



**Dokumentation und Würdigung
der Anhörung zum Entwurf der
Allgemeinen Methoden 6.0**

Version 1.0 vom 05.11.2020

Anschrift des Herausgebers

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
Im Mediapark 8
50670 Köln

Tel.: +49 221 35685-0

Fax: +49 221 35685-1

E-Mail: methoden@iqwig.de

Internet: www.iqwig.de

Inhaltsverzeichnis

	Seite
Tabellenverzeichnis	v
Abkürzungsverzeichnis	vi
1 Dokumentation der Anhörung	1
2 Würdigung der Anhörung	2
2.1 Würdigung der Stellungnahmen zu Abschnitt 1.3.3 „Methodische Standards in der Gesundheitsökonomie“	2
2.2 Würdigung der Stellungnahmen zu Abschnitt 2.1 „Produktspezifische Verfahrensabläufe“	2
2.3 Würdigung der Stellungnahmen zu Abschnitt 2.1.1 „Bericht“	3
2.4 Würdigung der Stellungnahmen zu Abschnitt 2.1.8 „Gesundheitsinformation“	3
2.5 Würdigung der Stellungnahmen zu Abschnitt 2.1.10 „HTA-Bericht“	3
2.6 Würdigung der Stellungnahmen zu Abschnitt 2.2.3 „Gewährleistung der fachlichen Unabhängigkeit“	4
2.7 Würdigung der Stellungnahmen zu Abschnitt 3.1.3 „Ermittlung des Schadens medizinischer Interventionen“	4
2.8 Würdigung der Stellungnahmen zu Abschnitt 3.2.1 „Auswirkung nicht publizierter Studienergebnisse auf das Fazit“	9
2.9 Würdigung der Stellungnahmen zum Abschnitt 3.2.2 „Dramatischer Effekt“ ..	10
2.10 Würdigung der Stellungnahmen zu Abschnitt 3.3.3 „Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V“	11
2.11 Würdigung der Stellungnahmen zu Abschnitt 3.4 „Nicht medikamentöse therapeutische Interventionen“	12
2.12 Würdigung der Stellungnahmen zu Abschnitt 3.5 „Diagnostische Verfahren“	13
2.13 Würdigung der Stellungnahmen zu Abschnitt 3.8 „Potenzialbewertung“	13
2.14 Würdigung der Stellungnahmen zu Kapitel 5 „Bewertungen zur Versorgung“	13
2.15 Würdigung der Stellungnahmen zu Abschnitt 5.1.4 „Bewertung der methodischen Leitlinienqualität“	14
2.16 Würdigung der Stellungnahmen zu Abschnitt 5.1.6 „Strukturierte Informationssynthese: Extraktion und Analyse der Empfehlungen“	14
2.17 Würdigung der Stellungnahmen zu Abschnitt 5.3.4 „Datenquellen“	15
2.18 Würdigung der Stellungnahmen zu Abschnitt 6.2 „Themensammlung“	15
2.19 Würdigung der Stellungnahmen zu Abschnitt 6.3.3 „1. Stufe des Auswahlverfahrens: Nominierung von Themen“	15

2.20	Würdigung der Stellungnahmen zu Abschnitt 6.3.4 „2. Stufe des Auswahlverfahrens: Auswahl von Themen, zu denen HTA-Berichte erstellt werden“	15
2.21	Würdigung der Stellungnahmen zu Abschnitt 7.2.1 „Themenkatalog gemäß dem Generalauftrag“	16
2.22	Würdigung der Stellungnahmen zu Abschnitt 7.9.3 „Nutzertesting“	16
2.23	Würdigung der Stellungnahmen zu Abschnitt 7.15.2 „Erfahrungsberichte“	17
2.24	Würdigung allgemeiner Stellungnahmen zu Abschnitt 8.1.1 „Recherche in bibliografischen Datenbanken“	17
2.25	Würdigung der Stellungnahmen zu Abschnitt 8.2.5 „Überprüfung der Vollständigkeit eines Studienpools“	18
2.26	Würdigung allgemeiner Stellungnahmen zu Abschnitt 8.4 „Recherche nach Leitlinien zur Erstellung von Leitliniensynopsen“	18
2.27	Würdigung der Stellungnahmen zu Abschnitt 9.1.2 „Zusammenhang zwischen Studientyp / -art und Fragestellung“	18
2.28	Würdigung der Stellungnahmen zu Abschnitt 9.3.3 „Beurteilung klinischer Relevanz“	18
2.29	Würdigung der Stellungnahmen zu Abschnitt 9.3.4 „Bewertung subjektiver Endpunkte bei offenen Studiendesigns“	23
2.30	Würdigung der Stellungnahmen zu Abschnitt 9.3.8 „Metaanalysen“	24
2.31	Würdigung der Stellungnahmen zu Abschnitt 9.3.10 „Subgruppenanalysen“	24
2.32	Würdigung der Stellungnahmen zu Abschnitt 9.3.11 „Umgang mit nicht oder nicht vollständig publizierten Daten“	25
2.33	Würdigung der Stellungnahmen zu Abschnitt 9.3.12 „Umgang mit unvollständigen Daten“	25
2.34	Würdigung der Stellungnahmen zu Abschnitt 9.3.13 „Umgang mit unterschiedlichen mittleren Beobachtungsdauern“	26
2.35	Würdigung der Stellungnahmen zu Abschnitt 9.3.15 „Auswertung abhängiger Daten“	27
2.36	Literatur.....	28
3	Dokumentation der wissenschaftlichen Erörterung – Teilnehmerliste, Tagesordnung und Protokoll.....	32
3.1	Teilnehmerliste der wissenschaftlichen Erörterung.....	32
3.2	Liste der Stellungnahmen, zu denen keine Vertreterin bzw. kein Vertreter an der wissenschaftlichen Erörterung teilgenommen hat.....	33
3.3	Tagesordnung der wissenschaftlichen Erörterung.....	33
3.4	Protokoll der wissenschaftlichen Erörterung.....	33
3.4.1	Begrüßung und Einleitung.....	34
3.4.2	Tagesordnungspunkt 1: Ausmaßbestimmung für SMDs (Abschnitt 3.3.3)	35
3.4.3	Tagesordnungspunkt 2: Festlegung der Responseschwelle von 15 % (Abschnitt 9.3.3).....	40

3.4.4	Tagesordnungspunkt 3: Festlegung von und Umgang mit Publikationsbias (Abschnitt 3.2.1).....	56
3.4.5	Tagesordnungspunkt 4: Aufbereitung und Darstellung von Schadensaspekten (Abschnitt 3.1.3).....	61
3.4.6	Tagesordnungspunkt 5: Umgang mit unvollständigen Daten (Abschnitt 9.2.12 und 9.3.13).....	65
3.4.7	Tagesordnungspunkt 6: Verschiedenes	71
Anhang A – Dokumentation der Stellungnahmen		75

Tabellenverzeichnis

Tabelle 1: Spannweite der empirisch erhobenen MIDs im Verhältnis zur Spannweite der jeweils verwendeten Skala in %	20
---	----

Abkürzungsverzeichnis

Abkürzung	Bedeutung
AGREE	Appraisal of Guidelines Research and Evaluation
AMNOG	Arzneimittelmarktneuordnungsgesetz
AOK	Allgemeine Ortskrankenkasse
BfArM	Bundesinstitut für Arzneimittel und Medizinprodukte
BMG	Bundesministerium für Gesundheit
COPD	Chronic obstructive pulmonary Disease (chronisch obstruktive Lungenerkrankung)
DMP	Disease-Management-Programm
DMP-A-RL	DMP-Anforderungen-Richtlinie
DVG	Digitale-Versorgung-Gesetz
EIRD	Implantateregister-Errichtungsgesetz
EMA	European Medicines Agency (Europäische Arzneimittel-Agentur)
EORTC	European Organisation for Research and Treatment of Cancer
EORTC QLQ-C30	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30
EPAR	European public Assessment Report
G-BA	Gemeinsamer Bundesausschuss
GKV	gesetzliche Krankenversicherung
GRC	Global Rating of Change
HTA	Health Technology Assessment
ICD	International Statistical Classification of Diseases and Related Health Problems (Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme)
IIT	Investigator initiated Trial
IQTIG	Institut für Qualitätssicherung und Transparenz im Gesundheitswesen
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ISR	Investigator sponsored Research
ITT	Intention to treat
MedDRA	Medical Dictionary for Regulatory Activities (Medizinisches Wörterbuch für Aktivitäten im Rahmen der Arzneimittelzulassung)
MID	Minimal important Difference
PRO	Patient-reported Outcome (patientenberichteter Endpunkt)
PRO-CTCAE	Patient-reported Outcome – Common Terminology Criteria for Adverse Events (patientenberichtete Version der Common Terminology Criteria for Adverse Events)
PT	Preferred Term (bevorzugter Begriff)

Abkürzung	Bedeutung
pU	pharmazeutischer Unternehmer
RCT	Randomized controlled Trial (randomisierte kontrollierte Studie)
SGB	Sozialgesetzbuch
SGRQ	St. George's Respiratory Questionnaire
SISAQOL	Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data
SMD	Standardized Mean Difference (standardisierte Mittelwertdifferenz)
SMQ	Standardized MedDRA Query (standardisierte MedDRA-Abfrage)
SOC	System Organ Class (Systemorganklasse)
TSVG	Terminservice- und Versorgungsgesetz
UE	unerwünschtes Ereignis
WIdO	Wissenschaftliches Institut der AOK

1 Dokumentation der Anhörung

Am 05.12.2019 wurde der Entwurf der Allgemeinen Methoden Version 6.0 veröffentlicht und zur Anhörung gestellt. Bis zum 31.01.2020 konnten schriftliche Stellungnahmen eingereicht werden. Es gingen 40 Stellungnahmen per E-Mail ein. Alle eingereichten Stellungnahmen sind im Anhang abgebildet.

Unklare Aspekte in den schriftlichen Stellungnahmen wurden in einer wissenschaftlichen Erörterung am 19.06.2020 im Hotel Park Inn, Köln diskutiert. Das Wortprotokoll der Erörterung befindet sich in Abschnitt 3.4.

2 Würdigung der Anhörung

Im Stellungnahmeverfahren zum Dokument „Allgemeine Methoden: Entwurf für Version 6.0“ vom 05.12.2019 gingen bis zum 31.01.2020 Stellungnahmen von 40 Organisationen, Institutionen, Firmen und Einzelpersonen ein. Alle Stellungnahmen wurden in der Überarbeitung des Methodenpapiers berücksichtigt. Sofern sich aus den Stellungnahmen Änderungen in diesen Abschnitten ergaben, ist dies im vorliegenden Dokument explizit beschrieben. Die angegebenen Abschnittsnummern beziehen sich – sofern nicht anders angegeben – auf den Entwurf für die Allgemeinen Methoden Version 6.0 vom 05.12.2019.

2.1 Würdigung der Stellungnahmen zu Abschnitt 1.3.3 „Methodische Standards in der Gesundheitsökonomie“

In 1 Stellungnahme wird die Frage aufgeworfen, warum wissenschaftliche Standards und Methoden der Gesundheitsökonomie durch Allokationsentscheidungen bestimmt werden.

In den Allgemeinen Methoden findet sich folgende Aussage: „Letztlich ist die Frage, auf Grundlage welcher Rechte, Ansprüche oder Bedürfnisse Ressourcen im Gesundheitswesen mit welchem Ziel und mit welcher Wirkung auf die Allokation und Distribution von Gütern bzw. Leistungen verteilt werden sollen, nur als Wertentscheidung aufzufassen, und sie bestimmt wiederum, welche wissenschaftlichen Standards und Methoden eingesetzt werden sollen.“ Hiermit ist u. a. gemeint, dass das Ziel der gesundheitsökonomischen Evaluation das methodische Vorgehen bestimmen kann (z. B. Wahl der Perspektive).

2.2 Würdigung der Stellungnahmen zu Abschnitt 2.1 „Produktspezifische Verfahrensabläufe“

1 Stellungnahme verweist auf eine inkonsistente Darstellung hinsichtlich der Beauftragung bzw. Initiierung von Gesundheitsinformationen zwischen Tabelle 1 und Abbildung 7 in Abschnitt 2.1.8 und regt an, weiterhin die Beauftragung von Gesundheitsinformationen durch das Bundesministerium für Gesundheit (BMG) und den Gemeinsamen Bundesausschuss (G-BA) zu ermöglichen.

Tatsächlich bestand eine Inkonsistenz zwischen Tabelle 1 und der Abbildung 7 in Abschnitt 2.1.8, die bereinigt wurde. Es wurde die Abbildung dahin gehend korrigiert, dass die Erstellung des Produkts Gesundheitsinformation im Generalauftrag [15] erfolgt bzw. die Themenfindung und Initiierung der Gesundheitsinformation entsprechend Institut-für-Qualität-und-Wirtschaftlichkeit-im-Gesundheitswesen(IQWiG)-intern verantwortet werden. Nichtsdestotrotz ist es dem G-BA bzw. dem BMG weiterhin und unverändert möglich, dem IQWiG Aufträge zur Erstellung evidenzbasierter Gesundheitsinformationen für Bürgerinnen und Bürger zu übertragen. Im Falle einer solchen Beauftragung erfolgt die Erarbeitung in Form eines Berichts, Rapid Reports oder Addendums, einschließlich entsprechender Produktspezifika hinsichtlich öffentlicher Stellungnahmeverfahren und Publikation. Das Produkt wird auf der IQWiG-Website veröffentlicht, die darin erarbeitete evidenzbasierte Gesundheitsinformation zudem auf www.gesundheitsinformation.de. Dieser Sachverhalt wurde in Abschnitt 2.1.8 klargestellt.

2.3 Würdigung der Stellungnahmen zu Abschnitt 2.1.1 „Bericht“

In 2 Stellungnahmen wird der Wegfall des Anhörungsverfahrens zu Berichtsplänen zur Bewertung von Untersuchungs- und Behandlungsmethoden nach § 139a Abs. 3 Nr. 1 Sozialgesetzbuch (SGB) V hinterfragt. In einer Stellungnahme wird zudem davon ausgegangen, dass der Wegfall des Anhörungsverfahrens eine Empfehlung des IQWiG sei.

Der Wegfall des Anhörungsverfahrens zu Berichtsplänen zur Bewertung von Untersuchungs- und Behandlungsmethoden nach § 139a Abs. 3 Nr. 1 SGB V ergibt sich zwangsläufig und wie im Methodenpapier beschrieben aus der Änderung des § 139a gemäß dem Implantateregister-Errichtungsgesetz (EIRD) [9]. Im Methodenpapier erfolgte die Klarstellung, dass das Anhörungsverfahren zum Berichtsplan entsprechend den gesetzlichen Regelungen im § 139a Abs. 5 SGB V erfolgt.

2.4 Würdigung der Stellungnahmen zu Abschnitt 2.1.8 „Gesundheitsinformation“

1 Stellungnahme regt die Einbindung von Expertinnen und Experten bzw. externen Sachverständigen auch in der Aktualisierung von Gesundheitsinformationen an. In diesem Zusammenhang regt sie zudem die Vereinheitlichung von Bezeichnungen zum Zwecke der Klarheit an.

Tatsächlich ist die Einbindung von externen Sachverständigen bei umfassenden oder grundlegenden inhaltlichen Änderungen in Form eines Reviews vorgesehen. Da der Abschnitt jedoch Unklarheiten aufwies, wurde er – auch in Hinblick auf die Beauftragung externer Reviews durch externe Sachverständige – umfassend redaktionell überarbeitet.

2.5 Würdigung der Stellungnahmen zu Abschnitt 2.1.10 „HTA-Bericht“

In 1 Stellungnahme wird gefragt, ob es sich bei Health Technology Assessment (HTA) kompakt um ein eigenständiges IQWiG-Produkt handele. Zudem wird darum gebeten, HTA kompakt in die Abbildung 9 in Abschnitt 2.1.10 HTA-Bericht aufzunehmen.

Bei HTA kompakt handelt es sich nicht um ein eigenständiges Produkt des IQWiG. HTA kompakt ist die allgemein verständliche Zusammenfassung des dazugehörigen HTA-Berichts und richtet sich an die bzw. den Themenvorschlagenden und andere Interessierte. Sie wird zusammen mit dem HTA-Bericht und der Dokumentation der Anhörung zum HTA-Bericht auf der Website des IQWiG veröffentlicht. Um diesen Sachverhalt verständlicher darzulegen, wurden der Abschnittstext und Abbildung 9 angepasst.

1 Stellungnahme verweist darauf, dass der G-BA und das BMG – im Gegensatz zu Angaben auf der IQWiG-Website – nicht als Adressaten des HTA-Berichts auf der IQWiG-Website genannt werden, und empfiehlt die Weiterleitung an G-BA und BMG ausdrücklich.

Die Inkonsistenz wurde beseitigt, d. h., es erfolgte die Klarstellung im Methodenpapier, dass neben den im Methodenpapier genannten IQWiG-Gremien auch der G-BA und das BMG den HTA-Bericht erhalten.

2.6 Würdigung der Stellungnahmen zu Abschnitt 2.2.3 „Gewährleistung der fachlichen Unabhängigkeit“

1 Stellungnahme empfiehlt Harmonisierungen der Nomenklatur zu „Interessenkonflikten“ mit internationalen Standards sowie die Aufnahme einer aktuellen Literaturstelle.

Der Empfehlung, die Begrifflichkeit „potenzieller Interessenkonflikt“ im Zusammenhang mit der bloßen Offenlegung von Beziehungen zu vermeiden, wird gefolgt. Die entsprechenden Ausführungen im Methodenpapier wie auch das entsprechende Formblatt wurden überarbeitet. Die Aufnahme neuer Literatur wurde geprüft, jedoch verworfen, da sich hierin keine neuen Aspekte fanden.

2.7 Würdigung der Stellungnahmen zu Abschnitt 3.1.3 „Ermittlung des Schadens medizinischer Interventionen“

Von mehreren Stellungnehmenden wird kritisiert, dass die in Abschnitt 3.1.3 beschriebene Methodik zur Auswahl der Endpunkte zu unerwünschten Ereignissen (UEs) nicht hinreichend konkretisiert sei. Insbesondere für den Weg der Auswahl über die Ergebnisse in den relevanten Studien seien die beschriebenen Anforderungen an die Unterschiede zwischen den Behandlungsgruppen oder die Mindesthäufigkeit bei nicht schweren UEs aus Sicht der Stellungnehmenden nicht nachvollziehbar und ermöglichten so keine systematische und vollständige Aufarbeitung der spezifischen UEs für die Nutzenbewertung.

Wie in Abschnitt 3.1.3 beschrieben ist es das Ziel, diejenigen potenziell relevanten unerwünschten Wirkungen zusammenzustellen, die bei der Entscheidungsfindung für oder gegen den Einsatz der zu evaluierenden Intervention einen wesentlichen Stellenwert haben. Die Fragestellungen der Nutzenbewertungen sind sehr heterogen, beispielsweise hinsichtlich der Schwere der zugrunde liegenden Erkrankungen. Aus diesem Grund werden sich bei der Auswahl der spezifischen UEs auf Basis der in den Studien aufgetretenen Ereignisse auch die Anforderungen an die Unterschiede zwischen den Behandlungsgruppen je nach Fragestellung unterscheiden. Weitere Konkretisierungen, wie zum Beispiel feste Grenzwerte für Unterschiede zwischen den Behandlungsgruppen, sind aus diesen Gründen in den Allgemeinen Methoden nicht beschrieben.

In den Stellungnahmen finden sich auch keine Vorschläge, wie diese Anforderungen zu konkretisieren seien. Auch auf Nachfrage in der Erörterung, ob und wie die Auswahl spezifischer UEs vor dem Hintergrund der sehr heterogenen Fragestellungen aus Sicht der Stellungnehmenden algorithmisch zu operationalisieren sei, wurden keine Vorschläge gemacht. Vielmehr wurde deutlich, dass es für die Stellungnehmenden Unklarheiten hinsichtlich verschiedener spezieller Aspekte bei der Auswahl gibt. So sei insbesondere unklar, wie das methodische Vorgehen bei der Auswahl der am besten geeigneten Operationalisierung und der Einschätzung der Ergebniskonsistenz von Ereignissen zum selben inhaltlichen Konstrukt ist. Um diese Aspekte deutlicher zu machen, wurden sie in den Allgemeinen Methoden anhand von Beispielen näher erläutert.

In 1 Stellungnahme wird vorgeschlagen, den im Vergleich zu den Methoden 5 gestrichenen Absatz zu den bei der Identifikation von relevanten UEs verwendeten Quellen beizubehalten. Die dort beschriebenen Quellen (wie beispielsweise epidemiologische Studien, Pharmakovigilanzdaten, Ergebnisse aus Tierexperimenten, Unterlagen der Zulassungsbehörden) sollten aus Sicht des Stellungnehmers weiterhin verwendet werden, da in Dossiers zur frühen Nutzenbewertung von Arzneimitteln sehr wichtige Daten häufig nicht aufgenommen würden.

Der im Vergleich zu den Methoden 5 gelöschte Absatz bezog sich ausschließlich auf Bewertungen nach § 139a SGB V, also explizit nicht auf Dossierbewertungen. Genannt wurden die Quellen, die zur Identifikation von möglicherweise relevanten Nebenwirkungen für die Bewertung dienen sollten. Die Erfahrung hat jedoch gezeigt, dass die genannten Quellen, wie beispielsweise Pharmakovigilanzdaten oder Ergebnisse aus Tierexperimenten, keine relevanten Informationen für die Identifikation von potenziell relevanten UEs liefern und eine Betrachtung dieser Quellen somit nicht notwendig ist. Die Wiederaufnahme des genannten Absatzes ist deshalb nicht sinnvoll. Unterlagen der Zulassungsbehörden, wie beispielsweise der öffentlich zugängliche Bewertungsbericht (European public Assessment Report, EPAR) der Europäischen Arzneimittel-Agentur (EMA), stellen hingegen weiterhin eine der relevanten Quellen für die Identifikation potenziell relevanter Nebenwirkungen dar, die für das Krankheitsbild oder die in der Studie bzw. den Studien eingesetzten Interventionen von besonderer Bedeutung sind. Sie werden bei der Zusammenstellung im Rahmen der Vorabrecherche herangezogen. Sie gehören auch zu den Unterlagen, die vom Hersteller mit dem Dossier gemäß Dossiervorlagen des G-BA [14] bereitgestellt werden müssen. Da sie aber nur eine von mehreren Quellen darstellen und für die bereitzustellenden Unterlagen primär die Verfahrensordnung des G-BA ausschlaggebend ist, wird auf eine explizite Auflistung in Abschnitt 3.1.3 verzichtet.

In 1 Stellungnahme wird es als problematisch angesehen, dass die Nutzenbewertung sich bei der Auswahl der spezifischen UEs ausschließlich auf die jeweils relevante Studie stützt und auf Berichte des pharmazeutischen Unternehmers (pU) zurückgegriffen wird. Zum einen werde nicht nur das neue Arzneimittel bewertet, zum anderen seien auch für neue Arzneimittel potenziell relevante UEs bereits erwartbar, weil sie beispielsweise einer schon bekannten Wirkstoffklasse angehören. Aus diesem Grund ist es aus Sicht des Stellungnehmenden erforderlich, die Vorabrecherche beizubehalten.

In der Dossierbewertung erfolgt die Auswahl der spezifischen UEs nicht allein auf Basis der in den eingeschlossenen Studien aufgetretenen UEs oder auf Basis von Berichten des pU selbst. Auch hier werden (zusätzlich) solche spezifischen unerwünschten Wirkungen ausgewählt, die für das Krankheitsbild oder die in der Studie bzw. den Studien eingesetzten Interventionen von besonderer Bedeutung sind. Dazu werden Quellen außerhalb der relevanten Studien oder von Berichten des pU (wie beispielsweise der EPAR oder Fachinformationen von zu bewertenden Arzneimitteln und zweckmäßiger Vergleichstherapie) herangezogen. Diese Unterlagen sind zum großen Teil bereits in Modul 5 des Dossiers enthalten, ggf. werden auch weitere Quellen (z. B. Publikationen) bei der Auswahl herangezogen. Auch Informationen, die sich aus Bewertungen zu Wirkstoffen aus der gleichen Wirkstoffklasse ergeben, fließen in die Auswahl

ein. Somit erfolgt in jedem Fall unabhängig von den Ergebnissen in den relevanten Studien und unabhängig vom Dossier des pU eine Auseinandersetzung damit, welche potenziell relevanten Nebenwirkungen für die jeweilige Bewertung von Bedeutung sind. Aufgrund der engen zeitlichen Fristen der Dossierbewertung erfolgt jedoch keine Vorabrecherche im Sinne von Abschnitt 8.1.1 der Allgemeinen Methoden. Insgesamt ergibt sich keine Änderung an den Allgemeinen Methoden.

In 1 Stellungnahme wird darauf hingewiesen, dass bei der Bewertung der Operationalisierung der ausgewählten UEs darauf Wert zu legen sei, dass diese a priori geplant war.

Die Überprüfung der Präspezifizierung ist Teil der Einschätzung der Ergebnissicherheit einer Operationalisierung. Ausgewählte Operationalisierungen, die lediglich einen bevorzugten Begriff (PT) oder eine Systemorganklasse (SOC) gemäß Medizinischem Wörterbuch für Aktivitäten im Rahmen der Arzneimittelzulassung (MedDRA) darstellen, werden in Studien regelmäßig erhoben und in Studienberichten regelmäßig dargestellt. Dies wird mit der Präspezifizierung gleichgestellt. Für alle anderen Operationalisierungen, die beispielsweise eine Zusammenstellung mehrerer PTs oder SOC darstellen (z. B. Standardized MedDRA Queries, SMQs), wird stets geprüft, ob sie in der Studie präspezifiziert wurden. In der Regel werden in die Dossierbewertung a priori festgelegte Operationalisierungen herangezogen. Insgesamt ergibt sich keine Änderung an den Allgemeinen Methoden.

In 1 Stellungnahme wird gefordert, dass bei der Bewertung der Ergebnisse zu UEs auch immer die kumulative Inzidenz unter Berücksichtigung des Schweregrads zu berücksichtigen sei. Dass bei einem Arzneimittel UEs später auftreten als bei einem anderen Arzneimittel, dürfe aufgrund der Betrachtung der Überlebenszeitanalysen nicht mit einem grundsätzlichen Vorteil verwechselt werden. Die Beobachtungsdauer müsse so gewählt werden, dass beispielsweise immunvermittelte UEs trotz späteren Auftretens sicher erfasst würden.

Es ist richtig, dass es bei der Betrachtung von Überlebenszeitanalysen vorkommen kann, dass sich ein statistisch signifikanter Unterschied allein aufgrund der Verzögerung des Auftretens von Ereignissen zeigt. Bei der Betrachtung von UEs werden Überlebenszeitanalysen allerdings in der Regel nur in den Fällen herangezogen, in denen unterschiedliche Beobachtungsdauern zwischen den Behandlungsarmen auftreten. Das ist regelmäßig der Fall bei der Bewertung von Studien zu onkologischen Fragestellungen, weil UEs auch heute noch zumeist nur während der Behandlung erhoben und bei Behandlungsabbruch oder Behandlungswechsel die entsprechenden Beobachtungen zensiert werden (siehe Abschnitt 9.3.12 der finalen Version 6.0 der Allgemeinen Methoden). Sind die Unterschiede zwischen den Beobachtungsdauern der einzelnen Studienarme zu groß, sind die Ergebnisse für UEs auf Basis der kumulativen Inzidenzen nicht sinnvoll interpretierbar. Damit in solchen Situationen überhaupt Schlussfolgerungen auf Basis einer adäquaten Inferenzstatistik möglich sind, wird in der Nutzenbewertung auf Verfahren für Überlebenszeiten zurückgegriffen. Da die kumulativen Inzidenzen bei zu stark unterschiedlichen Beobachtungsdauern nicht verwertbar sind, ist eine entsprechende Berücksichtigung auch nicht sinnvoll. Die Stellungnehmenden machen auch keinen Vorschlag, wie

die kumulativen Inzidenzen in der Situation unterschiedlicher Beobachtungszeiten berücksichtigt werden sollen.

In der Nutzenbewertung muss auf die vorhandenen Daten aus den relevanten Studien zurückgegriffen werden. Die Beobachtungsdauern für die verschiedenen Endpunkte, auch für UEs, sind durch das Studienprotokoll vorgegeben. Um die oben genannten Probleme zu vermeiden und auch die kumulativen Inzidenzen zu den UEs sinnvoll interpretieren zu können, wäre es notwendig, dass in Studien die Beobachtungsdauer so gewählt wird, dass auch UEs – wie das Überleben – über das Behandlungsende hinaus bis zum Studienende bzw. zum Versterben der Patientinnen und Patienten beobachtet werden. Dies muss jedoch in der Studienplanung berücksichtigt werden.

In 1 Stellungnahme wird gefordert, dass bei der Bewertung der Nebenwirkungen die Patientenperspektive berücksichtigt wird.

Die Perspektive der Betroffenen ist sehr wichtig, auch bei der Bewertung von Nebenwirkungen von Interventionen. Aus diesem Grund ist bei der Einbindung von Betroffenen die Identifikation von potenziell relevanten Nebenwirkungen der zu bewertenden Interventionen ein wesentlicher Bestandteil. Teilweise werden in relevanten Studien Instrumente zu patientenberichteten Endpunkten verwendet, die auch Nebenwirkungen der Behandlungen abbilden. Ein Beispiel ist der European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30), auch wenn er Nebenwirkungen nicht vollständig abbildet. Konzepte wie die patientenberichtete Version der Common Terminology Criteria for Adverse Events (PRO-CTCAE) [3] könnten dabei unterstützen, die Perspektive der Betroffenen bei der Erhebung und auch Bewertung von Nebenwirkungen besser zu berücksichtigen. Dies wäre auch aufgrund der teilweise geringen Übereinstimmung der Einschätzung zwischen Patientinnen und Patienten und behandelnden Ärztinnen und Ärzten wichtig [2]. Diese sind in den relevanten Studien der bisherigen Bewertungen aber nicht erhoben worden. Sollten zukünftig in den relevanten Studien Ergebnisse mittels PRO-CTCAE erhoben werden, werden diese auch in der Nutzenbewertung berücksichtigt.

In 1 Stellungnahme wird gefordert, dass bei der Weiterentwicklung der Methoden zur Bewertung des Schadens Patientenpräferenzen berücksichtigt werden sollen. Nur so könne gewährleistet werden, dass alle Ergebnisse medizinischer Interventionen miteinander vergleichbar sind und die Patientenpräferenzen respektiert werden.

Aus der Stellungnahme wird nicht deutlich, auf Basis welcher Informationen die angesprochenen Patientenpräferenzen in die Bewertung einfließen sollen. Patientenpräferenzstudien wurden im Rahmen der frühen Nutzenbewertung bereits mehrere Male vom jeweiligen pU vorgelegt (siehe beispielsweise [21-23]). Diese wurden jeweils daraufhin geprüft, ob sie geeignet waren, entweder die Patientenrelevanz von Endpunkten einzuschätzen oder eine Priorisierung der Endpunkte zu rechtfertigen. Geeignet waren sie bisher aber in keinem Fall. Der Hauptgrund dafür war, dass die Auswahl der in die Patientenpräferenzstudie eingehenden

Endpunkte nicht adäquat, da selektiv und unvollständig, war. In den Nutzenbewertungen, sowohl nach § 139a SGB V als auch in der frühen Nutzenbewertung nach § 35a SGB V, werden Betroffene bzw. Patientenorganisationen eingebunden, jeweils auf unterschiedlichen Wegen (siehe dazu den Abschnitt 2.2.1 der finalen Version 6.0 der Allgemeinen Methoden). Die Identifikation von potenziell relevanten Nebenwirkungen der zu bewertenden Interventionen ist bei der Einbindung der Betroffenen immer ein wesentlicher Bestandteil. Insgesamt ergibt sich keine Änderung an den Allgemeinen Methoden.

In 1 Stellungnahme wird argumentiert, dass die Behauptung, UEs würden am besten in randomisierten kontrollierten Studien (RCTs) dargestellt, nicht korrekt sei. Im Folgenden werden in der Stellungnahme verschiedene Ergänzungen in den Allgemeinen Methoden vorgeschlagen, die sich darauf beziehen, dass sich seltene UEs am ehesten in prospektiven Registern erfassen lassen und dass eine Abgrenzung interventionsgenerierter Effekte von gleichartigen indikationsspezifischen Effekten nur durch vergleichende Studien möglich sei.

Hier sind verschiedene Aspekte voneinander zu trennen. Es ist zwar richtig, dass sich seltene (und auch sehr spät auftretende) UEs in den üblichen randomisierten Zulassungsstudien nicht hinreichend abbilden lassen, weil diese Studien in der Regel auch gar nicht darauf ausgelegt sind, Effekte in seltenen UEs aufzudecken. Für die Signaldetektion, die im Rahmen der Zulassung von Bedeutung ist, können deshalb nicht randomisierte prospektive Studien hilfreich sein. Wichtig ist bei den Fragestellungen der Nutzenbewertung aber der Vergleich zwischen einer Intervention und einer Vergleichstherapie, der Aussagen zum Schaden zwischen den Therapiealternativen erlaubt. Dazu bedarf es hinreichend ergebnissicherer vergleichender Daten. Deshalb ist für die Nutzenbewertung die RCT nach wie vor das Mittel der Wahl. In einer RCT auftretende Effekte sind stets auf die untersuchten Interventionen zurückzuführen, da alle weiteren Einflussfaktoren durch die Randomisierung gleich verteilt sind. Dies ist in nicht randomisierten Studien nicht bzw. nur bei ausreichend großen oder deterministischen Effekten der Fall.

Die Abgrenzung von indikationsspezifischen Effekten von interventionsspezifischen Effekten ist auch in RCTs nur bedingt möglich, da dazu eine Einschätzung jedes einzelnen UE notwendig wäre. Diese Einschätzungen der Kausalität in Bezug auf eine Intervention werden in Studien zwar regelhaft vorgenommen, sind aber sehr subjektiv und daher mit sehr hoher Unsicherheit behaftet. Für eine genauere Einschätzung „echter“ Nebenwirkungen kann es eher sinnvoll sein, Ereignisse, die eindeutig auf die Progression der Grunderkrankung zurückzuführen sind (z. B. Exazerbationen bei chronisch obstruktiver Lungenerkrankung [COPD]), in den Gesamtraten der UEs, schweren und schwerwiegenden UEs nicht zu berücksichtigen. Dies spiegelt sich auch in den Dossievorlagen des G-BA wider [14]. Insgesamt ergibt sich kein Änderungsbedarf für die Allgemeinen Methoden.

In 1 Stellungnahme wird darauf hingewiesen, dass die Gesamtrate der Therapieabbrüche derjenigen der Abbrüche wegen schwerer UEs gegenüberzustellen sei, da ein Unterschied

bestehe zwischen „keine laufende Therapie mehr“ und „Schaden am Patienten durch die Therapie, der zum Abbruch führt“.

Beide genannten Werte, die Gesamtrate der Therapieabbrüche als auch die Rate der Abbrüche wegen UEs, sind für die Nutzenbewertung von Bedeutung, wenn auch an jeweils unterschiedlichen Stellen der Bewertung. Für die Bewertung des Schadens einer Intervention ist ausschließlich der Endpunkt Abbrüche wegen UEs relevant, und zwar aus dem vom Stellungnehmenden genannten Grund, dass die Patientinnen und Patienten ein UE erleiden, das dazu führt, dass sie die Therapie abbrechen. Es ist dabei allerdings nicht zwingend notwendig, nur die schweren UEs zu betrachten, die zum Abbruch geführt haben. Es kann auch nicht schwere bzw. nicht schwerwiegende UEs geben, die für Patientinnen und Patienten so belastend sind, dass sie sich dazu entscheiden, ihre gegenwärtige Therapie abzubrechen. Insgesamt ergibt sich keine Änderung an den Allgemeinen Methoden.

2.8 Würdigung der Stellungnahmen zu Abschnitt 3.2.1 „Auswirkung nicht publizierter Studienergebnisse auf das Fazit“

Mehrere Stellungnehmende kritisierten, dass die Ausführungen schwer verständlich seien. Insbesondere bleibe offen, wer als „Dritter“ für eine Datenübermittlung infrage komme. Auch wurde angeregt, die Tabelle aufzuteilen in Szenarien mit und ohne geplante Datenübermittlung durch Dritte.

Dem Hinweis der Stellungnehmenden folgend wurde im Methodenpapier ergänzt, dass sich die Datenübermittlung durch Dritte insbesondere auf Herstelleranfragen bezieht. Zur besseren Abgrenzung von Szenarien mit und ohne geplante Datenübermittlung durch Dritte wurden sowohl in der Tabelle als auch im Text Zwischenüberschriften eingefügt.

In mehreren Stellungnahmen wurde darum gebeten zu spezifizieren, anhand welcher Kriterien der Umfang der fehlenden Daten als mäßig bzw. erheblich eingestuft wird.

Im Umgang mit fehlenden, vermutlich unpublizierten Daten erscheinen feste Kriterien (z. B. Fehldatenanteil > 30 % ist erheblich) schwer zu rechtfertigen, weil die Gründe für und das Wissen um unpublizierte Daten je nach Themenfeld stark variieren können. Wie auch in der mündlichen Erörterung erläutert mussten bislang erst in einer IQWiG-Nutzenbewertung Grenzwerte projektspezifisch festgelegt werden [24,25]. Diese lagen bei < 10 % (vernachlässigbarer Fehldatenanteil), 10 % bis 30 % (mäßiger Fehldatenanteil) bzw. > 30 % (erheblicher Fehldatenanteil) und können der groben Orientierung dienen.

Mehrere Stellungnehmende äußerten die Sorge, dass Herstellerinnen und Hersteller trotz einer vollständigen Übermittlung ihrer Studiendaten benachteiligt werden könnten, wenn aufgrund des Fehlens weiterer Studiendaten eine insgesamt erhebliche Datenunvollständigkeit zustande käme und daher kein Nutzen für eine Intervention abgeleitet werden könne. Eine Herstellerin oder ein Hersteller – so die Stellungnehmenden – könne nicht für das Fehlen von Studiendaten

verantwortlich gemacht werden, die sich nicht in seinem Verantwortungs- und Zugriffsbereich befänden.

Ein wesentlicher Bestandteil einer Nutzenbewertung ist die Einschätzung der Aussage-sicherheit der vorliegenden Daten. Wenn für die Nutzenbewertung ein Großteil der relevanten Daten nicht zur Verfügung steht, sind Schlussfolgerungen auf Basis der zur Verfügung stehenden Daten potenziell hochgradig verzerrt. Dies gilt in der Regel auch dann, wenn einzelne Herstellerinnen und Hersteller die ihnen vorliegenden Daten vollständig zur Verfügung gestellt haben. Diese potenzielle Verzerrung ist in der Schlussfolgerung aus wissenschaftlichen Gründen entsprechend zu berücksichtigen und kann nicht als Benachteiligung einer Herstellerin oder eines Herstellers angesehen werden. Dieses Vorgehen ist im Übrigen bereits seit ca. 10 Jahren in den Allgemeinen Methoden unverändert verankert [20].

Dennoch kann es in bestimmten Fällen (z. B. je nach Datenübermittlungsverfahren und Datenanteil, Aufteilung der Nutzenbewertung in Teilfragestellungen) trotz Nachweises fehlender Daten möglich sein, das Ausmaß des Publikationsbias zu verringern, in dem ein nachweislich unverzerrter Teil-Studienpool bewertet wird. Falls zu einer Fragestellung Studien unterschiedlicher Herstellerinnen und Hersteller vorliegen, kann die Beschränkung auf die Daten einer einzelnen Herstellerin bzw. eines einzelnen Herstellers hierfür jedoch ungeeignet sein [19]. Dies wurde in den Allgemeinen Methoden ergänzt.

2.9 Würdigung der Stellungnahmen zum Abschnitt 3.2.2 „Dramatischer Effekt“

In mehreren Stellungnahmen wird die in Abschnitt 3.2.2 genannte Grenze eines relativen Risikos von größer als 10 für einen dramatischen Effekt kritisiert. Aus Sicht der Stellung-nehmenden sei eine solche pauschale Bewertung nicht zeitgemäß, weil sie in bestimmten Datenkonstellationen eine nicht zu erreichende Hürde darstelle.

Abhängig von der Ergebnissicherheit der zugrunde liegenden Evidenz muss ein beobachteter Effekt hinreichend groß sein, damit mit ausreichender Sicherheit darauf geschlossen werden kann, dass er nicht auf systematischer Verzerrung beruht. Die Angabe der von Glasziou et al. [16] vorgeschlagenen Näherung an einen sogenannten dramatischen Effekt ist dabei nur als grobe Richtschnur zu verstehen, die in den Nutzenbewertungen des IQWiG nicht algorithmisch angewendet wird. Vielmehr wird abhängig von der konkreten Datenlage in der speziellen Bewertungssituation eine entsprechende Abschätzung vorgenommen, ob der beobachtete Effekt durch systematische Verzerrung zu erklären ist. In der Situation eines qualitativ hochwertigen nicht randomisierten Vergleichs, in dem für alle wesentlichen Confounder adjustiert wurde, sind auch andere Grenzen denkbar [26]. Insofern ist den Stellungnehmenden zu wider-sprechen, dass es sich dabei um eine pauschale Bewertung handele. Dennoch kann es Situa-tionen geben, in denen auf Basis von unsicherer Evidenz, beispielsweise beim Vergleich einzelner Arme aus verschiedenen Studien, ein Zusatznutzen nicht abgeleitet werden kann, wenn der zugrunde liegende Effekt nicht groß genug ist. In einem solchen Fall bedarf es anderer Studiendesigns, mit denen sich auch kleinere Effekte hinreichend sicher ableiten lassen.

2.10 Würdigung der Stellungnahmen zu Abschnitt 3.3.3 „Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V“

Im Entwurf der Allgemeinen Methoden wurde eine neue Methode zur Feststellung des Ausmaßes des Zusatznutzens bei stetigen Endpunkten vorgeschlagen, was in den Stellungnahmen vielfach begrüßt wird. In vielen Stellungnahmen wird jedoch darauf hingewiesen, dass es Unklarheiten gibt bezüglich der Ableitung und Anwendung der neu vorgeschlagenen Schwellenwerte zur Feststellung des Ausmaßes des Zusatznutzens bei stetigen Endpunkten. In manchen Stellungnahmen wird konkret vorgeschlagen, eine ausführlichere Begründung im Anhang des Methodenpapiers vorzunehmen.

Die Begründung der neuen Schwellenwerte zur Feststellung des Ausmaßes des Zusatznutzens bei stetigen Endpunkten war im Entwurf der Allgemeinen Methoden relativ knapp dargestellt. Bei der Überarbeitung des Entwurfs wurde festgestellt, dass es bezüglich der Ableitung von Schwellenwerten zur Feststellung des Ausmaßes des Zusatznutzens bei stetigen Endpunkten noch Klärungsbedarf gibt. Das Vorgehen hierzu wird daher nicht in die finale Version 6.0 der Allgemeinen Methoden aufgenommen. Stattdessen wird ein neuer Vorschlag in einem zukünftigen Entwurf erneut zur Stellungnahme gegeben.

In 1 Stellungnahme wird vorgeschlagen, einen neuen Unterabschnitt zu Kosten einzuführen, in dem u. a. zu beschreiben wäre, welche Packungen für eine Kostenkalkulation heranzuziehen sind, welche Abzüge vom Apothekenverkaufspreis abgezogen werden, wann Kostenspannen angesetzt werden, welche Kostenpositionen saldiert werden, inwieweit Companion Diagnostics als zusätzliche Gesetzliche-Krankenversicherung(GKV)-Leistung berücksichtigt werden, welche Sonderregeln für die Kostendarstellung bei Orphan Drugs bestehen. Der Stellungnehmende nennt als Grund dafür Abweichungen von den G-BA-Beschlüssen und Inkonsistenzen in der Bewertung des IQWiG.

Die methodischen Grundlagen für die Bewertung sind in der Dossievorlage (Abschnitt 3.3.1 bis 3.3.4) dargestellt. So finden sich dort z. B. in Abschnitt 3.3.3 Angaben dazu, welche Packungen für die Kostenbestimmung herangezogen werden sollen oder um welche Rabatte der Apothekenverkaufspreis zu reduzieren ist. Des Weiteren finden sich u. a. in den Abschnitten 3.3.1 und 3.3.2 Angaben dazu, wann Spannen anzugeben sind.

In jeder Dossierbewertung wird zunächst die Herleitung der vom pU angegebenen Kosten in ihren Einzelschritten genau beschrieben. Die Bewertung dieses Vorgehens wird in der Dossierbewertung, spezifisch in Bezug auf die zu beantwortende Fragestellung, für die einzelnen Berechnungsschritte des pU transparent dargestellt und jeweils begründet.

Aus § 139a Abs. 1 SGB V ergibt sich, dass das IQWiG ein fachlich unabhängiges wissenschaftliches Institut ist. In § 139b Abs. 4 SGB V ist geregelt, dass das Institut seine Arbeitsergebnisse dem G-BA als Empfehlungen zuleitet. Zudem berücksichtigt der G-BA bei seinem Beschluss auch ggf. neue Erkenntnisse, die sich aus der Anhörung zur Dossierbewertung ergeben haben. Damit müssen die Bewertungsergebnisse des IQWiG und des G-BA nicht

notwendigerweise gleich ausfallen. Für das Methodenpapier ergibt sich daher kein Änderungs- oder Ergänzungsbedarf.

In 1 Stellungnahme wird darauf hingewiesen, dass die allgemeine Aussage „Je nach Qualität der Zielgröße muss das Konfidenzintervall vollständig unterhalb eines bestimmten Schwellenwertes liegen, um das Ausmaß als gering, beträchtlich oder erheblich anzusehen“ nicht für alle Zielgrößen richtig ist.

Dieser Hinweis ist korrekt; die Formulierung in den Allgemeinen Methoden wurde entsprechend angepasst.

2.11 Würdigung der Stellungnahmen zu Abschnitt 3.4 „Nicht medikamentöse therapeutische Interventionen“

Einige Stellungnehmende hinterfragen, ob die besonderen Herausforderungen und Schwierigkeiten klinischer Studien zu nicht medikamentösen Interventionen in den Methoden des Instituts ausreichend beachtet werden. Weitere Spezifikationen hierzu werden angeregt. Konkrete Vorschläge werden jedoch nur zur Definition objektiver Endpunkte und zum Stellenwert einer verblindeten Datenanalyse gemacht.

Die in der Stellungnahme geäußerte Sorge, dass es in manchen klinischen Fragestellungen keine objektiv erfassbaren, patientenrelevanten Endpunkte geben könne, erscheint weitgehend unbegründet. Gerade in der Zahnmedizin, worauf der Stellungnehmende sich bezieht, gibt es mit dem Zahnverlust einen objektiven Endpunkt, der bei sehr vielen zahnmedizinischen Erkrankungen gut messbar ist.

Der Vorschlag, die Verblindung der Datenanalyse als Aspekt der internen Validität mit zu betrachten, entspricht nicht dem aktuellen Stand des methodischen Wissens. Wichtiger als eine Verblindung der Datenanalyse ist die prospektive Planung der Analyse. Denn eine prospektive, also a priori geplante Datenanalyse ist deutlich besser geeignet, um datengetriebene und damit wenig valide statistische Auswertungen zu verhindern oder zumindest als solche abzugrenzen.

In 1 Stellungnahme wird darauf hingewiesen, dass für die Bewertung von E-Health-Anwendungen derzeit der gesetzliche Auftrag für eine Einbindung des Instituts fehle. Außerdem seien mit dem Digitale-Versorgung-Gesetz (DVG) andere, neuere Kriterien für die Bewertung von E-Health-Anwendungen zu beachten.

Es ist zutreffend, dass mit dem DVG das Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) die Aufgabe erhalten hat zu prüfen, ob neue digitale Gesundheitsanwendungen positive Versorgungseffekte aufweisen. Da sich diese Regelung (§ 33a SGB V) jedoch nur auf digitale Medizinprodukte niedriger Risikoklasse bezieht, ist es weiterhin möglich, dass Interventionen, die auf digitalen Medizinprodukten höherer Risikoklasse beruhen, vom IQWiG bewertet werden. In diesem Fall wären auch die Bewertungsmaßstäbe andere als die dem BfArM vorgegebenen. Daher ergibt sich hier kein Änderungsbedarf am Methodenpapier.

2.12 Würdigung der Stellungnahmen zu Abschnitt 3.5 „Diagnostische Verfahren“

In 1 Stellungnahme wird kritisiert, dass der Linked-Evidence-Ansatz „nur im Zusammenhang mit der Bewertung von Screening-Verfahren beschrieben, d. h. bei anderen diagnostischen Verfahren scheinbar nicht in Betracht gezogen“ werde.

Hier liegt ein Missverständnis vor. Der Linked-Evidence-Ansatz kann sowohl bei diagnostischen Interventionen als auch bei Screeninginterventionen sinnvoll eingesetzt werden. Da in Abschnitt 3.5 explizit auf den Linked-Evidence-Ansatz hingewiesen wird, ergibt sich kein Änderungsbedarf am Text des Methodenpapiers.

2.13 Würdigung der Stellungnahmen zu Abschnitt 3.8 „Potenzialbewertung“

1 Stellungnehmer regt an, die vom G-BA verwendete Definition von „Potenzial“ auch im Methodenpapier aufzunehmen. Hieraus ergäbe sich vor allem die Notwendigkeit, sich mit dem Wirkprinzip einer Untersuchungs- und Behandlungsmethode zu beschäftigen.

Im Methodenpapier wird bereits ausgeführt, dass unter dem „Potenzial“ einer Methode ein Wissensstand zu verstehen ist, gemäß dem „die bisher vorliegenden Erkenntnisse einen möglichen Nutzen erkennen lassen“. Diese verkürzte Formulierung entspricht inhaltlich der Potenzialdefinition des G-BA und umfasst auch die Berücksichtigung des Wirkprinzips einer Behandlungsmethode.

In 1 Stellungnahme wird angeregt, möglichst bald die Änderungen umzusetzen, die sich aus dem Terminservice- und Versorgungsgesetz (TSVG) ergeben. Konkret wird darauf verwiesen, dass für die Erprobung einer neuen Behandlungsmethode gemäß § 137h SGB V nicht weiter ein Potenzial erforderlich sei. In einer weiteren Stellungnahme wird darum gebeten, die Kriterien Schädlichkeit und Unwirksamkeit, die neu in den § 137h SGB V aufgenommen wurden, näher zu operationalisieren.

Am 16.07.2020 hat der G-BA den notwendigen Beschluss zur Umsetzung der Änderungen des § 137h SGB V durch das TSVG gefasst. Dementsprechend wurden nun die notwendigen Folgeänderungen im Methodenpapier umgesetzt. Hieraus ergaben sich eine Aufteilung des Abschnitts 3.8 in 2 Unterkapitel zur Potenzialbewertung und zur §-137h-Bewertung sowie mehrere Textänderungen zur §-137h-Bewertung. Insoweit eine Operationalisierung neuer Bewertungskriterien erforderlich war, wurde diese im Methodenpapier vorgenommen.

2.14 Würdigung der Stellungnahmen zu Kapitel 5 „Bewertungen zur Versorgung“

In 1 Stellungnahme wird kritisiert, dass im überarbeiteten Kapitel 5 auf eine Darstellung und Bewertung der Versorgungsqualität verzichtet wird. Die Bewertung der Versorgungsqualität sei aus Sicht der Patientinnen und Patienten ein wichtiges Kriterium, das in den Allgemeinen Methoden berücksichtigt und wiederaufgenommen werden sollte.

Wir verstehen das Anliegen. Wir haben jedoch in den letzten Jahren festgestellt, dass im Institut zu diesem Thema die Kapazitäten fehlen und nach der Gründung des Instituts für

Qualitätssicherung und Transparenz im Gesundheitswesen (IQTIG) diese Aufgabe auch prioritär dort angesiedelt ist.

2.15 Würdigung der Stellungnahmen zu Abschnitt 5.1.4 „Bewertung der methodischen Leitlinienqualität“

2 Stellungnehmende regen an, für die Bewertung der methodischen Leitlinienqualität einen definierten Cut-off-Wert anzuwenden, um zwischen qualitativ hoch- und niedrigwertigen Leitlinien zu differenzieren.

Die Festlegung auf einen spezifischen Cut-off-Wert wird international eher kontrovers diskutiert. Das IQWiG verzichtet bewusst auf eine solche Festlegung, da ansonsten viele Leitlinien, die für den deutschen Versorgungskontext relevant sind, aus der Leitliniensynapse ausgeschlossen würden. Um der Leserin oder dem Leser dennoch aufzuzeigen, wenn Empfehlungen aus methodisch schwachen Leitlinien einen Einfluss auf die Einschätzung „inhaltlich diskrepant zur DMP-A-RL“ haben, werden diese entsprechend im Bericht markiert.

In 1 Stellungnahme wird vorgeschlagen, die 3 Domänen des Appraisal-of-Guidelines-Research-and-Evaluation(AGREE)-II-Instruments, die für die methodische Bewertung der Leitlinien herangezogen werden, unterschiedlich zu gewichten und insbesondere die Domäne 3: Genauigkeit der Leitlinienbewertung stärker zu betonen. Begründet wird dies damit, dass Domäne 3 den größten Einfluss auf die AGREE-II-Gesamtbewertung habe.

Im Konzept des AGREE-II-Instruments ist keine A-priori-Gewichtung einzelner Domänen vorgesehen. Das bisher weder operationalisierte noch validierte Item zur Gesamtbewertung wurde von dem AGREE-Entwicklerteam als übergeordnete Bewertung ergänzt, aber nicht als Ersatz der Einzeldomänen oder Zusammenfassung der Einzeldomänen betrachtet. Auch die in der Stellungnahme zitierten systematischen Übersichtsarbeiten beschreiben aufgrund der subjektiven Erfassung keine systematisch klare Begründung für die unterschiedliche Gewichtung einzelner Domänen.

2.16 Würdigung der Stellungnahmen zu Abschnitt 5.1.6 „Strukturierte Informationssynthese: Extraktion und Analyse der Empfehlungen“

1 Stellungnehmende hinterfragt die Einführung des Begriffs „Code“ im Zusammenhang mit der Analyse der Empfehlungen.

Da der Begriff „Code“ trotz Erläuterung für die Leserin bzw. den Leser offenbar missverständlich ist, haben wir den Begriff ersetzt durch „Versorgungsaspekt“. Eine Erläuterung zum Begriff Versorgungsaspekt findet sich zu Beginn des Abschnitts 5.1.6. In diesem Zusammenhang wurde auch der Begriff Kernaussage durch den des Kerninhalts ersetzt.

2.17 Würdigung der Stellungnahmen zu Abschnitt 5.3.4 „Datenquellen“

In 1 Stellungnahme wird empfohlen, Daten aus elektronischen Patientenakten als Datenquelle zu erwägen, sofern diese die im Rapid Report A19-43 beschriebenen Herausforderungen adressieren können.

Im Rapid Report A19-43 über versorgungsnahe Daten zum Zwecke der Nutzenbewertung hat sich das IQWiG in Abschnitt 5.4.3 mit den derzeitigen Möglichkeiten und Grenzen der Nutzung von Daten aus elektronischen Patientenakten ausführlich auseinandergesetzt. Die hier ausführlich beschriebenen kritischen Aspekte sind auch für die Nutzung im Rahmen der Analyse von Versorgungsdaten zu berücksichtigen. Im Kern wird eine Erhebung versorgungsnaher Daten zum Zwecke der Nutzenbewertung nach § 35a SGB V aus elektronischen Patientenakten aktuell als nicht realistisch durchführbar angesehen.

2.18 Würdigung der Stellungnahmen zu Abschnitt 6.2 „Themensammlung“

In 1 Stellungnahme wird gefragt, warum nur Einzelpersonen Themen für den ThemenCheck Medizin des IQWiG vorschlagen können.

Die Eingrenzung des Kreises derjenigen, die Themen für den ThemenCheck Medizin vorschlagen können, ergibt sich aus § 139b Abs. 5 des SGB V. Dort ist festgelegt, dass „Versicherte und sonstige interessierte Einzelpersonen“ medizinische Verfahren und Technologien für eine Bewertung vorschlagen können. Es besteht kein Ergänzungs- oder Änderungsbedarf für die Allgemeinen Methoden 6.0.

2.19 Würdigung der Stellungnahmen zu Abschnitt 6.3.3 „1. Stufe des Auswahlverfahrens: Nominierung von Themen“

In 1 Stellungnahme wird darauf hingewiesen, dass die Ämter Patientenbeauftragte der Bundesregierung und Pflegebevollmächtigte getrennt wurden. Daraus ergebe sich die Notwendigkeit, den Zusatz „und Bevollmächtigte beziehungsweise Bevollmächtigter für Pflege“ aus einer Formulierung in Abschnitt 6.3.3 zu streichen.

Dieser Hinweis ist richtig. Die fehlerhafte Formulierung in Abschnitt 6.3.3 wurde wie in der Stellungnahme vorgeschlagen korrigiert.

2.20 Würdigung der Stellungnahmen zu Abschnitt 6.3.4 „2. Stufe des Auswahlverfahrens: Auswahl von Themen, zu denen HTA-Berichte erstellt werden“

In 2 Stellungnahmen wird gefragt, welche Rolle dem erweiterten Fachbeirat bei der Auswahl von Themen für die Erstellung eines HTA-Berichts zukommt. Hinsichtlich der Rolle des erweiterten Fachbeirats bei der Auswahl von Themen gäbe es Diskrepanzen zwischen der Darstellung in Abbildung 16 und der textlichen Darstellung in Abschnitt 6.3.4.

Im Fließtext in Abschnitt 6.3.4 wurde konkretisiert, dass die Auswahl der Themen für die Erstellung von HTA-Berichten durch das IQWiG im Benehmen mit dem erweiterten Fachbeirat erfolgt.

In 1 Stellungnahme wird gefragt, ob mit der Anpassung einer Formulierung von „Institutsleitung“ in „Institut“ eine Aufgabenübertragung hinsichtlich der endgültigen Wahl der Themen durch das IQWiG stattgefunden habe.

Bei der Anpassung der Formulierung handelt es sich nicht um eine Aufgabenübertragung, sondern um eine sprachliche Vereinheitlichung. Es besteht kein Ergänzungs- oder Änderungsbedarf für die Allgemeinen Methoden 6.0.

2.21 Würdigung der Stellungnahmen zu Abschnitt 7.2.1 „Themenkatalog gemäß dem Generalauftrag“

1 Stellungnehmer wirft die Frage auf, inwieweit das Codierungsverhalten die Daten des Versorgungsreports des Wissenschaftlichen Instituts der Allgemeinen Ortskrankenkasse (WIdO) beeinflusst respektive ob dieses immer korrekt ist.

Wie in den Methoden ausgeführt, ist ein grundlegender Aspekt der epidemiologischen Bedeutung einer Erkrankung deren Häufigkeit. Der alle 2 Jahre aktualisierte Versorgungsreport des WIdO ist dafür eine valide Quelle. Er beinhaltet auf Basis der Daten von etwa 24 Millionen Allgemeine-Ortskrankenkasse(AOK)-Versicherten Angaben zu Prävalenzen und Hospitalisierungsraten für die 1500 häufigsten Erkrankungsgruppen (3-stellige International-Statistical-Classification-of-Diseases-and-Related-Health-Problems[ICD]-10-Gruppen). Wir sind uns bewusst, dass die Codierungsqualität Einfluss auf die Daten des Reports und damit auch auf unsere Themenauswahl haben könnte. Es zeigt sich jedoch über die Jahre ein relativ stabiles Bild bei den etwa 200 häufigsten Themen, die Basis unserer Auswahl sind. Zudem wird als mögliches Korrektiv der Themenkatalog des Instituts regelmäßig überprüft. Wenn notwendig, kann er auch angepasst oder erweitert werden, beispielsweise in Abhängigkeit von Auftrags-themen des IQWiG oder um Erkrankungen, deren Bedeutung und Krankheitslast über die Prävalenz / Inzidenz nur unzureichend erfasst werden.

2.22 Würdigung der Stellungnahmen zu Abschnitt 7.9.3 „Nutzertesting“

1 Stellungnehmer bemängelt das Fehlen von Informationen zu den Auswahlkriterien der Nutzer sowie die Zusammensetzung, Repräsentativität und Anzahl der Teilnehmer an einem Test. Zudem fehlten die Beschreibung des Auswahlprozesses für die Fokusgruppen bzw. Einzelinterviews sowie die Gründe für die Auswahl.

Bei der Nutzertesting kommt ein Methodenmix aus dokumentierter Einzelbewertung und Gruppendiskussion bzw. Einzelinterviews zur Anwendung. Ob eine Gruppendiskussion oder Einzelinterviews durchgeführt werden, wird themenspezifisch entschieden, beispielsweise können sich bei tabuisierten Themen oder bestimmten Zielgruppen Einzelinterviews eher eignen. Anhand eines semistrukturierten Leitfadens testen in der Regel 5 Betroffene oder

potenzielle Nutzerinnen und Nutzer die Texte im Hinblick auf ihren Informationsgehalt und ihre Verständlichkeit. Die Testerinnen und Tester werden anhand der vom IQWiG vorgegebenen Kriterien vom externen Auftragnehmer oder von der externen Auftragnehmerin rekrutiert. Die Auswahlkriterien beziehen sich auf die in den zu testenden Informationsmaterialien gesetzten Schwerpunkte und umfassen in der Regel Angaben zu Diagnose, Geschlecht, Alter, Erfahrungen mit bestimmten Behandlungsoptionen und ggf. zum sozialen Status. Wir haben den Abschnitt entsprechend präzisiert.

Derselbe Stellungnehmer hält ein zeitgleiches Durchführen des Stellungnahmeverfahrens und der Nutzertestung für weniger sinnvoll. Die Überprüfung der fachlichen Richtigkeit der Inhalte solle in einem 1. Schritt erfolgen, die Verständlichkeit der Arbeit sowie die Sprache seien in einem 2. Schritt zu prüfen. Unklar sei auch, ob die Nutzer als Co-Autoren miteinbezogen werden oder ob sie die Gesundheitsinformationen lediglich auf ihre Verständlichkeit bzw. ihren Informationsgehalt prüfen. Informationen zur Überprüfung der Gesundheitsinformationen durch die Nutzer auf ihre Akzeptanz, Relevanz sowie Benutzerfreundlichkeit fehlten.

Aus unserer Sicht hat sich der bisherige zeitliche Ablauf bewährt und es besteht kein Änderungsbedarf. Zum einen erfolgt regelhaft bereits vor der zeitgleichen nicht öffentlichen Stellungnahme (Prüfung der fachlichen Inhalte) und Nutzertestung (Prüfung der Verständlichkeit) eine externe medizinisch-fachliche Begutachtung der Textentwürfe durch eine auf dem jeweiligen Gebiet praktisch tätige Fachärztin oder einen Facharzt. Die nicht öffentliche Stellungnahme ist eine Möglichkeit für unter anderem die Fachöffentlichkeit, die Texte noch einmal fachlich zu prüfen und zu kommentieren. Zum anderen besteht für den Fall, dass es im Rahmen der Stellungnahme zu einer so erheblichen Änderung des Textes kommen sollte, dass die Ergebnisse der Nutzertestung der Vorversion nicht übertragbar scheinen, die Möglichkeit, dass eine erneute Nutzertestung durchgeführt werden kann.

2.23 Würdigung der Stellungnahmen zu Abschnitt 7.15.2 „Erfahrungsberichte“

1 Stellungnehmer fragt, ob es Analysen gebe, wie sich Erfahrungsberichte zur Einschätzung der Behandler verhalten.

Uns sind keine entsprechenden Analysen bekannt. Unsere Erfahrungsberichte haben aber letztlich das Ziel, das subjektive Erleben der Betroffenen darzustellen, unabhängig von der Einschätzung oder Sichtweise der oder des Behandelnden.

2.24 Würdigung allgemeiner Stellungnahmen zu Abschnitt 8.1.1 „Recherche in bibliografischen Datenbanken“

In 1 Stellungnahme wird vorgeschlagen, bei der Beschreibung der Ziele der Vorabrecherche zu ergänzen, dass geprüft werden soll, ob es sich um einen dramatischen Effekt handeln könnte.

Die Vorabrecherche hat mehrere Ziele, die sich zudem auch projektspezifisch unterscheiden können. Die Ziele sollten daher etwas allgemeiner formuliert sein. Um dies klarzustellen, wurde

der Text etwas gekürzt und spezifische Beispiele (z. B. Testset zur Entwicklung von Suchstrategien) gelöscht.

2.25 Würdigung der Stellungnahmen zu Abschnitt 8.2.5 „Überprüfung der Vollständigkeit eines Studienpools“

In 1 Stellungnahme wird darauf hingewiesen, dass bei der Vollständigkeitsprüfung noch das Register PharmNet.Bund als Quelle gelistet wurde und eigentlich entfallen könnte.

Das Register wurde gelöscht.

2.26 Würdigung allgemeiner Stellungnahmen zu Abschnitt 8.4 „Recherche nach Leitlinien zur Erstellung von Leitliniensynopsen“

In 1 Stellungnahme wird vorgeschlagen, eine Festlegung zur Aktualität der Leitlinien bei Einschluss hinzuzufügen.

Solche Einschränkungen sind projektspezifisch zu treffen. Es wird daher kein Ergänzungsbedarf im Methodenpapier gesehen.

2.27 Würdigung der Stellungnahmen zu Abschnitt 9.1.2 „Zusammenhang zwischen Studientyp / -art und Fragestellung“

In 1 Stellungnahme wird darauf hingewiesen, dass ein Verweis auf den Abschnitt 3.2.1 falsch ist, und es wird vorgeschlagen, ein weiteres Beispiel für Krankheitsbilder mit einem deterministischen (oder nahezu deterministischen) Verlauf zu nennen.

Dieser Hinweis ist korrekt; der Querverweis im Methodenpapier wurde entsprechend korrigiert. Der Vorschlag wurde ebenfalls aufgegriffen und ein weiteres Krankheitsbild mit einem deterministischen (oder nahezu deterministischen) Verlauf wurde ergänzt (Kammerflimmern).

2.28 Würdigung der Stellungnahmen zu Abschnitt 9.3.3 „Beurteilung klinischer Relevanz“

Mit dem Entwurf zur Version 6.0 des Methodenpapiers wurde ein Vorschlag für ein geändertes Verfahren zur Auswahl von Responsekriterien für Responderanalysen bei komplexen Skalen vorgelegt. Eine Reihe von Stellungnahmen merken an, dass die wissenschaftliche Herleitung des Vorschlags nicht hinreichend klar sei.

Die Frage nach einer detaillierteren Beschreibung der Herleitung der generischen Responseschwelle von 15 % der Skalenspannweite ist nachvollziehbar. Das Vorgehen wird deshalb hier noch einmal detailliert dargestellt.

Im 1. Schritt wurde eine fokussierte Recherche nach systematischen Übersichtsarbeiten zu Minimal important Differences (MIDs) in MEDLINE durchgeführt (Suchdatum 22.05.2019). Die Übersichtsarbeiten wurden unter Berücksichtigung folgender Einschlusskriterien selektiert:

- Übersichtsarbeiten ab 2016
- Die Übersichtsarbeiten sollten Arbeiten einschließen, in denen MIDs empirisch ermittelt wurden.
- Die Übersichtsarbeiten sollten Arbeiten einschließen, in denen die MIDs für Instrumente ermittelt wurden, die patientenberichtet sind, mehr als 1 Item umfassten und deren Ausprägungen nicht direkt interpretierbar sind (abstrakte Skalen in Abgrenzung zu bspw. natürlichen Skalen, deren Ergebnis direkt interpretierbar ist).

Im 2. Schritt wurden aus den eingeschlossenen Übersichtsarbeiten die Volltexte der eingeschlossenen Referenzen geprüft. Die Volltexte wurden hinsichtlich der Methodik zur Ermittlung der MID gesichtet. Die Methodik zur Ermittlung der MID musste folgende Kriterien erfüllen:

- longitudinale Studie
- ankerbasierte MID
- patientenberichteter Anker (oder bei Kindern / Demenzkranken durch die Eltern bzw. Proxy)
- Global-Rating-of-Change(GRC)-Anker
- Der Cut-off für den GRC-Anker sollte bei minimal, small, little oder höchstens moderate liegen, um die Ermittlung einer MID zu gewährleisten.
- Es musste sich um abstrakte (Patient-reported-Outcome[PRO])-Multi-Item-Skalen handeln.

Durch die Recherche wurden 16 Übersichtsarbeiten identifiziert [1,5-8,11-13,17,18,27,29,32,33,35,37]. Aus 8 dieser Arbeiten [1,12,13,18,27,32,35,37] wurden die empirisch ermittelten MIDs, die die oben genannten Kriterien erfüllten, extrahiert. Auf die Extraktion der übrigen 8 Übersichtsarbeiten [5-8,11,17,29,33] wurde verzichtet, da diese ausschließlich orthopädische Indikationen abbilden und die Studien zum großen Teil mit denen der bereits eingeschlossenen Übersichtsarbeiten überlappen. Somit waren keine über die bereits extrahierten Übersichtsarbeiten hinausgehenden Informationen zu erwarten.

In einem nächsten Schritt wurden die extrahierten MIDs im Verhältnis zur Spannweite der jeweiligen Skala dargestellt (MID in % der Skalenspannweite). Die folgende Tabelle zeigt die Spannweite dieser Prozentsätze der empirisch erhobenen MIDs.

Tabelle 1: Spannweite der empirisch erhobenen MIDs im Verhältnis zur Spannweite der jeweils verwendeten Skala in %

Übersichtsarbeit	Therapiegebiet	Spannweite der erhobenen MID im Verhältnis zur Spannweite der Skala in %
Doganay 2016 [12]	Rheuma	< 1 %–38 %
St-Pierre 2016 [37]	Orthopädie	6 %–23 %
Ebrahim 2017 [13]	Pädiatrie	2 %–34 %
Nordin 2016 [32]	Fatigue	5 %–24 %
Jayadevappa 2017 [27]	Onkologie	3 %–18 %
	kardiovaskuläre Erkrankungen	9 %–16 %
	respiratorische Erkrankungen	3 %–17 %
	Erkrankungen des Urogenitaltrakts	6 %–18 %
	muskuloskelettale Erkrankungen	1 %–23 %
	Pädiatrie	< 1 %–38 %
	verschiedenes	< 1 %–23 %
Alma 2018 [1]	COPD	2 %–17 %
Ousmen 2018 [35]	Onkologie	< 1 %–27 %
Hao 2019 [18]	Orthopädie	4 %–22 %
COPD: chronisch obstruktive Lungenerkrankung; MID: Minimal important Differenz		

Die systematische Betrachtung von empirisch erhobenen MIDs zeigt noch einmal die Variabilität der Ergebnisse von Studien zur Ermittlung von MIDs auf. Dabei ist hervorzuheben, dass in diese Aufstellung bereits nur solche Studien eingingen, die gewisse Mindestqualitätskriterien erfüllten (z. B. patientenberichteter Anker, Längsschnittstudie, siehe oben). Diese Kriterien stellen einen wesentlichen Bestandteil des kürzlich von der Gruppe der McMaster-Universität vorgeschlagenen Instruments zur Beurteilung der Qualität von MIDs dar [10].

Die Werte am unteren Rand der erhobenen MIDs erscheinen unrealistisch klein und bezüglich der Abgrenzung von spürbarer Veränderung und Messunsicherheit wenig geeignet. Die Frage, ob die Werte am oberen Rand des Spektrums verlässlicher sind, lässt sich aufgrund fehlender akzeptierter Standards zur Qualitätsbewertung, wie auch fehlender Berichtsqualität von MID-Studien, nicht beantworten (siehe unten). Jedoch zeigt die systematische Betrachtung, dass der wesentliche Anteil empirisch ermittelter MIDs unterhalb von 20 % der jeweiligen Skalen-spannweite liegt. Auf Basis der Ergebnisse dieser systematischen Erhebung von MIDs wurde somit ein Schwellenwert von 15 % als plausibel geeignet angesehen, hinreichend sicher eine für Patientinnen und Patienten spürbare Veränderung abzubilden. Diese Responseschwelle soll eine eher kleine Veränderung darstellen, der Wert kann aber durchaus oberhalb einer minimalen Schwelle liegen. Das heißt, bei der Festlegung dieser Schwelle geht es explizit nicht darum, eine MID im Sinne der minimalen spürbaren Veränderung zu identifizieren.

Verschiedene Stellungnahmen wiesen darauf hin, dass das vorgeschlagene Responsekriterium von 15 % des Skalenrangs zu unspezifisch und zu hoch sei, um eine MID darzustellen. Darüber hinaus spiegle die Verwendung einer generischen Schwelle nicht den Stand der wissenschaftlichen Diskussion zur Verwendung von MIDs wider.

Hier liegt ein Missverständnis vor. Die generische Schwelle von 15 % der Skalenspannweite als Kriterium für eine Response soll eben keine MID beschreiben. Vielmehr soll dieser Wert auf Basis der Analyse systematischer Übersichten zu MIDs eine Schwelle darstellen, die hinreichend sicher eine für Patientinnen und Patienten spürbare Veränderung abbildet (siehe oben). Dabei wird davon ausgegangen, dass der Wert durchaus oberhalb der MID liegen kann. Diese Eigenschaft berücksichtigt auch die bestehende Variabilität von MIDs und trägt dazu bei, dass das Responsekriterium in unterschiedlichen Konstellationen nicht zu häufig unter die MID fällt. Gleichzeitig ist der Wert klein genug, um eine eher geringe Änderung abzubilden.

Einige Stellungnahmen halten es für kritisch, dass unter dem neuen Kriterium einer Response-schwelle von 15 % des Skalenrangs eine Reihe von bisher akzeptierten MIDs nicht mehr berücksichtigt würden.

Es ist richtig, dass die Definition der Responseschwelle von 15 % der Skalenspannweite dazu führt, dass einige der bisher verwendeten Responsekriterien nicht mehr berücksichtigt werden. Das gilt z. B. für die von den Stellungnehmenden angeführten Responsekriterien für den EORTC QLQ-C30 von 10 Punkten oder den St. George's Respiratory Questionnaire (SGRQ) von 4 Punkten. Zur Beurteilung dieses Sachverhalts ist es notwendig, die Qualität der infrage stehenden Responsekriterien zu betrachten. In den von den Stellungnehmenden genannten Beispielen gehen die Responsekriterien in der Regel auf Studien zurück, die dem aktuellen Stand der wissenschaftlichen Erkenntnis zu Studien zur Bestimmung von MIDs nicht entsprechen. So wird das Responsekriterium für den EORTC QLQ-C30 primär mit der Arbeit von Osoba et al. [34] begründet. In dem dort herangezogenen Datensatz erfüllte z. B. die Korrelation von Anker und Skala nicht die jetzt diskutierten Anforderungen [10]. In dieser Situation hat die European Organisation for Research and Treatment of Cancer (EORTC), die für das Instrument verantwortlich zeichnet, ein Projekt zur Bestimmung von MIDs für den EORTC QLQ-C30 und weitere EORTC-Fragebogen begonnen [31]. Für die MID von 4 Punkten für den SGRQ wird primär eine narrative Übersicht aus dem Jahr 2005 zitiert [28], andere Arbeiten aus den Jahren 2015 und 2018 beschreiben MIDs zwischen 6 und 9 Punkten [1,39]. Insgesamt ist die Methodik der Erhebung und Auswertung patientenberichteter Endpunkte auf Basis einer intensiven wissenschaftlichen Diskussion im Wandel, das führt notwendigerweise auch zu Anpassungen bisher etablierter Vorgehensweisen. Der für das Methodenpapier 6.0 vorgelegte Vorschlag trägt der Problematik der Änderung methodischer Vorgehensweisen Rechnung, indem noch einmal hervorgehoben wird, dass auch Auswertungen stetiger Daten herangezogen werden können, die in der Regel zur Verfügung stehen, und indem ein Vorschlag für eine Post-hoc-Auswertung unterbreitet wird, der eine Anpassung an die aktuellen methodischen Entwicklungen ermöglicht.

