikanz der Domäne kognitive Funktion nicht weiter ins Gewicht, so dass der Endpunkt insgesamt mit einem beträchtlichen ZN in die Nutzenbewertung einging.³⁰⁶ Positiv hervorzuheben ist auch im Falle von Crizotinib der kombinierte Einsatz von spezifischen und generischen HRQoL Instrumenten. Darüber hinaus sind die a priori im Studienprotokoll geplanten Responderanalysen positiv hervorzuheben. Insbesondere im Falle des Vorliegens einer validierten MID stellen Responderanalysen das Auswertungsverfahren der Wahl zur Bewertung der Relevanz von Gruppenunterschieden dar.³⁰⁷

²⁹⁹ Vgl. Klakow-Frank (2014), S. 154.

³⁰⁰ Vgl. Pfizer Pharma GmbH (2012), S. 30-32.

³⁰¹ Vgl. Ebd., S. 60.

³⁰² Vgl. Ebd., S. 59.

³⁰³ Vgl. Ebd., S. 22.

³⁰⁴ Vgl. Klakow-Frank (2014), S. 154.

³⁰⁵ Vgl. G-BA (2013a), S. 17f.

³⁰⁶ Vgl. G-BA (2013a), S. 7f.

³⁰⁷ IQWiG (2013a), S. 134.

8.4 Mängel in der Erhebung und Einreichung von HRQoL-Daten

In der überwiegenden Mehrheit der Verfahren wurde weder in der Dossierbewertung noch in der Beschlussfassung ein endpunktspezifischer ZN festgestellt. Das im Rahmen der vorliegenden Arbeit vorgestellte Kategoriensystem differenziert die Gründe für die Nichtberücksichtigung sowie das Versagen eines endpunktspezifischen ZN in die Oberkategorien *Mängel im Studiendesign*, *Mängel des Erhebungsinstruments*, *Mängel in der Auswertung*, *mangelnde Signifikanz/Relevanz der Daten* und *formale Mängel*.

8.4.1 Mängel im Studiendesign

Hinsichtlich der Attestierung von Mängeln im Studiendesign zeigten sich erhebliche Unterschiede zwischen der Dossierbewertung und der Beschlussfassung. Die Diskussion entsprechender Mängel erfolgte weitaus häufiger in der Dossierbewertung.

Unvollständige Dokumentation und Präsentation der Daten

Die Verfahren Pasireotid (Signifor[®]) und Abirateronacetat (Zytiga[®]) (neues AWG) illustrieren generelle Probleme, die sich aus einer mangelnden Konkretisierung, insbesondere der statistischen Testung des Endpunktes HRQoL im Rahmen des Studienprotokolls ergeben. Sowohl in der Dossierbewertung als auch in der Beschlussfassung wurde dieses Vorgehen bemängelt.³⁰⁸ Bereits im Rahmen der Zulassung fordert die FDA (2009) eine vollumfängliche Darlegung der geplanten statistischen Auswertungsverfahren und unterstreicht dabei, dass hierbei für PRO die gleichen Anforderungen bestehen wie auch für andere Endpunkte.³⁰⁹ Ungeplante bzw. post hoc festgelegte Auswertungen bergen folglich bereits auf Ebene der Zulassung die Gefahr einer Nichtberücksichtigung. Klose et al. (2013) verweisen in diesem Zusammenhang auch auf nachgelagerte Probleme, die aus einer post-hoc Bildung von Subgruppen resultieren können.³¹⁰ So kann die damit einhergehende reduzierte Stichprobengröße zu einer Verletzung des ITT-Prinzips führen (Vgl. 8.3.3.1 Verletzung des ITT-Prinzips).

Systematische Unterschiede zwischen den Behandlungsgruppen

In dem Verfahren Ruxolitinib (Jakavi[®]) wurden in der Dossierbewertung und der Beschlussfassung systematische Unterschiede zwischen den Behandlungsgruppen hinsichtlich der Geschlechts-, der Altersverteilung, des Gesundheitszustandes sowie der Vorbehandlung der eingeschlossenen Patienten festgestellt.³¹¹ Imbalancen zwischen den Studienarmen ergaben sich darüber hinaus durch die Möglichkeit eines *Cross-Over* der Patienten vom Placebo- in den

³⁰⁸ Vgl. G-BA (2012d), S. 33; IQWiG (2011a), S. 49; G-BA (2013b), S. 9f.

³⁰⁹ Vgl. FDA (2009), S. 27.

³¹⁰ Vgl. Klose et al. (2013), S. 14.e2.

³¹¹ Vgl. G-BA (2012e), S. 18.

Ruxolitinib-Arm.³¹² Dies kann dazu führen, dass Verum mit Verum verglichen wird, d.h. HRQoL-Daten nach dem *Cross-Over* schlicht nicht mehr interpretierbar sind.³¹³ Dass sich dennoch kein automatischer Ausschluss der Daten aus der Nutzenbewertung ergibt, zeigt das Verfahren Crizotinib. Auch hier wies das IQWiG zunächst auf einen verzerrenden Einfluss infolge des *Cross-Over* hin, sprach auf Basis der Daten jedoch einen Anhaltspunkt für einen HRQoL-spezifischen ZN aus.³¹⁴ Als „Ausweg“ hinsichtlich des Dilemmas einer möglichen Verzerrung erwies sich im Falle von Crizotinib die Messung der HRQoL bei Studienabbruch sowie -ende.³¹⁵ Somit standen für den Zeitraum bis zum Wechsel des Studienarms valide Daten zur Verfügung.

Hohes Verzerrungspotenzial

In den meisten Fällen bezog sich die Attestierung von Mängeln im Studiendesign auf ein hohes Verzerrungspotenzial der vorgelegten Studien. Diese Kategorie war in der Dossierbewertung wesentlich stärker ausgeprägt als in der Beschlussfassung, worauf sich auch die Varianz auf Ebene der Oberkategorie zurückführen lässt.

In den Verfahren Belatacept (Nulojix[®]), Vemurafenib (Zelboraf[®]) und Decitabin (Dacogen[®]) ergab sich ein hohes Verzerrungspotenzial infolge eines offenen Studiendesigns. Der pU gab an, eine vollständige Verblindung sei aus praktischen und medizinischen Gründen ausgeschlossen gewesen.³¹⁶ In der Dossierbewertung des OD Ruxolitinib (Jakavi[®]) führte das Auftreten charakteristischer Nebenwirkungen zu der Feststellung des G-BA, eine Verblindung sei nur schwer aufrechtzuerhalten.³¹⁷ Das Fehlen einer Kontrollgruppe sowie eine hohen Rate an Studienabbruchern führten in der Nutzenbewertung des OD Pasireotid (Signifor[®]) zu der Feststellung eines hohen Verzerrungspotenzials.³¹⁸ In den Verfahren Abirateronacetat (Zytiga[®]) und Decitabin (Dacogen[®]) war die unzureichende Darstellung des Umgangs mit *Lost-to-follow-up* Patienten und Protokollverletzern ursächlich für die diesbezügliche Kritik.³¹⁹ Im Falle des Verfahrens Axitinib (Inlyta[®]) ergab sich das hohe Verzerrungspotenzial neben dem offenen Studiendesign aus unterschiedlichen Behandlungszeiträumen der Axitinib- und Sorafenibgruppe.³²⁰

Die diesbezüglichen Beanstandungen des IQWiG und des G-BA decken sich mit den Empfehlungen der FDA (2009) und der EMA (2005) zur Erhebung von PRO.³²¹ So existiert

³¹² Vgl. G-BA (2013c), S. 7f.

³¹³ Vgl. Leverkus (2014), S. 115.

³¹⁴ Vgl. IQWiG (2013d), S. 3; S. 22.

³¹⁵ Vgl. Leverkus (2014).

³¹⁶ Vgl. BMS GmbH & Co. KGaA (2012a), S. 197; Roche Pharma AG (2012), S. 56; Janssen-Cilag GmbH (2012a), S. 43.

³¹⁷ Vgl. G-BA (2012e), S. 55.

³¹⁸ Vgl. G-BA (2012f), S. 5f.

³¹⁹ Vgl. IQWiG (2011a), S. 42; G-BA (2013d), S. 76.

³²⁰ Vgl. IQWiG (2012a), S. 4.

³²¹ Vgl. FDA (2009), S. 22; EMA (2005), S. 5.

empirische Evidenz für ein hohes Verzerrungspotenzial von in offenen Studiendesigns erhobenen, subjektiven Endpunkten.³²² Das IQWiG eröffnet jedoch grundsätzlich den Spielraum, einen Hinweis auf einen HRQoL-spezifischen ZN aus offenen Studien abzuleiten.³²³ Herstellerseitig sind in derartigen Fällen andere Anstrengungen nachzuweisen, um die Verzerrung der Daten zu minimieren, bzw. einer Einschätzung zugänglich zu machen. Das IQWiG führt exemplarisch die verblindete Endpunkterhebung und -bewertung an.³²⁴

Während im Rahmen der Dossierbewertung eine sehr differenzierte Bewertung von Mängeln im Studiendesign vorgenommen wurde, fand nur in den Verfahren Ruxolitinib, Pasireotid und Abirateronacetat eine Thematisierung entsprechender Mängel in der Beschlussfassung statt. Ein möglicher Grund hierfür könnte eine Staffelung von Ablehnungsgründen sein, die bspw. an dem Verfahren Fingolimod (Gilenya[®]) deutlich wird (Vgl. 8.3.3. Mängel in der Auswertung).

8.4.2 Mängel des Erhebungsinstruments

Hinsichtlich der Attestierung von Mängeln des Erhebungsinstruments zeigte sich ein homogenes Bild zwischen der Dossierbewertung und der Beschlussfassung.

Zu geringe Rücklaufquote des Erhebungsinstruments

Die OD-Verfahren Decitabin (Dacogen[®]) und Ruxolitinib (Jakavi[®]) illustrieren das generelle Problem zu geringer Rücklaufquoten von HRQoL-Erhebungsinstrumenten, aus dem sich Probleme hinsichtlich der Interpretation der Daten ergeben. Gründe für einen zu geringen Fragebogenrücklauf können systematisch sein, wenn sie auf ein Fortschreiten der Erkrankung zurückzuführen sind.³²⁵ Ein geringer Fragebogenrücklauf kann jedoch auch zufällig begründet sein und auf administrativen Fehlern, wie bspw. dem versehentlichen Vergessen des Aushändigens von Fragebögen oder der mangelnden Mitwirkung der Patienten beruhen.³²⁶ Marquis et al. (2006) sprechen in diesem Zusammenhang von der Notwendigkeit, den “buy-in” sämtlicher an der Erhebung beteiligter Personen sicherzustellen.³²⁷ Einen Ansatz hierfür präsentieren Osoba et al. (2007), die zunächst auf die Bedeutung der systematischen Verankerung der HRQoL-Erhebung im Studienprotokoll, insbesondere hinsichtlich der Rationale, des Zeitpunkts und der Verantwortlichkeiten hinweisen.³²⁸ Darüber hinaus empfehlen sie eine durchgängige Kommunikation des Stellenwerts der HRQoL-Erhebung gegenüber klinisch-wissenschaftlichen Mitarbeitern und Patienten.³²⁹

³²² Vgl. Wood et al. (2008), S. 601-605.

³²³ Vgl. IQWiG (2013a), S. 46.

³²⁴ Vgl. Ebd.

³²⁵ Vgl. Wood-Dauphinee (1999) S. 357.

³²⁶ Vgl. Moinpour/Lovato (1998), S. 647; Fairclough (2002).

³²⁷ Vgl. Marquis et al. (2006), S. 198.

³²⁸ Vgl. Osoba et al. (2007).

³²⁹ Vgl. Ebd.

Das Verfahren Ipilimumab (Yervoy[®]) verdeutlicht zudem die Problematik der HRQoL-Erhebung in der onkologischen Behandlung bzw. in hoch palliativen Situationen. Deutlich zeigt sich daran, dass der G-BA ein hohes Interesse an HRQoL-Daten im Rahmen der Palliativbehandlung hat und auch gewillt ist, diese zu berücksichtigen. Wie bereits im Falle von Ruxolitinib sprach der G-BA zwar keinen endpunktspezifischen ZN aus, bezog die HRQoL-Daten jedoch in die Interpretation der Bewertung anderer Endpunkte mit ein.

Im Zusammenhang mit der Problematik zu geringer Rücklaufquoten stellt sich jedoch die Frage nach der Zulässigkeit derartiger Schwellenwerte in Ermangelung ausreichender empirischer Evidenz. Das IQWiG legt den *Cut-off* für die Berücksichtigung von HRQoL-Daten bei 70 Prozent Rücklauf fest, diskutiert jedoch auch einen schärferen Schwellenwert von 80 Prozent. Der G-BA fordert in der Regel eine 80-prozentige Rücklaufquote.³³⁰ Formal handelt es sich jedoch bei beiden Schwellenwerten um eine sozialemprirische Konvention. Strech (2010) verweist diesbezüglich auf versteckte Werturteile im Kontext der frühen Nutzenbewertung.³³¹ Zwar sei die Festlegung von Schwellenwerten an verschiedenen Stellen des Assessment-Prozesses zulässig und sogar nötig; jedoch erfordern diese eine systematische Identifizierung, explizite Diskussion, Begründung und transparente Dokumentation.³³² Da HRQoL-Daten die Sicht des Patienten auf seinen Gesundheitszustand widerspiegeln und somit zentrale, zusätzliche Informationen für die Nutzenbewertung liefern können, gilt dies umso mehr. Anstatt Informationen zur HRQoL formal aufgrund eines zu geringen Rücklaufs zu verwerfen, sollten diese stärker als bisher diskutiert, kontextabhängig bewertet und ggf. in die Betrachtung anderer Endpunkte miteinbezogen werden.

Mangelnde Validierung des Erhebungsinstruments

Die Verfahren Abirateronacetat (Zytiga[®]), Vemurafenib (Zelboraf[®]), Ruxolitinib (Jakavi[®]) und Decitabin (Dacogen[®]) illustrieren vor allem formale Probleme des Nachweises der Validierung von HRQoL-Instrumenten, denn grundsätzlich gelten sämtliche eingesetzten Instrumente als hinreichend validiert. So bemängelte das IQWiG im Falle von Abirateronacetat bspw., die zitierte Validierungsstudie beziehe sich auf Version 3 des FACT-P, wohingegen die vorgelegten HRQoL-Daten mittels der Version 4 des FACT-P erhoben worden seien.³³³ Im Falle der Dossierbewertung von Vemurafenib kritisierte das IQWiG, der pU habe eine verkürzte Version des FACT-M verwendet, in der die Subskala zur chirurgischen Melanombehandlung ausgelassen wurde.³³⁴ Zwar erscheine dieses Vorgehen einleuchtend, da in der Studie keine chirurgischen Eingriffe vorgesehen waren, allerdings ergebe sich daraus das Problem, dass der FACT-M-Gesamtscore nicht mehr errechnet werden könne.³³⁵ Entsprechend urteilte auch der G-BA und

³³⁰ Vgl. Klakow-Franck (2014), S. 154.

³³¹ Vgl. Strech (2010).

³³² Vgl. Ebd., S. 174.

³³³ Vgl. IQWiG (2011a), S. 31.

³³⁴ Vgl. IQWiG (2012j), S. 40.

³³⁵ Vgl. Ebd., S. 40f.

sah die Auswertungen zum Gesamtscore als nicht valide an.³³⁶ Insgesamt ist jedoch festzuhalten, dass in keinem der Verfahren die mangelnde Validierung alleine den Ausschlag für die Nichtberücksichtigung der HRQoL-Daten gab.

Darüber hinaus zeigen die Verfahren Vemurafenib, Ruxolitinib und Decitabin die Schwierigkeiten auf, die sich durch das Auslassen bzw. die Darstellung einzelner Subskalen ergeben. Hierauf sollte bei Nichtvorliegenden entsprechender Validierungsstudien verzichtet werden. An den Verfahren Ruxolitinib und Decitabin wird zudem die Präferenz des G-BA hinsichtlich des Einsatzes krankheitsspezifischer Instrumente deutlich.

8.4.3 Mängel in der Auswertung

Auch bezüglich der Attestierung von Mängeln in der Auswertung zeigten sich Abweichungen zwischen der Dossierbewertung und der Beschlussfassung. Entsprechende Mängel wurden in der Dossierbewertung wesentlich häufiger diskutiert.

Verletzung des ITT-Prinzips

Die Verfahren Boceprevir (Victrelis[®]), Fingolimod (Gilenya[®]), Telaprevir (Incivo[®]), Tafamidis Meglumine (Vyndaqel[®]), Emtricitabin, Rilpivirin, Tenofoviridisoproxil (Eviplera[®]), Rilpivirin (Edurant[®]) Vemurafenib (Zelboraf[®]) und Crizotinib (Xalkori[®]) spiegeln generelle Probleme der Auswertung von Studiendaten wider, die sich aus Protokollverletzungen und Studienaustritten ergeben. Im Falle des OD Ruxolitinib (Jakavi[®]) bemängelt der G-BA in der Dossierbewertung, dass seitens des pU keine ITT-Analysen durchgeführt wurden.³³⁷ Um eine systematische Verzerrung infolge der Nichtberücksichtigung bestimmter Patientengruppen in der Auswertung (sog. *Attrition Bias*) ausschließen zu können, wird in der endpunktspezifischen Betrachtung die adäquate Umsetzung des ITT-Prinzips geprüft, wonach mindestens 90 Prozent aller randomisierten Patienten in der durch die Randomisierung zugeordneten Gruppe bei der Auswertung zu berücksichtigen sind.³³⁸

Auch dieser *Cut-off*-Wert ist als Konvention zu bewerten. Insofern greift auch in diesem Zusammenhang Strechs (2010) Kritik bezüglich versteckter Werturteile in der frühen Nutzenbewertung.³³⁹ Anhand der Nutzenbewertung von Fingolimod offenbart sich die Problematik derartiger Schwellenwerte im Kontext der unterschiedlichen Anforderungen von Zulassungs- und HTA-Behörden: In seiner Stellungnahme zur Dossierbewertung führte der pU an, die Bewertung des IQWiG widerspreche der ICH-E9-Richtlinie, die explizit einräume, Patienten aus der ITT-Population auszuschließen, wenn von ihnen nach der Randomisierung keine Daten vorliegen.³⁴⁰ Mit Berufung auf die ICH-E9-Richtlinie sah der pU im Falle von

³³⁶ Vgl. G-BA (2012g), S. 5f.

³³⁷ Vgl. G-BA (2012e), S. 55.

³³⁸ Vgl. IQWiG (2013a), S. 145.

³³⁹ Vgl. Strech (2010).

³⁴⁰ Vgl. G-BA (2012h), S. 260f.

Fingolimod eine adäquate Umsetzung des ITT-Prinzips als gewährleistet an und plädierte dafür, das Verzerrungspotenzial auf Endpunktebene als niedrig einzustufen.³⁴¹ Der G-BA stellte nach Prüfung des Sachverhaltes fest, dass das Verzerrungspotenzial durch das IQWiG zwar als hoch eingestuft wurde, dies jedoch keinen Einfluss auf die Nutzenbewertung habe, da die HRQoL-Daten keine statistische Signifikanz zeigten.³⁴²

Aus diesem Vorgehen wird eine pragmatisch wirkende Vorgehensweise des G-BA deutlich, die die statistische Signifikanz von HRQoL-Daten als notwendige Bedingung für deren Berücksichtigung im Rahmen der Nutzenbewertung voraussetzt, um weitere methodische Aspekte zu diskutieren.

Unzureichende Responderanalyse/MID

Die Verfahren Pirfenidon (Esbriet[®]), Abirateronacetat (Zytiga[®]), Pasireotid (Signifor[®]), Ruxolitinib (Jakavi[®]), Ivacaftor (Kalydeco[™]), Decitabin (Dacogen[®]) und die Erweiterung des AWG von Abirateronacetat (Zytiga[®]) illustrieren das Problem unzureichend durchgeführter Responderanalysen.

Im Rahmen der Dossierbewertung des OD Pirfenidon bemängelte das IQWiG, dass durch den pU nicht beschrieben wurde, ob die verwendete MID auch für den modifizierten SGRQ-I gültig sei.³⁴³ Darüber hinaus würde die Verknüpfung mit dem Ereignis Tod nicht begründet.³⁴⁴ Im Falle von Abirateronacetat führte der pU Responderanalysen ohne Nachweis der potenziellen Effektmodifikation des gewählten Subgruppenmerkmals durch.³⁴⁵ In dem OD-Verfahren Pasireotid kritisierte der G-BA, die für den CushingQoL verwendete MID basiere allein auf den Daten der Zulassungsstudie selbst und leite sich lediglich aus der Standardabweichung der Messwerte der Patienten ab.³⁴⁶ In der Dossierbewertung des OD Ruxolitinib wies der G-BA darauf hin, dass sich aus der zitierten Validierungsstudie lediglich eine intraindividuelle Relevanzschwelle für EORTC QLQ-C30 ableiten ließe.³⁴⁷ Im Falle des OD Ivacaftor war die Verwendung einer gepoolten MID für den verwendeten CFQ-R Gegenstand der Kritik des G-BA.³⁴⁸ In der Dossierbewertung des OD Decitabin führte der pU eine Differenz von 5 bis 10 Punkten als eine geringe, von 10 bis 20 Punkten als eine moderate Änderung und eine Differenz von mehr als 20 Punkten als eine sehr große Änderung des EORTC QLQ-30 zum Besseren oder Schlechteren an.³⁴⁹ Der G-BA monierte, dass diese Kategorien nicht im statistischen Analyseplan genannt wurden, sondern in der Auswertung eine vereinfachte Darstellung bzw. Einteilung von > -10 Veränderung bzw. ≤ -

³⁴¹ Vgl. G-BA (2012h), S. 260f.

³⁴² Vgl. Ebd.

³⁴³ Vgl. IQWiG (2011b), S. 39.

³⁴⁴ Vgl. Ebd.

³⁴⁵ Vgl. IQWiG (2011a), S. 42.

³⁴⁶ Vgl. G-BA (2012d), S. 18f.

³⁴⁷ Vgl. G-BA (2012e), S. 34.

³⁴⁸ Vgl. G-BA (2012c), S. 35.

³⁴⁹ Vgl. Janssen-Cilag GmbH (2012a), S. 58.

10 Veränderung gewählt wurde.³⁵⁰ Innerhalb des Verfahrens zur Erweiterung des Anwendungsgebiets von Abirateronacetat gab der pU als relevante Verschlechterung ein Absinken des FACT-P um 10 Punkte an.³⁵¹ Das IQWiG kritisierte, dass die verwendete MID nicht patientenrelevant sei, da das zu deren Herleitung eingesetzte, ankerbasierte Verfahren entweder auf Laborparametern oder auf einer Einschätzung des behandelnden Arztes beruhe.³⁵² Die Zusammenschau der Verfahren zeigt, dass derartige unzureichend durchgeführten Responderanalysen insbesondere im Falle von OD feststellbar sind. Gegenstand der Kritik war zumeist die verwendete MID. Dieses Ergebnis deckt sich mit der Feststellung von Kose et al. (2013), wonach u.a. mangelnde Relevanzkriterien zur Interpretation von Gruppenunterschieden die geringe Anzahl an HRQoL-spezifischen ZN-Belegen erklärten.³⁵³ An dem Verfahren Abirateronacetat (Zytiga[®]) (neues AWG) zeigt sich darüber hinaus der stringente Fokus des IQWiG auf den Aspekt der Patientenrelevanz, der nicht nur auf Endpunktebene, sondern zusätzlich auch hinsichtlich der verwendeten Responsekriterien sicherzustellen ist. Insgesamt kann festgestellt werden, dass Mängel in der Auswertung eine zentrale Rolle für die Nichtberücksichtigung von HRQoL-Daten in der Dossierbewertung einnahmen.