Einige Stellungnahmen beschreiben, dass der Vorschlag nicht auf alle Skalen anwendbar sei. Als Beispiel wird die 6-Minuten-Gehstrecke genannt.

Hier liegt ein Missverständnis vor. Es ist richtig, dass das vorgeschlagene Responsekriterium von 15 % des Skalenrangs z. B. für nach oben offene Skalen nicht herangezogen werden kann. Eine solche Anwendung ist aber auch nicht geplant. Die Festlegung adressiert komplexe Skalen (insbesondere psychometrische Skalen), deren Ergebnisse nicht ohne Weiteres interpretiert werden können. Das vorgeschlagene Responsekriterium von 15 % des Skalenrangs soll nicht bei Skalen eingesetzt werden, deren Ergebnisse in natürlichen Einheiten ausgedrückt werden und die deshalb einer inhaltlichen Bewertung der Relevanz eines Effekts zugänglich sind. Da komplexe psychometrische Skalen stets über einen Maximalwert verfügen, kann das von den Stellungnehmenden angeführte Problem nicht auftreten.

Einige Stellungnehmende befürchten, dass durch das vorgeschlagene Vorgehen die Anforderungen erhöht werden und der Nachweis eines Zusatznutzens von Interventionen erschwert würde.

Die Stellungnehmenden führen nicht aus, warum sie davon ausgehen, dass das vorgeschlagene Responsekriterium den Nachweis eines Zusatznutzens erschweren sollte, und legen keine Daten vor, die diese Annahme stützen würden. Generell ist nicht davon auszugehen, dass (robuste) Effekte nicht mehr nachweisbar sind, wenn sich alternativ eingesetzte Responsekriterien im unteren Bereich der Spannweite einer Skala bewegen. Diese Einschätzung wurde durch die Diskussion in der wissenschaftlichen Erörterung bestätigt.

Verschiedene Stellungnehmende schlagen eine Diskussion zu Kriterien für die Auswahl von MIDs vor und befürworten bis zum Ergebnis einer solchen Diskussion eine unverändert fortgeführte Bewertung von MIDs zum Einschluss in die Nutzenbewertung.

Generell werden in den Stellungnahmen wenig konkrete Vorschläge gemacht, wie mit der Variabilität von MIDs und den bestehenden Schwierigkeiten bei der Bewertung aktuell vorliegender Studien zu MIDs umgegangen werden kann. Es bleibt deshalb offen, welche Ansatzpunkte und Ziele die vorgeschlagene Diskussion haben kann.

Nach Abschluss der Stellungnahmefrist wurde ein Vorschlag für ein Instrument zur Beurteilung der Aussagekraft ankerbasierter Festlegungen von MIDs veröffentlicht [10]. In dieser Arbeit beschreibt eine Gruppe um Wissenschaftlerinnen und Wissenschaftler der McMaster-Universität die Entwicklung der Qualitätskriterien für Studien zur Ermittlung von MIDs und die Ergebnisse einer Studie zur Reliabilität des neu entwickelten Instruments. Grundsätzlich fokussieren sie dabei allein auf ankerbasierte Verfahren; verteilungsbasierte Verfahren werden per se als ineffektiv und als die Patientenperspektive vernachlässigend eingeschätzt. Die Qualitätskriterien, die die Autorinnen und Autoren der Arbeit für die Beurteilung der Aussagekraft einer MID benennen, sind vor dem Hintergrund der Diskussion der letzten Jahre nachvollziehbar. Das Instrument besteht aus 5 Kernkriterien (Beantwortung des Fragebogens, für den die MID

bestimmt werden soll, und des Ankers durch Patientinnen und Patienten (bzw. wenn notwendig eine Vertretung); leichte Verständlichkeit und Relevanz des Ankers für Patientinnen und Patienten; eine gute Korrelation zwischen Fragebogen und Anker; Präzision der MID; der Schwellenwert des Ankers für die Bestimmung der MID zeigt einen kleinen, aber relevanten Unterschied an) und 4 zusätzlichen Kriterien für MIDs, die mithilfe von Abfragen zu Änderungen des Ankers (Transition Ratings) erfolgen (optimale Zeitspanne zwischen den Messungen; zufriedenstellende Korrelation zwischen Fragebogen und Anker [Transition Item] bei der Follow-up-Messung; Korrelation des Ankers [Transition Item] mit dem Ausgangswert des Fragebogens; höhere Korrelation zwischen Änderung des Ankers [Transition Item] und Änderung des Fragebogens als zwischen Änderung des Ankers und Wert des Fragebogens zum Follow-up). Diese ergänzenden Kriterien sind relevant, weil ein großer Teil von MIDs über die Abfrage der Änderungen von Fragebogen und Anker erhoben wird.

Interessant für die Diskussion um den Methodenvorschlag des IQWiG sind die Ergebnisse der Reliabilitätsstudie. Für die Untersuchung der Reliabilität des neu entwickelten Instruments nutzte die Gruppe eine Zufallsstichprobe aus einer Datenbank empirisch erhobener MIDs. In dieser Stichprobe konnten die Autorinnen und Autoren die Reliabilität der Kriterien 2 bis 4 für MID-Ermittlungen über Abfragen zur Änderung des Ankers nicht ermitteln, weil die dafür notwendigen Angaben nur in einem sehr geringen Teil der Studien zur Ermittlung der MIDs berichtet wurden. Diese Daten decken sich mit Erfahrungen aus der Dossierbewertung, in der ebenfalls häufig nicht alle notwendigen Angaben zur Prüfung der Aussagekraft einer MID vorliegen. Der sofortige Einsatz dieses neuen Instruments würde deshalb voraussichtlich wegen fehlender Angaben zu den Qualitätskriterien zu einem hohen Anteil negativer Bewertungen der Aussagekraft der MIDs führen. Der Vorschlag der generischen Responseschwelle mit Bezug zum Skalenrange erscheint auch unter diesem Gesichtspunkt ein sinnvoller Schritt, zumindest für den Zeitraum, bis sich die Qualität der Studien zur Ermittlung von MIDs verbessert hat.

2.29 Würdigung der Stellungnahmen zu Abschnitt 9.3.4 „Bewertung subjektiver Endpunkte bei offenen Studiendesigns“

In 2 Stellungnahmen wird die Anwendung einer adjustierten Entscheidungsgrenze bei subjektiven Endpunkten und offenem Studiendesign kritisiert, da das zu einer „doppelten Bestrafung“ führt. Zum einen wird aufgrund der fehlenden Verblindung die Ergebnissicherheit herabgesetzt, zum anderen wird beim Signifikanztest die Nullhypothese verschoben (Anwendung einer adjustierten Entscheidungsgrenze). Es wird vorgeschlagen, dass eine fehlende Verblindung nur bei der Bewertung des Verzerrungspotenzials berücksichtigt wird.

Die Argumentation mit der „doppelten Bestrafung“ ist nicht korrekt, da bei Anwendung einer adjustierten Entscheidungsgrenze die Ergebnissicherheit aufgrund der vorliegenden großen Effektstärke wieder heraufgesetzt würde. Aufgrund aktueller Literatur aus dem Bereich der metaepidemiologischen Forschung zum Effekt fehlender Verblindung [30] lässt sich allerdings keine sinnvolle adjustierte Entscheidungsgrenze bei subjektiven Endpunkten und offenem Studiendesign mehr ableiten. Daher wurde der Abschnitt 9.3.4 aus den Allgemeinen Methoden entfernt.

2.30 Würdigung der Stellungnahmen zu Abschnitt 9.3.8 „Metaanalysen“

Die im Entwurf vorgenommenen Präzisierungen und Klarstellungen zur Durchführung von Metaanalysen werden in den Stellungnahmen begrüßt. In 1 Stellungnahme wird eine weitere Präzisierung des Begriffs „Varianzkorrektur“ vorgeschlagen.

Der Vorschlag wurde aufgegriffen und im betreffenden Abschnitt wurden entsprechende Umformulierungen vorgenommen.

In 1 Stellungnahme wird angeregt, als Kriterium zur Entscheidung für die Nichtzusammenfassbarkeit von Studien in einer Metaanalyse neben dem Heterogenitätstest und I^2 auch die Varianz zwischen den Studien τ^2 zu verwenden. Es wird ein Beispiel genannt, wie man bei standardisierten Mittelwertdifferenzen eine Schwelle für τ^2 definieren könnte, um eine klinisch bedeutsame Heterogenität abzubilden.

Um die Varianz zwischen den Studien τ^2 als Kriterium zur Entscheidung für die Nichtzusammenfassbarkeit von Studien in einer Metaanalyse zu verwenden, müssten sinnvolle Schwellenwerte nicht nur für standardisierte Mittelwertdifferenzen, sondern auch für alle anderen Effektmaße definiert werden. Hierbei müssten außerdem die jeweiligen Endpunkteigenschaften, wie z. B. die Variabilität in der Grundgesamtheit, berücksichtigt werden. Das erscheint in der Praxis nicht durchführbar. Zudem ist die Varianz zwischen den Studien bei Metaanalysen mit sehr wenigen Studien nicht verlässlich schätzbar. Es ergibt sich kein Änderungsbedarf für die Allgemeinen Methoden.

In 1 Stellungnahme wird eine Aktualisierung des Unterabschnitts bezüglich Metaanalysen von Studien zur diagnostischen Güte mit Referenz auf neuere Literatur vorgeschlagen. Es wurde zudem kritisiert, dass auch die Anwendung getrennter Modelle für Sensitivität und Spezifität als Möglichkeit genannt wird.

Der Unterabschnitt bezüglich Metaanalysen von Studien zur diagnostischen Güte wurde überarbeitet. Es wurden einige neuere Literaturzitate ergänzt und es wurde erläutert, dass in der Praxis aufgrund unzureichender Daten häufig getrennte metaanalytische Modelle für Sensitivität und Spezifität angewendet werden müssen, obwohl eine bivariate Metaanalyse vorzuziehen ist.

2.31 Würdigung der Stellungnahmen zu Abschnitt 9.3.10 „Subgruppenanalysen“

In 1 Stellungnahme wird angemerkt, dass die Regel zur Zusammenfassung von Paaren auf Basis signifikanter Interaktionstests beim Vorliegen von mehr als 2 Subgruppen nicht ganz eindeutig ist.

Es ist richtig, dass die angeführte Regel auf Basis signifikanter Interaktionstests nicht alle möglichen Fälle abdeckt. Dies geschah jedoch bewusst, da in Situationen, die sich rein auf Basis der Informationen zur Signifikanz von Interaktionstests nicht eindeutig entscheiden lassen, im Einzelfall entschieden werden muss. Dies wurde in den Allgemeinen Methoden ergänzt.

2.32 Würdigung der Stellungnahmen zu Abschnitt 9.3.11 „Umgang mit nicht oder nicht vollständig publizierten Daten“

In 1 Stellungnahme wird folgende Textpassage kommentiert: „Beim weitgehenden oder vollständigen Fehlen relevanter Informationen kann es vorkommen, dass eine Publikation nicht bewertet bzw. eine Studie nicht für die Analyse herangezogen werden kann. In diesen Fällen wird lediglich dargestellt, dass es zu einem bestimmten Thema weitere Daten gibt, die aber für eine Qualitätsbewertung bzw. für die Analyse nicht zur Verfügung stehen.“

Es wird kritisiert, dass obige Formulierung nahelegt, dass eine Studie, in der weitgehend oder vollständig relevante Informationen fehlen, völlig unberücksichtigt bleibt und es dadurch zu Publikationsbias kommt.

Es liegt hier ein Missverständnis vor. Die Textpassage wurde umformuliert und es wurde klargestellt, dass auch Studien, in der weitgehend oder vollständig relevante Informationen fehlen, in der studienübergreifenden Bewertung berücksichtigt werden.

2.33 Würdigung der Stellungnahmen zu Abschnitt 9.3.12 „Umgang mit unvollständigen Daten“

In mehreren Stellungnahmen werden die in diesem Abschnitt festgelegten Schwellenwerte von 70 % für in die Analysen einzuschließende Personen und 15 Prozentpunkte für den Unterschied der nicht in den Analysen eingegangenen Personen zwischen den Gruppen thematisiert. Zum einen wird eine Rationale sowie Literaturzitate für diese Schwellenwerte gefordert. Zum anderen wird moniert, dass die festgelegten 70 % in vielen klinischen Studien nicht sachgerecht seien, da insbesondere bei langen Studienlaufzeiten Follow-up-Verluste von 30 % im langfristigen Studienverlauf unumgänglich sind. Somit würden viele Studien nicht in die Bewertungen einfließen.

Die Schwellenwerte wurden pragmatisch festgelegt, wobei sich der Schwellenwert von 70 % an den 20 % orientiert, die in der Arbeit von Schulz und Grimes [36] genannt sind. Diese Arbeit wird auch an entsprechender Stelle im Abschnitt zitiert. Ein Zitat für den Schwellenwert 15 Prozentpunkte gibt es nicht. Die Sinnhaftigkeit und eine mögliche Rationale für die Schwellenwerte wurden in der mündlichen Erörterung ausführlich diskutiert. Keiner der anwesenden Stellungnehmenden konnte auf Nachfrage weitere Literaturstellen benennen oder einen konkreten Vorschlag zur Herleitung solcher Schwellenwerte unterbreiten. Die genannten Schwellenwerte werden in den Bewertungen des Instituts seit vielen Jahren herangezogen und haben sich als praktikabel erwiesen. In der mündlichen Erörterung wurde vereinbart, dass die Schwellenwerte nicht starr angewendet werden, sondern je nach Kontext als Orientierung dienen sollten. Dies wurde im Abschnitt entsprechend ergänzt.

Die Kritik, dass Follow-up-Verluste von 30 % in klinischen Studien häufig vorkommen und durch Anwendung des Schwellenwerts viele Studien nicht in die Bewertung einfließen, beruht auf einem Missverständnis, das durch ungenaue und nicht sauber strukturierte Formulierungen im Textentwurf hervorgerufen wurde. Die Vorgabe von 70 % für in die Analysen einzuschließende Personen bezieht sich nur auf die Personen, die in den Analysen vollständig

unberücksichtigt bleiben. Follow-up-Verluste hingegen sind Personen, für die im Laufe der Studie vorzeitig die Beobachtung beendet wird. Diese Personen gehen mit unvollständigen Daten in die Analyse ein, fehlen also nicht komplett und sind von der Vorgabe gar nicht betroffen. In der mündlichen Erörterung wurde dieser Sachverhalt noch einmal erläutert und das Missverständnis aufgelöst. Zur besseren Verständlichkeit wurde der Abschnitt überarbeitet.

In 1 Stellungnahme wird angemerkt, dass es auch zu Baseline und nicht nur durch Nicht-erscheinen einer Patientin oder eines Patienten zu fehlenden Werten kommen kann und dies in dem Abschnitt nicht adressiert wird.

Die Anmerkung ist richtig und ein entsprechender Absatz wurde ergänzt.

2.34 Würdigung der Stellungnahmen zu Abschnitt 9.3.13 „Umgang mit unterschiedlichen mittleren Beobachtungsdauern“

In mehreren Stellungnahmen wird vorgeschlagen, nicht nur auf das Problem von unterschiedlichen mittleren Beobachtungsdauern einzugehen, sondern allgemein auf den Umgang mit variablen Beobachtungszeiten, insbesondere auf das Problem konkurrierender Ereignisse.

Dem Vorschlag folgend wurde der Abschnitt umbenannt in „Umgang mit variablen Beobachtungszeiten“ und es wurde ein Absatz ergänzt, der sowohl auf die allgemeinen Methoden der Überlebenszeitanalyse eingeht als auch auf die Methoden zur Berücksichtigung konkurrierender Ereignisse.

In mehreren Stellungnahmen wird angeführt, dass die formulierte Forderung der vollständigen Erfassung UEs auch nach Behandlungsabbruch oder -wechsel dem üblichen Vorgehen in der klinischen Forschung widerspricht und zu einer systematischen Verzerrung führt. Eine Erfassung von UEs ausschließlich im Zeitraum der randomisierten Behandlung (zuzüglich 30 Tage) sei hinreichend und ließe valide Schlussfolgerungen auch im Rahmen der Nutzenbewertung zu.

Es ist richtig, dass in klinischen Studien UEs häufig bis maximal 30 Tage nach Absetzen der Therapie erfasst werden und dass in der Zulassung zur Auswertung UEs regelhaft lediglich vereinfachte Analysen zur Aufdeckung von Sicherheitssignalen angewendet werden [38]. Die Berücksichtigung von UEs nur während der Behandlungsphase (zuzüglich 30 Tage) widerspricht allerdings dem Intention-to-treat-Prinzip, führt zu systematischen Verzerrungen und ist daher auch im Bereich der Zulassung nicht für eine Benefit-Risk-Evaluation geeignet [40]. In der Nutzenbewertung ist – wie für alle anderen patientenrelevanten Endpunkte – auch für UEs der Treatment-Policy-Estimand ausschlaggebend, sodass hier die Berücksichtigung von UEs nur während der Behandlungsphase (zuzüglich 30 Tage) unzureichend ist [4,38]. In den Allgemeinen Methoden wurde ergänzt, dass für adäquate Auswertungen auf Basis des Treatment-Policy-Estimands und des Intention-to-treat-Prinzips eine vollständige Erfassung von UEs auch nach Behandlungsabbruch oder -wechsel erforderlich ist.

2.35 Würdigung der Stellungnahmen zu Abschnitt 9.3.15 „Auswertung abhängiger Daten“

In 1 Stellungnahme wird darauf hingewiesen, dass es im Entwurf der Allgemeinen Methoden 6.0 keine Ausführungen zum Umgang mit Cross-over-Studien gibt.

Es ist richtig, dass Cross-over-Studien im Methodenpapier bisher nur in Abschnitt 9.1.2 unter den gängigen Studiendesigns erwähnt wurden. Daher wurde in Abschnitt 9.3.15 ergänzt, dass bei Cross-over-Studien die Datenabhängigkeiten berücksichtigt werden müssen mit Verweis auf gängige Literatur.

2.36 Literatur

1. Alma H, De Jong C, Tsiligianni I, Sanderman R, Kocks J, Van der Molen T. Clinically relevant differences in COPD health status: systematic review and triangulation. *Eur Respir J* 2018; 52(3): 1800412.
2. Basch E, Dueck AC, Rogak LJ, Minasian LM, Kelly WK, O'Mara AM et al. Feasibility assessment of patient reporting of symptomatic adverse events in multicenter cancer clinical trials. *JAMA Oncology* 2017; 3(8): 1043-1050.
3. Basch E, Reeve BB, Mitchell SA, Clauser SB, Minasian LM, Dueck AC et al. Development of the National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). *J Natl Cancer Inst* 2014; 106(9): dju244
4. Bender R, Beckmann L, Lange S. Biometrical issues in the analysis of adverse events within the benefit assessment of drugs. *Pharm Stat* 2016; 15(4): 292-296.
5. Celik D, Coban O, Kilicoglu O. Minimal clinically important difference of commonly used hip-, knee-, foot-, and ankle-specific questionnaires: a systematic review. *J Clin Epidemiol* 2019; 113: 44-57.
6. Chung AS, Copay AG, Olmscheid N, Campbell D, Walker JB, Chutkan N. Minimum clinically important difference: current trends in the spine literature. *Spine (Phila Pa 1976)* 2017; 42(14): 1096-1105.
7. Copay AG, Chung AS, Eyberg B, Olmscheid N, Chutkan N, Spangehl MJ. Minimum clinically important difference; current trends in the orthopaedic literature; part I: upper extremity; a systematic review. *JBJS Reviews* 2018; 6(9): e1.
8. Copay AG, Eyberg B, Chung AS, Zurcher KS, Chutkan N, Spangehl MJ. Minimum clinically important difference; current trends in the orthopaedic literature; part II: Lower extremity; a systematic review. *JBJS Reviews* 2018; 6(9): e2.
9. Deutscher Bundestag. Gesetz zur Errichtung des Implantateregisters Deutschland und zu weiteren Änderungen des Fünften Buches Sozialgesetzbuch (Implantateregister-Errichtungsgesetz - EIRD): Gesetzesbeschluss des Deutschen Bundestages; Drucksache Bundesrat 506/19 [online]. 18.10.2019 [Zugriff: 07.09.2020]. URL: <http://dipbt.bundestag.de/dip21/brd/2019/0506-19.pdf>.
10. Devji T, Carrasco-Labra A, Qasim A, Phillips M, Johnston BC, Devasenapathy N et al. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. *BMJ* 2020; 369: m1714.
11. Devji T, Guyatt GH, Lytvyn L, Brignardello-Petersen R, Foroutan F, Sadeghirad B et al. Application of minimal important differences in degenerative knee disease outcomes: a systematic review and case study to inform BMJ Rapid Recommendations. *BMJ Open* 2017; 7(5): e015587.

12. Doganay Erdogan B, Leung YY, Pohl C, Tennant A, Conaghan PG. Minimal clinically important difference as applied in rheumatology: an OMERACT Rasch Working Group systematic review and critique. *J Rheumatol* 2016; 43(1): 194-202.
13. Ebrahim S, Vercammen K, Sivanand A, Guyatt GH, Carrasco-Labra A, Fernandes RM et al. Minimally important differences in patient or proxy-reported outcome studies relevant to children: a systematic review. *Pediatrics* 2017; 139(3): e20160833.
14. Gemeinsamer Bundesausschuss. Verfahrensordnung des Gemeinsamen Bundesausschusses [online]. URL: <https://www.g-ba.de/informationen/richtlinien/42/>.
15. Gemeinsamer Bundesausschuss. Beschluss des Gemeinsamen Bundesausschusses über die Anpassung der Beauftragung des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen zur Erstellung von Patienteninformationen [online]. 13.03.2008 [Zugriff: 18.03.2015]. URL: <http://www.g-ba.de/downloads/39-261-650/2008-03-13-IQWiG-Anpassung-Generalauftrag.pdf>.
16. Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. *BMJ* 2007; 334(7589): 349-351.
17. Goldsmith ES, Taylor BC, Greer N, Murdoch M, MacDonald R, McKenzie L et al. Focused evidence review: psychometric properties of patient-reported outcome measures for chronic musculoskeletal pain. *J Gen Intern Med* 2018; 33(Suppl 1): 61-70.
18. Hao Q, Devji T, Zeraatkar D, Wang Y, Qasim A, Siemieniuk RAC et al. Minimal important differences for improvement in shoulder condition patient-reported outcomes: a systematic review to inform a BMJ Rapid Recommendation. *BMJ Open* 2019; 9(2): e028777.
19. Heres S, Davis J, Maino K, Jetzinger E, Kissling W, Leucht S. Why olanzapine beats risperidone, risperidone beats quetiapine, and quetiapine beats olanzapine: an exploratory analysis of head-to-head comparison studies of second-generation antipsychotics. *Am J Psychiatry* 2006; 163(2): 185-194.
20. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden: Version 4.0 [online]. 23.09.2011 [Zugriff: 16.10.2018]. URL: http://www.iqwig.de/download/IQWiG_Methoden_Version_4_0.pdf.
21. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Ramucirumab (Lungenkarzinom): Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag A16-11 [online]. 30.05.2016 [Zugriff: 06.07.2016]. (IQWiG-Berichte; Band 398). URL: https://www.iqwig.de/download/A16-11_Ramucirumab_Nutzenbewertung-35a-SGB-V.pdf.
22. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Osimertinib (nicht kleinzelliges Lungenkarzinom): Nutzenbewertung gemäß § 35a SGB V (Ablauf Befristung); Dossierbewertung; Auftrag A17-20 [online]. 28.07.2017 [Zugriff: 11.08.2017]. (IQWiG-Berichte; Band 527). URL: https://www.iqwig.de/download/A17-20_Osimertinib_Nutzenbewertung-35a-SGB-V_V1-0.pdf.

23. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Abemaciclib (Mammakarzinom; Kombination mit einem Aromatasehemmer): Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag A18-72 [online]. 30.01.2019 [Zugriff: 06.02.2019]. (IQWiG-Berichte; Band 718). URL: https://www.iqwig.de/download/A18-72_Abemaciclib_Nutzenbewertung-35a-SGB-V_V1-0.pdf.
24. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Vakuumversiegelungstherapie von Wunden mit intendierter primärer Wundheilung: Abschlussbericht; Auftrag N17-01B [online]. 12.06.2019 [Zugriff: 15.08.2019]. (IQWiG-Berichte; Band 777). URL: https://www.iqwig.de/download/N17-01B_Vakuumversiegelungstherapie-von-Wunden-primare-Wundheilung_Abschlussbericht_V1-0.pdf.
25. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Vakuumversiegelungstherapie von Wunden mit intendierter sekundärer Wundheilung: Abschlussbericht; Auftrag N17-01A; Version 1.1 [online]. 25.06.2019 [Zugriff: 17.07.2019]. (IQWiG-Berichte; Band 713). URL: https://www.iqwig.de/download/N17-01A_Abschlussbericht_Vakuumversiegelungstherapie-von-Wunden-mit-intendierter-sekundaerer-Wundheilung_V1-1.pdf.
26. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Konzepte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V: Rapid Report; Auftrag A19-43; Version 1.1 [online]. 13.05.2020 [Zugriff: 02.06.2020]. (IQWiG-Berichte; Band 863). URL: https://www.iqwig.de/download/A19-43_Versorgungsnahe-Daten-zum-Zwecke-der-Nutzenbewertung_Rapid-Report_V1-1.pdf.
27. Jayadevappa R, Cook R, Chhatre S. Minimal important difference to infer changes in health-related quality of life-a systematic review. J Clin Epidemiol 2017; 89: 188-198.
28. Jones PW. St. George's Respiratory Questionnaire: MCID. COPD 2005; 2(1): 75-79.
29. MacKay C, Clements N, Wong R, Davis AM. A systematic review of estimates of the minimally clinically important difference and patient acceptable symptom state of the Western Ontario and McMaster Universities Osteoarthritis Index in patients who underwent total hip and total knee replacement. Osteoarthritis and Cartilage 2019; 27(13): S238-S239.
30. Moustgaard H, Clayton GL, Jones HE, Boutron I, Jorgensen L, Laursen DRT et al. Impact of blinding on estimated treatment effects in randomised clinical trials: meta-epidemiological study. BMJ 2020; 368: l6802.
31. Musoro ZJ, Hamel JF, Ediebah DE, Cocks K, King MT, Groenvold M et al. Establishing anchor-based minimally important differences (MID) with the EORTC quality-of-life measures: a meta-analysis protocol. BMJ Open 2018; 8(1): e019117.
32. Nordin A, Taft C, Lundgren-Nilsson A, Dencker A. Minimal important differences for fatigue patient reported outcome measures-a systematic review. BMC Med Res Methodol 2016; 16: 62.

33. Nwachukwu BU, Runyon RS, Kahlenberg CA, Gausden EB, Schairer WW, Allen AA. How are we measuring clinically important outcome for operative treatments in sports medicine? *Physician and Sportsmedicine* 2017; 45(2): 159-164.
34. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol* 1998; 16(1): 139-144.
35. Ousmen A, Touraine C, Deliu N, Cottone F, Bonnetain F, Efficace F et al. Distribution- and anchor-based methods to determine the minimally important difference on patient-reported outcome questionnaires in oncology: a structured review. *Health Qual Life Outcomes* 2018; 16(1): 228.
36. Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. *Lancet* 2002; 359(9308): 781-785.
37. St-Pierre C, Desmeules F, Dionne CE, Fremont P, MacDermid JC, Roy JS. Psychometric properties of self-reported questionnaires for the evaluation of symptoms and functional limitations in individuals with rotator cuff disorders: a systematic review. *Disabil Rehabil* 2016; 38(2): 103-122.
38. Unkel S, Amiri M, Benda N, Beyersmann J, Knoerzer D, Kupas K et al. On estimands and the analysis of adverse events in the presence of varying follow-up times within the benefit assessment of therapies. *Pharm Stat* 2019; 18(2): 166-183.
39. Welling JB, Hartman JE, Ten Hacken NH, Klooster K, Slebos DJ. The minimal important difference for the St George's Respiratory Questionnaire in patients with severe COPD. *Eur Respir J* 2015; 46(6): 1598-1604.
40. Yang F, Wittes J, Pitt B. Beware of on-treatment safety analyses. *Clin Trials* 2019; 16(1): 63-70.

3 Dokumentation der wissenschaftlichen Erörterung – Teilnehmerliste, Tagesordnung und Protokoll

3.1 Teilnehmerliste der wissenschaftlichen Erörterung

Name	Organisation / Institution / Firma / privat
Andreas, Jens-Otto	UCB Pharma GmbH
Bender, Ralf	IQWiG
Bierl, Michael	Flatiron Health, Inc.
Böhm, Doris	IQVIA Commercial GmbH & Co. OHG
Dheban, Srirangan	Janssen-Cilag GmbH
Dintsios, Charalabos-Markos	Bayer Vital GmbH
Filla, Christoph	Protokollant (Sitzungsprotokollarischer Dienst des Landtags NRW)
Flintrop, Jens	IQWiG
Friede, Tim	Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e. V. (GMDS) / Internationale Biometrische Gesellschaft – Deutsche Region (IBS-DR) – (gemeinsame Stellungnahme)
Kiencke, Peter	Novo Nordisk Pharma GmbH
Kohlmann, Thomas	Privat
Kucka, Annett	Bristol-Myers Squibb GmbH & Co. KGaA
Kuß, Oliver	Wissenschaftlicher Beirat
Lerche, Marianne	Geschäftsstelle der Beauftragten der Bundesregierung für die Belange der Patientinnen und Patienten
Leverkus, Friedhelm	Pfizer Deutschland GmbH
Loske, Sonja	SmartStep Consulting GmbH
Lützelberger, Uwe	AstraZeneca GmbH
Michnacs, Kristian	Bundesverband Medizintechnologie e. V. (BVMed)
Osowski, Ulrike	Merck Serono GmbH
Pfarr, Egon	Boehringer Ingelheim Pharma GmbH & Co. KG
Sauerbruch, Tilmann	Deutsche Gesellschaft für Innere Medizin e. V. (DGIM)
Sauerland, Stefan	IQWiG
Semrau, Frank	Privat
Skipka, Guido	IQWiG
Sternberg, Kati	AbbVie Deutschland GmbH & Co. KG
Ten Haaf, Petra	Bundesverband der Arzneimittel-Hersteller e. V. (BAH)
Thomas, Stefanie	IQWiG
Vervölgyi, Volker	IQWiG
Von Wulffen, Meike	Ecker + Ecker GmbH
Werner, Sebastian	Verband Forschender Arzneimittelhersteller e. V. (vfa)
Wieseler, Beate	IQWiG
Wille, Hans	Arzneimittelkommission der deutschen Ärzteschaft (AkdÄ)
Windeler, Jürgen	IQWiG (Moderation)

3.2 Liste der Stellungnahmen, zu denen keine Vertreterin bzw. kein Vertreter an der wissenschaftlichen Erörterung teilgenommen hat

In der folgenden Tabelle werden Stellungnahmen genannt, zu denen kein Stellungnehmender oder Vertreter bei der wissenschaftlichen Erörterung teilnehmen konnte.

Organisation / Institution / Firma / Privatperson
AMGEN GmbH
Bundesverband der Pharmazeutischen Industrie e. V. (BPI)
Deutsche Gesellschaft für Gesundheitsökonomie e. V. (dggö)
Deutsche Gesellschaft für Hämatologie und Medizinische Onkologie e. V. (DGHO)
Deutsche Gesellschaft für Orthopädie und Unfallchirurgie e. V. (DGOU)
Deutsche Krankenhausgesellschaft e. V. (DKG)
GlaxoSmithKline GmbH & Co. KG
Kassenzahnärztliche Bundesvereinigung K. d. ö. R. (KZBV)
Lilly Deutschland GmbH
Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen e. V. (MDS)
MSD Sharp & Dohme GmbH
Novartis Pharma GmbH
Roche Pharma AG
Sanofi-Aventis Deutschland GmbH
Takeda Pharma Vertrieb GmbH & Co. KG

3.3 Tagesordnung der wissenschaftlichen Erörterung

	Begrüßung und Einleitung
TOP 1	Ausmaßbestimmung für SMDs (Abschnitt 3.3.3)
TOP 2	Festlegung der Responseschwelle von 15 % (Abschnitt 9.3.3)
TOP 3	Festlegung von und Umgang mit Publikationsbias (Abschnitt 3.2.1)
TOP 4	Aufbereitung und Darstellung von Schadaspekten (Abschnitt 3.1.3)
TOP 5	Umgang mit unvollständigen Daten (Abschnitt 9.3.12 und 9.3.13)
TOP 6	Verschiedenes

3.4 Protokoll der wissenschaftlichen Erörterung

Datum: 19.06.2020, 10:30 bis 12:50 Uhr

Ort: Park Inn by Radisson Köln City West, Innere Kanalstraße 15, 50823 Köln

Moderation: Jürgen Windeler

3.4.1 Begrüßung und Einleitung

Moderator Jürgen Windeler: Meine Damen und Herren! Ich begrüße Sie sehr herzlich zur wissenschaftlichen Erörterung der Stellungnahmen zum Methodenpapier Version 6.0. Ich freue mich, dass Sie erschienen sind, freue mich, dass wir mit Ihnen diskutieren, uns mit Ihnen austauschen können, vielleicht unsere Fragen beantwortet finden. Ich will zu der konkreten, für uns alle offensichtlich besonderen Situation kein weiteres Wort verlieren. Sie sehen und merken, dass die Situation heute speziell ist.

Das Spezielle – das vielleicht vorweg – hat folgende zwei Auswirkungen: Erstens. Sie finden eine Teilnehmerliste, auf der nur Ihr Name steht. Diese legen Sie bitte nach der Veranstaltung in den Kasten am Counter.

Zweitens finden Sie einen Zettel vom Hotel, den Sie bitte ausfüllen. Das sind die Regularien speziell in Nordrhein-Westfalen; in anderen Bundesländern sind sie vermutlich ähnlich. Den füllen Sie bitte aus, auch in Ihrem Interesse, mit einer Telefonnummer bzw. Kontaktadresse, unter der man Sie auch kurzfristig erreichen kann. Lassen Sie diesen Zettel bitte einfach auf Ihrem Tisch liegen. Das ist für das Hotel, damit das Hotel uns erreichen kann, falls es die Notwendigkeit dafür sieht.

So viel zu diesen speziellen Dingen. Ansonsten sind Sie auf Ihren Plätzen hoffentlich ausreichend versorgt für die drei Stunden, und wir können starten.

Für diejenigen, die das zum ersten Mal mitmachen – ich sehe eine ganze Menge bekannter Gesichter –: Wir führen diese Erörterungen durch, um offene Fragen unsererseits zu klären. Die Erörterungen haben also nicht den Charakter einer mündlichen Wiederholung oder des Vorlesens von Stellungnahmen. Das brauchen wir hier nicht. Vielmehr versuchen wir, die Dinge, die für uns offengeblieben sind, mit Ihnen zu erörtern, zu diskutieren, um sie möglichst zu klären. Gehen Sie also bitte auf dieser Basis davon aus, dass wir Ihre Stellungnahmen sehr, sehr sorgfältig gelesen haben, dass wir glauben, die Dinge, die wir heute nicht thematisieren, auch verstanden zu haben, und sie berücksichtigen. Berücksichtigen heißt, wir setzen uns mit Ihren Stellungnahmen auseinander und setzen sie um, wenn sie uns überzeugen. Insofern hoffe ich, dass wir die offenen Fragen heute noch klären können.

Die Erörterung wird aufgezeichnet. Das kennen die meisten von Ihnen. Sie haben Ihr Einverständnis erklärt, indem Sie hierhergekommen sind, dass diese Aufzeichnung veröffentlicht wird. Ich weise an dieser Stelle aus einem Anlass der letzten Tage darauf hin, dass eigene Aufzeichnungen nicht erlaubt sind. Die Erörterung wird, wie gesagt, protokolliert, sowohl per Band als auch von einem Stenografen. Es ist daher erforderlich, damit alle Beiträge den jeweiligen Redenden zugewiesen werden können, dass Sie bitte bei jeder Wortmeldung – also nicht nur beim ersten Mal, um sich vorzustellen – Ihren Namen sagen – ich werde mir erlauben, Sie ab und zu daran zu erinnern, falls das in Vergessenheit geraten sollte –, sodass wir auch im Protokoll festhalten können, wer was gesagt hat.

Wir werden keine Pause machen; das machen wir bei Erörterungen sowieso nicht. Wenn jemand dringende Bedürfnisse hat, dann möge er diese irgendwie zwischendurch einschieben und diesen unauffällig – so ganz unauffällig wird das nicht sein; denn ich habe Sie alle im Blick – nachgehen.

Gibt es Fragen zum Ablauf? – Das ist nicht der Fall.

Herr Flintrop wird schauen, wer sich zu Wort meldet, und wird Sie dann auch aufrufen, sodass wir uns diese Arbeit ein bisschen teilen können. Das ist bei diesem großen Raum nicht so ganz einfach; aber das kriegen wir hin.

Wir haben Ihnen eine Tagesordnung zugeschickt, auch was die Themen angeht. Gibt es Fragen zur Tagesordnung? – Dass man unter „Verschiedenes“ wie üblich noch Verschiedenes ansprechen kann, ist klar. Ansonsten sehe ich im Moment keine weiteren Unklarheiten und möchte daher jetzt in die Tagesordnung einsteigen. Wir haben drei Stunden Zeit. Das bedeutet bei sechs Tagesordnungspunkten etwa 30 Minuten pro Tagesordnungspunkt. Ich werde ein bisschen gucken, dass wir nicht gleich beim ersten Tagesordnungspunkt sozusagen völlig aus dem Ruder laufen. Aber ich denke, wir haben gute Chancen, das in dieser Zeit zu bewältigen.

3.4.2 Tagesordnungspunkt 1: Ausmaßbestimmung für SMDs (Abschnitt 3.3.3)

Ralf Bender übernimmt den Punkt.

Ralf Bender: TOP 1 lautet „Ausmaßbestimmung für SMDs“. Wir haben mit dem Entwurf zum Methodenpapier Version 6.0 zum ersten Mal als Vorschlag eine Tabelle vorgelegt, die Schwellenwerte beinhaltet, mit deren Hilfe wir das Ausmaß des Zusatznutzens auch bei stetigen Daten, nämlich für das Effektmaß, standardisierte Mittelwertdifferenz, bestimmen können.

Dieser Vorschlag wurde von den Stellungnehmenden weitgehend begrüßt und auch als sinnvoll bezeichnet. Es gab von fast allen aber auch den Wunsch nach mehr Details und Erläuterungen. Wir haben das relativ knapp gehalten, weil die Methodik, die dahintersteckt, eigentlich analog zu der Methodik für binäre Daten, die wir hier seit 2011 anwenden, ist, aber wir sehen, dass doch der Wunsch nach mehr Details da ist. Diesem Wunsch werden wir nachkommen. Es spricht einiges dafür, dass wir den Vorschlag der Biometrischen Gesellschaft GMDS aufnehmen, das im Anhang zu tun. – Das vielleicht erst einmal vorweg.

Es gab aber auch durchaus Kritik an diesem Vorschlag, insbesondere von der pharmazeutischen Industrie. Dort wurde gesagt, dass das nicht nachvollziehbar ist und dass hier die Notwendigkeit besteht, doch Kriterien und weitere Quellen als Grundlage für diesen Vorschlag zu benennen. Wir haben uns bei diesem Vorschlag auf die groben Grenzen bezogen, die von Cohen vor langer Zeit vorgeschlagen worden sind. Offensichtlich reicht vielen diese Quelle nicht aus. Und meine Frage ist jetzt an Pfizer, aber im Grunde genommen an die Vertreter der pharmazeutischen Industrie, welche Quellen denn bekannt sind, die wir da außer dem Lehrbuch von Cohen noch zu nennen haben.

Friedhelm Leverkus: Leverkus, Pfizer. – Ich finde es auch gut, dass die stetigen Daten jetzt eine stärkere Rolle in der Nutzenbewertung spielen können. Wir wissen auch keine anderen Quellen, würden uns aber andere Quellen wünschen. Es ist sicherlich so, dass der Vorschlag, den das IQWiG da macht, ein sehr konservativer Vorschlag ist.

Abgesehen davon gibt es aus Sicht der Industrie natürlich generell eine Kritik an der Vorgehensweise. Dazu kann Sebastian Werner vielleicht noch etwas sagen. Uns haben hier im Prinzip insbesondere Transparenz und Klarheit gefehlt. – Danke.

Sebastian Werner: Ich würde auch gerne etwas dazu sagen, wenn das geht.

Jens Flintrop: Gerne, Herr Werner.

Sebastian Werner: Sebastian Werner, vfa. – Auch wir begrüßen das Ziel, dass es eine Methode gibt, die die Quantifizierung jetzt erlaubt. Wir haben aber tatsächlich Bedenken bei den Schwellen, die Sie da angegeben haben – aber nur teilweise. Denn die nachvollziehbaren Schwellen sehen wir schon in der üblichen Einteilung von Cohen, also 0,2, 0,5, 0,8, und das sind aus unserer Sicht auch die Werte, die wir ganz klar unterstützen, weil sie international akzeptiert sind und aus unserer Sicht auch dem Stand der Erkenntnisse oder den Standards der evidenzbasierten Medizin zuzuordnen sind.

Für die anderen Werte, die Sie in der Tabelle ausgelegt haben, insbesondere zur Zielgrößenkategorie „schwerwiegend“, sehen wir das tatsächlich nicht so. Das halten wir für schwer nachvollziehbar und sehen das deswegen als problematisch an.

Aber was wir eigentlich als noch viel problematischer ansehen, ist – das hat Herr Leverkus auch schon angesprochen –, dass Sie das Konfidenzintervall bei Ihrer Festlegung berücksichtigen und damit die Methode, wie Sie sie vorschlagen, weitaus konservativer ist als die von Cohen vorgeschlagene Methode, die eigentlich auf den Punktschätzer abzielt. Deswegen würden wir uns eigentlich wünschen, dass man die akzeptierte Einteilung von Cohen anwendet, nicht modifiziert. Aus unserer Sicht ist deswegen auch eine Unterteilung in die Zielgrößenkategorien eigentlich nicht notwendigerweise erforderlich. Das Wichtigste ist eigentlich, dass hier noch einmal überlegt werden sollte, ob man, wenn man sich auf Cohen bezieht, dann unbedingt die Konfidenzintervallgrenze berücksichtigen muss oder ob ein Punktschätzer ausreicht. – Danke.

Ralf Bender: Ich komme einmal zurück zu meiner Ausgangsfrage. Also, andere Quellen wurden jetzt nicht benannt. Dann gehe ich davon aus, dass auch hier keine anderen Quellen bekannt sind, die wir im Methodenpapier zu zitieren haben. Wenn das doch der Fall sein sollte, kann man das ja hinterher noch mal sagen.

Kurz zu den anderen Punkten. Ja, das haben wir im Grunde genommen auch vor Jahren schon mal diskutiert. Das ist analog zum Vorgehen bei binären Daten. Ich glaube nicht, dass wir das hier an dieser Stelle noch mal diskutieren müssen, warum es diese Unterscheidung zwischen schwerwiegenden und nicht schwerwiegenden Ereignissen gibt. Es ist selbstverständlich, dass

wir diese Unterteilung gemäß AMNOG machen müssen. Das ist so wie bei binären Daten, und das sehen wir auch bei stetigen Daten nicht anders.

Noch eine kurze Erläuterung. Natürlich müssen wir von den Konfidenzintervallen ausgehen und nicht einfach Schwellen für den Punktschätzer beschreiben. Auch das ist analog zum Vorgehen bei binären Daten. Zur Ableitung dieser Schwellenwerte haben wir aber sehr wohl sozusagen auf den wahren Wert geschaut. Wir sehen diese Grenzen von Cohen's d natürlich bezogen auf den theoretisch wahren Wert, und daraus leiten wir dann gemäß dem Vorgehen mit diesen zwei pivotalen Studien in der Metaanalyse usw. wie bei binären Daten die Schwellenwerte auch für SMDs ab. Das ist nichts anderes. Das haben wir schon damals diskutiert, und ich glaube, wir müssen das alles hier jetzt nicht noch mal im Detail neu aufrollen.

Jens Flintrop: Herr Skipka.

Guido Skipka: Mich erstaunt dieser Einwand auch ein bisschen, dass man hier bei stetigen Daten nicht Konfidenzintervallgrenzen zur Ausmaßbestimmung heranziehen sollte. Wenn wir das jetzt bei stetigen Daten anders machen, kriegen wir natürlich eine extreme Diskrepanz zu binären Daten, wo wir mittlerweile seit 2011 anhand von Konfidenzintervallgrenzen das Ausmaß bestimmen. Ich denke, wir sind uns einig, dass sich das auch irgendwie etabliert hat. Diese Diskrepanz zwischen den Messniveaus könnte ich überhaupt nicht verstehen und auch nicht vertreten.

Moderator Jürgen Windeler: Gibt es weiteren Gesprächsbedarf? Wir liegen noch deutlich unter 30 Minuten. – Die Frage zu den Quellen ist durch die Äußerung von Herrn Leverkus unmissverständlich beantwortet worden. Die andere Frage zur Konsistenz des Vorgehens bei stetigen und binären Daten ist auch noch einmal verdeutlicht worden, sodass wir, wenn wir jetzt über das Vorgehen in den Eckpunkten, die gerade beschrieben worden sind, diskutieren, das, wie Ralf Bender gesagt hat, noch mal ganz von vorne diskutieren, auch was die gesamte Systematik angeht, woran ich niemanden hindern will, was aber nicht so richtig irgendwohin führt. – Herr Dintsios, bitte.

Charalabos-Markos Dintsios: Als die Matrix – ich nenne sie mal so für das Protokoll – für die binären Endpunkte entwickelt wurde, war vielleicht die Erfahrung hinsichtlich der Anzahl an pivotalen Studien noch eine andere. Wir waren relativ früh im AMNOG. Es gab auch zwischenzeitlich eine Veröffentlichung von IQWiG-Mitarbeitern bzw. von Stefan Lange, der sich unter anderem auch die Studienanzahl in der Onkologie angesehen hat. Ich kann verstehen, dass man, wenn man etwas „from the scratch“, also von Anfang an, entwickelt, auch idealtypisierte Annahmen dazu verwendet, und hier waren es die zwei oder mindestens zwei pivotale Studien.

Mittlerweile sind wir aber in der ersten Dekade beim AMNOG, und wir wissen, dass in vielen Indikationsgebieten die Anzahl der pivotalen Studien nicht $n = 2$ oder größer ist, sondern durchaus auch bei $n = 1$ liegen kann. Wenn ich das aber sozusagen perpetuiere, gehe ich bei

dieser Annahme an der Realität durchaus in Teilen vorbei. Das würde ich mir wünschen, dass es irgendwie auch mit berücksichtigt wird, weil es nun mal eine andere Realität ist als in der Annahme, die man vielleicht 2011 gemacht hat, die sehr idealtypisch war.

Ralf Bender: Auch diesen Punkt haben wir damals wiederholt ausführlich diskutiert. Natürlich wissen wir auch, dass wir diese Situation von zwei Studien nicht immer haben. Sehr häufig haben wir tatsächlich gerade in der frühen Nutzenbewertung leider nur eine Studie. Wir wissen aber auch, dass, um aus dieser einen Studie einen Beleg zu erhalten, diese Studie besondere Anforderungen erfüllen muss. Sie muss besonders groß sein und in der Regel mehr Power haben als eine Studie, die nicht zu einem Beleg ausreicht. Wir haben aber genauso die Situation, dass wir manchmal drei oder vier Studien haben. Dann kommt noch die Möglichkeit hinzu, indirekte Vergleiche durchzuführen. Wir verwenden dieses Modell mit diesen zwei Studien, um diese Schwellenwerte abzuleiten, und ob wir diese Schwellen dann erreichen oder nicht, hängt sehr vom Stichprobenumfang dieser einzelnen Studien ab, von der konkreten Situation. Häufig hat man zu wenig Power. Aber das ist dann eben so, dass man zu wenig Power hat, um ein gewisses Ausmaß zu erreichen. Dann ist man in der Ausmaßstufe eine Stufe niedriger. Wenn man einen Nachweis haben will, dass man in einer bestimmten Ausmaßkategorie landet, haben wir eben gewisse Power-Anforderungen, und wir sehen keine Möglichkeit, das irgendwie zu ändern. Jetzt hier sehr liberal sozusagen als Regelfall nur von einer Studie auszugehen, halten wir nicht für gerechtfertigt.

Jens Flintrop: Guido Skipka.

Guido Skipka: Markos, ich habe eine Rückfrage dazu. Verstehe ich dich richtig, dass du dafür plädiert, dass wir die Ausmaßgrenzen in Abhängigkeit der Anzahl von Studien bestimmen? Das würde das Ganze natürlich deutlich verkomplizieren.

Charalabos-Markos Dintsios: Nein, das habe ich so nicht behauptet oder behaupten wollen. Ich habe nur darauf hingewiesen, dass einige Annahmen natürlich sehr idealtypisch sind und dass dadurch das, was dann herauskommt, simulationstechnisch eben nicht vorlag oder analog zu den binären Endpunkten natürlich auch zu Schwellen führt, die vielleicht aus einigen Indikationsgebieten betrachtet durchaus aufgrund der Pragmatik, dass sehr oft nur eine Studie vorhanden ist, hyperkonservativ ausfallen. Ich gebe Ralf Bender auch recht. Es gibt Indikationsgebiete oder auch Fälle, wo wir drei oder vier hatten. Nicht, dass ich mich hier falsch ausgedrückt habe oder falsch verstanden wurde. Aber trotzdem ist es natürlich ein sehr, sehr konservatives Vorgehen, weil es idealtypisch ist.

Ich möchte noch zu bedenken geben, dass andere HTA-Agenturen nicht unbedingt hinsichtlich der Quantifizierung des Zusatznutzens – ich nehme das Beispiel Frankreich als Nachbarland, das letztendlich nichts anderes als relative effectiveness analysis macht mit einer anderen Kategorisierung, mit ASMR (Amélioration du service médical rendu), die ohne Verwendung der Präzision auskommt – ... wir dort sozusagen diese Vorgehensweise gesehen haben und trotzdem auch in diesem Land eine Quantifizierung des Zusatznutzens möglich erscheint.

Ralf Bender: Vielleicht kurz als Ergänzung: Klar, es gibt andere Länder, die machen das anders. Es gibt aber auch Länder, die haben sogar den IQWiG-Vorschlag übernommen. Wir sind hier aber in Deutschland, und das AMNOG gibt es in der Form erst mal nur in Deutschland. Wir sehen unsere Aufgabe darin, hierfür entsprechende Schwellenwerte vorzulegen, um eine Ausmaßbestimmung hier zu ermöglichen.

Jens Flintrop: Herr Dintsios.

Charalabos-Markos Dintsios: Ich bestreite das nicht, Ralf. Natürlich bewegen wir uns hier in Deutschland. Aber wir machen als international tätige Pharmaindustrie Studien für die ganze Welt oder für Europa. Wir machen sie nicht explizit für Deutschland. Sie können es gerne aufnehmen, Herr Stenograf – ich sage es sehr langsam, sehr deutlich –: Am deutschen HTA-Wesen kann nicht der Rest Europas genesen! – Meinen Namen können Sie dahinter setzen.

Moderator Jürgen Windeler: Das hat der Stenograf sorgfältig mitgeschrieben. – Gibt es weitere Wortmeldungen? – Herr Friede.

Tim Friede: Vielen Dank. – Tim Friede, Universitätsmedizin Göttingen. Ich vertrete hier heute die Präsidiumskommission von GMDS und IBS-DR. Vielen Dank für die Möglichkeit, die neue Version des Methodenpapiers kommentieren zu dürfen und heute auch an der Veranstaltung teilnehmen zu können.

Wie schon vorhin einleitend erwähnt, haben wir den entsprechenden Abschnitt kommentiert. Unser Verständnis war, dass man hier Dinge auf diesen Outcome der standardisierten Mittelwertdifferenz übertragen will, und wir haben es nicht so verstanden, dass man generell das Vorgehen diskutieren will. Deswegen haben wir uns in den Kommentaren auch so darauf bezogen und den Gesamtrahmen erst einmal so akzeptiert.

Ich würde an dieser Stelle auch noch mal bestätigen wollen, dass die genannte Referenz eigentlich die ist, die ich auch standardmäßig verwenden würde in diesem Kontext. Da können wir uns wahrscheinlich alle bei Herrn Cohen bedanken, dass er das mal so aufgeschrieben hat. Mir ist auch kein weiterer Ansatz bekannt.

Ich meine, wir als Fachgesellschaften sind natürlich immer offen für Diskussionen, wie man so ein Konzept in der Zukunft vielleicht auch anders gestalten kann. Aber wir haben das jetzt erst einmal so verstanden, dass das zumindest hier nicht angedacht war, und haben das deswegen auch im Rahmen der anderen Endpunkte so kommentiert. Wir begrüßen das Vorgehen, würden uns aber freuen, wenn man im Appendix die Ausführungen, wie man zu diesem Algorithmus hingelangt, noch etwas präzisieren würde. – Vielen Dank.

Moderator Jürgen Windeler: Gibt es weitere Wortmeldungen zu diesem Thema? – Das ist nicht der Fall.

Damit schließe ich diesen Tagesordnungspunkt und komme zu:

3.4.3 Tagesordnungspunkt 2: Festlegung der Responseschwelle von 15 % (Abschnitt 9.3.3)

Frau Wieseler.

Beate Wieseler: Wir haben mit dem Entwurf zum Methodenpapier einen Vorschlag vorgelegt, wie wir Responseschwellen für patientenberichtete Endpunkte bzw. für komplexe Skalen definieren möchten. Das ist eine Änderung im Vergleich zu den Vorgängerversionen, was die Definition der Responseschwellen angeht. Es ist keine Änderung bezüglich des Stellenwerts von Responderanalysen für die Relevanzbewertung. Wir haben dazu eine ganze Reihe von Stellungnahmen bekommen.

Es gibt in diesen Stellungnahmen, glaube ich, zwei Missverständnisse, die ich gern zu Beginn adressieren möchte, einfach um die Diskussion zu erleichtern. Also, was wir mit diesem Vorschlag von 15 % der Skalenrange meinen, ist nicht eine MID. Wir behaupten nicht, dass dieser Wert eine MID darstellt, sondern eine Responseschwelle, bei der wir hinreichend sicher sind, dass wir eine für Patientinnen und Patienten spürbare Änderung abbilden. Dieses Verfahren mit diesen 15 % soll auch nicht auf jede beliebige Skala angewendet werden, insbesondere nicht auf solche mit natürlichen Einheiten, die man auch anders interpretieren kann, die man einfach in einer inhaltlichen Bewertung interpretieren kann, sondern es geht um komplexe Skalen, die im Ergebnis in einem Punktwert enden, der so erst einmal nicht interpretierbar ist. Also, es geht nicht um eine 6-Minuten-Gehstrecke oder solche Dinge, die ich in ihrer Relevanz inhaltlich diskutieren kann, sondern es geht um ein EORTC, wo ich am Schluss sehe, dass im Mittel zwei Punkte Änderung vorliegen, und ich keinerlei Vorstellung davon habe, ob das relevant und spürbar ist. – Das vielleicht zur Klärung vorab.

Warum haben wir diesen neuen Vorschlag gemacht? In den letzten Jahren – vielleicht seit 2015 – wird diese Methode ganz intensiv diskutiert, vermutlich auch deshalb, weil die patientenberichteten Endpunkte einen höheren Stellenwert im Bewertungsverfahren haben. Die wissenschaftliche Erkenntnis, die sich in den letzten Jahren herausgebildet hat, zeigt, dass diese erste Idee der Minimal Important Difference nicht trivial in die Praxis zu überführen ist. Wenn wir uns die Empirie zu den Studien anschauen, in denen MIDs bestimmt werden, dann sehen wir, dass es für einzelne Instrumente sehr viele verschiedene MIDs mit einer weiten Range gibt. Wir sehen, dass diese MID von allen möglichen Faktoren abhängig ist, von methodischen Faktoren. Also, mit welchem statistischen Verfahren ermittle ich den Wert? Die MID ist abhängig von Patientencharakteristika, vom Schweregrad. Sie ist abhängig davon, in welche Richtung ich schaue: eine Verbesserung, eine Verschlechterung? – Also, wir haben eine ganz hohe Variabilität und Heterogenität in diesen MID-Werten. – Herr Kohlmann nickt.

In dieser Situation und auch ausgehend von den Erfahrungen, die wir in den Dossierbewertungen regelhaft machen, mit den Schwierigkeiten und den Diskussionen, die wir da haben, uns auf eine MID festzulegen, haben wir die Notwendigkeit gesehen, dieses Verfahren anzupassen. Daraus resultiert dieser Vorschlag, der auf Empirie beruht und ein pragmatischer Vorschlag ist, um das Verfahren mit hinreichender Sicherheit durchführen zu können. Also,

wie gesagt: keine MID, nicht für alle Skalen, aber aus unserer Sicht auf Basis der Empirie, die wir uns angeschaut haben, ein pragmatischer Vorschlag für die Definition von Response-schwellen, die hinreichend sicher spürbare Veränderungen abbilden.

Was der Vorschlag auch enthält, ist die Option einer sehr definierten Post-hoc-Analyse. Wir denken, dass wir dadurch auch erreichen, dass Studien, in denen die MID oder die Responderanalyse nicht primär so angelegt wurde, dass sie dieses neue Kriterium trifft, trotzdem verwendet werden können. Das ist ein Aspekt, den wir durchaus in unserem Vorschlag abgebildet sehen. – Das vielleicht noch mal zur Erläuterung.

Meine Frage in diesem Zusammenhang wäre: Relativ viele Stellungnahmen haben vorgeschlagen, dass es Arbeitsgruppen geben sollte, die sich mit diesem Vorschlag noch einmal beschäftigen. Herr Andreas, wir haben ja bereits im vergangenen Jahr über ein solches Treffen diskutiert. Sie haben die Funktion als Vertreter einer Arbeitsgruppe des vfa. Mit welchem Methodenvorschlag würden Sie denn in so eine Diskussion hineingehen?

Jens-Otto Andreas: Jens-Otto Andreas, UCB. – Vielen Dank, Frau Wieseler. Es ist richtig: Wir hatten im letzten Jahr schon telefoniert zu dem Thema, weil wir uns ja auch aufgrund der Diskussionen in den Nutzenbewertungen mit dem Thema befasst hatten. Und unser Vorschlag wäre, methodisch darauf aufzubauen. Es ist ja nicht so, dass wir uns da im luftleeren Raum bewegen. Wir hatten ja auch über die McMaster-Universität gesprochen, Guyatt; da gibt es ja die Vorschläge. Es gibt auch andere Publikationen diesbezüglich, die Kriterien wie zum Beispiel Ankerbasiertheit, Längsschnitt, Korrelation beinhalten. Diese Dinge sind alle nicht ganz neu, sondern mehrfach publiziert und auch von Ihnen referenziert oder als Kritik angegeben worden, dass kein Längsschnitt stattgefunden hat oder ein Anker nicht gültig ist für MIDs in früheren Verfahren. Daher wäre das die Basis, auf der man meiner Meinung nach für das, was da ist, im gemeinsamen Dialog Kriterien entwickeln sollte, mit der man dann die Güte einer MID-Validierung dann beurteilen könnte.

Zu den 15 %. Sie haben gesagt, das ist keine MID. Nichtsdestotrotz ist mir nicht so richtig klar – wir reden ja über patientenrelevante MIDs, also patientenrelevante Messungen, die da stattfinden –, wie da die Patientenrelevanz mit ins Spiel kommt. Das wäre dann eine Art Obergrenze, weil viele MIDs, die publiziert sind, bewegen sich ja unterhalb dieser Schwelle. Daher kann man sagen: Okay, das ist irgendwas, das passt auf jeden Fall. – Aber letztendlich legt man die Hürde doch verdammt hoch. Und dieses Pauschale, zu sagen: „Einer für alle“, steht für mich ein bisschen im Gegensatz zu der Patientenrelevanz. Daher wäre es schön, wenn wir da im gemeinsamen Dialog Kriterien entwickeln könnten, die dann auch in das Methodenpapier einfließen sollten.

Beate Wieseler: Sie haben jetzt verschiedene Punkte angesprochen, also zum einen die Frage der Patientenrelevanz. Die Patientenrelevanz ist erst mal primär durch den Endpunkt selbst gegeben. Also, wenn Sie eine adäquat entwickelte Skala haben, ist die Patientenrelevanz des Endpunkts aus meiner Sicht erst einmal durch die Tatsache, dass Sie dieses Instrument

verwenden, gegeben. Hier geht es jetzt um die Frage, welcher individuelle Unterschied, welche Änderung relevant für den Patienten wäre. Da haben wir uns in unserem Vorschlag durchaus an der vorhandenen Evidenz orientiert.

Also vielleicht ganz kurz im Überblick: Wir haben eine fokussierte, systematische Recherche nach systematischen Reviews von MIDs gemacht und haben aus diesen Arbeiten die MIDs identifiziert, die nach akzeptierten Verfahren ermittelt wurden, und haben uns auf dieser empirischen Basis angeschaut, in welcher Range, in welchem Anteil der Skalen-Range diese MIDs liegen, die nach den aktuellen Verfahren empirisch ermittelt werden. Wir haben dann einen Wert gewählt mit diesen 15 %, der etwas über den MIDs, die aktuell empirisch ermittelt werden, liegt. Das ist nicht irgendwie eine obere Grenze oder so – es gibt durchaus MIDs aus unterschiedlichen Arbeiten, die über diesen 15 % liegen –, sondern es ist ein pragmatischer Vorschlag, den wir für plausibel halten.

Das liegt etwas oberhalb der MID, weil wir damit auch versuchen, einen Teil der Variabilität, die durch andere Faktoren entsteht, mit abzufangen, also zum Beispiel die Variabilität, die durch die Richtung der Veränderung mit hineinkommt, die Variabilität, die dadurch hineinkommt, welches statistische Verfahren ich verwende, um diese MID zu bestimmen. Also, es handelt sich in der Tat um einen empirisch gestützten pragmatischen Vorschlag.

Sie haben darauf Bezug genommen, dass das verdammt hoch ist. Also, das finde ich nicht. Wenn man sich die Ranges anschaut, die wir sehen, dann gehen die zum Beispiel bis 35 % der Skalen-Range für eine MID. Da finde ich 15 % jetzt nicht verdammt hoch.

Die Frage wäre ja auch: Welche Konsequenz hat das denn, dass diese Responseschwelle jetzt ein bisschen höher ist als eine mit vielen Fragezeichen versehene potenzielle MID? – Ich kenne jetzt keine Daten, die zeigen würden, dass ich mit einem etwas höheren, aber immer noch kleinen Responsekriterium auf einmal keine Effekte mehr sehe. Ich weiß nicht, ob Sie solche Daten kennen. Das ist für uns auch schwer zu beantworten, weil Sie müssten diese Analysen machen. Das kann man nur auf Basis der individuellen Patientendaten machen. Also, ich kann dieses „Das ist verdammt hoch“ so noch nicht nachvollziehen. Ich weiß nicht, ob Sie diesbezüglich noch Daten haben, die das bestätigen würden, dass das verdammt hoch ist.

Jens-Otto Andreas: Darf ich direkt antworten?

Moderator Jürgen Windeler: Ich würde gerne ein bisschen auf die Rednerliste achten, damit wir kein Zwiegespräch machen. Sie sind zwar direkt angesprochen worden – das habe ich gesehen –, aber es sind noch zwei Redner vor Ihnen dran, und Sie können sich bitte noch einen Moment gedulden.

Jens Flintrop: Die Rednerliste ist sogar noch länger. Zunächst ist Herr Friede dran, danach Frau Thomas, Herr Kohlmann und Herr Werner. Danach wären Sie dran, Herr Andreas.

Tim Friede: Wir haben diesen Punkt auch in der Präsidiumskommission diskutiert, ihn letztendlich aber nicht aufgenommen in unserer formalen Stellungnahme. Ich denke, wenn ich die Diskussion so erinnere und kurz zusammenfasse, dann ist das, was ich durchaus begrüße, dieses systematische Vorgehen, die empirische Evidenz hier zusammenzutragen. Ich glaube, der Schritt, der dann die meisten Kolleginnen und Kollegen erstaunt hat, ist, dann zu sagen: Wenn wir denken, typischerweise liegt es im Bereich 10 bis 20 %, dann gehen wir einfach auf die Mitte. – Das ist ein bisschen ... Ja, warum geht man jetzt auf 15 %? Wenn man sagt, dass das typischerweise im Bereich 10 bis 20 % liegt, dann würde ich erwarten, dass das oberhalb von 10 % liegen muss, weil sonst ist es seltsam und müsste besonders begründet werden. Also, ich glaube, der Schritt auf die Mitte war etwas, was zumindest von einigen vielleicht als falsch, als eine zu große Vereinfachung angesehen wurde.

Jens Flintrop: Frau Thomas.

Stefanie Thomas: Ich würde gerne versuchen, das, was Beate Wieseler gerade schon unterstrichen hat in ihrer ersten Antwort, noch mal ein Stück weit auf den Punkt zu bringen, und noch mal die Frage unterstreichen: Was ist da eigentlich zu hoch und warum?

Tatsächlich sind wir uns einig: Mit unserem Vorschlag eines Responsekriteriums wollen wir nicht eine MID treffen, sondern wir wollen etwas treffen, was eben nicht – möglicherweise mit vielen Fragezeichen versehen – einen minimalen relevanten Unterschied abbildet, sondern wir wollen etwas festlegen, bei dem wir uns ausreichend sicher sind, dass das ein Responsekriterium ist, das sinnvoll und kein sehr großes ist.

Gerade hieß es: „Boah, das ist aber eine hohe Hürde“, und auch in den Stellungnahmen war beschrieben, dass das eine Schwierigkeit für den Effektnachweis ist. Da habe ich mich auch nach genau solchen Daten gefragt, weil wir sind uns doch hoffentlich einig, dass ein Effektnachweis nur dann nicht davon abhängig ist, dass man genau exakt eine MID gewählt hat für diesen Effektnachweis, sondern man kann ja plausibler Weise auch mal andere Responsekriterien wählen. Wir gehen doch hoffentlich davon aus, dass man auch bei diesen anderen Responsekriterien einen Effekt sehen würde. Ich glaube, wir wissen alle, es gibt für jede Studie, die man mit einem Effekt beobachtet, ein optimales Responsekriterium, und das liegt, glaube ich, bei der Hälfte des mittleren Unterschiedes von der Verteilung, wenn man das mal so überschlägt.

Das heißt, das optimale Responsekriterium, um möglichst viel Power zu haben, ist sowieso etwas ganz anderes. Wir reden hier darüber, dass wir ein plausibles, für alle transparent bekanntes, nachvollziehbares, a priori festlegbares Kriterium haben, von dem wir denken, dass es das abbildet, dass man sagen kann: Im Mittel ist das für Patienten spürbar.

Noch ein Wort zur MID – ich denke, die Literatur dazu zu kennen Sie alle –: Die gibt es ja nicht für jeden gleich. Das heißt, bei der MID sind wir schon in dem Problem: Wir haben die nicht exakt, wir werden die auch nie haben, und wir werden immer Verfahren haben.

Also noch einmal die Frage: Wo ist die hohe Hürde? Wo sind Ihre Bedenken zum Beispiel bei den 15 %? Bei der Einigkeit ist es vielleicht etwas größer, als es vielleicht einmal eine MID sein könnte für eine Skala.

Jens Flintrop: Herr Kohlmann, bitte.

Thomas Kohlmann: Thomas Kohlmann, Universitätsmedizin Greifswald. – Mein Anliegen ist relativ schnell erzählt: Wenn man es im Bereich von Studienplanung oder Datenanalyse oder Ex-post-Interpretation von Daten mit minimal bedeutsamen Unterschieden oder mit Schwellenwerten oder so was zu tun hat, dann ist man in einigen Fällen leider Gottes darauf angewiesen, Anhaltswerte zu verwenden, wie beispielsweise das, was das IQWiG vorgeschlagen hat. Und das war auch meine langjährige Praxis. Also, ich habe mich dann sehr gefreut, dass so was dann eben auch die Weihe des IQWiG bekommen hat.

Nur, in meiner bisherigen Praxis lagen diese Anhaltswerte immer niedriger. Also, so eine halbe Standardabweichung ist der Wert, bei dem man meistens ankommt. Wenn man unterschiedliche betrachtet – auch bei der großen Heterogenität der Ergebnisse; Frau Wieseler hatte darauf hingewiesen –, dann kommt man immer bei diesem nach Cohen mittleren Effekt heraus; manchmal ist es sogar ein bisschen weniger. Deswegen habe ich mich gefragt – und ich bin stutzig geworden –, wieso man 50 % über dieser Schwelle liegt. In dem Entwurf des Methodenpapiers ist es halt nicht so, dass eine systematische Übersicht gemacht worden ist, sondern es sind zwei Übersichtsarbeiten, zwei Reviews, einmal zu Rotatorenmanschettenrupturen, einmal zu Fatigue, dargestellt worden, und da habe ich halt einfach vorgeschlagen, man sollte es doch noch mal ein bisschen systematisieren. Ich behaupte nicht, dass ich mit meiner halben Standardabweichung genau auf den Punkt getroffen habe, obwohl ich relativ sicher bin, dass man da nicht so sehr daneben liegt. Das war eben mein Vorschlag.

Letzter Punkt, wenn Sie gestatten, zu Herrn Friede. Sie hatten ja Cohen erwähnt. Wir Verhaltenswissenschaftler und Psychologen, Sozialwissenschaftler dürfen ja sehr stolz sein, dass Jacob Cohen ein Buch geschrieben hat, „Statistical Power Analysis for the Behavioral Sciences“. Diese 0,2, 0,5, 0,8 Cohen's d sind weitläufig akzeptiert in den Neurowissenschaften bis hin zu den Hirnforschern, die mit Voxels und solchen Dingen arbeiten. Das ist eine gewissermaßen allgemein akzeptierte Größe geworden, und deswegen würde ich sagen, 0,5 Standardabweichungseinheit ist eine relativ gute Marge, und das wären 10 Prozentpunkte, vielleicht sogar ein bisschen weniger. Das ist mein Anliegen gewesen. – Danke schön.

Jens Flintrop: Herr Werner.

Sebastian Werner: Sebastian Werner, vfa. – Sie hatten jetzt wiederholt gefragt: Warum ist das zu hoch? – Wir haben uns das einmal angeschaut. Wir haben uns alle in den AMNOG-Nutzenbewertungen bisher als etabliert bzw. validiert akzeptierten MIDs angeschaut. Da kommt heraus, dass Sie mit Ihrem 15-%-Wert praktisch bei 90 % dieser MIDs gut über dem bisher validierten und als etabliert geltenden Wert liegen. Wir haben das auch ausgerechnet.

Das ist gut ein Faktor von 1,1. Das ist jetzt nicht so weit, aber bis 8,3 im Faktor nach oben. Also, das ist doch eine erhebliche Erhöhung, also deutlich konservativer anzusehen.

Was uns besonders Sorgen macht, ist, dass das die laut AMNOG-Bewertung durchaus hinreichend sichere Entscheidung darüber ist, ob das eine spürbare Veränderung war. Wenn man das jetzt sozusagen annehmen würde, dann würden übermorgen fast alle MIDs im AMNOG nicht mehr bestehen. Das halten wir für höchst bedenklich. Deswegen war eigentlich unser Vorschlag oder unser Wunsch, dass man sich in einer Kooperation in einer Arbeitsgruppe einmal genau anschaut, was hier mögliche Maßnahmen sein können. Wir denken auch, dass man an dieser Sache arbeiten sollte. Aber vielleicht ist eben ein starrer, ein generischer Schwellenwert nicht die Lösung. Deswegen noch mal der Wunsch ausgedrückt, dass man mögliche Kooperationen auf Plattformen von Fachgesellschaften macht, dass man noch einmal ein Brainstorming macht, was es möglicherweise auch an alternativen Möglichkeiten gäbe.

Jens Flintrop: Herr Andreas, jetzt dürfen Sie.

Jens-Otto Andreas: Jens-Otto Andreas, UCB. – Herr Werner hat schon eine ganze Menge gesagt, wo auch die Aussage von mir herkommt, dass es relativ hoch ist. Wir haben uns das angeschaut, und der Großteil der MIDs liegt wirklich darunter.

Aber um noch mal auch auf die Größenordnung oder die Frage: „Kann man was damit zeigen oder nicht?“ zurückzukommen. Darum geht es natürlich nicht, irgendwie zu sagen: Okay, mit 15 % können wir vielleicht schlechter einen Unterschied zeigen als mit 10 oder 12 %. – Es geht grundsätzlich um den festen Wert, weil das Ganze hängt ja auch wirklich von der Lage ab. Und wenn es wirklich darum geht: „Ich kann es vielleicht gerade so mit 12 % zeigen, mit 15 % nicht“, dann hat das natürlich eine Schwäche. Es geht wirklich mehr um die starre Festsetzung eines Wertes, zu dem Sie sagen: Okay, das ist irgendwie eine obere Grenze. Dann nehme ich die Patientenrelevanz scheinbar mit. – Aber das Ganze ist doch sehr individuell, und man sollte da auch die Untersuchungen und Arbeiten, die gemacht werden, bei denen mithilfe eines Ankers doch klar gezeigt wurde oder klar gezeigt werden konnte, dass gewisse Unterschiede, die auch niedriger sind, patientenrelevant sind, auch weiterhin berücksichtigen können.

Jens Flintrop: Als Nächstes habe ich Herrn Leverkus auf der Liste.

Friedhelm Leverkus: Leverkus, Pfizer. – Ausgangspunkt war ja, dass es einige MIDs gibt, bei denen Sie gesagt haben: Hm, das ist aber nicht so richtig abgeleitet. – Es gibt einige oder viele, zu denen es im Prinzip keine Kritik gab. Daher ist es wichtig, dass die auf alle Fälle erhalten bleiben und man nicht das gesamte Feld umwirft. Bei denen, zu denen jetzt im Prinzip Kritik geäußert ist, hat Jens-Otto gezeigt, wo diese Kritikpunkte im Prinzip waren, und das ist sicherlich auch lösbar. Und dieser Ansatzpunkt, dass man gemeinsam versucht, bei den Skalen, deren Herleitung ein bisschen weich ist, die Herleitung etwas härter zu machen und sich zu Kriterien zu äußern, erscheint uns sinnvoller zu sein, als jetzt alle mit einem Maß zu messen, was den Skalen im Prinzip möglicherweise auch nicht zustande kommt.

Es ist sicherlich schön, dass Sie ein systematisches Review gemacht haben, aber bei dem systematischen Review sind natürlich auch die dabei, die jetzt ein bisschen schlechter sind, die auch noch mal Variationen bringen. Und Sie haben im Prinzip noch mal eine Variation drin. Wenn man den Einzelfallansatz wählt, dann hat man für den Einzelfall im Prinzip eine Lösung geschaffen, die dafür auch nach den Kriterien, wenn man Kriterien festlegt, das besser trifft, als wenn ich einen Mittelwert nehme und diesen ansetze, wohl wissend, dass der ja auch nur eine Streuung ist, dass der eine Population ist. Daher möchte ich noch mal bekräftigen, dass man sich vielleicht noch mal zusammensetzt und versucht, gemeinsam Kriterien zu erarbeiten, die das Ganze auch zum Fliegen bringen.