8.4.4 Mangelnde Signifikanz/Relevanz der Daten

Auch hinsichtlich der Attestierung mangelnder Signifikanz bzw. Relevanz offenbarten sich Unterschiede zwischen der Dossierbewertung und der Beschlussfassung.

Kein statistisch signifikanter Unterschied zwischen den Behandlungsgruppen

In den Verfahren Pirfenidon (Esbriet[®]), Rilpivirin (Edurant[®]), Ivacaftor (Kalydeco[™]), Axitinib (Inlyta[®]), Decitabin (Dacogen[®]), Telaprevir (Incivo[®]), Tafamidis Meglumini (Vyndaqel[®]), Belatacept (Nulojix[®]) und Ipilimumab (Yervoy[®]) wiesen die eingesetzten HRQoL-Instrumente keinen statistisch signifikanten Unterschied zwischen den Behandlungsgruppen nach.³⁵⁴ Der pU beanspruchte jeweils keinen endpunktspezifischen ZN. Im Falle von Tafamidis Meglumini begründete der pU dieses Ergebnis mit einer zu geringen Studienpower.³⁵⁵ Da es sich um die erste RCT in dieser Indikation handelte, standen keine Vergleichswerte zu Verfügung.

Im Falle des Verfahrens Emtricitabin, Rilpivirin, Tenofoviridisoproxil (Eviplera[®]) beanspruchte der pU auf Basis metaanalytischer Berechnungen einen ZN.³⁵⁶ Das IQWiG stellte hingegen fest, weder für den *Physical Component Summary Score* noch für den *Mental Component Summary Score* des

³⁵⁰ Vgl. G-BA (2013d), S. 37f.

³⁵¹ Vgl. Janssen-Cilag GmbH (2013), S. 100.

³⁵² Vgl. IQWiG (2013c), S. 44.

³⁵³ Vgl. Klose et al. (2013), S. 14.e2.

³⁵⁴ Vgl. InterMune Deutschland GmbH (2011), S. 84; Janssen-Cilag GmbH (2012b), S. 99; Vertex Pharmaceuticals GmbH (2012), S. 20; IQWiG (2012a), S. 95-97; G-BA (2013d), S. 111; G-BA (2012i), S. 9; Pfizer Pharma GmbH (2011), S. 64; G-BA (2012j), S. 13; BMS GmbH & Co. KGaA (2012b) S. 91f.

³⁵⁵ Vgl. Pfizer Pharma GmbH (2011), S. 139.

³⁵⁶ Vgl. Gilead Sciences GmbH (2012), S. 73f.

verwendeten SF-36v sei ein statistisch signifikanter Unterschied zwischen den Behandlungsgruppen feststellbar.³⁵⁷ Der G-BA folgte der Argumentation des IQWiG.³⁵⁸

In der Nutzenbewertung von Fingolimod (Gilenya[®]) zeigte der EQ-5D Index zwar signifikante Ergebnisse in der FREEDOMS-Studie ($p=0,037$), in der TRANSFORMS-Studie hingegen nicht ($p=0,101$).³⁵⁹ Seitens des IQWiG wurde nur die TRANSFORMS-Studie für die NB herangezogen.³⁶⁰ Da diese keine signifikanten Ergebnisse zum Endpunkt HRQoL enthielt, stellte das IQWiG fest, ein endpunktspezifischer ZN sei nicht belegt.³⁶¹ Der G-BA folgte der Argumentation des IQWiG.³⁶²

Die Problematik widersprüchlicher HRQoL-Befunde verdeutlicht das Verfahren Vemurafenib (Zelboraf[®]). Lediglich für zwei der fünf Subskalen des FACT-M zeigten sich statistisch signifikante Ergebnisse unterschiedlicher Effektrichtung. Während die Subskala psychisches Wohlbefinden Ergebnisse zugunsten von Vemurafenib zeigte, wies die Subskala emotionales Wohlbefinden Ergebnisse zuungunsten von Vemurafenib auf.³⁶³

In der Gesamtschau der Verfahren wird deutlich, dass die mangelnde statistische Signifikanz eine dominante Rolle für die Nichtberücksichtigung von HRQoL-Daten in der Dossierbewertung, vor allem jedoch in der Beschlussfassung spielte. Die Abweichung auf Ebene der Oberkategorie ist vor allem auf die unterschiedliche Verfahrenszahl in der Dossierbewertung und der Beschlussfassung zurückzuführen. Im Falle der Verfahren Extrakt aus Cannabis Sativa (Sativex[®]), Belimumab (Benlysta[®]) sowie Linagliptin (Trajenta[®]) (Wiedereinreichung gemäß § 35a Abs. 5b) nahm der G-BA eine eigene Bewertung der Studiendaten vor. Das IQWiG hatte hingegen keine Dossierbewertung durchgeführt mit der Begründung, der pU habe keine für die Nutzenbewertung relevanten Studien vorgelegt.³⁶⁴ Jedoch zeigte sich in keinem der drei Verfahren ein statistisch signifikanter Vorteil gegenüber der ZVT.³⁶⁵

Wie das Verfahren Tafamidis Meglumini illustriert, könnten mögliche Gründe hierfür in der Fallzahlplanung bestehen. So setzt die Planung der Studienpower die Kenntnis der Effektstärke bzw. des Unterschieds zwischen den Behandlungsgruppen voraus, der als klinisch relevant erachtet werden kann.³⁶⁶ Einen Ansatzpunkt hierfür kann die Fallzahlplanung ähnlicher Studien bieten, die nachweisliche Therapieunterschiede gezeigt haben.³⁶⁷ Dies setzt jedoch das Vorhandensein entsprechender Studien voraus. Problematisch erscheinen vor diesem Hintergrund die begriffliche sowie die konzeptionelle Vielfalt der HRQoL, die eine

³⁵⁷ Vgl. IQWiG (2012c), S. 17f.

³⁵⁸ Vgl. G-BA (2012k), S. 14.

³⁵⁹ Vgl. Novartis Pharma GmbH (2011), S. 101f.

³⁶⁰ Vgl. IQWiG (2012d), S. 3.

³⁶¹ Vgl. Ebd., S. 24.

³⁶² Vgl. G-BA (2012h), S. 12.

³⁶³ Vgl. Roche Pharma AG (2012), S. 107-111.

³⁶⁴ Vgl. IQWiG (2012e), S. 5; IQWiG (2012f), S. 3-4; IQWiG (2012g), S. 6.

³⁶⁵ Vgl. G-BA (2012b), S. 16; G-BA (2012l), S. 15f.; G-BA (2013e), S. 17.

³⁶⁶ Vgl. Neiß (1990), S. 43.

³⁶⁷ Vgl. Ebd.

entsprechende Zergliederung der Forschungslandschaft bewirkt. Spitzer (1995) sprach in diesem Zusammenhang bereits in den 90er Jahren von einer „Epidemie“ der HRQoL-Werkzeuge.³⁶⁸ Es brauche nicht mehr, sondern eine Fokussierung auf das bereits entwickelte Instrumentarium, um dessen Validierung, Verständnis und Einsatz im klinischen Kontext voranzubringen.³⁶⁹

Darüber hinaus zeigte die Auswertung, dass der Endpunkt HRQoL in 66,7 Prozent der eingeschlossenen Verfahren sekundär erhoben wurde (Vgl. 8.1 Studiencharakteristika und Messinstrumentarium). Da die Bestimmung der erforderlichen Patientenzahl jedoch in der Regel auf den primären Endpunkt hin erfolgt, könnte ein weiterer möglicher Grund für den hohen Anteil der Verfahren, die keine signifikanten HRQoL-Daten vorweisen konnten, in einer zu geringen Fallzahlplanung für die die Erhebung des Endpunktes HRQoL bestehen.

Kein klinisch relevanter Unterschied zwischen den Behandlungsgruppen

Das Verfahren Belatacept (Nulojix[®]) illustriert als einziges Verfahren einen Fall, in dem das IQWiG und der G-BA trotz vorheriger Bestätigung des Vorliegens statistisch signifikanter und gegenüber der ZVT vorteiliger HRQoL-Daten keinen endpunktspezifischen ZN feststellten. In Ermangelung skalenspezifisch validierter bzw. etablierter Relevanzkriterien griff der pU zur Beurteilung der klinischen Relevanz auf die standardisierte Mittelwertdifferenz in Form von Hedges' g zurück. Für den SF-36 Summenscore körperliche Gesundheit ergab sich ein Hedges' g von 0,21 (KI95% 0,06; 0,36, p=0,006). Das Ergebnis lag somit nicht vollständig oberhalb der definierten Irrelevanzschwelle von 0,2 weshalb das IQWiG feststellte, dass ein ZN nicht belegt sei.³⁷⁰ Der G-BA folgte der Empfehlung des IQWiG zur Bewertung der klinischen Relevanz.

Anhand von Belatacept wird die doppelte Hürde deutlich, die sich in Form des Nachweises der Signifikanz und Relevanz von HRQoL-Daten im Rahmen der Nutzenbewertung stellt. Auch in der Festlegung von Signifikanzleveln ist ein Werturteil im Rahmen des HTA-Assessments zu sehen. Der zugrundeliegende Konflikt zwischen der „Glaubwürdigkeit von Informationen“ und dem „Informationsgewinn per se“, wird laut Strech (2010) maßgeblich durch die Risikobereitschaft des jeweiligen Entscheidungsträgers bestimmt.³⁷¹ Legitimität erhalte dessen Urteil jedoch erst durch die explizite Begründung.³⁷² Entsprechend fordert auch der Verband Forschender Arzneimittelhersteller e.V. (vfa), die klinische Relevanz von HRQoL-Daten nicht ausschließlich anhand eines statistischen Maßes zu bewerten.³⁷³

Im Falle der Erweiterung des AWG des Wirkstoffs Abirateronacetat (Zytiga[®]) zeigte sich für Patienten der Günstige-Prognose-Gruppe eine signifikant verlängerte Zeit bis zur Verschlechterung des FACT-P Gesamtscore (HR=0,706; 95% KI: 0,564; 0,885; p=0,0025). Das IQWiG wertete das verwendete Responsekriterium jedoch als nicht patientenrelevant und kam

³⁶⁸ Vgl. Spitzer (1995), S. 63-64.

³⁶⁹ Vgl. Ebd.

³⁷⁰ Vgl. IQWiG (2012h), S. 25.

³⁷¹ Vgl. Strech (2010), S. 171.

³⁷² Vgl. Ebd.

³⁷³ Vgl. G-BA (2012j), S. 65f.

insgesamt zu der Feststellung, der pU habe keine verwertbaren Daten zur HRQoL vorgelegt (Vgl. 7.2.3. Mängel in der Auswertung). Der G-BA folgte der Argumentation des IQWiG, hielt jedoch fest, dass hinsichtlich der klinischen Relevanz der Ergebnisse unterschiedliche Auffassungen bestünden. Insgesamt bleibe die Gesamtaussage zum Ausmaß des ZN davon jedoch unberührt. Das Verfahren Abirateronacetat stellt insofern einen Sonderfall dar, da die Frage der klinischen Relevanz der vorgelegten HRQoL-Daten offen blieb.

8.4.5 Formale Mängel

Hinsichtlich der Attestierung formaler Mängel zeigte sich weitestgehend eine Übereinstimmung zwischen der Dossierbewertung und der Beschlussfassung.

Aus formalen Gründen nicht in die NB einbezogen

Im Falle von Telaprevir (Incivo[®]) schloss das IQWiG die unterstützend vorgelegte Studie ILLUMINATE mit der Begründung aus, dass diese die Einschlusskriterien der ZVT nicht erfülle.³⁷⁴ Auch der G-BA bezog sich nicht explizit auf die ILLUMINATE-Studie. Eine Beurteilung der Konsequenzen dieses Studienausschlusses ist nicht möglich, da der pU im VD keine Angaben zur statistischen Signifikanz der Ergebnisse machte.

Im Falle des OD Tafamidis Meglumin (Vyndaqel[®]) schlossen das IQWiG und der G-BA die offene, multizentrische, einarmige Studie Fx-006 aus, da für die untersuchte Teilpopulation in Form der Studie Fx-005 bereits eine RCT vorlag. Die Fx-006 ist eine einjährige Extensionsstudie der Fx-005. Über den Gesamtbetrachtungszeitraum von 30 Monaten wurde die HRQoL aufrechterhalten. Im VD verwies der pU ausdrücklich darauf, nicht-kontrollierte Studien allenfalls im Sinne unterstützender Evidenz präsentieren zu wollen. Dieses Vorgehen deckt sich mit den Empfehlungen von Windeler und Lange (2008), für Aussagen zum ZN von seltenen Erkrankungen auch auf nicht-vergleichende Studien zurückzugreifen.³⁷⁵

Im Falle von Fingolimod (Gilenya[®]) schlossen das IQWiG und der G-BA die Placebo-kontrollierte Studie FREEDOMS nicht in die Nutzenbewertung ein. Im Gegensatz zu der eingeschlossenen Studie TRANSFORMS zeigte diese in der Auswertung des EQ-5D (Index+VAS) signifikante Unterschiede zwischen den Behandlungsgruppen.

In der Dossierbewertung von Crizotinib (Xalkori[®]) berücksichtigten das IQWiG und der G-BA die Ergebnisse der Studie PROFILE 1005 nicht. Das IQWiG begründete sein Vorgehen damit, dass sämtliche weiteren vorgelegten Studien dieselbe Fragestellung adressieren wie die RCT, jedoch keine zusätzlichen Informationen lieferten.³⁷⁶

Der darüber hinausgehende formale Ausschluss von Studien mit Angaben zur HRQoL resultiert aus der nachträglichen durch den G-BA durchgeführten NB im Falle der Verfahren Extrakt aus

³⁷⁴ Vgl. IQWiG (2012i), S. 71f.

³⁷⁵ Vgl. Windeler/Lange (2008).

³⁷⁶ Vgl. IQWiG (2013d), S. 9.

Cannabis Sativa (Sativex[®]) und der Wiedereinreichung von Linagliptin (Trajenta[®]) gemäß § 35a Abs. 5b.

Darüber hinaus schloss der G-BA im Falle des Verfahrens Belatacept (Nulojix[®]) die Studie BENEFIT aus, da sie hinsichtlich der Patientencharakteristika und Spenderkriterien, als nur bedingt repräsentativ für die entsprechende europäische Patientenpopulation angesehen wurde.³⁷⁷

Keine Nutzenbewertung erfolgt

Im Falle der Verfahren Linagliptin (Trajenta[®]), Retigabin (Trobalt[®]), Vandetanib (Caprelsa[®]), Perampanel (Fycompa[®]), Acridiniumbromid (Eklira[®] Genuair[®] / Bretaris[®] Genuair[®]), Aflibercept (Eylea[®]), Ingenolmebutat (Picato[®]) erfolgte weder eine Nutzenbewertung durch das IQWiG noch durch den G-BA.

In der Gesamtschau der Oberkategorie *formale Mängel* wird deutlich, dass HRQoL-Daten in einem wesentlichen Anteil der Verfahren keinen Eingang in die Nutzenbewertung bzw. die Bemessung des ZN fanden. Diese Verfahren lassen jedoch keinen unmittelbaren Schluss auf die Qualität der eingereichten Daten zu.

³⁷⁷ G-BA (2012j), S. 6.

9 Zusammenfassung und Ausblick

Trotz einer zwischenzeitlich recht regen Diskussion über den Stellenwert und die Charakteristika des patientenrelevanten Endpunktes HRQoL innerhalb des Verfahrens der frühen Nutzenbewertung in Deutschland existieren bisher kaum systematische Auseinandersetzungen mit dem bisherigen Stand der Berücksichtigung entsprechender Daten durch das IQWiG und den G-BA. Im Rahmen der vorliegenden Arbeit konnte gezeigt werden, dass pU innerhalb des Beobachtungszeitraums vom 01.01.2011 bis 31.08.2013 bereits in großem Umfang von der Möglichkeit Gebrauch machten, Daten zu dem patientenrelevanten Endpunkt HRQoL im Rahmen der Nutzenbewertung einzureichen.

Hinsichtlich der ausgewerteten Verfahren konnte festgestellt werden, dass die Nutzenbewertung nahezu ausschließlich auf Zulassungsstudien beruhte. Dabei ist insbesondere in den unterschiedlichen Anforderungen, die Zulassungs- und HTA-Behörden an die einzureichenden Daten stellen, eine Hürde in Bezug auf eine breitere HRQoL-Datenbasis im Rahmen der Nutzenbewertung zu sehen. Bezüglich des verwendeten Messinstrumentariums zeigte sich, dass nahezu ausschließlich etablierte, validierte Instrumente eingesetzt wurden. Problematisch erscheint die sich abzeichnende, neue Bewertungspraxis des EQ-5D durch das IQWiG und den G-BA in Anbetracht seines breiten Einsatzes. Ein Ausschluss des EQ-5D bedeutet zugleich einen erheblichen Ausschluss von HRQoL-spezifischer Evidenz. Diese Ergebnisse deuten auf zentrale Herausforderungen hinsichtlich der systematischen Verankerung von HRQoL als Nutzenkriterium hin und sollten daher an anderer Stelle intensiver analysiert und diskutiert werden.

Insgesamt konnten lediglich zwei Verfahren identifiziert werden, in denen ein HRQoL-spezifischer ZN durch den G-BA ausgesprochen wurde. Als *Best-Practice* der HRQoL-Erhebung ließen sich die kombinierte Anwendung spezifischer wie generischer Instrumente sowie der Einsatz von Profilinstrumenten identifizieren. Ebenfalls explizite Wertschätzung erfuhr die Durchführung von Responderanalysen unter Verwendung validierter Responsekriterien, was jedoch das Vorhandensein oder die Entwicklung und Validierung einer für das jeweilige Messinstrument spezifischen MID voraussetzt.

In 26 Verfahren wurde weder in der Dossierbewertung noch in der Beschlussfassung ein endpunktspezifischer ZN festgestellt. Die vorliegende Arbeit bietet somit einen ersten Vorschlag zur Systematisierung der Diskussion über die Gründe für die geringe Anzahl HRQoL-spezifischer ZN im Rahmen der frühen Nutzenbewertung.

Die inhaltsanalytische Auswertung der Verfahrensdokumente zeigte, dass in der Regel keine monokausale Begründung für das Versagen eines HRQoL-spezifischen ZN vorlag, sondern den eingeschlossenen Studien Mängel in der Konzeption, der Anwendung von Erhebungsinstrumenten und der Ergebnisauswertung attestiert wurden. Eine dominante Rolle für die Nichtberücksichtigung und das Versagen eines HRQoL-spezifischen ZN auf Ebene der Dossierbewertung spielten Mängel in der Auswertung. Ursachen hierfür scheinen vor allem in der mangelnden Verfügbarkeit validierter Schwellenwerte bzw. Relevanzkriterien zur Interpretation von Gruppenunterschieden zu liegen. Im Rahmen der Beschlussfassung bezog

sich die am häufigsten geäußerte Kritik auf die mangelnde statistische Signifikanz der Daten. Ursachen hierfür sind in einem Dreiklang aus einer unzureichenden Fallzahlplanung, mangelnder empirischer Anhaltspunkte zur Festlegung von Effektgrößen sowie in dem geringen Stellenwert des Konzepts der HRQoL auf Eben der Zulassung zu vermuten. Formale Mängel, die zum Ausschluss von HRQoL-Daten führten, bildeten sowohl in der Dossierbewertung als auch in der Beschlussfassung die zweithäufigste Ursache für deren Nichtberücksichtigung. Diese Kategorie lässt jedoch keinen unmittelbaren Schluss auf die Qualität der vorgelegten HRQoL-Daten zu. Weitere charakteristische Mängel der HRQoL-Erhebung zeigten sich hinsichtlich des Studiendesigns und der verwendeten Erhebungsinstrumente.

Insgesamt waren die Nichtberücksichtigung von HRQoL-Daten sowie das Versagen eines endpunktspezifischen ZN durch das IQWiG und den G-BA nur in einem geringen Anteil der Verfahren auf eine mangelnde Validierung der eingesetzten Instrumente zurückzuführen. Stattdessen ließen sich eher formale Beanstandungen an den eingesetzten Instrumenten feststellen. Dies lässt den Schluss zu, dass das Methodenrepertoire grundsätzlich zur Verfügung steht und bereits eingesetzt wird. Jedoch sind weitere Anstrengungen erforderlich, um Theorie und Praxis der HRQoL-Forschung enger zu verzahnen. Schwierigkeiten in der Argumentation über einen HRQoL spezifischen ZN sind insbesondere vor dem Hintergrund der bestehenden begrifflichen und konzeptionellen Vielfalt der HRQoL zu sehen. Erhebliches Potenzial birgt daher die Entwicklung einer einheitlichen Begriffsdefinition der HRQoL sowie die gezieltere theoretische Fundierung und die Etablierung von HRQoL als primärem Endpunkt in klinischen Studien. Diese Konzentration der HRQoL-Forschung könnte die Grundlage für die Erzeugung entsprechender (metaanalytischer) Evidenz liefern und schließlich die notwendigen Referenzpunkte für die Fallzahlplanung bzw. die Antizipation von Effektstärken bereitstellen.

Zusammenfassend lässt sich festhalten, dass die Integration der Patientensicht im Kontext der frühen Nutzenbewertung in Deutschland formal und insbesondere verbal bereits gewährleistet ist. Hinsichtlich der tatsächlichen Berücksichtigung von HRQoL-Daten offenbart sich hingegen ein gegenteiliges Bild. Vielfach scheitern die herstellerseitigen Bemühungen noch, den Nutzen neuer Medikamente aus Patientensicht abzubilden. Da bisher lediglich zwei Verfahren erfolgreich bezüglich eines HRQoL-spezifischen ZN entschieden wurden, bieten sich kaum Orientierungspunkte zur Etablierung einer *Best Practice* der HRQoL-Erhebung. Zugleich erscheint die bisherige Bewertungspraxis durch das IQWiG und den G-BA eher hemmend denn konstruktiv. Statt sich im Rahmen der Nutzenbewertung auf die reine Kontrolle der Einhaltung methodischer Standards zu beschränken, wäre ein akteursübergreifender Dialog zu den methodischen Herausforderungen der HRQoL-Erhebung wünschenswert. In diesem Zusammenhang sollten auch die Probleme adressiert werden, die sich aus einer formalen Festlegung von Schwellenwerten in Ermangelung ausreichender empirischer Evidenz ergeben. Dies betrifft vor allem die derzeit verwendeten *Cut-off* -Werte für den erforderlichen Fragebogenrücklauf, die Einhaltung des ITT-Prinzips und die Beurteilung der klinischen Relevanz. In Analogie zu den *Guidelines* der Zulassungsbehörden sollte für die Nutzenbewertung eine *Good Clinical Practice* der HRQoL-Erhebung formuliert und Lösungsvorschläge für gängige methodische Probleme diskutiert werden.