Eine solche Gruppe, eine solche Initiative ist im Prinzip nicht neu für Deutschland. Wir haben das ja schon mal gemacht, als sich Industrie, IQWiG, BFAM, Universitäten zusammengesetzt haben und ein Paper über Estimands bei unerwünschten Ereignissen und unterschiedlichen Beobachtungszeiten. Also, wir haben schon Strukturen, wir können in den Strukturen auch denken.

Man kann das Ganze im Prinzip auch bei den Fachgesellschaften anhängen und auch die Academia mit einbeziehen.

Daher wäre das ein Weg, auf dem man wirklich gemeinsam die Probleme, die wir haben, ausräumen und jeder seinem Auftrag nachgehen könnte. – Danke.

Jens Flintrop: Vielen Dank. – Ich habe jetzt drei Wortmeldungen aus dem IQWiG, zunächst Frau Wieseler, dann Herr Windeler, dann Frau Thomas.

Beate Wieseler: Vielleicht zunächst mal zu dem Punkt des Vergleichs der bisher akzeptierten MID's zu diesem Vorschlag. Das sind teilweise Studien bzw. diese Responsewerte beruhen auf Studien, die den heutigen Kriterien nicht mehr standhalten würden. Das fängt schon an mit der MID für die EORTC, diese Arbeit von Osoba, 1998. Das war eine Arbeit, die den jetzt aktuell diskutierten Qualitätskriterien für solche Studien nicht mehr standhält. Ein anderes Beispiel ist SGRQ, eine Arbeit aus 2005. Die würde den heutigen Kriterien nicht mehr standhalten. Wir haben eine Weiterentwicklung in dem Bereich. Als diese Diskussion anging, waren wir sehr froh, dass überhaupt jemand mal den Patienten gefragt hat, ob er eine Veränderung spürt, und das nicht dem Urteil des Arztes überlassen hat.

Aus dieser Zeit sind wir heraus. Wir haben sehr viel gelernt über diese Art der Bewertung. Es ist ja vor etwa zwei Wochen diese Arbeit aus der McMaster-Gruppe im „The British Medical Journal“ erschienen, in der die erstmals ein Reliabilitätsinstrument für Studien zur Ermittlung der MID vorlegen. Das ist eine Arbeit, auf die wir lange gewartet haben. Wir sind mit der Gruppe seit 2015 immer mal wieder in der Diskussion gewesen. Die haben auch sehr systematisch gearbeitet. Wenn ich mir diese Kriterien anschau, die aus meiner Sicht auch sehr sinnvoll sind, dann treffen diese Arbeiten, auf denen diese MID's beruhen, die jetzt verglichen wurden mit unserem 15-%-Kriterium, die durchgängig nicht. Das heißt, wir müssen vielleicht

auch konstatieren, dass wir hier eine Entwicklung haben, dass wir besser werden müssen, was diese Kriterien angeht. Das machen wir in anderen Bereichen auch. Wir haben auch das Verfahren zur Meta-Analyse geändert. Ich glaube, wir müssen hier konstatieren, es gibt eine Weiterentwicklung, und da können wir nicht auf den Dingen beharren, die irgendwann Anfang der 2000er-Jahre an fünf bis acht Patienten ermittelt wurden.

Deshalb wäre das für mich erst mal kein primärer Grund, zu sagen, diese ganz alten MIDs liegen unterhalb der 15 %. Insbesondere dann finde ich das auch gar nicht entscheidungsleitend, wenn, wie Herr Andreas ja auch sagt, das auf die Bestimmung von Effekten eigentlich gar nicht primär den Einfluss hat. Das wäre ja sowieso eher der kritische Punkt, der zu diskutieren wäre, und da habe ich jetzt aber aus allen Beiträgen verstanden, dass wir da nicht erwarten, dass wir da, wenn wir eine Responseschwelle in diesem Bereich anpassen, auf einmal überhaupt keine Effekte mehr sehen können.

Ein weiterer Punkt, der in dieser Arbeit klar wird: Die haben das Instrument und auch die Reliabilität dieses Instruments untersucht und stellen fest, dass in den aktuell vorliegenden Studien zu MIDs ganz viele Informationen, die man für die Bewertung der Qualität dieser MIDs benötigen würde, nicht vorliegen. Zum Beispiel liegen Korrelationen zwischen Anker und Instrument in vielen Fällen nicht vor. Das deckt sich mit unserer Erfahrung. Das heißt, wir würden erwarten, dass wir, wenn wir diesen Weg weitergehen, dass wir eine hochqualitative MID wollen, mit den aktuell vorliegenden Studien da nicht zum Erfolg kommen. Auch deshalb verfolgen wir diesen eher pragmatischen Ansatz, der es uns ermöglicht, mit Responderanalysen zu arbeiten, auch in einer Situation, in der die Studien noch nicht die Qualität haben, die im Grunde genommen auf Basis der aktuellen Kenntnis gefordert wird.

Jens Flintrop: Herr Windeler.

Moderator Jürgen Windeler: Vielen Dank. – Ich möchte, auch im Anschluss an das, was Frau Wieseler gerade gesagt hat, noch mal zwei generelle Punkte sagen. Ich bin ein bisschen irritiert über die Rückmeldungen, man möge jetzt doch nicht so einen allgemeinen, pragmatischen Ansatz machen, sondern das müsse man ganz individuell betrachten. Also, normalerweise ist die Industrie dabei, auf Berechenbarkeit zu setzen, und dieses ist ein extrem berechenbares Vorgehen. Man weiß ganz genau, was man machen muss. Man weiß ganz genau, was wir fordern.

Wenn wir jetzt wirklich ein individuelles Verfahren auf der Basis von in gewisser Weise – und das wird beliebig schwierig, wie Frau Wieseler beschrieben hat – validierten MIDs gehen, dann ist das in gar keiner Weise mehr berechenbar, und ich würde auch – das mache ich in Erörterungen gerne mal – auf die Konsequenzen hinweisen. Denn das wird für Ihre AMNOG-Bewertungen und die Begründungen der Eignung von MIDs einen ganz erheblichen und möglicherweise nicht zu leistenden Aufwand bedeuten. Nur mal so als Hinweis.

Der andere Punkt: Wir sind für Zusammenarbeiten, Kooperation und AGs immer zu haben, aber wir reden hier jetzt über unser Methodenpapier und über eine Fassung, die wir veröffentlichen wollen, und über eine Fassung, zu der wir jetzt ein Stellungnahmeverfahren durchführen. Ich sage also in aller Klarheit: Wir werden uns für dieses Methodenpapier jetzt nicht erst in allgemeine Konsultationen mit welchen Playern auch immer begeben.

Jens Flintrop: Herr Werner, bitte.

Sebastian Werner: Frau Wieseler und Herr Windeler, wir wollen ja auch daran arbeiten, und wir sehen ja auch, dass sich da was tun muss. Wir sind nur unterschiedlicher Auffassung darüber, was jetzt hier der richtige Weg ist, ob man hier eine generische Schwelle macht, die möglicherweise pragmatisch ist und uns möglicherweise auch helfen könnte.

Auf der anderen Seite – und da sehen wir jetzt das Problem – haben Sie diese Kriterien offensichtlich noch nicht aufgeschrieben, nach denen man normalerweise eine Validierung von MIDs vornehmen und gut beurteilen könnte.

Wenn man das gemacht hätte und dann zu dem Urteil käme, dass das alles schwer zu bewerten ist, weil das in den Studien gar nicht rapportiert wurde, dann könnte man möglicherweise einen zweiten Schritt wagen und eine pragmatische Lösung vorschlagen. Aber ich habe das Gefühl, Sie probieren hier, den zweiten Schritt vor dem ersten zu machen. Und deswegen glaube ich, wir brauchen diese Kriterien. Man müsste sich wirklich einmal über die Kriterien Gedanken machen. Sie sagten gerade, vor zwei Wochen ist dazu eine Publikation herausgekommen. Na ja, wir sind da offenbar genau an der Spitze der Wissenschaft, und deswegen halte ich es auch möglicherweise für verfrüht, jetzt hier den Deckel draufzumachen, auch im Rahmen dieses Methodenpapiers. Also, Herr Windeler, warum drehen wir nicht noch mal eine Runde, in der wir das gerne noch mal zusammen diskutieren? Wir probieren, die Kriterien einmal zu fassen in Ihrem Methodenpapier, und dann gucken wir, ob wir eine pragmatische Lösung brauchen.

Im Übrigen ist die pragmatische Lösung, die Sie mit der 15%-Schwelle vorschlagen, auch nicht so einfach umsetzbar für die forschenden Pharmaunternehmen. Dann wird sozusagen eine neue Schwelle in Deutschland publiziert, die aber in internationalen Studien wahrscheinlich kaum untergebracht werden kann, weil die natürlich auf die validierten und etablierten MIDs abzielt. Also, ich bin mir nicht sicher, ob wir da wirklich das Ziel erreichen.

Jens Flintrop: Ich habe auf meiner Rednerliste jetzt Frau Kucka.

Annett Kucka: Annett Kucka von Bristol-Myers Squibb. – Ich habe da auch noch einen Kommentar, aber Herr Werner hat es eigentlich schon gut mit erörtert. Frau Wieseler, wenn Sie bereits Kriterien haben, an denen man Validierungsstudien sozusagen bewerten kann, wäre es natürlich gut, wenn man diese aufschreiben und gegebenenfalls diskutieren könnte.

Jens Flintrop: Als Nächstes Herr Sauerbruch, bitte.

Tilman Sauberbruch: Sauerbruch von der DGIM. – Das hier ist jetzt eine sehr statistische und methodische Diskussion; es ist ja auch ein Methodenpapier. Aber dennoch wäre ich interessiert – denn es geht um klinische Relevanz –, welche Kriterien oder Parameter der MID unterliegen. Sind das Patientenkriterien? Das habe ich nicht ganz verstanden. Können Sie das noch mal erläutern?

Jens Flintrop: Frau Wieseler.

Beate Wieseler: Ja, wir reden hier über Instrumente, mit denen patientenberichtete Endpunkte erhoben werden.

Tilman Sauberbruch: Nur patientenberichtete Endpunkte?

Beate Wieseler: Ja, komplexe Skalen, die so in ihrer Ausprägung ohne Weiteres nicht interpretierbar sind.

Tilman Sauberbruch: Danke.

Beate Wieseler: Gerne. – Da ich schon dran bin: Es gibt zwei Punkte. Also, wir sollen Kriterien aufschreiben, und dann sollen wir die Kriterien auch noch mal diskutieren. Es ist ja nicht so, dass wir bisher keine Kriterien angewendet haben. Es gab aber bis vor zwei Wochen keinen Vorschlag für ein Instrument zur Bewertung der Qualität dieser MID-Studien. Dieser Vorschlag liegt jetzt vor. Der ist nicht von uns, sondern von einer großen internationalen Arbeitsgruppe, die sich seit 30 Jahren mit dem Thema beschäftigt, aufgeschrieben worden. Die ist für jeden zugänglich, im „BMJ“ veröffentlicht. Deshalb ist das jetzt nicht die Lösung, dass wir diese Kriterien in unserem Methodenpapier aufschreiben.

Wenn wir auf die Kriterien schauen und uns die Studien vor Augen führen, die wir in der Regel zum Beispiel in den Dossiers zur Nutzenbewertung sehen, dann erfüllen diese Studien in der Regel diese Kriterien nicht. Das ist auch in der Untersuchung der Reliabilität dieses Dokuments sichtbar geworden, dass von den 200 Studien, die sie sich angeschaut haben, nur ein sehr geringer Teil Informationen zu allen Kriterien geliefert hat. Das heißt, wir können jetzt natürlich diese Kriterien aufschreiben und dann noch zehn Jahre warten, bis wir diese Daten haben. Wir müssen aber jedes Jahr 80 Dossierbewertungen machen. Da ist natürlich die Frage: Was machen wir in der Zwischenzeit? Lehnen wir alles ab? – Deshalb haben wir jetzt hier einen pragmatischen Vorschlag gemacht, weil wir diese Responderanalysen grundsätzlich für einen sinnvollen Weg halten, um diese Daten interpretierbar zu machen.

Unsere Frage ist: Was ist Ihr Vorschlag jetzt konkret in der Zeit, bis diese Studien diese Qualität haben, dass wir vielleicht wieder darüber nachdenken könnten, mit MIDs zu arbeiten? Was ist Ihr Vorschlag für diese Zwischenzeit, um mit diesen Daten in dem Verfahren realistisch arbeiten zu können, ohne dass Sie für das Dossier, für jede Bewertung erst einmal ein systematisches Review machen und die ganzen MID-Vorschläge identifizieren, dann in die Validierung dieser Studien gehen, um dann im Dossier – denn das ist Ihre Aufgabe –

darzulegen, dass das eine valide MID ist? Was ist Ihr Vorschlag für die Zwischenzeit, bis wir die Studien haben, die diese Kriterien treffen?

Jens Flintrop: Herr Werner, direkt dazu?

Werner: Ich kann leider nichts zu den Kriterien sagen. Ich meine, wir haben in der Arbeitsgruppe zum Beispiel Längsschnitt, ankerbasierte, hinreichende Korrelationen, mindestens zwei Messwerte diskutiert. Und jetzt noch mal: Wenn man diese Kriterien aber nicht in das Methodenpapier gießt, wie glauben Sie denn, dass Sie dann eine Bewertung machen können?

Beate Wieseler: Unser Vorschlag ist im Moment, dass wir nicht auf die MID gehen, sondern auf einen Wert, der empirisch ermittelt etwas über der MID liegt. Wir würden einschätzen, dass das hinreichend sicher eine spürbare Veränderung abbildet. Und wir würden mit diesem Wert arbeiten, ohne dass wir die Qualität dieser MID-Studien bewerten

Wir bieten Ihnen an, dass Sie auch in Studien, in denen das nicht präspezifiziert vorliegt, eine sehr definierte Post-hoc-Analyse vorlegen. Das ist notwendig, damit wir nicht in eine ergebnisgesteuerte Berichterstattung kommen. Also, wir eröffnen eine Möglichkeit, mit Responderanalysen zu arbeiten, aber da müssen wir natürlich eine ergebnisgesteuerte Berichterstattung ausschließen.

Jens Flintrop: Stefanie Thomas hat sich gerade gemeldet.

Stefanie Thomas: Ich denke, Frau Wieseler hat jetzt alles noch mal deutlich gemacht. Das reicht. Ich ziehe meine Wortmeldung zurück.

Jens Flintrop: Dann Herr Dintsios, bitte.

Charalabos-Markos Dintsios: Ich richte die Frage an Beate Wieseler. Wenn es darum geht, sozusagen einen Sicherheitsabstand zu den nach Auffassung des IQWiG heute nicht mehr standhaltenden Kriterien der Validierungsstudie aus Anno Domini – so war es ja auch 1998 beim EORTC QLQ-C30, wenn ich mich richtig erinnere – – wieso dann nicht der Versuch eines Sicherheitsabstands als Aufschlag auf die vorhandenen MIDs zu generieren, falls möglich? Und wieso dieses quasi inverse Gießkannenprinzip, das ja ein Mittelwert ist zwischen 10 und 20 %? Zumindest im Methodenvorschlag wurden ja zwei Referenzen genannt. Die beziehen sich auf spezifische Indikationsstellungen und spezifische Symptomatiken, Fatigue. Ich meine, man sollte nicht einen Mangel, der da ist, durch einen anderen Mangel ersetzen. Wieso geht man nicht den Weg zusammen und versucht vielleicht – – Ich akzeptiere ja das Argument, dass einiges überholt ist. Da sind wir alle hier, auch die Industrie, nicht anderer Auffassung. Aber wieso diesen Generalcharakter?

Und zu Herrn Windelers Kommentar wegen der Planungssicherheit. Man löst hier nicht planungstechnisch ein Problem, indem man exogen quasi eine numerische Schwelle setzt, die keine MID ist, sondern sie ist, so wie ich sie richtig interpretiere, eine Sicherheitsschwelle. Wir

haben laufende Studien, wir haben unterschiedliche Instrumente, unterschiedliche Eigenschaften der Instrumente. Auch bei komplexen Skalen sind die Eigenschaften untereinander durchaus unterschiedlich. Je nach Instrument gibt es andere Wertungen, andere Gewichtungungen. Ich glaube, es wäre hier besser, doch mal in die Details zu gehen und vielleicht diesen Ansatz zu verfolgen. Den würde ich nicht mal ablehnen wollen, aber auf eine MID sozusagen individuell betrachtet vielleicht einen Aufschlag versuchen. Das wäre auch nicht schlecht. Dann wären wir bei Herrn Kohlmann, der ja Effekte gesehen hat, die weit unter 10 % lagen und nicht immer bei 15 %. Vielleicht gibt es auch Sachen, die bei 30 % liegen. Ich will es ja nicht in Abrede stellen.

Jens Flintrop: Herr Werner, Sie hatten gerade noch gezuckt.

Sebastian Werner: Uns würde auch zur Planungssicherheit helfen, wenn wir die Kriterien genau kennen würden, wonach wir eine MID sinnvoll als validiert einschätzen können, ja oder nein.

Uns ist im Vergleich zur AMNOG-Bewertungspraxis auch aufgefallen, dass in bestimmten AMNOG-Nutzenbewertungen, die Sie verfasst haben, die Darlegung bezüglich der Kriterien weitaus detaillierter ist als das, was jetzt sozusagen im Methodenpapier steht. Ich glaube, dass es unbedingt erforderlich ist, den derzeitigen Stand der wissenschaftlichen Erkenntnisse diesbezüglich auch in Ihrem Methodenpapier zu fassen, damit man einen zweiten Schritt gehen und sagen kann, wie Sie jetzt sagen, dass alle MIDs nicht mehr als validiert gelten können. Das sehen wir derzeit noch nicht, und wir können es auch nicht sehen. Wir können es auch nicht als solche bewerten, weil Sie das nicht in Ihrem Methodenpapier hineinschreiben. Sie gehen den zweiten Schritt vor dem ersten.

Moderator Jürgen Windeler: Also, ich muss gestehen, dass ich einige Punkte schwer verstehen kann. Unabhängig davon, was die Guyatt-Gruppe vor 14 Tagen publiziert hat, ist eigentlich seit Jahren klar, dass die Validierung von MIDs über die Anker große Probleme macht. Das ist seit Jahren klar, was den aktuellen wissenschaftlichen Stand angeht. Das kann man wissen, um das mal auf den Punkt zu bringen.

Wir können gerne – und das ist ja auch der Punkt, den ich eben schon angesprochen habe und den auch Frau Wieseler noch mal angesprochen hat – den aktuellen Stand der wissenschaftlichen Erkenntnisse, insbesondere das Guyatt-Papier, noch mal ergänzend darstellen, um zu sagen, dass das die Grundlage dafür ist, wie man heute MIDs validieren sollte oder ermitteln sollte. Dann gibt es die Option, zu sagen: Das machen wir jetzt in Dossiers bitte immer so. – Oder es gibt die Möglichkeit, zu sagen: Na ja, das machen wir jetzt mal die nächsten fünf oder zehn Jahre nicht genau so, weil wir wissen, dass das nicht geht, und weil wir die Auflagen, die Sie machen, und der Prüfaufwand, den wir machen müssen, und das Ergebnis dieses Prozesses völlig klar sind.

Vielmehr bieten wir Ihnen an eine allgemeine und – ich sage es noch mal – berechenbare Lösung an. Selbstverständlich ist die Festsetzung von Dingen berechenbar, um sich dann

danach zu orientieren. Wir bieten Ihnen ergänzend an, das sogar post hoc zu machen. Insofern fällt das Argument, dass es in internationalen Studien noch nicht so angewendet worden ist, auch weg. Das ist unser Angebot, das empirisch gestützt ist, das unseres Erachtens sinnvoll ist und das – ich sage es mal etwas vereinfacht – uns allen das Leben erleichtert und trotzdem für die Patienten relevante Bewertungsergebnisse produzieren kann, sodass man noch einmal darstellt, was die Guyatt-Gruppe beschrieben hat und was die eigentlichen Anforderungen wären, um dann zu sagen: Na ja, lass uns mal die Kirche im Dorf lassen. – Das halte ich für einen eigentlich sinnvollen Vorschlag.

Wir haben jetzt aber Ihre Argumente gehört. Wie werden sie selbstverständlich noch mal erwägen. Wir werden uns vielleicht noch mal insbesondere einer unterstützenden – Herr Werner hat das angesprochen – Darlegung des derzeitigen Erkenntnisstandes, was die Defizite und speziell die Anforderungen der Guyatt-Gruppe angeht, widmen, und dann werden wir gucken, wie wir mit dem Ergebnis umgehen.

Gibt es dazu Wortmeldungen? – Bitte.

Egon Pfarr: Pfarr, Boehringer Ingelheim. – Wenn Sie sich schon ein einfaches Kriterium überlegen, was prinzipiell durchaus zu begrüßen ist, haben Sie sich Gedanken gemacht, ob man dieses einfache Kriterium nicht doch noch ein bisschen an die Datenlage anpassen kann? Also, was passiert mit Patienten, die schon relativ gute Ausgangswerte haben? Da kann ich eine Response von 15 % der empirischen Spannweite zum Teil gar nicht erreichen. Das Gleiche gilt, wenn ich nach unten gucke. Haben Sie sich Gedanken gemacht, ob man sich nicht auch mal die Verteilung der Population, also die Werte der Population zur Baseline, anguckt, ob das Patienten sind, die in der konkreten Studie überhaupt auch nur annähernd an diesen empirischen Bereich heranreichen? Oder ist es nur eine Teilpopulation oder eine Population, die 50 % der Spannweite überhaupt erst zur Baseline hat, sodass man auch hier, zumindest wenn man ein grobes Kriterium ableitet, noch ein bisschen weiter differenziert?

Das gilt auch für die Frage, ob man immer nur in eine Richtung blicken kann, also nur Response, nur positive Response, nur Verschlechterung, oder ob man sich letztendlich nicht auch, was natürlich vom Punktschätzer nicht direkt zu interpretieren ist, die drei Kriterien „Responder“, also Verbesserung, „Verschlechterung“ und „weitgehend unverändert“, gemeinsam anguckt. Denn hier könnte man auch eine Signifikanz sehen, die in der Einzelbetrachtung – nur positive Response, nur negative Response – nicht zum Tragen kommt. Das ist also die Frage: Haben Sie da noch ein bisschen mehr an Gedanken investiert?

Jens Flintrop: Frau Wieseler.

Beate Wieseler: Ja, wir haben alle möglichen Gedanken investiert. Sie haben jetzt den Baseline-Schweregrad und die Richtung der Veränderung als zwei Punkte genannt. Es gibt zig andere Punkte, die dann zu berücksichtigen wären. Das ist ja genau das Problem, dass die MID selbst

von so vielen Faktoren potenziell abhängig ist. Das ist ein Teil des Problems, den wir hier zu lösen versuchen.

Was generell andere Auswertungsmöglichkeiten neben der Responderanalyse angeht, besteht natürlich nach wie vor die Möglichkeit, Mittelwertsauswertungen zu machen. Dann legen wir ein anderes Response-, ein anderes Relevanzkriterium an. Das ist in unserem Methodenpapier auch beschrieben.

Ich würde Ihnen auch recht geben, dass es sich sicherlich lohnt, weiter darüber nachzudenken, welche anderen Arten von Auswertungen für diesen Bereich sinnvoll sind. Da gibt es auch internationale Arbeitsgruppen; das wissen Sie. Die SISAQOL-Gruppe hat die Arbeit gerade beendet. Es wird ein Nachfolgeprojekt geben, an dem wir uns auch beteiligen. Dieses Feld ist aktuell extrem in Bewegung, und wir beobachten das natürlich und würden das dann immer wieder berücksichtigen. Aber unter den Daten, die wir jetzt aktuell sehen, decken wir eigentlich den größten Teil dessen, was wir sehen, mit Responderanalysen und Mittelwertsbetrachtungen ab.

Egon Pfarr: Vielleicht darf ich zwei Kommentare nachlegen. – Erstens. Selbst wenn in der Vergangenheit vielleicht die Methode oder die Studien, auf die eine Methode angewandt wurde, nicht den heutigen Kriterien entspricht, heißt es nicht, dass man ausschließen kann, dass sich diese abgeleiteten MIDs – oder wie immer man die jetzt bezeichnet – nicht auch in der Praxis bewährt haben. Es gibt durchaus MIDs, die sich aus medizinischer Sicht wahrscheinlich sehr wohl etabliert haben. Aber auch diese sollten in Zukunft durch diese 15%-Regelung außer Kraft gesetzt werden, wenn ich das richtig verstanden habe. Dann stellt sich zum Beispiel auch die Frage, ob man nicht da noch mal die Diskussion mit den entsprechenden Gesellschaften führen sollte, ob sich nicht auch empirisch bewährte Kriterien dann im AMNOG-Verfahren niederschlagen.

Zweitens kurz zu Herrn Windeler und „post hoc“. Ich habe es nicht ganz verstanden, was uns das bringt, wenn AMNOG oder Ihr Methodenpapier jetzt sagt, post hoc ist zulässig. Ich meine, wenn Sie etwas definieren, wo ist der Vorteil, wenn man das als Post-hoc-Analyse definieren darf? Es ist ja eh klar, dass das post hoc ist. Ich sehe jetzt nicht den Vorteil durch dieses Label „post hoc“.

Moderator Jürgen Windeler: Das Argument war vorhin, dass man das wegen der internationalen Studien, wenn wir jetzt sozusagen mit der 15%-Schwelle kommen, gar nicht mehr machen kann, weil das nicht geplant ist, weil andere das nicht machen usw. Ja, wir legen nicht fest, dass das quasi im Prüfplan schon vor zwei Jahren präspezifiziert gewesen sein muss, sondern wir lassen post hoc zu, dass solche Auswertungen eben auch mit den vorhandenen Daten gemacht werden können. Das wollte ich damit sagen.

Jens Flintrop: Mir liegen jetzt noch Wortmeldungen von Herrn Lützelberger, Herrn Michnacs und Herrn Kohlmann vor. Danach schließe ich die Rednerliste.

Uwe Lützelberger: Lützelberger, AstraZeneca. – Die Diskussion ist total spannend, auch vom Methodischen her. Aber abgesehen vom methodischen Aspekt, der jetzt hier diskutiert wird, möchte ich einfach noch mal ein Argument erwähnen, das für uns im Hintergrund auch so ein bisschen bei der Diskussion mitschwingt. Das ist diese mögliche Abkopplung von den Zulassungsanforderungen. Wir wissen, dass sozusagen auch von Zulassungsbehörden dieses Thema „Patient-reported outcomes“ intensiv diskutiert wird, dass da viele Veranstaltungen stattfinden, Paper publiziert werden. Es ist rein theoretisch auch denkbar, dass mal eine Studie da ist, in der eine solch komplexe Skala dann koprimärer Endpunkt oder key secondary Endpunkt ist, also eine wichtige Grundlage vielleicht auch für die Zulassungsentscheidung von dem betreffenden Arzneimittel war. Und wenn dieses Arzneimittel dann in diese Bewertung beim IQWiG kommt und dann aufgrund dieses normativ festgesetzten Kriteriums dann vielleicht ein anderes Ergebnis herauskommt, dann steht man vor der Situation: Es wurde zugelassen, weil es bessere Lebensqualität bringen soll. IQWiG sagt, dass es keine bessere Lebensqualität bringt. – Das ist natürlich ein anderes Problem, mit dem man dann auch noch zu tun bekäme. Das wollte ich einfach noch mal ansprechen.

Sie sagen jetzt natürlich, Sie sehen keine Hinweise darauf, dass die Effekte anders wären. Gut, das kann ich jetzt auch nicht beurteilen ohne weitere Daten. Aber es ist theoretisch möglich, dass es zu solchen Situationen kommt, wenn man so ein universelles „One size fits all“-Kriterium dann immer anlegen will.

Jens Flintrop: Herr Michnacs.

Kristian Michnacs: Kristian Michnacs für den BVMed. – Zunächst muss ich sagen, dass die Patientenrelevanz in der klinischen Forschung wahrscheinlich einer der wichtigsten Aspekte ist und dass die Beurteilung dessen durch die Tatsache, dass das IQWiG das mit einem Instrument aufgreift, um es besser darzustellen, erst einmal grundsätzlich zu begrüßen ist.

Wir haben es in unserer Stellungnahme auch nicht explizit erwähnt, aber tatsächlich ist Ihr pragmatisch gewähltes Ergebnis der 15 % für uns auch nicht ganz ableitbar gewesen. Ich muss aber auch ergänzen, dass ein pragmatischer Ansatz in dem Moment vielleicht auch gar nicht der dümmste ist, weil eine eindeutige Festlegung vielleicht so auch gerade gar nicht geht.

Was mich bei der pragmatischen Herangehensweise Ihrerseits interessieren würde, ist die Frage, ob Sie vielleicht unterschiedliche, pragmatisch wie auch immer ausgesuchte Ansätze mal bei Studien simuliert haben, bei denen Sie sich vielleicht überlegt haben: Ja, hätten wir so einen Ansatz vorher schon gehabt, hätte uns das vielleicht weitergeholfen. – Und wenn ja, was kam da bei der Simulation heraus?

Sofern sich das IQWiG vielleicht doch dazu entscheidet, das noch mal irgendwie dem Kuratorium zuzuleiten und im Rahmen einer Beratung zu diskutieren, bekunden natürlich auch wir unser Interesse an der Teilnahme.

Jens Flintrop: Frau Wieseler, direkt dazu?

Beate Wieseler: Wir können das nicht simulieren, so wie Sie das nennen. Dafür bräuchten wir individuelle Patientendaten aus den Studien, um dann zu errechnen, welche Effekte sich mit unterschiedlichen Responsekriterien ergeben würden. Das wäre etwas, was die Unternehmen selbst machen müssten. Das in einer gewissen Systematik mal vorzulegen, wäre sicherlich interessant.

Jens Flintrop: Herr Kohlmann.

Thomas Kohlmann: Zum Zeitpunkt der Abfassung des Entwurfs des Methodenpapiers war diese Studie von Guyatt und anderen, die jetzt publiziert worden ist, noch gar nicht bekannt; sie konnte also gar nicht einfließen. Deswegen wäre mein Vorschlag, noch mal zu überlegen, ob man dieses Instrument, das dort vorgeschlagen wurde, insofern mit einbezieht, als man sagt, diese 15-%-Schwelle ist für all diejenigen Studien, bei denen es unklar ist, ob Response-schwellen in der MID belastbar sind, die Maßgabe. Wenn aber gezeigt werden kann, dass nach der Anwendung dieses neuen Beurteilungsinstruments eben belastbare MIDs oder Response-schwellen vorliegen, dann können diese verwendet werden. Das würde dem so gewissermaßen entgegenkommen, dass man sagt: Wenn gute Evidenz da ist, die zeigt, dass die 15 % vielleicht ein bisschen zu hoch gewählt sind, dann kann man auch heruntergehen.

Moderator Jürgen Windeler: Danke für diese Anregung. – Ralf.

Ralf Bender: Ich wollte noch kurz auf Herrn Lützelberger antworten. Ganz allgemein: Zulassung ist eben nicht gleich Nutzenbewertung, und die Tatsache, dass etwas zugelassen ist, bedeutet nicht automatisch, dass ein Zusatznutzen da ist. Das gilt ganz allgemein.

Moderator Jürgen Windeler: Das kam jetzt nicht so überraschend.

(Heiterkeit)

Bitte, Herr Lützelberger.

Uwe Lützelberger: Das Argument verstehe ich natürlich, es kommt ja auch immer wieder. Es gibt aber doch häufig die Situation, in der der Zulassungsentscheidung genau dieselbe Studie zugrunde liegt wie der HTA-Bewertung, zum Beispiel bei Placebo-Vergleich oder Best-Supportive-Care-Vergleich. Und dann ist es dasselbe, was da bewertet wird.

Ralf Bender: Nein, ist es nicht.

Moderator Jürgen Windeler: Jetzt dürfen Sie noch mal „Doch“ sagen. Ansonsten schlage ich vor, dass wir mit den Anregungen, den Hinweisen, den Kommentaren, die wir jetzt bekommen haben, diesen Tagesordnungspunkt schließen.

Wir werden noch mal die Argumente bewegen und diskutieren und dann gucken, was wir dann für das Methodenpapier, so wie Frau Wieseler das beschrieben hat, für die nächsten zehn oder

jedenfalls für die nächsten fünf Jahre daraus machen. Wenn sich in zwei Jahren bei der nächsten Revision ergeben sollte, dass wir da etwas ändern müssen, dann werden wir das ändern.

Damit komme ich zu:

3.4.4 Tagesordnungspunkt 3: Festlegung von und Umgang mit Publikationsbias (Abschnitt 3.2.1)

Ich übergebe an Stefan Sauerland.

Stefan Sauerland: Im Folgenden wird es um Publikationsbias gehen. Das ist das Kapitel 3.2.1 im Methodenpapier, „Auswirkung nicht publizierter Studienergebnisse auf das Fazit“. Wir haben relativ viele Stellungnahmen erhalten, die darum baten, bestimmte Aspekte noch mal ein bisschen präziser auszuformulieren. Ich kann die gerne, falls da noch Missverständnisse sind, gleich auflösen. Uns interessieren hier in dem Feld aber jetzt zwei Punkte, die wir gerne mit Ihnen diskutieren würden, und das sind beides Fragen, die sich sowohl an Industrie als auch an die akademischen Kollegen richten.

Die eine Frage ist eine Frage, die Sie uns ebenfalls gestellt haben: Wo ist denn überhaupt eine sinnvolle Abgrenzung zwischen einem vielleicht irrelevanten Anteil von fehlenden Daten und einem relevanten Anteil von fehlenden Daten möglich? – Wir hatten in unsere Tabelle die Begriffe „mäßiger Fehldatenanteil“ und „erheblicher Fehldatenanteil“ eingeführt. Das war vorher, in der früheren Tabelle, so nicht vorhanden. Wir hatten im letzten Jahr – das haben einige von Ihnen bemerkt, andere aber auch nicht – in einem Bericht zur Vakuumwundversiegelung für diese beiden Kategorien auch Grenzwerte angesetzt. Das waren 10 % Fehldatenanteil für einen mäßigen Umfang fehlender Daten, und wir hatten eine Grenze von 30 % angesetzt für einen erheblichen Fehldatenanteil. Das sind aus unserer Sicht als Erstes aber nur Pi-mal-Daumen-Vorgaben, sodass wir die nicht ins Methodenpapier hineingenommen haben. Da würde uns aber interessieren, ob diese Zahlen aus Ihrer Sicht als grobe Abschätzung erst einmal geeignet erscheinen.

Die zweite Frage, die wir an Sie hätten, wäre: Wie kann man es denn schaffen, in den Bewertungen den Anteil von fehlenden Daten – und da geht es jetzt vor allen Dingen um komplett fehlende Studienergebnisse – zu reduzieren? – Das sind natürlich Dinge, die sich weniger auf die Dossierbewertung beziehen, sondern vielmehr auf die konventionelle Nutzenbewertung, also wo wir Studien von beispielsweise Medizinprodukteherstellern brauchen, wo wir Studien aber auch von Universitäten oder sonstigen Forschungseinrichtungen einfließen lassen wollen. Da wäre unsere eher organisatorische Frage, wie man es gerade in den heutigen Zeiten, wo die Bewertung immer schneller werden soll, schaffen kann, rechtzeitig und frühzeitig Ihre Studienergebnisse auch wirklich komplett vorliegen zu haben.

Moderator Jürgen Windeler: Wer möchte? – Bitte.

Kristian Michnacs: Kristian Michnacs für den BVMed. – Wir haben das Thema auch in unserer Stellungnahme aufgenommen, und als wir die für diesen Punkt bearbeitet haben, ist tatsächlich die Unbestimmtheit der Begrifflichkeiten hochgekommen. Als Beispiel: Für mich erklärt sich immer noch nicht, was der Unterschied zwischen „mäßig möglich“ und „erheblich wahrscheinlich“ sein soll. Hätten Sie Zahlen dieser Art angegeben, wie Sie es eben kurz erörtert haben, hätte sich zumindest schon mal diese Fragestellung etwas aufgelöst. Ob man mit der Abgrenzung, die Sie mit den beiden Werten vornehmen, einverstanden wäre? Das müsste man noch einmal diskutieren. Aber diese Frage hätte sich durchaus anders gestellt.

In der jetzigen Fassung Ihres Methodenpapiers beantwortet sich dieser Widerspruch der Unbestimmtheit der Begriffe eben nicht, weil diese Zahlen oder Eckpunkte, an denen man sich orientieren kann, nicht aufgenommen sind. Auf Ihren Hinweis, für ein pragmatisch oder gemeinsam vereinbartes Vorgehen eventuell weitere Daten zu gewinnen, sofern ein Publikationsbias vollständiger Studien vorliegt, haben wir auch einen Hinweis gegeben in unserer Stellungnahme, und zwar haben wir in Ihrer Fußnotenerörterung noch eine Fußnote C eingefügt, die eine weitere Rückfrage an Medizinprodukte oder an die Hersteller von Produkten – das können Arzneimittel sein, Medizinprodukte etc. – beinhaltet. Okay, man könnte jetzt sagen, dass wäre eine Rolle rückwärts und man würde das Verfahren neu beginnen, aber wenn das in dem Moment wieder als erheblich auffällt, ist das zumindest ein gangbarer Weg. Denn es kann durchaus sein, dass es Studien gibt, die zum Zeitpunkt, als Dossiers erarbeitet wurden, noch liefen, dann keine Veröffentlichung stattfinden konnte und es da zu zeitlichen Überschneidungen kam. In solchen Fällen wäre das ein pragmatischer Vorgang.

Stefan Sauerland: Erst mal danke für das Angebot, da Daten liefern zu wollen. – Sie hatten jetzt als Letztes Bezug genommen auf die laufenden Studien. Uns geht es ja darum, überhaupt den Studienstatus erst einmal feststellen zu können. Wenn eine Studie läuft, dann ist das ja völlig in Ordnung.

Und es ist uns natürlich auch bekannt, dass es häufig passiert, dass Studien dann doch länger laufen, als es eigentlich geplant war, und dass im Studienregister ein Datum steht, das initial mal optimistischer Weise festgelegt wurde, sich aber so nicht halten lässt. Wichtig ist uns, dass dann im Studienregister oder im Kontakt mit den Herstellern auch wirklich klar ist, dass diese Studie eben länger läuft als eigentlich geplant. Sie wissen alle, dass es seit etwa 2005 die Vorgabe gibt, Studien zu registrieren. Es ist aber auch wichtig, diese Studienregistereinträge zu aktualisieren, damit wir erkennen können, dass nicht irgendwelche Studienergebnisse seit ein, zwei, drei Jahren eigentlich fertig sind, aber nicht publiziert werden, weil einem die Ergebnisse eben nicht gefallen. Darum geht es uns.

Kristian Michnacs: Ich will nur kurz ergänzen, was Sie zuletzt gesagt haben: Dann haben die Ergebnisse vielleicht nicht gefallen. – Ich glaube, keiner von uns ist hier im Raum und macht Cherry Picking bei den Ergebnissen, und es schickt sich auch nicht; das ist ja klar.

Also, es ist durchaus kontrovers, dass die Aktualisierung der eigenen Dokumentation völlig unabhängig davon, wer nun gerade verantwortlich dafür ist, immer rechtzeitig in den Studienregistern erfolgt; ohne Frage. Es kann aber auch beispielsweise zu Situationen kommen, in denen Studien zu einem Zeitpunkt initiiert werden, in dem Sie beispielsweise in der Bewertung sind und, als das alles fertiggestellt wurde, dieses Protokoll in den Registern noch gar nicht einsehbar war. Das wäre auch möglich, und diese Situation würden Sie zumindest durch die Fußnote, die wir Ihnen in der Formulierung angeboten haben, auch auflösen können. Ich kann jetzt nicht sagen, wie oft sie vorkommen würde oder wie oft sie schon vorgekommen ist – das weiß ich nicht –, aber möglich ist das.

Moderator Jürgen Windeler: Vielleicht zum letzten Punkt noch eine Frage und dann zum ersten noch mal eine Rückfrage.

Zum letzten Punkt. Herr Sauerland hat das gerade angesprochen, angedeutet: Unsere Möglichkeiten, im Rahmen der normalen Nutzenbewertungen in irgendeiner Weise mit Stakeholdern, Studiendurchführenden in Kontakt zu treten und sie um Daten zu bitten, werden zunehmend beschränkt, und zwar massiv beschränkt. Wir haben schlicht keine Zeit mehr dafür. Und deswegen will ich auch in aller Deutlichkeit sagen: Unsere Möglichkeiten – nicht so sehr unser Interesse, sondern unsere Möglichkeiten –, an Hersteller heranzutreten, wenn wir irgendwo Probleme sehen, Daten fehlen, Studieninformationen fehlen, werden zum Beispiel durch Rechtsverordnungen massiv beschnitten.

Insofern, glaube ich, möchte ich noch mal in aller Deutlichkeit sagen: Es ist im Interesse der Stakeholder, studiendurchführenden Hersteller, Informationen entweder in den Registern aktuell zu halten oder uns zeitnah zur Verfügung zu stellen, und zwar von sich aus. Die Beratungsthemen im G-BA sind ja alle bekannt, unsere Beauftragungen auch. Unsere Möglichkeiten, nachzufragen und nachzuforschen, werden sehr begrenzt werden, und insofern sind wir dann darauf angewiesen, Informationen mehr oder weniger ungefragt zu bekommen und uns nicht lange auf die Suche außerhalb von üblichen Quellen zu begeben. Das möchte ich noch mal zum Ausdruck bringen.

Insofern kann man Nachfragemöglichkeiten als Optionen aufschreiben, aber wir werden die entweder gar nicht oder – das ist der andere Punkt – in einer Weise nutzen können, dass wir sagen müssen: Wir haben genau eine Woche Zeit, um eine Rückmeldung zu geben, und zwar suffizient zu geben. Mehr Zeit steht nicht zur Verfügung. – Das wird für uns und, ich glaube, für einige andere auch noch eine schwierige Situation werden.

Der andere Punkt: Ich habe noch einmal eine Frage zu den Zahlen. Also, habe ich Sie jetzt so verstanden, dass Sie da Zahlen drinstehen haben möchten? Wollen Sie da 10 und 30 % stehen haben?

Kristian Michnacs: Also, ich habe gesagt, dass die Widersprüchlichkeit in der Unbestimmtheit der Begriffe durch Zahlen aufgelöst werden könnte. Ich habe aber auch erwähnt, dass ich zum

jetzigen Zeitpunkt nicht sagen kann, ob wir mit 10 und 30 % einverstanden wären. Dass solche Zahlen eine Orientierung anbieten oder diese Unbestimmtheit etwas auflösen, ist unbestritten. Aber wie die sich differenziert, ist eine andere Frage.

Stefan Sauerland: Das nehme ich erst mal so mit. Vielen Dank. Ich glaube auch, dass diese Zahlen als Orientierung hilfreich sind, und vielleicht werden wir einfach alle mal ein wenig abwarten und Erfahrungen damit sammeln.

Der andere Punkt, den ich vielleicht noch nennen wollte, ist, dass es zwischen dem IQWiG und dem BVMed und anderen Medizinprodukteherstellerverbänden Gespräche in den letzten Jahren gab, die darauf abzielten, Musterverträge zu entwickeln, damit eben nicht erst über die Übermittlung von Studiendaten ein Vertrag langwierig ausgehandelt werden muss, sondern damit diese Verträge von den Unternehmen dann relativ rasch akzeptiert werden können und dann auch rasch erfüllt werden können, sodass diese Studiendaten, die wir dann benötigen, auch wirklich innerhalb kurzer Zeit zur Verfügung gestellt werden können.

Moderator Jürgen Windeler: Da sehe ich eine Wortmeldung. Bitte.

Annett Kucka: Annett Kucka von Bristol-Myers Squibb. – Ich habe eine Rückfrage dazu. Für die Arzneimittelbewertung betrifft die Konkretisierung dieser Studien sozusagen die Studien, die in der systematischen Literaturrecherche gefunden wurden. Wie sieht das bei indirekten Vergleichen aus, wenn auf Basis von Publikationen nicht alle Endpunkte zur Verfügung stehen?

Stefan Sauerland: Das ist eine komplizierte Frage. Da gebe ich gerne ab an die Arzneimittelkollegin.

Beate Wieseler: Im Bereich Arzneimittel ist die Übermittlung von Studiendaten relativ gut etabliert. Wir gehen so vor, dass wir im Rahmen der Informationsbeschaffung von Anfang an vollständige Studienübersichten von den Herstellern anfordern und für die Studien, die von Herstellern gesponsert sind, gar nicht auf die Studienpublikationen in Journals zurückgreifen, sondern auf die vollständigen Studienberichte. Deshalb würden wir da dann die vollständige Information haben. Das ist sicherlich auch unter dem Aspekt, den Herr Windeler angesprochen hat, ein beschleunigender Faktor, dass wir nicht erst irgendwelche Publikationen lesen, herausfinden, was da vielleicht fehlt, und dann in eine Informationsbeschaffung gehen, sondern dass wir von vornherein zunächst eine vollständige Übersicht aller Studien und dann den vollständigen Studienbericht für die von uns identifizierten relevanten Studien bekommen. Das ist sicher etwas, was das Verfahren insgesamt beschleunigt.

Volker Vervölgyi: Trotzdem kann es auch den Fall geben, dass es IITs gibt, die in die Dossierbewertung eingehen. Den Fall hatten wir auch schon häufiger, und zwar nicht nur bei indirekten Vergleichen; ich nenne als Beispiel Abirateron bei metastasiertem Prostatakarzinom. Da hatten wir die STAMPEDE-Studie, und die hat einen großen Anteil der Daten ausgemacht, im Vergleich zu der zusätzlich vom Hersteller gelieferten Studie. Und wenn wir die jetzt nicht gehabt hätten, aber identifiziert hätten, wäre es trotzdem für die Bewertung schwierig gewesen,

Aussagen zu treffen. Daher kann es auch da zum Tragen kommen, dass dann Daten fehlen und die dann noch Auswirkungen aufs Fazit haben.

Moderator Jürgen Windeler: Herr Sauerbruch.

Tilman Sauerbruch: Herr Windeler, noch mal zu Ihren Ausführungen. Mir ist doch nicht klar, warum Sie bei Studienregistern nicht nachfragen können oder wollen. Ich meine, wenn wir dieses Instrument schon haben und Ihnen auffällt, dass da möglicherweise die Daten nicht ausreichend gepflegt sind, dann ist es doch wert, da nachzufragen. Ich meine, Sie akzeptieren ja dann von vornherein einen Bias.

Stefan Sauerland: Herr Sauerbruch, natürlich fragen wir nach. Bislang haben wir das gemacht. Unsere Sorge ist, dass die Zeit hierfür und vor allen Dingen die Reaktionszeit, die dann auf Ihrer Seite verbleibt, immer kürzer werden. Das soll eben auch dadurch verhindert werden, dass Registereinträge aktuell gehalten werden. Unser Hauptproblem sind Registereinträge, die Jahre alt sind, und von akademischen Institutionen irgendwann einmal geplante Studien, bei denen fraglich ist, ob es diese Studie wirklich jemals gegeben hat. Dann ist ein Autorenkontakt natürlich auch äußerst schwierig.

Ich würde auch gerne einmal etwas zu den verschiedenen Studienquellen sagen, weil das gerade so ein bisschen den Eindruck hatte, als würden wir eventuell in einer Bewertung eines Produktes erhebliche Probleme kriegen, wenn die akademischen Studien, die IITs, unvollständig verfügbar sind. Das ist ein Problem, aber da ist im Methodenpapier jetzt ein Absatz vorgesehen, dass wir uns im Zweifelsfalle auch einen kompletten Studienpool herausgreifen und den anderen Studienpool dann komplett ignorieren. Das könnte konkret bedeuten, dass es einen Hersteller gibt, der ein Produkt hat. Dieses Produkt hat er in drei, vier Studien untersucht, und diese Studien sind publiziert. Es gibt aber drei, vier Studien aus der akademischen Welt, die nicht publiziert sind. Dann würden wir zum Beispiel sagen: Okay, wir machen diese Bewertung allein auf der Basis des Herstellerpools und lassen die Studien von einem anderen Sponsor, von einer anderen Quelle komplett außen vor. – Dann tun wir dies aber komplett.

Moderator Jürgen Windeler: Vielleicht nur noch eine Anmerkung. Es ging nicht darum, dass wir bei Registern nicht nachfragen. Also, Register gucken wir uns sowieso systematisch an. Es geht darum, ob wir bei Unklarheiten, Unvollständigkeiten, irgendwelchen Fragen bei Studienautoren oder Sponsoren nachfragen. Diese Möglichkeiten haben wir in der Vergangenheit mehr oder weniger intensiv genutzt. Wir halten sie auch grundsätzlich für sinnvoll, jedenfalls wenn man nicht ein halbes Jahr auf eine Antwort warten muss. Also, wir gucken dann schon, dass das zügig gehen sollte. Aber mit den Fristen, die der Gesetzgeber dem GB-A und uns im Methodenbewertungsverfahren noch lässt, wird das nicht mehr oder nur noch sehr, sehr eingeschränkt möglich sein. Das ist einfach eine ganz banale Konsequenz aus dem, was jetzt da passiert.

Tilman Sauerbruch: Darf ich darauf antworten? – Ja, natürlich könnten Sie dann auch in der akademischen Welt Ihre Fristen vorgeben und das dann in Ihrem Report eben entsprechend darstellen. Das führt natürlich auch dort zu einer – wie soll ich sagen? – Verbesserung der Stringenz. Das muss man im Dialog machen. Es ist natürlich auch ein Schlechtes, wenn das nicht gepflegt wird. Das ist ein Pingpongspiel. Man sollte aber nicht ganz darauf verzichten.

Moderator Jürgen Windeler: Dem stimme ich zu. – Ich sehe dort hinten noch eine Wortmeldung.

Annett Kucka: Annett Kucka von Bristol-Myers Squibb. – Ich habe noch eine Rückfrage zu den ISRs. Wenn die Daten für den pU in welcher Form auch immer sozusagen nicht verfügbar sind, wenn man die Daten also einfach nicht bekommt, läuft das dann auch unter „fehlende Daten“?

Volker Vervölgyi: Der Hersteller ist erst einmal dazu verpflichtet, seine eigenen Studien und auch alle verfügbaren Daten zu Studien Dritter zu liefern. Ich nehme die STAMPEDE-Studie als Beispiel. Wenn es gar keine Publikationen und gar keine verfügbaren Daten dazu gäbe, dann gilt das natürlich nicht. Hätte es sie aber gegeben und wäre sie nicht im Dossier gewesen – das war der Punkt, den ich eben meinte –, dann wäre es eine andere Situation.

Moderator Jürgen Windeler: Ich sehe keine weiteren Wortmeldungen; Herr Flintrop hat auch keine registriert. Dann schlage ich vor, dass wir diesen Tagesordnungspunkt verlassen und zum TOP 4 übergehen.

3.4.5 Tagesordnungspunkt 4: Aufbereitung und Darstellung von Schadensaspekten (Abschnitt 3.1.3)

Herr Vervölgyi, bitte.

Volker Vervölgyi: Hier geht es um den Abschnitt 3.1.3 im Methodenpapier. Was haben wir da gemacht? Wir haben konkretisiert, sowohl für Bewertungen nach § 139a, also die ausführlichen Bewertungen, als auch für die Dossierbewertungen, welche Schadensendpunkte wir eigentlich heranziehen. Das haben wir ein bisschen weiter ausgeführt, als es vorher dort stand. Die Stellungnehmenden insbesondere von der pharmazeutischen Industrie haben in ihrer Stellungnahme darauf hingewiesen, dass die Ausführungen nicht konkret genug seien, dass nicht genau klar werde, wie wir denn da vorgehen würden. Ich entnehme dem so ein bisschen, dass der Wunsch nach einem mehr oder weniger algorithmischen Vorgehen besteht. Das zieht sich durch die Diskussion heute: Braucht man Zahlen, oder braucht man keine Zahlen? Kann man das eindeutig quantifizieren, wie man vorgehen sollte, oder nicht?

Meine Frage wäre jetzt vor dem Hintergrund dessen, dass wir ja eigentlich relativ heterogene Fragestellungen haben, insbesondere in der Dossierbewertung, mit verschiedensten Erkrankungen und verschiedensten Schweregraden der Erkrankung, ob Sie sich ein algorithmisches

Vorgehen überhaupt vorstellen können. Denn ich habe da tatsächlich Schwierigkeiten. Wenn ja, wie würde das aussehen und wie sollte man das dann genau operationalisieren?

Moderator Jürgen Windeler: Wer möchte? – Herr Lützelberger.

Uwe Lützelberger: Wir haben von unserer Seite aus kein algorithmisches Vorgehen oder so was gefordert oder vorgeschlagen. Uns ging es in einer Stellungnahme primär darum, in den konkreten Nutzenbewertungsberichten etwas transparenter darzustellen, wie Sie das gemacht haben. Wir haben schon einige Beispiele, wo Sie das gemacht haben. Ich jedenfalls konnte nicht nachvollziehen, warum bestimmte Symptome einbezogen wurden, andere nicht, obwohl beides über 10 % Häufigkeit hatte, verschiedene Kriterien jedenfalls. Also, es ist nicht gut nachvollziehbar gewesen. Wenn man nicht weiß, wie das zustande gekommen ist, dann kann man sich im Stellungnahmeverfahren auch nicht sinnvoll dazu äußern.

Bei dem Fall, den ich mir angeguckt hatte, war das kein großes Problem. Das war die Studie gegen beobachtetes Abwarten. Was sich da ergab, war dann tatsächlich auch gut kompatibel mit dem, was in der Fachinformation zu dem Produkt stand. Aber wie gesagt, man konnte es nicht gut nachvollziehen, wie bestimmte Sachen da hineingekommen sind, andere nicht. In einem Fall wurde zum Beispiel ein Symptom, das mal zugunsten des Produktes ausgegangen war, nicht aufgenommen. Ein anderes wurde aufgenommen. Dann kann man natürlich nicht Stellung dazu nehmen, wenn man das nicht verstanden hat.

Der andere Punkt war vielleicht ein bisschen nachrangig, aber einfach so der Hinweis: Das, was Sie hier machen, ist ja ein bisschen was anderes als das, was wir a priori in den Studien als unerwünschte Ereignisse von spezifischem Interesse definieren. Das klingt hier sehr ähnlich. Die Anregung wäre sozusagen, das vielleicht ein bisschen anders zu bezeichnen, was Sie jetzt hier tun, um das nicht zu Verwirrung führen zu lassen.

Volker Vervölgyi: Zum zweiten Punkt. In der Tat ist es etwas anderes, weil wir natürlich eine andere Fragestellung haben als für die Zulassung, als Sie für Ihre Studie, weil es natürlich darum geht, die Nebenwirkungen des neuen Produktes quasi genauer zu betrachten. Was wir in der Dossierbewertung oder überhaupt in den Nutzenbewertungen machen, ist ein direkter Vergleich zwischen zwei Therapieoptionen, und deswegen kann es schon allein deshalb nicht das Gleiche sein, weil wir beide Seiten betrachten müssen.

Zur Konkretheit in den Bewertungen. Es ist schon so, dass wir – und das ist ja auch im Methodenpapier beschrieben – zwei Wege haben, auch in der Dossierbewertung, um diese Endpunkte zu diesen spezifischen unerwünschten Ereignissen – auf die will ich mich jetzt fokussieren – zu identifizieren. Einerseits gucken wir, welche Nebenwirkungen oder potenziellen Nebenwirkungen auf jeden Fall relevant sind. Denen müssen wir uns widmen, die müssen wir auf jeden Fall mit hineinnehmen. Das geht aus der Dossierbewertung meines Erachtens schon hervor, welche das sind. Nehmen wir als Beispiel Diabetes. Es ist ganz klar, dass wir uns Hypoglykämien angucken, und das ist dann in der Liste der betrachteten

Endpunkte auch aufgelistet. Ein anderes Beispiel sind Blutungen bei Thrombozytenaggregationshemmern. Das sind Nebenwirkungen, denen wir uns auf jeden Fall widmen werden.

Der Rest ist dem Umstand geschuldet, dass wir nur eine geringe Zeitspanne von drei Monaten haben und dass wir zumeist neue Arzneimittel haben, die man noch gar nicht so antizipieren kann. Da geht es darum, zu schauen: Was ist in den Studien eigentlich an Ereignissen aufgetreten? Wo gibt es große Unterschiede? – Das ist der eine Punkt.

Der andere Punkt ist – und das wundert mich –, dass Sie Ihren Stellungnahmen sagen, das kann man gar nicht nachvollziehen. Nach meiner Erfahrung war es eigentlich in den allerseltensten Fällen so, dass Unklarheit geäußert wurde, warum jetzt genau diese Nebenwirkungen hereinkommen und diese nicht. Das würde ich dann erwarten, wenn es tatsächlich überhaupt nicht nachvollziehbar wäre.

Moderator Jürgen Windeler: Gibt es weitere Wortmeldungen? – Herr Dintsios.

Charalabos-Markos Dintsios: Ich glaube, das liegt auch daran, dass es sehr konzise ist, aber drei Aussagen enthält. Das ist dieser Abschnitt 3.1.3. Andererseits ist sozusagen vorgetragen worden, dass die am besten geeignete Operationalisierung in Bezug auf Mess- und Ergebnissicherheit herangezogen werden soll. Es gibt überhaupt kein Kriterium, oder gehen wir mal davon aus, das IQWiG würde die selber aussuchen oder der Hersteller bietet mehrere an und man pickt sich die, von der man glaubt, sie sei die am besten geeignete, heraus. Dann geht man zurück auf das inhaltliche Konstrukt, wenn es inkonsistente Ergebnisse sind. Und dann kommt noch der dritte Aspekt der nicht geeigneten Operationalisierung. Und dann kommt es noch mal zurück auf die Konsistenz oder die Inkonsistenz. Das heißt, hier sind ja szenariotechnisch fast neun Szenarien abgebildet. Es kann ja sein, dass ich zwei geeignete Operationalisierungen habe, aber die Ergebnisse inkonsistent sind. Wieso soll ich sie mir nicht genau anschauen? Vielleicht gibt es eine Ätiologie für die Inkonsistenz. Es kann ja sein, dass beide Operationalisierungen gleich gut geeignet sind. Und dann wird die Begründung der Inkonsistenz sozusagen völlig ignoriert. Den vorigen Abschnitt des Satzes kann jeder nachvollziehen: Wenn es keine geeignete gibt ... – Ich glaube, da sind wir alle d'accord. Das macht die Lesbarkeit dieses Abschnitts sehr, sehr schwer. Der ist sehr konzise, hat aber sehr viele Aussagen, und ich glaube nicht, dass hier das Algorithmische, sondern eher eine etwas entzerrendere Darstellung gefordert ist.

Ich lese Sätze normalerweise einmal und verstehe sie. Das musste ich wirklich zweimal lesen, um die mehreren Aussagen, die hier verkappt oder verschachtelt – nicht beabsichtigt – enthalten sind, zu verstehen. Ich glaube, das würde uns alle ein bisschen weiterbringen bei diesem spezifischen Abschnitt.

Jens Flintrop: Herr Dintsios, Sie sprechen jetzt von der Lesbarkeit. Wir haben ein bisschen ein Problem mit Ihrer Verständlichkeit, insbesondere der Stenograf. Wenn es beim nächsten Mal ein bisschen langsamer wäre, wäre das sehr hilfreich.

Charalabos-Markos Dintsios: Also, da ich ehemaliger IQWiG-Mitarbeiter bin, wissen Sie doch, dass ich heute sehr langsam rede. Ich kann noch 20-mal schneller.

Jens Flintrop: Sie sind gerade schneller geworden. – Herr Vervölgyi, bitte.

Volker Vervölgyi: Zu dem Punkt, was die bestgeeignete Operationalisierung ist, kann ich kurz ausführen: Wenn man sich die Dossiervorlagen anschaut, ist klar, was wir geliefert bekommen. Das sind einmal die PTs und SOC's nach MedDRA, jetzt, nach den neuen Dossiervorlagen, auch die Standardised MedDRA Queries, also die SMQs, in denen noch mal Konzepte zusammengefasst werden. Und in seltenen Fällen ist es so – das sind mehr oder weniger die beiden Beispiele, die ich eben genannt habe, nämlich Bluthochdruck und Hypoglykämien –, dass Beispiele auch außerhalb von MedDRA noch mal in der Studie operationalisiert werden.

Und was wir machen, ist, uns zu einem Konstrukt – nehmen wir zum Beispiel mal Hypoglykämien – anzuschauen, ob es in der Studie tatsächlich außerhalb von MedDRA eine Operationalisierung gibt. In den meisten Diabetesstudien ist das so. Dann wird das wahrscheinlich so sein, dass das eigentlich auch die beste Operationalisierung ist, die wir dann auch auswählen würden. In anderen Fällen, in denen es so etwas nicht gibt, sondern man sich auf SOC, PT oder SMQ beschränken muss, ist es tatsächlich eine Abwägungssache, und man muss sich einmal die SMQs angucken. Da gibt es auch viele verschiedene Wege: Ist es die Broad or Narrow Scope? – Da gibt es ganz viele Spielarten, und das ist eine Abwägungssache, welche dann für das, was ich mir eigentlich anschauen möchte, die Nebenwirkung, zu der ich eine Aussage treffen möchte, die beste ist. Das ist, glaube ich, sehr individuell, je nachdem, welche Fragestellung ich habe.

Inkonsistenz. Da ist eigentlich etwas anderes mit gemeint. Nehmen wir mal zwei PTs als Beispiel. Ich habe einmal das PT Bronchitis und einmal irgendwie Husten oder Infektion der oberen Atemwege. Wenn das eine in die eine Richtung, das andere in die andere Richtung geht, dann ist es schwierig, das zu interpretieren. Natürlich kann das eine Ätiologie haben – das ist ganz klar –, aber es ist anhand der Daten, die wir haben, die auch nur eine Sammlung der Ereignisse sind, die in der Studie aufgetreten sind und auf vielleicht ganz vielen verschiedenen Verbatims beruhen, trotzdem schwierig, daraus irgendwie eine gerichtete Aussage zu machen. Das ist mit Inkonsistenz an der Stelle gemeint.

Wenn ich das jetzt zusammenzähle, komme ich auf drei Szenarien: Das eine ist die Operationalisierung, das andere ist Inkonsistenz, und wenn beides nicht gegeben ist, dann kann ich das nicht betrachten. – So ist es gemeint.

Moderator Jürgen Windeler: Ich habe jetzt den Eindruck, dass die Wortmeldungen einheitlich in die Richtung gehen, dass einerseits die Darstellung im Methodenpapier noch mal überdacht

und vielleicht auch verbessert werden kann und dass andererseits – das war der Hinweis von Ihnen – die jeweiligen Begründungen in den Dossiers noch mal überdacht und gegebenenfalls verbessert werden können. Jedenfalls geht es nicht so sehr um konzeptionelle Kritik, sondern eher um Kritik an den Ausführungen. Und Herr Dintsios musste es ja nur zweimal lesen. Insofern kann es nicht so schlimm gewesen sein, kann der Verbesserungsbedarf nicht so groß sein. Ich denke, dem können wir uns dann widmen.

Charalabos-Markos Dintsios: Herr Windeler, Sie müssen bedenken: Ich bin nicht nativ Deutsch sprechend.

Moderator Jürgen Windeler: Gibt es weitere Wortmeldungen? – Auch nach ein bisschen Bedenkzeit nicht. Dann können wir den TOP 4 schließen und kommen zu:

3.4.6 Tagesordnungspunkt 5: Umgang mit unvollständigen Daten (Abschnitt 9.2.12 und 9.3.13)

Guido.

Guido Skipka: Das Thema „unvollständige Daten“ treibt uns seit vielen Jahren um. Jeder, der sich mit der Bewertung von Studien beschäftigt, wird wissen, wovon ich rede. Wir haben im Entwurf des Methodenpapiers für die Version 6.0 zwei neue Abschnitte eingefügt, in denen wir ein paar Gedanken von uns formuliert haben. In Abschnitt 9.3.12 geht es um das Thema „fehlende Werte“, in Abschnitt 9.3.13 etwas spezieller um unterschiedliche mittlere Beobachtungsdauern.

Insbesondere den ersten Abschnitt möchte ich mit Ihnen diskutieren. Es gab zu beiden Abschnitten eine ganze Reihe von Stellungnahmen.

In 9.3.12 haben wir formuliert, dass ein unvollständiges Follow-up das Verzerrungspotenzial erhöht. Das erstaunt, glaube ich, niemanden hier im Raum; dazu gab es auch keine Kritik in den Stellungnahmen. Wir haben aber relativ konkret formuliert, dass wir Ergebnisse in der Bewertung nicht berücksichtigen, falls diese auf weniger als 70 % der einzuschließenden Patienten basieren bzw. wenn der Unterschied zwischen den Gruppen größer als 15 Prozentpunkte ist. Wir haben also ganz konkrete Zahlen genannt, und wir haben unter TOP 3 diskutiert; da haben wir keine konkreten Zahlen genannt. Ich habe zwischen den Zeilen herausgehört – das kam auch in den Stellungnahmen –, da hätte man sich konkrete Zahlen gewünscht. Hier, sozusagen auf individueller Ebene von fehlenden Werten, haben wir also Zahlen genannt. Es wurde in vielen Stellungnahmen angemerkt, dass die Grenzen vielleicht zu starr sind. Insbesondere wurde angemerkt, dass die Rationale für diese Grenzen – so will ich sie mal nennen – fehlen würde.

Bevor wir das diskutieren, noch zwei, drei Anmerkungen dazu. Diese Grenze – wir nennen das 70%-Regel – verwenden wir seit vielen Jahren, man kann schon sagen, seit mindestens einem Jahrzehnt. Das ist eine Setzung, die wir schon sehr lange haben. Die hatten wir bis jetzt nicht

im Methodenpapier formuliert, sondern in den Berichtsplänen. Also, das gibt es schon wirklich sehr lange. Ich glaube, da war Herr Dintsios sogar noch im IQWiG, als wir diese Grenzen festgelegt haben. Wir haben sie jetzt ins Methodenpapier gehoben, weil wir sie nicht jedes Mal in die Berichtspläne schreiben wollten und es für uns so etabliert ist, dass es würdig ist, ins Methodenpapier gehoben zu werden. Also, wir sagen, das ist jetzt erst mal nichts Neues, wir haben das jetzt nur ins Methodenpapier geschrieben.

Das andere ist ein Missverständnis, das ich vielleicht hier schon mal ausräumen kann. Es kam in den Stellungnahmen heraus, dass diese Regelung vor allem längeren Studien nicht gerecht wird. Wenn wir also mehrjährige Studien haben, ist es doch ganz klar, dass man mit der Zeit Follow-up-Verluste hat und dass die irgendwann auch größer als 30 % sind. Das kann doch nicht dazu führen, dass wir diese Ergebnisse nicht betrachten. Da gebe ich Ihnen völlig recht. Das ist aber ein Missverständnis. Ich kann vorwegnehmen, das werden wir in dem Abschnitt etwas klarer formulieren, dass zukünftige Leserinnen und Leser diesem Missverständnis nicht aufliegen.

Diese 70%-Regel bezieht sich dezidiert auf Patienten, die überhaupt nicht in der Auswertung sind oder, anders ausgedrückt, die zur Baseline zensiert werden. Es geht nicht um Patienten, die im Verlauf herausfallen, also unvollständig beobachtet sind, aber trotzdem mit ihren Werten eingehen. Das muss man sauber differenzieren. Das ist im Abschnitt nicht passiert. Das werden wir sauberer differenzieren. Aber ich kann es hier schon einmal klarstellen: Wir betrachten bei langjährigen Studien natürlich auch die Ergebnisse, selbst wenn am Ende mehr als 30 % unvollständig beobachtet sind. Das ist nicht unsere 70%-Regel, wie wir sie leben und zumindest formulieren wollten.

Eine Frage kam von vielen Stellungnehmenden. Ich hätte mir jetzt einfach mal die Vertretung der biometrischen Fachgesellschaften, also konkret dich, Tim, herausgepickt, aber die Frage richtet sich an das ganze Auditorium. Wir haben diese Grenzen genannt. Wie gesagt, sie sind nicht ganz neu. Es kam in vielen Stellungnahmen die Frage nach der Rationalen, nach Literatur, die das stützt. Wir haben keine Literatur dazu. Meine konkrete Frage an Sie und auch an dich, Tim, wäre: Haben Sie bzw. hast du Literaturhinweise, die das stützen würden? – Ich nehme an, dass das nicht der Fall ist. Dann wäre meine Anschlussfrage: Wie würden Sie sich solchen Grenzwerten nähern, wenn Sie denn Grenzwerte festlegen müssten?

Jens Flintrop: Herr Dheban.

Srirangan Dheban: Dheban von der Firma Janssen. – Ich denke, es ist durchaus wichtig, dass man in diesem Fall herzuleiten versucht, wie man auf solche Grenzen kommen kann. Dazu könnte man vielleicht in irgendeiner Form versuchen, Simulationsstudien aufzusetzen, in denen man dann Stück für Stück den Anteil der fehlenden Daten erhöht, um zu schauen, welchen Impact das auf den Bias hat.

Guido Skipka: Ja, das könnte man in der Tat machen.

Moderator Jürgen Windeler: Guido, bitte.

Guido Skipka: Das sollte jetzt nicht flapsig sein. Solche Simulationsstudien gab es in der Vergangenheit. Es gibt sogar eine Arbeit von Herrn Windeler mit Frau Unnebrink – die ist schon etwas betagter –, in der solche Simulationen gemacht wurden. Das Problem ist, das hat so viele Parameter, Stellschrauben, Dimensionen, dass man je nach Simulationsaufbau alles zeigen oder auch widerlegen kann. Also, das ist ohne weitere Pflöcke, die man sich setzt, sehr schwer durchführbar.

Moderator Jürgen Windeler: Ralf, bitte.

Ralf Bender: Als Ergänzung hierzu: In einer ganz konkreten Datensituation könnte man das machen, also gucken. Wenn der Effekt genau so ist und man einen Stichprobenumfang von genau soundso viel hat, dann wäre so was auch zielführend. Dann muss man sich ein Kriterium vorgeben. Man erlaubt sich ein Bias von maximal relativ 5 % oder so was. Dann könnte man ermitteln: Aha, hier können wir noch etwas ableiten, wenn zum Beispiel sogar 35 % der Daten fehlen.

Aber für eine andere Situation gilt das wieder nicht. Also, wir brauchen doch hier aus pragmatischen Gründen sozusagen einen Grenzwert für alles. Und diesen kann man nicht mit Simulationen belegen, weil es so viele verschiedene Situationen gibt, dass das einfach nicht zielführend ist, hier eine Simulationsstudie aufzusetzen.

Jens Flintrop: Guido.

Guido Skipka: Also, ich habe eben – ich hoffe, erfolgreich – versucht, dieses Missverständnis auszuräumen. Diese 70-%-Regel kommt gar nicht so oft zum Tragen bei uns. Ein viel größeres Problem haben wir mit Patienten, die mit gewissen Informationen eingehen, aber unvollständig. Da werden wir in der Regel Ergebnisse nicht einfach zur Seite legen.

Wir schauen uns Ersetzungsstrategien an. Bei binären Daten können wir sogar selber Ersetzungsstrategien rechnen, weil wir dazu keine individuellen Patientendaten brauchen. Wir schauen uns weitere Faktoren an. Also, diese Follow-up-Verluste im Verlauf schauen wir uns sehr genau an, und da wird es auch sehr, sehr individuell. Diese 70-%-Regel, die jetzt für alles gilt, kommt zum einen nicht so oft zum Tragen, und wenn sie zum Tragen kommt, sind die Fälle meistens relativ klar.

Das viel größere und viel relevanter vorkommende Problem sind unvollständige Daten im Verlauf. Da gehen wir zum Teil sehr individuell heran. Wir schauen uns alle Informationen an. Welche Gründe führen zu Ausschlüssen? Also, jeder, der sich mit fehlenden Werten beschäftigt, landet sofort bei diesen Konzepten „Missing completely at random“ oder informative Zensierung usw. Wer unsere Bewertungen liest, stellt hoffentlich fest, dass wir da nicht hemdsärmelig und pauschal herangehen, sondern versuchen, jeweils den Kontext zu berücksichtigen.

Jens Flintrop: Herr Dintsios.

Charalabos-Markos Dintsios: Dazu vielleicht ein Kommentar – ich habe es auch unserer Stellungnahme angefügt –: Ich glaube, mit dem 70%-Wert kann man sehr gut leben. Der ist sehr pragmatisch; das habe ich auch nicht in Abrede gestellt.

Ich tue mich schwerer mit den 15%-Werten zwischen den beiden Armen. Denn da kann es durchaus Konstellationen geben, die auch zu Baseline, sozusagen zu solchen Sachen führen. Ich denke an offene Studiendesigns. Die Leute merken, wo sie landen. Dann geben sie ihre Lebensqualitätsbögen nicht ab, Staff oversight, klassische Probleme bei klinischen Studien. Da bin ich ein bisschen zurückhaltender, muss ich sagen. Und vielleicht nicht eine Rationale, sondern irgendwo praktiziert auf der Welt. Was macht man in anderen Systemen, dass man so einen hohen Anker setzt, ohne jetzt den Vorwurf zu machen, dass die 15 falsch sein muss? Aber es ist schwierig; denn es gibt eindeutige Fälle in der Bewertung, aber es gab ein paar Fälle, wo ich es auch gesehen habe und dachte: Oje, da ist jemand bei 18 % gelandet. Was machen wir jetzt mit diesen Daten nun?

Jens Flintrop: Herr Leverkus.

Friedhelm Leverkus: Leverkus. – Ich habe eine Nachfrage. Die protokollbedingten Studienenden werden im Prinzip nicht mit zu den 70 % gezählt. Denn es ist gerade in der Onkologie so, dass das Studienende durch einen Progress oder so was definiert wird und die Patienten mit dem Medikament aufhören und ein anderes Medikament nehmen. Die sind im Prinzip nicht mehr im Risk Set drin. Das heißt, die würden bei euch auch nicht mehr in den Nenner hineinfallen.

Jens Flintrop: Guido Skipka.

Guido Skipka: Um es noch mal auszuführen: Wir schauen uns das Ganze schrittweise an. Wir haben, sagen wir mal, 100 Patienten in einem Arm, gucken in die Auswertung, beispielsweise in eine Cox-Regression, und da sind zur Baseline nur noch 60 % unter Risiko, weil 40 % zur Baseline zensiert sind. Da haben wir ein großes Problem, es sei denn, wir können sehr gut nachvollziehen, dass diese 40 Personen oder 40 %, die da herausfallen, keine informativen Zensierungen sind. Das klassische Beispiel ist die multizentrische Studie. Einem Zentrum ist irgendwie der Datenzugriff abgeraucht, kann keine Daten mehr liefern, und selbst wenn dieses Zentrum ein Gewicht von über 30 % hätte, würden wir die Zahl der anderen Zentren nehmen, weil das Abrauchen des Datacenters vermutlich wenig mit den Patientencharakteristika oder den Behandlungen zu tun haben wird.

Sprich, wenn wir sehen, es gibt einen Anteil, der relativ klar „Missing completely at random“ ist, um mal den Fachbegriff zu nennen, dann kommt der nicht im Nenner. Also, die ziehen wir sofort ab, die sind überhaupt kein Problem für uns. Ein Problem haben wir, wenn mehr als 30 % des dann übrig bleibenden Nenners aus potenziell informativen Gründen gar nicht in der Auswertung sind. Dann haben wir ein Problem.

Zu dem, was du angesprochen hast, Friedhelm. Diese Follow-up-Verluste nach Progression sind aus unserer Sicht Patienten, die mit gewissen Informationen einfließen. Die sind von dieser Regel überhaupt nicht betroffen. Das schauen wir uns völlig separat an in einem zweiten Schritt.