Anhang A

Bestandteile der Übersichtsmatrix	
Bestandteil	Ausprägung
Wirkstoff	<i>individuell</i>
Handelsname	<i>individuell</i>
Indikation	<i>individuell</i>
Datum der Beschlussfassung	<i>individuell</i>
pU	<i>individuell</i>
ZN-Ergebnissicherheit/Wahrscheinlichkeit laut G-BA-Beschluss	<i>Beleg, Hinweis, Anhaltspunkt</i>
ZN-Ausmaß laut G-BA-Beschluss	<i>erheblich, beträchtlich, gering, nicht quantifizierbar, kein, geringerer</i>
Vorgelegte Studien des pU, die HRQoL Daten enthalten	<i>individuell</i>
Studienkategorie (Zulassungsstudie)	<i>ja, nein</i>
Endpunkt und Endpunktebene	<i>primär/co-primär, sekundär, tertiär, explorativ, nicht näher spezifiziert</i>
HRQoL-Instrument	<i>individuell</i>
HRQoL-Instrumententyp	<i>generisch, spezifisch</i>
Berücksichtigung durch das IQWiG	<i>ja, berücksichtigt, aber kein patientenrelevanter ZN, nein</i>
Begründung für Nichtberücksichtigung/kein patientenrelevanter ZN	<i>individuell</i>
Berücksichtigung durch den G-BA	<i>ja, berücksichtigt, aber kein patientenrelevanter ZN, nein</i>
Begründung für Nichtberücksichtigung/kein patientenrelevanter ZN	<i>individuell</i>

Anhang B

Interpretationsregeln der zusammenfassenden qualitativen Inhaltsanalyse nach Mayring (2010), S. 70.
Z1: Paraphrasierung
Z1.1: Streiche alle nicht (oder wenig) inhaltstragenden Textbestandteile wie ausschmückende, wiederholende, verdeutlichende Wendungen!
Z1.2: Übersetze die inhaltstragenden Textstellen auf eine einheitliche Sprachebene!
Z1.3: Transformiere sie auf eine grammatikalische Kurzform!
Z2: Generalisierung auf das Abstraktionsniveau
Z2.1: Generalisiere Gegenstände der Paraphrasen auf die definierte Abstraktionsebene, sodass die alten Gegenstände in den neu formulierten impliziert sind!
Z2.2: Generalisiere die Satzaussagen (Prädikate) auf die gleiche Weise!
Z2.3: Belasse die Paraphrasen, die über dem angestrebten Abstraktionsniveau liegen!
Z2.4: Nimm theoretische Vorannahmen bei Zweifelsfällen zu Hilfe!
Z3: Erste Reduktion
Z3.1: Streiche bedeutungsgleiche Paraphrasen innerhalb der Auswertungseinheiten!
Z3.2: Streiche Paraphrasen, die auf dem neuen Abstraktionsniveau nicht als wesentlich inhaltstragend erachtet werden!
Z3.3: Übernehme die Paraphrasen, die weiterhin als zentral inhaltstragend erachtet werden (Selektion)!
Z3.4: Nimm theoretische Vorannahmen bei Zweifelsfällen zu Hilfe!
Z4: Zweite Reduktion
Z4.1: Fasse Paraphrasen mit gleichem (ähnlichem) Gegenstand und ähnlicher Aussage zu einer Paraphrase (Bündelung) zusammen!
Z4.2: Fasse Paraphrasen mit mehreren Aussagen zu einem Gegenstand zusammen (Konstruktion/Integration)!
Z4.3: Fasse Paraphrasen mit gleichem (oder ähnlichem) Gegenstand und verschiedener Aussage zu einer Paraphrase zusammen (Konstruktion/Integration)!
Z4.4: Nimm theoretische Vorannahmen bei Zweifelsfällen zu Hilfe!

Anhang C

Nutzenbewertungsverfahren gemäß § 35a SGB V, in denen der Endpunkt HRQoL erhoben wurde (Zeitraum: 1. Januar 2011 bis 31. August 2013)				
Lfd. Nr.	Wirkstoff	Handelsname	Indikation	pU
1	Ticagrelor	Brilique®	Akutes Koronarsyndrom	AstraZeneca GmbH
2	Boceprevir	Victrelis®	Chronische Hepatitis C	MSD SHARP & DOHME GmbH
3	Pirfenidon (OD)	Esbriet®	idiopathische pulmonale Fibrose	InterMune Deutschland GmbH
4	Abirateronacetat	Zytiga®	metastasiertes kastrationsresistentes Prostatakarzinom	Janssen-Cilag GmbH
5	Linagliptin	Trajenta®	Diabetes mellitus Typ 2	Boehringer Ingelheim International GmbH
6	Fingolimod	Gilenya®	Multiple Sklerose	Novartis Pharma GmbH
7	Telaprevir	Incivo®	Hepatitis C	Janssen Cilag GmbH
8	Retigabin	Trobalt®	Epilepsie	GlaxoSmithKline GmbH & Co. KG
9	Tafamidis Meglumine (OD)	Vyndaqel®	Amyloidose	Pfizer Pharma GmbH
10	Extrakt aus Cannabis Sativa	Sativex®	Spastik bei Multipler Sklerose	Almirall Hermal GmbH
11	Belatacept	Nulojix®	Nierentransplantation	Bristol-Myers Squibb GmbH & Co. KGaA
12	Emtricitabin, Rilpivirin, Tenofoviridisoproxil	Eviplera®	HIV Infektion	Gilead Sciences GmbH
13	Rilpivirin	Edurant®	HIV-Infektion	Janssen-Cilag GmbH
14	Belimumab	Benlysta®	Systemischer Lupus erythematodes	GlaxoSmithKline GmbH & Co. KG
15	Ipilimumab	Yervoy®	Melanom	Bristol-Myers Squibb GmbH & Co. KGaA
16	Vemurafenib	Zelboraf®	Melanom	Roche Pharma AG
17	Vandetanib	Caprelsa®	Schilddrüsenneoplasien	AstraZeneca GmbH
18	Pasireotid (OD)	Signifor®	Hypophysendysfunktion	Novartis Pharma GmbH
19	Ivacaftor (OD)	Kalydeco™	Mukoviszidose mit G551D-Mutation	Vertex Pharmaceuticals

Anhang

				GmbH
20	Linagliptin (§ 35a Abs. 5b)	Trajenta®	Diabetes mellitus Typ 2	Boehringer Ingelheim International GmbH
21	Perampanel	Fycompa®	partielle Epilepsie	Eisai GmbH
22	Ruxolitinib (OD)	Jakavi®	Chronisch myeloproliferative Erkrankungen	Novartis Pharma GmbH
23	Acidiniumbromid	Eklira® Genuair® / Bretaris® Genuair®	COPD	Amirall Hermal GmbH
24	Axitinib	Inlyta®	Nierenzellkarzinom	Pfizer Pharma GmbH
25	Decitabin (OD)	Dacogen®	myeloische Leukämie	Janssen-Cilag GmbH
26	Crizotinib	Xalkori®	Nicht kleinzelliges Bronchialkarzinom	Pfizer Pharma GmbH
27	Aflibercept	Eylea®	neovaskuläre altersabhängige Makuladegeneration	Bayer Vital GmbH
28	Ingenolmebutat	Picato®	aktinische Keratose	LEO Pharma GmbH
29	Abirateronacetat (neues AWG)	Zytiga®	Prostatakarzinom	Janssen-Cilag GmbH

Anhang D

Kategoriensystem der Gründe für die Nichtberücksichtigung und das Versagen eines HRQoL-spezifischen ZN
1. Mängel im Studiendesign
1.1. eingeschränkte Bewertbarkeit aufgrund unvollständiger Dokumentation und Präsentation der Daten
1.2. systematische Unterschiede zwischen den Behandlungsgruppen
1.3. hohes Verzerrungspotenzial
1.3.1. hohes Verzerrungspotenzial aufgrund mangelnder Verblindung
1.3.2. hohes Verzerrungspotenzial aufgrund hoher Rate an Studienabbrechern
1.3.3. hohes Verzerrungspotenzial aufgrund unzureichender Darstellung des Umgangs mit <i>Lost-to-follow-up</i> Patienten und Protokollverletzern
1.4.4. hohes Verzerrungspotenzial aufgrund informativer Zensierung
2. Mängel des Erhebungsinstruments
2.1. zu geringe Rücklaufquote des Erhebungsinstruments
2.2. mangelnde Validierung des Erhebungsinstruments
3. Mängel in der Auswertung
3.1. Verletzung des ITT-Prinzips
3.2. unzureichende Responderanalyse/MID
4. Mangelnde Signifikanz/Relevanz der Daten
4.1. kein klinisch relevanter Unterschied zwischen den Behandlungsgruppen
4.2. kein statistisch signifikanter Unterschied zwischen den Behandlungsgruppen
5. formale Mängel
5.1. keine NB erfolgt
5.2. aus formalen Gründen nicht in die NB einbezogen

Literaturverzeichnis

Ärzte Zeitung (ÄZ) (2013): Hecken schlägt vor: Nutzenkategorie Lebensqualität, 2013. Online verfügbar unter http://www.aerztezeitung.de/politik_gesellschaft/anzneimittelpolitik/nutzenbewertung/article/847652/hecken-schlaegt-nutzenkategorie-lebensqualitaet.html, zuletzt geprüft am 24.05.2014.

Bortz, J.; Schuster, C. (2010): Statistik für Human- und Sozialwissenschaftler. 7. Aufl.: Springer.

Brenner, H. M. (1995): Quality of Life Assessment in Medicine: A Historical View of Basic Science and Applications. In: I. Guggenmoos-Holzmann (Hg.): Quality of life and health. Concepts, methods, and applications. Berlin, Boston: Blackwell Wissenschafts-Verlag, S. 41–58.

Bristol-Myers Squibb (BMS) GmbH & Co. KGaA (2012a): Dossier zur Nutzenbewertung gemäß § 35a SGB V - Belatacept (Nulojix™). Online verfügbar unter https://www.g-ba.de/downloads/92-975-82/2012-01-13_Modul4A_Belatacept.pdf, zuletzt geprüft am 24.05.2014.

Bristol-Myers Squibb (BMS) GmbH & Co. KGaA (2012b): Dossier zur Nutzenbewertung gemäß § 35a SGB V - Ipilimumab (Yervoy®). Online verfügbar unter https://www.g-ba.de/downloads/92-975-105/2012-01-30_Modul4A_Ipilimumab.pdf, zuletzt geprüft am 24.05.2014.

Bullinger, M. (2000a): Erfassung der gesundheitsbezogenen Lebensqualität mit dem SF-36-Health Survey. In: *Bundesgesundheitsbl - Gesundheitsforsch - Gesundheitsschutz* 43 (3), S. 190–197. DOI: 10.1007/s001030050034.

Bullinger, M. (Hg.) (2000b): Lebensqualitätsforschung aus medizinpsychologischer und -soziologischer Perspektive. Göttingen [u.a.]: Hogrefe, Verl. für Psychologie.

Bullinger, M. (2006): Methodische Grundlagen und Aspekte der Lebensqualität. In: *Dtsch Med Wochenschr* 131 Suppl. 1, S. 5–7. DOI: 10.1055/s-2006-941732.

Bullinger, M. (2014): Das Konzept der Lebensqualität in der Medizin – Entwicklung und heutiger Stellenwert. In: *ZEFQ* 108 (2-4), S. 97–103. DOI: 10.1016/j.zefq.2014.02.006.

Bullinger, M.; Schmidt, S. (2006): Methoden zur Lebensqualitätsbewertung in der Onkologie. In: H.-J. Schmoll, K. Höffken und K. Possinger (Hg.): Kompendium internistische Onkologie. Standards in Diagnostik und Therapie. 4. Aufl. Heidelberg: Springer Medizin, S. 2505–2516.

Caron, M.; Emery, M.; Marquis, P.; Piau, E.; Scott, J. (2008): Recent Trends in the Inclusion of Patient-Reported Outcome (PRO) Data in Approved Drugs Labeling by the FDA and EMEA. In: *PRO Newsletter* (40). Online verfügbar unter http://www.pro-newsletter.com/images/PDF_articles/prolabels.pdf, zuletzt geprüft am 24.05.2014.

Deutsches Institut für Medizinische Dokumentation und Information (DIMDI) (2011): Der Stellenwert von Patient Reported Outcomes (PRO) im Kontext von Health Technology Assessment (HTA) (Health Technology Assessment, 109).

Deuschinoff, G.; Friedrich, C.; Thiem, U.; Voigtmann, R.; Pientka, L. (2005): Lebensqualität in der Onkologie. Status Quo und Ausblick. In: *Der Onkologe* 11 (2), S. 164–172. DOI: 10.1007/s00761-004-0825-x.

Doward, L. C.; Gnanasakthy, A.; Baker, M. G. (2010): Patient reported outcomes: looking beyond the label claim. In: *Health Qual Life Outcomes* 89 (8), S. 1–9. DOI: 10.1186/1477-7525-8-89.

Doward, L. C.; McKenna, S. P. (2004): Defining patient-reported outcomes. In: *Value in Health* 7 Suppl. 1, S. 4–8. DOI: 10.1111/j.1524-4733.2004.7s102.x.

European Medicines Agency (EMA) (2005): Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicine products. Online verfügbar unter http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003637.pdf, zuletzt geprüft am 24.05.2014.

European Medicines Agency (EMA) (2013): CHMP assessment report. Stivarga. Online verfügbar unter http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/002573/WC500149166.pdf, zuletzt geprüft am 24.05.2014.

Fairclough, D. L. (2010): Design and analysis of quality of life studies in clinical trials. 2. Aufl. Boca Raton: CRC Press.

Fayers, P. M.; Machin, D. (2007): Quality of life. The assessment, analysis and interpretation of patient-reported outcomes. 2. Aufl. Chichester, England, Hoboken, NJ: John Wiley & Sons.

Fitzpatrick, R.; Davey, C.; Buxton, M.; Jones, D. R. (1998): Evaluating patient-based outcome measures for use in clinical trials. In: *Health Technol Assess* 2 (14). DOI: 10.3310/hta2140.

Fitzpatrick, R.; Davey, C.; Buxton, M. J.; Jones, D. R. (2001): Criteria for Assessing Patient Based Outcome Measures for Use in Clinical Trials. In: A. Stevens (Hg.): The advanced handbook of methods in evidence based healthcare. London, Thousand Oaks: SAGE Publications, S. 181–194.

Flechtner, H. (2001): Lebensqualität in onkologischen Studien. In: *Oncol Res Treat* 24 Suppl. 5, S. 22–27. DOI: 10.1159/000055183.

Food and Drug Administration (FDA) (2009): Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Online verfügbar unter <http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf>, zuletzt geprüft am 24.05.2014.

Gemeinsamer Bundesausschuss (G-BA) (2011a): Zusammenfassende Dokumentation über die Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII-Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Ticagrelor. Online verfügbar unter https://www.g-ba.de/downloads/40-268-1828/2011-12-15_AM-RL-XII_Ticagrelor_TrG.pdf, zuletzt geprüft am 24.05.2014.

Gemeinsamer Bundesausschuss (G-BA) (2012a): Zusammenfassende Dokumentation über die Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII-Beschlüsse über die Nutzenbewertung von

Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Pirfenidon. Online verfügbar unter https://www.g-ba.de/downloads/40-268-1904/2012-03-15_AM-RL-XII_Pirfenidon_TrG.pdf, zuletzt geprüft am 24.05.2014.

Gemeinsamer Bundesausschuss (G-BA) (2012b): Zusammenfassende Dokumentation über die Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII-Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Extrakt aus Cannabis Sativa (Wirkstoffkombination Delta-9-Tetrahydrocannabinol und Cannabidiol). Online verfügbar unter https://www.g-ba.de/downloads/40-268-1953/2012-06-21_AM-RL-XII_Extrakte%20aus%20Cannabis_TrG.pdf, zuletzt geprüft am 24.05.2014.

Gemeinsamer Bundesausschuss (G-BA) (2012c): Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V. Bewertung von Arzneimitteln für seltene Leiden nach § 35a Absatz 1 Satz 10 i.V.m. 5. Kapitel § 12 Nr. 1 Satz 2 Verfo - Ivacaftor. Online verfügbar unter <https://www.g-ba.de/downloads/92-975-130/Nutzenbewertung%20G-BA.pdf>, zuletzt geprüft am 24.05.2014.

Gemeinsamer Bundesausschuss (G-BA) (2012d): Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V. Bewertung von Arzneimitteln für seltene Leiden nach § 35a Absatz 1 Satz 10 i.V.m. 5. Kapitel § 12 Nr. 1 Satz 2 Verfo - Pasireotid. Online verfügbar unter <https://www.g-ba.de/informationen/nutzenbewertung/34/>, zuletzt geprüft am 24.05.2014.

Gemeinsamer Bundesausschuss (G-BA) (2012e): Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V. Bewertung von Arzneimitteln für seltene Leiden nach § 35a Absatz 1 Satz 10 i.V.m. 5. Kapitel § 12 Nr. 1 Satz 2 Verfo - Ruxolitinib. Online verfügbar unter <https://www.g-ba.de/downloads/92-975-141/Nutzenbewertung%20G-BA.pdf>, zuletzt geprüft am 24.05.2014.

Gemeinsamer Bundesausschuss (G-BA) (2012f): Zusammenfassende Dokumentation über die Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII-Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Pasireotid. Online verfügbar unter https://www.g-ba.de/downloads/40-268-2299/2012-12-06_AM-RL-XII_Pasireotid_ZD.pdf, zuletzt geprüft am 24.05.2014.

Gemeinsamer Bundesausschuss (G-BA) (2012g): Zusammenfassende Dokumentation über die Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII-Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Vemurafenib. Online verfügbar unter https://www.g-ba.de/downloads/40-268-2133/2012-09-06_AM-RL-XII_Vemurafenib_ZD.pdf, zuletzt geprüft am 24.05.2014.

Gemeinsamer Bundesausschuss (G-BA) (2012h): Zusammenfassende Dokumentation über die Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII-Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Fingolimod. Online verfügbar unter https://www.g-ba.de/downloads/40-268-1947/2012-03-29_AM-RL-XII_Fingolimod_ZD.pdf, zuletzt geprüft am 24.05.2014.

Gemeinsamer Bundesausschuss (G-BA) (2012i): Zusammenfassende Dokumentation über die Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII-Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Telaprevir. Online verfügbar unter https://www.g-ba.de/downloads/40-268-1949/2012-03-29_AM-RL-XII_Telaprevir_ZD.pdf, zuletzt geprüft am 24.05.2014.

Gemeinsamer Bundesausschuss (G-BA) (2012j): Zusammenfassende Dokumentation über die Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII-Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Belatacept. Online verfügbar unter https://www.g-ba.de/downloads/40-268-2076/2012-07-05_AM-RL-XII_Belatacept_ZD.pdf, zuletzt geprüft am 24.05.2014.

Gemeinsamer Bundesausschuss (G-BA) (2012k): Zusammenfassende Dokumentation über die Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII-Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Emtricitabin, Rilpivirin, Tenofoviridisoproxil. Online verfügbar unter https://www.g-ba.de/downloads/40-268-2096/2012-07-05_AM-RL-XII_Emtricitabin-Rilpivirin-Tenofovir_ZD.pdf, zuletzt geprüft am 24.05.2014.

Gemeinsamer Bundesausschuss (G-BA) (2012l): Zusammenfassende Dokumentation über die Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII-Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Belimumab. Online verfügbar unter https://www.g-ba.de/downloads/40-268-2102/2012-08-02_AM-RL-XII_Belimumab_ZD.pdf, zuletzt geprüft am 24.05.2014.

Gemeinsamer Bundesausschuss (G-BA) (2013a): Zusammenfassende Dokumentation über die Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII-Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Crizotinib. Online verfügbar unter https://www.g-ba.de/downloads/40-268-2589/2013-05-02_AM-RL-XII_Crizotinib_ZD.pdf, zuletzt geprüft am 24.05.2014.

Gemeinsamer Bundesausschuss (G-BA) (2013b): Zusammenfassende Dokumentation über die Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII-Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Abirateronacetat. Online verfügbar unter https://www.g-ba.de/downloads/40-268-2762/2013-07-04_AM-RL-XII_Abirateronacetat-neuesAWG_ZD.pdf, zuletzt geprüft am 24.05.2014.

Gemeinsamer Bundesausschuss (G-BA) (2013c): Zusammenfassende Dokumentation über die Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII-Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Ruxolitinib. Online verfügbar unter https://www.g-ba.de/downloads/40-268-2482/2013-03-07_AM-RL-XII_Ruxolitinib_ZD.pdf, zuletzt geprüft am 24.05.2014.

Gemeinsamer Bundesausschuss (G-BA) (2013d): Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V. Bewertung von Arzneimitteln für seltene Leiden nach § 35a Absatz 1 Satz

10 i.V.m. 5. Kapitel § 12 Nr. 1 Satz 2 VerfO - Decitabin. Online verfügbar unter https://www.g-ba.de/downloads/92-975-163/Decitabin_Nutzenbewertung%20G-BA.pdf, zuletzt geprüft am 24.05.2014.

Gemeinsamer Bundesausschuss (G-BA) (2013e): Zusammenfassende Dokumentation über die Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII-Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Linagliptin (neues Anwendungsgebiet). Online verfügbar unter https://www.g-ba.de/downloads/40-268-2405/2013-02-21_AM-RL-XII_Linagliptin_ZD.pdf, zuletzt geprüft am 24.05.2014.

Gemeinsamer Bundesausschuss (G-BA) (2014): Verfahrensordnung - Stand: 8. Mai 2014. Online verfügbar unter https://www.g-ba.de/downloads/62-492-873/VerfO_2014-03-20.pdf, zuletzt geprüft am 18.05.2014.

Gibis, B. (2008): HTA und Entscheidungsfindung - Regulation von Technologien. In: M. Perleth, R. Busse, A. Gerhardus, B. Gibis und D. Lühmann (Hg.): Health technology assessment. Konzepte, Methoden, Praxis für Wissenschaft und Entscheidungsfindung. Berlin: MWV Med. Wiss. Verl.-ges., S. 23–65.

Gilead Sciences GmbH (2012): Dokumentvorlage, Version vom 20.01.2011 Dossier zur Nutzenbewertung gemäß § 35a SGB V - Emtricitabin/Rilpivirin/Tenofoviridisoproxil (Eviplera®). Online verfügbar unter https://www.g-ba.de/downloads/92-975-86/2012-01-10_Modul4A_Emtricitabin-Rilpivirin-Tenofoviridisoproxil.pdf, zuletzt geprüft am 24.05.2014.

Graf-Baumann, T. (1990): "Lebensqualität": Gesundheitsökonomische Folgerungen. In: P. Schölmerich und G. Thews (Hg.): "Lebensqualität" als Bewertungskriterium in der Medizin. Symposium der Akademie der Wissenschaften und der Literatur, Mainz ; mit 18 Tabellen. Stuttgart [u.a.]: Fischer (2), S. 271–282.