Also, wir schauen uns erst mal: Was ist überhaupt der Nenner, den wir haben wollen? Wer geht überhaupt mit einer Information größer Epsilon in die Auswertung? – Das ist die 70%-Regel. Der Rest sind Patienten, die vollständig oder unvollständig beobachtet sind, aber in gewisser Weise beobachtet sind, und da gehen wir mit unserer ITT-Bewertung heran. Wir schauen: Was haben wir da an Sensitivitätsanalysen, Ersetzungsstrategien? Was sind die Gründe für das vorzeitige Follow-up-Ende? – Da jonglieren wir weniger mit Prozentzahlen.

Moderator Jürgen Windeler: Kannst du noch eine Antwort auf den Kommentar von Herrn Dintsios geben?

Guido Skipka: Die 15 % sind nicht durch Literatur gestützt, weil es dazu auch keine Literatur gibt. Das ist eine Setzung von uns, wie gesagt, die wir seit über zehn Jahren verfolgen.

Uns ist es wichtig, dass wir ein Kriterium haben, wenn wir zwischen den Gruppen differentielle Anteile von Personen haben, die in die Auswertung kommen. Jetzt kann man sich lange streiten, ob das 10, 15, 20 % oder etwas anderes sein sollte. Wir haben gesagt: Na gut, wenn insgesamt nicht mehr als 30 % in der Auswertung sind, haben wir ein Problem, oder wenn der Unterschied die Hälfte ist.

Das ist, wie wir das manchmal so machen, etwas hemdsärmelig, aber ich wüsste nicht, welche andere Grenze viel sinnvoller wäre.

Moderator Jürgen Windeler: Stefan.

Stefan Sauerland: Ich möchte noch einen Satz zu diesen Simulationsstudien ergänzen, weil wir im Methodenpapier eine Arbeit von einer Frau Kristman zitieren, in der genau das gemacht wurde und sich zeigte, dass bei einem „Missing not at random“-Mechanismus bereits 20 % Datenverlust zu erheblichem Bias führen; das steht da so drin. Wenn ich einen differentiellen Follow-up-Verlust habe, der 15 % erreicht, dann bin ich schon relativ sicher, dass da ein Mechanismus dahinter ist, der erst einmal formal nicht random ist, und dementsprechend ist man dann mit den 15 % schon relativ nah an den 20 %. Man könnte jetzt auch sagen, diese 20 %, die Frau Kristman da gefunden hat, sind vielleicht eher für unsere 30 % maßgeblich. Da sieht man aber, dass unsere 30%-Grenze durchaus liberal ist.

Jens Flintrop: Herr Skipka.

Guido Skipka: Vielleicht noch als Bemerkung – wer seit vielen Jahren unsere Erörterungen miterlebt, kennt das auch –: Wir schreiben ganz bewusst – auch an dieser Stelle –, dass wir in der Regel diese 70%-Regel anwenden. Also, wir gehen da schon mit Augenmaß heran. Wir rechnen das nicht aus, und wenn wir bei 70,1 % sind, haben wir kein Problem, und wenn wir

bei 69,9 % sind, schauen wir uns das alles nicht mehr an. Das Ganze hängt natürlich vom Kontext ab, von der Größe der Effekte. Auch bei großen Effekten wird man da vielleicht liberaler sein als bei kleinen Effekten. – So viel dazu.

Moderator Jürgen Windeler: Gibt es noch Fragen? – Ja, Herr Friede.

Tim Friede: Weil ich direkt angesprochen wurde, wollte ich noch kurz antworten. Also, uns ist auch keine Arbeit bekannt, die sich hier anbieten würde, jetzt konkrete Grenzen zu nennen; sonst hätten wir sie auch so benannt. Aber ich denke, was für uns hilfreich ist, ist diese Erläuterung, die wir heute bekommen haben, dass es sich hier wirklich auf die Patienten bezieht, von denen überhaupt keine Daten in der Nachbeobachtung vorhanden sind. Das war uns in der Form nicht bekannt, und jeder kennt Indikationen, denke ich, wo es im Verlauf zum Teil doch zu deutlichem Drop-out kommt. Insofern hat man sich gefragt, wie sinnvoll das hier mit den 30 % ist.

Aber auch mal aus der eigenen Erfahrung gesprochen: Also, bei Studien, in denen für jeden dritten Patienten gar keine Information im Verlauf vorliegt, würde ich denken, dass dort im Design etwas falsch angelegt ist. Entweder waren die Visiten nicht regelmäßig genug, oder man hat vielleicht die Randomisierung an den falschen Zeitpunkt gesetzt oder so etwas. Insofern würde ich von strukturellen Problemen in der Versuchsanlage, wenn man so will, ausgehen.

Ich glaube, diese Erläuterungen, die wir jetzt bekommen haben, sind sinnvoll. Vielleicht ist das aber auch ein Punkt, wo man sagen kann: Ich weiß jetzt nicht, ob es so gut ist, eine Formulierung zu wählen, wo man strikt 15 Prozentpunkte oder so etwas formuliert. Vielleicht ist es besser, dass man das mehr so als eine Zielvorgabe formuliert, wo man sich auch vorstellen kann, dass es unter besonderen Gesichtspunkten vielleicht auch Anlässe gibt, davon abzurücken. – Danke.

Moderator Jürgen Windeler: Guido.

Guido Skipka: Wie gesagt, ich habe mir den Abschnitt auch gestern noch mal durchgelesen und kann gut verstehen, dass es da zu Missverständnissen kommt, weil wir von Follow-up-Verlusten reden und ohne Absatz dann auf einmal von Patienten, die gar nicht in der Auswertung sind. Wir werden das klarer formulieren, damit diese Aufteilung der verschiedenen Arten von fehlenden Werten klarer wird. Vielleicht sollten wir auch noch klarer formulieren, dass das keine strikten Grenzen sind, sondern Orientierungen, an denen man sich entlang-handelt.

Tim Friede: Sehr gut. Vielen Dank. – Und die Arbeit mit der Kollegin Unnebrink liegt so lange zurück, dass sie noch nicht einmal Unnebrink hieß.

(Heiterkeit)

Moderator Jürgen Windeler: So ist es. Ich möchte auf die Arbeit jetzt nicht weiter eingehen. „Statistics in Medicine“, 1998 oder so in der Größenordnung.

Wer möchte sich noch zu der Thematik äußern? – Ich sage nur noch mal: Die Arbeit löst das Problem in der Tat nicht. Guido hat das ja beschrieben. – Bitte.

Guido Skipka: Ich möchte noch etwas zu dem zweiten Abschnitt sagen, wenn es zu diesem Abschnitt keine weiteren Fragen mehr gibt. – Es gibt noch den zweiten Abschnitt, wo wir uns über Unterschiede in den mittleren Beobachtungsdauern auslassen. Das ist ein etwas speziellerer Abschnitt. Der resultiert aus der Beschäftigung mit den onkologischen Studien in den letzten Jahren, wo man das typischerweise sieht, dass nach Progression oftmals die Beobachtung beendet wird und man, sobald man Progressionseffekte hat, aus unserer Sicht sofort in das Problem läuft, dass man zwischen den Gruppen unterschiedlich kumuliert, unterschiedlich lange Beobachtungsdauern hat. Da ist man direkt in der Estimand-Diskussion. Für die Fachleute unter Ihnen: „While on Treatment“ oder „Treatment Policy“. Wir haben in den letzten zwei, drei Jahren eine Arbeitsgruppe gehabt – Herr Friedhelm Leverkus hat sie schon erwähnt –, an der ich auch beteiligt war. Dabei ist ein schönes Paper herausgekommen.

In den Stellungnahmen wurde noch mal angemerkt, dass es diese Arbeit gibt, die wir zwar zitieren, aber etwas allgemeiner eingebracht werden sollte. Zum Beispiel kam der Vorschlag, dass wir uns noch mal zum Thema „konkurrierende Risiken“ äußern, was wir bis jetzt in diesem Abschnitt nicht tun. Ich habe jetzt keine Frage an Sie, sondern wollte nur signalisieren, dass wir uns das noch einmal überlegen, speziell zum Thema „konkurrierende Ereignisse“. Das kam nicht nur von den biometrischen Fachgesellschaften, sondern auch von der AkdÄ, das gegebenenfalls noch ein bisschen zu erweitern.

Moderator Jürgen Windeler: Gibt es Wortmeldungen hierzu? – Herr Kohlmann.

Thomas Kohlmann: Wenn ich zum letzten Abschnitt noch einen kurzen redaktionellen Hinweis geben darf. In 9.3.12 wird in der vierten und fünften Zeile als ein Grund für Drop-out Widerruf der Einverständniserklärung genannt. In der Ethikkommission in Greifswald sagen uns die Juristen immer, es muss juristisch korrekt „Einwilligungserklärung“ heißen.

Moderator Jürgen Windeler: Für juristische Hinweise sind wir immer sehr dankbar. Wir bemühen uns, das juristisch korrekt zu machen.

Gibt es weitere Wortmeldungen zu diesem TOP? – Das ist nicht der Fall. Damit rufe ich auf:

3.4.7 Tagesordnungspunkt 6: Verschiedenes

Wer von Ihnen möchte dazu das Wort ergreifen? – Herr Sauerbruch, bitte.

Tilman Sauerbruch: Wir haben in unserer Stellungnahme die Frage aufgeworfen, ob das IQWiG die Auswertung von Daten aus Social Media in Zukunft nicht auch methodisch mit hineinnehmen sollte. Die Corona-App wäre zum Beispiel ein ganz aktuelles Beispiel. Wie gehen Sie da methodisch heran?

Moderator Jürgen Windeler: Wir sind bisher nicht im Entferntesten in die Situation gekommen, uns mit solchen Daten zu beschäftigen, wobei ich schon die Abgrenzung zwischen Social Media und Corona-App schwierig finde. Ich möchte versuchen, eine sibyllinische Antwort zu geben: Auf der einen Seite würden wir mit solchen Daten, wenn wir sie denn hätten und sie in eine Bewertung einbeziehen würden, genauso umgehen wie mit anderen Dingen auch. Und wenn sich dort – das werden wir aber nicht mehr im Methodenpapier 6.0 umsetzen – erstens Bedarf ergibt, uns im Rahmen unserer Aufgaben – das sage ich ausdrücklich dazu – überhaupt mit diesen Daten und deren Auswertung zu beschäftigen, und zweitens auch Notwendigkeiten ergeben, hier methodisch oder prozedural bestimmte Wege zu beschreiten, dann werden wir das sicherlich als Anregung für die nächste Revision des Methodenpapiers aufnehmen. Im Moment sind wir mit solchen Dingen wenig oder eigentlich gar nicht befasst.

Gibt es weitere Punkte? – Bitte.

Kristian Michnacs: Meine Frage unter „Verschiedenes“ bezieht sich auf den Abschnitt 3.4 im Methodenpapier. Da geht es um die nicht medikamentösen Interventionen, und da sprechen Sie auch Digitalprodukte an. Ich habe keine Differenzierung auf Basis der Risikoklasseneinordnung für diese Produktart gesehen, die therapeutisch einsetzbar ist. Deshalb meine Frage zur Klarstellung: Meinen Sie da hohe Risikoklassen, oder was genau meinen Sie? Denn die niedrigen Risikoklassen sind ja der Bewertung des BfArM zugeschrieben. Da wäre eine Klarstellung vielleicht hilfreich.

Stefan Sauerland: Die gesetzlichen Grundlagen sind uns bekannt, in welchen Risikoklassen welche Bewertungspfade möglich sind. Gerade mit dem Digitale-Versorgung-Gesetz, DVG, ist jetzt auch dem BfArM eine Aufgabe in diesem Bereich zugeteilt worden. Das Methodenpapier verstehen wir als ein allgemeines Papier, das erst einmal den Rahmen für Bewertungen insgesamt setzt und eine Möglichkeit für Eventualitäten offenlässt. Denn methodisch ist es kein Unterschied, ob ich jetzt ein digitales Medizinprodukt der Klasse II a oder ein Klasse II b bewerte. Daher wäre aus meiner Sicht die Ergänzung nicht erforderlich, da die gesetzlichen Grundlagen immer gelten und wir das Methodenpapier nicht noch in dieser Hinsicht verkomplizieren müssen, indem wir da noch mal unterteilen.

Moderator Jürgen Windeler: Gibt es weitere Wortmeldungen? – Herr Lützelberger.

Uwe Lützelberger: Danke. – Ich habe eine Frage aus aktuellem Anlass zur Struktur Ihrer Berichte. Das war kein Punkt in unserer Stellungnahme. Mir ist aufgefallen, dass seit Mitte Mai die Nutzenbewertungsberichte des IQWiG zu den Arzneimittelbewertungen in einer verkürzten Version veröffentlicht werden, ohne den Kommentarteil, in dem Sie immer erläutert haben, warum Sie beispielsweise den einen Studienendpunkt akzeptiert haben, den anderen aber nicht. Ist das jetzt Corona-bedingt, oder ist das eine grundsätzliche Änderung Ihrer Vorgehensweise?

Moderator Jürgen Windeler: Ich glaube, es steht explizit in den Berichten drin, dass es Corona-bedingt ist.

Ich sage es hier noch mal: Es ist Corona-bedingt. Es resultiert eigentlich aus den Arbeitsbedingungen, die wir im Rahmen der Pandemie hatten und teilweise noch haben. Das Vorgehen ist mit dem G-BA so abgestimmt. Wir machen es immer noch weiter im Moment; die Pandemie ist ja noch ausgerufen. Aber es ist keine Grundsatzentscheidung, die Berichte grundsätzlich weiterhin so erstellen.

Jens Flintrop: Herr Bierl.

Michael Bierl: Michael Bierl, Flatiron Health. – Ich habe eine Frage zu Kapitel 5.3, „Analyse von Versorgungsdaten“. Insbesondere vor dem Hintergrund des Rapid Reports stellt sich die Frage, inwieweit aus diesem noch Informationen in das Methodenpapier gerade in Bezug auf Datenquellen und die Nutzung von internationalen Daten einfließen.

Moderator Jürgen Windeler: Da die unmittelbar Zuständigen für diesen Bereich jetzt nicht hier sind, schlage ich vor, dass wir das als Hinweis mitnehmen. Sie meinen also den Rapid Report zu versorgungsnahen Daten – der heißt bei uns A19-43 – und fragen, ob die Ergebnisse bzw. Erkenntnisse noch in das Methodenpapier eingebaut werden. Habe ich das richtig verstanden?

Michael Bierl: Genau, ob es da noch einen Abgleich gibt oder ein Einfließen.

Moderator Jürgen Windeler: Das werden wir mal prüfen. Ich werde es jetzt so mitnehmen und die Anregung quasi aufnehmen. Wir werden schauen, inwieweit die Dinge, die im Rapid Report stehen, nicht ohnehin schon im Methodenpapier drinstehen; was Datenquellen angeht, gilt das vielleicht nicht. Ansonsten werden wir aber in dem jetzigen Stadium – das werden Sie sicherlich sofort nachvollziehen können – auch keine grundlegenden Änderungen, neue Kapitel, neue Abschnitte in dieser Version des Methodenpapiers mehr vorsehen. Denn sonst würden einige von Ihnen – vielleicht sogar zu Recht – sagen, wir sollten uns in einem halben Jahr wieder zum Stellungnahmeverfahren treffen. Das wollen wir nicht. Wir werden allerdings noch mal gucken, ob wir nicht doch noch niederschwellige Ergänzungen vornehmen können. Insofern bin ich dankbar für die Anregung.

Michael Bierl: Danke.

Moderator Jürgen Windeler: Herr Werner.

Sebastian Werner: Herr Windeler, wenn ich noch um etwas bitten darf: Sie haben das eben schon wunderbar gesagt, dass Sie das nicht machen wollen. Aber wenn das wirklich einfließen sollte, dann würden wir uns noch mal ein Stellungnahmeverfahren diesbezüglich wünschen.

Moderator Jürgen Windeler: Gut. Sie haben das gehört, was der vfa gesagt hat? – Also, die kommen nicht in das aktuelle Methodenpapier – jedenfalls wenn wir uns dem Wunsch des vfa anschließen, was ich mir noch überlegen werde.

Gibt es weitere Wortmeldungen? – Das sehe ich nicht.

Dann bedanke ich mich ganz herzlich, dass Sie hier waren, dass Sie hierhergekommen sind, dass Sie mitdiskutiert und uns an einigen Stellen jetzt noch mal Hinweise gegeben haben, wie wir das Methodenpapier noch besser machen können. Ich bin auch froh, dass wir dieses Stellungnahmeverfahren pandemiebedingt mit einiger Verzögerung abschließen können, die Endfassung des Methodenpapiers fertigstellen und dann veröffentlichen können, sodass dieser Prozess endlich zu Ende geht und wir uns – nach 6.0 ist vor 7.0; Ralf Bender freut sich schon – an die nächste Stufe der Revision machen können.

Ich erinnere noch mal daran, dass Sie bitte die Teilnahmeerklärung vorne am Counter in den Kasten legen und den Zettel mit Ihren Kontaktdaten auf dem Platz liegen lassen. Die Mitarbeiter des Park Inn werden diese dann einsammeln.

Ansonsten wünsche ich Ihnen allen einen erfreulichen, unkomplizierten Heimweg. Ich bedanke mich dafür, dass Sie da waren, und wünsche Ihnen allen eine gute Zeit.

(Beifall)

Anhang A – Dokumentation der Stellungnahmen

Inhaltsverzeichnis

	Seite
A.1 – Stellungnahmen von Organisationen, Institutionen und Firmen	A 3
A.1.1 – AbbVie Deutschland GmbH & Co. KG	A 3
A.1.2 – AMGEN GmbH	A 9
A.1.3 – Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) e. V.	A 15
A.1.4 – Arzneimittelkommission der deutschen Ärzteschaft (AkdÄ).....	A 20
A.1.5 – AstraZeneca GmbH.....	A 30
A.1.6 – Bayer Vital GmbH.....	A 41
A.1.7 – Boehringer Ingelheim Pharma GmbH & Co. KG.....	A 50
A.1.8 – Bristol-Myers Squibb GmbH & Co. KGaA	A 56
A.1.9 – Bundesverband der Arzneimittel-Hersteller e. V. (BAH).....	A 64
A.1.10 – Bundesverband der Pharmazeutischen Industrie e. V. (BPI)	A 70
A.1.11 – Bundesverband Medizintechnologie e. V. (BVMed)	A 75
A.1.12 – Deutsche Gesellschaft für Gesundheitsökonomie e. V. (dggö)	A 85
A.1.13 – Deutsche Gesellschaft für Hämatologie und Medizinische Onkologie e. V. (DGHO).....	A 91
A.1.14 – Deutsche Gesellschaft für Innere Medizin e. V. (DGIM).....	A 96
A.1.15 – Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e. V. (GMDS) / Internationale Biometrische Gesellschaft – Deutsche Region (IBS-DR) – (gemeinsame Stellungnahme).....	A 103
A.1.16 – Deutsche Gesellschaft für Orthopädie und Unfallchirurgie e. V. (DGOU) / Deutsche Gesellschaft für Orthopädie und Orthopädische Chirurgie e. V. (DGOOC) / Deutsche Gesellschaft für Unfallchirurgie e. V. (DGU) – (gemeinsame Stellungnahme).....	A 107
A.1.17 – Deutsche Gesellschaft für Pneumologie und Beatmungsmedizin e. V. (DGP).....	A 109
A.1.18 – Deutsche Krankenhausgesellschaft e. V. (DKG)	A 112
A.1.19 – Ecker + Ecker GmbH.....	A 121
A.1.20 – Flatiron Health, Inc.	A 129
A.1.21 – Geschäftsstelle der Beauftragten der Bundesregierung für die Belange der Patientinnen und Patienten	A 133
A.1.22 – GlaxoSmithKline GmbH & Co. KG. (GSK)	A 137
A.1.23 – IQVIA Commercial GmbH & Co. OHG.....	A 144
A.1.24 – Janssen-Cilag GmbH.....	A 152
A.1.25 – Kassenzahnärztliche Bundesvereinigung K. d ö. R. (KZBV).....	A 160

A.1.26 – Lilly Deutschland GmbH	A 173
A.1.27 – Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen e. V. (MDS)	A 181
A.1.28 – Merck Serono GmbH	A 193
A.1.29 – MSD Sharp & Dohme GmbH.....	A 201
A.1.30 – Novartis Pharma GmbH.....	A 207
A.1.31 – Novo Nordisk Pharma GmbH.....	A 219
A.1.32 – Pfizer Deutschland GmbH.....	A 224
A.1.33 – Roche Pharma AG.....	A 232
A.1.34 – Sanofi-Aventis Deutschland GmbH.....	A 239
A.1.35 – SmartStep Consulting GmbH.....	A 246
A.1.36 – Takeda Pharma Vertrieb GmbH & Co. KG.....	A 252
A.1.37 – UCB Pharma GmbH	A 257
A.1.38 – Verband Forschender Arzneimittelhersteller e. V. (vfa)	A 265
A.2 – Stellungnahmen von Privatpersonen.....	A 275
A.2.1 – Kohlmann, Thomas	A 275
A.2.2 – Semrau, Frank	A 279

A.1 – Stellungnahmen von Organisationen, Institutionen und Firmen

A.1.1 – AbbVie Deutschland GmbH & Co. KG

Autorinnen und Autoren

- Manzke, Ulrike
- Sternberg, Kati

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Manzke, Ulrike
Sternberg, Kati; Dr.
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
☒ im Namen folgender Institution / Organisation: AbbVie Deutschland GmbH & Co. KG
☐ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

<i>Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.</i>
AbbVie begrüßt die Überarbeitung des Methodenpapiers, würde jedoch aufgrund umfassender methodischer Änderungen, u.a. im Bereich der Beurteilung der klinischen Relevanz der Studienergebnisse, eine Übergangsregelung der Gültigkeiten des Methodenpapiers Version 5.0 und Version 6.0 erstrebenswert finden.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
Abschnitt 3.1.3 „Ermittlung des Schadens medizinischer Interventionen“, S. 49-50	<u>Anmerkung:</u> In Abschnitt 3.1.3 wird die Aufarbeitung und Darstellung von Schadensaspekten in Nutzenbewertungen ergänzt. Das IQWiG führt aus: • „Grundsätzlich werden die Gesamtraten der schwerwiegenden unerwünschten Ereignisse, ggf. der schweren unerwünschten Ereignisse (z. B. solche mit Common-Terminology-Criteria-for-Adverse-Events[CTCAE]-Grad ≥ 3) und der Abbrüche wegen unerwünschter Ereignisse herangezogen. • Zudem werden für die Bewertung potenziell relevante spezifische unerwünschte Wirkungen ausgewählt. Dies erfolgt auf 2 verschiedenen Wegen: - o Zum einen werden solche spezifischen unerwünschten Wirkungen ausgewählt, die für das Krankheitsbild oder die in der Studie bzw. den Studien eingesetzten Interventionen von besonderer Bedeutung sind. Für Nutzenbewertungen gemäß § 139a SGB V ist dies der primäre Weg der Identifizierung solcher unerwünschten Wirkungen. Die Zusammenstellung erfolgt in diesem Fall im Rahmen der Vorabrecherche zur jeweiligen Fragestellung und durch die Einbindung von Betroffenen. - o Zum anderen werden für die Nutzenbewertung relevante spezifische unerwünschte Wirkungen basierend auf den

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	<i>in der relevanten Studie bzw. den relevanten Studien aufgetretenen unerwünschten Ereignissen identifiziert. Dieser Weg ist insbesondere für Nutzenbewertungen von Arzneimitteln gemäß § 35a SGB V (siehe Abschnitt 3.3.3) von Bedeutung. Erstens werden aufgrund des Verfahrens keine Vorabrecherchen durchgeführt und zweitens werden regelhaft neue Arzneimittel bewertet, sodass die Identifizierung unbekannter bzw. unerwarteter unerwünschter Wirkungen hier einen höheren Stellenwert hat. Zur Auswahl werden in diesem Fall die unerwünschten Ereignisse, schwerwiegenden unerwünschte Ereignisse, ggf. schweren unerwünschte Ereignisse zugrunde gelegt, die vom pharmazeutischen Unternehmer gemäß der Anforderungen in der Verfahrensordnung des G-BA im Dossier übermittelt werden [271]. Die Kriterien für die Auswahl stellen die Patientenrelevanz der Ereignisse und die Unterschiede zwischen den Behandlungsarmen dar. Für nicht schwerwiegende bzw. nicht schwere unerwünschte Ereignisse werden zudem ggf. zusätzliche Anforderungen an die Mindesthäufigkeit der Ereignisse gestellt, sofern dies die Gesamtabwägung zum Zusatznutzen nicht beeinflusst.</i> <i>Für die ausgewählten spezifischen unerwünschten Ereignisse wird jeweils die am besten geeignete Operationalisierung in Bezug auf Mess- und Ergebnissicherheit herangezogen. Auch wird geprüft, ob die ausgewählten Ereignisse zum selben inhaltlichen Konstrukt konsistent sind. Liegt keine geeignete Operationalisierung vor oder sind die Ergebnisse nicht konsistent, werden diese in der Regel nicht für Aussagen zum Schaden einer Intervention herangezogen.“</i> AbbVie begrüßt grundsätzlich, dass sich das IQWiG mit der wichtigen Thematik zur Auswahl und Darstellung unerwünschter Ereignisse befasst und Informationen im Abschnitt mit dem Titel „Aufbereitung und Darstellung von Schadensaspekten in Nutzenbewertungen“ im Entwurf des Allgemeinen Methodenpapiers ergänzt wurden. Die sorgfältige Auswahl und Bewertung der unerwünschten Ereignisse sind vor allem für die Patienten von sehr großer Bedeutung. Auch vor dem Hintergrund der Anpassung der Verfahrensordnung des G-BA und des Inkrafttretens der neuen Modulvorlagen ab April 2020, stellt die Auswahl und Darstellung von unerwünschten Ereignissen vor allem bei der Nutzenbewertung nach § 35a SGB V einen essenziellen Themenblock dar (1, 2).

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	Entsprechend den Anforderungen gemäß neuen Modulvorlagen werden von den pharmazeutischen Herstellern eine große Menge an Studienergebnissen zu den unerwünschten Ereignissen aufbereitet, auf deren Basis das IQWiG und der G-BA eine Auswahl an relevanten unerwünschten Ereignissen vornimmt. Die Methodik der Auswahl dieser Ereignisse war aus Sicht von AbbVie sowohl aus den Bewertungen des Instituts als auch aus den bisherigen Methodenpapieren bis jetzt nicht hinreichend nachvollziehbar. Im Entwurf des neuen Methodenpapiers werden nun vom IQWiG die Patientenrelevanz der Ereignisse sowie Unterschiede zwischen den Behandlungsarmen als Auswahlkriterien genannt. Konkrete Angaben zur Größe der Unterschiede werden jedoch nicht gemacht. Weiterhin wird im Methodenpapier ausgeführt, dass bei nicht schwerwiegenden bzw. nicht schweren Ereignissen ggf. zusätzliche Anforderungen an die Mindesthäufigkeit gestellt werden, sofern diese die Gesamtabwägung zum Zusatznutzen nicht beeinflussen. Diese Anforderungen bzw. das methodische Vorgehen werden ebenfalls nicht konkretisiert. Um jedoch eine vollständige, systematische und G-BA- und IQWiG-konforme Aufarbeitung der spezifischen unerwünschten Ereignisse für die Nutzenbewertung zu gewährleisten, sind eindeutige methodische Spezifizierungen zwingend erforderlich. <u>Vorgeschlagene Änderung:</u> Für die Auswahl der unerwünschten Ereignisse sollten die im Entwurf des Methodenpapiers genannten Kriterien, wie die Unterschiede zwischen den Behandlungsgruppen oder Anforderung an die Mindesthäufigkeit konkretisiert und darüber hinausgehende Kriterien genannt werden, um eine differenzierte und nachvollziehbare methodische Herangehensweise zu ermöglichen.
Abschnitt 9.3.3 „Beurteilung klinischer Relevanz“, S. 177-178	<u>Anmerkung:</u> Ergänzungen in Abschnitt 9.3.3 hinsichtlich eines Responsekriteriums von 15 % des Skalenrangs basierend auf Übersichtsarbeiten zu MIDs. AbbVie begrüßt, dass das IQWiG die Problematik hinsichtlich der Bewertung und Akzeptanz von Responseschwellen im Zuge der Nutzenbewertung adressiert und ein pragmatisches Vorgehen zum Umgang mit den Relevanzschwellen anstrebt. Nichtsdestotrotz besteht aus Sicht von AbbVie großer Diskussionsbedarf, da die vom IQWiG vorgeschlagene generische Responseschwelle von 15 % indikationsspezifische Gegebenheiten und Skalencharakteristika nicht berücksichtigt. Es ist nicht

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	hinreichend klar, wie eine vorgeschlagene Responseschwelle mit validierten und etablierten Responseschwellen in etablierten Fragebögen vorangegangener AMNOG Nutzenbewertungen (wie beispielsweise PedsQL, MDASI) in Einklang steht und zum Teil zu einer deutlichen Erhöhung der Schwellenwerte führen kann. Insgesamt ist die Herleitung der Responseschwelle von 15 % aus Sicht von AbbVie nicht hinreichend klar. Das IQWiG benennt für die Begründung dieser Schwelle exemplarisch zwei Übersichtsarbeiten hinsichtlich der MIDs zur Symptomatik „Fatigue“ und zu Patienten mit „rotator cuff disorder“, die auf einer Erhebungsskala eine MID-Spannweite von 10-20 % beschreiben. Vermutlich wird hier mit 15 % der Mittelwert der Spannweite herangezogen. Das Methodenpapier, welches die wissenschaftliche Grundlage respektive Methodik für die Nutzenbewertung darstellt, sollte seinem Sinn und Zweck entsprechend Empfehlungen beschreiben, die auf einer wissenschaftlich nachvollziehbaren Rationale basieren. In diesem Zusammenhang wäre eine nähere Begründung und Diskussion der vorgeschlagenen Schwelle wünschenswert. Insgesamt sollte eine Neuregelung nicht zur Ablehnung von bisher verwendeten und akzeptierten Responseschwellen bei validierten und etablierten Fragebögen / Skalen führen. <u>Vorgeschlagene Änderungen:</u> Aus Sicht von AbbVie sollte eine Festlegung einer generischen Responseschwelle, wie sie im derzeitigen Entwurf des Methodenpapiers vorgeschlagen wird, entfallen. Hier besteht ein großer Bedarf an einem intensiven, wissenschaftlichen Austausch, beispielsweise im Rahmen eines Workshops, mit dem Ziel eines wissenschaftlich nachvollziehbaren Konsens. Bis zur Einigung und Implementierung im Methodenpapier sollten bisher in AMNOG Nutzenbewertungsverfahren akzeptierte validierte MIDs und etablierte Fragebögen / Skalen weiterhin ihre Gültigkeit behalten.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

1. (G-BA) GB. Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Verfahrensordnung (VerfO): Änderung der Anlage II zum 5. Kapitel Berlin 2019.
2. (G-BA) GB. Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Verfahrensordnung (VerfO): Änderung der Anlage II zum 5. Kapitel. Berlin 2019.

A.1.2 – AMGEN GmbH

Autorinnen und Autoren

- Dornstauder, Eugen
- Floßmann, Carmen
- Lebioda, Andrea

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Lebioda, Andrea; Lead Health Economics and Outcomes Research
Floßmann, Carmen; Dr.; Health Economics and Outcomes Research Senior Manager
Dornstauder, Eugen; Health Economics and Outcomes Research Manager
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
☒ im Namen folgender Institution / Organisation: Amgen GmbH
☐ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

<i>Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.</i>

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
9.3.3 (S. 178)	„[...] Eine Sichtung von systematischen Übersichtsarbeiten zu MIDs [z. B. [523,655]] hat ergeben, dass sich MIDs in vielen Fällen zwischen 10 % und 20 % der Spannweite der Scores eines Erhebungsinstruments bewegen. Daraus ergibt sich regelhaft folgendes Vorgehen: 1) Falls in einer Studie Responderanalysen unter Verwendung einer MID präspezifiziert sind und das Responsekriterium mindestens 15 % der Skalenspannweite des verwendeten Erhebungsinstruments entspricht, werden diese Responderanalysen ohne weitere Prüfung des Responsekriteriums für die Bewertung herangezogen. [...]“ <u>Anmerkung:</u> Die Festlegung einer MID von mindestens 15 % der Skalenspannweite als klinisch relevante MID wirkt arbiträr. Es wird daher um eine wissenschaftliche Begründung gebeten. Das vorgeschlagene Vorgehen würde zudem zu einer Inkonsistenz bei allen Verfahren zur frühen Nutzenbewertung führen. Beispielsweise wurde in der Indikation „Multiples Myelom“ bisher eine MID von 10 Punkten (entspricht 10% der Skalenspannweite) zur Beurteilung einer patientenrelevanten Veränderung in den Skalen der Fragebögen EORTC-QLQ-C30 und EORTC-QLQ-MY20 als klinisch relevant akzeptiert (G-BA 2018a, G-BA 2018b, G-BA 2018c, G-BA 2018d; G-BA 2019a; G-BA 2019b, IQWiG 2017, IQWiG 2018, IQWiG 2019). Die Festlegung einer neuen und gleichzeitig höheren MID führt zu einer Inkonsistenz in der zukünftigen Bewertung und Beurteilung des Zusatznutzens im Vergleich zu bereits abgeschlossenen Verfahren. Diese neue MID

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	erhöht deutlich die Hürde, einen patientenrelevanten Vorteil in selbstberichteten Symptomen und der Lebensqualität zeigen zu können und birgt das Risiko einer Benachteiligung gegenüber vorangegangenen Verfahren. <u>Vorgeschlagene Änderung:</u> Es wäre zu begrüßen, wenn aus den oben genannten Gründen, das bisherige Vorgehen einheitlich beibehalten wird.
2.1.1 (S.20)	<i>„Der Berichtsplan wird zur Anhörung gestellt. Die Anhörung erfolgt mittels schriftlicher Stellungnahmen, die innerhalb einer Frist von mindestens 4 Wochen ab Bekanntmachung auf der Website des Instituts abgegeben werden können. Gegenstand der Anhörung ist insbesondere das projektspezifische methodische Vorgehen zur Beantwortung der Fragestellung. Die Fragestellung selbst ist i. d. R. durch den Auftrag vorgegeben und nicht Gegenstand der Anhörung.“</i> <u>Anmerkung:</u> Die aktuellen Ausführungen lassen wenig Spielraum für Diskussion offen. Dies gilt insbesondere für Fälle, in denen die Fragestellung selbst einen zusätzlichen wissenschaftlichen Diskurs erfordert. Im Verfahren „[A19-10] Nutzenbewertung von Bisphosphonaten, Teriparatid und Denosumab zur Behandlung der postmenopausalen Osteoporose“ z.B., erfolgte bereits bei Beauftragung durch den G-BA eine Einschränkung der Population auf Frauen mit postmenopausaler Osteoporose. Dies geschah, obwohl bei Männern mit Osteoporose keine wesentlich unterschiedlichen Therapieempfehlungen bestehen (siehe dazu auch die Ausführungen seitens Amgen in der dezidierten Stellungnahme). In solch gelagerten Fällen wäre es angesichts der vorgeschlagenen Ausführungen im Methodenpapier nicht möglich, die PICO Kriterien einer angemessenen wissenschaftlichen Diskussion zu unterziehen. Dies wird insbesondere durch den Umstand erschwert, dass der Auftrag des G-BA selbst keine Anhörung vorsieht. <u>Vorgeschlagene Änderung:</u> <i>„Der Berichtsplan wird zur Anhörung gestellt. Die Anhörung erfolgt mittels schriftlicher Stellungnahmen, die innerhalb einer Frist von mindestens 4 Wochen ab Bekanntmachung auf der Website des Instituts abgegeben werden können. Gegenstand der Anhörung ist insbesondere das projektspezifische methodische Vorgehen zur Beantwortung der Fragestellung. Die Fragestellung selbst ist i. d. R. durch den Auftrag vorgegeben und nicht Gegenstand der</i>

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<i>Anhörung. Die Ausführungen von Aspekten der Fragestellung selbst können Gegenstand der Anhörung werden, wenn hierzu eine fundierte wissenschaftliche Begründung dargelegt werden kann“.</i>
2.1.1 (S.20)	„Bei der Bearbeitung von Aufträgen zur Bewertung von Untersuchungs- und Behandlungsmethoden nach § 139a Abs. 3 Nr. 1 SGB V findet zum Berichtsplan keine Anhörung statt, sondern lediglich zum Vorbericht [168].“ <u>Anmerkung:</u> Das IQWiG empfiehlt bei Themen zu „Recherche, Darstellung und Bewertung des aktuellen medizinischen Wissensstandes zu diagnostischen und therapeutischen Verfahren bei ausgewählten Krankheiten“ Anhörungen zum Berichtsplan entfallen zu lassen. Eine offene wissenschaftliche Diskussion ist jedoch bereits bei der Planung des Berichts wertvoll. Denn eine reine Anhörung zum Vorbericht erschwert dabei eine Diskussion der grundlegenden Annahmen deutlich. Sollte zudem bei der Diskussion zum Zeitpunkt des Vorberichts ein Konsens entstehen, der die wesentliche Planung und Annahmen betrifft, würde dies auch mit einem deutlichen Mehraufwand für das IQWiG selbst einhergehen. Der Änderungsvorschlag kann daher nicht nachvollzogen werden. <u>Vorgeschlagene Änderung:</u> Der Änderungsvorschlag sollte nicht übernommen werden. An den Anhörungen nach § 139a Abs. 3 Nr. 1 SGB V sollte festgehalten werden.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

1. G-BA, Gemeinsamer Bundesausschuss 2018a. Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V – Carfilzomib (Neubewertung eines Orphan – Drugs nach Überschreitung der 50 Mio. Euro-Grenze); Datum: 15.02.2018 [Zugriff: 07.01.2020]. URL: https://www.g-ba.de/downloads/39-261-3220/2018-02-15_AM-RL-XII_Carfilzomib_D-302_BAnz.pdf.
2. G-BA, Gemeinsamer Bundesausschuss 2018b. Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen

- nach § 35a SGB V – Daratumumab (neues Anwendungsgebiet; Neubewertung eines Orphan Drugs nach Überschreitung der 50 Mio. Euro Grenze); Datum: 15.02.2018 [Zugriff: 07.01.2020]. URL: https://www.g-ba.de/downloads/39-261-3222/2018-02-15_AM-RL_XII_Daratumumab_D-310_BAnz.pdf.
3. G-BA, Gemeinsamer Bundesausschuss 2018c. Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V – Carfilzomib (Neubewertung eines Orphan-Drugs nach Überschreitung der 50 Mio. Euro Grenze); Datum: 15.02.2018 [Zugriff: 07.01.2020]. URL: https://www.g-ba.de/downloads/40-268-4802/2018-02-15_AM-RL-XII_Carfilzomib_D-302_TrG.pdf.
 4. G-BA, Gemeinsamer Bundesausschuss 2018d. Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V – Daratumumab (neues Anwendungsgebiet; Neubewertung eines Orphan Drugs nach Überschreitung der 50 Mio. Euro Grenze); Datum: 15.02.2018 [Zugriff: 07.01.2020]. URL: https://www.g-ba.de/downloads/40-268-4804/2018-02-15_AM-RL_XII_Daratumumab_D-310_TrG.pdf.
 5. G-BA, Gemeinsamer Bundesausschuss 2019a. Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V Pomalidomid (neues Anwendungsgebiet: Kombinationstherapie Multiples Myelom); Datum: 05.12.2019 [Zugriff: 07.01.2020]. URL: https://www.g-ba.de/downloads/39-261-4063/2019-12-05_AM-RL-XII_Pomalidomid_D-456_BAnz.pdf.
 6. G-BA, Gemeinsamer Bundesausschuss 2019b. Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V Pomalidomid (neues Anwendungsgebiet: Kombinationstherapie Multiples Myelom); Datum: 05.12.2019 [Zugriff: 07.01.2020]. URL: https://www.g-ba.de/downloads/40-268-6198/2019-12-05_AM-RL-XII_Pomalidomid_D-456_TrG.pdf.
 7. IQWiG, Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen 2017. IQWiG-Berichte – Nr. 562: Daratumumab (multiples Myelom) – Nutzenbewertung gemäß § 35a SGB V; Datum: 13.11.2017 [Zugriff: 29.01.2020]. URL: https://www.g-ba.de/downloads/92-975-2047/2017-08-15_Nutzenbewertung-IQWiG_Daratumumab-D-310.pdf.
 8. IQWiG, Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen 2018. IQWiG-Berichte – Nr. 588: Carfilzomib (multiples Myelom) – Addendum zum Auftrag A17-38; Datum: 01.02.2018 [Zugriff: 29.01.2020]. URL: https://www.g-ba.de/downloads/92-975-2183/2018-02-15_AM-RL-XII_Carfilzomib_D-302_Addendum.pdf.
 9. IQWiG, Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen 2019. IQWiG-Berichte – Nr. 814: Pomalidomid (multiples Myelom) – Nutzenbewertung gemäß § 35a SGB V; Datum: 12.09.2019 [Zugriff: 29.01.2020]. URL: https://www.g-ba.de/downloads/92-975-3117/2019-06-15_Nutzenbewertung-IQWiG_Pomalidomid-D-456.pdf.

A.1.3 – Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) e. V.

Autorinnen und Autoren

- Blödt, Susanne
- Kopp, Ina B.
- Kreienberg, Rolf
- Nothacker, Monika
- Spies, Claudia

Berlin, 31.01.2020

**Stellungnahme
der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen
Fachgesellschaften (AWMF)
zum Entwurf des IQWiG „Allgemeine Methoden -
Entwurf für Version 6.0“ vom 05.12.2019**

Die AWMF wurde am 05.12.2019 um eine Stellungnahme zu dem oben genannten Entwurf gebeten. Die AWMF hat ihrerseits ihre Mitgliedsfachgesellschaften gebeten, bei Bedarf eine eigene Stellungnahme zu verfassen. Die bis einschließlich 31.01.2020 bei der AWMF eingegangenen vier Stellungnahmen von insgesamt sieben Fachgesellschaften sind der Stellungnahme der AWMF als Anlage beigelegt (s. Anlage 1).

Die AWMF begrüßt die regelmäßige Überarbeitung der Allgemeinen Methoden des IQWiG, da hierdurch aktuellen methodischen Entwicklungen Rechnung getragen werden kann (s.a. Stellungnahme von GMDS & IBS-DR). In Bezug auf die Abgabe von Stellungnahmen ist festzustellen, dass die Wissenschaftlichen Medizinischen Fachgesellschaften aktuell in einer hohen Dichte für Stellungnahmen angefragt werden und die Ressourcen dafür bei den ehrenamtlich arbeitenden Mitgliedern begrenzt sind.

I. Allgemeine Anmerkungen

Da das IQWiG erfreulicherweise künftig regelhaft Leitliniengruppen der Wissenschaftlichen Medizinischen Fachgesellschaften in der AWMF durch die Erstellung von Evidenzberichten unterstützt, halten wir Ausführungen zur geplanten Vorgehensweise in Ergänzung zu den allgemeinen Methoden für hilfreich. Seitens der AWMF sollte hier neben der Verwendung einer international anerkannten Methodik der Evidenzbewertung (GRADE) vor allem der klinisch-methodische Austausch zwischen der betreffenden Leitliniengruppe und Mitarbeitern des IQWiG für jeden Arbeitsschritt festgeschrieben werden. Dieses Vorgehen hat sich in dem 2017-2019 durchgeführten „Pilotprojekt“¹ bewährt.

¹ Erstellung von Evidenzberichten für die S3 Leitlinie „Vaginale Geburt am Termin“

II. Stellungnahme zu spezifischen Aspekten

Kapitel/ Abschnitt im Entwurf	Stellungnahme mit Begründung sowie vorgeschlagene Änderung
2.1.1 Produktspezifische Verfahrensabläufe – Bericht	<u>Anmerkung:</u> Die AWMF bedauert, dass die Stellungnahme- möglichkeit zu Berichtsplänen bei Verfahren nach §139 künftig wegfallen soll <u>Vorgeschlagene Änderung:</u> Stellungnahmemöglichkeit beibehalten
2.2.3: Darlegung von Interessen und Umgang mit Interessenkonflikten	<u>Anmerkung:</u> Die AWMF hält eine Harmonisierung mit internationalen (G-I-N) und nationalen (AWMF) Entwicklungen in Bezug auf die Nomenklatur und eine Vereinheitlichung des Interessen- erklärungsformulars für erforderlich. Die Literatur zum Thema sollte aktualisiert werden. <u>Vorgeschlagene Änderung:</u> „Formblatt zur Offenlegung potenzieller Interessenkonflikte“ ändern in „Formblatt zur Offenlegung von Interessen“. Vermeidung des Ausdrucks „potentielle Interessenkonflikte“, da Interessenkonflikte per se nur ein Risiko einer unangemessenen Beeinflussung darstellen Aufnahme u.a. der Literaturstelle: Schünemann HJ, Al-Ansary LA, Forland F, Kersten S, Komulainen J, Kopp IB, Macbeth F, Phillips SM, Robbins C, van der Wees P, Qaseem A; Board of Trustees of the Guidelines International Network.Guidelines International Network: Principles for Disclosure of Interests and Management of Conflicts in Guidelines. Ann Intern Med. 2015 Oct 6;163(7):548-53. doi: 10.7326/M14-1885.
3.14 Endpunktbezogene Bewertung	<u>Anmerkung:</u> Das IQWiG führt aus: Die Nutzenbewertung und die Einschätzung der Stärke der Ergebnis(un)sicherheit orientieren sich an internationalen Standards der evidenzbasierten Medizin, wie sie z. B. von der GRADE-Gruppe erarbeitet werden [26]. Aus Sicht der AWMF stellt das GRADE System derzeit am ehesten den internationalen Standard der Einschätzung des Vertrauens in einen Effektschätzer dar. <u>Vorgeschlagene Änderung:</u> Das IQWiG verwendet zur Nutzenbewertung und der Einschätzung der Stärke der Ergebnis(un)sicherheit den internationale Standard in der evidenzbasierten Medizin, wie er von der GRADE Gruppe erarbeitet wurde. Als erforderlich gesehene Abweichungen oder zusätzliche Regelungen werden begründet
5.1.4 Bewertung der methodischen Leitlinienqualität	<u>Anmerkung:</u> Dass das IQWiG AGREE II zur methodischen Beurteilung der Leitlinienqualität einsetzt, wird weiterhin begrüßt. Aufgrund des Zweckes der erstellten Leitliniensynopsen ist Eingrenzung auf die beschriebenen 3 Domänen nachvollziehbar. Auch ist es zu begrüßen, dass das IQWiG nun über die rein deskriptive methodische Beurteilung der eingeschlossenen Leitlinien hinausgeht und in den Leitliniensynopsen Leitlinien, die in 1 oder mehreren der 3 betrachteten Domänen standardisierte Domänenwerte < 30 % erreichen, markiert, bzw. auch die Empfehlungen, die ausschließlich auf solchen Leitlinien beruhen.

	Allerdings zeigen bisherige Auswertungen, dass die Interrater-Reliabilität von AGREE II nicht hoch ist, was ggf. der 7 stufigen Einteilung ohne feste Zuordnungen geschuldet ist. Die VerfasserInnen einer systematischen Übersicht über den internationalen Umgang mit AGREE II Bewertungen (Hoffmann-Eßer W, Siering U, Neugebauer EAM, Lampert U, Eikermann M, Systematic review of current guideline appraisals performed with the Appraisal of Guidelines for Research & Evaluation II instrument-a third of AGREE II users apply a cut-off for guideline quality.J Clin Epidemiol. 2018 Mar;95:120-127. doi: 10.1016/j.jclinepi.2017.12.009.) kommen zu dem Schluss: „Many users might welcome a clear distinction between high- and low-quality guidelines; specifying a cut-off for this purpose might be useful.“ <u>Vorgeschlagene Änderung:</u> Die AWMF überarbeitet derzeit ihre Regeln für evidenzbasierte Leitlinien und wendet dabei einen solchen cut-off an. Dabei werden nicht %Anteile gewertet, die eine stärkeren subjektiven Beurteilung unterliegen, sondern konkrete methodische Vorgehensweisen abgefragt. Wir schlagen vor, dass das IQWiG prüft, ob es diesen übernehmen kann, wir übermitteln die aktualisierten Regeln baldmöglichst. Für das Methodenpapier könnte aktuell eine Formulierung eingefügt werden, die die Anwendung eines fest definierten cut-offs ermöglicht.
--	--

III. Stellungnahmen der Mitgliedsfachgesellschaften

Die beigegefügtten Stellungnahmen von Mitgliedsfachgesellschaften äußern sich auch zu weiteren Aspekten, die wir ebenfalls zu berücksichtigen bitten. Diese betreffen spezifische methodische Fragestellungen (siehe Stellungnahmen der GMDS-IB-DR und der DGP).

Für die Frühe Nutzenbewertung von Arzneimitteln und die Kommentierungs-/Beteiligungsmöglichkeit bei Addenda sowie die Einschätzung zu den Ausführungen des IQWiG zu „dramatischen Effekten“ verweisen wir auf die Stellungnahme der DGHO.

Als allgemeine Forderung möchten wir hervorheben, dass cut-off Werte oder Mindestzahlen gut begründet werden müssen, auch die Abweichung von gängiger klinischer Praxis (s.a. Stellungnahme der DGHO, DGP und GMDS-IBDR).

Desweiteren verweisen wir in Bezug auf den Umgang mit klinischer Unsicherheit sowie die ggf. erforderliche Berücksichtigung von Fallberichten auf die Stellungnahme der DGIM.

Das ausreichende Einbringen von Fachexpertise unterstützen wir weiterhin als unverzichtbar (s. Stellungnahme der DGIM und der DGOU-DGOOC-DGU). Dafür sind ausreichende zeitliche und finanzielle Ressourcen bereitzustellen.

Für Rückfragen stehen wir jederzeit gerne zur Verfügung.

Dr. med. Monika Nothacker, MPH [REDACTED]

Prof. Dr. med. Ina B. Kopp [REDACTED]

Dr. rer. medic. Susanne Blödt [REDACTED]

Prof. Dr. med. Claudia Spies [REDACTED]

Prof. Dr. med. Rolf Kreienberg [REDACTED]

Anlage 1: Stellungnahmen der Fachgesellschaften

1. Gemeinsame Stellungnahme Deutsche Gesellschaft für Orthopädie und Unfallchirurgie (DGOU), Deutsche Gesellschaft für Orthopädie und Orthopädische Chirurgie (DGOOC), und Deutsche Gesellschaft für Unfallchirurgie (DGU)
2. Deutsche Gesellschaft für Pneumologie und Beatmungsmedizin (DGP)
3. Gemeinsame Stellungnahme Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS) und Deutsche Region der Internationalen Biometrischen Gesellschaft (IBS-DR)
4. Deutsche Gesellschaft für Hämatologie und Medizinische Onkologie (DGHO)

Information, dass keine Abgabe einer Stellungnahme erfolgt:

1. Deutsche Gesellschaft für Nephrologie (DGfN)
2. Deutsche Gesellschaft für für Thorax-, Herz- und Gefäßchirurgie (DGTHG)
3. Deutsche Gesellschaft für Hals-Nasen-Ohren-Heilkunde, Kopf- und Halschirurgie (DGHNO-KC)

A.1.4 – Arzneimittelkommission der deutschen Ärzteschaft (AkdÄ)

Autorinnen und Autoren

- Arzneimittelkommission der deutschen Ärzteschaft (AkdÄ)

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Arzneimittelkommission der deutschen Ärzteschaft (AkdÄ) Herbert-Lewin-Platz 1, 10623 Berlin, www.akdae.de
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
☒ im Namen folgender Institution / Organisation: AkdÄ
☐ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
3.1.3 Aufarbeitung und Darstellung, S. 49	<u>Anmerkung:</u> Folgende Passage wurde aus der Vorversion gestrichen: „Die Zusammenstellung erfolgt im Rahmen der Vorabrecherche zur jeweiligen Fragestellung insbesondere auf Grundlage der Daten kontrollierter Interventionsstudien, in denen zielgerichtet der Nutzen der Intervention untersucht wurde, sowie ggf. auf Basis vorliegender epidemiologischer Studien (zum Beispiel Kohorten- oder Fall-Kontroll-Studien), von Pharmakovigilanzdaten, Informationen von Zulassungsbehörden etc. Im Einzelfall können hier auch Ergebnisse aus Tierexperimenten sowie aus Experimenten zur Überprüfung eines pathophysiologischen Konstrukts hilfreich sein. Diese Zusammenstellung potenziell relevanter unerwünschter Wirkungen bildet die Grundlage für die Bewertung des Schadens auf Basis der in die Nutzenbewertung eingehender Studien.“ Diese Passage sollte beibehalten und die aufgeführten Quellen weiter verwendet werden, da in der Praxis oft sehr wichtige Daten nicht in das Dossier aufgenommen werden, wie z. B. Todesfälle aufgrund von unerwünschten Wirkungen. Diese Daten finden sich dann in anderen Quellen (s. o.), z. B. im EPAR. <u>Vorgeschlagene Änderung:</u> Beibehaltung der alten Passage zu den Quellen.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
3.1.3 Aufarbeitung und Darstellung, S. 49	<u>Anmerkung:</u> Zu Spiegelstrich 2, Unterpunkt 2: „Zum anderen werden für die Nutzenbewertung relevante spezifische unerwünschte Wirkungen basierend auf den in der relevanten Studie bzw. den relevanten Studien aufgetretenen unerwünschten Ereignissen identifiziert. Dieser Weg ist insbesondere für Nutzenbewertungen von Arzneimitteln gemäß § 35a SGB V (siehe Abschnitt 3.3.3) von Bedeutung. Erstens werden aufgrund des Verfahrens keine Vorabrecherchen durchgeführt und zweitens werden regelhaft neue Arzneimittel bewertet, sodass die Identifizierung unbekannter bzw. unerwarteter unerwünschter Wirkungen hier einen höheren Stellenwert hat. Zur Auswahl werden in diesem Fall die unerwünschten Ereignisse, schwerwiegenden unerwünschte Ereignisse, ggf. schweren unerwünschte Ereignisse zugrunde gelegt, die vom pharmazeutischen Unternehmer gemäß der Anforderungen in der Verfahrensordnung des G-BA im Dossier übermittelt werden [271]. ...“ Bei den spezifischen unerwünschten Ereignissen (spezifische UE) nur auf die für die jeweils anstehende Nutzenbewertung relevante/n Studie/n zurückzugreifen und sich nur auf Berichte des pharmazeutischen Unternehmers zu verlassen, erscheint problematisch und unzureichend: 1. werden nicht nur „regelhaft neue Arzneimittel“ bewertet, sondern mittlerweile oft auch bereits bekannte in neuen oder erweiterten Indikationen und 2. gibt es für viele „neue Arzneimittel“ spezifische UEs, die mehr oder weniger zu erwarten sind, wenn diese beispielsweise einer schon bekannten Wirkstoffklasse angehören. Aktuell trifft dies sicher für viele „neue“ Immuntherapeutika in der Onkologie zu. Hier vollständig auf eine „Vorrecherche“, wie sie regelhaft für allgemeine Nutzenbewertungen gemäß § 139a SGB V vorgesehen ist (siehe Spiegelstrich 2, Unterpunkt 1), zu verzichten, könnte zu kurz gegriffen sein, zumal wie bereits oben beschrieben wichtige Daten in den Dossiers oft fehlen. <u>Vorgeschlagene Änderung:</u> Vorrecherche beibehalten/durchführen.
3.1.3 Aufarbeitung und Darstellung, S. 50	<u>Anmerkung:</u>

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	„Für die ausgewählten spezifischen unerwünschten Ereignisse wird jeweils die am besten geeignete Operationalisierung in Bezug auf Mess- und Ergebnissicherheit herangezogen. Auch wird geprüft, ob die ausgewählten Ereignisse zum selben inhaltlichen Konstrukt konsistent sind. Liegt keine geeignete Operationalisierung vor oder sind die Ergebnisse nicht konsistent, werden diese in der Regel nicht für Aussagen zum Schaden einer Intervention herangezogen.“ Bei der Operationalisierung von unerwünschten Ereignissen ist darauf Wert zu legen, dass dies a priori und nicht post-hoc geschieht. Ferner ist immer auch die kumulative Inzidenz unter Berücksichtigung des Schweregrades der Nebenwirkungen zu analysieren. Die Zeit bis zum Auftreten von unerwünschten Ereignissen zum Beispiel ist bei immunvermittelten Nebenwirkungen typischerweise länger als bei klassischen Chemotherapeutika. Dies darf nicht mit einem grundsätzlichen Vorteil verwechselt werden. Die Beobachtungsdauer muss so gewählt werden, dass immunvermittelte Nebenwirkungen trotz typischerweise späterem Auftreten sicher erfasst und bewertet werden. Die Bewertung von unerwünschten Ereignissen muss die Patientenperspektive berücksichtigen. PROs (Patient-reported Outcomes) sind als wichtige Voraussetzung für die Beurteilung patientenrelevanten Zusatznutzens zu fordern. <u>Vorgeschlagene Änderung:</u> Überprüfung der Validität neu vorgeschlagener Operationalisierungen.
3.2.1 Auswirkung nicht publizierter Studienergebnisse auf das Fazit Tabelle 4, S. 56/57	<u>Anmerkung:</u> Die Erweiterung des Abschnitts zu möglichen Konsequenzen bei unvollständigen Daten erscheint grundsätzlich sinnvoll. Jedoch: 1. Tabelle 4 erscheint für sich allein (ohne den erklärenden Text) schwer verständlich, z. B. „Datenübermittlung durch Dritte“ in der ersten Zeile. Wer ist „Dritter“? Geht es nicht um eine Antragstellerin (die eine Intervention bewertet haben möchte)? Muss es immer einen Hersteller geben (beispielsweise wenn es nicht um Arzneimittel, Medizinprodukte oder Diagnostika geht – sondern bspw. um OP-Methoden)? 2. Die Kopfnote b im Szenario 4 ist im Begleittext nicht genügend adressiert.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	3. Die einleitenden Sätze für Szenario 5–7 (3. Absatz S. 57) sind etwas schwer verständlich. 4. Es stellt sich die Frage, ob beide Grundszenarien (Bewertung mit bzw. ohne Datenübermittlung) in nur einer Tabelle darstellbar sind; die Überschrift „Nachweis, dass insgesamt Daten fehlen“ passt eigentlich nur zur Bewertung mit Datenübermittlung – im Falle ohne Datenübermittlung wäre „Nachweis, dass Daten vorhanden sind, die nicht verfügbar sind“ treffender. 5. Der letzte Absatz von 3.2.1 „In bestimmten Fällen (...) für die Bewertung nicht zu berücksichtigen, sofern diese Daten durch Publikationsbias gefährdet sind“ erscheint problematisch. Wenn der Nachweis erbracht ist, dass vorhandene Daten nicht (adäquat) verfügbar sind, muss als Option auch genannt werden, dass eine Bewertung anhand der verfügbaren Daten gar nicht möglich ist (beispielsweise wenn ein Großteil der Daten durch Publikationsbias gefährdet ist). <u>Vorgeschlagene Änderung:</u> 1. Besser beispielsweise „Datenübermittlung aus Sicht des Antragstellers ...“ und 2. Kopfnote b im Szenario 4 im Text ergänzen. 3. Besser die ersten Sätze im 3. Absatz S. 57: „Bestehen bei einer Nutzenbewertung ohne Datenübermittlung durch den Antragsteller keine Hinweise darauf, dass vorhandene Daten nicht verfügbar sind, ...“ und „Lässt sich jedoch nachweisen, dass Daten vorhanden sein müssen, die nicht verfügbar sind (z. B. durch ...)“. 4. Zwei Tabellen. 5. Entsprechende Ergänzung.
3.3.3 Binäre Zielgrößen, S. 65	<u>Anmerkung:</u> Für bestimmte Krankheitsbilder mit einer sehr geringen oder sehr hohen Mortalität sind die Grenzen zur Bewertung des Ausmaßes nie zu erreichen. Wenn die Mortalität unter einer Standardtherapie bei 100 % liegt und unter einer neuen Intervention nur noch bei 90 % wäre das relative Risiko 0,9, die obere Grenze des Konfidenzintervalls läge bei n = 200 Patienten bei (0,84;0,96), das bedeutet einen geringen Zusatznutzen, obwohl es hier doch eine erhebliche Verbesserung gibt. Es sollte ein Satz eingefügt werden, der die Bewertung des Ausmaßes eines Zusatznutzens bei zu erwartenden Ereignisraten im Randbereich (nahe 0 oder 1) relativiert.
3.3.3 Stetige oder quasistetige Zielgrößen mit	<u>Anmerkung:</u>

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
jeweils vorliegenden SMDs, S. 67	Unter D) ist neu aufgenommen, wie bei stetigen Effektmaßen mit vorliegender SMD (Standardized Mean Difference) eine Kategorisierung der Zusatznutzens über die SMD in erheblich, beträchtlich und gering erfolgen soll. 1. Hierbei wird zunächst zugrunde gelegt, dass bei Verwendung von Cohen's d (oder Hedges' g) als Effektmaß für die SMD ein Bereich zwischen 0,2–0,5 kleine, ein Bereich zwischen 0,5–0,8 mittlere und Bereiche > 0,8 große Effekte darstellen. - Cohen's d oder Hedges' g sind nur zwei Beispiele von Maßen für die Effektstärke bei stetigen Zielgrößen. - Die auf Cohen zurückgehenden Kriterien für kleine, mittlere und große Effekte sind nicht unumstritten – und gelten als empirisch schwach belegt. - Für einige Forschungs- oder Anwendungsbereiche werden Kategorisierung der Effektstärke kleinere Maßzahlen herangezogen. - Andererseits dürften Cohen's d oder Hedges' g und die genannten Kriterien in praxi die am häufigsten verwendeten sein. 2. Für die SMDs werden Schwellenwerte definiert, die unter Berücksichtigung der Streuung für die SMD erreicht werden müssen, um die Kriterien für kleine, mittlere und große Effekte zu erfüllen. Dazu werden die Konfidenzintervalle der SMD herangezogen (analog zum Vorgehen bei binären Maßen). - Es ist nicht klar, welche Konfidenzintervalle gemeint sind – vermutlich sind die 95 % Konfidenzintervalle gemeint. - Laut Text kann man „zeigen, dass die für ein bestimmtes Ausmaß wünschenswerte Effektstärke ungefähr halbiert werden muss, um den entsprechenden Schwellenwert abzuleiten“ (...oberhalb dessen das 95 % Konfidenzintervall liegen muss). Eine Begründung oder Herleitung für dieses Vorgehen fehlt. Es wird nur auf das Verfahren nach Skipka 2016 für binäre Daten verwiesen. Dort werden Schwellenwerte für stetige Daten aber gar nicht thematisiert oder auch nur mitdiskutiert. Durch was soll eine Halbierung erfolgen? Cohen's d ist ein Maß für die Effektstärke zwischen Intervention und Kontrolle. Man kann sich fragen, wie stark die Intervention den standardisierten Mittelwert im Kontrollarm reduzieren muss, aber in Cohen's d gehen bereits Mittelwert von Intervention und Kontrolle ein. Der Satz sollte umformuliert werden. Die Klassifikationen von Cohen's d in kleine, mittlere und starke Effekte ohne einer inhaltlichen Betrachtung des Endpunkts und des Krankheitsbilds erlauben keine sinnvolle Beurteilung der

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>									
	Ausmaßkategorie. Diese Aussage trifft auf alle anderen Endpunkttypen ebenso zu. - Die tatsächliche Formulierung der Schwellenwerte (Tabelle 6) zur Ausmaßbestimmung für die SMD geht nicht konform mit den eigenen Vorgaben (Halbierung der gewünschten Effektstärke): Danach müssten die Schwellen für <table><tr><td>kleine Effekte</td><td>0,2–0,5 x 0,5</td><td>0,175</td></tr><tr><td>mittlere Effekte</td><td>0,5–0,8 x 0,5</td><td>0,325</td></tr><tr><td>große Effekte</td><td>> 0,8 x 0,5</td><td>0,4</td></tr></table> lauten – und so auch die Schwellen bei der Zielgrößenkategorie „schwerwiegende Symptome ...“ für die Ausmaßkategorien gering, beträchtlich und erheblich. Außer bei den großen Effekten (= erheblich) sind die Abweichungen zwar nur gering (und die Ausmaßkategorien in der AmNutzen-V ohnehin nur sehr unscharf definiert); dennoch ist der Grund für die Abweichung unklar. - Für die Zielgrößenkategorie „nicht schwerwiegende Symptome ...“ wird die Schwelle bei der Ausmaßkategorie beträchtlich deutlich höher gesetzt (0,5), bei der Kategorie gering jedoch wie bei „schwerwiegende Symptome ...“ (0,2). Höhere Schwellen erscheinen hier zwar plausibel (wie bei binären Daten) – die konkrete Umsetzung hier jedoch nicht bzw. eher arbiträr. <u>Vorgeschlagene Änderung:</u> Genauere Begründung des Vorgehens, Umformulierungen.	kleine Effekte	0,2–0,5 x 0,5	0,175	mittlere Effekte	0,5–0,8 x 0,5	0,325	große Effekte	> 0,8 x 0,5	0,4
kleine Effekte	0,2–0,5 x 0,5	0,175								
mittlere Effekte	0,5–0,8 x 0,5	0,325								
große Effekte	> 0,8 x 0,5	0,4								
9.3.3 Beurteilung klinischer Relevanz, S. 177–178	<u>Anmerkung:</u> Das IQWiG präzisiert vor diesem Hintergrund seine Hierarchie zum Umgang mit MIDs (Minimal important Difference) für die Responderanalysen: 1. Responderanalysen mit MIDs ≥ 15 % der Skalenspannweite werden akzeptiert. 2. Responderanalysen mit MIDs < 15 % der Skalenspannweite werden nicht akzeptiert; in diesem Fall können die stetigen Daten herangezogen werden, wobei dann eine Irrelevanzschwelle von 0,2 für die SMD gilt (s. o.), d. h. das 95 % Konfidenzintervall der SMD muss vollständig $> 0,2$ sein. Alternativ werden post hoc durchgeführte Responderanalysen mit einer MID von 15 % der Skalenspanne akzeptiert. 3. Liegen Responderanalysen mit o. g. MIDs und auch Analysen der stetigen Daten mittels SMD vor, werden erstere bevorzugt angewendet.									