Greiner, W.; Klose, K. (2014): Lebensqualitätsmessung und Utilities in der Gesundheitsökonomie. In: *ZEFQ* 108 (2-3), S. 120–125. DOI: 10.1016/j.zefq.2014.02.004.

Hazard Vallerand, A.; Breckenridge, D. M.; Hodgson, N. A. (2001): Theorien und Begriffsmodelle als Leitlinien lebensqualitätsbezogener Forschung. In: C. R. King (Hg.): Lebensqualität. Pflege- und Patientenperspektiven ; Theorie - Forschung - Praxis. 1. Aufl. Bern, Göttingen, Toronto, Seattle: Huber, S. 77–94.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2011a):

Abirateronacetat - Nutzenbewertung gemäß § 35a SGB V. Online verfügbar unter https://www.g-ba.de/downloads/92-975-26/2011-10-01-D-023_Abirateronacetat_IQWiG-Nutzenbewertung.PDF, zuletzt geprüft am 24.05.2014.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2011b): Pirfenidon - Nutzenbewertung gemäß § 35a SGB V. Online verfügbar unter https://www.g-ba.de/downloads/92-975-16/2011-09-15-D-020_Pirfenidon_IQWiG-Nutzenbewertung.pdf, zuletzt geprüft am 24.05.2014.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2012a): Axitinib – Nutzenbewertung gemäß § 35a SGB V. Online verfügbar unter https://www.g-ba.de/downloads/92-975-157/2012-12-21_A12-14_Axitinib_Nutzenbewertung-%C2%A735a.pdf, zuletzt geprüft am 24.05.2014.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2012b): Ivacaftor - Nutzenbewertung gemäß § 35a Abs. 1 Satz 10 SGB V. Online verfügbar unter <https://www.g-ba.de/downloads/92-975-131/Bewertung%20der%20Therapiekosten%20und%20Patientenzahlen%20IQWiG.pdf>, zuletzt geprüft am 24.05.2014.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2012c): Addendum zum Auftrag A12-02 (Rilpivirin/Emtricitabin/Tenofovir). Online verfügbar unter https://www.iqwig.de/download/A12-10_Addendum_zum_Auftrag_A12-02_Rilpivirin-Emtricitabin-Tenofovir.pdf, zuletzt geprüft am 24.05.2014.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2012d): Fingolimod - Nutzenbewertung gemäß § 35a Abs. 1 Satz 10 SGB V. Online verfügbar unter https://www.g-ba.de/downloads/92-975-42/2011-04-15-D-004_Fingolimod_IQWiG-Nutzenbewertung.PDF, zuletzt geprüft am 24.05.2014.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2012e): Extrakt aus Cannabis Sativa - Nutzenbewertung gemäß § 35a Abs. 1 Satz 10 SGB V. Online verfügbar unter https://www.g-ba.de/downloads/92-975-78/2011-07-01-D-010_Cannabis-Sativa_IQWiG-Nutzenbewertung.pdf, zuletzt geprüft am 24.05.2014.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2012f): Belimumab - Nutzenbewertung gemäß § 35a Abs. 1 Satz 10 SGB V. Online verfügbar unter https://www.g-ba.de/downloads/92-975-106/2011-07-27-D-012_Belimumab_IQWiG-Nutzenbewertung.pdf, zuletzt geprüft am 24.05.2014.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2012g): Linagliptin - Erneute Nutzenbewertung gemäß § 35a Absatz 5b SGB V. Online verfügbar unter https://www.g-ba.de/downloads/92-975-136/2012-11-29_A12-11_Linagliptin_Erneute%20Nutzenbewertung%20%C2%A7%2035a%20Absatz%205b%20SGB%20V.pdf, zuletzt geprüft am 24.05.2014.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2012h): Belatacept - Nutzenbewertung gemäß § 35a Abs. 1 Satz 10 SGB V. Online verfügbar unter https://www.g-ba.de/downloads/92-975-91/2012-07-15-D-011_Belatacept_IQWiG-Nutzenbewertung.pdf, zuletzt geprüft am 24.05.2014.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2012i): Telaprevir - Nutzenbewertung gemäß § 35a Abs. 1 Satz 10 SGB V. Online verfügbar unter https://www.g-ba.de/downloads/92-975-106/2011-07-27-D-012_Telaprevir_IQWiG-Nutzenbewertung.pdf, zuletzt geprüft am 24.05.2014.

ba.de/downloads/92-975-37/2011-10-15-D-022_Telaprevir_IQWiG-Nutzenbewertung.PDF, zuletzt geprüft am 24.05.2014.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2012j): Vemurafenib - Nutzenbewertung gemäß § 35a Abs. 1 Satz 10 SGB V. Online verfügbar unter https://www.g-ba.de/downloads/92-975-117/2012-03-15-D-029_Vemurafenib_IQWiG-Nutzenbewertung.pdf, zuletzt geprüft am 24.05.2014.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2013a): Allgemeine Methoden: Version 4.1. Köln. Online verfügbar unter https://www.iqwig.de/download/IQWiG_Methoden_Version_4-1.pdf, zuletzt geprüft am 24.05.2014.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2013b): Addendum Auftrag: Version: Stand : A 13 - 13 1.0 15.04. 20 13 Addendum zum Auftrag A12-15 (Crizotinib). Online verfügbar unter https://www.iqwig.de/download/A13-13_Addendum-zum-Auftrag-A12-15_Crizotinib.pdf, zuletzt geprüft am 24.05.2014.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2013c): Abirateronacetat (neues Anwendungsgebiet) - Nutzenbewertung gemäß § 35a SGB V. Online verfügbar unter https://www.g-ba.de/downloads/92-975-229/2013-04-11_A13-06_Abirateronacetat_neues-Anwendungsgebiet_Nutzenbewertung-IQWiG.pdf, zuletzt geprüft am 24.05.2014.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2013d): Crizotinib - Nutzenbewertung gemäß § 35a Abs. 1 Satz 10 SGB V. Online verfügbar unter https://www.g-ba.de/downloads/92-975-168/2013-02-13_A12-15_Crizotinib_Nutzenbewertung%20%C2%A7%2035a%20SGB%20V.pdf, zuletzt geprüft am 24.05.2014.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2014): Kuration vs. Palliation: Versuch einer Begriffsklärung. Online verfügbar unter https://www.iqwig.de/download/GA12-01_Arbeitspapier_Kuration-versus-Palliation.pdf, zuletzt geprüft am 24.05.2014.

InterMune Deutschland GmbH (2011): Dossier zur Nutzenbewertung gemäß § 35a SGB V - Pirfenidon (Esbriet®). Online verfügbar unter https://www.g-ba.de/downloads/92-975-15/2011-09-14_Modul4A_Pirfenidon.pdf, zuletzt geprüft am 24.05.2014.

International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (1998): Statistical Principles for Clinical Trials E9. Online verfügbar unter http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf, zuletzt geprüft am 24.05.2014.

Janssen-Cilag GmbH (2012a): Dossier zur Nutzenbewertung gemäß § 35a SGB V - Decitabin (Dacogen®). Online verfügbar unter https://www.g-ba.de/downloads/92-975-161/2012-10-31_Modul4A_Decitabin.pdf, zuletzt geprüft am 24.05.

- Janssen-Cilag GmbH (2012b):** Dossier zur Nutzenbewertung gemäß § 35a SGB V - Rilpivirin (Edurant®). Online verfügbar unter https://www.g-ba.de/downloads/92-975-90/2012-01-15_Modul4A_Rilpivirin.pdf, zuletzt geprüft am 24.05.2014.
- Janssen-Cilag GmbH (2013):** Dossier zur Nutzenbewertung gemäß § 35a SGB V - Abirateronacetat (Zytiga®). Online verfügbar unter https://www.g-ba.de/downloads/92-975-20/2011-09-30_Modul4A_Abirateronacetat.pdf, zuletzt geprüft am 24.05.2014.
- Johnston, B. C.; Patrick, D. L.; Busse, J. W.; Schünemann, H. J.; Agarwal, A.; Guyatt, G. H. (2013):** Patient-reported outcomes in meta-analyses--Part 1: assessing risk of bias and combining outcomes. In: *Health Qual Life Outcomes* 11 (1), S. 109. DOI: 10.1186/1477-7525-11-109.
- King, C. R. (Hg.) (2001):** Lebensqualität. Pflege- und Patientenperspektiven ; Theorie - Forschung - Praxis. 1. Aufl. Bern, Göttingen, Toronto, Seattle: Huber.
- Kitchen, H.; Rofail, D.; Caron, M.; Emery, M.-P. (2011):** Oncology patient-reported claims: maximizing the chance for success. In: *Eccancermedicalscience* 5, S. 212. DOI: 10.3332/ecancer.2011.212.
- Klag, M. J.; MacKenzie, E. J.; Carswell, C. I.; Bridges, J. F. P. (2008):** The Role of The Patient in Promoting Patient-Centered Outcomes Research. In: *Patient* 1 (1), S. 1–3. DOI: 10.2165/01312067-200801010-00001.
- Klakow-Franck, R. (2014):** Die Bedeutung von Lebensqualität für die Arbeit des Gemeinsamen Bundesausschusses. In: *ZEFQ* 108 (2-3), S. 151–156. DOI: 10.1016/j.zefq.2014.02.007.
- Klose, K.; Witte, J.; Kreimeier, S.; Greiner, W. (2013):** Lebensqualität als Wettbewerbsvorteil in der frühen Nutzenbewertung. In: *Public Health Forum* 21 (81), S. 14 e 1–14 e3. DOI: 10.1016/j.phf.2013.09.011.
- Kohlmann, T. (2010):** Patientenberichtete Studienendpunkte - Stand in Forschung und Praxis. In: *ZEFQ* 104 (3), S. 259–263. DOI: 10.1016/j.zefq.2010.03.014.
- Kohlmann, T. (2014):** Messung von Lebensqualität: So einfach wie möglich, so differenziert wie nötig. In: *ZEFQ* 108 (2-3), S. 104–110. DOI: 10.1016/j.zefq.2014.03.015.
- Lange, S.; Kaiser, T.; Beatrice-Schüler, Y.; Skipka, G. (2010):** Ein formalisiertes Vorgehen zur Bewertung der Relevanz von Gruppenunterschieden auf Skalen. IQWiG. Online verfügbar unter https://www.iqwig.de/download/Abstracts_IQWiG_im_Dialog_2010.pdf, zuletzt geprüft am 24.05.2014.
- Leverkus, F. (2014):** Methodik für die Auswertung von Lebensqualitätsdaten im Rahmen der Nutzenbewertung von Arzneimitteln. In: *ZEFQ* 108 (2-3), S. 111–119. DOI: 10.1016/j.zefq.2014.02.008.
- Levine, S. (1995):** The meanings of health, illness and quality of life. In: I. Guggenmoos-Holzmann (Hg.): *Quality of life and health. Concepts, methods, and applications*. Berlin, Boston: Blackwell Wissenschafts-Verlag, S. 7–14.