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	- Der Ansatz erscheint auf Basis der publizierten Literatur zur Bewertung von MIDs plausibel und ein begründeter Kompromiss, allerdings auch arbiträr. Simple Beispiele: ADAScog mit einer Skalenbreite von 70: Hier würde für eine Responderanalyse eine MID von $\geq 10,5$ gefordert, d. h. zwischen den Gruppen wird der Anteil der Patienten mit einer Änderung von $\geq 10,5$ Punkten im ADAScog miteinander verglichen. Hier wird allgemein (z. B. FDA) davon ausgegangen, dass bei den Mittelwertdifferenzen schon Unterschiede von 3 oder 4 Punkten als relevant anzusehen sind. VAS z. B. zu Schmerzen mit einer Skalenbreite von 10 bzw. 100: Hier würde für Responderanalysen eine MID von $\geq 1,5$ bzw. ≥ 15 gefordert. EDSS für MS mit einer Skalenbreite von 10: Hier würde für die Responderanalysen ebenfalls eine MID von $\geq 1,5$ gefordert. Die „Forderungen“ des IQWiG erscheinen somit „streng“ beim Vergleich mit den Mittelwertdifferenzen, die bei Verwendung der Skalen üblicherweise als relevant akzeptiert werden. Außerdem kann infrage gestellt werden, dass das Heranziehen der gesamten Skalenbreite für die Bestimmung der MID unter klinischen Aspekten Sinn ergibt: Erfahrungsgemäß weisen die Patienten, die für die verschiedensten Indikationen in Studien eingeschlossen werden, die Skalen zur Erfassung von Therapieeffekten verwenden, in aller Regel Skalenwerte im mittleren Bereich auf – zumindest sind Patienten mit Extremwerten am oberen oder unteren Ende von Skalen zu bspw. Symptomen äußerst selten. <u>Vorgeschlagene Änderung:</u> Interessant wären empirische Daten, wie sich die mithilfe der MID ermittelten Ergebnisse für einen Response zu Ergebnissen verhalten, die sich mithilfe der Mittelwertdifferenzen ergeben würden.
9.3.12 Umgang mit unvollständigen Daten S. 191–192	<u>Anmerkung:</u> Es ist unklar, warum das Problem unvollständiger Werte nur auf Follow-Up-Verluste bezogen wird. Fehlwerte können auch bei der Baseline-Erhebung entstehen, etwa wenn Messungen vergessen werden oder Proben verloren gehen. Auch deutet die Formulierung „bei jeder Nachbeobachtung innerhalb einer Studie kann es unvermeidlich passieren, dass Probandinnen und Probanden für die Datenerhebung zu Endpunkten nicht mehr zur Verfügung stehen“ darauf hin, dass Fehlwerte ausschließlich auf das Nichterscheinen eines Patienten zurückzuführen sind. Tatsächlich gibt es viele Gründe für Fehlwerte, die nicht dem Patienten selbst zuzuschreiben sind. Auch wenn der Patient zum Follow-Up

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	erscheint, können einzelne Merkmale aus verschiedenen Gründen nicht erhoben werden.
9.3.13 Umgang mit unterschiedlichen mittleren Beobachtungsdauern S. 192–193	<u>Anmerkung:</u> Informative Zensierungen entstehen nicht nur bei unvollständiger Datenerfassung. Wenn in einer Behandlungsgruppe mehr Menschen versterben, so sind diese Todesfälle als informative Zensierungen für alle nicht fatalen Todesereignisse zu bewerten. Grundsätzlich sollte für die Bewertung von UEs und schwerwiegende UEs (SUE) Methoden der Überlebenszeitanalysen angewandt werden, die auch für informative Zensierungen geeignet sind. Diese Empfehlung sollte sich nicht nur auf den Fall unterschiedlicher mittlerer Beobachtungsdauern beschränken. <u>Vorgeschlagene Änderung:</u> Insgesamt erscheinen die Formulierungen in diesem Abschnitt sehr vage. Es müssten konkretere Vorgaben formuliert werden.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

A.1.5 – AstraZeneca GmbH

Autorinnen und Autoren

- Bergner, Sven
- Lützelberger, Uwe

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Uwe Lützelberger; Director Payer Excellence
Dr. Sven Bergner; Director Medical Benefit Assessment & RWE Oncology
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
☒ im Namen folgender Institution / Organisation: AstraZeneca GmbH
☐ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

<i>Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.</i>

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
Abschnitt 9.3.3 „Beurteilung klinischer Relevanz“, S. 177/178	<u>Anmerkung:</u> In Abschnitt 9.3.3 wird auf den Seiten 177/178 ein neues Vorgehen zur Durchführung von Responder-Analysen eingeführt. Das IQWiG führt aus, dass es bei der Bestimmung der individuellen Minimal important Difference (MID), welche den Ausgangspunkt für Responderanalysen darstellt, vielfältige methodisch bedingten Unsicherheiten auftreten. Als Konsequenz sieht das IQWiG nun einen generellen Wert von 15% der Skalenspannweite des jeweiligen Erhebungsinstrumentes als MID vor; dieser kann überschritten (aber nicht unterschritten, auch dann nicht, wenn bereits eine validierte MID vorliegt) werden, sofern Responderanalysen mit einer entsprechenden MID präspezifiziert sind. Betroffen hiervon wären ein Großteil der Analysen zur Morbidität (bspw. patientenberichtete Symptomatik oder Gesundheitszustand) und zur gesundheitsbezogenen Lebensqualität. In der Konsequenz würde die vorgeschlagene neue Regelung bedeuten, dass Analysen, wie sie bislang vom IQWiG und vom G-BA im Rahmen von Nutzenbewertungen nach § 35a SGB V zur Bestimmung des (Netto-) Zusatznutzens herangezogen wurden, für die überwiegende Mehrheit der verwendeten Fragebogeninstrumente und Skalen künftig nicht mehr berücksichtigt würden. Für zahlreiche Messinstrumente würden so die in der Wissenschaft allgemein hin akzeptierten (validierten) MCIDs ignoriert. Betroffen wären davon auch so etablierte Messinstrumente wie der SF-36, die EQ-5D-VAS, die EORTC-

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	QLQ- sowie die FACT-Fragebögen, der FACIT-Fatigue oder bspw. auch die seit vielen Jahren etablierten Instrumente für den Bereich der Atemwegserkrankungen (AQLQ, SGRQ, ...). Vor diesem Hintergrund ist dies mehr als nur eine „Akzentverschiebung“, wie vom IQWiG auf seiner Website unterstellt. Der Vorschlag des IQWiG ist aus mehreren Gründen nicht akzeptabel: • MCIDs sind u.a. abhängig vom Skalenkonstrukt (bzw. vom Fragebogenkonstrukt), vom Gegenstand der Messung (Endpunkt), ggf. von der Effektrichtung (Verbesserung oder Verschlechterung) und ggf. von anderen Faktoren (bspw. vom Leidensdruck in der jeweiligen Indikation). Eine einzige „Universal-MCID“ wird diesem Umstand nicht gerecht. Die individuellen Schwellenwerte für eine klinische Relevanz können somit niedriger oder auch höher als 15% liegen. • Die vom IQWiG vorgeschlagene konkrete MCID von 15% der Skalenspannweite ist ein arbiträr festgesetztes Kriterium und soll eine Art „Durchschnitt“ über die MCIDs sämtlicher Messinstrumente über alle Indikationen suggerieren. Letzteres trifft allein schon deshalb nicht zu, weil die vom IQWiG als Quellen herangezogenen Übersichtsarbeiten auf ausgewählte Anwendungsgebiete beschränkt sind. • Die ausschließliche Akzeptanz einer einzigen „Universal-MCID“ bedeutete eine willkürliche normative Vorgabe für PRO-Analysen im Rahmen der Nutzenbewertung nach § 35a SGB V. Es darf bezweifelt werden, dass dies im Einklang mit der gesetzlichen Anforderung steht, wonach die Nutzenbewertung auf Basis des aktuellen Stands der wissenschaftlichen Erkenntnisse sowie der international anerkannten Kriterien und Standards der evidenzbasierten Medizin zu erfolgen habe. Denn zum einen existieren wissenschaftlich validierte – und bislang auch vom IQWiG selbst sowie vom G-BA akzeptierte - MCIDs für zahlreiche Messinstrumente und Anwendungsgebiete, und zum anderen würde das vorgeschlagene Vorgehen einer (endgültigen) Abkoppelung von den laufenden wissenschaftlichen Entwicklungen auf diesem aktuell sehr dynamischen Forschungsgebiet gleichkommen. • Letzteres beträfe auch Entwicklungen im Bereich künftiger Zulassungsanforderungen an PRO-Analysen. Dies umso mehr, als zahlreiche Nutzenbewertungen dieselben Fragestellungen wie auch zuvor schon die

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	entsprechenden Zulassungsverfahren untersuchen (d.h. basierend auf denselben relevanten klinischen Studien). Somit könnte es zu sehr unterschiedlichen Studienergebnis-Interpretationen und Entscheidungen zwischen Zulassung und Nutzenbewertung kommen. Im Übrigen dürfen gemäß AM-NutzenV Entscheidungen des G-BA die Schlussfolgerungen der Zulassungsbehörde bzgl. Nutzen und Verträglichkeit des infrage stehenden Arzneimittels nicht widersprechen; auch insofern ist die Eignung einer einzigen „Universal-MCID“ vermutlich nicht zielführend, wenn es um PRO-Endpunkte geht. Zumal eine willkürlich festgelegte „Universal-MCID“, wie jetzt vom IQWiG vorgeschlagen, unseres Wissens seitens EMA und FDA derzeit nicht zur Diskussion steht. Stattdessen werden verschiedene andere Möglichkeiten zur MCID-Ermittlung und zur PRO-Datenaufbereitung diskutiert, welche vermutlich eine transparentere und somit auch bessere Entscheidungsgrundlage bieten könnten (bspw. auch verteilungsbasierte Ergebnisdarstellungen: https://www.fda.gov/media/132505/download [1]). AstraZeneca erkennt jedoch an, dass für einen Teil der PRO-Messinstrumente bislang keine validierten MCIDs existieren und dass es dem einzelnen Hersteller in den meisten Fällen dann auch nicht möglich sein wird, dieses Problem kurzfristig zu „heilen“ (bspw., wie zuweilen vorgeschlagen, durch eine Validierung im Rahmen der Phase-2-Studien). In solchen Fällen würde man analysetechnisch zunächst auf Unterschiede in den Gruppen-Mittelwerten abheben müssen. Da diese Ergebnisse jedoch schwieriger zu interpretieren sind, ist es ausdrücklich begrüßenswert, auch über ergänzende Datenaufbereitungen auf Basis individueller Responsekriterien nachzudenken, um dem G-BA eine gut informierte Entscheidung trotz Unsicherheiten zu ermöglichen. Der Weg über eine willkürlich festgesetzte „One-Size-fits-All“-MCID ist aus Sicht von AstraZeneca diesbezüglich jedoch nicht zielführend. <u>Fazit:</u> Die vorgesehenen Änderungen werden von AstraZeneca abgelehnt. Selbst wenn die Bestimmung einer MID methodisch komplex sein kann, so ist es doch absolut nicht zielführend, als Reaktion eine universell geltende MID einzuführen, die alle vorliegende Evidenz bzgl. skalenspezifischer MIDs komplett ignoriert. Es würden damit etablierte Schwellenwerte von seit vielen Jahren eingesetzten Skalen infrage gestellt, ohne eine

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	individuelle, sachbezogene Evaluation der jeweiligen MID voranzustellen. Diese grob verallgemeinernde Ablehnung aller Studien und Evidenz zu MIDs ist nicht zielführend und erschwert die Berücksichtigung patientenrelevanter Endpunkte (wie z.B. Symptomskalen) in der Nutzenbewertung erheblich. Zusätzlich ist die Ableitung der vorgeschlagenen minimalen MID von 15% der Skalenspannweite wissenschaftlich nicht nachvollziehbar. Der Wert wird in dem vorliegenden Methoden-Entwurf mit Verweis auf zwei hochspezifische Übersichtsarbeiten begründet (Fatigue bzw. Rotator cuff disorders). Aus diesen Arbeiten eine Mindest-MID für sämtliche andere Skalen abzuleiten, entbehrt einer wissenschaftlichen Grundlage. <u>Vorgeschlagene Änderung:</u> Die vorgeschlagenen Änderungen zur MCID-Bewertungsmethodik sollten gänzlich entfallen. Aus Sicht von AstraZeneca sollten mögliche neue Standards zur MCID-Bewertung nur nach intensivem Austausch mit Vertretern aus Wissenschaft, Praxis, ggf. Zulassungsbehörde und pharmazeutischer Industrie entwickelt werden. Dabei sollte ein wissenschaftlicher Konsens angestrebt werden. Erst danach sollte eine Änderung des Methodenpapiers in der Sache angestrebt werden.
Abschnitt 3.3.3 „Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V“, S. 67	<u>Anmerkung:</u> In Abschnitt 3.3.3 wird unter der Überschrift „D) Stetige oder quasistetige Zielgrößen mit jeweils vorliegenden SMDs“ eine Methodik zur Quantifizierung des Effektausmaßes auf Basis von standardisierten Mittelwertdifferenzen (SMDs) und ihren 95% Konfidenzintervallen eingeführt. Im Kern begrüßt AstraZeneca das Bemühen, auch auf Basis von SMDs zu einer Quantifizierung des Zusatznutzens zu kommen. Die vorgeschlagenen Grenzen erscheinen aber willkürlich und werden wissenschaftlich nicht angemessen begründet. So verwendet das IQWiG zwar die für Cohen's d üblichen Bereiche (kleine Effekte ab 0.2; mittlere Effekte ab 0.5; starke Effekte ab 0.8) als Ausgangsbasis, argumentiert aber für die Findung von Schwellenwerten für die untere 95% Konfidenzintervallgrenze, man könne zeigen dass „[...] die für ein bestimmtes Ausmaß wünschenswerte Effektstärke ungefähr halbiert werden muss, um den entsprechenden Schwellenwert abzuleiten.“ Dieses Vorgehen sei an das Vorgehen zur Ableitung des Zusatznutzens bei binären Punkten angelehnt, wie von Skipka et al. [2] publiziert. Es werden aber keine weiteren Ausführungen gemacht, wie (und unter Annahme welcher Rahmenbedingungen) sich eine Halbierung der

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	wünschenswerten Effektstärke zur Findung des Schwellenwertes mit diesem Ansatz wissenschaftlich nachvollziehbar begründen lässt. Zudem stellt der bei Skipka et al. dargestellte Ansatz ohnehin keine allgemein wissenschaftlich anerkannte Vorgehensweise dar. Desweiteren ist nicht nachvollziehbar, in welcher Art und Weise das IQWiG die beiden unterschiedlichen Schweregrade der betrachteten Endpunkte in die Berechnung der Schwellenwerte hat einfließen lassen. Das gesamte Vorgehen erscheint somit willkürlich und nicht in Übereinstimmung mit dem Ziel einer ergebnisoffenen Nutzenbewertung von Arzneimitteln. Aufgrund der Anforderung, dass das komplette Konfidenzintervall der SMD oberhalb des jeweiligen Schwellenwertes liegen muss, müssen die real geschätzten Effektstärken (Punktschätzer) in der Praxis deutlich oberhalb der jeweiligen Schwellenwerte liegen. Ein nach üblicher Einteilung mittelstarker Effekt (z.B. Cohen's $d = 0.55$) würde demnach z.B. in den allermeisten Fällen als geringer Effekt eingestuft werden, da die Konfidenzintervallgrenzen nicht komplett oberhalb von 0.5 liegen. <u>Vorgeschlagene Änderung:</u> Die vorgeschlagenen Schwellenwerte sollten wissenschaftlich nachvollziehbar begründet werden. Die verwendete Methodik sollte sicherstellen, dass die Effektstärken nicht durch unrealistische Anforderungen an Konfidenzintervallgrenzen systematisch unterschätzt werden.
Abschnitt 9.3.13 „Umgang mit unterschiedlichen mittleren Beobachtungsdauern“, S. 192/193	<u>Anmerkung:</u> Im Abschnitt "9.3.13 Umgang mit unterschiedlichen mittleren Beobachtungsdauern" führt das IQWiG aus, dass unerwünschte Ereignisse vollständig erfasst werden sollten, d.h. auch nach einem Behandlungsabbruch oder -wechsel. Aus Sicht von AstraZeneca widerspricht dieses dem üblichen Vorgehen in klinischen Studien, in denen unerwünschte Ereignisse meist bis maximal 30 Tage nach Absetzen der Therapie erfasst werden. AstraZeneca vertritt die Ansicht, dass die Aussagekraft der Endpunkte zu unerwünschten Ereignissen durch eine Erfassung ausschließlich im Zeitraum der randomisierten Behandlung (zuzüglich 30 Tage) hinreichend gegeben ist und valide Schlussfolgerungen im Rahmen der Nutzenbewertung zulässt. Die vom IQWiG geforderte Erhebung weit über das Absetzen der Medikation hinaus erscheint wenig zielführend, da hierbei einige Limitationen ignoriert werden, die sich vor allem aus der Interpretation der Ergebnisse ergäben: Grundsätzlich würde eine Erfassung der unerwünschten Ereignisse über die festgelegten 30 Tage nach Therapieende hinaus zu einer systematischen Verzerrung führen, da durch das Einleiten weiterer

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	Folgetherapien oder im natürlichen Krankheitsverlauf ebenfalls unerwünschte Ereignisse auftreten, die jedoch für gewöhnlich in keinerlei Zusammenhang mehr mit der ursprünglich betrachteten Behandlung stehen. Die zusätzliche Betrachtung dieser Ereignisse würde daher in der Konsequenz den Vergleich des Sicherheitsprofils von Prüfsubstanz und ZVT verschleiern, unabhängig davon ob diese in beiden Armen aufgrund ähnlicher Folgetherapien mit gleicher Häufigkeit auftreten. Man müsste in den meisten Therapieszenarien daher nur ausreichend lange nachbeobachten, um keine (oder zumindest deutlich weniger) statistisch signifikante Unterschiede zwischen den Behandlungsgruppen feststellen zu können. Dies kann aber weder im Rahmen der Arzneimittelzulassung, noch in der Nutzenbewertung Ziel der Erhebung von unerwünschten Ereignissen sein. Ein Nachbeobachtungszeitraum von 30 Tagen ist des Weiteren im Einklang mit den für internationale klinische Studien in der Regel relevanten ‚Good Review Practices‘ der US-amerikanischen Zulassungsbehörde FDA. Selbst zur Feststellung von Todesfällen, die im Zusammenhang mit dem Sicherheitsprofil eines Arzneimittels stehen könnten, wird für jene mit raschem Wirkeintritt und relativ kurzer Eliminationshalbwertszeit eine Beobachtung nach Behandlungsende von 4 Wochen empfohlen [3]. Darüber hinaus wären engmaschige Visiten notwendig, um unerwünschte Ereignisse einer Folgetherapie akkurat erfassen und bewerten zu können. Das wäre vielen Patienten nicht zumutbar, vor allem in Überlebensnachbeobachtungen in onkologischen Studien. Eine 1- bis 2-mal jährliche Erfassung hingegen, erscheint, aufgrund vorwiegend retrospektiver und unvollständiger Daten, stark verzerrt und somit wenig erkenntnisbringend. <u>Vorgeschlagene Änderung:</u> Die Analyse und Bewertung von unerwünschten Ereignissen sollte vor dem Hintergrund des jeweiligen Bewertungsverfahrens erfolgen. Die generelle Forderung einer Erhebung von unerwünschten Ereignissen weit über das Ende einer Therapie hinaus erscheint nicht zielführend und sollte gestrichen werden.
Abschnitt 3.1.3 „Ermittlung des Schadens medizinischer Interventionen“,	<u>Anmerkung:</u> Hinsichtlich der Aufbereitung und Darstellung von Schadensaspekten in Nutzenbewertungen definiert das IQWiG die Betrachtung „spezifische unerwünschte Wirkungen“ neu. Das IQWiG definiert zwei Wege für die Auswahl der ‚potenziell relevanten spezifischen unerwünschten Wirkungen‘ je nach

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
S. 49	Nutzenbewertungsverfahren gemäß 139 a SGB V bzw. § 35 a SGB V. • Im Rahmen der Nutzenbewertung gemäß §139a SGB V erfolgt die Auswahl der spezifischen UEs vorrangig aufgrund der Bedeutung i) für das Krankheitsbild oder ii) die in der Studie bzw. in den Studien eingesetzten Interventionen und basiert auf einer Vorabrecherche zur jeweiligen Fragestellung sowie der Einbindung von Betroffenen. • Im Rahmen der Nutzenbewertung von neuen Arzneimitteln gemäß § 35a SGB V wird für die Auswahl der spezifischen UEs die folgende Herangehensweise postuliert: <i>„Zum anderen werden für die Nutzenbewertung relevante spezifische unerwünschte Wirkungen basierend auf den in der relevanten Studie bzw. den relevanten Studie aufgetretenen unerwünschten Ereignissen identifiziert. [...] Erstens werden aufgrund des Verfahrens keine Vorabrecherchen durchgeführt und zweitens werden regelhaft neue Arzneimittel bewertet, sodass die Identifizierung unbekannter bzw. unerwarteter unerwünschter Wirkungen hier einen höheren Stellenwert hat. Zur Auswahl werden in diesem Fall die unerwünschten Ereignisse, schwerwiegenden unerwünschte Ereignisse, ggf. schweren unerwünschte Ereignisse zugrunde gelegt, die vom pharmazeutischen Unternehmer gemäß der Anforderungen in der Verfahrensordnung des G-BA im Dossier übermittelt werden [271]. Die Kriterien für die Auswahl stellen die Patientenrelevanz der Ereignisse und die Unterschiede zwischen den Behandlungsarmen dar. Für nicht schwerwiegende bzw. nicht schwere unerwünschte Ereignisse werden zudem ggf. zusätzliche Anforderungen an die Mindesthäufigkeit der Ereignisse gestellt, sofern dies die Gesamtabwägung zum Zusatznutzen nicht beeinflusst.“ [4]</i> Die Betrachtung von spezifischen unerwünschten Ereignissen ist nicht neu bei der Generierung von Evidenz für (neue) Arzneimittel und deren Bewertung. Üblicherweise werden spezifische UEs im Studienprotokoll à priori definiert und im Studienbericht separat berichtet („unerwünschte Ereignisse von spezifischem Interesse (AESI)“). Diese spezifischen UEs wurden in der Vergangenheit auch im Rahmen der Nutzenbewertung vorgelegt und bewertet. Die Betrachtung der spezifischen UEs gemäß des vorliegenden IQWiG-Methodenpapiers, zielt hingegen auf die Identifizierung der Unterschiede in den Nebenwirkungsprofilen der Behandlungsarme ab. Insofern unterscheiden sich die

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	prädefinierten spezifischen UEs im Studienprotokoll per se von den retrospektiv ermittelten spezifischen UEs im Sinne des neuen IQWiG-Ansatzes. Im Hinblick auf die Eindeutigkeit des Begriffes „spezifische unerwünschte Ereignisse“, könnte eine alternative und geeignetere Bezeichnung der spezifischen UEs, welcher das Vorgehen des IQWiG besser abbildet, hilfreich sein. Das Vorgehen des IQWiGs bei den retrospektiv zu ermittelnden spezifischen UEs für die Darstellung von negativen Effekten in der Nutzenbewertung von neuen Arzneimitteln basierend auf einem „Screening“ der in der relevanten Studie bzw. den relevanten Studien aufgetretenen unerwünschten Ereignissen, schwerwiegenden unerwünschten Ereignissen, ggf. schweren unerwünschten Ereignisse und der Auswahl anhand von i) Patientenrelevanz und ii) Unterschieden zwischen den Behandlungsarmen, ist aus Sicht von AstraZeneca nach Betrachtung verschiedener Nutzenbewertungen bislang nicht hinreichend nachvollziehbar. Die zugrunde gelegte Evidenz wurde jeweils nicht angeführt, wie auch die Begründung der Patientenrelevanz oder der besonderen Relevanz für das Krankheitsbild nicht dargelegt wurde [5, 6]. Es ist bspw. aus den bisherigen IQWiG-Bewertungen nicht ersichtlich, warum bestimmte unerwünschte Ereignisse in die Betrachtung einbezogen wurden und andere nicht. Es entsteht der Eindruck einer eher beliebigen und weniger systematischen Anlegung von Auswahlkriterien. <u>Fazit:</u> Die zugrundeliegende Methodik bzw. das Vorgehen des IQWiGs bei den retrospektiv zu ermittelnden spezifischen UEs für die Darstellung von negativen Effekten in der Nutzenbewertung von neuen Arzneimitteln ist aus Sicht von AstraZeneca nicht hinreichend nachvollziehbar und die Aussagekraft dieser Betrachtung daher infrage zu stellen. In den Nutzenbewertungsberichten sollte daher das jeweils konkrete Vorgehen genauer beschrieben und dargestellt werden als bisher (bspw. warum wurden bestimmte Nebenwirkungen in die Betrachtung einbezogen und andere nicht), um auf dieser Grundlage eine fundierte Diskussion im Rahmen einer Stellungnahme zu ermöglichen. <u>Vorgeschlagene Änderung:</u> Es sollte im Methodenpapier klargestellt werden, dass die jeweils konkrete Ableitung der spezifischen UEs in den Nutzenbewertungen darzulegen ist (bspw. via Anhang). Im Übrigen sollte eine ggf. alternative Bezeichnung eingeführt

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	werden, um mögliche Missverständnisse hinsichtlich der üblichen à priori definierten unerwünschten Ereignisse von spezifischem Interesse (AESI) zu vermeiden.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

1. **Food and Drug Administration (FDA). 2019.** Incorporating clinical outcome assessments into endpoints for regulatory decision-making. [Zugriff: 31.01.2020] <https://www.fda.gov/media/132505/download>
2. **Skipka G, Wieseler B, Kaiser T, Thomas S, Bender R, Windeler J et al.** Methodological approach to determine minor, considerable, and major treatment effects in the early benefit assessment of new drugs. *Biom J* 2016; 58(1): 43-58.
3. **Food and Drug Administration (FDA). 2010.** Good Review Practice: Clinical Review Template. [Zugriff 31.01.2020]. <https://www.fda.gov/media/72472/download>
4. **Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2019).** Allgemeine Methoden. Entwurf für Version 6.0. Vom 05.12.2019. [Zugriff: 31.01.2020]. <https://www.iqwig.de/de/methoden/methodenpapier.3020.html>
5. **Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2019).** IQWiG-Berichte Nr. 826 Olaparib (Ovarialkarzinom; Erstlinie Erhaltung) –Nutzenbewertung gemäß § 35a SGB V. Version 1. Vom 11.10.2019 [Zugriff: 31.01.2020]. https://www.g-ba.de/downloads/92-975-3160/2019-07-15_Nutzenbewertung-IQWiG_Olaparib-D-464.pdf
6. **Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2019).** IQWiG-Berichte Nr. 825 Olaparib (Mammakarzinom) – Nutzenbewertung gemäß § 35a SGB V. Version 1. Vom 11.10.2019 [Zugriff 31.01.2020]. https://www.g-ba.de/downloads/92-975-3159/2019-07-15_Nutzenbewertung-IQWiG_Olaparib_D-459.pdf

A.1.6– Bayer Vital GmbH

Autorinnen und Autoren

- Dintsios, Charalabos-Markos
- Kaup, Marisa

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Dintsios, Charalabos-Markos; Dr.
Kaup, Marisa
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen) ☒ im Namen folgender Institution / Organisation: <u>Bayer Vital GmbH</u> ☐ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
Abschnitt 3.1.3 „Ermittlung des Schadens medizinischer Interventionen“ (S. 50)	<u>Anmerkung:</u> Am Ende des Abschnitts geht das IQWiG auf die Operationalisierung von unerwünschten Ereignissen ein und führt diesbezüglich aus <i>“Für die ausgewählten spezifischen unerwünschten Ereignisse wird jeweils die am besten geeigneten Operationalisierung in Bezug auf Mess- und Ergebnissicherheit herangezogen. Auch wird geprüft, ob die ausgewählten Ereignisse zum selben inhaltlichen Konstrukt konsistent sind. Liegt keine geeignete Operationalisierung vor oder sind die Ergebnisse nicht konsistent, werden diese in der Regel nicht für Aussagen zum Schaden einer Intervention herangezogen.”</i> Die Kategorisierung der „am besten geeigneten Operationalisierung“ in Bezug auf Reliabilität und Ergebnissicherheit wird im Methodenentwurf relativ abstrakt gehalten. Des Weiteren wird nicht auf die Ätiologie einer potenziellen Ergebnisinkonsistenz eingegangen, sondern bei Vorliegen jedweder Ergebnisinkonsistenz die Einbeziehung der Ergebnisse in der Regel abgelehnt. Eine beispielhafte Darstellung der am besten geeigneten Operationalisierung anhand gängiger Konstellationen der Operationalisierung von Sicherheitsendpunkten in klinischen Studien würde die Forderung des IQWiG konkretisieren. Ferner sind Aussagen zum Schaden der Intervention unter Einbeziehung der Begründung potenzieller Ergebnisinkonsistenz und nachfolgender Abwägung hinsichtlich einer Einbeziehung der Ergebnisse nicht per se wegen ihrer Inkonsistenz auszuschließen.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<u>Vorgeschlagene Änderung:</u> Einfügung einer tabellarischen Darstellung der am besten geeigneten Operationalisierung anhand gängiger Konstellationen der Operationalisierung von Sicherheitsendpunkten in klinischen Studien. Einfügung eines Abschnittes zur Ätiologie von Ergebnisinkonsistenz und davon abhängiger potenzieller Einbeziehung inkonsistenter Ergebnisse in die Betrachtung von Schadensaspekten.
Abschnitt 3.3.3 „Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V“, S. 67	<u>Anmerkung:</u> In diesem Abschnitt ergänzt das IQWiG die Methodik zur Quantifizierung der Effektausmaße um die Vorgehensweise bei stetigen oder quasistetigen Zielgrößen mit standardisierten Mittelwertdifferenzen (SMDs) und definiert hierbei eine Reihe an Schwellenwerten zur Ausmaßbestimmung bei SMDs in Abhängigkeit von der Zielgrößenkategorie. Das IQWiG führt aus: <i>„In Anlehnung an das Verfahren für binäre Daten nach Skipka et al. [646] lassen sich auch Schwellenwerte für das Effektmass SMD (Cohen's d bzw. Hedges' g) ableiten. Man kann zeigen, dass die für ein bestimmtes Ausmaß wünschenswerte Effektstärke ungefähr halbiert werden muss, um den entsprechenden Schwellenwert abzuleiten. Als Orientierung für die theoretisch wünschenswerten Effektstärken bei schwerwiegenden Symptomen können die Mittelwerte der Bereiche für die übliche Einteilung von Cohen's in kleine (SMD zwischen 0,2 und 0,5) mittlere (SMD zwischen 0,5 und 0,8) und große Effekte (SMD > 0,8) dienen. Tabelle 6 gibt einen Überblick über die verwendeten Schwellenwerte. Im Gegensatz zu den Schwellenwerten für die relativen Maße ist bei der SMD zu beachten, dass das Konfidenzintervall vollständig oberhalb des entsprechenden Schwellenwerts liegen muss und der Nulleffekt hier nicht durch die Zahl 1 sondern durch die Zahl 0 gegeben ist.“</i> Zwar mag auch in diesem Vorschlag des IQWiG wie bei der entsprechenden Herleitung der Schwellenwerte zur Ausmaßbestimmung für das relative Risiko bei binären Zielgrößen eine anzurechnende intellektuelle Leistung erkennbar sein, dennoch gelten auch für das aktuell vorgeschlagene Vorgehen zu stetigen oder quasistetigen Zielgrößen mit jeweils vorliegenden SMD die damaligen Kritikpunkte. Diese wurden in den Stellungnahmen zum entsprechenden Methodenentwurf sowie zum Anhang A der Nutzenbewertung von Ticagrelor festgehalten und auch in der diesbezüglichen Literatur (Röhmel 2012, Witte & Greiner 2013, Vach 2014, Herpers & Dintsios 2019) dargelegt. Diesen Schwellenwerten liegen durch die Aufteilung der Ausmaßkategorien in Abhängigkeit von den kategorisierten Zielgrößen implizite normative Entscheidungen zugrunde, auch wenn das Institut hierzu keine

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	begleitende, wissenschaftlich fundierte Rationale liefert und auf das Vorgehen für binäre Zielgrößen verweist; wobei darauf hinzuweisen ist, dass solche Entscheidungen explizit dem auftraggebenden Gemeinsamen Bundesausschuss vorbehalten sind, wie Letzterer auch in seiner Beschlussfassung bzw. den dazugehörigen Tragenden Gründen zur Erstbewertung von Ticagrelor ausführt. Somit erlangen die Schwellenwerte den Charakter von Setzungen und nicht von Methoden, die internationale Standards der Evidenz-basierten Medizin, wie im Sozialgesetzbuch gefordert, wiedergeben. Anders als bei der Zielgrößenkategorie 2, wo zumindest eine Analogie zur Cohen's d Trichotomisierung hinsichtlich der Effektstärke konstatiert werden kann (SMD zwischen 0,2 und 0,5 kleine Effekte; SMD zwischen 0,5 und 0,8 mittlere und SMD > 0,8 große Effekte), ist insbesondere für die Zielkategorie 1 (schwer-wiegende Symptome, Nebenwirkungen und gesundheitsbezogene Lebensqualität) die Schwellenwertsetzung nicht nachvollziehbar, da eine genaue Herleitung nicht im Methodenentwurf 6.0 präsentiert wird. Darüber hinaus gestaltet sich das vorgeschlagene Vorgehen als hyperkonservativ, da durch die Schwellensetzungen für die Konfidenzintervalle des jeweiligen Punktschätzers und nicht für die Punktschätzer selber, auf welche sich die konventionelle Einteilung zur Effektstärke nach Cohen's d bezieht, stärkere Effekte als in der tradierten Einteilung erzielt werden müssen. Erst wenn es sich um tradierte Setzungen mit mittlerweile Konventionscharakter wie bspw. einen Alphafehler von 0,05 im Rahmen von klinischen Studien handelt, und diese Setzungen auch international Akzeptanz erreichen, kann von Methoden gesprochen werden. Bei Instituts-eigenen Setzungen allerdings, ohne einen vorausgegangenen oder begleitenden internationalen erkenntnistheoretischen wissenschaftlichen Diskurs hierzu, stellt sich die Frage inwieweit diese einen verbindlichen Charakter besitzen, um auch im Rahmen von klinischen Studienplanungen für Studien, die neben der Zulassung auch in die frühe Nutzen-bewertung Eingang finden sollen, berücksichtigt zu werden. Richtigerweise merkt das Institut in seinen Methoden im selben Abschnitt an, dass „Für den 3. Schritt der Operationalisierung der Gesamtaussage zum Ausmaß des Zusatznutzens bei gemeinsamer Betrachtung aller Endpunkte ist eine strenge Formalisierung nicht möglich, da für die hierauf zutreffenden Werturteile gegenwärtig keine ausreichende Abstraktion bekannt ist“. Der einzige Unterschied im Vorgehen zwischen Schritt 2 zur Endpunkt-bezogenen Zusatznutzenausprägung und Schritt 3 zur Gesamt-aussage zum Zusatznutzen liegt darin, dass im ersten Fall die

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	Werturteile implizit über die eingehenden Effektgrößen enthalten sind, im zweiten explizit auf diese bereits im Methodenpapier hingewiesen wird. Nichtsdestotrotz bleiben es in beiden Fällen Werturteile, für die das IQWiG kein Mandat vom Entscheidungsträger hat. <u>Vorgeschlagene Änderung:</u> Beibehaltung der ursprünglichen Trichotomisierung der Effektstärke nach Cohen's mit direktem Bezug auf den Punktschätzer und ohne weitergehende Differenzierung nach Zielgrößenkategorie zur Vermeidung impliziter Werturteile, die ohne den Charakter einer Konvention zu erlangen, nicht zu den internationalen Standards der Evidenz-basierten Medizin hinzugezählt werden können. Alternativ: Transparenzmachung der impliziten Werturteile mit der Möglichkeit entsprechende Gewichtungen für die jeweiligen Schwellengrößen indikationsbezogen unter Beteiligung von Patienten oder des Entscheidungsträgers durchzuführen, um willkürliche indikationsübergreifende Setzungen zu vermeiden.
Abschnitt 9.3.3 „Beurteilung klinischer Relevanz“, S. 177/178	<u>Anmerkung:</u> Während in den Vorversionen des Methodenpapiers bei der Anwendung von Responderanalysen „<i>ein validiertes bzw. etabliertes Responsekriterium</i>“ (im Sinne einer individuellen Minimal important Difference [MID]) erforderlich war, um einen Nachweis eines relevanten Effekts zu erbringen, ändert das Institut diese Anforderungen und führt ein Responsekriterium ein, das „<i>mindestens 15 % der Skalenspannweite des verwendeten Erhebungsinstruments</i>“ entspricht und zudem „<i>präspezifiziert</i>“ sein muss. Für den Fall, dass präspezifizierte Responsekriterien im Sinne einer MID unterhalb von 15 % der Skalenspannweite liegen, sollen diese nicht herangezogen werden. In diesen Fällen und solchen, in denen gar keine Responsekriterien präspezifiziert wurden, soll entweder auf stetige Daten mit standardisierten Mittelwertdifferenzen (SMDs, in Form von Hedges' g) zurückgegriffen werden, oder alternativ post hoc spezifizierte Analysen mit einem Responsekriterium von genau 15 % des Skalenrangs berücksichtigt werden. Laut Institut sei Ausgangspunkt für die Änderung, dass es derzeit keine akzeptierten Standards gäbe, mit denen die Qualität von Studien bewertet und die Aussagekraft der ermittelten MIDs abgeschätzt werden könne. Begründet wird diese Festlegung mit der Sichtung von zwei Übersichtsarbeiten zu MIDs, die sich in vielen Fällen zwischen 10 % und 20 % der Spannweite der Scores eines Erhebungsinstruments bewegen würden. Die vorgeschlagene „präspezifizierte“ Mindestresponderschwelle von 15% der Skalenbreite entbehrt jeglicher methodischen

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	Fundierung und erscheint erneut als eine Instituts-eigene Setzung mit normativem Charakter. Selbst unter der Prämisse, dass es derzeit keine akzeptierten Standards gibt, mit denen die Qualität von Studien bewertet und die Aussagekraft der ermittelten MIDs abgeschätzt werden könne, was auch wiederum zu hinterfragen wäre, würde dieser Mangel durch einen anderen Mangel ersetzt werden. Der neue Vorschlag des IQWiG lässt weder die Skaleneigenschaften noch die Änderungssensitivität der Instrumente in die Betrachtung einfließen und setzt somit ein willkürliches Kriterium an, das sich als arithmetisches Mittel von MIDs, die sich in vielen Fällen zwischen 10 % und 20 % der Spannweite der Scores eines Erhebungsinstruments bewegen u.a. auf Basis der beiden herangezogenen Übersichtsarbeiten zu MIDs (Nordin 2016, St-Pierre 2016) erweist. Letztere beziehen sich allerdings auf ausgewählte und sehr spezifische Symptome wie „fatigue“ und „symptoms and functional limitations in individuals with rotator cuff disorders“, sodass deren Ergebnisse nicht allgemeine Geltung über alle Instrumente hinweg genießen können. Letztlich handelt sich bei diesem Vorgehen des Instituts um einen Alleingang, der die jahrelange Arbeit von Arbeitsgruppen an Validierungsstudien schlicht negiert. Ferner stellt sich eher die Frage nach verteilungsbasierten MIDs (1/2 SD), falls ankerbasierte MIDs im Rahmen ihrer Validierungsstudien gravierende Mängel aufweisen sollten, um diese Mängel zu heilen. Jedenfalls ist der Ansatz einer exogen festgelegten 15% Schwelle der entsprechenden Skalenbreite nicht zielführend, da damit bereits validierte MIDs zu etablierten Instrumenten, die auch für entsprechende Studienplanungen herangezogen wurden, insofern diese niedriger ausfallen, hinfällig werden und somit eine Reihe von bereits erfolgten Nutzenbewertungen auf Basis dieser Instrumente mit den jeweilig angesetzten MIDs im Nachgang ihre Geltung verlieren würde. Beispielsweise würde für den in der Onkologie breit angewendeten EORTC QLQ-C30 eine nicht nachvollziehbare Erhöhung der MID um 50 % von bisher 10 auf 15 Punkte erfolgen, welche dessen bereits nachgewiesene klinische Relevanz unbegründet streitig machen würde. <u>Vorgeschlagene Änderung:</u> Das ursprüngliche Vorgehen sollte beibehalten werden. Für Fälle, in denen vorhandene Validierungsstudien mit gravierenden Mängeln behaftet sind, kann erst nach intensivem methodischem Diskurs auf internationaler Ebene eine konsenterte Alternative in Erwägung gezogen werden.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
Abschnitt 9.3.12 „Umgang mit unvollständigen Daten“, S. 191-192	<u>Anmerkung:</u> Das IQWiG führt hierzu folgendes aus: „Das Verzerrungspotenzial steigt mit steigendem Anteil fehlender Werte. Ergebnisse fließen daher in der Regel nicht in die Nutzenbewertung ein, wenn diese auf weniger als 70 % der in die Auswertung einzuschließenden Patientinnen und Patienten basieren, das heißt, wenn der Anteil der Patientinnen und Patienten, die nicht in der Auswertung berücksichtigt werden, größer als 30 % ist. In der Literatur werden zum Teil bereits Auswertungen, in denen 20 % der Patientinnen und Patienten nicht berücksichtigt werden, als nicht mehr aussagekräftig betrachtet [618]. Die Ergebnisse werden ferner auch dann nicht in die Nutzenbewertung einbezogen, wenn der Unterschied der Anteile nicht berücksichtigter Patientinnen und Patienten zwischen den Gruppen größer als 15 Prozentpunkte ist.“ Die europäische Zulassungsbehörde EMA nennt keine expliziten Richtwerte zur Einschätzung der Anteile fehlender Werte und selbst in der vom Institut zitierten Quelle mit einem Richtwert von 20% kommen die Autoren zum Schluss, dass bspw. eine genaue Beschreibung des Patientenflusses bei der Interpretation hilft. Für den Richtwert des Unterschieds zwischen den Gruppen mit 15% wiederum fehlt jegliche Quelle, auf die das IQWiG verweisen kann. Abgewichen von den genannten Richtwerten kann laut IQWiG im Falle von erwiesenem missing at random oder durch Einsatz von Imputationen bzw. Ausgleich unter Verwendung entsprechender statistischer Methoden, ohne jedoch diese zu konkretisieren. <u>Vorgeschlagene Änderung:</u> Während eine Richtgröße von 70% fallbezogen pragmatisch erscheint, ist die Anwendung der 15%-Regel zwischen Patientengruppen weder nachvollziehbar noch wissenschaftlich fundiert bzw. literaturbelegt, sodass von dieser Abstand genommen werden sollte. Eine Konkretisierung akzeptierter fallbezogener Imputationsverfahren sollte im Rahmen des Methodenpapiers in Erwägung gezogen werden.

Literaturverzeichnis

Röhmel, J.: Gutachten zum Vorschlag des IQWiG zur Bewertung des Ausmaßes des Zusatznutzens im Rahmen der Nutzenbewertung von Arzneimitteln nach § 35a SGB V, Bremen (2012)

Witte, J., Greiner, W.: Problembefunde der Quantifizierung des Zusatznutzens im Rahmen der frühen Arzneimittelnutzenbewertung. *Gesundh ökon Qual manag.*, 18(05), 226–234 (2013)

Vach W. Quantifizierung des Ausmaßes des Zusatznutzens von neuen Arzneimitteln: „gering“ – „beträchtlich“ – „erheblich“ – 6 Anmerkungen aus der Sicht eines Biometrikers. *Gesundheitswesen*, 76(11): 757-762 (2014)

Herpers, M., Dintsios CM. Methodological problems in the method used by IQWiG within early benefit assessment of new pharmaceuticals in Germany. *EJHE*, 20(1):45–57 (2019)

A.1.7 – Boehringer Ingelheim Pharma GmbH & Co. KG

Autorinnen und Autoren

- Henschel, Andreas
- Minarzyk, Anette
- Pfarr, Egon

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Anette Minarzyk
Egon Pfarr
Dr. Andreas Henschel
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen) ☒ im Namen folgender Institution / Organisation: Boehringer Ingelheim Pharma GmbH & Co. KG ☐ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.
Am 05. Dezember 2019 hat das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) einen Entwurf mit inhaltlichen Aktualisierungen und Ergänzungen seiner allgemeinen Methoden veröffentlicht und ein Stellungnahmeverfahren eröffnet. Boehringer Ingelheim (BI) begrüßt die kontinuierliche Weiterentwicklung dieser allgemeinen Methoden und nimmt im folgenden dazu Stellung.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
3.2.1 (S. 55 ff.)	<i>Auswirkung nicht publizierter Studienergebnisse auf das Fazit</i> • Es ist unklar, inwiefern "Dritte" hier den Hersteller des zu bewertenden Arzneimittels (pU) meint, oder andere Parteien wie universitäre oder staatliche Forschungsgruppen und andere pU. "Dritte" im Sinne des pU ist für Szenario 1-4 nachvollziehbar, wirft aber mit Blick auf die Szenarien 5-7 im Zusammenhang der frühen Nutzenbewertung Fragen auf, da diese ohne übermittelte Daten in jedem Fall zu "keinem Beleg für Zusatznutzen" ohne Ableitung der Aussagesicherheit führt, gleich ob Daten nicht vorhanden waren oder nicht übermittelt wurden. • Das Kapitel sollte im Hinblick auf Nachvollziehbarkeit und Klarheit <u>analogisiert werden</u>. Eine Definition von "Dritte" wäre in diesem Kontext hilfreich. <u>Anmerkung: Fazit unter Vorbehalt</u> <i>"Wird bei einer vollständigen Datenübermittlung durch die Herstellerin oder den Hersteller durch andere Recherchen der Nachweis erbracht, dass Daten in einem mäßigen oder erheblichen Umfang (z. B. aus Investigator-initiierten Studien oder Studien anderer Hersteller) fehlen, auf die die Herstellerin oder der</i>

Kapitel/Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<i>Hersteller keinen Zugriff hat, liegt zwar keine selektive Datenübermittlung durch die Herstellerin oder den Hersteller vor, ist aber eine Verzerrung durch die anderweitig fehlenden Daten möglich (Szenario 2) oder sogar wahrscheinlich (Szenario 3). Eine Nutzensaussage im Fazit (im Sinne von Beleg, Hinweis oder Anhaltspunkt für Nutzen oder Schaden) wird deshalb in ihrer Aussagesicherheit entsprechend abgeschwächt oder gar nicht abgeleitet." (S. 56, 57)</i> • Version 5.0 des Methodenpapiers sah ein Fazit unter Vorbehalt vor, wenn die Übermittlung der Herstellerdaten vollständig war, aber Daten anderer Hersteller oder Forscher nicht bereitgestellt werden konnten. Diese Möglichkeit fällt in der neuen Version 6.0 weg. Hier führt dieses Szenario zur Herabstufung der Aussagesicherheit bis hin zur Schlussfolgerung "kein Beleg für einen Nutzen". <u>Vorgeschlagene Änderung:</u> Das Fehlen von Daten aus Quellen, auf die ein pU keinen Zugriff hat, sollte maximal zu einer Herabstufung der Aussagesicherheit und/oder Fazit unter Vorbehalt, aber nicht zu einer Aberkennung des Zusatznutzens führen.
9.3.3 (S. 175 ff.)	• <u>Anmerkung:</u> Eine MID von (mindestens) 15 % der (theoretisch möglichen) Skalenspannweite als "regelhaftes Vorgehen" festzulegen wird vom IQWiG weder ausreichend wissenschaftlich begründet noch nachvollziehbar abgeleitet. Die beispielhaft angeführten Literaturstellen beziehen sich auf zwei spezifische Indikationen (Fatigue PROs bzw. Rotator Cuff Disorders) [1, 2]. Die Ableitung des IQWiG, "dass sich MIDs in vielen Fällen zwischen 10 % und 20 % der Spannweite der Scores eines Erhebungsinstruments bewegen" ist unspezifisch bzw. basiert nicht auf Empfehlungen der Autoren. Vielmehr schließen Nordin et al. bezüglich Fatigue: "MIDs vary substantially by estimation method, patient population and context both across and within fatigue PROMs. In light of this variation, published MIDs should be applied judiciously, after carefully considering their applicability to characteristics of the study in question." • Das IQWiG schreibt auf S. 177: "...dass MIDs von Charakteristika der Patientenpopulation, in denen das Instrument eingesetzt wird, und von anderen Kontextfaktoren abhängig sind: So scheinen der

Kapitel/Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<i>Schweregrad der Erkrankung, die Art der eingesetzten Intervention oder die Frage ob die Patientinnen und Patienten eine Verbesserung oder Verschlechterung ihrer Erkrankung erfahren, Einfluss auf die MID zu haben"</i> und widerspricht mit dieser Aussage einer regelhaften Ableitung der MID von 15 % der Skalenspannweite. • Auch das an dieser Stelle zitierte Paper von Alma et al 2019 spricht sich in seiner Schlussfolgerung für die Notwendigkeit von entsprechend angepassten MIDs aus. [3] • Der Entwurf sieht nicht vor, dass schon lange etablierte MIDs verwendet werden könnten. Damit werden auch wissenschaftlich sauber abgeleitete und validierte MIDs in Frage gestellt. BI teilt die Auffassung, dass MIDs auch abhängig von Charakteristika der Patientenpopulation und anderen Kontextfaktoren, zum Beispiel der Art der erhobenen Endpunkte sind. Ein starrer Cut-off von 15% wird als nicht zweckdienlich angesehen. Vorgeschlagene Änderung: • Eine MID, die basierend auf wissenschaftlichem Konsens, für eine spezifische Fragestellung von EMA und FDA explizit anerkannt ist, sollte auch als Responsekriterium für das AMNOG Dossier verwendet werden. • Zur Beurteilung des Zusatznutzen können dann ggf. weitere AMNOG-spezifische Kriterien (Vergleichstherapie, spezifische Analysekonzepte) Anwendung finden, ohne schon akzeptierte Responsekriterien neu ableiten zu müssen. • Sofern es noch keine anerkannten MIDs gibt, sollten die spezifischen MIDs z.B. mit Experten und Fachgesellschaften erörtert und Lösungen erarbeitet werden, die der Abhängigkeit der MIDs von den jeweiligen Kontextfaktoren Rechnung tragen. So wären z.B. für Patienten, deren Ausgangswerte zu Baseline schon an der Ober- bzw. Untergrenze der Skalen liegen, die Responsekriterien entsprechend anzupassen.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

1. Nordin A, Taft C, Lundgren-Nilsson A, Dencker A. Minimal important differences for fatigue patient reported outcome measures: a systematic review. BMC Med Res Methodol 2016; 16: 62. (PMID 27387456)
2. St-Pierre C, Desmeules F, Dionne CE, Fremont P, MacDermid JC, Roy JS. Psychometric properties of self-reported questionnaires for the evaluation of symptoms and functional limitations in individuals with rotator cuff disorders: a systematic review. Disabil Rehabil 2016; 38(2): 103-122. (PMID: 25801922)
3. Alma H, De Jong C, Jelusic D, Wittmann M, Schuler M, Kollen B et al. Baseline health status and setting impacted minimal clinically important differences in COPD: an exploratory study. J Clin Epidemiol 2019; 116: 49-61 (PMID: 31362055)

Abkürzungsverzeichnis

Abkürzung	Erklärung
AMNOG	Arzneimittelmarktneuordnungsgesetz
BI	Boehringer Ingelheim
EMA	European Medicines Agency
FDA	U.S. Food and Drug Administration
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
MID	Minimal Important Difference
PRO	Patient-Reported Outcome
PROMs	Patient-Reported Outcome Measures
pU	pharmazeutisches Unternehmen

A.1.8 – Bristol-Myers Squibb GmbH & Co. KGaA

Autorinnen und Autoren

- Kucka, Annett
- Kupas, Katrin

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Kucka, Annett
Kupas, Katrin; Dr.
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
☒ im Namen folgender Institution / Organisation: Bristol-Myers Squibb
☐ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

<i>Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.</i>

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
Abschnitt 3.2.1 (S. 55)	<u>Anmerkung:</u> Das IQWiG hat den Abschnitt zur Auswirkung nicht publizierter Studienergebnisse auf das Fazit überarbeitet und um einige Szenarien erweitert, bei denen zwar durch den Hersteller eine vollständige Übermittlung der Studiendaten erfolgt, aber dennoch die Aussagesicherheit entweder eingeschränkt wird, oder gar kein Nutzen oder Schaden abgeleitet wird (siehe Tabelle 4, Seite 56). BMS widerspricht diesem Vorgehen. Bei einer vollständigen Übermittlung der Daten seitens des Herstellers ist es unverständlich, dass kein oder nur ein eingeschränktes Fazit in der Nutzenbewertung gezogen werden kann. Dies gilt nur für den Fall der indirekten Vergleiche, wenn die publizierten Daten der Vergleichstherapie nicht ausreichen, um einen indirekten Vergleich sinnvoll durchführen zu können. In allen anderen Fällen ist nicht nachvollziehbar, dass bei vollständiger Datenlage der Herstellerstudien kein Fazit über Nutzen oder Schaden ausgesprochen werden kann. <u>Vorgeschlagene Änderung:</u> Bei vollständiger Datenübermittlung durch den Hersteller kann eine Bewertung des Zusatznutzens auf Basis dieser Daten erfolgen, und unabhängig von anderen Datenquellen kann Nutzen oder Schaden auf Basis der vom Hersteller vollständig übermittelten Daten festgestellt werden. Die Szenarien sollten für die Fälle konkretisiert werden, in denen tatsächlich keine Ableitung des Zusatznutzens möglich ist, wie z.B. bei indirekten Vergleichen.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
Abschnitt 3.3.3 (S. 67)	<u>Anmerkung:</u> Im Abschnitt 3.3.3 wird die Methodik zur Quantifizierung der Effektausmaße um die Vorgehensweise bei stetigen oder quasistetigen Zielgrößen mit standardisierten Mittelwertdifferenzen (SMDs) ergänzt. Das IQWiG definiert hierfür eine Reihe an Schwellenwerten zur Ausmaßbestimmung bei SMDs in Abhängigkeit von der Zielgrößenkategorie. Das IQWiG führt aus: <i>„In Anlehnung an das Verfahren für binäre Daten nach Skipka et al. [646] lassen sich auch Schwellenwerte für das Effektmaß SMD (Cohen's d bzw. Hedges' g) ableiten. Man kann zeigen, dass die für ein bestimmtes Ausmaß wünschenswerte Effektstärke ungefähr halbiert werden muss, um den entsprechenden Schwellenwert abzuleiten. Als Orientierung für die theoretisch wünschenswerten Effektstärken bei schwerwiegenden Symptomen können die Mittelwerte der Bereiche für die übliche Einteilung von Cohen's in kleine (SMD zwischen 0,2 und 0,5) mittlere (SMD zwischen 0,5 und 0,8) und große Effekte (SMD > 0,8) dienen. Tabelle 6 gibt einen Überblick über die verwendeten Schwellenwerte. Im Gegensatz zu den Schwellenwerten für die relativen Maße ist bei der SMD zu beachten, dass das Konfidenzintervall vollständig oberhalb des entsprechenden Schwellenwerts liegen muss und der Nulleffekt hier nicht durch die Zahl 1 sondern durch die Zahl 0 gegeben ist.“</i> Es ist grundsätzlich zu begrüßen, dass das IQWiG ein methodisches Vorgehen anstrebt, welches dem Ziel dient, auch bei stetigen Daten eine Ausmaß-Quantifizierung der Effektgrößen zu ermöglichen. Jedoch müssen die festgelegten Schwellenwerte und ihre Herleitung grundsätzlich hinterfragt werden, da diese teilweise nicht nachvollziehbar sind bzw. nicht wissenschaftlich begründet wurden. So lassen sich lediglich die Schwellenwerte der Zielgrößenkategorie für nicht schwerwiegende Symptome und Nebenwirkungen auf Basis der üblichen Einteilung von Cohen's d nachvollziehbar herleiten [1]. Diese sehen einen kleinen Effekt mit „0.2“ (Ausmaßkategorie gering) und einen mittleren Effekt „0.5“ (Ausmaßkategorie beträchtlich) vor. Im Gegensatz zu den Schwellenwerten der Zielgrößenkategorie für nicht schwerwiegende Symptome und Nebenwirkungen, sind die Schwellenwerte in der Zielkategorie für schwerwiegende Symptome, Nebenwirkungen und gesundheitsbezogene Lebensqualität nicht nachvollziehbar und auch nicht wissenschaftlich begründet.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	Für die Ableitung der entsprechenden Schwellenwerte kann man laut IQWiG zeigen, dass für ein bestimmtes Ausmaß die wünschenswerte Effektstärke ungefähr halbiert werden muss. Für dieses Vorgehen findet sich keine wissenschaftliche Rationale. Zusätzlich basiert die Ableitung des Ausmaßes des Zusatznutzens darauf, dass das gesamte Konfidenzintervall vollständig oberhalb des entsprechenden Schwellenwerts liegen muss (sog. Verschiebung der Nullhypothese). Dieser methodische Ansatz entspricht zum einen nicht dem nach Cohen vorgeschlagenen Vergleich mit Punktschätzer und zum anderen sind die Werte nicht auf die Zielkategorien des Behandlungseffektes von Symptomen, Nebenwirkungen und gesundheitsbezogene Lebensqualität ausgelegt [1]. In gleicher Weise ist die Erforderlichkeit einer Unterteilung in unterschiedliche Zielgrößenkategorien nicht ersichtlich. Auch werden die vom IQWiG vorgeschlagenen Schwellenwerte weit höher liegen, als die nach Cohen's vorgeschlagenen Werte auf Basis der Punktschätzer. In der Praxis kann die Orientierung an Konfidenzintervallgrenzen dazu führen, dass die tatsächlich beobachteten Punktschätzer 2- bis 3-mal über den nach Cohen's vorgeschlagenen Werten (z. B. „0,2“) liegen können. Damit lägen die Punktschätzer in einem Bereich, der nach Cohen bereits als mittlerer Effekt gilt, jedoch nach IQWiG-Ansatz als geringer Effekt klassifiziert werden müsste. Mit diesem konservativen Ansatz werden so die Hürden einen relevanten und quantifizierbaren Vorteil zeigen zu können, deutlich erhöht. <u>Vorgeschlagene Änderung:</u> Dem vom IQWiG geschilderte Vorgehen sollte eine wissenschaftlichen fundierte Rationale unterliegen. So sollten die Schwellenwerte zur Ausmaßbestimmung bei stetigen Daten nachvollziehbar und wissenschaftlich begründet werden und den international anerkannten Standards und Kriterien der evidenzbasierten Medizin genügen. In gleicher Weise sollte berücksichtigt werden, dass die Schwellenwerte nach Cohen's für den Punktschätzer gelten und nicht für das Konfidenzintervall intendiert sind.
Abschnitt 9.3.3 (S.178)	<u>Anmerkung:</u> Folgender Abschnitt 9.3.3 (Beurteilung klinischer Relevanz) wurden im Entwurf der Allg. Methoden v6.0 ergänzt bzw. geändert, zu der BMS Stellung nimmt:

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	„Eine Sichtung von systematischen Übersichtsarbeiten zu MIDs [z. B. [523,655]] hat ergeben, dass sich MIDs in vielen Fällen zwischen 10 % und 20 % der Spannweite der Scores eines Erhebungsinstruments bewegen. Daraus ergibt sich regelhaft folgendes Vorgehen: 1) Falls in einer Studie Responderanalysen unter Verwendung einer MID präspezifiziert sind und das Responsekriterium mindestens 15 % der Skalenspannweite des verwendeten Erhebungsinstruments entspricht, werden diese Responderanalysen ohne weitere Prüfung des Responsekriteriums für die Bewertung herangezogen. 2) Falls präspezifiziert Responsekriterien im Sinne einer MID unterhalb von 15 % der Skalenspannweite liegen, werden diese regelhaft nicht herangezogen. In diesen Fällen und solchen, in denen gar keine Responsekriterien präspezifiziert wurden, aber stattdessen Analysen kontinuierlicher Daten zur Verfügung stehen, bestehen verschiedene Möglichkeiten. Entweder können die Analysen der kontinuierlichen Daten herangezogen werden, für die Relevanzbewertung wird dabei auf ein allgemeines statistisches Maß in Form von standardisierten Mittelwertdifferenzen (SMDs, in Form von Hedges' g) zurückgegriffen. Dabei wird eine Irrelevanzschwelle von 0,2 verwendet: Liegt das zum Effektschätzer korrespondierende Konfidenzintervall vollständig oberhalb dieser Irrelevanzschwelle, wird davon ausgegangen, dass die Effektstärke nicht in einem sicher irrelevanten Bereich liegt. Dies soll gewährleisten, dass der Effekt hinreichend sicher mindestens als klein angesehen werden kann [233]. Alternativ können post hoc spezifizierte Analysen mit einem Responsekriterium von genau 15 % des Skalenrangs berücksichtigt werden. 3) Liegen sowohl geeignete Responderanalysen (Responsekriterium präspezifiziert mindestens 15 % des Skalenrangs oder post hoc genau 15 % des Skalenrangs) als auch Analysen stetiger Daten vor, werden die Responderanalysen herangezogen.“ BMS stimmt zu, dass die Beurteilung klinischer Relevanz ein wichtiges Thema ist und eine geeignete Definition gefunden werden sollte. Jedoch ist das vorgeschlagene Vorgehen nicht nachvollziehbar. Grundsätzlich kann es für die Beurteilung klinischer Relevanz keine allgemeingültige Responderschwelle geben, da Fragebögen und Indikationen individuell sind. Zum einen wirkt die Basis der vom IQWiG gewählten MID von 15 % willkürlich. Zur Herleitung dieser Schwelle bezieht sich das IQWiG auf „eine Sichtung von systematischen Übersichtsarbeiten“.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	Hier bleibt unklar, ob diese Sichtung vollständig ist, da sich die beiden einzigen vom IQWiG exemplarisch angegebenen Quellen (Nordin 2016, St-Pierre 2016) lediglich auf ausgewählte und sehr spezifische Symptome wie „fatigue“ und „symptoms and functional limitations in individuals with rotator cuff disorders“ beziehen. Zum anderen kann auf Basis der MID von 15 % keine klinische Beurteilung für bestimmte Zahlenbereiche, wie z.B. 6MGT (6-Minuten-Gehstrecke), stattfinden. Zudem sollte die Beurteilung einer klinischen Relevanz auf Basis einer individuellen Skala auch durch eine individuelle Validierungsstudie vertreten werden. So widerspricht eine allgemeingültige MID von 15 % den vom IQWiG genannten Anforderungen an Validierungsstudien zur Beurteilung klinischer Relevanz in den Nutzenbewertungen. So schrieb das IQWiG in der Nutzenbewertung von Bosutinib: <i>„Bei der MID handelt es sich jedoch um eine von den Patientinnen und Patienten auf individueller Ebene selbst als wichtig wahrgenommene Änderung im interessierenden Endpunkt (hier der Gesundheitszustand) [26-28] ... Diese Anforderungen werden von der aktuellen methodischen Diskussion zur Bestimmung einer validen MID unterstrichen [29-33].“ [2]</i> Diese aktuellen Diskussionen sollte fortgeführt werden, um so den Anspruch an Validierungsstudie definieren zu können. Hierzu könnten Kriterien sein: • Anker-basiert • Anzahl von Veränderungen des PROs • Korrelation zwischen Anker und Veränderung des PROs [3] Auch zeigt sich auf Basis von bisher in AMNOG-Verfahren akzeptierten MIDs, dass die vorgeschlagene 15 %-Schwelle in vielen Fälle zu einer Erhöhung der bisher validierten und in der Anwendung akzeptierten MIDs führen würde. So würde z.B. für das in der Onkologie sehr häufig verwendete Erhebungsinstrument EORTC QLQ-C30 eine sehr deutliche Erhöhung der MID um 50 % resultieren (von bisher 10 auf 15) [4]. In diesem speziellen Fall und in allen anderen ähnlich gelagerten Fällen, wäre es nicht nachvollziehbar, warum etablierte MIDs, deren klinische Relevanz bereits belegt ist, im Rahmen der Nutzenbewertung nicht verwendet werden sollten. <u>Vorgeschlagene Änderung:</u>

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	Es sollten Standards für Validierungsstudien zur Beurteilung klinischer Relevanz definiert werden. Hierzu beteiligt sich BMS gern an den weiteren aktuellen Diskussionen und schlägt vor an Kriterien, wie z.B. Anker-basiert oder Anzahl der Messungen (Längsschnitt) weiterzuarbeiten. Bis dahin sollten etablierte MIDs auch weiterhin in der Nutzenbewertung auf Basis einer Responderanalyse zur Ableitung des Zusatznutzen herangezogen werden können.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

[1]. Cohen J. Statistical Power Analysis in the Behavioral Sciences (2nd edition). Hillsdale (NJ): Lawrence Erlbaum Associates, Inc., 1988

[2]. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). IQWiG-Berichte – Nr. 660. Bosutinib (chronische myeloische Leukämie)– Nutzenbewertung gemäß § 35a SGB V vom 03.09.2018. https://www.g-ba.de/downloads/92-975-2442/2018-06-01_Nutzenbewertung-IQWiG_Bosutinib_D-355.pdf Zugriff am 29.01.2020

[3]. Coon CD, Cook KF. Moving from significance to real-world meaning: methods for interpreting change in clinical outcome assessment scores. Qual Life Res 2018; 27(1): 33-40.

[4]. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. J Clin Oncol. 1998;16(1):139-44. Epub 1998/01/24.

A.1.9 – Bundesverband der Arzneimittel-Hersteller e. V. (BAH)

Autorinnen und Autoren

- Poß, Rudolf

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Poß, Rudolf
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
☒ im Namen folgender Institution / Organisation:
Bundesverband der Arzneimittel-Hersteller e.V., Friedrichstraße 134, 10117 Berlin
☐ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

<i>Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.</i>

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
Abschnitt 3.3.3 „Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V“, S. 67, Abschnitt D) Stetige oder quasistetige Zielgrößen mit jeweils vorliegenden SMDs	<u>Anmerkung:</u> Ein methodisches Vorgehen, welches das Ziel verfolgt, auch bei stetigen Daten die Quantifizierung der Effekte standardisiert darzustellen, ist grundsätzlich zu begrüßen. Das Vorgehen sollte jedoch wissenschaftlich begründet und konsistent zur Vorgehensweise in vergleichbaren Situationen sein. So sind die Schwellenwerte in der Zielkategorie „schwerwiegende Symptome, Nebenwirkungen und gesundheitsbezogene Lebensqualität“ nicht nachvollziehbar und eine wissenschaftliche Herleitung dazu war nicht zu finden. Gemäß Vorschlag des IQWiG sollen für die Ableitung der entsprechenden Schwellenwerte die Mittelwerte der üblichen Cohen's d-Bereiche gebildet und danach ungefähr halbiert werden. Eine wissenschaftliche Rationale für dieses Vorgehen ist nicht erkennbar. Aus unserer Sicht entspricht es damit auch nicht den international anerkannten Kriterien oder Standards der evidenzbasierten Medizin. Zugleich ist nicht nachvollziehbar, warum eine Unterteilung in unterschiedliche Zielgrößenkategorien überhaupt erforderlich sein sollte. Die beiden Zielgrößenkategorien in „Tabelle 6: Schwellenwerte zur Ausmaßbestimmung für die SMD“ sollten wissenschaftlich gleich behandelt werden. <u>Vorgeschlagene Änderung:</u> Die gängige und in der Wissenschaft akzeptierte Einteilung nach Cohen's d sollte in gleicher Weise für unterschiedliche

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	Zielgrößenkategorien bei stetigen Zielgrößen mit SMD Anwendung finden. Das Vorgehen sollte nachvollziehbar und wissenschaftlich begründet werden und den international anerkannten Standards und Kriterien der evidenzbasierten Medizin genügen.
Abschnitt 9.3.3 „Beurteilung klinischer Relevanz“, S. 177/178	<u>Anmerkung:</u> In Abschnitt 9.3.3 behandelt das IQWiG das methodische Vorgehen zur Beurteilung klinischer Relevanz von Gruppenunterschieden. Im Fall von Responderanalysen werden dazu sogenannte Minimal Important Differences (MIDs) verwendet. Im Gegensatz zum bisherigen Vorgehen will das IQWiG künftig nur noch MIDs akzeptieren, die über 15 % der Skalenspannweite des jeweiligen Messinstruments liegen, ansonsten soll als einheitliche MID 15 % der Skalenspannweite verwendet werden. Diese Festlegung ist aus verschiedenen Gründen höchst problematisch: - Eine MID, die sich direkt mittels eines mathematischen Algorithmus aus der Skalenspannweite berechnet, lässt die klinische Relevanz von Unterschieden völlig außer Acht. - Eine einheitliche Grenze wird den Unterschieden zwischen vorkommenden Skalen und Indikationen nicht gerecht. - Die resultierenden MID würden in vielen Fällen erheblich von bereits anerkannten Werten abweichen. - Die Begründung der gewählten Grenze von 15 % durch das IQWiG ist methodisch nicht nachvollziehbar. - Für Endpunkte mit offenen Skalen ist die Methode prinzipiell nicht anwendbar (z. B. 6-Minuten-Gehtest - Strecke in Meter, Skala nach oben offen). Auch wenn die Wahl der Skalenspannweite im Regelfall begründet und nachvollziehbar sein mag, gibt es keine grundsätzliche Korrelation zwischen Skalenspannweite und MID, wie sie der IQWiG-Vorschlag beinhaltet. Eine wissenschaftliche Herleitung für das Konzept liefert das Institut nicht. Würden Relevanzgrenzen auf diese Weise bestimmt, so wären sie beliebig und zur Beurteilung der klinischen Relevanz von Gruppenunterschieden völlig ungeeignet. Nach dem aktuellen Stand der Wissenschaft besteht ein breiter Konsens darüber, dass MIDs auch indikations- oder domänenspezifisch sein können, da die Relevanz von Veränderungen von der Patientenpopulation, der

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	Krankheitsschwere und anderen kontextuellen Faktoren abhängt. Das schreibt auch das IQWiG selbst im vorliegenden Entwurf für das Methodenpapier (S. 177 dritter Abschnitt unten). Ein spezifisches Instrument kann daher in verschiedenen Indikationen unterschiedliche MIDs aufweisen. Ein Beispiel ist die Fatigue Severity Scale (FSS) mit einer MID von 20,2 bei rheumatoider Arthritis und einer MID von 2 – 7 in der Indikation Lupus (entsprechen 37,4 % bzw. 3,7 – 13,0 % der Skalenspannweite). Dies zeigt, dass ein festgesetzter Prozentsatz der Skalenspannweite nicht als Grenzwert zur Bestimmung einer MID für alle Instrumente in allen Indikationen geeignet sein kann. Die resultierenden MIDs würden erheblich von den in der Nutzenbewertung bislang anerkannten MIDs abweichen, obwohl diese mit Hilfe von in der Wissenschaft weithin akzeptierten Methoden validiert wurden. Einige Beispiele von MIDs, die bislang durch das IQWiG und den G-BA akzeptiert worden sind und die mit dem Vorschlag des IQWiG für nichtig erklärt würden sind FACT-P-Gesamtscore (6,4 %), FACT-G-Gesamtscore (4,6 %), FACIT-Fatigue (7,7 %) oder EORTC QLQ-C30 (10,0 %). In Klammern ist jeweils der entsprechende Wert des Verhältnisses MID zu Skalenspannweite als prozentualer Anteil angegeben. Die Abweichungen zu der postulierten Grenze von 15% sind erheblich. Die Herleitung des Wertes 15 % ist nicht nachvollziehbar. So ist die Mehrheit der in der zitierten Literatur aufgeführten MIDs kleiner als 15 % und würde daher künftig als zu klein abgelehnt. Die angeführte Literatur kann daher nicht als Begründung der 15 %-Grenze dienen. Es bleibt unklar, wie das IQWiG aus der angegebenen Literatur den Wert von 15 % abgeleitet hat. Darüber hinaus ist nicht nachvollziehbar, warum das IQWiG in seine eigene wissenschaftliche Argumentationskette unterbricht. Auf S. 177 wird ausgeführt, derzeit gebe es keine akzeptierten Standards, mit denen die Aussagekraft ermittelter MIDs abgeschätzt werden könne. Es werde an der Entwicklung eines solchen Standards gearbeitet, mit offenem Ergebnis. Anstatt auf den Stand der internationalen wissenschaftlichen Diskussion einzugehen und daraus einen logisch nachvollziehbaren Ansatz zur Operationalisierung aussagekräftiger MIDs abzuleiten, präsentiert das IQWiG daraufhin ohne Herleitung seinen 15%-Ansatz. Dabei sind, basierend auf der Diskussion der wissenschaftlichen Community, verschiedene Ansätze denkbar, um die Aussagekraft von MIDs zu steigern. Beispielsweise gibt es die grundsätzlich verschiedenen Ansätze der ankerbasierten und verteilungsbasierten Berechnungsmethoden, die beide Vor- und

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	Nachteile haben. Eine Kombination bzw. Synthese beider Methoden könnte möglicher Weise die Probleme der Einzelansätze minimieren. Da MIDAs auf patient reported outcomes (PRO) beruhen, könnten alternativ Tools zur strukturierten Bewertung von PRO auch für die Abschätzung der Validität von MIDAs hilfreich sein. Auf jeden Fall sollte ein Verfahren zur Bewertung der MID-Relevanz die Abhängigkeit vom Kontext berücksichtigen, um eine faire Beurteilung zu gewährleisten. <u>Vorgeschlagene Änderung:</u> Die vorgeschlagenen Änderungen zur MID-Bewertungsmethodik sollten gänzlich entfallen. Die in bisherigen Nutzenbewertungsverfahren bereits akzeptierten MIDAs sollten auch weiterhin herangezogen werden. Eine strukturierte Methodik sollte auf dem aktuellen Stand der Wissenschaft aufsetzen. Sie sollte in einem offenen Austausch mit wissenschaftlichen Experten entwickelt werden. Dabei ist ein wissenschaftlicher Konsens anzustreben. Erst danach ist eine Anpassung des Methodenpapiers in diesem Zusammenhang sinnvoll.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

A.1.10 – Bundesverband der Pharmazeutischen Industrie e. V. (BPI)

Autorinnen und Autoren

- Wilken, Matthias

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Wilken, Matthias; Dr.
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
☒ im Namen folgender Institution / Organisation: Bundesverband der Pharmazeutischen Industrie e. V.
☐ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
Abschnitt 9.3.3	<u>Anmerkung:</u> Im Abschnitt 9.3.3 wird das methodische Vorgehen zur Beurteilung der klinischen Relevanz von Gruppenunterschieden beschrieben. Für Responderanalysen werden dazu sogenannte Minimal Important Difference (MID) verwendet, die üblicherweise mit ankerbasierten Methoden, ggf. auch in Verbindung mit verteilungsbasierten Methoden, bestimmt werden. Abweichend von der bisherigen Praxis ist beabsichtigt, zukünftig nur noch MID zu akzeptieren, die $\geq 15\%$ der Skalenspannweite des jeweiligen Messinstruments sind. Ansonsten soll als einheitliche MID 15% der Skalenspannweite verwendet werden. Diese Neufestlegung erscheint aus mehreren Gründen problematisch. Zum einen werden die Skalenspannweiten nicht im Hinblick auf die klinische Relevanz von Unterschieden ausgewählt. Zudem wird die vorgesehene einheitliche Grenze Unterschieden zwischen vorkommenden Skalen und Indikationen nicht gerecht. Im Vergleich zum status quo würden die resultierenden MID in vielen Fällen erheblich von bereits anerkannten Werten abweichen. Die gewählte Grenze von 15% erscheint wissenschaftlich nicht hinreichend unterlegt und insofern gegriffen. Für Endpunkte mit offenen Skalen ist die Methode prinzipiell nicht anwendbar. MID können auch indikations- oder domänenspezifisch sein, da die Relevanz von Veränderungen u. a. von der Patientenpopulation, der Krankheitsschwere und anderen kontextuellen Faktoren abhängt. Für verschiedene Indikationen kann ein identisches Instrument daher unterschiedliche MID aufweisen. Daher kann ein

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	festgesetzter Prozentsatz der Skalenspannweite nicht als Grenzwert zur Bestimmung einer MID übergreifend für alle Instrumente sowie alle Indikationen geeignet sein. Bleibe das IQWiG bei der Festlegung der aktuell vorgesehenen Schwellenwerte, würden die resultierenden MID zum Teil erheblich von im Rahmen von Nutzenbewertungsverfahren bislang anerkannten MID abweichen, obwohl diese mit Hilfe von in der Wissenschaft weithin akzeptierten Methoden validiert und in der Vergangenheit von IQWiG und G-BA auch akzeptiert wurden. Es bleibt unklar, auf welcher Grundlage aus der angegebenen Literatur ein Wert von 15 % abgeleitet worden ist. Darüber hinaus ist nicht nachvollziehbar, warum gerade die beiden angeführten systematischen Übersichtsarbeiten zur Ableitung eines einheitlichen Grenzwerts herangezogen worden sind. Eine Limitation der Methode besteht darüber hinaus bei Skalen, die nach oben oder unten (oder beiden Seiten) unbegrenzt sind. Für diese lässt sich keine Spannweite und daher nach der vorgeschlagenen Methode auch keine MID bestimmen. Die vorgesehenen Neufestlegungen bedeuten im Ergebnis die Abkehr von einer etablierten Vorgehensweise, wobei die neue Methodik derzeit auf Basis der zur Verfügung stehenden Informationen weder als evidenzbasiert noch wissenschaftlich anerkannt bezeichnet werden kann. <u>Vorgeschlagene Änderung:</u> Eine erneute Überarbeitung des Abschnitts 9.3.3 „Beurteilung klinischer Relevanz“ wird für erforderlich gehalten. Dabei ist insbesondere darzulegen, wie die methodische Qualität von Studien zur Validierung von MID sowie die Aussagekraft der ermittelten MID aus diesen etablierten Verfahren (anker- und verteilungsbasierte Herleitung) bewertet werden soll. Etablierte MIDs, deren klinische Relevanz bereits belegt ist, sollten unter Berücksichtigung neuer wissenschaftlicher Erkenntnisse im Rahmen der Nutzenbewertung weiterhin auf Akzeptanz stoßen. Insbesondere sollten krankheits- und populationsspezifische Einflüsse ebenso wie Skalencharakteristika nicht ignoriert werden. Die aktuell vorgeschlagene Methodik könnte aufgrund der dargestellten Limitationen allenfalls als Notlösung für den Fall Anwendung finden, dass keine geeignete MID vorliegt. Die Wahl des Grenzwertes sollte in diesem Fall evidenzbasiert erfolgen und nachvollziehbar sein.
	<u>Anmerkung:</u>

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<u>Vorgeschlagene Änderung:</u>

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

*Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenz-
liste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.*

A.1.11 – Bundesverband Medizintechnologie e. V. (BVMed)

Autorinnen und Autoren

- Winkler, Olaf

Bundesverband
Medizintechnologie e.V.
Reinhardtstraße 29b
10117 Berlin
Tel. +49 (0)30 246 255 - 0
Fax +49 (0)30 246 255 - 99
info@bvmed.de
www.bvmed.de

Berlin, 31. Januar 2020

[REDACTED]

Stellungnahme

zum

**Entwurf des allgemeinen Methodenpapiers Version 6.0
des IQWiG**

Die vorliegende Stellungnahme gibt die Auffassung der Mitglieder des Fachbereiches Market Access der AG Methodenbewertung des Bundesverbands Medizintechnologie wieder.

[REDACTED]

[REDACTED]

I. Grundsätzliche Anmerkung

Die Methoden des IQWiG zur Bewertung neuer Untersuchungs- und Behandlungsmethoden mit Medizinprodukten muss sich an die Dynamik und Besonderheiten der Technologien kontinuierlich anpassen. Die derzeitige Ausgestaltung des Methodenpapieres ist im vorliegenden Entwurf nicht vollumfänglich ausreichend, um die Besonderheiten der Bewertung neuer Untersuchungs- und Behandlungsmethoden mit Medizinprodukten in der Arbeitsweise des IQWiG zu berücksichtigen. Wir geben mit dieser Stellungnahme Hinweise zur sachgerechten Anpassung der methodischen Werkzeuge mit dem Angebot des weiteren Dialoges.

Bei kritischen Fragestellungen, die derzeit auf der alleinigen Expertise des IQWiGs beruhen (wie zum Beispiel die Einschätzung der qualitativen Ergebnissicherheit) empfehlen wir die Einbeziehung von Fachgesellschaften. So kann eine objektivere Entscheidung beruhend auf der bestmöglichen Fachexpertise getroffen werden.

BVMed bietet weiterhin Expertise für das Kuratorium an

Die Besonderheiten der Bewertung neuer Untersuchungs- und Behandlungsmethoden mit Medizinprodukten, insbesondere hoher Risikoklassen, sollen, im Gegensatz zu den Methoden der Nutzenbewertung von Arzneimitteln, ausreichend in der Arbeitsweise des IQWiG berücksichtigt werden. Wir halten es weiterhin für angezeigt, dass der BVMed zukünftig für die Medizinprodukteindustrie aktiv beratend im Kuratorium der Stiftung vertreten ist, um hier die Expertise der Medizinprodukteindustrie zur Verfügung zu stellen.

II. Anmerkungen zu den einzelnen Kapiteln

1.2 Evidenzbasierte Medizin

Zu 1.2.5 Die Verbindung zwischen Ergebnissicherheit und Alltagsnähe

Änderungsvorschlag:

Großen Wert auf die Abschätzung der Ergebnissicherheit zu legen, wird oft kritisiert. Ein Argument ist, dass Studien mit hoher Ergebnissicherheit, insbesondere randomisierte kontrollierte Studien, zwar eine hohe interne Validität besäßen, jedoch oft die Versorgung von Patientinnen und Patienten im Alltag nicht abbildeten und deshalb nicht „übertragbar“ seien, also nur eine geringe externe Validität vorhanden sei. Dabei ist zu prüfen, wie gut die in Studien untersuchte Patientenpopulation, die eingesetzten Interventionen und die Zielkriterien dem Versorgungsalltag entsprechen. Diese Kritik ist dann oft mit der Forderung verbunden, andere Studientypen ohne Randomisierung hinzuzunehmen, um Alltagsbedingungen besser einfließen zu lassen.

Diese Kritik vermischt jedoch Argumentationsebenen, die klar voneinander zu trennen sind. Folgende Aspekte müssen Beachtung finden:

- 1) Grundlage einer Nutzenbewertung ist der Nachweis von Kausalität. Unverzichtbare Bedingung für den Nachweis von Kausalität ist ein vergleichendes Experiment, das so angelegt sein muss, dass ein Unterschied zwischen Interventionsgruppen – ein Effekt – nur auf eine einzige Einflussgröße – die geprüfte Intervention – zurückgeführt werden kann. Dieses Ziel macht für klinische Studien erhebliche Anstrengungen nötig, weil es zahlreiche unerwünschte Einflüsse gibt, die einen Effekt vortäuschen oder auch verbergen können (Verzerrung). Der stärkste dieser verzerrenden Einflüsse sind ungleiche Ausgangsbedingungen zwischen den Vergleichsgruppen. Die Randomisierung zusammen mit einer sorgfältigen Verdeckung der Zuteilung (Allocation Concealment) ist das beste gegenwärtig verfügbare Instrument, um diese Verzerrung zu minimieren. Die zufällige Zuteilung der Teilnehmer zu den Gruppen stellt sicher, dass es keine systematischen

Unterschiede zwischen den Gruppen gibt, und zwar weder bezüglich bekannter Eigenschaften (z. B. Alter, Geschlecht, Schwere der Erkrankung), noch bezüglich unbekannter Faktoren. Damit ist in randomisierten kontrollierten Studien (Randomized controlled Trials, RCTs) eine grundlegende Voraussetzung für einen Kausalitätsnachweis gegeben. Allerdings ist auch eine Randomisierung allein kein Garant für eine hohe Ergebnissicherheit. Dazu sind z. B. auch eine unverzerrte Erhebung, Zusammenfassung und Veröffentlichung der Ergebnisse nötig.