- Marquis, P.; Arnould, B.; Acquadro, C.; Roberts, W. M. (2006):** Patient-reported outcomes and health-related quality of life in effectiveness studies: pros and cons. In: *Drug Dev. Res.* 67 (3), S. 193–201. DOI: 10.1002/ddr.20077.
- Matthias, K. (2010):** Bewertung und Bedeutung von Studienendpunkten - Konsequenzen für evidenzbasierte Entscheidungen im Gesundheitswesen. In: *ZEFQ* 104 (3), S. 272–276. DOI: 10.1016/j.zefq.2010.03.008.
- Mayring, P. (2010):** Qualitative Inhaltsanalyse. Grundlagen und Techniken. 11. Aufl. Weinheim: Beltz.
- Moinpour, C. M.; Lovato, L. C. (1998):** Ensuring the quality of quality of life data: the Southwest Oncology Group experience. In: *Stat Med* 17 (5-7), S. 641–651.
- Mokkink, L. B.; Terwee, C. B.; Patrick, D. L.; Alonso, J.; Stratford, P. W.; Knol, D. L. et al. (2010):** The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. In: *Qual Life Res* 19 (4), S. 539–549. DOI: 10.1007/s11136-010-9606-8.
- Mühlbauer, V.; Teupen, S. (2014):** Gemeinsamer Bundesausschuss und Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Aufgaben und Struktur unter Berücksichtigung der Patientenbeteiligung. In: *Dtsch Med Wochenschr* 139 (4), S. 147–151. DOI: 10.1055/s-0033-1359933.
- Neiß, A. (1990):** Biometrische Aspekte in der Messung von „Lebensqualität“. In: P. Schölmerich und G. Thews (Hg.): "Lebensqualität" als Bewertungskriterium in der Medizin. Stuttgart [u.a.]: Fischer (2), S. 41–58.
- Novartis Pharma GmbH (2011):** Dossier zur Nutzenbewertung gemäß § 35a SGB V - Fingolimod (Gilenya®). Online verfügbar unter https://www.g-ba.de/downloads/92-975-43/2011-09-21_Modul4_Fingolimod.pdf, zuletzt geprüft am 24.05.2014.
- Osoba, D.; Bezjak, A.; Brundage, M.; Pater, J. (2007):** Evaluating health-related quality of life in cancer clinical trials: the National Cancer Institute of Canada Clinical Trials Group experience. In: *Value Health* 10 Suppl. 2, S. 138–145. DOI: 10.1111/j.1524-4733.2007.00278.x.
- Patient Centered Outcomes Research Institute (PCORI) (2012):** The Design and Selection of Patient-Reported Outcomes Measures (PROMs) for Use in Patient Centered Outcomes Research. Online verfügbar unter <http://www.pcori.org/assets/The-Design-and-Selection-of-Patient-Reported-Outcomes-Measures-for-Use-in-Patient-Centered-Outcomes-Research1.pdf>, zuletzt geprüft am 24.05.2014.
- Perleth, M. (2008):** Grundlagen und Prinzipien von Health Technology Assessment (HTA). In: M. Perleth, R. Busse, A. Gerhardus, B. Gibis und D. Lühmann (Hg.): Health technology assessment. Konzepte, Methoden, Praxis für Wissenschaft und Entscheidungsfindung. Berlin: MWV, S. 1–22.
- Pfizer Pharma GmbH (2011):** Dossier zur Nutzenbewertung gemäß § 35a SGB V - Tafamidis Meglumine (Vyndaqel®). Online verfügbar unter https://www.g-ba.de/downloads/92-975-71/2011-12-12_Modul4A_Tafamidis.pdf, zuletzt geprüft am 24.05.2014.

- Pfizer Pharma GmbH (2012):** Dossier zur Nutzenbewertung gemäß § 35a SGB V - Crizotinib (Xalkori®). Online verfügbar unter https://www.g-ba.de/downloads/92-975-167/2012-11-07_Modul4A_Crizotinib.pdf, zuletzt geprüft am 24.05.2014.
- Radoschewski, M. (2000):** Gesundheitsbezogene Lebensqualität - Konzepte und Maße. Entwicklung und Stand im Überblick. In: *Bundesgesundheitsbl - Gesundheitsforsch - Gesundheitschutz* 43 (3), S. 165–189. DOI: 10.1007/s001030050033.
- Ramsenthaler, C. (2013):** Was ist „Qualitative Inhaltsanalyse?“. In: M. Schnell, C. Schulz, H. Kolbe und C. Dunger (Hg.): *Der Patient am Lebensende*. Wiesbaden: Springer Fachmedien Wiesbaden, S. 23–42.
- Rapley, M. (2003):** *Quality of life research. A critical introduction*. London, Thousand Oaks, Calif: SAGE Publications.
- Raspe, H.-H. (1990):** Zur Theorie und Messung der "Lebensqualität" in der Medizin. In: P. Schölmerich und G. Thews (Hg.): "Lebensqualität" als Bewertungskriterium in der Medizin. Symposium der Akademie der Wissenschaften und der Literatur, Mainz ; mit 18 Tabellen. Stuttgart [u.a.]: Fischer (2), S. 23–40.
- Renneberg, B.; Hammelstein, P. (2006):** *Gesundheitspsychologie*. Heidelberg: Springer Medizin.
- Roche Pharma AG (2012):** Dossier zur Nutzenbewertung gemäß § 35a SGB V - Vemurafenib (Zelboraf®). Online verfügbar unter https://www.g-ba.de/downloads/92-975-112/2012-02-17_Modul4A_Vemurafenib.pdf, zuletzt geprüft am 24.05.2014.
- Rothman, M.L.; Beltran, P.; Cappelleri, J. C.; Lipscomb, J.; Teschendorf, B. (2007):** Patient-reported outcomes: conceptual issues. In: *Value in Health* 10 Supplement 2. DOI: 10.1111/j.1524-4733.2007.00269.x.
- Ruof, J.; Schwartz, F. W.; Schulenburg, J.-M.; Dintsios, C.-M. (2013):** Early benefit assessment (EBA) in Germany: analyzing decisions 18 months after introducing the new AMNOG legislation. In: *Eur J Health Econ*. DOI: 10.1007/s10198-013-0495-y.
- Schöffski, O. (2007a):** Lebensqualität als Ergebnisparameter in gesundheitsökonomischen Studien. In: O. Schöffski und Graf von der Schulenburg, J.-M. (Hg.): *Gesundheitsökonomische Evaluationen*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg, S. 321–334.
- Schöffski, O. (2007b):** Nutzentheoretische Lebensqualitätsmessung. In: O. Schöffski und Graf von der Schulenburg, J.-M. (Hg.): *Gesundheitsökonomische Evaluationen*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg, S. 335-385.
- Schölmerich, P.; Thews, G. (Hg.) (1990):** "Lebensqualität" als Bewertungskriterium in der Medizin. Symposium der Akademie der Wissenschaften und der Literatur. Stuttgart [u.a.]: Fischer (2).
- Schumacher, J.; Klaiberg, A.; Brähler, E. (2003):** *Diagnostische Verfahren zu Lebensqualität und Wohlbefinden*. Göttingen: Hogrefe.
- Schwalm, A.; Perleth, M.; Matthias, K. (2010):** Der Umgang des Gemeinsamen Bundesausschusses mit „schwacher“ oder fehlender Evidenz. In: *ZEFQ* 104 (4), S. 323–329. DOI: 10.1016/j.zefq.2009.12.029.

- Siegrist, J. (1990):** Grundannahmen und gegenwärtige Entwicklungsperspektiven einer gesundheitsbezogenen Lebensqualitätsforschung. In: P. Schölmerich und G. Thews (Hg.): "Lebensqualität" als Bewertungskriterium in der Medizin. Stuttgart [u.a.]: Fischer (2), S. 59–66.
- Sloan, J. A.; Halyard, M. Y.; Frost, M. H.; Dueck, A. C.; Teschendorf, B.; Rothman, M. L. (2007):** The Mayo Clinic manuscript series relative to the discussion, dissemination, and operationalization of the Food and Drug Administration guidance on patient-reported outcomes. In: *Value in Health* 10 Suppl. 2, S. 59–63. DOI: 10.1111/j.1524-4733.2007.00267.x.
- Spitzer, W. O. (1995):** Quality of Life and Functional Status as Target Variables for Research. In: I. Guggenmoos-Holzmann (Hg.): *Quality of life and health. Concepts, methods, and applications*. Berlin, Boston: Blackwell Wissenschafts-Verlag, S. 59–67.
- Stratmann-Schöne, D. (2001):** Ein präferenzbasierter Index zur ökonomischen Evaluation von Gesundheitsleistungen. 1. Aufl. Baden-Baden: Nomos-Verl.-Ges.
- Strech, D. (2010):** Werturteile in der Evidenzanalyse. Meist intransparent, oft konfliktbehaftet, nie vermeidbar. In: *ZEFQ* 104 (3), S. 168–176. DOI: 10.1016/j.zefq.2010.03.001.
- Testa, M. A.; Simonson, D. C. (1996):** Assessment of quality-of-life outcomes. In: *N Engl J Med* 334 (13), S. 835–840. DOI: 10.1056/NEJM199603283341306.
- Valderas, J. M.; Alonso, J. (2008):** Patient reported outcome measures: a model-based classification system for research and clinical practice. In: *Qual Life Res* 17 (9), S. 1125–1135. DOI: 10.1007/s11136-008-9396-4.
- Vertex Pharmaceuticals GmbH (2012):** Dossier zur Nutzenbewertung gemäß § 35a SGB V - Dokumentvorlage, Version vom 20.01.2011 Stand: 1 3.0 8.2012 Vertex Pharmaceuticals - Ivacaftor (Kalydeco™). Online verfügbar unter https://www.g-ba.de/downloads/92-975-129/2012-08-13_Modul4A_Ivacaftor.pdf, zuletzt geprüft am 24.05.2014.
- Walters, S. J. (2009):** *Quality of life outcomes in clinical trials and health-care evaluation. A practical guide to analysis and interpretation*. Chichester, U.K: John Wiley & Sons.
- Windeler, J.; Lange, S. (2008):** Nutzenbewertung in besonderen Situationen – Seltene Erkrankungen. In: *ZEFQ* 102 (1), S. 25–30. DOI: 10.1016/j.zgesun.2007.12.005.
- Wintner, L. M.; Giesinger, J. M.; Kemmler, G.; Sztankay, M.; Oberguggenberger, A.; Gamper, E.-M. et al. (2012):** Verwendung und Nutzen von Patient-Reported Outcomes in der onkologischen Behandlung: eine Übersicht. In: *Wien Klin Wochenschr* 124 (9-10), S. 293–303. DOI: 10.1007/s00508-012-0168-3.
- Wood, L.; Egger, M.; Gluud, L. L.; Schulz, K. F.; Juni, P.; Altman, D. G. et al. (2008):** Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. In: *BMJ* 336 (7644), S. 601–605. DOI: 10.1136/bmj.39465.451748.AD.

Wood-Dauphinee, S. (1999): Assessing quality of life in clinical research: from where have we come and where are we going? In: *J Clin Epidemiol* 52 (4), S. 355–363. DOI: 10.1016/S0895-4356(98)00179-6.

World Health Organization (WHO) (1948): Constitution of the World Health Organization. New York. Online verfügbar unter http://whqlibdoc.who.int/hist/official_records/constitution.pdf, zuletzt geprüft am 24.05.2014.

World Health Organization (WHO) (1993): Study protocol for the World Health Organization project to develop a Quality of Life assessment instrument (WHOQOL). In: *Qual Life Res* 2 (2), S. 153–159. DOI: 10.1007/BF00435734.

Zapf, W. (1984): Individuelle Wohlfahrt: Lebensbedingungen und wahrgenommene Lebensqualität. In: W. Glatzer und W. Zapf (Hg.): *Lebensqualität in der Bundesrepublik. Objektive Lebensbedingungen und subjektives Wohlbefinden*. Frankfurt am Main: Campus Verlag (Diagnostik für Klinik und Praxis), S. 13–26.

Zentner, A.; Busse, R.; Velasco Garrido, M. (2011): Aktuelle internationale Initiativen der evidenzbasierten Bewertung von Arzneimitteln. Implikationen für Zulassung und Health Technology Assessment in Deutschland und Europa. Berlin (Working Papers in Health Policy and Management). Online verfügbar unter http://www.mig.tu-berlin.de/fileadmin/a38331600/2011.publications/2011.zentner_HTA-Initiativen_FGReihe_vol6_FINAL.pdf, zuletzt geprüft am 24.05.2014.