- 2) Andere Studientypen als RCTs sind in der Regel für einen Kausalitätsnachweis nicht geeignet. In nicht randomisierten vergleichenden Studien kann grundsätzlich nicht von einer Strukturgleichheit der Gruppen ausgegangen werden. Sie liefern daher immer ein potenziell verzerrtes Ergebnis und können zumeist die maßgebliche Frage nicht hinreichend sicher beantworten, ob ein Unterschied auf der geprüften Intervention beruht. Daher bedarf die Verwendung nicht randomisierter Studien, als Nachweis der Kausalität, einer Intervention einer besonderen Begründung bzw. besonderer Voraussetzungen und spezieller Qualitätsanforderungen.
- 3) Es trifft zu, dass viele randomisierte Studien Aspekte des Alltags der Patientenversorgung nicht abbilden, z. B. Patientinnen und Patienten mit Begleiterkrankungen ausschließen, die im Alltag häufig sind. Dies ist jedoch keine Konsequenz der Technik der Randomisierung, sondern die Folge anderer Faktoren wie z. B. der Festlegung enger Ein- und Ausschlusskriterien der Studie, der Auswahl der Interventionen oder der Zielkriterien. Zudem werden Patientinnen und Patienten in randomisierten Studien oft anders (intensiver und enger) betreut als im Alltag. Das alles sind aber bewusste Entscheidungen derjenigen, die in einer Studie eine bestimmte Frage beantworten wollen. Durch einen Verzicht auf die Randomisierung werden diese Entscheidungen nicht geändert. Auch in nicht randomisierten Studien findet eine Selektion der Teilnehmer durch Ein- und Ausschlusskriterien und ggf. andere Designcharakteristika statt, sodass externe Validität ebenso wenig per se gegeben ist wie bei RCTs.
- 4) Aber selbst wenn sich Patientengruppen einer RCT vom Versorgungsalltag unterscheiden, muss das die externe Validität des Ergebnisses nicht infrage stellen. Entscheidend ist vielmehr die Frage, ob zu erwarten ist, dass der in einer Population festgestellte Therapieeffekt in einer anderen Population anders ist.
- 5) Es hängt vom Einzelfall ab, wie sich die Intensität der Betreuung in einer Studie auswirken könnte. Zum Beispiel ist denkbar, dass ein Nutzen tatsächlich nur in den Händen besonders qualifizierter Ärztinnen und Ärzte vorhanden ist, weil unter Alltagsbedingungen ansonsten zu häufig Komplikationen auftreten. Alternativ bieten sich auch in diesen Fällen Studiendesigns deskriptiver Beobachtungsstudien in Form von prognostischen Korrelations- und Querschnittstudien aus hoch standardisierten Real-World-Evidenz Datenbanken an, die Unterschiede auf der Ebene der Arztqualifikation und Erfahrung, auf Grundlage der großen Datenmengen statistisch ermittelbar machen können. Es ist aber auch möglich, dass eine enge Betreuung Unterschiede zwischen Gruppen eher verringert. Zum Beispiel könnten Unterschiede in der Therapieadhärenz in Studien kleiner ausfallen, in denen Patientinnen und Patienten grundsätzlich sehr eng betreut werden.
- 6) Die Festlegung der Studienbedingungen liegt jedoch in der Hand der Initiatorinnen und Initiatoren einer klinischen Studie. Sie können die Fragestellung und Endpunkte definieren, die sie als so wichtig einschätzen, dass sie in der Studie beobachtet werden sollen. Wenn eine Herstellerin oder ein Hersteller eines Arzneimittels z. B. Therapieadhärenz für einen wichtigen Aspekt des Nutzens seines Produkts hält, wäre die nahe liegende Konsequenz, dann Studien aufzulegen, die diesen Aspekt mit möglichst großer Ergebnissicherheit und Nähe zu Alltagsbedingungen messen können und gleichzeitig die Patientenrelevanz demonstrieren.

Diese Ausführungen zeigen, dass Ergebnissicherheit und Alltagsnähe (oder interne und externe Validität) nicht fest miteinander verknüpft sind; Hohe Ergebnissicherheit und Alltagsnähe schließen einander nicht aus, sondern erfordern nur eine geeignete Kombination aus Studientyp, Design und Durchführung.

Auch wenn die Kritik mangelnder Alltagsnähe tatsächlich für viele Studien zutreffen mag, wäre nichts dadurch gewonnen, wenn man eine hohe Ergebnissicherheit für eine größere Alltagsnähe aufgeben würde. Man würde einen Mangel durch Inkaufnahme eines zweiten, noch schwerwiegenderen Mangels zu kompensieren versuchen.

Es sind Studien wünschenswert und machbar, die Alltagsnähe und hohe Ergebnissicherheit vereinigen. RCTs, in denen weder über den Behandlungsalltag hinausgehende Anforderungen an Patientinnen und Patienten gestellt noch definierte Studienvisiten vorgegeben werden, sind sehr wohl möglich. Sie werden international als Real-World Trials, Practical Trials oder Pragmatic Trials diskutiert. Allerdings können solche sogenannten pragmatischen Studien ihrerseits ebenfalls zu Interpretationsproblemen führen. Werden z. B. sehr breite Einschlusskriterien gewählt, stellt sich die Frage, ob die (Gesamt-)Studienergebnisse auf die gesamte Studienpopulation anwendbar sind, was letztlich über adäquate Subgruppenanalysen zumindest ansatzweise zu beantworten wäre. Der entscheidende Vorteil deskriptiver Beobachtungsstudien in Form von prognostischen Korrelations- und Querschnittstudien aus hoch standardisierten Real-World-Evidenz Datenbanken ist, dass klinische Effekte ermittelbar sind die sich in kohortenspezifischen Trends zeigen, auf Basis derer man Potentiale für einen Nutzen oder Schaden ableiten kann.

Begründung:

Mit dem Implantateregister - Errichtungsgesetz hat der Gesetzgeber eine Grundlage geschaffen, die neben dem Schutz der Gesundheit und Sicherheit von Patientinnen und Patienten, von Anwendern und von Dritten sowie der Abwehr von Risiken durch Implantate, explizit die Informationsgewinnung über die Qualität, und die Qualitätssicherung der Implantate und der medizinischen Versorgung mit Implantaten in den verantwortlichen Gesundheitseinrichtungen verfolgt. Dieser Zweck führt in der Berücksichtigung von Datenquellen und den daraus abzuleitenden Studiendesigns zu den deskriptiven Beobachtungsstudien in Form von prognostischen Korrelations- und Querschnittstudien. Insbesondere in der Bewertung von Nutzen und Schaden nicht-Medikamentöser Methoden und Interventionen unter der Anwendung fortschrittlicher Medizinprodukte darf die Berücksichtigung von Studienergebnissen aus diesem Kontext nicht ausbleiben und ist entsprechend in das Methodenpapier aufzunehmen.

3.1 Patientenrelevanter medizinischen Nutzen und Schaden

Zu 3.1.4 Endpunktbezogene Bewertung

B) Ableitung der Beleglage und Aussagesicherheit

Änderungsvorschlag:

Aus lediglich 1 Studie kann im Ausnahmefall für einen Endpunkt ein Beleg für eine spezifische (Teil-)Population abgeleitet werden, wenn ein Studienbericht gemäß den ICH-Richtlinien vorliegt, die auch sonst für einen Beleg geforderten Bedingungen erfüllt sind und darüber hinaus die Studie folgende besondere Anforderungen erfüllt:

- Die Studie ist multizentrisch, in jeden Studienarm wurden der wissenschaftlichen Fragestellung entsprechend Patientinnen oder Patienten eingeschlossen.
- Das Ergebnis ist innerhalb der Studie konsistent: Für die interessierende (Teil-)Population liegen Auswertungen verschiedener weiterer Teilpopulationen (insbesondere Teilmengen von Studienzentren) vor, die jeweils bewertbare und ausreichend homogene Effektschätzer ergeben. Diese Konsistenzbewertung ist bei binären Daten nur möglich, falls eine gewisse Mindestanzahl von Ereignissen vorliegt.

- Die im vorherigen Punkt adressierten Auswertungen für Teilpopulationen liegen für alle relevanten Endpunkte vor, d. h. diese Auswertungen sind nicht auf einzelne ausgewählte Endpunkte beschränkt.

Es ist möglich, dass sich die Beleglage beim Vorliegen von lediglich einer Studie, die für sich genommen nur einen Hinweis oder Anhaltspunkt ergibt, durch zusätzliche indirekte Vergleiche ändert. Hierfür sind allerdings hohe methodische Anforderungen an die indirekten Vergleiche zu stellen (siehe Abschnitt 9.3.9). Außerdem ist es bei entsprechend homogener Datenlage möglich, dass sich durch Ergänzung indirekter Vergleiche die Präzision der Effektschätzung erhöht, was bei der Feststellung des Ausmaßes des Zusatznutzens eine wichtige Rolle spielt (siehe Abschnitt 3.3.3).

Begründung:

Der Umfang der Stichprobengröße steht in direkter Verbindung zur wissenschaftlichen Fragestellung und den daraus abzuleitenden statistischen Testverfahren. Entsprechende Kalkulationsmethoden sind in der Humanwissenschaft validiert anzuwenden und transparent zu machen. (Röhrig B et al. Fallzahlplanung in klinischen Studien, *Deutsches Ärzteblatt Int* 2010; 107(31–32): 552–6 DOI: 10.3238/arztebl.2010.0552) Das Voraussetzen von Stichprobengrößen und Zentrenzahlen ist unwissenschaftlich und methodisch nicht haltbar. In der biostatistischen Wissenschaftsgemeinde ist die Debatte um das Signifikanzniveau eine derzeit im Schwerpunkt der „scientific method“ geführte Auseinandersetzung, die jedoch noch keinen einheitlichen Konsens gefunden hat (Ioannidis JPA. The Proposal to Lower P Value Thresholds to .005. *JAMA*. 2018; 319(14):1429–1430. doi:10.1001/jama.2018.1536). Mit der Aufnahme derart sensibler Signifikanzniveaus setzt das Methodenpapier einen Standard, der in der wissenschaftstheoretischen Herangehensweise der Humanwissenschaften in diesem Maß noch keinen Konsens hat. Das IQWiG sollte in seinem Methodenpapier die etablierten Standards berücksichtigen. Die derzeitige Diskussion bezüglich des Sachverhaltes sollte durch das Methodenpapier nicht weiteren kontroversen Zuspruch bekommen, bis ein Sensitivitätsniveau für den p-Wert konsentiert ist. (0.05 or 0.005? P-value Wars Continue. The p-value is under fire yet again, but this time with some quick-and-dirty solutions (and some long-and-onerous ones too) to the problems created by relying on this quick-and-dirty test. <https://sciencebasedmedicine.org/0-05-or-0-005-p-value-wars-continue/>) Dem gegenübergestellt ist die Gewichtung des p-Wert alleine unvollständig. Der p-Wert bzw. die Nullhypothese muss in der Prüfgröße getestet werden, welches über Einseitige, oder im Fall eines ungerichteten Tests, über zweiseitige Signifikanztests durchführbar ist. Erst hierüber kann innerhalb der statistischen Ergebnisse entsprechende Validität abgeleitet werden, die eine haltbare Bewertung von Nutzen oder Schaden im Falle einer 1-Studien-situation zulässt.

3.2 Spezielle Aspekte der Nutzenbewertung

Zu 3.2.1 Auswirkung nicht publizierter Studienergebnisse auf das Fazit

Änderungsvorschlag:

Die folgende Tabelle 4 beschreibt, in welchen Konstellationen ein Risiko für eine Verzerrung von Bewertungsergebnissen besteht und welche Konsequenzen sich daraus für das Fazit der Nutzenbewertung ergeben. Die Szenarien 1 bis 3 beziehen sich auf Nutzenbewertungen mit Datenübermittlung durch Dritte, Szenarien 4 bis 6 auf solche ohne zunächst vorgesehene Datenübermittlung.

Tabelle 4: Szenarien der Datenvollständigkeit und Konsequenzen für das Fazit einer Nutzenbewertung

Szenario	Datenübermittlung durch Dritte (z. B. Daten von Herstellerinnen / Herstellern)	Nachweis, dass insgesamt Daten fehlen	Verzerrung durch Publikationsbias	Bewertung / Einfluss auf das Fazit
1	vollständig	nein oder in vernachlässigbarem Umfang	unwahrscheinlich	keine Einschränkung des Fazits
2	vollständig	Ja	möglich	Einschränkung der Aussagesicherheit
3	unvollständig	ja oder unklar	wahrscheinlich	kein Beleg (und kein Hinweis oder Anhaltspunkt) für einen Nutzen bzw. Schaden
4	keine Datenübermittlung vorgesehen	nein oder in vernachlässigbarem Umfang	unwahrscheinlich	keine Einschränkung des Fazits
5	keine Datenübermittlung vorgesehen	Ja	möglich	Einschränkung der Aussagesicherheit
6	keine Datenübermittlung vorgesehen	Ja oder unklar	wahrscheinlich	kein Beleg (und kein Hinweis oder Anhaltspunkt) für einen Nutzen bzw. Schaden

a: Bei Nutzenbewertungen mit eigenem Stellungnahmeverfahren kann das Fazit der Bewertung auch unter Vorbehalt gestellt werden.

b: Sofern im Ausnahmefall vom Datenübermittelnden nachgewiesen wird, dass die Menge der fehlenden Daten sicher irrelevant ist, wird das Fazit nicht eingeschränkt.

Ab Szenario 3: Sofern Unklarheit über die Vollständigkeit der Daten besteht, fordert das IQWiG den Hersteller oder die Herstellerin zur Datenübermittlung auf, kommt dieser oder diese der Aufforderung nicht nach, erfolgt die kategorisierte Einschätzung.

War in einer Nutzenbewertung mit Datenübermittlung durch Dritte diese Datenübermittlung vollständig und liegt kein Nachweis vor, dass Daten in einem mehr als vernachlässigbaren Umfang fehlen (Szenario 1), wird eine Verzerrung als unwahrscheinlich eingestuft. Die Schlussfolgerungen aus der Bewertung der Daten werden ohne Einschränkung in das Fazit übernommen. Wird bei einer vollständigen Datenübermittlung durch die Herstellerin oder den Hersteller durch andere Recherchen der Nachweis erbracht, dass Daten in einem mäßigen oder erheblichen Umfang (z. B. aus Investigator-initiierten Studien oder Studien anderer Hersteller) fehlen, auf die die Herstellerin oder der Hersteller keinen Zugriff hat, liegt zwar keine selektive Datenübermittlung durch die Herstellerin oder den Hersteller vor, ist aber eine Verzerrung durch die anderweitig fehlenden Daten möglich (Szenario 2). Eine Nutzensaussage im Fazit (im Sinne von Beleg) wird deshalb in ihrer Aussagesicherheit entsprechend abgeschwächt und als Hinweis oder Anhaltspunkt für Nutzen oder Schaden interpretiert.

Führte eine unvollständige Datenübermittlung durch Dritte zum Fehlen von Daten in einem erheblichen, mäßigen oder unklaren Umfang (Szenario 3), wird von einer selektiven Datenübermittlung ausgegangen. Weitere relevante unpublizierte Daten, die durch die Rechenschritte nicht zugänglich sind, können existieren. Eine weitere Auswertung der verfügbaren limitierten Daten und eine daraus abgeleitete Aussage zum Nutzen oder Schaden sind in dieser Situation wahrscheinlich hochgradig verzerrt und stellen damit keine valide Bewertungsgrundlage dar. In dieser Situation kann deshalb kein Beleg (und kein Hinweis oder Anhaltspunkt) für einen Nutzen oder Schaden der zu bewertenden Intervention festgestellt werden, unabhängig von den beobachteten Ergebnissen in den vorgelegten Daten.

Lässt sich in einer Nutzenbewertung ohne vorgesehene Datenübermittlung durch Dritte ein Fehlen von Daten nicht erkennen (Szenario 4), ergibt sich keine Einschränkung für das Fazit. Lassen sich jedoch fehlende Daten nachweisen (z. B. durch Abgleich zwischen registrierten und publizierten Studien), so ist eine Verzerrung durch Publikationsbias möglich (Szenario 5) oder sogar wahrscheinlich (Szenario 6). Die Abgrenzung zwischen beiden Szenarien erfolgt anhand der eingeschätzten Höhe des Anteils fehlender Daten. Bei möglichem Publikationsbias wird das Fazit in seiner Aussagesicherheit herabgestuft. Alternativ kann das Fazit unter Vorbehalt gestellt werden, wenn im Rahmen eines vom IQWiG durchgeführten Stellungnahmeverfahrens zu einer Nutzenbewertung die Möglichkeit besteht, dass fehlende Daten nachgeliefert werden. Bei wahrscheinlichem Publikationsbias wird nach erfolgloser Datenabfrage beim Hersteller oder der Herstellerin, kein Beleg (und kein Hinweis oder Anhaltspunkt) für einen Nutzen bzw. Schaden abgeleitet. In diesem Fall werden lediglich die vorhandenen und die fehlenden Informationen dargestellt.

Begründung:

Es ist in der Humanwissenschaft evident, dass nur etwa 36% der Abstract-Einreichungen mit vorläufigen Studienergebnisse zu einer Publikation führen. (Sprague S. et al. Barriers to full-text publication following presentation of abstract at annual orthopaedic meetings. *J Bone Joint Surg Am* 2003;85(1):158-163) Hinsichtlich einer behaupteten Ergebnisverzerrung der Outcomes von bisher nicht veröffentlichten Daten aus Studien, spiegelt die belastbare Datenlage solcher Data Bias, insbesondere bei den Outcomes, eine vollkommen andere Realität wider. In nur etwa 2% der Fälle von preminilary abstracts, die zur Publikation kamen, lagen Outcome Bias vor. (Guryel E. et al. From presentation to publication: the natural history of orthopaedics abstracts in the United Kingdom. *Postgrad Med J* 2006;82(963):70-72) Es liegen also hinreichende wissenschaftliche Erkenntnisse zur entsprechenden Würdigung solcher Daten vor. Zudem muss dem Hersteller oder der Herstellerin die Möglichkeit der Datenübermittlung gegeben werden, insofern der Bedarf dazu deutlich wird.

Zu 3.2.4 Patientenberichtete Endpunkte

Änderungsvorschlag:

Die im Fließtext dieses Abschnitts zitierte Referenz 251 ist zu ersetzen, denn die inhaltliche Ableitung aus der Referenz ist unzutreffend. Es liegt dementsprechend ein Zitationsfehler vor.

Begründung:

Die Referenz 251 verweist auf die Fundstelle der FDA "*Developing Medical Imaging Drug and Biological Products, Part 2: Clinical Indications (I)*". Die zutreffende Referenz für das Interpretieren von geeigneten Studien und deren Ergebnisse ist aber ausdrücklich die Fundstelle der FDA "*Developing Medical Imaging Drug and Biological Products, Part 3: Design, Analysis, and Interpretation of Clinical Studies (I)*". Bei allen Fundstellen dieser Art muss erwähnt werden, dass es keine Guidelines sondern Guidances sind. Bei allen diesen Guidances dieser Art vermerkt die veröffentlichende FDA zudem "Contains Nonbinding Recommendations". Die inhaltliche Gewichtung der wissenschaftstheoretischen Ableitung des Abschnittes ist dementsprechend zu übertragen.

Zu 3.4 Nicht medikamentöse therapeutische Interventionen

Vorbemerkung:

Die aufgezählten Felder der Anwendung der Methodik bedürfen einer gesonderten Betrachtung, wie beispielsweise das IDEAL –Statement für chirurgische Verfahren (Literaturhinweis, <https://doi.org/10.1016/j.zefq.2015.05.013>) es bereits aufgezeigt hat. Es muss gleiche Bedingungen für alle Interventionen geben, d. h. alle Felder der Medizin müssen einen Nutzen nachweisen, aber die Anforderung an evidenzbasierte Medizin kann sich nicht ausschließlich auf RCTs stützen, so dass für spezielle Felder (Chirurgie, etc.) auch entsprechend angepasste Evidenzgenerierung akzeptiert werden sollte.

Die pauschale Benennung der Nutzensnachweise für E-Health-Anwendungen ist nicht ausreichend, da bereits durch das Digitale-Versorgung-Gesetz (Drucksache 19/14867) für die dort genannten digitalen Gesundheitsleistungen für Medizinprodukte mit niedrigen Risikoklassen neben den Nutzensnachweisen auch positive Versorgungseffekte genannt werden, die in dem Entwurfspapier bisher nicht berücksichtigt werden. Zudem fehlen für die Bewertung von weiteren E-Health-Anwendungen derzeit der gesetzliche Auftrag bzw. die Anforderungen in der Verfahrensordnung durch den G-BA. Die besonderen Herausforderungen für die Konzeption und Durchführung von klinischen Studien bei Untersuchungs- und Behandlungsmethoden mit Medizinprodukten sind erstmals die Herausforderungen und Schwierigkeiten in der Entwurfsversion erwähnt worden, jedoch ist nicht erkennbar, inwiefern allein diese Benennung dazu beiträgt, eine sachgerechte Interpretation der genannten Schwierigkeiten zu gewährleisten. Es ist derzeit nicht erkenntlich, wie die genannten einzelnen Punkte (Lernkurve, Schrittinnovationen und Verblindungsmöglichkeiten/ subjektive Endpunkte) in welcher Form konkret berücksichtigt werden. Hier bedarf es weiterer Spezifikationen unter Berücksichtigung internationaler Erfahrungen und Anwendungen.

Zu 9.3.3 Beurteilung klinischer Relevanz

Punkte 1, 2 und 3: Spezifikation und Ausführungen zum Responsekriterium 15%

Änderungsvorschlag:

Wir halten eine starre Festlegung des Responsekriteriums auf 15 % zwar für pragmatisch aber nicht sinnvoll. Wir schlagen stattdessen die Verwendung einer angemessenen und begründeten, für die Therapie und den jeweiligen Outcome Parametern, definierten MID vor.

Begründung:

Wie das IQWiG selbst schreibt: "Eine Sichtung von systematischen Übersichtsarbeiten zu MIDs [z. B. [523,655]] hat ergeben, dass sich MIDs in vielen Fällen zwischen 10 % und 20 % der Spannweite der Scores eines Erhebungsinstruments bewegen."

Mit einer starren Festlegung des Responsekriteriums auf 15 % wird wissentlich in Kauf genommen, dass damit falsch zu hoch bzw. falsch zu niedrig kalkulierte Responderanalysen erfolgen, die die Evidenz für den Nutzen einer medizinischen Intervention einschränken können. In einem Worst Case Szenario wird damit eine wirkungsvolle Therapie aufgrund eines fehlerhaften Responsekriteriums als nicht besser als bestehende Therapien eingestuft und damit den Patienten ggf. eine wirksame Therapieoption vorenthalten.

Zu 9.3.12 Umgang unvollständigen Daten

Textabschnitt aus dem IQWiG Methodenentwurf 6.0:

„Ergebnisse fließen daher in der Regel nicht in die Nutzenbewertung ein, wenn diese auf weniger als 70% der in die Auswertung einzuschließenden Patientinnen und Patienten basieren, das heißt, wenn der Anteil der Patientinnen und Patienten, die nicht in der Auswertung berücksichtigt werden, größer als 30% ist.“

Änderungsvorschlag:

Wir schlagen die folgende Ergänzung vor: "Kann nachgewiesen werden, dass die jeweiligen Ergebnisse nicht empfindlich gegenüber fehlenden Daten sind, fließen die Ergebnisse in die Nutzenbewertung ein."

Begründung:

Wir halten eine generelle Festlegung, dass Ergebnisse in der Regel nicht in die Nutzenbewertung einfließen, wenn diese auf weniger als 70 % der in die Auswertung einzuschließenden Patientinnen und Patienten basieren, für nicht sinnvoll. Dies gilt insbesondere für Patienten mit terminalen Erkrankungen. Als Beispiel: eine Studie zur Lebensqualität bei Krebspatienten im terminalen Stadium kann trotz einer sehr hohen Todesrate, dennoch wertvolle Ergebnisse liefern. Eine Standard Kaplan-Meier Analyse zur Mortalität, zum Beispiel nach 24 Monaten, ist auch valide mit deutlich mehr als 30% der Patienten, zensiert nach 24 Monaten.

Abschließende Anmerkung:

In der Prüfung nicht-Medikamentöser Methoden und Interventionen unter der Anwendung fortschrittlicher Medizinprodukte, muss die Bewertung von Nutzen und Schaden den teils erschwerten Umständen einer klinischen Prüfung Rechnung getragen werden. Bei allen durchgeführten klinischen Prüfungen nicht-Medikamentöser Methoden und Interventionen unter der Anwendung fortschrittlicher Medizinprodukte im regulatorischen Umfang oder vor einer Nutzenbewertung im Zuge der Verfahren nach §§135, 137c, 137e und 137h SGB V, wurden bereits Ethikprüfungen nach humanwissenschaftlichen Standards durchgeführt. Gleiches gilt für die Durchführung von prospektiv durchgeführten Beobachtungsstudien mit Kontrollgruppen vor Verfahren der o.g. Rechtsgrundlagen, woraus in der Berücksichtigung dessen das Methodenpapier keine höhergestellten Anforderungen ex Post stellen kann. Das Methodenpapier kann lediglich Maßgabe für die Ableitung eines Nutzens oder Schadens sein, die sich alleine auf die Inhalte der Studien innerhalb der vorliegenden Designs stützt, und sollte keinen Standard als Voraussetzung für eine patientenrelevante Gewichtung der Endpunkte anlegen. Die Würdigung humanwissenschaftlicher Studienergebnisse bei nicht-Medikamentösen Methoden und Interventionen unter der Anwendung fortschrittlicher Medizinprodukte, muss daher an der verfügbaren Evidenz, und nicht an einer ex Ante behauptet vorliegenden Evidenz mit bestimmten Designfaktoren, beurteilt werden. Wissenschaftstheoretische Ansätze folgen oft dem Umstand des Möglichen und nicht der Maßgabe des Vermeintlichen, alles andere käme einer unangemessenen wissenschaftlichen Zäsur gleich und wäre als solche zu bewerten.

Mit freundlichen Grüßen

BVMed – Bundesverband
Medizintechnologie e. V.

Dr. Marc-Pierre Möll
Geschäftsführer

Olaf Winkler
Leiter Referat Industrieller Gesundheitsmarkt

A.1.12 – Deutsche Gesellschaft für Gesundheitsökonomie e. V. (dggö)

Autorinnen und Autoren

- Mühlbacher, Axel C.
- Vorstand dggö

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Vorstand DGGÖ
Prof. Dr. Axel C. Mühlbacher
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
☒ im Namen folgender Institution / Organisation: Deutsche Gesellschaft für Gesundheitsökonomie (DGGÖ)
☐ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.

Die Stellungnahme bezieht sich, wie gewünscht, nur auf Änderungen im Methodenpapier 6.0. Diese Praxis sollte kritisch hinterfragt werden, da bei einem solchen Prozess nicht von einer methodischen Weiterentwicklung des Methodenpapiers ausgegangen werden kann. Vielmehr wird der Status Quo kleinteilig fortgeschrieben. Vor diesem Hintergrund ist es der Deutschen Gesellschaft für Gesundheitsökonomie (dggö) ein Anliegen, die folgenden allgemeinen Anmerkungen zu machen:

Gesundheitsökonomische Methoden sollten nicht nur für die Kostenbewertung, sondern auch für die Nutzenbewertung herangezogen werden. Gerade bei der Nutzenbewertung bietet die Ökonomie internationale Standards und wissenschaftliche Reflektion, die im Methodenpapier nur unzureichend berücksichtigt werden. Die gesundheitsökonomische Literatur, entgegen der Literatur zur evidenzbasierten Medizin, diskutiert ausgiebig die Operationalisierung des Nutzens. Neben der Messung der klinischen Effekte, bedarf es der wissenschaftlich transparenten Bewertung der gemessenen erwünschten und unerwünschten Effekte einer Therapie oder Technologie. Diese Abwägung von Nutzen und Schaden erfolgt auf Basis von Werturteilen der Entscheidungsträger. Das Methodenpapier zeigt nur unzureichend auf, wie die Präferenzen der betroffenen Patienten bei Entscheidungen berücksichtigt werden.

Die folgende Tabelle zeigt, dass auch bei der Überarbeitung des Methodenpapiers 6.0 die implizite normative Setzung von Werturteilen vorgezogen wird. Die explizite und transparente Berücksichtigung der Patientenpräferenzen auf Basis wissenschaftlicher Studien ist nicht vorgesehen. Die dggö plädiert, wie medizinische Fachgesellschaften zuvor, für die systematische und transparente Einbeziehung der Patientenperspektive in die Dokumentation des Patientennutzens.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung
	<i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
3.1.3 Ermittlung des Schadens medizinischer Interventionen S. 47-50	<u>Anmerkung:</u> Das IQWiG stellt auf S. 47 fest: „Die Darstellung des Schadens ist ein wesentlicher und gleichberechtigter Bestandteil der Nutzenbewertung einer Intervention. Sie gewährleistet eine informierte populationsbezogene, aber auch individuelle Nutzen-Schaden-Abwägung.“ Unklar bleibt, wie populationsbezogene und individuelle Abwägungen erfolgen. Auch die nachfolgenden Ausführungen von

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	S. 50 des Methodenpapiers treffen keinerlei inhaltliche Aussagen hinsichtlich der verwendeten Methode, sodass die Operationalisierung der Abwägung multipler Schadensdimensionen offenbleibt. „Für die ausgewählten spezifischen unerwünschten Ereignisse wird jeweils die am besten geeignete Operationalisierung in Bezug auf Mess- und Ergebnissicherheit herangezogen. Auch wird geprüft, ob die ausgewählten Ereignisse zum selben inhaltlichen Konstrukt konsistent sind.“ <u>Vorgeschlagene Änderung:</u> Für die Weiterentwicklung der Methoden sollte sich das IQWiG dringend die Berücksichtigung von Patientenpräferenzen vornehmen. Nur so kann gewährleistet werden, dass alle Ergebnisse medizinischer Interventionen miteinander vergleichbar sind und die Patientenpräferenzen respektiert werden. Die Methoden des IQWiG fallen hier weit hinter internationale Standards zurück.
3.3.3 Nutzenbewertung von Arzneimitteln gemäß § 35a SGBV S. 63-68	<u>Anmerkung:</u> Abschnitt 3.3.3 ergänzt die Methodik zur Quantifizierung der Effektmaße um die Vorgehensweise bei stetigen oder quasi-stetigen Zielgrößen. Hierzu werden standardisierte Mittelwertdifferenzen (SMDs) herangezogen. Schwellenwerte zur Ausmaßbestimmung werden entsprechend der Zielgrößen bestimmt. Hierbei handelt es sich um eine normative Setzung. Werturteile werden in Abhängigkeit der Zielgrößenkategorie fixiert, ohne das Ausmaß der gewünschten Effektstärke aus der Patientenperspektive zu beleuchten. Der Patientennutzen orientiert sich damit nicht notwendigerweise an den Bedürfnissen und Bedarfen der Patienten. Auf S. 67 heißt es: „Man kann zeigen, dass die für ein bestimmtes Ausmaß wünschenswerte Effektstärke ungefähr halbiert werden muss, um den entsprechenden Schwellenwert abzuleiten. Als Orientierung für die theoretisch wünschenswerten Effektstärken bei schwerwiegenden Symptomen können die Mittelwerte der Bereiche für die übliche Einteilung von Cohen's d in kleine (SMD zwischen 0,2 und 0,5) mittlere (SMD zwischen 0,5 und 0,8) und große Effekte (SMD > 0,8) dienen. Tabelle 6 gibt einen Überblick über die verwendeten Schwellenwerte. Im Gegensatz zu den Schwellenwerten für die relativen Maße ist bei der SMD zu beachten, dass das Konfidenzintervall vollständig oberhalb des entsprechenden Schwellenwerts liegen muss und der Nulleffekt hier nicht durch die Zahl 1 sondern durch die Zahl 0 gegeben ist.“ Schwellenwerte können als normative Setzung vorgegeben werden und beeinflussen damit in erheblichem Maße die zukünftige Methodik zur Dokumentation des Gesamtnutzens. Aus der Perspektive der gesundheitsökonomischen Forschung sind Schwellenwerte abhängig vom (empfundenen) Nutzenbeitrag aus

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	der Perspektive der Patienten. Eine Aufteilung der Schwellenwerte entsprechend der Ausmaßkategorien ist eine implizite normative Entscheidung ohne die Berücksichtigung der Patientenpräferenzen. Zudem wird hier die Trennung von „Assessment“ und „Appraisal“ seitens des IQWiG und GBA nicht berücksichtigt. Der GBA ist institutionell vorgesehen, die Werturteile von Kostenträgern, Leistungserbringern und Patienten bei der Nutzenbewertung zu berücksichtigen. Inwiefern das IQWiG zur normativen Setzung legitimiert ist, ist offen. Siehe dazu auch die Stellungnahme der dggö zum Methodenpapier 4.2 aus dem Jahr 2014 (http://file.dggoe.de/2014-07-29_dggö_Stellungnahme_IQWiG_Allgemeine_Methoden_4_2.pdf) <u>Vorgeschlagene Änderung:</u> Im Methodenpapier wird angemerkt, dass für die „Operationalisierung der Gesamtaussage zum Ausmaß des Zusatznutzens bei gemeinsamer Betrachtung aller Endpunkte (ist) eine strenge Formalisierung nicht möglich (ist), da für die hierauf zutreffenden Werturteile gegenwärtig keine ausreichende Abstraktion bekannt ist“. Dieser Aussage kann insofern zugestimmt werden, da es hierzu keine einheitlichen Standards gibt bzw. bestehende Standards/ Richtlinien kritisch diskutiert werden. Dies gilt aber ebenso für die Konzeption klinischer Studien sowie der Berücksichtigung von Evidenz. Notwendig aus Sicht der dggö ist, dass Methoden zur Operationalisierung des Zusatznutzens Gegenstand der wissenschaftlichen Diskussion des IQWiG sind.
3.8 Potenzialbewertung S. 77-79	Auf S. 77 heißt es: „Potenzialbewertungen gemäß § 137e SGB V zielen im Gegensatz zu Nutzenbewertungen darauf ab zu prüfen, ob neue Untersuchungs- oder Behandlungsmethoden möglicherweise einen Nutzen aufweisen. Potenzial bedeutet hierbei, dass erstens die bisher vorliegenden Erkenntnisse einen möglichen Nutzen erkennen lassen (...)“. Trotz erheblicher Unsicherheit, muss ein zusätzlicher Patientennutzen erwartet werden. Insbesondere vor dem Hintergrund der Unsicherheit ist es nicht nachvollziehbar, dass die Präferenzen der Patienten nicht adäquat berücksichtigt werden. <u>Vorgeschlagene Änderung:</u> Das Potenzial einer neuen Untersuchungs- oder Behandlungsmethode muss immer auch vor dem Hintergrund der Patientenpräferenzen beleuchtet werden. Eine explizite und transparente Dokumentation der Patientenpräferenzen ist maßgeblich für die nachvollziehbare Analyse des Potenzials.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

**A.1.13 – Deutsche Gesellschaft für Hämatologie und Medizinische Onkologie e. V.
(DGHO)**

Autorinnen und Autoren

- Wörmann, Bernhard

DGHO e.V. • Alexanderplatz 1 • 10178 Berlin

**Institut für Qualität und Wirtschaftlichkeit
im Gesundheitswesen (IQWiG)**
Im Mediapark 8 (KölnTurm)
50670 Köln

Prof. Dr. med. Lorenz Trümper
Geschäftsführender Vorsitzender

Prof. Dr. med. Hermann Einsele
Vorsitzender

Prof. Dr. med. Maike de Wit
Mitglied im Vorstand

PD Dr. med. Ingo Tamm
Mitglied im Vorstand

Alexanderplatz 1 • 10178 Berlin
Tel.: 030 27876089- 0
Fax: 030 27876089-18
info@dgho.de

31. Januar 2020

Stellungnahme zu IQWiG Allgemeine Methoden Entwurf Version 6.0

Zusammenfassung

Hämatologie und Medizinische Onkologie haben in den letzten Jahren vor allem bei der frühen Nutzenbewertung neuer Arzneimittel, aber auch bei anderen Themen wie Diagnostik oder Methoden Kontakt mit dem IQWiG gehabt. Leider fallen bei dieser Aktualisierung des Methodenpapiers vor allem die kritischen Themen auf, die nicht adressiert wurden:

- Niedriger Anteil von Betroffenen (Patienten), die in die Abfassung der Bericht einbezogen sind
- Hohe Rate von Addenda, die im Nachgang zu den Berichten vom G-BA in Auftrag gegeben werden müssen. Diese Rate ist auch ein Maß für die Qualität der IQWiG Berichte.
- Fehlende Weiterentwicklung der Methodik in Richtung einer stärkeren Berücksichtigung von Parametern der Morbidität bei chronisch verlaufenden Erkrankungen
- Defizit bei der Erfassung dramatischer Effekte bei Erhöhung der Langzeitüberlebensrate.

Einleitung

Das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) legt nach der Veröffentlichung seines Methodenpapiers 5.0 vom Juli 2017 jetzt den Entwurf der nächsten, eigentlich jährlich geplanten Aktualisierung vor. Es wurde eine Reihe inhaltlicher Änderungen bei unterschiedlichen Themen vorgenommen.

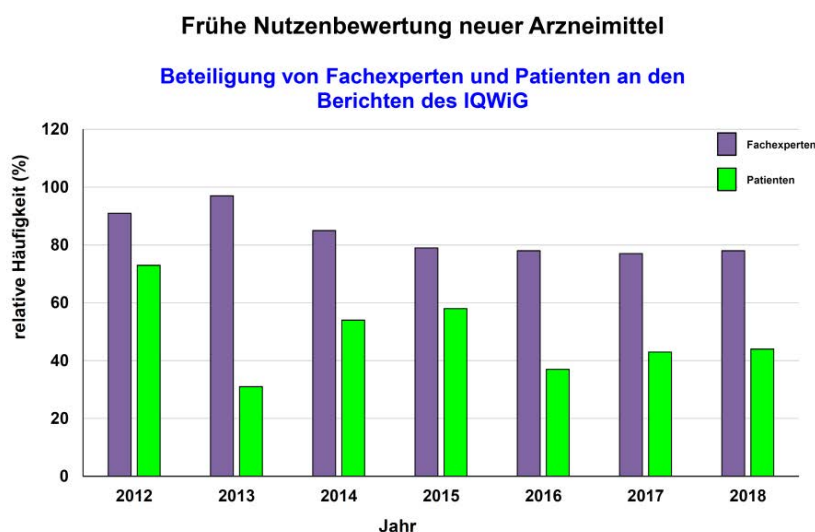
Hämatologie und Medizinische Onkologie haben in den letzten Jahren vor allem bei der frühen Nutzenbewertung neuer Arzneimittel aber auch bei anderen Themen wie Diagnostik oder Methoden Kontakt mit dem IQWiG gehabt. Dabei haben sich wiederholt Kritikpunkte an Methodik und Vorgehen des Instituts ergeben.

Es ist bedauerlich, dass die viele der von uns als relevant angesehenen Probleme in dieser Überarbeitung der Allgemeinen Methoden nicht oder nicht angemessen adressiert werden.

Beteiligung von Betroffenen (Kapitel 3. 1. 3. Darstellung von Schadensaspekten)

Dem IQWiG ist die Relevanz der Beteiligung von Betroffenen bewusst. In der letzten Aktualisierung der Methodik war die Änderung der Einbindung von „tatsächlich Betroffenen“ in das Methodenpapier aufgenommen worden. Im Kapitel 3. 1. 3. wird jetzt nachvollziehbar formuliert: *„Zum einen werden solche spezifischen unerwünschten Wirkungen ausgewählt, die für das Krankheitsbild oder die in der Studie bzw. den Studien eingesetzten Interventionen von besonderer Bedeutung sind. Für Nutzenbewertungen gemäß § 139a SGB V ist dies der primäre Weg der Identifizierung solcher unerwünschten Wirkungen. Die Zusammenstellung erfolgt in diesem Fall im Rahmen der Vorabrecherche zur jeweiligen Fragestellung und durch die **Einbindung von Betroffenen**.“*

Das ist inhaltlich korrekt, aber die Umsetzung erscheint schwierig, siehe Abbildung 1.



Der Anteil von IQWiG-Berichten im Rahmen der frühen Nutzenbewertung neuer Arzneimittel, die ohne Patienten erstellt werden, liegt durchgehend unter 50%.

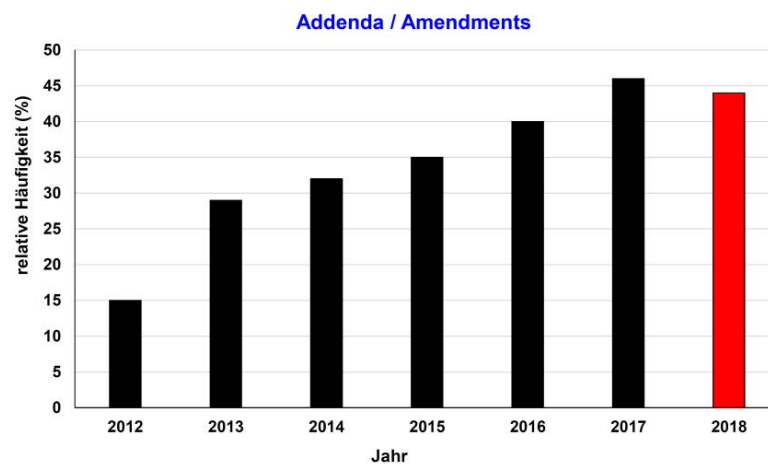
Fazit: Das IQWiG ist derzeit nicht in der Verfassung, die neuen Ansprüche zu 3. 1. 3. umzusetzen.

Qualität der Berichte

Der G-BA kann zusätzliche Aufträge an das IQWiG zur Erstellung eines Addendums erteilen oder diese selbst als Amendment erstellen. Grundsätzlich begrüßen wir die Beauftragung von Addenda als Reaktion auf Stellungnahmen und auf die Diskussionen im Rahmen der Anhörungen. Problematisch am derzeitigen Verfahren ist, dass die wissenschaftlichen Fachgesellschaften und andere Stellungnehmer keine Gelegenheit zur Diskussion der Addenda haben. Sie gehen nur dem G-BA zu und werden erst nach Abschluss des Verfahrens publiziert. Die Methodik, auch innerhalb des Methodenpapiers 6.0, wird der Bedeutung von Addenda nicht gerecht. Es fehlen z. B. die obligate Beteiligung von Experten, die frühzeitige Publikation des Berichtes (nicht erst nach der Festlegung des G-BA) und Festlegungen zur internen Qualitätssicherung.

Der Anteil von Addenda/Amendments an der Gesamtzahl der Verfahren ist in Abbildung 2 dargestellt.

Frühe Nutzenbewertung neuer Arzneimittel



Der Anteil von Verfahren mit Addenda/Amendments liegt bei 40 - 50%.

Fazit: Der Anteil von erforderlichen Addenda ist auch ein Maß für die Qualität der IQWiG Berichte. Diese ist in den letzten Jahren deutlich gesunken.

Erfassung von Morbidität

Die aktuelle Methodik der frühen Nutzenbewertung kommt vor allem Krankheitsindikationen mit kurzer Lebenserwartung und Arzneimitteln mit Orphan-Drug-Status entgegen. Benachteiligt sind chronische Erkrankungen mit langem Therapie- und Krankheitsverlauf. Die starken Unterschiede zwischen den Fachgebieten sind in Abbildung 3 dargestellt.



Diese Ungleichgewichtigkeit wirkt sich langfristig negativ auf die Entwicklung neuer Arzneimittel aus.

Fazit: Erforderlich ist eine Weiterentwicklung der Methodik in Richtung einer stärkeren Berücksichtigung von Parametern der Morbidität bei chronisch verlaufenden Erkrankungen.

Dramatische Effekte

Im Methodenpapier ist festgelegt, dass „Aussagen zum Nutzen auf Basis von Studien mit niedrigerer Evidenzstufe (retrospektive vergleichende Kohortenstudien, historisch vergleichende Studien, nicht vergleichende Studien) nur in Verbindung mit einem dramatischen Effekt möglich sind.“ Hierzu wird auf eine Publikation aus dem Jahr 2007 referenziert. Darin wurde nach Auswertung historischer Beispiele aus der ganzen Breite der Medizin als Maß für einen dramatischen Effekt in nicht-vergleichenden Studien ein „Signal to Noise“-Verhältnis von ≥ 10 vorgeschlagen.

Eine solche pauschale Bewertung ist nicht zeitgemäß. Ein wesentliches Ziel der modernen Onkologie ist nicht mehr allein die Verlängerung der Gesamtüberlebenszeit als Median, sondern die Erhöhung der Überlebensrate nach einem definierten Zeitpunkt, z. B. nach 2, 3 oder 5 Jahren. Praktisch bedeutet die die Anwendung des IQWiG-Ansatzes, dass bei einer 3-Jahres-Überlebensrate von 10% in der Vergleichsgruppe die Interventionsgruppe auf eine Überlebensrate von mindestens 100% kommen müsste. Aus klinischer Sicht können eine Verdoppelung oder Verdreifachung der Überlebensrate ein relevanter Effekt sein, und für betroffene Patienten ganz sicher „dramatisch“.

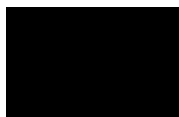
Fazit: Die aktuelle Methodik zur Erfassung dramatischer Effekte ist im Hinblick auf das Ziel einer Erhöhung der Langzeitüberlebensrate nicht angemessen.

Für Rückfragen stehen wir jederzeit gern zur Verfügung.

Mit freundlichen Grüßen



Prof. Dr. med. Lorenz Trümper
Geschäftsführender Vorsitzender



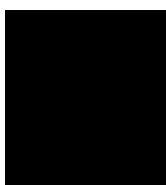
Prof. Dr. med. Hermann Einsele
Vorsitzender



Prof. Dr. med. Maike de Wit
Mitglied im Vorstand



PD Dr. med. Ingo Tamm
Mitglied im Vorstand

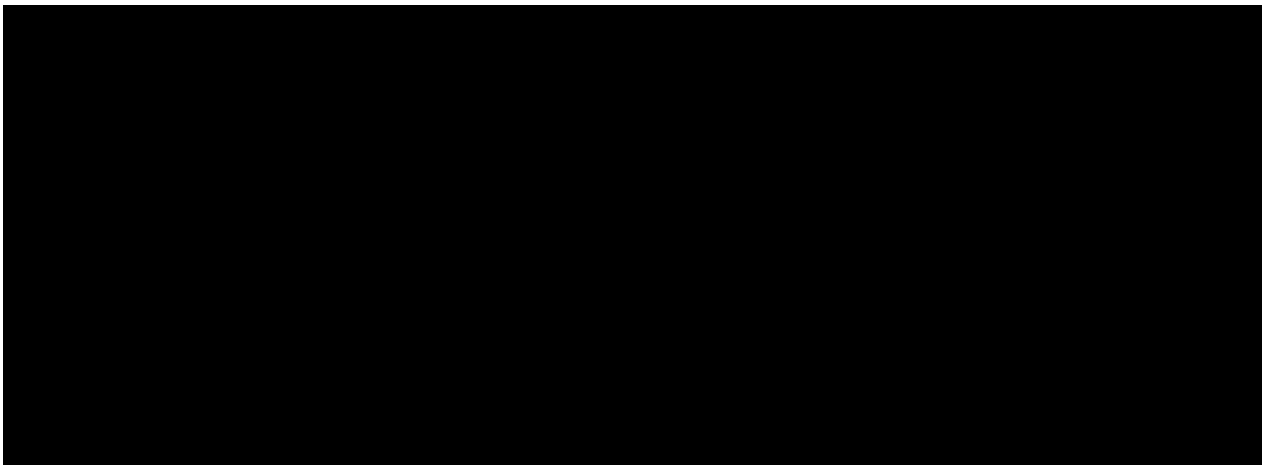


Prof. Dr. med. Bernhard Wörmann
Medizinischer Leiter der DGHO

A.1.14 – Deutsche Gesellschaft für Innere Medizin e. V. (DGIM)

Autorinnen und Autoren

- Fölsch, Ulrich
- Sauerbruch, Tilman



Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Prof. Dr. med. Tilman Sauerbruch
Prof. Dr. med. Ulrich Fölsch
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
☒ im Namen folgender Institution / Organisation: Deutsche Gesellschaft für Innere Medizin e.V.
☐ als Privatperson(en)

Kommentare zum IQWiG – Methoden-Report

Version 6.0 vom 05.1.2019

Seite 2: Die DGIM begrüßt den expliziten Hinweis auf die Einbeziehung externer Sachverständiger. Hier bringt sich die DGIM jederzeit gerne ein.

Seite 5: Wir begrüßen, dass Sie darauf hinweisen, dass eine Evidenz im strengen Sinne gelegentlich lückenhaft ist. Hier ist die Unsicherheit des Instituts möglicherweise größer als die Unsicherheit von Ärztinnen und Ärzten, die die Patienten direkt behandeln und einschätzen können.

Seite 6: 1) Es gibt in der Medizin durchaus (nicht selten) Situationen, in denen nur eine korrekte Therapie gewählt werden kann/muss.

Seite 9: 1) Wir sind uns nicht sicher, ob Ihr Kausalitätsbegriff hier wirklich korrekt ist. Im Grunde werden auch in randomisierten Studien nur eine mehr oder weniger enge Assoziationen zu einem Krankheitsverlauf beschrieben (bei manchen Patienten kann eine Intervention kausalbeeinflussen, bei anderen Patienten dann wieder nicht)

Seite 10: Vorletzter Absatz: Man braucht Ergebnissicherheit und Alltagsnähe. Die Alltagsnähe kann von dem Institut nur schwer abgeschätzt werden. Hier muss es auf direkt betroffene Akteure zurückgreifen.

Seite 11: Wir begrüßen die Betrachtung der Nutzenbewertung im Einzelfall. Hier wäre es durchaus sinnvoll, wenn sich das Institut mit seinen vielen Möglichkeiten einmal systematisch Kasuistiken vornehmen würde, bzw. entsprechende Netzwerke analysiert.

Seite 14, vorletzter Absatz: Wir verstehen nicht genau, warum wissenschaftliche Standards und Methoden durch Allokationsentscheidungen bestimmt werden da?

Seite 16, zweiter Absatz: Wir halten es für ein Problem, dass für die Potenzialbewertung nach § 137e eine Anhörung durch das Institut nicht vorgesehen ist (darauf haben wir schon einmal hingewiesen).

Seite 23, letzter Absatz: Hier empfehlen wir auch bei diesen wissenschaftlichen Fachgesellschaften Hinweise zur externen Sachverständigung einzuholen neu.

Seite 31, Abb. 7: Was heißt externes Review bei Begleitinformationen nicht einzuholen, wie genau sind Begleitinformationen definiert?

Seite 32, Abb. 8: Bei der Themenauswahl empfehlen wir wissenschaftliche Fachgesellschaften einzubeziehen (hier findet schon ein Dialog zwischen DGIM und IQWiG statt).

Seite 36, mittlerer Absatz (Auswahl von Betroffenen): Wie valide ist diese?? Ist das überprüft?

Seite 43: Hier wird davon ausgegangen, dass die zweckmäßige Vergleichstherapie immer von Nutzen ist. Dies muss aber nicht der Fall sein.

Seite 45: Hierarchisierung der Endpunkte: Sind hier Mortalität, Morbidität und Lebensqualität gemeint? Nach welchen Kriterien soll das geschehen?

Seite 46: Definition der Surrogat-Endpunkte: Nach der Erfahrung der DGIM sollte hier besondere Sorgfalt hinsichtlich der Rücksprache mit wissenschaftlichen Fachgesellschaften erfolgen.

Seite 53: Mitte, mindestens 1.000 Patienten: Diese Zahl wird zwar relativiert. Es stellt sich dennoch die Frage, wie das Institut zu dieser Zahl gekommen ist. Solche Zahlen hängen doch sehr von der Hypothese einer Studie, den bekannten/den mutmaßlichen Effekten einer jeweiligen Maßnahme und der Fallzahlschätzung ab.

Seite 60: extrem seltene Erkrankungen. Welche Erfahrungen hat das Institut mit der Zusammenstellung von Kasuistiken aus der vorhandenen Literatur für diese Erkrankungen? Siehe auch Kommentar oben!

Seite 61: Zulassungsstatus: Wie steht das Institut zu Fragen des Drug Repurposing?

Seite 62 oben: Warum diese Einschränkung bei Altersgruppen außerhalb der Zulassung? Hier könnten durchaus neue Erkenntnisse/Hypothesen generiert werden.

Seite 76: Der Begriff „Überdiagnosen“ ist zwar definiert, erscheint uns aber nicht ganz passend. Eine Diagnose ist die Feststellung einer Erkrankung. Unabhängig von der Auswirkung der Erkrankung. Diese Auswirkung kann auch bei jedem Patienten anders sein. Das wird im Prinzip auch in der Ref. 120 diskutiert, auf die sich das Institut bezieht. Ev. kann man auf Dauer hier einen geeigneteren Terminus finden.

„Da über Diagnosen als schädliche Wirkung jedes Screenings unvermeidlich sind ...“
Was heißt das? Diese Aussage impliziert von vornherein, dass die Relevanz einiger Diagnosen durch das behandelnde Personal falsch eingeschätzt wird.

Seite 80 ff: Die verschiedenen Modelle zur Kosten-Nutzen-Bewertung sind uns nur teilweise verständlich. Wie häufig wird diese Fragestellung im IQWiG bearbeitet?

Seite 92, QUALYS: Der Wert der QUALYS wird infrage gestellt. Hierbei sollte allerdings beachtet werden, dass QUALYS in der Literatur gemeinhin verwendet werden. Also international ein benutzter Standard sind.

Seite 103: Datengrundlage: Hat sich das Institut schon einmal damit beschäftigt, welche Qualität die Daten des Robert-Koch-Instituts bzw. die GKV Routinedaten wirklich haben? Wie kommen sie zustande?

Seite 117 ff. (Leitliniensynopsen): Hier sollte geprüft werden, inwieweit zuständige Mitarbeiter des IQWiG bei der Erstellung von Leitlinien einbezogen werden können.

Seite 123: letzter Spiegelstrich (potenzieller Forschungsbedarf):

Hier ist eine Rücksprache mit den wissenschaftlichen Fachgesellschaften unerlässlich.

Seite 125, Tab. 9: Siehe auch Bemerkung oben, Wert der Routinedaten aus dem Pool der GKV oder auch der Kassenärztlichen Vereinigung?

Seite 126: Themenvorschläge: Warum können nur Einzelpersonen entsprechende Vorschläge einreichen? Oder was heißt Versicherte und interessierte Einzelpersonen?

Seite 127, letzter Absatz: Die Relevanz des Auswahlbeirates ist uns nicht klar bzw. die Frage, wie er zustande kommt.

Seite 129, vorletzter Absatz: „Sokratischer Ansatz“. Kurze Erläuterung?

Seite 132: Spiegelstrich Förderung der Unterstützung der Patientinnen und Patienten zum Beispiel durch Angehörige ... Wie ist das gemeint? Informationen von Angehörigen?

Seite 133, Themenkatalog: Hier stellt sich die Frage, inwieweit das Kodierungsverhalten die Daten dieses Reports beeinflusst, bzw. ob dieses immer korrekt ist.

Seite 146, Erfahrungsberichte: Gibt es Analysen, wie sich solche Erfahrungsberichte zur Einschätzung der Behandler verhalten? Solche vergleichende Analysen wären wichtig.

Seite 162, Tatsächliche Wirkung von Therapien und Interventionen:

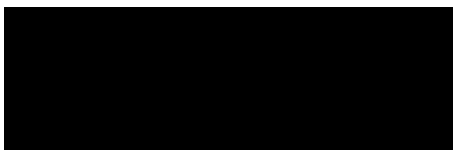
Diese lassen sich eigentlich nur heruntergebrochen auf den einzelnen Patienten abschätzen.

Seite 176, klinische Relevanz: Hier begibt sich das Institut auf einen Bereich, der vor allem die Einschätzung der direkt behandelnden Ärzte betrifft. Eine entsprechende Rücksprache ist unerlässlich.

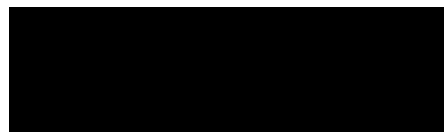
Seite 203, Tab. 13: Lebensqualität ist die gesundheitsbezogene Lebensqualität von der Gesamt-Lebensqualität zu trennen?

Allgemeine Bemerkung: Das Institut sollte sich zukünftig auch methodisch mit der Frage beschäftigen, wie Daten aus sozialen Medien(Social Media), Apps zu bewerten sind.

Wiesbaden, 17. Januar 2020



Prof. Dr. Dr. h.c. Ulrich R. Fölsch
Generalsekretär a. D. der DGIM e.V.

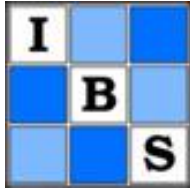


Prof. Dr. med. Tilman Sauerbruch
Beauftragter der DGIM

A.1.15 – Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e. V. (GMDS) / Internationale Biometrische Gesellschaft – Deutsche Region (IBS-DR) – (gemeinsame Stellungnahme)

Autorinnen und Autoren

- Friede, Tim



**Stellungnahme zum Entwurf der Version 6.0 des IQWiG-Papiers
„Allgemeine Methoden“
durch die gemeinsame Präsidiumskommission
„Methodenaspekte in der Arbeit des IQWiG und IQTiG“ der GMDs und IBS-DR**

Im „Allgemeinen Methodenpapier“ werden die gesetzlichen und wissenschaftlichen Grundlagen der Arbeit am Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) erläutert. Die derzeit gültige Fassung ist die Version 5.0 vom 10.07.2017. Am 5. Dezember 2019 hat nun das IQWiG einen Entwurf der neuen Version 6.0 vorgelegt. In der Version 6.0 wurden im Vergleich zur Version 5.0 neben redaktionellen Modifikationen auch inhaltliche Änderungen durchgeführt. Grundlegende Überarbeitungen erfolgten in den Kapiteln zur *Nutzenbewertung medizinischer Interventionen* (Kapitel 3), *Bewertung der Versorgung* (Kapitel 5), *HTA-Berichte* (Kapitel 6) und *Informationsbewertung* (Kapitel 9). Aus biometrischer Sicht sind insbesondere einige Änderungen in den Kapiteln 3 und 9 relevant. Auf diese gehen wir unten näher ein.

Die Kommission begrüßt die regelmäßige Überarbeitung der „Allgemeinen Methoden“. Die stetige Fortentwicklung relevanter Methoden macht dies unerlässlich. Wie aus den spezifischen Kommentaren unten deutlich wird, gibt es einige Abschnitte, die einer Aktualisierung bedürfen. Die Kommission ist offen für eine fachliche Diskussion mit dem IQWiG in Bezug auf geeignete Methoden und unterstützt das IQWiG in dieser Hinsicht gerne.

Die Kommission hat zu den folgenden speziellen Punkten des Papiers Kommentierungen durchgeführt:

Abschnitt 3.3.3 Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V

Der Abschnitt 3.3.3 widmet sich der Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V. Unter Punkt D) werden Schwellenwerte für die Ausmaßbestimmung für stetige oder quasistetige Zielgrößen mit jeweils vorliegenden standardisierten Mittelwertdifferenzen (SMD) behandelt. Diese sind in der Tabelle 6 zusammengefasst. Aus Sicht der Kommission ist dies eine sinnvolle Erweiterung. Allerdings ist der folgende Satz unverständlich: „Man kann zeigen, dass die für ein bestimmtes Ausmaß wünschenswerte Effektstärke ungefähr halbiert werden muss, um den entsprechenden Schwellenwert abzuleiten.“ Hier wünscht sich die Kommission eine nachvollziehbare Begründung. Dies könnte z.B. durch eine Erweiterung des Anhangs A im Hinblick auf SMD geschehen.

Abschnitt 9.3.8 Metaanalysen

Im Abschnitt 9.3.8 wird das Vorgehen bei Metaanalysen detailliert. Unter anderem wird hier festgestellt, dass bei Vorliegen einer ausreichenden Anzahl von Studien (interpretiert als wenigstens fünf Studien) die Knapp-Hartung-Methode für Metaanalysen mit zufälligen Effekten zur Anwendung kommen sollte. Bei der Knapp-Hartung-Methode werden zur Konstruktion von Konfidenzintervallen t-Quantile und ein skaliertes Varianzschätzer verwendet. Die Skalierung des Varianzschätzers kann in seltenen Fällen dazu führen, dass das Konfidenzintervall beruhend auf der Knapp-Hartung-Methode kürzer ist als die Intervalle, die z.B. aus einer Metaanalyse mit festen Effekten oder eine Metaanalyse

der DerSimonian-Laird-Methode resultieren. Hier kann man sich mit einer einfachen ad hoc Korrektur behelfen, indem der Skalierungsfaktor für den Varianzschätzer auf einen Bereich von 1 oder größer beschränkt wird. Diese ad hoc Korrektur wurde von Hartung und Knapp im Kontext mit Metaregressionen beschrieben (Referenz [421]). Die Kommission hat keine Einwände in Bezug auf die Verwendung dieser ad hoc Korrektur. Allerdings ist die Kommission der Meinung, dass einige Formulierungen in dem Absatz beginnend mit „Bei Anwendung der Knapp-Hartung-Methode für Metaanalysen mit zufälligen Effekten“ (Seite 183) potentiell irreführend sind. So ist es z.B. sachlich nicht richtig, dass Knapp und Hartung in dem zitierten Aufsatz die Anwendung dieser Korrektur für die beschriebene Situation vorgeschlagen haben. Zudem ist der Hinweis auf die „Knapp-Hartung-Methode mit Varianzkorrektur“ nicht eindeutig, da diese Methode verglichen z.B. mit der DerSimonian-Laird-Methode immer eine Varianzkorrektur beinhaltet. Dies sollte eindeutiger formuliert werden, z.B. durch die Verwendung eines Begriffs wie „ad hoc Korrektur“ mit einem eindeutigen Bezug auf die Literatur.

Unter Punkt D) im Abschnitt 9.3.8 finden sich Erläuterungen zu Metaanalysen von Studien zur diagnostischen Güte. Aus Sicht der Kommission sollte dieser Punkt dringend aktualisiert werden, da einige methodische Entwicklungen der letzten Jahre sich hier leider nicht wiederfinden und sich somit Widersprüche zu Empfehlungen z.B. der Cochrane Collaboration ergeben. Dies betrifft insbesondere die Berechnung von summarischen ROC (SROC)-Kurven, sowie die Verwendung von getrennten Modellen zur Schätzung der Sensitivität und Spezifität.

Für die Berechnung von SROC-Kurven ist eine entsprechende Datengrundlage notwendig, die Angaben zu allen in den Einzelstudien geschätzten Sensitivitäten und Spezifitäten enthält. Die unter Punkt D) genannten statistischen Verfahren erlauben zwar Metaregressionen, gehen aber davon aus, dass jeweils ein diagnostischer Schwellenwert pro Studie selektiert wird. Damit sind sie für die komplexere Situation von mehreren Schwellenwerten auf Studienebene nicht geeignet. Hinzu kommt, dass die resultierenden SROC-Kurven nur schwer interpretierbar sind (Arends et al, 2008). Hierfür gibt es neuere statistische Verfahren, die zitiert werden sollten (Steinhauser et al, 2016; Hoyer et al, 2018; Jones et al, 2019). Weiterhin wird unter Punkt D) aufgeführt, dass zur metaanalytischen Zusammenfassung getrennte Modelle für Sensitivität und Spezifität verwendet werden können. Der Hinweis auf diese Möglichkeit ist insofern irreführend, als dass er im Gegensatz zu den Empfehlungen der Cochrane Collaboration steht, welche die Verwendung von bivariaten bzw. hierarchischen Modellen nahelegt (Macaskill et al, 2010).

Abschnitt 9.3.12 Umgang mit unvollständigen Daten

Im Abschnitt 9.3.12 wird der Umgang mit unvollständigen Daten behandelt. Hierbei werden konkrete Verfahrensregeln festgelegt. Diese sehen vor, dass Studien, in denen weniger als 70% der eingeschlossenen Patienten (d.h. der Intent-to-Treat-Population) zur Auswertung beitragen oder der Anteil der in die Auswertung eingeschlossenen Patienten zwischen den Behandlungsgruppen sich um mindestens 15 Prozentpunkte unterscheiden, nicht berücksichtigt werden. Aus Sicht der Kommission fehlt eine Rationale für die festgelegten Grenzen von 70% und 15 Prozentpunkten. Zudem werden auch keine belastbaren Referenzen genannt. Die Kommission empfiehlt darüber hinaus, Formulierungen zu verwenden, die Ausnahmen zulassen. Zum Beispiel könnte sich eine Situation ergeben, in der unter Standardtherapie aufgrund von unerwünschten Ereignissen sehr viele Patienten vorzeitig die Studie beenden, während dies im experimentellen Arm nicht der Fall ist. Dann würde hier u.U. die experimentelle Therapie zu Unrecht bestraft. Diese Situationen sind in der Darstellung im Nutzendossier bezüglich ihrer Auswirkungen auf die Effektschätzer und den zugrundeliegenden Estimand zu diskutieren. Insbesondere ist aus Sicht der Kommission eine grundlegende Überarbeitung

dieses Abschnitts dringend geboten, da dieser der aktuellen Estimand-Diskussion und neueren Literatur nicht gerecht wird (siehe z.B. Referenz [699]).

Abschnitt 9.3.13 Umgang mit unterschiedlichen mittleren Beobachtungsdauern

Der Abschnitt 9.3.13 behandelt den Umgang mit unterschiedlichen mittleren Beobachtungsdauern. Die Kommission begrüßt, dass dieser Abschnitt um neuere Entwicklungen bei der Analyse von unerwünschten Ereignissen ergänzt wurde. So ist nun z.B. ein Verweis auf den kürzlich erschienenen Artikel von Unkel et al. enthalten (Referenz [699]), in dem Ergebnisse einer gemeinsamen Projektgruppe von GMDS und IBS-DR zur Auswertung von unerwünschten Ereignissen bei variierenden Nachbeobachtungszeiten und konkurrierenden Ereignissen zusammengetragen wurden. Allerdings sieht die Kommission als problematisch an, dass hier lediglich die Situation unterschiedlicher mittlerer Beobachtungszeiten zwischen Behandlungsgruppen behandelt wird. Auch im Falle unterschiedlicher individueller Beobachtungszeiten innerhalb einer Behandlungsgruppe können ungeeignete Analysemethoden zu verzerrten Schätzungen der Wahrscheinlichkeit für ein unerwünschtes Ereignis innerhalb der Behandlungsgruppe führen. Dies kann einerseits zu einer falschen Beurteilung von Sicherheitsrisiken innerhalb der Behandlungsgruppen und andererseits ebenfalls zu Verzerrungen bei der Betrachtung von Gruppenunterschieden führen. Des Weiteren sollte aus Sicht der Kommission unbedingt betont werden, dass konkurrierende Ereignisse (wie Tod oder vorzeitiges Ende der Nachbeobachtung bzgl. unerwünschter Ereignisse) in der Analyse durch adäquate Methoden berücksichtigt werden müssen. Ein globaler Verweis auf Techniken der Überlebenszeitanalyse ist nicht ausreichend. Dieser globale Hinweis ist sogar irreführend, da der häufig verwendete Kaplan-Meier-Schätzer (d.h. 1 – Kaplan-Meier-Kurve zum entsprechenden Zeitpunkt) die Wahrscheinlichkeit eines unerwünschten Ereignisses bei Vorliegen konkurrierender Ereignisse überschätzt. Darüber hinaus sieht die Kommission als besonders problematisch an, dass in der Dokumentvorlage für das Dossier zur Nutzenbewertung (Version vom 21.02.2019) explizit die Erstellung von Kaplan-Meier-Kurven gefordert wird. Es sollte darauf hingewiesen werden, dass bei Vorliegen von konkurrierenden Ereignissen der Aalen-Johansen-Schätzer für die Wahrscheinlichkeit eines unerwünschten Ereignisses zu verwenden ist.

Referenzen

- Allignol A, Beyersmann J, Schmoor C (2016) Statistical issues in the analysis of adverse events in time-to-event data. *Pharmaceutical Statistics* 15: 297-305
- Arends LR, Hamza TH, van Houwelingen JC, Heijenbrok-Kal MH, Hunink MGM, Stijnen T (2008) Bivariate Random Effects Meta-Analysis of ROC Curves. *Medical Decision Making* 28: 621-638
- Steinhäuser S, Schumacher M, Rücker G (2016) Modelling multiple thresholds in meta-analysis of diagnostic test accuracy studies. *BMC Medical Research Methodology* 16: 97
- Hoyer A, Hirt S, Kuss O (2018) Meta-analysis of full ROC curves using bivariate time-to-event models for interval-censored data. *Research Synthesis Methods* 9: 62-72
- Jones HE, Gatsonis CA, Trikalinos TA, Welton NJ, Ades AE (2019) Quantifying how diagnostic test accuracy depends on threshold in a meta-analysis. *Statistics in Medicine* 38: 4789-4803
- Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and Presenting Results. In Deeks JJ, Bossuyt PM, Gatsonis C (editors), *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0*. The Cochrane Collaboration, 2010. Available from: <http://srdta.cochrane.org/>

**A.1.16 – Deutsche Gesellschaft für Orthopädie und Unfallchirurgie e. V. (DGOU) /
Deutsche Gesellschaft für Orthopädie und Orthopädische Chirurgie e. V.
(DGOOC) / Deutsche Gesellschaft für Unfallchirurgie e. V. (DGU) –
(gemeinsame Stellungnahme)**

Autorinnen und Autoren

- Kladny, Bernd
- Pennig, Dietmar

GESCHÄFTSSTELLE

DGOU-/ DGOOC-/ DGU-Geschäftsstelle · Straße des 17. Juni 106-108 · 10623 Berlin

Frau

Institut für Qualität und Wirtschaftlichkeit
im Gesundheitswesen (IQWiG)
Im Mediapark 8

50670 Köln

Per E-Mail: methoden@iqwig.de

DGOU e. V. / DGOOC e. V. / DGU e. V.

Straße des 17. Juni 106-108

(Eingang Bachstraße)

10623 Berlin

Tel.: +49 30 3406036-00

Fax: +49 30 3406036-01

office@dgou.de

www.dgou.de

Berlin, 22.01.2020

**Gemeinsame Stellungnahme
der Deutschen Gesellschaft für Orthopädie und Unfallchirurgie (DGOU),
der Deutschen Gesellschaft für Orthopädie und Orthopädische Chirurgie (DGOOC) und
der Deutschen Gesellschaft für Unfallchirurgie (DGU)
zum Entwurf der Allgemeinen Methoden Version 6.0**

Sehr ,

die Deutsche Gesellschaft für Orthopädie und Unfallchirurgie (DGOU), die Deutsche Gesellschaft für Orthopädische Chirurgie (DGOOC) und die Deutsche Gesellschaft für Unfallchirurgie nehmen Bezug auf die Anfrage des IQWiG zum Entwurf der Allgemeinen Methoden, Version 6.0, vom 05.12.2019.

Hierzu verweisen wir auf unsere Stellungnahme zur Version 5.0 aus dem Jahre 2017, welche wir diesem Schreiben nochmals beifügen und bitten um dessen Beachtung.

Mit freundlichen Grüßen

Prof. Dr. Dieter C. Wirtz
Generalsekretär der DGOU

Prof. Dr. Bernd Kladny
Stellv. Generalsekretär der DGOU

Vorstand (gemäß §26 BGB Abs. 1)

Präsident: Prof. Dr. Dieter C. Wirtz, Stellvertretender Präsident: Prof. Dr. Michael J. Raschke
Generalsekretär: Prof. Dr. Dieter C. Wirtz, Stellvertretender Generalsekretär: Prof. Dr. Bernd Kladny

A.1.17 – Deutsche Gesellschaft für Pneumologie und Beatmungsmedizin e. V. (DGP)

28. Januar 2020

**Anmerkungen der Deutschen Gesellschaft für Pneumologie
und Beatmungsmedizin e.V. (DGP) zu
IQWiG Allgemeine Methoden
Entwurf für Version 6.0 vom 05.12.2019**

1. Die statistischen Methoden der IQWiG Analysen sind aus Sicht der DGP adäquat. Bei den Arzneimittelbewertungen (z.B. Biologika, Triple Therapie für COPD) sehen wir Probleme mit der „Zweckmäßigen Vergleichstherapie“, die nicht vom IQWiG, sondern vom G-BA mit wenig transparenter Methodologie bestimmt und im Laufe der Bewertung ggf. mehrfach geändert wird. Häufig ist die Vergleichstherapie in Rahmen einer Zulassungs-RCT nicht praktikabel. Bei der frühen Bewertung nach Marktzulassung können in der Regel keine „head-to-head“-Vergleichsstudien (für Biologika) vorliegen. Der G-BA akzeptiert zwar auch indirekte Vergleiche, scheitert aber jedes Mal an der Verschiedenheit der Einschlusskriterien der entsprechenden Studien. Die Kriterien für die geforderte Vergleichbarkeit von Studienpopulationen sind im Vergleich zum Einschluss verschiedener Studien in eine Metaanalyse sehr streng angesetzt. In den „allgemeinen Methoden“ könnte ein Hinweis aufgenommen werden, der über das Problem der Vergleichbarkeit verschiedener Studien für einen indirekten Vergleich (zum Beispiel zweier Asthma-Biologika) Kriterien beschreibt.
2. Die ausdrücklich aufgeführte Akzeptanz von nicht doppelblinden oder nicht randomisierten Studien für die Bewertung nicht - medikamentöser Therapien (unter Berücksichtigung des niedrigeren Evidenzgrades) trotz eines geringeren möglichen Nutzengrades wird begrüßt.
3. Die Potenzialbewertung bedarf keiner Diskussion, da bis zur Einführung der neuen Methodologie das beschriebene Verfahren obsolet sein wird (aufgrund der TSVG). Die neue Methode könnte zum Beispiel eher für die Bewertung von FeNO praktikabel sein.
4. Hinsichtlich der Bewertung von Leitlinien stehen wir vor dem Problem, dass sich durch die Dauer des Bewertungsprozesses eines Medikamentes, der mit der ersten Konsultation des Herstellers oft 2-3 Jahre vor der G-BA Beurteilung beginnt, die Leitlinien ändern und dadurch der G-BA Beschluss bereits zum Zeitpunkt der Veröffentlichung obsolet ist. Das war zum Beispiel der Fall bei der Bewertung der Triple-Therapie für COPD gegenüber ICS/LABA anstatt LABA/LAMA. Es musste eine neue Prüfung durchgeführt werden.
5. Wir begrüßen, dass die Effektstärke nach anerkannten und jetzt festgelegten Kriterien mit in die Beurteilung einfließt (Cohen's d und Hedge's g). Bei der Beurteilung der patientenbezogenen klinischen Relevanz (MCID bzw. MID) sehen wir mit der pauschalen Verwendung von 15% des Skalawertes in den am meisten verwendeten pneumologischen Fragebögen Schwierigkeiten. St George's RQ, ACQ, CAT etc. haben eine klinisch (einigermaßen validierte, besser gesagt weit akzeptierte) MCID

ANSCHRIFT

GESCHÄFTSFÜHRENDER VORSTAND

VEREINSREGISTER

STEUERNUMMER & GLÄUBIGER-ID

STELLUNGNAHME

von weit unter 15%. Bei der Dupilumab/Asthma Diskussion wurde auch ACQ 0.5 wie üblich in der Diskussion akzeptiert. Wird sich dies jetzt ändern?

Seite 2 | 2

6. Bei dem Verzerrungspotenzial der Fragebögen sollte dezidiert auf das Problem der Linearität der Scores hingewiesen werden (dort ist nur von Schweregradabhängigkeit die Rede).

A.1.18 – Deutsche Krankenhausgesellschaft e. V. (DKG)

Autorinnen und Autoren

- Brenske, Michael

Institut für Qualität und
Wirtschaftlichkeit im Gesundheitswesen
-Stellungnahme zu den „Allgemeinen
Methoden Version 6.0“-
Im Mediapark 8 (Köln Turm)
50670 Köln

Stellungnahme der Deutschen Krankenhausgesellschaft zum Entwurf „Allgemeine Methoden Version 6.0“ des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) vom 05.12.2019

Sehr geehrter Herr Professor Windeler,
sehr geehrte Damen und Herren,

mit dem Entwurf der „Allgemeinen Methoden Version 6.0“ legen Sie erneut eine ausführliche Darstellung Ihrer Vorgehens- und Arbeitsweisen vor. Die damit geschaffene Transparenz und Nachvollziehbarkeit Ihrer unterschiedlichen Produkte begrüßen wir sehr. Dies trifft gleichermaßen auf den Umstand zu, dass die Allgemeinen Methoden damit einer regelmäßigen Überarbeitung und Weiterentwicklung unterliegen und die Möglichkeit erhalten geblieben ist, über ein Stimmabgabeverfahren in einen Austausch über deren Inhalte zu treten.

Gern möchten wir daher die Gelegenheit nutzen, Ihnen Hinweise und Anregungen für eine Überarbeitung des vorgelegten Entwurfes zu übermitteln. Dazu sei angemerkt, dass es durch den mittlerweile erreichten Umfang des Methodenpapiers nicht möglich war, jede ihrer Darstellungen einer ausführlichen Prüfung unterziehen zu können. Somit ist nicht ausgeschlossen, dass sich neben den nachfolgend aufgeführten Aspekten noch weiterer Bedarf für Anmerkungen ergeben kann. Insofern ist unsere Stellungnahme als nicht abschließend anzusehen und wir behalten uns vor, ggf. weitere Aspekte in zukünftige Diskussionsprozesse einzubringen. Im Einzelnen möchten wir an dieser Stelle auf folgende Punkte eingehen:

Nutzenbeleg/ Endpunktbezogene Bewertung

Zu diesem Thema möchten wir einen Aspekt erneut aufgreifen, der bereits Gegenstand einer früheren Stellungnahme unsererseits war. Demnach erfolgt im Kapitel 3.1 (Patien-

tenrelevanter medizinischer Nutzen und Schaden) eine ausführliche Darlegung der vom IQWiG angewendeten Maßstäbe bei der Nutzenbewertung. Den bereits in einer Vorgängerversion der „Allgemeinen Methoden“ eingeführten Versuch, Aussagesicherheiten mit einer Kategorisierung, verbunden mit konstanten Begrifflichkeiten zu versehen, können wir dem Grunde nach weiterhin gut nachvollziehen (Kapitel 3.1.4, S. 50 ff.). Es wurden dabei die Begriffe „Beleg“, „Hinweis“ und „Anhaltspunkt“ gewählt. Die Nutzung des Begriffs „Beleg“ bedarf dabei aus unserer Sicht allerdings ergänzender Erläuterungen. Dies hängt damit zusammen, dass für das Vorliegen eines „Belegs“ für den Nutzen im Methodenpapier regelhaft mindestens zwei randomisierte Studien gefordert werden. Dies kann jedoch als im Widerspruch zum Sozialgesetzbuch V stehend angesehen werden. Werden nämlich die gesetzlichen Regelungen der §§ 137c, 137e und 137h SGB V im Hinblick auf die Situation der Durchführung einer Erprobung (mögliche Entscheidung zur Durchführung einer Erprobung, ohne dass bislang ein RCT existiert) durchdekliniert, so ist der hinreichende Beleg eines Nutzens auch auf Grundlage eines einzigen RCT erbringbar.

An anderer Stelle des Methodenpapiers wird zwar auf einen Sonderfall eingegangen, bei dem auch ein RCT ausreichen könnte. Allerdings stellen die durch das SGB V vorgegebenen Bedingungen keinen „Sonderfall“ dar. Darüber hinaus erscheint auch die für den „Sonderfall“ vorgesehene Mindestzahl von 1000 Patienten je Studienarm wenig kompatibel mit den bereits genannten gesetzlichen Regelungen zu sein. Beispielweise ist für eine Erprobung im Kontext des § 137h SGB V in der Regel ein Zeitraum von zwei Jahren vorgesehen. Es erscheint wenig wahrscheinlich, in dieser kurzen Zeit regelhaft 2000 Patienten rekrutieren zu können.

Die jüngere Praxis zeigt zudem, dass das IQWiG von diesen Zahlen bereits deutlich abgewichen ist. Beispielweise wird im IQWiG-Addendum im Kontext des 137h-Verfahrens zur sonografiegesteuerten hochfokussierten Ultraschalltherapie bei bösartigen Neubildungen der Leber und der intrahepatischen Gallengänge vom August 2019 dargelegt, dass das IQWiG im Zuge der Kostenschätzung für den Fall einer Erprobung gemäß § 137e SGB V hier offenbar von einer Studie in der Größenordnung von ca. 150 einzuschließenden Patientinnen und Patienten ausgeht (s. a. <https://www.iqwig.de/de/projekte-ergebnisse/projekte/nichtmedikamentoese-verfahren/h-projekte/h19-03-sonografiegesteuerte-hochfokussierte-ultraschalltherapie-bei-boesartigen-neubildungen-der-leber-und-der-intrahepatischen-gallengaenge-2-addendum-zum-auftrag-h16-02d.12208.html>).

Ein weiteres diesbezügliches Beispiel, zu einer im Übrigen weit häufigeren Krankheitsentität, stellt die Kostenschätzung zu einer möglichen Erprobung im kürzlich veröffentlichten Addendum zur sonografiegesteuerten hochfokussierten Ultraschalltherapie beim Leiomyom des Uterus dar. Dort wurden für Studien mit mittelgroßer Fallzahl ca. 400 einzuschließende Patientinnen veranschlagt (s. a. <https://www.iqwig.de/de/projekte-ergebnisse/projekte/nichtmedikamentoese-verfahren/h-projekte/h19-02-sonografiegesteuerte-hochfokussierte-ultraschalltherapie-beim-leiomyom-des-uterus-2->

addendum-zum-auftrag-h16-02b.12207.html). Auch der Gemeinsame Bundesausschuss (G-BA) ging in allen von ihm bisher bereits beschlossenen Erprobungsstudien von Patientenzahlen aus, die sehr deutlich unter 2000 (sogar unter 600) Teilnehmerinnen bzw. Teilnehmern liegen.

Aus unserer Sicht wäre es daher zum Einen wichtig und zielführend, wenn ergänzende Erläuterungen dahingehend aufgenommen werden würden, dass es sich bei der vom IQWiG gewählten Terminologie in diesem Fall um eine eigene Kategorisierung des IQWiG handelt, deren Begrifflichkeiten zwar teilweise gleichlautend („Beleg“) zu gesetzlich verwendeten sind, diesen aber in Bedeutung und Wertung nicht gleichzusetzen sind. Des Weiteren erscheint es zweckmäßig, in ergänzenden Erläuterungen darzustellen, dass in die Beantwortung der Frage nach der geeigneten Anzahl an Studienteilnehmern immer auch eine Reflexion der zur Studienfrage vorliegenden, zum Teil vielschichtigen Rahmenbedingungen enthalten ist. Das dies ein nicht ganz unerheblicher Punkt ist, lässt sich auch daran ablesen, dass sogar der Gesetzgeber dies indirekt adressiert hat. In § 137e Abs. 2 Satz 4 SGB V ist nunmehr die Regelung aufgenommen, dass die Anforderungen an die Erprobung unter Berücksichtigung der Versorgungsrealität zu gewährleisten haben, dass die Erprobung und die Leistungserbringung durchgeführt werden können.

Überdies sehen wir in anderen Punkten noch einen Überarbeitungsbedarf. Legt man Maßstäbe für einen einzelnen RCT als Beleg (mindestens 2000 Teilnehmer/innen mindestens 1000 je Studienarm) insbesondere auch im Bereich seltener oder sehr seltener Erkrankungen zugrunde, so wird einmal mehr deutlich, dass diese schwer realisierbar sein können. Zudem erscheint uns die Fokussierung auf den p-Wert mit $< 0,001$ insofern nicht verständlich, da aus dem p-Wert keine Effektstärke und keine Präzision ablesbar ist, sondern lediglich die Aussage eines statistisch signifikanten Unterschieds. Deshalb sollte an dieser wie auch an anderen Stellen des Methodenpapiers u. E. erwogen werden, grundsätzlich die Effektschätzung mit p-Wert und Konfidenzintervallen in die Betrachtung einzubeziehen, da Konfidenzintervalle den Vorteil bieten, über die Stärke und die Präzision des Effekts unmittelbar Auskunft geben zu können.

Es fällt daneben außerdem auf, dass gerade in diesem wichtigen, entscheidungsrelevanten Kapitel 3.1.4 nur sehr wenige Referenzierungen vorgenommen werden. Dies erscheint umso ungewöhnlicher, da in vielen anderen Teilen des Papiers umfangreich auf Quellen bei einzelnen Aussagen hingewiesen wird. Daher wäre es aus unserer Sicht hilfreich, wenn der in Kapitel 3.1.4 erfolgte methodisch-wissenschaftliche Diskurs, und hier insbesondere die Ausführungen zum Nutzen, mit mehr Referenzierungen auf entsprechende Literatur versehen werden würde. Lediglich ein allgemeiner orientierender Hinweis auf die GRADE-Gruppe am Anfang des zentralen Kapitels 3.1.4 (S. 50) erscheint aus unserer Sicht hier nicht weitreichend genug, dies insbesondere auch, da an dieser Stelle GRADE nur in Teilen und nicht vollständig herangezogen wird. Beispielsweise werden Aspekte aus GRADE, die beobachtende Studien betreffen, hier nicht aufgegriffen. Für den Leser könnte es in seiner Orientierung zudem hilfreich sein,

bereits an dieser Stelle auch auf Ausführungen zu besonderen Gegebenheiten zu verweisen, die in weiteren Kapiteln abgehandelt werden (z. B. dramatischer Effekt, Umgang mit kleinen Populationen bei seltenen Erkrankungen).

Auswirkungen nicht publizierter Studienergebnisse auf das Fazit bei Nutzenbewertungen

In Kapitel 3.2.1 widmet sich das Methodenpapier dem bedeutsamen Thema der Auswirkungen nicht publizierter Studienergebnisse auf das Fazit bei Nutzenbewertungen. Auf S. 55 ff. wird jedoch nicht deutlich, wie der darin aufgeführte Nachweis fehlender Daten anhand des Abgleichs von registrierten und publizierten Studien geführt werden soll. So stellt sich z. B. die Frage, welche Zeiträume bis zu einer Publikation zugrunde gelegt werden, wie mit abgebrochenen Studien oder vorab nicht registrierten Studien umgegangen wird oder wie ein „*wahrscheinlicher Publikationsbias*“ genau definiert ist, der dazu führt, dass eine Bewertung von Nutzen und Schaden hinsichtlich Beleg, Hinweis oder Anhaltspunkt nicht vorgenommen wird.

Die für das Maß der Verzerrung durch Publikationsbias angeführten Kategorien „*vollständiger*“ und „*unvollständiger Datenüberübermittlung*“ und das „*Fehlen von Daten*“ in „*vernachlässigbarem*“, „*erheblichem*“ oder „*mäßigem Umfang*“ erscheinen in der Perspektive des äußeren Betrachters unpräzise. Überdies kann beim Abgleich von registrierten und publizierten Studien gerade die Registrierung zum Nachteil gereichen, während vorab nicht registrierte Studien gar nicht erst entdeckt und einbezogen werden, obwohl sie ebenfalls zum Publikationsbias beitragen können. Da die Studienregistrierung aus vielerlei Gründen ausgesprochen wünschenswert ist (A-Priori-Planung mit Studienprotokollen, Transparenz, frühzeitige Information, Vermeidung von Doppelarbeit etc./ siehe auch internationale Empfehlungen z.B. der „Good Clinical Practice“ (GCP) u. ä.) ist es ratsam, etwaige Nachteile aufgrund kontraproduktiver Anreizstrukturen zu vermeiden – zumal die Definition eines „*wahrscheinlichen Publikationsbias*“ im IQWiG-Methodenpapier ohnehin sehr vage bleibt und erheblichen Auslegungsspielraum lässt.

Die Frage des Ausmaßes von Publikationsbias kann aus grundsätzlichen Plausibilitätsbetrachtungen nicht abschließend beantwortet werden, da ohne Registrierung und/ oder ohne Publikation eine unbestimmbare Dunkelziffer bleibt. Zweifelsohne sind die Diskussion von Publikationsbias und dessen möglicher Einfluss auf die Aussagekraft der Ergebnisse wichtig und zu untersuchen (siehe z. B. auch weitere Kapitel der Allgemeinen Methoden des IQWiG zur Darstellung von Publikationbias mittels Funnelplot und Metaregression oder das Cochrane Collaboration Handbook). Allzu apodiktische Antworten jedoch, mit kategorischen Ausschlüssen von Nutzenbewertungen, erscheinen angesichts der Limitationen immanenter Bewertungsunsicherheiten widersprüchlich. Vielmehr erscheinen Aussagen auf dem Kontinuum mehr oder weniger wahrscheinlich verzerrter Ergebnisse im Sinne einer Annäherung als sinnvollerer Ansatz – dies auf der Grundlage umfangreicher wissenschaftlicher Studiensuche, die im vorlie-

genden Methodenpapier ausführlich beschrieben wird und die Einbeziehung identifizierter publizierter als auch nicht publizierter Daten berücksichtigt.

Dramatischer Effekt

Im Kapitel 3.2.2 wird dargelegt, bei welcher Größenordnung von einem dramatischen Effekt ausgegangen werden kann. Die genannte Größenordnung von 10 soll laut Methodenpapier dabei keine starre Grenze darstellen.

Der Anspruch, dass ein Effekt so groß sein muss, dass dies nicht allein durch Verzerrung erklärt werden kann, ist nachvollziehbar. Die Nennung einer spezifischen Größenordnung im Methodenpapier, auch wenn sie nicht als starr bezeichnet wird, erscheint hingegen wenig sachdienlich. Die Vergangenheit hat auch gezeigt, dass diese Größenordnung in Nutzenbewertungen des IQWiG selbst in Erkrankungssituationen „orientierend“ herangezogen wurde, in denen ein Wert 10 medizinisch gar nicht erreichbar ist. Denn die Höhe einer Grenze hängt ganz maßgeblich von der Erkrankungssituation ab. Hat man beispielsweise bei einer onkologischen Erkrankung bei einem schicksalhaften Verlauf 1-Jahres-Gesamtüberlebensraten von 10 %, in der Interventionsgruppe hingegen Raten von 40 % ist dies aus medizinischer Sicht ein dramatischer Effekt, der sich weit weg von der Größenordnung 10 bewegt. Ein dramatischer Effekt in Anlehnung an das Methodenpapier würde hingegen 1-Jahres-Überlebensraten von annähernd 100 % erfordern, was absolut unrealistisch ist.

Auch kann man sich der Frage des dramatischen Effekts nicht immer primär methodisch nähern. Wenn sich beispielsweise beim schicksalhaften Vergleich in einer Palliativsituation Patientinnen und Patienten in Studien oder in der Beobachtung befinden, die schon aufgrund ihres Allgemeinzustands einer Intervention nicht zugänglich sind, sind die Patientengruppen nicht wirklich vergleichbar. Die Verzerrung wird aus methodischer Sicht nachvollziehbar als groß eingestuft; mit der Folge, dass ein besonders hoher Effekt erwartet wird, um die Verzerrung zu kompensieren, was wiederum medizinisch gar nicht erreichbar ist. Das Werkzeug dramatischer Effekt stößt also in derartigen Erkrankungssituationen stark an seine Grenzen.

Diese Limitationen sollten in diesem Kapitel unbedingt Berücksichtigung finden. Auch sollte in den Berichtsprodukten des IQWiG dann jeweils dargelegt werden, ab welchem Wert und mit welcher Begründung ein Effekt als dramatisch bezeichnet werden kann.

Beurteilung klinischer Relevanz

Wie im Methodenpapier in Kapitel 9.3.3 aufgeführt wird, gibt es noch kein breit akzeptiertes methodisches Vorgehen für die Beurteilung der klinischen Relevanz von Studienergebnissen. Da in den letzten Jahren auch im Rahmen der Nutzenbewertung nach

§ 35 a SGB V vermehrt Responderanalysen auf Basis eines als Minimal Important Difference (MID) definierten Responsekriteriums durchgeführt wurden, besteht hier ein besonderer Bedarf, der Problematik unterschiedlicher MIDs zum gleichen Instrument zu begegnen. Grundsätzlich wird der Ansatz einer pragmatischen Vorgehensweise begrüßt, die eine Einbeziehung von Studienergebnissen zumindest nicht erschwert. Dennoch bleiben beim gewählten Ansatz Fragen offen:

So ist der gewählte Wert von 15 % vor dem Hintergrund der Literatur auf die (exemplarisch) verwiesen wird, nicht offenkundig nachvollziehbar. Er scheint sich eher in einem mittleren bis oberen Bereich der gefundenen Werte zu befinden, was dann jedenfalls als deutlicher Effekt zu interpretieren wäre. Besonderheiten, die sich beispielsweise durch unterschiedliche Schweregrade einer Erkrankung ergeben, wird durch eine derartige Festlegung weiterhin nicht Rechnung getragen.

Insbesondere ergibt sich die Frage, weshalb bei post-hoc spezifizierten Analysen nur solche mit einem Responsekriterium von genau 15 % des jeweiligen Skalenranges berücksichtigt werden. Zum einen liegt ein Kriterium von mindestens 15 % bereits eher am oberen Bereich der gefundenen MIDs und Unterschiede bei einem höheren Wert würden einen deutlichen Effekt nicht in Frage stellen. Zum anderen stellt sich die Frage, wie mit dieser präzisen Festlegung umgegangen werden soll, wenn ein genauer Wert von 15 % in der betreffenden Skala gar nicht abzubilden ist.

Umgang mit unterschiedlichen mittleren Beobachtungsdauern und Darstellung von Verzerrungsarten

Zum Umgang mit unterschiedlichen mittleren Beobachtungsdauern in Kapitel 9.3.13 auf S. 192 wäre es sinnvoll, einige Verfahren zu nennen, welche hier als adäquat für Überlebenszeiten erachtet werden. Gleiches gilt für die Darstellung von Verzerrungsarten in Kapitel 9.3.14 auf S. 194. Auch hier wäre es hilfreich, im zweiten Absatz verschiedene Imputationsverfahren beispielhaft anzuführen.

Diagnostische Verfahren

Im Kontext der Bewertung von diagnostischen Methoden erfolgt in den IQWiG-Methoden auch eine Befassung mit dem Werkzeug des „Linked-Evidence-Ansatzes“. Die Bausteine der diagnostisch-therapeutischen Kette/ Screeningkette können in Teilen mittels Linked-Evidence-Ansatz oder als Ganzes mittels komplexer Interventionsstudiendesigns untersucht werden. Angesichts des häufigen Mangels an Interventionsstudien zur vollständigen diagnostisch-therapeutischen Kette/ Screeningkette können mit dem Linked-Evidence-Ansatz Daten von Studien zur diagnostischen Güte herangezogen werden, die mit vergleichenden Therapiestudien verbunden werden, um auf die-

se Weise die Ergebnisse der diagnostisch-therapeutischen Kette/ Screeningkette als Bausteine zusammenführen.

Dieser Ansatz wird in Kapitel 3.5 und 3.6 allerdings nur im Zusammenhang mit der Bewertung von Screening-Verfahren beschrieben, d. h. bei anderen diagnostischen Verfahren scheinbar nicht in Betracht gezogen. Da es sich beim Screening ebenfalls um diagnostische Verfahren handelt, bleibt es unklar, warum hier offenbar unterschiedliche Bewertungsmaßstäbe zur Anwendung kommen. Immerhin betreffen Screeningverfahren einschließlich ihres Schadenspotenzials bei der Früherkennung von Erkrankungen viel größere Bevölkerungsgruppen. Gerade vor diesem Hintergrund stellt sich die Frage, warum bei der gezielten klinischen Diagnostik höhere Bewertungsmaßstäbe mit Interventionsstudien zur gesamten diagnostisch-therapeutischen Kette ohne die grundsätzliche Möglichkeit des Linked-Evidence-Ansatzes angelegt werden.

Potenzialbewertung

Bei den Darstellungen im Kapitel 3.8 „Potenzialbewertung“ fällt auf, dass an keiner Stelle die vom G-BA in seiner Verfahrensordnung niedergelegte Definition des Potenzials einer erforderlichen Behandlungsalternative wörtlich wiedergegeben ist. Gemäß 2. Kapitel § 14 Abs. 3 der Verfahrensordnung des G-BA ist dieses wie folgt definiert:

„Das Potenzial einer erforderlichen Behandlungsalternative kann sich etwa ergeben, wenn sie aufgrund ihres Wirkprinzips und der bisher vorliegenden Erkenntnisse mit der Erwartung verbunden ist, dass andere aufwändigere, für den Patienten invasivere oder bei bestimmten Patienten nicht erfolgreich einsetzbare Methoden ersetzt werden können, die Methode weniger Nebenwirkungen hat, sie eine Optimierung der Behandlung bedeutet oder die Methode in sonstiger Weise eine effektivere Behandlung ermöglichen kann. ...“

Die Ausführungen im Methodenpapier des IQWiG heben jedoch faktisch allein auf die „bisher vorliegenden Erkenntnisse“ ab, sodass unklar bleibt, ob überhaupt eine Auseinandersetzung mit dem Wirkprinzip erfolgt. Insofern ist hier ein Anpassungsbedarf gegeben, um dem Leser einen hinreichenden Überblick zu geben.

Diese Transparenz erscheint auch vor dem Hintergrund geboten, da es um die Auslegung dieses Potenzialbegriffes im Zusammenhang mit vier Entscheidungen des G-BA zur Methode „Ultraschallgesteuerter hoch-intensiver fokussierter Ultraschall“ im Rahmen von Bewertungen nach § 137c SGB V strittige Beschlussfassungen gab, welche anschließend vom Bundesministerium für Gesundheit im Rahmen seiner Prüfung nach § 94 SGB V beanstandet wurden (siehe z. B. https://www.g-ba.de/downloads/40-268-4962/2018-02-15_KHMe-RL_USg-HIFU_Endometriose_BMG.pdf). Allen diesen Beschlüssen war gemein, dass das Potenzial vom G-BA, nicht zuletzt auf Grundlage von Bewertungen des IQWiG, verneint wurde.

Wenngleich in Ausführungen am Ende des Kapitels darauf hingewiesen wird, dass es im Zuge des Terminservice- und Versorgungsgesetzes zu maßgeblichen Änderungen in § 137h SGB V in Bezug auf das Potenzial gekommen ist und ein zukünftiger Anpassungsbedarf des Methodenpapiers avisiert wird, erscheint es zweckmäßig, bereits auch in dieser Version schon Anpassungen vorzunehmen. Es wirkt unglücklich und ggf. verwirrend, wenn der Leser in den Erläuterungen zur Potenzialbewertung an den Regelungen des § 137h SGB V entlang geführt wird, um dann am Ende zu erfahren, dass diese zukünftig in dieser Form nicht mehr einschlägig sind.

Umgang mit unvollständigen Daten

Im Kapitel 9.3.12 wird der Ausschluss von Studienergebnissen für die Nutzenbewertung postuliert, wenn weniger als 70% der eingeschlossenen Patientinnen und Patienten in der Auswertung berücksichtigt werden. Ebenso sollen Studienergebnisse ausgeschlossen werden, wenn der Unterschied der Anteile nicht berücksichtigter Patientinnen und Patienten zwischen den Gruppen größer als 15 Prozentpunkte ist. Statt harter Cut-Offs mit gänzlichem Ausschluss möchten wir alternativ anregen, sie in den Auswertungen/ Metaanalysen durchaus zu berücksichtigen und anhand von Sensitivitätsanalysen sowohl mit als auch ohne Studien/ Daten, welche hohe Follow-Up-Verluste verzeichnen, die Robustheit respektive Stabilität des Gesamteffektes zu prüfen (siehe auch Cochrane Collaboration Handbook). Im Vergleich zum kompletten Ausschluss sind bei diesem Verfahren die Datenverluste in Metaanalysen deutlich geringer; zugleich kann diese Vorgehensweise Auskunft über Art und Umfang möglicher Verzerrungen durch Studien mit hohen Follow-Up-Verlusten geben.

Wir hoffen, dass wir Ihnen eine Reihe von wertvollen Hinweisen für die Überarbeitung des Entwurfes geben konnten. Diese Stellungnahme ist als institutionelle Meinungsäußerung aufzufassen, an der verschiedene Mitarbeiterinnen und Mitarbeiter der DKG beteiligt waren. Im Falle einer mündlichen Anhörung gehen wir davon aus, eine Vertretung der DKG zu entsenden.

Mit freundlichen Grüßen
Der Hauptgeschäftsführer
In Vertretung



Dr. Michael Brenske
Geschäftsführer Dezernat VI - Medizin II

A.1.19 – Ecker + Ecker GmbH

Autorinnen und Autoren

- Ecker, Christof
- Ecker, Thomas
- Pütz, Claudia
- Radke, Mareike
- von Wulffen, Meike

**Stellungnahme der Ecker + Ecker GmbH zu
Allgemeine Methoden – Entwurf 6.0 vom 05.12.2019**

Autoren:

Dr. Christof Ecker

Dr. Thomas Ecker

Dr. Claudia Pütz

Mareike Radke

Dr. Meike von Wulffen

Einleitung

Die Ecker + Ecker GmbH unterstützt pharmazeutische Unternehmen regelmäßig im Verfahren der frühen Nutzenbewertung. Die Ecker + Ecker GmbH nimmt aus drei Kompetenzbereichen Stellung zu dem Entwurf des IQWiG.

Kompetenzbereich Nutzenbewertung

1. Nicht medikamentöse therapeutische Interventionen

Kompetenzbereich Biostatistik

2. Beurteilung klinischer Relevanz

Kompetenzbereich Gesundheitsökonomie

3. Kostendarstellung

Kompetenzbereich Nutzenbewertung:

1 Nicht medikamentöse therapeutische Interventionen

Bezug: Abschnitt 3.4 der Allgemeinen Methoden, Entwurf für Version 6.0

Problembeschreibung:

In Abschnitt 3.4 finden sich Erläuterungen zur Bewertung von nicht medikamentösen therapeutischen Interventionen. Das Institut stellt Problemfelder bei der Durchführung von Studien zu nicht medikamentösen Therapien dar und bewertet diese als „meist zumindest teilweise überwindbar“. Wie die Bewertung nicht medikamentöser therapeutischer Interventionen methodisch erfolgt, wird in dem oben genannten Abschnitt nicht erläutert.

Handlungsbedarf:

Der Sinn und Zweck des IQWiG-Methodenpapiers ist die Darlegung der methodischen Vorgehensweise des Instituts bei der Prüfung unterschiedlicher Sachverhalte. Der vorliegende Abschnitt erfüllt diesen Zweck gerade nicht. Vielmehr beschreibt und bewertet das Institut nur die Herausforderungen und Schwierigkeiten, die sich bei Studien-Designs zu nicht medikamentösen therapeutischen Interventionen stellen. Das methodische Vorgehen zur Bewertung nicht medikamentöser therapeutischer Interventionen bleibt offen.

Vor diesem Hintergrund wäre es wichtig, im Abschnitt 3.4 das Prüfungsvorgehen des IQWiG zu konkretisieren.

Lösungsvorschlag:

Es wird daher eine Überarbeitung des vorgenannten Abschnitts 3.4 angeraten. Es sollte in diesem Abschnitt eine Beschreibung der Vorgehensweise bei der Bewertung von Studien zu nicht medikamentösen therapeutischen Interventionen eingefügt werden, die transparent die Methoden des Instituts bei dieser Prüfung darlegt.

Kompetenzbereich Biostatistik:

2 Beurteilung klinischer Relevanz

Bezug: Abschnitt 9.3.3 der Allgemeinen Methoden, Entwurf für Version 6.0

Problembeschreibung:

In Abschnitt 9.3.3 behandelt das IQWiG das methodische Vorgehen zur Beurteilung der klinischen Relevanz von Gruppenunterschieden. Im Fall von Responderanalysen werden dazu sogenannte MID (*Minimal Important Difference*) verwendet, die üblicherweise mit ankerbasierten Methoden, ggf. auch in Verbindung mit verteilungsbasierten Methoden, bestimmt werden [1-3]. Im Gegensatz zu dem bisherigen Vorgehen möchte das IQWiG künftig nur noch MID akzeptieren, die mindestens 15 % der Skalenspannweite des jeweiligen Messinstruments entsprechen, ansonsten soll als einheitliche MID 15 % der Skalenspannweite verwendet werden.

Diese Festlegung ist insbesondere aufgrund der folgenden fünf Aspekte problematisch:

- 1) Die Skalenspannweite wird nicht im Hinblick auf die klinische Relevanz von Unterschieden ausgewählt.
- 2) Eine einheitliche Grenze wird den Unterschieden zwischen vorkommenden Skalen und Indikationen nicht gerecht.
- 3) Die resultierenden MID würden in vielen Fällen erheblich von bereits anerkannten Werten abweichen.
- 4) Die Begründung der gewählten Grenze von 15 % durch das IQWiG ist methodisch fragwürdig.
- 5) Für Endpunkte mit offenen Skalen ist die Methode prinzipiell nicht anwendbar.

Zu 1) Die Problematik lässt sich am Beispiel der Körpertemperatur darstellen. Eine physikalisch sinnvolle Spanne ist 0°C bis 100°C, die minimale relevante Differenz wäre somit dem IQWiG zufolge 15°C. Eine medizinisch sinnvollere Wahl wäre die Spanne von 15°C bis 45°C, bei Anwendung der vom IQWiG vorgeschlagenen Methodik wäre dann ein Unterschied von 4,5°C relevant. Je nach Wahl der Skalenspannweite würde die MID also um den Faktor 3,3 und damit in erheblicher Weise variieren. Die Wahl der Skalenspannweite bei der Fragebogenerstellung ist individuell im Hinblick auf den Kontext begründet und somit nicht beliebig. Sie erfolgt jedoch nicht im Hinblick auf die Festlegung der MID im Sinne des IQWiG-Vorschlags und ist damit für diesen Zweck nicht sachgerecht. Würden Relevanzgrenzen auf diese Weise bestimmt, so wären sie weitgehend beliebig und zur Beurteilung der klinischen Relevanz von Gruppenunterschieden völlig ungeeignet.

Zu 2) Nach dem aktuellen Stand der Wissenschaft besteht ein breiter Konsens darüber, dass MID auch indikations- oder domänenspezifisch sein können, da die Relevanz von Veränderungen von der Patientenpopulation, der Krankheitsschwere und anderen kontextuellen Faktoren abhängt (s. auch vorliegender Entwurf für Version 6.0 des IQWiG-Methodenpapiers) [3, 4]. Ein- und dasselbe Instrument kann daher in verschiedenen Indikationen unterschiedliche MID aufweisen. So zeigen Studien, dass für den generischen SF-36-Fragebogen je nach Indikation unterschiedliche MID ermittelt wurden. Für Patienten mit idiopathischer Lungenfibrose wurde beispielsweise auf der Subskala zum allgemeinen Gesundheitszustand eine MID von 2 Punkten ermittelt [5], während für Patienten mit Prostatakrebs eine MID für die gleiche Subskala von 8 Punkten bestimmt wurde [6]. Bei beiden Studien wurden sowohl verteilungsbasierte als auch ankerbasierte Methoden verwendet. Umgerechnet auf die Spannweite der Skala von 0 – 100 würden somit – neben den deutlichen Abweichungen zwischen den MID in den beiden Indikationen – beide MID weit unter der vom IQWiG vorgeschlagenen 15 %-Grenze liegen. Beispiele wie dieses unterstreichen, dass ein festgesetzter Prozentsatz der Skalenspannweite nicht als Grenzwert zur Bestimmung einer MID für alle Instrumente in allen Indikationen geeignet sein kann.

Zu 3) Die resultierenden MID würden erheblich von den in der Nutzenbewertung bislang anerkannten MID abweichen, obwohl diese mit Hilfe von in der Wissenschaft weithin akzeptierten Methoden ermittelt wurden. Einige Beispiele von MID, die bislang durch das IQWiG und den G-BA akzeptiert worden sind und die mit dem Vorschlag des IQWiG für nichtig erklärt würden, sind:

- FACT-P-Gesamtscore (6,4 %) [7, 8]
- FACT-G-Gesamtscore (4,6 %) [9, 10]
- FACIT-Fatigue (7,7 %) [11, 12]
- EORTC QLQ-C30 (10,0 %) [13, 14]

In Klammern ist jeweils der entsprechende Wert als prozentualer Anteil der jeweiligen Skalenspannweite gezeigt. Die Abweichungen zu der Grenze von 15 % sind erheblich.

Zu 4) Die Herleitung des Wertes 15 % ist rechnerisch nicht nachvollziehbar. So ist die Mehrheit (etwa 70 %) der in der zitierten Literatur aufgeführten MID kleiner als 15 % und würde daher künftig als zu klein abgelehnt. Darüber hinaus kann die angeführte Literatur als Begründung der 15 %-Grenze nicht herhalten. Es bleibt unklar, wie das IQWiG aus der angegebenen Literatur den Wert von 15 % abgeleitet hat. Ebenso ist nicht nachvollziehbar, warum gerade die beiden angeführten systematischen Übersichtsarbeiten ausgewählt wurden, um einen einheitlichen Grenzwert abzuleiten. Ein evidenzbasiertes Vorgehen, beispielsweise auf Basis einer systematischen Literaturrecherche, wäre hier wünschenswert.

Zu 5) Eine Limitation der Methode besteht darüber hinaus bei Skalen, die nach oben oder unten (oder beiden Seiten) unbegrenzt sind. Für diese lässt sich keine Spannweite und daher nach der vorgeschlagenen Methode auch keine MID bestimmen. Beispiele hierfür sind:

- 6-Minuten-Gehtest (Strecke in Meter, Skala nach oben offen)
- 9-Hole Peg Test (Zeit in Sekunden, Skala nach oben offen)
- Timed 25-Foot Walk (Zeit in Sekunden, Skala nach oben offen)

Handlungsbedarf:

Der derzeitige Entwurf des IQWiG sieht vor, dass eine international anerkannte und etablierte Vorgehensweise durch eine Methodik abgelöst wird, die weder evidenzbasiert noch anerkannt oder geeignet ist.

Lösungsvorschlag:

Im Rahmen dieser Stellungnahme wird die erneute Überarbeitung des Abschnitts 9.3.3 „Beurteilung klinischer Relevanz“ vorgeschlagen. Dabei ist insbesondere darzulegen, wie die methodische Qualität von Studien zur Validierung von MID sowie die Aussagekraft der ermittelten MID aus diesen etablierten Verfahren (anker- und verteilungsbasierte Herleitung) bewertet werden soll. Die derzeit vorgeschlagene Methode könnte allenfalls als Notlösung für den Fall, dass keine geeignete MID vorliegt, Anwendung finden. Die Wahl des Grenzwertes sollte in diesem Fall evidenzbasiert erfolgen und nachvollziehbar sein.

Kompetenzbereich Gesundheitsökonomie:**3 Kostendarstellung**

Bezug: Abschnitt 3.3.3 der Allgemeinen Methoden, Entwurf für Version 6.0

Problembeschreibung:

In diesem Abschnitt erläutert das IQWiG, dass im Rahmen der Nutzenbewertung eines Arzneimittels gemäß § 35a SGB V der pharmazeutische Unternehmer u. a. Angaben zu den Kosten der Therapie für die gesetzliche Krankenversicherung machen muss. In Abschnitt 3.3.3 finden sich jedoch keine methodischen Anmerkungen, wie das Institut die Angaben zu den Kosten der Therapie für die gesetzliche Krankenversicherung überprüft. Dies wurde bereits in den beiden Stellungnahmeverfahren zu Allgemeine Methoden, Entwurf für die Version 4.2 vom 18.06.2014 und Entwurf für die Version 5.0 vom 07.12.2016, angeführt.

Handlungsbedarf:

Die Bewertung der Angaben zu den Kosten der Therapie für die gesetzliche Krankenversicherung durch das IQWiG weist Inkonsistenzen auf:

- Kostendarstellungen in vergleichbaren therapeutischen Situationen werden ungleich bewertet.
- Die Bewertungen durch das IQWiG weichen nicht unerheblich vom G-BA in den Beschlüssen ab.

Diese Inkonsistenzen sind wissenschaftlich nicht nachvollziehbar.

Lösungsvorschlag:

Im Rahmen dieser Stellungnahme wird die Einführung eines neuen Abschnitts „3.4 Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V“ vorgeschlagen. In diesem Abschnitt sollte ein Unterabschnitt „3.4.5 Kosten“ eingeführt werden. In diesem Abschnitt wäre u. a. zu beschreiben,

- welche Packungen für eine Kostenkalkulation herangezogen werden,
- welche Abzüge vom Apothekenverkaufspreis abgezogen werden,
- wann Kostenspannen angesetzt werden,
- welche Kostenpositionen saldiert werden,
- inwieweit *Companion Diagnostics* als zusätzliche GKV-Leistung berücksichtigt werden,
- welche Sonderregeln für die Kostendarstellung bei Orphan Drugs bestehen.

Ecker + Ecker GmbH

24. Januar 2020

Literaturverzeichnis

1. Coon CD, Cook KF (2018): Moving from significance to real-world meaning: methods for interpreting change in clinical outcome assessment scores. Qual Life Res; 27(1):33-40.
2. U.S. Department of Health and Human Services Food and Drug Administration (2009): Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. [Zugriff: 17.01.2020]. URL: <https://www.fda.gov/media/77832/download>.
3. European Medicines Agency (2016): Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man. [Zugriff: 17.01.2020]. URL: https://www.ema.europa.eu/en/documents/other/appendix-2-guideline-evaluation-anticancer-medicinal-products-man_en.pdf.
4. Revicki D, Hays RD, Cella D, Sloan J (2008): Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. Journal of Clinical Epidemiology; 61(2):102-9.
5. Swigris JJ, Brown KK, Behr J, du Bois RM, King TE, Raghu G, et al. (2010): The SF-36 and SGRQ: validity and first look at minimum important differences in IPF. Respir Med; 104(2):296-304.
6. Jayadevappa R, Malkowicz SB, Wittink M, Wein AJ, Chhatre S (2012): Comparison of distribution- and anchor-based approaches to infer changes in health-related quality of life of prostate cancer survivors. Health Serv Res; 47(5):1902-25.
7. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2019): Apalutamid (Prostatakarzinom) - Nutzenbewertung gemäß § 35a SGB V. IQWiG-Berichte - Nr. 762. [Zugriff: 17.01.2020]. URL: https://www.g-ba.de/downloads/92-975-2909/2019-02-01_Nutzenbewertung-IQWiG_Apalutamid_D-437.pdf.
8. Gemeinsamer Bundesausschuss (2019): Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V Apalutamid. [Zugriff: 17.01.2020]. URL: https://www.g-ba.de/downloads/39-261-3907/2019-08-01_AM-RL-XII_Apalutamid_D-437_BAnz.pdf.
9. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2018): Tivozanib (Nierenzellkarzinom) - Addendum zum Auftrag A17-58. IQWiG-Berichte - Nr. 608. [Zugriff: 17.01.2020]. URL: https://www.g-ba.de/downloads/92-975-2269/2018-11-01_Addendum-IQWiG_Tivozanib_D-323_Addendum.pdf.
10. Gemeinsamer Bundesausschuss (2018): Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Tivozanib. [Zugriff: 17.01.2020]. URL: https://www.g-ba.de/downloads/39-261-3291/2018-04-19_AM-RL-XII_Tivozanib_D-323_BAnz.pdf.
11. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2017): Sarilumab (Rheumatoide Arthritis) - Nutzenbewertung gemäß § 35a SGB V. [Zugriff: 17.01.2020]. URL: https://www.g-ba.de/downloads/92-975-2032/2017-08-15_Nutzenbewertung-IQWiG_Sarilumab_D-299.pdf.
12. Gemeinsamer Bundesausschuss (2018): Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Sarilumab. [Zugriff: 17.01.2020]. URL: https://www.g-ba.de/downloads/39-261-3221/2018-02-15_AM-RL-XII_Sarilumab_D-299_BAnz.pdf.
13. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2019): Pembrolizumab (plattenepitheliales NSCLC, Kombinationschemotherapie) - Nutzenbewertung gemäß § 35a SGB V. IQWiG-Berichte - Nr. . [Zugriff: 17.01.2020]. URL: https://www.g-ba.de/downloads/92-975-3029/2019-04-01_Nutzenbewertung-IQWiG_Pembrolizumab_D-448.pdf.
14. Gemeinsamer Bundesausschuss (2019): Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V Pembrolizumab (neues Anwendungsgebiet: nicht-kleinzelliges Lungenkarzinom, plattenepithelial, Erstlinie, Kombination mit Carboplatin und (nab-) Paclitaxel). [Zugriff: 17.01.2020]. URL: https://www.g-ba.de/downloads/39-261-3958/2019-09-19_AM-RL-XII_Pembrolizumab_D-448_nAWG_BAnz.pdf.

A.1.20 – Flatiron Health, Inc.

Autorinnen und Autoren

- Bierl, Michael

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Bierl, Michael
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
☒ im Namen folgender Institution / Organisation: Flatiron Health, Inc.
☐ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.

Flatiron Health, Inc. bedankt sich für die Möglichkeit zur Stellungnahme zu den Allgemeinen Methoden: Entwurf für Version 6.0 des IQWiG. Die Anmerkungen berücksichtigen auch den Rapid Report A19-43.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
5.3.2 (S. 123)	<u>Anmerkung:</u> Ein möglicher Ansatz zur Evaluation von Fragen, die durch die Analyse von Versorgungsdaten beantwortet werden können, besteht darin Kooperationen mit Organisationen im Bereich der Datenanalyse und Registerbetreibern einzugehen, um Chancen und Risiken der Nutzung alternativer Evidenzquellen zu bestimmen. Solche Kooperationen sind Teil des Real-World-Evidence Rahmenwerks der U.S. Food and Drug Administration (FDA) mit Demonstrationsprojekte von diversen Vertretern aus Forschung, politischen Organisationen und Industrie, die darauf zielen besser zu verstehen wie und wann Versorgungsdaten für die Entscheidungsfindung eingesetzt werden können.¹ Ein ähnlicher Ansatz wurde im Bericht 2020 HMA-EMA Joint Big Data Taskforce Phase II report: Evolving Data-Driven Regulation vorgeschlagen. <u>Vorgeschlagene Änderung:</u> Erwägung von Kooperationen.
5.3.4 (S. 124)	<u>Anmerkung:</u> Wir empfehlen, Daten aus elektronischen Patientenakten als Datenquelle zu erwägen, sofern diese die im Rapid Report A19-43 beschriebenen Herausforderungen adressieren können. Solche Daten bieten klinische Tiefe, die derzeit nicht in Register- und Abrechnungsdaten erfasst wird, und ermöglichen Einblicke ohne zusätzlichen Dokumentationsaufwand durch Leistungserbringer.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	Zudem sollten Umstände berücksichtigt werden, unter denen internationale Daten aus elektronischen Patientenakten als Quelle von Sekundärdaten in Betracht kommen könnten. In Fällen, in denen ein Medikament oder eine Intervention bereits außerhalb der EU zugelassen wurde, könnten solche Daten beispielsweise relevante Nachweise für die Wirksamkeit und Nutzung in der klinischen Routinepraxis liefern.^{2,3} Dies gilt insbesondere in Fällen von seltenen Erkrankungen oder bei Populationen, die über Biomarker definiert werden und für die geringe Patientenzahlen vorliegen. <u>Vorgeschlagene Änderung:</u> Ergänzung von Daten aus elektronischen Patientenakten und internationalen Daten.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

1. Khozin, S et al. 2019. Real-World Progression, Treatment, and Survival Outcomes During Rapid Adoption of Immunotherapy for Advanced Non-Small Cell Lung Cancer. Cancer. 2019 Nov 15;125(22):4019-4032. doi: 10.1002/cncr.32383. Epub 2019 Aug 5.
2. Parikh, RB et al Association Between FDA Label Restriction and Immunotherapy and Chemotherapy Use in Bladder Cancer. JAMA. 2019 Sep 24;322(12):1209-1211. doi: 10.1001/jama.2019.10650.
3. Bennette, CS et al. Diffusion of innovation in oncology: A case study of immuno-oncology (IO) adoption for advanced non-small lung cancer (aNSCLC) patients across practices in the US. DOI: 10.1200/JCO.2018.36.15_suppl.6537 *Journal of Clinical Oncology* 36, no. 15_suppl (May 20, 2018) 6537-6537.

A.1.21 – Geschäftsstelle der Beauftragten der Bundesregierung für die Belange der Patientinnen und Patienten

Autorinnen und Autoren

- Penning-Poggenbeck, Doreen

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Penning-Poggenbeck, Doreen,
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
☒ im Namen folgender Institution / Organisation: Geschäftsstelle der Beauftragten der Bundesregierung für die Belange der Patientinnen und Patienten
☐ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.

Wegfall der Bewertung der Versorgungsqualität

Das Kap.5 wurde von Leitliniensynopsen und Versorgungsanalyse in Bewertungen der Versorgung unbenannt und grundlegend überarbeitet. Unter 5.3.2 wird nunmehr statt einer Versorgungsanalyse der Fokus auf die Analyse von Versorgungsdaten gelegt und das methodische Vorgehen beschrieben.

Während bislang bei der Versorgungsanalyse das übergeordnete Ziel die Beurteilung der Versorgungsqualität war, steht nunmehr die Analyse der Versorgungsdaten und die Beschreibung der Gesundheitsversorgung im Vordergrund. Auf eine Darstellung und Bewertung der Versorgungsqualität wird in diesem Abschnitt verzichtet. Das heißt, dass die Versorgungsqualität weder in diesem Kapitel noch an einer anderen Stelle der Allgemeinen Methoden geprüft wird, was kritisch zu sehen ist. Die Bewertung der Versorgungsqualität ist aus Sicht der Patientinnen und Patienten ein wichtiges Kriterium, das auch in den Allgemeinen Methoden Berücksichtigung finden und wieder aufgenommen werden sollte.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
6.3.3 (S.127)	<u>Anmerkung:</u> Die Ämter: Patientenbeauftragte und Pflegebevollmächtigter wurden getrennt; Der Zusatz Pflegebevollmächtigter wurde aus § 139a Abs. 5 SGB V gestrichen. Dies wurde im übrigen Text auch bereits bedacht. <u>Vorgeschlagene Änderung:</u> Den Zusatz „und Bevollmächtigte beziehungsweise Bevollmächtigter für Pflege“ bitte streichen

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
6.3 Abbildung 16 (S. 126) und 6.3.4 (S.128)	<u>Anmerkung:</u> Nach der Änderung im Entwurf für Version 6.0. ergibt sich in der Beschreibung nicht eindeutig, ob in der 2. Stufe des Auswahlverfahrens bei der Auswahl das Benehmen mit dem erweiterten Fachbeirat herzustellen ist. Hingegen ergibt sich aus der Abb. 16, dass ein Benehmen mit dem Erweiterten Fachbeirat bei der Auswahl hergestellt wird. Diese Unklarheit sollte klargestellt werden. <u>Vorgeschlagene Änderung:</u> Streichung des Begriffs Benehmen aus der Abbildung 16 oder Klarstellung in 6.3.4.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

A.1.22 – GlaxoSmithKline GmbH & Co. KG. (GSK)

Autorinnen und Autoren

- Hennig, Michael

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
PD Dr. Michael Hennig, Leiter Biostatistik & Epidemiologie
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
☒ im Namen folgender Institution / Organisation: GlaxoSmithKline GmbH&Co. KG
☐ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
9.3.3 Beurteilung der klinischen Relevanz	<u>Anmerkung:</u> Das vom IQWiG vorgeschlagene Vorgehen, eine pauschale MID-Schwelle von 15 % der theoretisch möglichen Skalenspannweite für alle Erhebungsinstrumente anzuwenden, ist aus unserer Sicht wissenschaftlich nicht begründet und daher für uns nicht nachvollziehbar. Zur Herleitung dieser Schwelle bezieht sich das IQWiG auf „eine Sichtung von systematischen Übersichtsarbeiten“. Hier bleibt unklar, ob diese Sichtung vollständig und systematisch ist, da sich die beiden einzigen vom IQWiG exemplarisch angegebenen Quellen [Nordin 2016 (1), St-Pierre 2016 (2)] lediglich auf ausgewählte und sehr spezifische Symptome wie „<i>fatigue</i>“ bzw. „<i>symptoms and functional limitations in individuals with rotator cuff disorders</i>“ beziehen. Im Folgenden wird argumentiert, dass „<i>sich MIDs in vielen Fällen zwischen 10 % und 20 % der Spannweite der Scores eines Erhebungsinstruments bewegen</i>“. Auch hier bleibt völlig unklar, was unter „<i>vielen Fällen</i>“ zu verstehen ist und wie valide die angeführten Grenzen von 10 % bzw. 20 % wirklich sind. Anschließend wird auf dieser Basis mehr oder weniger willkürlich die 15 %-Schwelle für alle MIDs abgeleitet. Dieser Wert ergibt sich wohl als Mittelwert der o.g. Grenzen von 10 % und 20 % - eine konkrete Begründung für die Wahl von 15 % liegt allerdings nicht vor.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	Neben der Willkürlichkeit dieser Setzung bezweifeln wir die Sinnhaftigkeit einer pauschalen Schwelle. In der auch vom IQWiG zitierten Arbeit von Revicki [Revicki 2008 (3)] heißt es dazu aus unserer Sicht völlig treffend: <i>„The MID for a PRO instrument is not an immutable characteristic, but may vary by population and context, and no one MID may be valid for all study applications. MID estimates should be based on multiple approaches and triangulation of methods. Anchor-based methods applying various relevant patient-rated, clinician-rated, and disease-specific variables provide primary and meaningful estimates of an instrument’s MID.“</i> Die Autoren kommen in ihrer Arbeit daher zu folgender Empfehlung <i>„We recommend that the MID is based primarily on relevant patient-based and clinical anchors, with clinical trial experience used to further inform understanding of MID.“</i> Dieser Empfehlung schließen wir uns an, da nur durch die Berücksichtigung von Patienten- und Arzt-basierten Ankern sichergestellt werden kann, dass die abgeleitete MID auch klinisch relevant ist. Diese Sichtweise wird unter anderem auch von der EMA geteilt [EMA 2014 (4)]. Auch aktuelle Arbeiten zur Herleitung von MIDs in spezifischen Indikationen basieren im Wesentlichen auf ankerbasierten Verfahren und benutzen weitere Methoden zur Triangulation (z. B. Sully 2019 (5) für die Indikation Multiples Myelom). Für das in der Indikation der chronisch obstruktiven Lungenerkrankung (COPD) sehr häufig verwendete Erhebungsinstrument Saint-George’s Respiratory Questionnaire (SGRQ) würde die Anwendung der 15 % Schwelle zu einem MID von 15 führen, da die Skalenspannweite 0-100 beträgt. Die wissenschaftlich etablierte und in bisherigen Nutzenbewertungsverfahren akzeptierte MID liegt jedoch bei einem Wert von 4 [Jones 2005 (6)]. Somit würde durch die 15 % Schwelle eine Erhöhung um den Faktor 3,75 resultieren (von bisher 4 auf 15). Eine Überprüfung von bisherigen AMNOG-Verfahren zeigt, dass die vorgeschlagene 15 %-Schwelle in ca. 90 % der Fälle zu einer Erhöhung der bisher validierten und in der Anwendung akzeptierten MIDs führen würde. Es ist nicht nachvollziehbar, warum etablierte MIDs, deren klinische Relevanz bereits belegt ist, im Rahmen der Nutzenbewertung nicht weiterhin verwendet werden sollten. <u>Vorgeschlagene Änderung:</u> Die vom IQWiG vorgeschlagenen Änderungen sollen entfallen.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	Bei der Bestimmung einer validen MID sollte im Einklang mit der dazu vorhandenen wissenschaftlichen Literatur [Revicki 2008(3)] auf ankerbasierte Verfahren zurückgegriffen werden, da diese Verfahren die von Patienten wahrgenommenen Veränderungen direkt abbilden. Darüber hinaus beteiligt sich GSK gerne an einem konstruktiven Dialog mit allen Beteiligten, inklusive IQWiG, G-BA und Akademie, um einen für die Nutzenbewertung verbindlichen und wissenschaftlich fundierten Katalog von Kriterien zu definieren, die bei der Bestimmung von MIDs regelhaft anzuwenden sind.
3.3.3 Ausmaßbestimmung bei stetigen Daten Abschnitt D) Stetige oder quasistetige Zielgrößen mit jeweils vorliegenden SMDs	<u>Anmerkung:</u> Wir begrüßen den Versuch des IQWiG, Kriterien für das Ausmaß des Zusatznutzens auch für stetige oder quasistetige Zielgrößen abzuleiten. Dazu orientiert sich das IQWiG an dem Verfahren für binäre Daten nach Skipka et al.[Skipka 2016(7)] und an der Einteilung von Cohen's d und leitet Schwellenwerte für das Effektmaß SMD (Cohen's d bzw. Hedges' g) ab. Das vom IQWiG vorgeschlagene Vorgehen ist aus unserer Sicht nicht vollständig nachvollziehbar. So geht aus der Passage <i>„Man kann zeigen, dass die für ein bestimmtes Ausmaß wünschenswerte Effektstärke ungefähr halbiert werden muss, um den entsprechenden Schwellenwert abzuleiten.“</i> leider nicht eindeutig und für Außenstehende nachvollziehbar hervor, wie der Nachweis einer „ungefähren Halbierung“ erfolgt(e), da dazu auch keine weiteren Referenzen angegeben sind. Lediglich die Schwellenwerte für nicht schwerwiegende Symptome und Nebenwirkungen lassen sich auf Basis der üblichen Einteilung von Cohen's d nachvollziehbar herleiten, während die Schwellenwerte für schwerwiegende Symptome, Nebenwirkungen und gesundheitsbezogene Lebensqualität nicht nachvollziehbar und auch nicht wissenschaftlich begründet sind. <u>Vorgeschlagene Änderung:</u> Die Schwellenwerte zur Ausmaßbestimmung bei stetigen Daten sollten nachvollziehbar und wissenschaftlich begründet werden und den international anerkannten Standards und Kriterien der evidenzbasierten Medizin genügen. Zudem sollte bei der Herleitung berücksichtigt werden, dass Schwellenwerte nach Cohen's d für den Punktschätzer gelten und nicht für Konfidenzintervalle entwickelt wurden. Die nach diesen Kriterien hergeleiteten Schwellenwerte sollten in einem geeigneten statistischen Journal veröffentlicht werden.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
9.3.12 Umgang mit unvollständigen Daten	<u>Anmerkung:</u> Wir stimmen der grundsätzlichen Einschätzung zu, dass das Verzerrungspotenzial mit steigendem Anteil der fehlenden Werte zunimmt. Die vom IQWiG definierten Schwellenwerte für die Nichtberücksichtigung von Studienergebnissen sind jedoch nicht hinreichend mit Publikationen belegt. Insbesondere für den Schwellenwert von 15% fehlt eine Literaturquelle. <u>Vorgeschlagene Änderung:</u> Die vorgeschlagenen Schwellenwerte für die Nichtberücksichtigung von Studienergebnissen sollten mit Literaturangaben belegt werden.
9.3.8 Anwendung der Knapp-Hartung-Methode bei Metaanalysen mit zufälligen Effekten	<u>Anmerkung:</u> Wir begrüßen die vorgenommene Klarstellung und Präzisierung. <u>Vorgeschlagene Änderung:</u> Keine.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

Literaturverzeichnis

1. Nordin Å, Taft C, Lundgren-Nilsson Å, Dencker A. Minimal important differences for fatigue patient reported outcome measures-a systematic review. BMC medical research methodology. 2016;16:62.
2. St-Pierre C, Desmeules F, Dionne CE, Fremont P, MacDermid JC, Roy JS. Psychometric properties of self-reported questionnaires for the evaluation of symptoms and functional limitations in individuals with rotator cuff disorders: a systematic review. Disability and rehabilitation. 2016;38(2):103-22.
3. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. Journal of Clinical Epidemiology. 2008;61(2):102-9.
4. (EMA) EMA. Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man
The use of patient-reported outcome (PRO) measures in oncology studies. European Medicines Agency (EMA); 2014. Report No.: EMA/CHMP/292464/2014

5. Sully K, Trigg A, Bonner N, Moreno-Koehler A, Trennery C, Shah N, et al. Estimation of minimally important differences and responder definitions for EORTC QLQ-MY20 scores in multiple myeloma patients. *European Journal of Haematology*. 2019;103(5):500-9.
6. Jones PW. St. George's respiratory questionnaire: MCID. COPD: *Journal of Chronic Obstructive Pulmonary Disease*. 2005;2(1):75-9.
7. Skipka G, Wieseler B, Kaiser T, Thomas S, Bender R, Windeler J, et al. Methodological approach to determine minor, considerable, and major treatment effects in the early benefit assessment of new drugs. *Biometrical journal Biometrische Zeitschrift*. 2016;58(1):43-58.

A.1.23 – IQVIA Commercial GmbH & Co. OHG

Autorinnen und Autoren

- Batscheider, Ariane
- Böhm, Doris
- Greiner, Roger-Axel

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Doris Böhm
Dr. Roger-Axel Greiner
Dr. Ariane Batscheider
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
☒ im Namen folgender Institution / Organisation: IQVIA Commercial GmbH & Co. OHG
☐ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.

Allgemeine Anmerkungen

Das IQWiG hat gemäß § 139a SGB V zu gewährleisten, dass die Nutzenbewertung nach den internationalen Standards der evidenzbasierten Medizin erfolgt (BMJV 2019). In diesem Zusammenhang überarbeitet das IQWiG auf regelmäßiger Basis sein Methodenpapier. IQVIA hält das grundsätzlich für sinnvoll und begrüßt den wissenschaftlichen Austausch, der dadurch angestoßen wird.

Einige der vom IQWiG im Entwurf für Version 6.0 vorgeschlagenen „inhaltlichen Aktualisierungen und Ergänzungen“ stellen allerdings nicht nur Anpassungen, wie etwa Anpassungen an den neuesten Stand der Erkenntnisse oder die Angleichung der Methodik an internationale Standards dar, sondern sind – in unseren Augen – als Vorschlag für neue Standards zu sehen. Inwieweit die vorgeschlagenen Änderungen für die Nutzenbewertung zur Anwendung kommen oder auch Basis der Entscheidungen und Beschlüsse des G-BA werden, lässt sich derzeit schwer abschätzen. Als Dienstleister im Bereich der Nutzenbewertung von Arzneimitteln, Medizinprodukten und neuen Methoden hat IQVIA einen gewissen Überblick über historische und aktuelle Bewertungen des IQWiG und des G-BA. Vor diesem Hintergrund fragen wir uns, inwieweit durch die vorgeschlagenen Änderungen eine Konsistenz und Gleichbehandlung in der Bewertung, z.B. über verschiedene Wirkstoffe in demselben Anwendungsgebiet hinweg oder auch bei der Wiedervorlage von des Nutzendossiers für einen Wirkstoff im Falle einer Befristung gegeben ist.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
9.3.3 S. 176ff	<u>Anmerkung:</u> Bei patientenrelevanten Endpunkten, die mithilfe psychometrischer Skalen gemessen werden, ist zur Bewertung der beobachteten Effekte neben der statistischen Signifikanz auch die klinische Relevanz der Effekte entscheidend (Revicki et al. 2007). In der Nutzenbewertung nach §35a SGB V werden zum Zweck der Relevanzbewertung in der Regel Responderanalysen auf Basis einer Minimal Important Difference (MID) gegenüber einer Analyse von Mittelwertunterschieden bevorzugt (G-BA 2018).

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	Mit dem Hinweis, dass es derzeit keine akzeptierten Standards gibt, mit denen die Qualität von Studien zur Ermittlung von MIDs bewertet und die Aussagekraft der ermittelten MIDs abgeschätzt werden kann, definiert das IQWiG in seinem neuen Methodenpapier einen neuen MID Schwellwert. Es sollen regelhaft nur noch Responder-Auswertungen bewertet werden, die eine MID als Responsekriterium verwenden, die mindestens so groß ist wie 15 % der Skalenspannweite. Basis für dieses 15%-Kriterium ist eine Sichtung von Übersichtsarbeiten zur MID, bei der sich zeigte, „dass sich MIDs in vielen Fällen zwischen 10 % und 20 % der Spannweite der Scores eines Erhebungsinstruments bewegen“. Mit 15% wählt das IQWiG die Mitte des identifizierten Intervalls. Gibt es keine präspezifizierten Responder-Auswertungen, so berücksichtigt das IQWiG die Analysen kontinuierlicher Daten und bewertet die Relevanz des beobachteten Effekts anhand der Lage des Konfidenzintervalles der standardisierten MWD (Hedges'g). Alternativ – und das ist eine Neuerung - können (post-hoc) Analysen mit einem Responsekriterium von genau 15 % des Skalenrangs berücksichtigt werden. Die Festlegung des IQWiG überrascht aus den folgenden Gründen: • Das Vorgehen des IQWiG lässt die ansonsten gewohnte Systematik vermissen. („Sichtung“ von Übersichtsarbeiten, „in vielen Fällen bewegte“ sich die MID zwischen 10% und 20%). Bei den exemplarisch genannten Übersichtsarbeiten handelt es sich um: - o eine Untersuchung zur MID für <i>Patient Reported Outcome Measures</i> (PROM) zur Bewertung der Fatigue; die Arbeit schließt sowohl generische als auch krankheitsspezifische PROMs zur Bewertung der Fatigue ein (Nordin et al. 2016); - o eine Untersuchung der psychometrischen Eigenschaften von Fragebögen zur Bewertung von Symptomen und funktionellen Einschränkungen bei Personen mit Rotatorenmanschettenstörungen (St-Pierre et al. 2016). • Die aufgeführten Beispiele spiegeln das Spektrum der in der frühen Nutzenbewertung bislang eingeschlossenen PROMS nur bedingt wider. • Es bleibt unklar, inwieweit das IQWiG die Aussagesicherheit der Einzelstudien bei seiner Sichtung der Übersichtsarbeiten zur MID bei seiner Festlegung des 15%-Kriteriums berücksichtigt hat. • Entgegen dem sonstigen Vorgehen nimmt das IQWiG keine krankheits- und populationsspezifische Differenzierung bei der Ermittlung/Bewertung der MID vor.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	• Es ist zweifelhaft, inwieweit das vorgeschlagene Bewertungskriterium (mindestens 15% des Skalenranges) den Besonderheiten von Skalen gerecht werden kann. Die Komplexität spezifischer Skalen kann durch ein generisches Kriterium nicht abgebildet werden. • Entgegen aktuellen Empfehlungen, wonach die MID-Bestimmung auf mehreren Ansätzen und einer Triangulation der Methoden basieren sollte (Revicki et al. 2008), legt das IQWiG ein singuläres Kriterium zur Bewertung der MID fest. • Weiterhin findet keine Differenzierung zwischen MIDs für relevante Verbesserungen und Verschlechterungen statt, wenngleich systematische Übersichtsarbeiten nahelegen, dass MIDs für Verschlechterungen tendenziell größer sind als die für Verbesserungen (Nordin et al. 2016). Ausdrücklich gutheißen möchte IQVIA den Vorschlag des IQWiG, in dem Fall, dass keine anerkannte MID existiert, post-hoc Analysen mit einer MID von 15% des Skalenrangs als Responsekriterium für die Bewertung heranzuziehen. Diese Option bietet dem pharmazeutischen Unternehmer Planungssicherheit; insbesondere dann, wenn aufgrund unterschiedlicher Beobachtungsdauern in den Behandlungsarmen Ereigniszeitanalysen für diesen Endpunkt gefordert sind, die eine MID benötigen. <u>Vorgeschlagene Änderung:</u> IQVIA schlägt vor, das bisherige Vorgehen bei der Bewertung der klinischen Relevanz beizubehalten, d.h. Reponderanalysen auf der Basis eines validierten Responsekriteriums heranzuziehen. Für den Fall, dass ein etabliertes Kriterium nicht existiert, bietet der Vorschlag des IQWiG eine pragmatische Lösung und erscheint uns somit akzeptabel.
9.3.4 S. 179	<u>Anmerkung:</u> Das IQWiG gibt an, dass es bei der Bewertung subjektiver Endpunkte im offenem Studiendesign keinen Beleg für einen Nutzen oder Schaden aus statistisch signifikanten Ergebnissen ableiten wird. Die Ergebnisse zu solchen Endpunkten sind potenziell hochgradig verzerrt. Daher müssen, so das IQWiG, Effekte einen bestimmten Abstand zum Nulleffekt aufweisen, um für die Ableitung eines Zusatznutzens herangezogen zu werden. Aus Sicht von IQVIA führt ein solches Vorgehen zu einer „doppelten Bestrafung“. Zum einen wird die Betrachtung einer verschobenen Nullhypothese gefordert und zum anderen führt das offene Studiendesign zu einer Reduktion der Aussagesicherheit.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	Inwieweit sich die Betrachtung verschobener Nullhypothesen auf die Quantifizierung des Effekts und damit die Ausmaßbestimmung des Zusatznutzens auswirkt bleibt unklar. Darüber hinaus erscheint die prospektive Festlegung des konkreten Grenzwerts in einem globalen Studiensetting nicht praktikabel, ebenso wie die Wahl des Grenzwerts durch empirische Daten aus metaepidemiologischer Forschung. <u>Vorgeschlagene Änderung:</u> Keine
3.3.3 S.67	<u>Anmerkung:</u> Das IQWiG legt in Anlehnung an das Verfahren für binäre Daten Schwellenwerte für die Ausmaßabschätzung des Zusatznutzens fest. Insgesamt ist es zu begrüßen, dass das IQWiG jetzt auch bei stetigen und quasistetigen Zielgrößen einen „quantifizierbaren“ Zusatznutzen ableiten kann, wenn jedoch Schwellenwerte im Einzelfall zu diskutieren und in den jeweiligen Kontext zu stellen sein. Das IQWiG nennt die Arbeit von Skipka (Skipka et al. 2016) als Grundlage für die Herleitung der Schwellenwerte. Hierzu ist anzumerken, dass die Arbeit von Skipka von zwei pivotalen Studien ausgeht - somit von einer Situation, die bei der frühen Nutzenbewertung insbesondere im Therapiegebiet onkologischer Erkrankungen kaum vorkommt. <u>Vorgeschlagene Änderung:</u> Keine
9.3.8 /S. 181 ff	<u>Anmerkung</u> Es ist positiv zu bemerken, dass das IQWiG nun präzisiert, wie bei Berechnung von Metaanalysen mit potentiell heterogenen Studien methodisch umzugehen ist (IQWiG 2019). Es wird empfohlen, dass bei einer ausreichenden Zahl an Studien die Durchführung von Metaanalysen mit zufälligen Effekten regelhaft die Knapp-Hartung-Methode (Knapp und Hartung 2003) angewendet werden soll. Da es in homogenen Datensituationen zu irreführend schmalen Konfidenzintervallen kommen kann, sollen immer Sensitivitätsanalysen mithilfe des Modells mit festem Effekt bzw. der DerSimonian-Laird-Methode erfolgen, um zu untersuchen, ob die Anwendung der Varianzkorrektur für die Knapp-Hartung-Methode angezeigt ist (Jackson et al. 2017; Wiksten et al. 2016). Ist das Konfidenzintervall nach der Knapp-Hartung-Methode schmäler als das der DerSimonian-Laird-Methode, so sollte die Knapp-Hartung-Methode mit Varianzkorrektur angewendet werden (IQWiG 2019).

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	Es ist auch hilfreich, dass das IQWiG nun angibt, dass vor allem im Fall von nur 2 Studien das Modell mit festem Effekt gewählt werden soll, wenn keine Gründe dagegensprechen (Bender et al. 2018). Falls dies nicht vertretbar ist, wird eine qualitative Evidenzsynthese empfohlen. Darüber hinaus kann es dann erforderlich sein, mehrere alternative Verfahren kontextabhängig als Sensitivitätsanalysen durchzuführen wie z.B. bayessche Verfahren oder generalisierte lineare Modelle (IQWiG 2019). <u>Vorgeschlagene Änderung:</u> Keine

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

1. Bender R., Friede T., Koch A., Kuss O., Schlattmann P., Schwarzer G. und Skipka, Guido 2018. *Methods for evidence synthesis in the case of very few studies*. Research synthesis methods 9 (3), S. 382–392.
2. Bundesministerium der Justiz und für Verbraucherschutz (BMJV) 2019. *Das Fünfte Buch Sozialgesetzbuch – Gesetzliche Krankenversicherung – (Artikel 1 des Gesetzes vom 20. Dezember 1988, BGBl. I S. 2477, 2482), das zuletzt durch Artikel 1 des Gesetzes vom 21. Dezember 2019 (BGBl. I S. 2913) geändert worden ist*. Verfügbar unter: https://www.gesetze-im-internet.de/sgb_5/SGB_5.pdf, abgerufen am: 31.01.2020.
3. Gemeinsamer Bundesausschuss (G-BA) 2018. *Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V – Bosutinib (neues Anwendungsgebiet: neu diagnostizierte Philadelphia-Chromosompositive chronische myeloische Leukämie in der chronischen Phase): Vom 22. November 2018*. Verfügbar unter: https://www.g-ba.de/downloads/40-268-5449/2018-11-22_AM-RL-XII_Bosutinib_D-355_TrG.pdf, abgerufen am: 31.01.2019.
4. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) 2019. *Allgemeine Methoden: Entwurf für Version 6.0 vom 05.12.2019*. Verfügbar unter: https://www.iqwig.de/download/Allgemeine-Methoden_Entwurf-fuer-Version-6-0.pdf, abgerufen am: 30.01.2020.
5. Jackson D., Law M., Rucker G. und Schwarzer, Guido 2017. *The Hartung-Knapp modification for random-effects meta-analysis: A useful refinement but are there any residual concerns?* Statistics in medicine 36 (25), S. 3923–3934.
6. Knapp G. und Hartung, Joachim 2003. *Improved tests for a random effects meta-regression with a single covariate*. Statistics in medicine 22 (17), S. 2693–2710.

7. Nordin A., Taft C., Lundgren-Nilsson A. und Dencker, Anna 2016. *Minimal important differences for fatigue patient reported outcome measures-a systematic review*. BMC Medical Research Methodology 16 (62), S. 1–16.
8. Revicki D., Hays R. D., Cella D. und Sloan, Jeff 2008. *Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes*. Journal of clinical epidemiology 61 (2), S. 102–109.
9. Revicki D. A., Erickson P. A., Sloan J. A., Dueck A., Guess H. und Santanello, Nancy C. 2007. *Interpreting and reporting results based on patient-reported outcomes*. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research 10 (Suppl 2), S. S116-24.
10. Skipka G., Wieseler B., Kaiser T., Thomas S., Bender R., Windeler J. und Lange, Stefan 2016. *Methodological approach to determine minor, considerable, and major treatment effects in the early benefit assessment of new drugs*. Biometrical journal. Biometrische Zeitschrift 58 (1), S. 43–58.
11. St-Pierre C., Desmeules F., Dionne C. E., Fremont P., MacDermid J. C. und Roy, Jean-Sebastien 2016. *Psychometric properties of self-reported questionnaires for the evaluation of symptoms and functional limitations in individuals with rotator cuff disorders: a systematic review*. Disability and Rehabilitation 38 (2), S. 103–122.
12. Wiksten A., Rucker G. und Schwarzer, Guido 2016. *Hartung-Knapp method is not always conservative compared with fixed-effect meta-analysis*. Statistics in medicine 35 (15), S. 2503–2515.

A.1.24 – Janssen-Cilag GmbH

Autorinnen und Autoren

- Mehnert, Angelika

Stellungnahme Janssen-Cilag GmbH (nachfolgend: Janssen) IQWiG Allgemeine Methoden – Entwurf der Version 6.0

IQWiG Methodenpapier 6.0 Kommentare zu den Änderungen bezüglich Abschnitt 3.3.3 Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V Ergänzung zur Ausmaßbestimmung bei stetigen Daten

Folgende Passage wurde in Version 6.0 ergänzt (S. 67-68):

„D) Stetige oder quasistetige Zielgrößen mit jeweils vorliegenden SMDs

In Anlehnung an das Verfahren für binäre Daten nach Skipka et al. [646] lassen sich auch Schwellenwerte für das Effektmaß SMD (Cohen's d bzw. Hedges' g) ableiten. Man kann zeigen, dass die für ein bestimmtes Ausmaß wünschenswerte Effektstärke ungefähr halbiert werden muss, um den entsprechenden Schwellenwert abzuleiten. Als Orientierung für die theoretisch wünschenswerten Effektstärken bei schwerwiegenden Symptomen können die Mittelwerte der Bereiche für die übliche Einteilung von Cohen's d in kleine (SMD zwischen 0,2 und 0,5) mittlere (SMD zwischen 0,5 und 0,8) und große Effekte (SMD > 0,8) dienen. Tabelle 6 gibt einen Überblick über die verwendeten Schwellenwerte. Im Gegensatz zu den Schwellenwerten für die relativen Maße ist bei der SMD zu beachten, dass das Konfidenzintervall vollständig oberhalb des entsprechenden Schwellenwerts liegen muss und der Nulleffekt hier nicht durch die Zahl 1 sondern durch die Zahl 0 gegeben ist.“

Tabelle 6: Schwellenwerte zur Ausmaßbestimmung für die SMD

		Zielgrößenkategorie	
		schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen und gesundheitsbezogene Lebensqualität ^a	nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen
Ausmaß-kategorie	erheblich	0,6	nicht besetzt
	beträchtlich	0,3	0,5
	gering	0,2	0,2
a: Voraussetzung ist – wie für alle patientenberichteten Endpunkte – die Verwendung eines validierten bzw. etablierten Instruments.			

Kommentar:

Zur Quantifizierung eines Zusatznutzens bei stetigen Variablen wurde vom IQWiG ein neuer Abschnitt in die Version 6.0 des Methodenpapiers eingefügt. Die Ableitung der Schwellenwerte für stetige Variablen wurde in Anlehnung an das Verfahren für binäre Daten nach Skipka et al. vorgenommen.

Janssen begrüßt eine mögliche Quantifizierung der Ausmaßbestimmung stetiger Variablen auf Basis des Hedges' g . Aus Sicht von Janssen sind die Schwellenwerte für die Quantifizierung eines Zusatznutzens bei stetigen Variablen jedoch nicht komplett nachvollziehbar. So sind die Schwellenwerte für nicht schwerwiegende Symptome für kleine (0,2) und mittlere Effekte (0,5) analog zur unteren Grenze des Cohen's d festgesetzt worden, die restlichen Schwellenwerte erscheinen jedoch nicht nachvollziehbar. Es fehlt eine Begründung, warum die für ein bestimmtes Ausmaß wünschenswerte Effektstärke ungefähr halbiert werden muss, um den entsprechenden Schwellenwert abzuleiten.

Für eine Ausmaßbestimmung des Zusatznutzens sollte geeignete wissenschaftlich belegte Literatur verwendet werden, um eine möglichst optimale Einteilung der Schwellenwerte vorzunehmen.

IQWiG Methodenpapier 6.0 Kommentare zu den Änderungen bezüglich Abschnitt 9.3.3 zur Beurteilung der klinischen Relevanz

Folgende Passagen wurden in Version 6.0 ergänzt (S. 175-178):

„Eine Sichtung von systematischen Übersichtsarbeiten zu MIDs [z. B. [523,655]] hat ergeben, dass sich MIDs in vielen Fällen zwischen 10 % und 20 % der Spannweite der Scores eines Erhebungsinstruments bewegen. Daraus ergibt sich regelhaft folgendes Vorgehen:

1) Falls in einer Studie Responderanalysen unter Verwendung einer MID präspezifiziert sind und das Responsekriterium mindestens 15 % der Skalenspannweite des verwendeten Erhebungsinstruments entspricht, werden diese Responderanalysen ohne weitere Prüfung des Responsekriteriums für die Bewertung herangezogen.

2) Falls präspezifiziert Responsekriterien im Sinne einer MID unterhalb von 15 % der Skalenspannweite liegen, werden diese regelhaft nicht herangezogen. In diesen Fällen und solchen, in denen gar keine Responsekriterien präspezifiziert wurden, aber stattdessen Analysen kontinuierlicher Daten zur Verfügung stehen, bestehen verschiedene Möglichkeiten. Entweder können die Analysen der kontinuierlichen Daten herangezogen werden, für die Relevanzbewertung wird dabei auf ein allgemeines statistisches Maß in Form von standardisierten Mittelwertdifferenzen (SMDs, in Form von Hedges' g) zurückgegriffen. Dabei wird eine Irrelevanzschwelle von 0,2 verwendet: Liegt das zum Effektschätzer korrespondierende Konfidenzintervall vollständig oberhalb dieser Irrelevanzschwelle, wird davon ausgegangen, dass die Effektstärke nicht in einem sicher irrelevanten Bereich liegt. Dies soll gewährleisten, dass der Effekt hinreichend sicher mindestens als klein angesehen werden kann [233]. Alternativ können post hoc spezifizierte Analysen mit einem Responsekriterium von genau 15 % des Skalenrangs berücksichtigt werden.

3) Liegen sowohl geeignete Responderanalysen (Responsekriterium präspezifiziert mindestens 15 % des Skalenrangs oder post hoc genau 15 %

des Skalenrangs) als auch Analysen stetiger Daten vor, werden die Responderanalysen herangezogen".

Kommentar:

Bezüglich der Anforderungen zur Beurteilung der klinischen Relevanz im Rahmen einer Anwendung von Responderanalysen wurde vom IQWiG ein neuer Abschnitt in die Version 6.0 des Methodenpapiers aufgenommen. Im bisherigen Methodenpapier 5.0 wurde für die Anwendung von Responderanalysen dargelegt, dass ein „validiertes bzw. etabliertes Responsekriterium“ im Sinne einer Minimal Important Difference [MID] erforderlich ist, um den Nachweis eines klinisch relevanten Effekts vorzunehmen. Das IQWiG hat diese Forderung in seiner neuen Version 6.0 des Methodenpapiers aufgehoben und schlägt stattdessen ein allgemeingültiges Responsekriterium vor, welches mindestens 15 % der theoretisch möglichen Skalenspannweite des verwendeten Erhebungsinstruments entsprechen muss.

Janssen begrüßt, dass sich das IQWiG mit der wichtigen Thematik zur Methode der Ableitung einer klinisch relevanten MID befasst. Die zahlreichen internationalen Initiativen (SISAQOL, IMI uvm.) unterstreichen die Bedeutung dieses Themas. Leider bleiben die bisherigen Ausarbeitungen der genannten Arbeitsgruppen in dem Entwurf des Methodenpapiers unberücksichtigt.

Die vorgeschlagene Responseschwelle von mindestens 15 % ist nicht nachvollziehbar bzw. willkürlich festgelegt und kann nicht als allgemein anerkannter Schwellenwert angesehen werden. Die Ableitung des Schwellenwertes wurde auf Basis einer Sichtung systematischer Übersichtsarbeiten vorgenommen. Es ist unklar, ob die Sichtung vollständig ist, da die vom IQWiG ausgewiesenen Quellen sich lediglich auf eine ausgewählte, spezifische Symptomatik (z. B. Fatigue) und eine spezielle Patientengruppe (rotator cuff disorders) beziehen. Weiterhin wird argumentiert, dass sich die MID *„in vielen Fällen zwischen 10 % und 20 % der Spannweite der Scores eines Erhebungsinstrumentes bewegen“*. Der verwendeten Begrifflichkeit *„in vielen Fällen“* kann keine konkrete Definition entnommen werden, sodass die Frage resultiert, wie repräsentativ die dargelegten Grenzen von 10 % und 20 % sind. Auf Basis dieser Angaben wird der Schwellenwert von 15 % abgeleitet. Eine Angabe zur konkreten Herleitung bzw. eine Begründung findet sich nicht; es ist anzunehmen, dass der vorgeschlagene Schwellenwert den Mittelwert der Angaben 10 % und 20 % darstellt.

Mit dem vom IQWiG vorgeschlagenen Schwellenwert hielte eine nicht wissenschaftlich begründete Responseschwelle für die patientenberichteten Endpunkte in zur Zulassung durchgeführten klinischen Studien Einzug in die Auswertung derselben.

Der vom IQWiG vorgeschlagene Schwellenwert erscheint insgesamt nicht für die frühe Nutzenbewertung geeignet zu sein. Die Bewertung von Arzneimitteln im Rahmen der frühen Nutzenbewertung wird auf Basis aktueller wissenschaftlicher Erkenntnisse sowie den Grundlagen der evidenzbasierten Medizin vorgenommen. Durch eine Berücksichtigung der nicht wissenschaftlich abgeleiteten Responseschwelle von mindestens 15 % würde die bisherige Praxis in der AMNOG-

Bewertung ausgesetzt und alle bisher als validiert und etabliert geltenden MID abgelehnt werden.

Darüber hinaus konterkariert die vorgeschlagene Responseschwelle den bisherigen Gedanken, für verschiedene Therapiegebiete einen spezifischen Schwellenwert zu validieren, um den verschiedenen Indikationen gerecht zu werden.

Des Weiteren ist anzumerken, dass es sich bei einer MID um einen minimalen bedeutsamen klinischen Unterschied handelt. Eine generelle Responseschwelle unabhängig vom Erhebungsinstrument ist mit dem Konzept einer MID nicht in Einklang zu bringen.

Eine Auswertung der bisherigen AMNOG-Verfahren hinsichtlich akzeptierter MID von etablierten Erhebungsinstrumenten (SF-36, EORTC QLQ-Instrumente, EQ-5D-VAS, FACT- und FACIT-Instrumente etc.) hat ergeben, dass durch Anwendung einer Responseschwelle von 15 % insgesamt ca. 90 % der verwendeten MID zu einer Erhöhung des Schwellenwerts kommen würde. Hieraus wird ersichtlich, dass eine generelle Responseschwelle von mindestens 15 % etablierte MID-Effektmaße von im AMNOG-Prozess häufig verwendeten Erhebungsinstrumenten sehr häufig überschätzt, sodass mit der Anwendung des Schwellenwertes ein tatsächlich vorliegender – minimal bedeutsamer – Effekt überdeckt würde.

Durch die Einführung eines für alle Erhebungsinstrumente gleichen Schwellenwertes besteht das Risiko, dass sich ein künstlicher Bedeutungsverlust für die Ergebnisse zu patientenberichteten Endpunkten ergibt und sich damit das Ableiten eines Zusatznutzens erschwert: relevanter Patientennutzen bliebe unerkant. In der Folge würden letztlich auch die Gültigkeit von Schlussfolgerungen zum Vorliegen und Ausmaß eines Zusatznutzens hinsichtlich patientenberichteter Effekte infrage gestellt und damit der allgemeingültige Anspruch, patientenberichtete Endpunkte in der Nutzenbewertung adäquat einzubeziehen, verloren gehen. Diese Tatsache kann weder im Sinne der Bewertungsorgane noch der Patienten sein.

Insgesamt scheint der Vorschlag zur Änderung der Responseschwelle nicht sachgerecht, da die Ableitung einer Responseschwelle von mindestens 15 % nicht auf Basis wissenschaftlicher Erkenntnisse beruht und alle bisher als validiert und etabliert geltenden MID unter dieser Voraussetzung als nicht geeignet für die Nutzenbewertungen eingestuft würden. Janssen schlägt vor, die Thematik im Rahmen eines konstruktiven Dialogs zwischen IQWiG, G-BA und vfa unter Beteiligung von Vertretern der Akademie zu vertiefen, um gemeinsam einen wissenschaftlich anerkannten Kriterienkatalog abzuleiten.

IQWiG Methodenpapier 6.0 **Kommentare zu den Änderungen bezüglich** **Abschnitt 9.3.12 zum Umgang mit unvollständigen Daten**

Folgende Passage wurde in Version 6.0 ergänzt (S. 191-192):

„Bei jeder Nachbeobachtung innerhalb einer Studie kann es unvermeidlich passieren, dass Probandinnen und Probanden für die Datenerhebung zu Endpunkten nicht mehr zur Verfügung stehen. Für diese sogenannten Follow-up-

Verluste kommen verschiedenste Gründe in Betracht, z. B. Tod, Abbruch oder Wechsel der Intervention (arzt- oder patientenseitig), Widerruf der Einverständniserklärung aufgrund von Nebenwirkungen oder ausbleibender Wirksamkeit, neu auftretende Komorbidität oder der belastende Charakter der Nachuntersuchung selbst. Die Standards zum Berichten klinischer Studien sehen vor, dass die Gründe für Follow-up-Verluste – soweit bekannt – für jede Interventionsgruppe getrennt anzugeben sind (siehe Abschnitt 9.1.1). Für eine Intention-to-treat-Analyse wiederum sind möglichst alle randomisierten Patientinnen und Patienten auch in der Analyse zu berücksichtigen [240].

Fehlende Daten erhöhen nicht nur die statistische Unsicherheit der Effektschätzung, sondern führen auch zu einem Verzerrungspotenzial, wenn nicht anhand der Gründe für Follow-up-Verlust klar ersichtlich ist, dass die Verluste („drop-outs“) zufällig erfolgten („missing completely at random“) [460]. Das Verzerrungspotenzial steigt mit steigendem Anteil fehlender Werte. Ergebnisse fließen daher in der Regel nicht in die Nutzenbewertung ein, wenn diese auf weniger als 70 % der in die Auswertung einzuschließenden Patientinnen und Patienten basieren, das heißt, wenn der Anteil der Patientinnen und Patienten, die nicht in der Auswertung berücksichtigt werden, größer als 30 % ist. In der Literatur werden zum Teil bereits Auswertungen, in denen 20 % der Patientinnen und Patienten nicht berücksichtigt werden, als nicht mehr aussagekräftig betrachtet [618].

Die Ergebnisse werden ferner auch dann nicht in die Nutzenbewertung einbezogen, wenn der Unterschied der Anteile nicht berücksichtigter Patientinnen und Patienten zwischen den Gruppen größer als 15 Prozentpunkte ist.

Ausnahmen von diesen Regeln werden dann gemacht, wenn die Follow-up-Verluste nach-weislich zufällig erfolgten oder wenn die möglichen Probleme durch Follow-up-Verlust durch adäquate Ersetzungsstrategien oder statistische Auswertungsmethoden ausgeglichen werden [40,246]. Hierzu ist es erforderlich, den Mechanismus, der zu den fehlenden Daten geführt hat, im Detail zu kennen und beschreiben zu können. Dies wird in der Praxis unmöglich sein, wenn der Anteil fehlender Daten zu hoch ist.“

Kommentar:

Zum Umgang mit unvollständigen Daten wurde vom IQWiG ein neuer Abschnitt in die Version 6.0 des Methodenpapiers eingefügt.

Aus dem Entwurf geht hervor, dass Ergebnisse nicht in die Nutzenbewertung einfließen sollen, wenn sie auf weniger als 70 % der in die Auswertung einzuschließenden Patienten basieren. Hierfür wird [618] zitiert und darauf verwiesen, dass in der Literatur auch strengere Schwellenwerte, in der vorliegenden Publikation 20 %, gefordert werden. Die zitierte Publikation [618] diskutiert verschiedene Szenarien unter welchen Patienten aus der Analyse ausgeschlossen werden und die entsprechenden Auswirkungen auf das Ergebnis. Der vom IQWiG zitierte Schwellenwert von 20 % wird von den ursprünglichen Autoren selbst als „five-and-20 rule of thumb“ bezeichnet und bezieht sich ausschließlich auf Follow-Up Verluste. Die Autoren kommen selbst zu dem Schluss, dass dieser Schwellenwert der Komplexität der Fragestellung nicht gerecht werden würde und, dass eine Betrachtung des Patientenflusses hilfreich ist. Aus Sicht von Janssen wird auch durch einen nicht weiter referenzierten oder hergeleiteten

Schwellenwert von 30 % diese Fragestellung nicht sachgemäß adressiert. Darüber hinaus ist festzuhalten, dass weder die Guidelines der FDA noch der EMA solch einen Schwellenwert für die Nicht-Verwertbarkeit benennen.

Ferner wird im Entwurf des Methodenpapiers Version 6.0 gefordert, Daten nicht zu verwerten, wenn der Anteil, der nicht in die Analyse eingegangenen Patienten sich um 15 Prozentpunkte zwischen den beiden Studienarmen unterscheidet. Auch hier wird die Herleitung des Schwellenwertes weder begründet noch referenziert und erscheint deshalb willkürlich.

Janssen begrüßt die Diskussion um das Thema des Umgangs mit unvollständigen Daten. Sofern aber potenzielle Schwellenwerte definiert werden, sollten diese nachvollziehbar hergeleitet und begründet werden.

IQWiG Methodenpapier 6.0 Kommentare zu den Änderungen bezüglich: **Abschnitt 3.2.1 Auswirkung nicht publizierter Studienergebnisse auf das Fazit**

Folgende Passage wurde in Version 6.0 ergänzt (S. 56-57):

„Die folgende Tabelle 4 beschreibt, in welchen Konstellationen ein Risiko für eine Verzerrung von Bewertungsergebnissen besteht und welche Konsequenzen sich daraus für das Fazit der Nutzenbewertung ergeben.“

Tabelle 4: Szenarien der Datenvollständigkeit und Konsequenzen für das Fazit einer Nutzenbewertung

Szenario	Datenübermittlung durch Dritte (z. B. Daten von Herstellerinnen / Herstellern)	Nachweis, dass insgesamt Daten fehlen	Verzerrung durch Publikationsbias	Bewertung / Einfluss auf das Fazit
1	vollständig	nein oder in vernachlässigbarem Umfang	unwahrscheinlich	keine Einschränkung des Fazits
2	vollständig	ja, in mäßigem Umfang	möglich	Einschränkung der Aussage-sicherheit ^a
3	vollständig	ja, in erheblichem Umfang	wahrscheinlich	kein Beleg (und kein Hinweis oder Anhaltspunkt) für einen Nutzen bzw. Schaden
4	unvollständig	ja oder unklar ^b	wahrscheinlich oder möglich	kein Beleg (und kein Hinweis oder Anhaltspunkt) für einen Nutzen bzw. Schaden
5	keine Datenübermittlung vorgesehen	nein oder in vernachlässigbarem Umfang	unwahrscheinlich	keine Einschränkung des Fazits
6	keine Datenübermittlung vorgesehen	ja, in mäßigem Umfang	möglich	Einschränkung der Aussage-sicherheit ^a
7	keine Datenübermittlung vorgesehen	ja, in erheblichem Umfang	wahrscheinlich	kein Beleg (und kein Hinweis oder Anhaltspunkt) für einen Nutzen bzw. Schaden
a: Bei Nutzenbewertungen mit eigenem Stellungsverfahren kann das Fazit der Bewertung auch unter Vorbehalt gestellt werden. b: Sofern im Ausnahmefall vom Datenübermittelnden nachgewiesen wird, dass die Menge der fehlenden Daten sicher irrelevant ist, wird das Fazit nicht eingeschränkt.				

War in einer Nutzenbewertung mit Datenübermittlung durch Dritte diese Datenübermittlung vollständig und liegt kein Nachweis vor, dass Daten in einem mehr als vernachlässigbaren Umfang fehlen (Szenario 1), wird eine Verzerrung als unwahrscheinlich eingestuft. Die Schlussfolgerungen aus der Bewertung der Daten werden ohne Einschränkung in das Fazit übernommen. Wird bei einer vollständigen Datenübermittlung durch die Herstellerin oder den Hersteller durch andere Recherchen der Nachweis erbracht, dass Daten in einem mäßigen oder erheblichen Umfang (z. B. aus Investigator-initiierten Studien oder Studien anderer Hersteller) fehlen, auf die die Herstellerin oder der Hersteller keinen Zugriff hat, liegt zwar keine selektive Datenübermittlung durch die Herstellerin oder den Hersteller vor, ist aber eine Verzerrung durch die anderweitig fehlenden Daten möglich (Szenario 2) oder sogar wahrscheinlich (Szenario 3). Eine Nutzensaussage im Fazit (im Sinne von Beleg, Hinweis oder Anhaltspunkt für Nutzen oder Schaden) wird deshalb in ihrer Aussagesicherheit entsprechend abgeschwächt oder gar nicht abgeleitet.

Kommentar

Es werden keine Angaben gemacht, nach welchen Kriterien und Maßstäben das Ausmaß möglicher fehlender (unveröffentlichter) Daten festgestellt wird und wie auf die Wahrscheinlichkeit eines Publikationsbias geschlussfolgert wird. Ohne diese Kriterien und Maßstäbe ist eine Einteilung in Kategorien von „mäßigem“ oder „erheblichen“ Umfang fehlender (unpublizierter) Studiendaten sowie von „möglicher“ gegenüber „wahrscheinlicher“ Verzerrung durch Publikationsbias mit dem daraus abgeleiteten Nutzenbewertungsfazit problematisch, denn sie ist nicht nachvollziehbar und es besteht die Gefahr willkürlich wirkender Bewertungen.

A.1.25 – Kassenzahnärztliche Bundesvereinigung K. d ö. R. (KZBV)

Autorinnen und Autoren

- Beck, Jörg

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Beck, Jörg Dr. MHA, Leiter Abteilung Qualitätsinstitut, Leitlinien
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
☒ im Namen folgender Institution / Organisation: Kassenzahnärztliche Bundesvereinigung K. d. ö. R. (KZBV)
☐ als Privatperson(en)

**Stellungnahme der KZBV zum
Entwurf der Allgemeinen Methoden
des IQWiG
Version 6.0, Stand: 05.12.2019**

Das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) stellte seine Allgemeinen Methoden (Version 6.0) am 5. Dezember 2019 öffentlich zur Stellungnahme. Diese Überarbeitung wird aus Sicht der KZBV als sinnvoll erachtet und ist daher sehr zu begrüßen.

Die vorliegende Stellungnahme beinhaltet die Anmerkungen der KZBV zu den Allgemeinen Methoden (Version 6.0). In den Kästchen zu den Anmerkungen sind wortwörtliche Zitate aus den Allgemeinen Methoden und die zugehörigen Seitenangaben enthalten. Die zitierte Literatur ist im Literaturverzeichnis aufgelistet.

Anmerkungen der KZBV

Seite 17:

Tabelle 1: Übersicht über die Produkte des Instituts

Im Vergleich zur Version 5.0 der Allgemeinen Methoden wurden in der Version 6.0, in der Tabelle 1 bei den Produkten „Gesundheitsinformationen“ und „Arbeitspapiere“ unkommentierte Änderungen hinsichtlich der Beauftragungen vorgenommen. Bei den Gesundheitsinformationen wird jetzt ausschließlich der Generalauftrag durch den G-BA benannt, nicht mehr das BMG bzw. die Initiierung durch das Institut. Hingegen werden auf den Seiten 30 und 31 (Abbildung 7), die Initiierung durch das Institut bzw. die Auftragserteilung durch G-BA oder BMG zur Erstellung von Gesundheitsinformationen weiterhin beschrieben.

- Die KZBV regt an, die Inkonsistenzen zu bereinigen und die bisherigen Formulierungen hinsichtlich der Beauftragung bzw. Initiierung von Gesundheitsinformationen aus der Version 5.0 der Allgemeinen Methoden unverändert zu belassen.

Seite 19:

Abbildung 1: Ablauf der Erstellung eines Berichts

In der Abbildung 1 ist der Ablauf der Erstellung eines IQWiG-Berichts dargestellt. Es wäre sehr hilfreich für den Leser, wenn die Zeitangaben zu den Fristen der Stellungnahmen bzw. der Freigabe der Dokumente in der Abbildung mitabgebildet werden. Dies gilt ggf. auch für die Abbildungen anderer IQWiG-Produkte.

- Die KZBV regt an, die Zeitangaben und Fristen zu Stellungnahmen bzw. Veröffentlichung der IQWiG-Produkte in den Abbildungen zu integrieren.

Seite 19:

Abbildung 1: Ablauf der Erstellung eines Berichts

Optional wird eine mündliche wissenschaftliche Erörterung mit Stellungnehmenden durchgeführt. Diese Erörterung dient der gegebenenfalls notwendigen Klärung von Inhalten der schriftlichen Stellungnahmen (Seite 20). Die mündliche wissenschaftliche Erörterung ist in der Abbildung 1 nicht abgebildet (s. auch andere Abbildungen).

- Die KZBV regt an, die mündliche Erörterung in Abbildung 1 sowie ggf. andere Abbildungen aufzunehmen.

Seite 30:

Nach Festlegung der innerhalb eines Themas zu behandelnden Aspekte erfolgt die systematische Informationsbeschaffung, gefolgt von der wissenschaftlichen Bewertung der identifizierten Publikationen. Der Prozess der Themenwahl sowie die Methode der Informationsbeschaffung für Gesundheitsinformationen, der wissenschaftlichen Bewertung und darüber hinaus der Einbeziehung von Betroffenen werden in Kapitel 7 beschrieben.

Die Informationsbeschaffung zur Erstellung von Gesundheitsinformationen wird auf Seite 30 sowie im Kapitel 7 beschrieben. Es ist sicherzustellen, dass neben den eigenen Publikationen auch die gültigen Richtlinien des G-BA, als auch zeitnah zu erwartende Aktualisierungen und Neuerstellungen von Richtlinien berücksichtigt werden.

- Die KZBV regt an, die Informationsbeschaffung zur Erstellung von Gesundheitsinformationen grundsätzlich um relevante Informationen aus anderen Quellen zusätzlich zu den Literaturrecherchen zu erweitern.

Seiten 31 und 142:

Eine Gesundheitsinformation durchläuft außerdem vor der Veröffentlichung – in der Regel zeitgleich mit dem Stellungnahmeverfahren – eine standardisierte externe Nutzertestung. Die während der in der Regel 4-wöchigen Frist eingegangenen Stellungnahmen und das Ergebnis der Nutzertestung werden zusammengefasst und im Hinblick auf einen daraus resultierenden inhaltlichen oder redaktionellen Änderungsbedarf der Gesundheitsinformation geprüft.

In den Allgemeinen Methoden wird der Ablauf der Erstellung einer Gesundheitsinformation beschrieben. Zur Nutzertestung im Rahmen der Erstellung von Gesundheitsinformationen fehlen allerdings Informationen zu den Auswahlkriterien der Nutzer, Zusammensetzung und Repräsentativität sowie Anzahl der Teilnehmer an einem Test. In den Allgemeinen Methoden fehlt außerdem eine Beschreibung des Auswahlprozesses für die Fokusgruppen bzw. Einzelinterviews sowie Gründe für die Auswahl (EBM-Netzwerk 2016).

Außerdem sind ein zeitgleiches Durchführen des Stellungnahmeverfahrens und der Nutzertestung weniger sinnvoll. Die Begutachtung wissenschaftlicher Arbeiten umfasst in erster Linie die Überprüfung fachlicher Richtigkeit der Inhalte (z.B. Vollständigkeit und Korrektheit). Die Verständlichkeit der Arbeit sowie die Sprache werden in einem 2. Schritt geprüft. Unklar ist auch, ob die Nutzer als Coautoren miteinbezogen werden oder ob sie die Gesundheitsinformationen lediglich auf ihre Verständlichkeit bzw. ihren Informationsgehalt prüfen. Informationen zur Überprüfung der Gesundheitsinformationen durch die Nutzer auf ihre Akzeptanz, Relevanz sowie Benutzerfreundlichkeit fehlen. Laut der „Leitlinie evidenzbasierte Gesundheitsinformation“ (Lühnen 2017): „Die Einbeziehung kann zum Zeitpunkt der Planung, Erstellung, Pilotierung und / oder Evaluation in einem unterschiedlichen Ausmaß erfolgen.“. Unklar ist es, warum das IQWiG die Bewertung von Experten und Nutzern gleichzeitig durchführt. Laut Gute Praxis Gesundheitsinformation (EBM-Netzwerk 2016): „Eine unverzichtbare Anforderung ist aber, dass Verfasser und Herausgeber von Gesundheitsinformationen ihr Vorgehen nachvollziehbar darlegen und durch eine transparente Beschreibung der zugrundeliegenden Methoden und Prozesse begründen.“ Unklar ist auch, dass dies „in der Regel“ durchgeführt wird. Kommt es zu Ausnahmen?

- Die KZBV regt an, die Nutzertestung ausführlicher zu beschreiben und die Prozesse transparent darzulegen.
- Die KZBV regt an, das gleichzeitige Durchführen von Stellungnahmen von Experten und Nutzertestung zu begründen und zu erklären, ob es zu Ausnahmen von der dargestellten Praxis kommt.
- Die KZBV sieht die gleichzeitige Durchführung des Stellungnahmeverfahrens und der Nutzertestung als ungeeignet an und schlägt stattdessen folgende Reihenfolge vor:

1. Fachliche Überprüfung (Experten),
2. Nutzertestung,
3. Ggf. Rücksprache mit den Experten (s. hierzu Abbildung 1 in Lühnen 2017).

Seite 32:

Korrekturen, Verbesserungen und Aktualisierungen der veröffentlichten Gesundheitsinformationen werden in der Regel intern vorgenommen. Bei umfassenden oder grundlegenden inhaltlichen Änderungen werden externe Sachverständige eingebunden. In solchen Fällen können auch eine erneute nicht öffentliche Stellungnahme und externe Nutzertestung erfolgen.

Die Aktualisierung veröffentlichter Gesundheitsinformationen erfolgt intern. Bei größeren inhaltlichen Änderungen werden Experten eingebunden. Allerdings können in solchen Fällen eine nicht öffentliche Stellungnahme und externe Nutzertestung erneut durchgeführt werden. Die KZBV erachtet die Einbindung von externen Experten in das Stellungnahmeverfahren auch bei Aktualisierungen als eine *conditio sine qua non*. Das haben die Aktualisierungen der Gesundheitsinformationen zu den epidemiologisch und versorgungsrelevanten „dynamischen“ Themen, Karies und Parodontitis gezeigt. Dies ist bisher nur bei Neuerstellungen von Gesundheitsinformationen vorgesehen.

- Die KZBV regt die verstärkte Einbindung von Experten – auch in der Aktualisierung veröffentlichter Gesundheitsinformationen – an.
- Die KZBV regt ebenfalls an, die verschiedenen Begriffe „Reviewer“, „Sachverständige“ und „Experte“ nicht als Synonyme zu verwenden, da es beim Leser zu Unklarheiten führen kann.

Seiten 34 und 35:

Der Ablauf der Erstellung der HTA-Berichte gemäß § 139b Abs. 5 SGB V ist in Abbildung 9 schematisch dargestellt. Der HTA-Bericht wird von den externen Sachverständigen (siehe Abschnitt 2.2.2) unter Anwendung der Methodik des IQWiG erstellt und vom IQWiG durch einen Herausgeberkommentar ergänzt.

Zu Beginn des Kapitels 2.1.10 wurde gegenüber der Version 5.0 der Allgemeinen Methoden das „eigene Fazit“ gestrichen. Es ist nicht immer klar ersichtlich, ob es sich hierbei um ein „eigenes“ Fazit des IQWiG oder der externen Sachverständigen handelt.

- Die KZBV regt an, das Fazit sprachlich eindeutiger zuzuordnen.

Seite 35:

Dem HTA-Bericht wird ein Herausgeberkommentar vorangestellt, in dem eine Einordnung der Ergebnisse durch das Institut erfolgt. Zusätzlich wird HTA kompakt, eine allgemein verständliche Zusammenfassung des HTA-Berichts, erstellt.

Der HTA-Bericht mit ergänzendem Herausgeberkommentar, die Dokumentation der Anhörung zum vorläufigen HTA-Bericht und HTA kompakt werden zunächst dem Vorstand der Stiftung sowie dem Stiftungsrat und dem Kuratorium der Stiftung zugeleitet. In der Regel weitere 4 Wochen später erfolgt die Veröffentlichung dieser Dokumente auf der Website des Instituts.

Zusätzlich zu dem HTA-Bericht und dem Herausgeberkommentar soll auch eine allgemein verständliche Zusammenfassung des Berichts „HTA kompakt“ erstellt werden. Unklar ist, ob HTA kompakt ein neues Produkt des IQWiG ist und ob das HTA kompakt die Patienten als Zielgruppe hat. In der Abbildung 9 sind die Schritte der Erstellung der HTA-Berichte beschrieben u.a. (vorläufiger) HTA-Bericht und Herausgeberkommentar. Der HTA kompakt fehlt in dieser Abbildung; ähnlich wie Begleitinformation in Abbildung 7.

- Die KZBV regt an, zu erklären, ob das HTA kompakt ein neues Produkt des IQWiG ist, ggf. Aufnahme in die Tabelle 1 (Übersicht über die Produkte des Instituts). Zu klären wäre auch, ob das HTA kompakt eine Information für Patienten ist.
- Die KZBV regt an, die Abbildung 9 um „HTA kompakt“ zu ergänzen.

Seite 35

Der HTA-Bericht mit ergänzendem Herausgeberkommentar, die Dokumentation der Anhörung zum vorläufigen HTA-Bericht und HTA kompakt werden zunächst dem Vorstand der Stiftung sowie dem Stiftungsrat und dem Kuratorium der Stiftung zugeleitet. In der Regel weitere 4 Wochen später erfolgt die Veröffentlichung dieser Dokumente auf der Website des Instituts.

Auf der Website zu *ThemenCheck Medizin* wird aufgeführt: „Die HTA-Berichte leitet das IQWiG an Institutionen in Deutschland weiter, die etwa über Leistungen und Struktur des Gesundheitswesens entscheiden, z. B. den Gemeinsamen Bundesausschuss (G-BA) und das Bundesministerium für Gesundheit (BMG). Auf diese Weise sollen die Ergebnisse von HTA-Berichten direkte Auswirkungen auf die Patientenversorgung in Deutschland haben, z. B. bei Behandlungsentscheidungen im Gespräch von Patient und Arzt oder auch bei Systementscheidungen von Selbstverwaltung und Politik“. Diese Regelung sollte auch in die Allgemeinen Methoden (Version 6.0) aufgenommen werden; zusätzlich zur Veröffentlichung auf der Webseite des Instituts.

- Die KZBV regt an, die Weiterleitung der HTA-Berichte an wichtige Institutionen im Gesundheitswesen in die Allgemeinen Methoden aufzunehmen.

Seiten 40 - 41:

Optional wird ergänzend eine wissenschaftliche Erörterung mit Stellungnehmenden durchgeführt mit dem Ziel, Inhalte schriftlicher Stellungnahmen zu klären. Auf Anfrage können Mitarbeiterinnen und Mitarbeiter der Geschäftsstelle des G-BA an der Erörterung teilnehmen.

Eine Ergänzung im Vergleich zur Version 5.0 der Allgemeinen Methoden ist, dass Mitarbeiter der Geschäftsstelle des G-BA an der wissenschaftlichen Erörterung teilnehmen können. Unklar ist es, welche Rolle sie dabei einnehmen.

- Die KZBV regt an, die Gründe für die geplante Teilnahme zu erläutern.
- Die KZBV regt an, die Rolle der Mitarbeiter der Geschäftsstelle des G-BA zu konkretisieren, ob sie stille Beobachter oder aktive Teilnehmer an der Erörterung sein sollen.

Seite 67:

Stetige oder quasistetige Zielgrößen mit jeweils vorliegenden SMDs

In Anlehnung an das Verfahren für binäre Daten nach Skipka et al. [646] lassen sich auch Schwellenwerte für das Effektmaß SMD (Cohen's d bzw. Hedges' g) ableiten. Man kann zeigen, dass die für ein bestimmtes Ausmaß wünschenswerte Effektstärke ungefähr halbiert werden muss, um den entsprechenden Schwellenwert abzuleiten.

Die neuen Ausführungen zum Standardized Mean Difference (SMD) beschreiben die Schwellenwerte für stetige oder quasistetige Zielgrößen. Diese Ergänzung erachtet die KZBV als sinnvoll. Die Untergrenze ist 0,2 (Tabelle 6). Es ist allerdings unklar, wie die Schwellenwerte 0,3, 0,5 sowie 0,6 bestimmt wurden.

- Die KZBV regt an, zu klären, wie die Schwellenwerte in der Tabelle 6 definiert wurden.

Seite 68:

Die Grundsätze von evidenzbasierter Medizin und Nutzenbewertung sind daher auch für spezielle Felder der Medizin gültig, so beispielsweise für die Chirurgie, die Zahnheilkunde, die Psychotherapie, die Alternativ- / Komplementärmedizin oder auch für digitale Gesundheitsanwendungen („E-Health“).

In den Allgemeinen Methoden werden ausschließlich randomisierte kontrollierte Studien (RCTs) als relevante wissenschaftliche Quelle für die Nutzenbewertung betrachtet, wenn dies theoretisch realisierbar erscheint. Als Begründung wird angeführt, dass **gut durchgeführte** RCTs mit der geringsten Ergebnisunsicherheit behaftet sind. In dem Abschnitt auf Seite 68 der Allgemeinen Methoden werden explizit bestimmte Felder der Medizin im Hinblick auf der Umsetzung der EbM-Kriterien hervorgehoben, wie z.B. die Zahnheilkunde.

In der die Zahnheilkunde existieren Studien auf dem höchsten Evidenzniveau (RCT-Niveau). Für bestimmte Fragestellungen allerdings nicht, da z.B. für eine besonders hohe Evidenz placebo-kontrollierte Langzeitstudien sinnvoll wären. Dass eine RCT-Studie zuverlässige Ergebnisse für die Nutzenbewertung einer Therapie liefern kann, ist unumstritten; vorausgesetzt ist eine **gute methodische Qualität**. Hier sind einige Aspekte hervorzuheben. In der Zahnmedizin existieren gut durchgeführte Kohortenstudien – das im Kontext nichtmedikamentöser Verfahren oft maximal zu realisierende Studiendesign – die zuverlässige Ergebnisse für die Nutzenbewertung abbilden können. Die Betonung der Bedeutung **methodischer Qualität** bei der Durchführung von Kohortenstudien kann dazu führen, dass die Qualität dieser Studien verbessert wird. Die Bewertung und Betrachtung der Qualität wissenschaftlicher Literatur der letzten 40 Jahre in bestimmten Feldern der Zahnmedizin zeigte einen Qualitätssprung (u.a. hinsichtlich der Reporting quality), insbesondere bei den Kohortenstudien auf. Das bestätigt auch der Cochrane Review von Anglemyer 2014.

Jedes Studiendesign verfügt über Vor- und Nachteile. RCTs können aufgrund ihrer Einschränkungen (Logistik, Dauer, Kosten, Repräsentativität der Studienteilnehmer sowie der Behandler und soziale Konsequenzen der Studie) Fragen zu vielen Themen in der Medizin nicht beantworten. In manchen Fällen sind RCT nicht durchführbar bzw. nicht möglich (Black 1996). Auch die Anwendung bestimmter (patienten) relevanter Endpunkte in RCTs ist nicht immer möglich (true endpoints) (Hujoel 2004). Deshalb ist es sinnvoll zu überlegen, was andere, **verfügbare** Studien an (zusätzliche) Informationen liefern können, z.B. Untersuchung der Bereiche, die durch ein RCT nicht abgedeckt werden können und Lieferung von inhaltlichen bzw. methodischen Erklärungen bei der Interpretation der Ergebnisse von RCTs.

Hier ist zu betonen, dass die Grundsätze der Evidenzbasierten Medizin die „**bestverfügbare**“ und nicht die theoretisch „bestmögliche“ Evidenz zur Bewertung heranzuziehen sind. Daraus resultiert einerseits, dass zu manchen Fragestellungen „kein Anhaltspunkt für einen (Zusatz-)Nutzen oder (höheren) Schaden“ einer Intervention

aufgefunden werden kann, andererseits widerspricht ein solches Vorgehen den Vorgaben der Verfahrensordnung des G-BA (VerfO). Der hier geforderte höhere Begründungsaufwand, wenn von der Evidenzstufe I abgewichen wird, rechtfertigt nach Einschätzung der KZBV nicht, nichtmedikamentöse therapeutische Interventionen ausschließlich auf Basis der Evidenzstufe I zu betrachten.

- Die KZBV regt an, diese Besonderheiten in die Allgemeinen Methoden aufzunehmen. Gerne darf der vorstehende Text verwendet werden.

Seite 69:

Studien im nicht medikamentösen Bereich sind im Vergleich zu Arzneimittelstudien häufig mit besonderen Herausforderungen und Schwierigkeiten verbunden [490], die aber meist zumindest teilweise überwindbar sind:

- Oft wird die Verblindung des die Intervention ausführenden Personals unmöglich und die der Patientinnen und Patienten schwierig oder ebenfalls unmöglich sein. Durch eine verblindete Endpunkterfassung oder durch die Verwendung objektiver Endpunkte lässt sich eine Verringerung der Aussagekraft einer Studie dennoch abmildern.

Das IQWiG schildert, dass Studien im nichtmedikamentösen Bereich häufig mit besonderen Herausforderungen und Schwierigkeiten verbunden sind und schlägt deshalb die Verwendung objektiver Endpunkte vor. Objektive Endpunkte werden erst auf Seite 166 definiert „Falls eine verblindete Zielgrößenerhebung nicht möglich ist, sollte ein möglichst objektiver Endpunkt gewählt werden, der in seiner Ausprägung und in der Stringenz der Erfassung so wenig wie möglich durch diejenige Person, die den Endpunkt (unverblindet) erhebt, beeinflusst werden kann.“. Unklar ist es, ob objektive Endpunkte immer möglich und patientenrelevant sind. Konkrete Beispiele für objektive Endpunkte werden in den Allgemeinen Methoden nicht aufgeführt (Seite 166). Außerdem ist eine Verblindung bei der Datenaufbereitung bzw. Analyse ebenfalls wesentlich für die Zuverlässigkeit der Ergebnisse. Dieser Aspekt könnte noch ergänzt werden.

In der Zahnmedizin werden außerdem zumeist Studien im Split-Mouth-Design, bei denen die Kontrolle und Intervention in **einer** Mundhöhle verglichen werden, durchgeführt. Viele dieser relevanten und qualitativ hochwertigen Studien werden jedoch mit dem Argument ausgeschlossen, dass die Abhängigkeiten der erhobenen Daten unklar sind. Dies ist nicht nachvollziehbar, da gerade dieses Studiendesign interindividuelle Variabilitäten z.B. durch unterschiedliches Mundhygieneverhalten der Patienten ausschließt und somit Verzerrungen minimiert werden. Bekannte und unbekannte personengebundene Störgrößen werden quasi gleichmäßig auf „Experimental- und Kontrollgruppen“ verteilt.

- Die KZBV regt an, konkrete Beispiele für objektive Endpunkte zu benennen und zu erklären, ob diese patientenrelevant sind.
- Die KZBV regt an, das Split-Mouth-Design als Evidenzquelle zu berücksichtigen.

Seite 128:

Die vom Auswahlbeirat in der 1. Stufe des Auswahlverfahrens nominierten Themen werden in der 2. Stufe mit dem erweiterten Fachbeirat **diskutiert**. Dieser ist mit Vertreterinnen und Vertretern der den Stiftungsrat der Stiftung für Qualität und Wirtschaftlichkeit bildenden Organisationen sowie einer Vertreterin beziehungsweise einem Vertreter des BMG besetzt. Anschließend wählt das **Institut** aus den vom Auswahlbeirat nominierten Themen bis zu 5 aus, für die HTA-Berichte erstellt werden.

Die Rolle des Fachbeirats wird hier reduziert. Die Themenauswahl erfolgte bislang im Benehmen mit dem Fachbeirat. Nun wird der Fachbeirat nur noch in die Diskussion eingebunden (2. Stufe). Die Einschränkung der Rolle des Fachbeirates ist aus Sicht der KZBV nicht nachvollziehbar und wird nicht begründet. In der Abbildung 16 (Seite 126) steht weiterhin „Auswahl von bis zu 5 Themen, zu denen HTA-Berichte erstellt werden, im **Benehmen** mit dem erweiterten Fachbeirat“. Die KZBV verweist auf die unterschiedlichen Angaben und bittet um eine Begründung für diese Änderung. Des Weiteren wäre zu klären, ob eine Aufgabenübertragung hinsichtlich endgültiger Wahl der Themen im IQWiG stattgefunden hat, im Vergleich zu Version 5.0 der Allgemeinen Methoden (Institutsleiter → Institut): Wer im Institut übernimmt diese Aufgabe? Ist das immer eine Person oder sind es wechselnde Personen bzw. Abteilungen?

- Die KZBV regt an, die o.g. Aspekte zu erläutern.

Seite 128:

Umgang mit nominierten aber nicht für eine HTA-Erstellung ausgewählten Themen

Die vom Auswahlbeirat nominierten aber vom **Institut** nicht für eine HTA-Erstellung ausgewählten Themen gehen einmalig erneut in das nächste Auswahlverfahren im Folgejahr ein. Sie durchlaufen dort erneut die in den Abschnitten 6.3.3 und 6.3.4 beschriebenen Auswahlstufen.

In diesem Abschnitt wird der Umgang mit den nominierten aber nicht für eine HTA-Erstellung ausgewählten Themen beschrieben. Auch hier wählt nun das **Institut** – und nicht wie bislang der Institutsleiter – die Themen für die Erstellung der HTA-Berichte aus. Die KZBV bittet zu erklären, ob eine Aufgabenübertragung hinsichtlich endgültiger Wahl

der Themen im IQWiG stattgefunden hat (vgl. Allgemeine Methoden Version 5.0: Institutsleiter → Institut). Wer im Institut übernimmt diese Aufgabe? Ist das immer eine Person oder wechselnde Personen bzw. Abteilungen?

- Die KZBV regt an, zu klären, ob eine Aufgabenübertragung hinsichtlich Wahl der HTA-Themen stattgefunden hat und wer im Institut diese Aufgabe übernimmt.

Seite 142:

Nach Ablauf der Stellungnahmefrist sichten die Mitglieder des an der Erstellung beteiligten Teams und ein Mitglied der Ressortleitung die in den Stellungnahmen vorgebrachten Argumente hinsichtlich ihrer Relevanz für die Gesundheitsinformation. Im Rahmen einer Stellungnahmeredaktionskonferenz werden die Kommentare und Argumente der Stellungnehmenden diskutiert und soweit vorhanden, die entsprechende Evidenz bewertet. Ein möglicher Änderungsbedarf wird konsentiert und dokumentiert. Zeitnah zur Veröffentlichung des finalen Textes werden die Stellungnahmen gewürdigt.

Die Würdigung der schriftlichen Stellungnahmen wird den Stellungnehmenden parallel zur Veröffentlichung bzw. nach der Veröffentlichung der Gesundheitsinformationen übermittelt. Die Würdigung durch das IQWiG sollte in Interaktion erfolgen und nicht erst nach Abschluss und anschließender Veröffentlichung der Arbeit. Wenn ein weiterer Überarbeitungsbedarf besteht, sollte eine zweite Rückkopplungsschleife erfolgen, d.h. eine erneute Überarbeitung und Kommentierung der Gesundheitsinformationen (s. hierzu Abbildung 1 in Lühnen 2017). Außerdem tragen Rückmeldungen beteiligter Organisationen und Fachgesellschaften sowie die konstruktive Zusammenarbeit mit diesen zur fachlichen Qualitätssteigerung der Gesundheitsinformationen bei.

- Die KZBV regt an, die Würdigung der Stellungnahmen **vor** der Veröffentlichung der Gesundheitsinformationen an die beteiligten Organisationen zurückzumelden; ggf. ist eine zweite Kommentierung beteiligter Organisationen notwendig.
- Die KZBV regt eine Transparente Darlegung der beteiligten Organisationen und Fachgesellschaften an, z.B. durch einen Hinweis auf den Gesundheitsinformationen („... erstellt mit fachlicher Unterstützung der KZBV ...“).

Literatur

Anglemyer, A.; Horvath, H. T.; Bero, L. (2014): Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials 4. DOI: 10.1002/14651858.MR000034.pub2.

Black, N. (1996): Why we need observational studies to evaluate the effectiveness of health care. In: BMJ (Clinical research ed.) 312 (7040), S. 1215–1218. DOI: 10.1136/bmj.312.7040.1215.

Deutsches Netzwerk Evidenzbasierte Medizin e.V. (2016): GUTE PRAXIS GESUNDHEITSINFORMATION. Ein Positionspapier des Deutschen Netzwerks Evidenzbasierte Medizin e.V. Hg. v. Deutsches Netzwerk Evidenzbasierte Medizin e.V.

Hujoel, P. P. (2004): Endpoints in periodontal trials: the need for an evidence-based research approach. Periodontology 2000. 36, 196–204.

Lühnen, J.; Albrecht, M.; Mühlhauser, I.; Steckelberg, A. (2017): Leitlinie evidenzbasierte Gesundheitsinformation. Evidenzbasierte Leitlinie Version: 1.0. Hg. v. Deutsches Netzwerk Evidenzbasierte Medizin e.V. Hamburg.

A.1.26 – Lilly Deutschland GmbH

Autorinnen und Autoren

- Lehmann, Tanja
- Nicolay, Claudia
- Rämsch, Christiane
- Rößler, Tatjana
- Wagner, Uwe
- Zimmermann, Thomas Markus
- Zschocke, Jürgen

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Lehmann, Tanja; Dr.
Nicolay, Claudia; Dr.
Rämsch, Christiane
Rößler, Tatjana; Dr.
Wagner, Uwe; Dr.
Zimmermann, Thomas Markus; Dr.
Zschocke, Jürgen; Dr.
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
☒ im Namen folgender Institution / Organisation: Lilly Deutschland GmbH
☐ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
3.2.1 Auswirkung nicht publizierter Studienergebnisse auf das Fazit (S. 55ff)	<u>Anmerkung:</u> Tabelle 4 fasst die Szenarien der Datenvollständigkeit und Konsequenzen für das Fazit einer Nutzenbewertung zusammen. Szenarien 2 und 3 beschreiben dabei den Fall, dass Daten fehlen, auf die der Hersteller keinen Zugriff hat, wodurch eine Verzerrung möglich bzw. wahrscheinlich ist. In so einem Fall wird „eine Nutzensaussage im Fazit (im Sinne von Beleg, Hinweis oder Anhaltspunkt für Nutzen oder Schaden) [...] deshalb in ihrer Aussagesicherheit entsprechend abgeschwächt oder gar nicht abgeleitet.“ In diesen beiden Szenarien hat der Hersteller alle Daten seinerseits vollständig eingereicht, aber keinen Zugriff auf möglicherweise fehlende Daten. Trotzdem wird u.U. kein Zusatznutzen abgeleitet, obwohl die vorhandenen Daten womöglich einen Zusatznutzen zeigen, weil Daten, auf die von Herstellerseite kein Zugriff möglich ist, einen Einfluss haben könnten. Es ist unklar, wie dieses Paradoxon im Rahmen einer Nutzenbewertung aufzulösen ist. Zudem ist unklar, anhand welcher Kriterien der Umfang der fehlenden Daten als „mäßig“ bzw. „erheblich“ eingestuft wird. <u>Vorgeschlagene Änderung:</u> Zum einen sollte konkretisiert werden, was es bedeutet, wenn Daten im mäßigen bzw. erheblichen Umfang fehlen.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	Des Weiteren sollte der Einfluss auf das Fazit im Falle fehlender Daten, auf die der Hersteller keinen Zugriff hat, entschärft werden. Wenn eine Verzerrung durch Publikationsbias wahrscheinlich ist, dann ist eine Abschwächung der Aussagesicherheit eine mögliche Vorgehensweise, diese Verzerrung in der Nutzenbewertung zu berücksichtigen. Es sollte davon abgesehen werden, keinen Nutzen abzuleiten, wenn von Herstellerseite alle Daten vollständig eingereicht wurden und auf die fehlenden Daten kein Zugriff besteht.
3.2.1 Auswirkung nicht publizierter Studienergebnisse auf das Fazit (S. 57)	<u>Anmerkung:</u> Es herrscht Unklarheit bezüglich der folgenden Aussage: "In bestimmten Fällen (z. B. je nach Datenübermittlungsverfahren und Datenanteil) kann es sinnvoll sein, einen Teil des Studienpools (z. B. sämtliche Studien, für die ein Hersteller als Sponsor verantwortlich ist) für die Bewertung nicht zu berücksichtigen, sofern diese Daten durch Publikationsbias gefährdet sind." Zum einen ist zu klären, wann eine Gefährdung durch Publikationsbias vorliegt bzw. auf Basis welcher Kriterien diese Gefährdung bestimmt wird. Auf der anderen Seite ist nicht nachvollziehbar, welchen Einfluss der Publikationsbias auf Studien des Herstellers haben sollte, da diese auch für die Nutzenbewertung herangezogen werden können, wenn noch keine Publikation vorliegt. <u>Vorgeschlagene Änderung:</u> Eine Konkretisierung der oben genannten Aussage, insbesondere zur Einschätzung der Gefährdung durch Publikationsbias und die Ergänzung weiterer, eindeutiger Beispiele zum besseren Verständnis, wäre wünschenswert.
3.3.3 Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V / D) Stetige oder quasistetige Zielgrößen mit jeweils vorliegenden SMDs (S. 67)	<u>Anmerkung:</u> Prinzipiell ist eine Klassifizierung nach Zusatznutzen-Kategorie für stetige Variablen analog der binären Variablen zu begrüßen, allerdings ist die Wahl der Schwellenwerte in Tabelle 6 nicht hinreichend begründet und entsprechende Quellen bzw. eine ausführliche Herleitung fehlen. Des Weiteren ist anzumerken, dass es sich bei der zugrunde gelegten Methode der Verschiebung der Nullhypothese um eine sehr konservative Methode handelt, zumal die von Cohen ursprünglich eingeführte Einteilung der Effekte nach Stärke den Punktschätzer betreffen und nicht den Intervallschätzer (1).

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	<u>Vorgeschlagene Änderung:</u> Es wäre wünschenswert, die entsprechende Herleitung der in Tabelle 6 präsentierten Schwellenwerte transparent darzustellen und entsprechend mit Quellen zu belegen. Es ist dabei zudem zu berücksichtigen, dass die von Cohen eingeführten Grenzwerte lediglich eine Klassifizierung des Punktschätzers darstellen und sich nicht auf den Intervallschätzer beziehen.
9.3.3 Beurteilung klinischer Relevanz (S. 175 ff)	<u>Anmerkung:</u> Dass in Responderanalysen im Rahmen der Nutzenbewertung geeignete Responseschwellen eingesetzt werden sollten, steht außer Frage. Allerdings ist es aus unserer Sicht nicht zielführend, ohne Berücksichtigung bisher etablierter MIDs, die schon in Nutzenbewertungsverfahren als Responsekriterien genutzt und auch vom G-BA sowie der Zulassungsbehörde akzeptiert wurden, Responseschwellen lediglich auf Basis zweier Übersichtsarbeiten ausnahmslos auf 15 % der Spannweite des Scores festzulegen. Es stellt sich die Frage, inwieweit dieser generische Schwellenwert mit höherer Aussagekraft eine minimale klinische Relevanz anzeigt als ein Schwellenwert, der im Rahmen einer Studie mit dem Ziel, eine MID zu validieren, ermittelt wurde (und zwar basierend auf der relevanten Patientenpopulation). Zur Herleitung der 15%-Schwelle bezieht sich das IQWiG auf „eine Sichtung von systematischen Übersichtsarbeiten“. Die Sichtung ist nicht dokumentiert und so bleibt unklar, an welchen Kriterien sich diese orientiert und ob diese vollständig ist. Die beiden vom IQWiG beispielhaft angegebenen Quellen gehen zudem lediglich auf ausgewählte und sehr spezifische Symptome wie „<i>fatigue</i>“ und „<i>symptoms and functional limitations in individuals with rotator cuff disorders</i>“ ein. Das vorgeschlagene Vorgehen betrifft auch den in der Onkologie regelhaft verwendeten Lebensqualitätsfragebogen EORTC QLQ-C30. Dieser weist eine Spannweite von 0 bis 100 Punkten auf, das etablierte und bisher in Nutzenbewertungsverfahren genutzte und akzeptierte Responsekriterium liegt bei 10 Punkten. Nach dem Vorschlag in den Allgemeinen Methoden 6.0. würde hier nun statt eines Schwellenwertes von 10 Punkten plötzlich ein Schwellenwert von 15 Punkten gelten.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	Des Weiteren bildet ein generisches Kriterium von 15 % der Spannweite des Scores weder Einflüsse der Krankheitsschwere und der Patientencharakteristika noch die Indikation ab. Es stellt sich daher die Frage, ob eine verteilungsbasierte und pauschale Schwelle für die MIDs aller Erhebungsinstrumente überhaupt sinnvoll sein kann. Auch in der vom IQWiG zitierten Arbeit von Revicki et al. (3) heißt es dazu: <i>„The MID for a PRO instrument is not an immutable characteristic, but may vary by population and context, and no one MID may be valid for all study applications. MID estimates should be based on multiple approaches and triangulation of methods. Anchor-based methods applying various relevant patient-rated, clinician-rated, and disease-specific variables provide primary and meaningful estimates of an instrument's MID.“</i> <u>Vorgeschlagene Änderung:</u> Das im Methodenpapier vorgeschlagene Vorgehen sollte so angepasst werden, das etablierte und bisher in Nutzenbewertungsverfahren genutzte und vom G-BA und der Zulassungsbehörde akzeptierte Responsekriterien weiterhin in der Nutzenbewertung Anwendung finden. Dies dient auch der Gewährleistung der Vergleichbarkeit vergangener mit zukünftigen Nutzenbewertungen. Zudem sollte beim Vorliegen von Validierungsstudien, die die Mindestqualitätsanforderungen für eine Validierung erfüllen (beispielsweise nach Kriterien der FDA-Guidance (2)), den darüber ermittelten MIDs Vorrang bei der Nutzung als Responsekriterium gegeben werden. Revicki et al. (3) sprechen sich ebenfalls dafür aus: <i>„We recommend that the MID is based primarily on relevant patient-based and clinical anchors, with clinical trial experience used to further inform understanding of MID.“</i> Liegt keinerlei Evidenz für eine anwendbare MID vor, dann eignet sich die vorgeschlagene Methodik, diese auf 15% der Spannweite des Scores festzulegen, um trotzdem Responderanalysen durchführen zu können. Anzustreben wäre hier ein konstruktiver Dialog zwischen IQWiG, G-BA, Akademie und dem vfa, um gemeinsam einen wissenschaftlich fundierten Kriterienkatalog zu etablieren, der für MIDs im Rahmen von Nutzenbewertungen regelhaft anzuwenden ist.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
9.3.12 Umgang mit unvollständigen Daten (S. 191f)	<u>Anmerkung:</u> Besonders bei ereignisgetriebenen Studien mit Zensierung in der Onkologie, die zum Teil sehr lange laufen, ist mit "unvollständigen Daten" durch Tod, Studienabbruch u.a. zu rechnen. Dass ab einem bestimmten Zeitpunkt nur noch weniger als 70 % der Patienten in der Analysepopulation enthalten sind, ist unter diesen Umständen wahrscheinlich. Ein Vorgehen für solche Studien ist aus dem Abschnitt 9.3.12 nicht unbedingt abzuleiten. Für die folgende Aussage „Die Ergebnisse werden ferner auch dann nicht in die Nutzenbewertung einbezogen, wenn der Unterschied der Anteile nicht berücksichtigter Patientinnen und Patienten zwischen den Gruppen größer als 15 Prozentpunkte ist“ fehlt zudem eine Quelle, welche diesen Schwellenwert erklärt bzw. belegt. <u>Vorgeschlagene Änderung:</u> Eine Konkretisierung zum Umgang mit unvollständigen Daten insbesondere bei ereignisgetriebenen Studien mit Zensierung wäre wünschenswert.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

- (1) Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed). Hillsdale, N.J: L. Erlbaum Associates.
- (2) FDA. Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Guidance for Industry. December 2009. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims>

- (3) Dennis Revicki, Ron D. Hays, David Cella, Jeff Sloan (2008). Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology*. Volume 61, Issue 2, 102-109.

**A.1.27 – Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen e. V.
(MDS)**

Autorinnen und Autoren

- Eikermann, Michaela
- Eyding, Dirk
- Gehrman, Ulrich
- Herrmann-Frank, Annegret
- Janatzek, Sandra
- Lipperheide, Kerstin

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Eikermann, Michaela; Dr.
Gehrmann, Ulrich; Dipl. Stat.
Eyding, Dirk; Dr.
Herrmann-Frank, Annegret; PD Dr.
Janatzek, Sandra; Dr.
Lipperheide, Kerstin; Dr.
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen) ☒ im Namen folgender Institution / Organisation: Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen e. V. (MDS) ☐ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
3.3.3 (S. 65 oben)	<u>Anmerkung:</u> In der Operationalisierung der Feststellung des Ausmaßes des Zusatznutzens heißt es: „Entscheidend ist also, dass die obere Grenze des Konfidenzintervalls kleiner als der jeweilige Schwellenwert ist.“ Dies ist insbesondere für binäre Zielgrößen zutreffend. Für stetige Zielgrößen ist die untere Grenze entscheidend, wie in Unterpunkt D (S. 67) beschrieben. <u>Vorgeschlagene Änderung:</u> Es wird vorgeschlagen, an dieser Stelle eine allgemeinere Formulierung, die für alle in den Abschnitten A) bis E) genannten Skalen gültig ist, zu wählen.
3.3.3, Unterpunkt D (S. 67)	<u>Anmerkung:</u> Die Herleitung der Tabelle 6 (Schwellenwerte zur Ausmaßbestimmung für die SMD) ist für uns nicht nachvollziehbar. Es wird beschrieben, dass die Herleitung angelehnt an das Verfahren von Skipka et al. (2016) für binäre Daten erfolgte. Als Orientierung diene die übliche Einteilung von Cohen's d. Daraus wird allerdings nicht klar, welche Kriterien analog zu Table 1 aus Skipka et al. (2016) hier verwendet wurden, um die Schwellenwerte zu bestimmen. Des Weiteren würde man eine Strukturgleichheit der Tabellen 5 und 6 bezüglich der Werte in den Spalten der schwerwiegenden und nicht-schwerwiegenden Symptome erwarten. In Tabelle 5 (bezüglich RR) unterscheiden sich die beiden Schwellenwerte für geringes Ausmaß (1,00 und 0,90), in Tabelle 6 jedoch nicht (0,2 in beiden

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	Spalten). Insbesondere fällt auf, dass der Nulleffekt bei schwerwiegenden Symptomen für RR als geeigneter Schwellenwert für geringes Ausmaß angesehen wird, für SMD jedoch nicht. Eine weitere Diskrepanz liegt darin, dass die Werte für beträchtlich/schwerwiegend und gering/nicht schwerwiegend in Tabelle 5 identisch sind (0,9), in Tabelle 6 jedoch nicht (0,3 und 0,2). <u>Vorgeschlagene Änderung:</u> Eine genauere Herleitung der Werte in Tabelle 6 wäre hilfreich.
3.8, S. 78	<u>Anmerkung:</u> Im Rahmen des Abschnitts zur Potenzialbewertung wird beschrieben, dass eigene Recherchen zur Unterstützung der Bewertung gegebenenfalls durchgeführt werden. Es wird gleichzeitig darauf hingewiesen, dass es jedoch nicht das Ziel sei, die vorgelegten Unterlagen zu vervollständigen. <u>Vorgeschlagene Änderung:</u> Es verbleibt unklar, welchen Einfluss die Ergebnisse der durchgeführten Recherchen auf die Durchführung der Potenzialbewertung haben und wie sie dokumentiert werden. Es wird daher vorgeschlagen, konkreter die Zielsetzung dieser Recherchen zu formulieren und darzulegen, welchen Einfluss sie (z.B. im Falle der Identifikation neuer Studien) auf das Ergebnis der Potenzialbewertung haben können.
3.8 (S. 79)	<u>Anmerkung:</u> Es wird beschrieben, dass bei Bewertungen gemäß § 137h SGB V der G-BA lediglich die offensichtliche Schädlichkeit oder Unwirksamkeit einerseits oder den offensichtlichen Nutzen andererseits prüfen muss und die Methodik des Instituts entsprechend angepasst werde, sobald der G-BA diese Vorgaben entsprechend umsetzt. Es wird darum gebeten, diese Operationalisierung explizit als Ergänzung der Methodik zu publizieren, da es um die Prüfung kritischer Kriterien wie „offensichtliche Schädlichkeit“ oder „Unwirksamkeit“ geht. <u>Vorgeschlagene Änderung:</u> keine

5.1.1 (S. 117)	<u>Anmerkung:</u> Im letzten Abschnitt des Kapitels wird formuliert, dass Leitliniensynopsen „u. a. dem G-BA als Grundlage für die Entwicklung und Aktualisierung der erkrankungsspezifischen Anforderungen hinsichtlich der medizinischen Diagnostik und Behandlung von chronisch kranken Menschen [dienen] (Disease-Management-Programme [DMPs]).“ In der DMP-Anforderungen-Richtlinie (DMP-A-RL) werden auch über Diagnose und Therapie hinausgehende Bereiche geregelt, wie z.B. Schulung oder Kooperation der Versorgungssektoren. <u>Vorgeschlagene Änderung:</u> Änderung der Formulierung in „...Grundlage für die Entwicklung und Aktualisierung der Anforderungen an strukturierte Behandlungsprogramme für chronisch kranke Menschen...“.
5.1.4 (S. 118 f.)	a) Bewertung einzelner Domänen <u>Anmerkung:</u> Es ist eine größere Änderung des Vorgehens bei der Bewertung der Leitlinienqualität vorgesehen, indem nicht mehr das ganze AGREE II-Instrument angewendet wird, sondern nur noch eine Bewertung einzelner Domänen erfolgt: ▪ Domäne 2: Beteiligung von Interessengruppen ▪ Domäne 3: Genauigkeit der Leitlinienentwicklung ▪ Domäne 6: Redaktionelle Unabhängigkeit Dieses Vorgehen ist grundsätzlich sinnvoll, wenn man akzeptiert, dass damit ein etabliertes und validiertes Instrument nicht mehr in seiner ursprünglichen Form genutzt wird. In der Tat ist es so, dass Autorengruppen, die Leitlinien zu unterschiedlichen Zwecken bewerten, teilweise nur einzelne Domänen dazu nutzen. Darüber hinaus zeigte sich sowohl in einer Befragung von Leitlinienexperten [3] als auch in einer systematischen Übersichtsarbeit von Publikationen, in denen Leitlinien bewertet wurden [2], dass die Domäne 3 den größten Einfluss auf die Gesamtbewertung der Leitlinien hatte. Bei der Befragung wurde zusätzlich Domäne 6 genannt, in der systematischen Übersichtsarbeit wurde ein Einfluss von Domäne 5 (Anwendbarkeit) identifiziert. Vor dem Hintergrund der im Methodenentwurf angesprochenen Zielsetzung der Leitliniensynopsen, eine evidenzbasierte Grundlage für die Aktualisierung bestehender bzw. Entwicklung neuer DMPs zur Verfügung zu stellen, kann die Bewertung der methodischen Qualität der eingeschlossenen Leitlinien dazu dienen, eine über die Einschlusskriterien hinausgehende Prüfung der zugrunde liegenden Entwicklungsmethoden differenziert darzustellen und somit eine Differenzierung in methodisch hochwertige und methodisch schwächere Leitlinien vorzunehmen, vergleichbar einer Gesamtbewertung. Domäne 2 betreffend sind sicherlich die Fragen 4 (Die Entwicklergruppe der Leitlinie schließt Mitglieder aller

	relevanten Berufsgruppen ein) und 5 (Die Ansichten und Wünsche der Patienten wurden ermittelt) am relevantesten, wenn man die Kriterien zur Entwicklung glaubwürdiger Leitlinien [1] des Institute of Medicine (IOM) zugrunde legt. Darüber hinaus erfolgt im Rahmen der Beratungen im Gemeinsamen Bundesausschuss sowohl in der Arbeit in den Arbeitsgruppen und im Unterausschuss als auch im Rahmen der Stellungnahmeverfahren eine breite Einbindung von Experten und Betroffenen, so dass dadurch eine Beteiligung der Interessengruppen im Kontext des deutschen Gesundheitssystems erfolgt. Daher ist die Auswahl der Domänen zwar intuitiv einleuchtend, aber nicht unmittelbar aus den Ergebnissen der genannten Publikationen und der Zielsetzung der Leitliniensynopsen ableitbar. <u>Vorgeschlagene Änderung:</u> Es sollte insbesondere für Domäne 2 überlegt werden, ob diese den gleichen Stellenwert hat wie die anderen Domänen, bzw. die besondere Stellung der Domäne 3 sollte stärker herausgehoben werden.
5.1.4 (S. 118 f.)	b) Bewertung der methodischen Gesamtqualität <u>Anmerkung:</u> In der Konsequenz der nun nicht mehr vollständigen Anwendung von AGREE II kann auch keine Bewertung der methodischen Gesamtqualität mehr erfolgen. Da AGREE II bisher keine Operationalisierung der Gesamtbewertung hat und diese auch von anderen Autorengruppen unterschiedlich gehandhabt wird, ist die Gesamtbewertung ohnehin kritisch zu sehen. Auch das IQWiG hat dazu bisher keine Operationalisierung veröffentlicht, obwohl eine Gesamtbewertung in Leitliniensynopsen angegeben wurde. Aufgrund dieser bisher fehlenden Standardisierung erscheint diese Empfehlung entbehrlich und die Änderung somit sinnvoll. <u>Vorgeschlagene Änderung:</u> keine
5.1.4 (S. 118 f.)	c) Einführung eines Schwellenwertes <u>Anmerkung:</u> Es wird anstatt einer Gesamtbewertung ein Schwellenwert für methodisch gute und methodisch schwache Leitlinien eingeführt. Der genannte Schwellenwert von 30% erscheint auf Basis der in der Literatur verwendeten Systeme und Werte sinnvoll, wenn es primär darum geht, die besonders schlechten Leitlinien zu identifizieren, und nicht darum, die besonders guten von den mittelmäßigen Leitlinien zu unterscheiden.

	In der Literatur findet man unterschiedliche Operationalisierungen von Schwellenwerten. In Publikationen, in denen Schwellenwerte zur Einstufung der Leitlinienqualität genutzt werden, beziehen sich die verwendeten Schwellenwerte entweder auf die standardisierten Domänenwerte einzelner Domänen, mehrerer Domänen oder aber aller Domänen. Verwendet wird aber auch ein Schwellenwert auf Basis der Gesamtbewertung. In einigen Publikationen bleibt der Bezug unklar [4]. Da es also kein einheitliches Vorgehen gibt, sollte noch einmal darüber nachgedacht werden, ob die Domänen 2, 3, und 6 wirklich als gleichermaßen relevant für die Einschätzung der Leitlinienqualität angesehen werden (s. Ausführungen dazu unter a)). Es wäre ggf. sinnvoll, die Domäne 3 hier stärker zu gewichten und beispielsweise eine stärkere Konsequenz aus der Bewertung zu ziehen. So könnten Leitlinien, die in dieser Domäne einen standardisierten Domänenwert von <30% erzielen, aus der Leitliniensynopse ausgeschlossen werden, da dies doch auf starke methodische Mängel hindeutet, so dass hier in der Konsequenz auch die Evidenzbasierung in Frage zu stellen ist und damit ein wesentliches Einschlusskriterium für die Leitlinien in die Leitliniensynopsen. <u>Vorgeschlagene Änderung:</u> Ein Nicht-Erreichen des Schwellenwertes in Domäne 3 sollte in der Regel zum Ausschluss der Leitlinien aus der Synopse führen, in den anderen Domänen sollte eine Markierung erfolgen, wie im Entwurf des Methodenpapiers vorgeschlagen.
5.1.6 (S. 120 f.)	a) Einführung des Begriffes „Code“ <u>Anmerkung:</u> Es ist unklar, warum der Begriff „Code“ eingeführt wird und was sich am bisherigen Vorgehen der Erstellung von Kernaussagen generell ändert. <u>Vorgeschlagene Änderung:</u> Es sollte geprüft werden, ob die Bezeichnung „Code“ wirklich sinnvoll ist. Alternativ sollte hierzu näher erläutert werden. Insgesamt wäre es wünschenswert, wenn die Formulierung der Kernaussagen transparenter und nachvollziehbarer gestaltet wird und insbesondere nicht zu viele einzelne Aspekte in einer Kernaussage zusammengefasst werden.
5.1.6 (S. 120 f.)	b) Bearbeitung nur der zur DMP-A-Richtlinie diskrepanten Empfehlungen/Themen <u>Anmerkung:</u> Es werden bei Aktualisierungen nur die Empfehlungen/Themen weiter bearbeitet, in denen Diskrepanzen zwischen den aktuellen

	Leitlinien und der DMP-A-Richtlinie bestehen. Dies erscheint aus Ressourcengründen zunächst verständlich. Es birgt jedoch die Gefahr, dass neue Evidenz, die nicht zu einer Änderung der Empfehlung geführt hat, nicht weiter beachtet wird und damit auch nicht für die Diskussion innerhalb der Gremien und Formulierung z. B. der Tragenden Gründe verfügbar ist. Damit liefe man Gefahr, dass die DMP-A-RL zwar dem aktuellen Stand der Evidenz entspricht, die jeweiligen Begründungen aber nicht. Sollte in einer späteren Aktualisierung die Empfehlung doch einmal von der DMP-A-RL abweichen, ist es dann womöglich deutlich schwieriger und aufwendiger rekonstruierbar. <u>Vorgeschlagene Änderung:</u> Es sollten daher weiterhin alle Empfehlungen mit Bezug zur DMP-A-RL bearbeitet werden und Eingang in die Leitliniensynopsen finden.
5.1.6 (S. 120 f.)	c) Primäre Berücksichtigung des Grade of Recommendation (GoR) bei den Kernempfehlungen <u>Anmerkung:</u> In diesem Punkt erfolgte keine Änderung der Methoden. Trotzdem ist anzumerken, dass auch grundsätzlich evidenzbasiert entwickelte Leitlinien Empfehlungen beinhalten können, die nicht auf hoher Evidenz beruhen, wenn z. B. ▪ keine hochwertige Evidenz vorhanden / identifiziert ist oder ▪ keine hochwertige Evidenz für die Empfehlung zugrunde gelegt wurde bzw. die Interpretation der Daten nicht dem eigentlichen Evidenzniveau entspricht. Auch starke Empfehlungen (hoher GoR) können somit auf schwacher Evidenz beruhen bzw. auf einem reinen Expertenkonsensus. Aktuell erfolgt keine Plausibilitätsprüfung hinsichtlich der Ableitbarkeit des GoR. Auch werden GoR in gleicher Weise genutzt, wenn in der Leitlinie der zugehörige Level of Evidence (LoE) und / oder die zugrunde liegende Literatur nicht aufgeführt ist. Entsprechend der Kriterien zur Entwicklung glaubwürdiger Leitlinien [1] des Institute of Medicine (IOM) sollten aber sowohl LoE als auch GoR in den Leitlinien dargestellt werden. Darüber hinaus ist eine eigene Plausibilitätsprüfung für die an den Arbeitsgruppen bzw. im Unterausschuss Beteiligten nicht ohne Aufwand durchführbar, da die den Empfehlungen zugrunde liegende Literatur nicht direkt aus den Leitliniensynopsen entnommen werden kann. <u>Vorgeschlagene Änderung:</u> Der LoE sollte nicht nur bei Unklarheiten, sondern generell auch bei den Kernaussagen angegeben werden bzw. auf fehlende Angaben zum LoE und / oder der Evidenzbasis zu dem entsprechenden GoR hingewiesen werden.

	Insgesamt wäre es wünschenswert, wenn eine Plausibilitätsprüfung erfolgen würde, ob der hohe GoR plausibel aus der Evidenz ableitbar ist.
8.4 (S. 160)	<u>Anmerkung:</u> Es gibt keine Festlegung zur Aktualität der Leitlinien bei Einschluss in die Leitliniensynopsen. Um sicherzustellen, dass nicht bereits bei Fertigstellung des Berichtes bereits veraltete Leitlinien vorliegen, und um eine Einheitlichkeit sicherzustellen, wäre ein Vorgehen wie im aktuellen Berichtsplan „Leitliniensynopse für die Aktualisierung des DMP Brustkrebs“ (V19-06) [5] wünschenswert, siehe nächsten Absatz. <u>Vorgeschlagene Änderung:</u> Es wird empfohlen, die Formulierung „Leitlinien, die älter als 5 Jahre sind, entsprechen in der Regel nicht mehr dem aktuellen Erkenntnisstand. Es werden deshalb nur Leitlinien berücksichtigt, die vom im Auftrag des G-BA genannten Abgabetermin des Abschlussberichts [oder Rapid Reports] ausgehend maximal 5 Jahre zuvor veröffentlicht wurden“ in die Allgemeinen Methoden aufzunehmen.
9.3.3, Punkt 2 (S. 178)	<u>Anmerkung:</u> Zur Beurteilung der klinischen Relevanz werden zwei Alternativen angeboten, falls keine präspezifizierten Responsekriterien im Sinne einer MID von mindestens 15 % der Skalenspannweite vorliegen: SMD-Analysen mit einer Irrelevanzschwelle von 0,2 oder post-hoc spezifizierte Responderanalysen mit 15% der Skalenspannweite als Kriterium. Konsequenter wäre unseres Erachtens, auch hier die Präferenz der Responderanalysen zu betonen. <u>Vorgeschlagene Änderung:</u> Wir schlagen vor, im Falle des Vorliegens von ausschließlich kontinuierlichen Daten der Nachforderung von Responderanalysen mit der Schwelle von genau 15% des Skalenwertebereichs den Vorzug zu geben. Wir schlagen weiter vor, nur in Fällen, wo diese Responderanalysen nicht vorliegen (können), im Sinne einer Hilfslösung auf die SMD-Analyse auszuweichen.
9.3.3, Punkte 1-3 (S. 178)	<u>Anmerkung:</u> Zur Beurteilung der klinischen Relevanz sollen nun vorrangig Responderanalysen herangezogen werden. Dieses Vorgehen begrüßen wir. Gleichzeitig erscheint es uns angesichts des Informationsverlustes durch die Dichotomisierung wichtig, ergänzend weitere Analysen für die Bewertung heranzuziehen.

	<u>Vorgeschlagene Änderung:</u> Wir empfehlen, auch beim Vorliegen und Verwenden von Responderanalysen den durch diese bedingten Informationsverlust durch das Heranziehen / Vorlegen weiterer Analysen im Sinne einer Stützung der primären Analyse zu ergänzen. Vorstellbar wären der Mittelwertsunterschied oder weitere Responderanalysen mit anderen Schwellenwerten.
9.3.6 (S. 180)	<u>Anmerkung:</u> Es wird beschrieben, dass Äquivalenzbereiche bzw. Nichtunterlegenheitsgrenzen a priori festgelegt werden müssen. Jedoch werden keine Kriterien zur Festlegung dieser Grenzen beschrieben. <u>Vorgeschlagene Änderung:</u> Falls anderweitige Vorteile der neuen Intervention Anlass zu einer Nichtunterlegenheitsfragestellung geben, so sollte beschrieben werden, wie dies bei der Festlegung der Grenzen eingehen soll. Die Bedeutung der jeweiligen Endpunkte sollte hierbei eine Rolle spielen.
9.3.8, Unterpunkt B (S. 184)	<u>Anmerkung:</u> Als Kriterium für die Nicht-Zusammenfassbarkeit von Studien in einer Metaanalyse wird der p-Wert ($<0,05$) angegeben, auch I^2 wird als mögliches Kriterium erwähnt. In [6] wurde stattdessen vorgeschlagen, τ^2 als Kriterium zu verwenden, da dieses Maß die Einheit der Zielgröße zum Quadrat hat und somit eine klinische Einschätzung der Heterogenität ermöglicht. So könnte man zum Beispiel bei einer Auswertung in Hedges' g (mit 0,2 als kleinem Effekt) den Schwellenwert 0,04 ($0,2^2$) für τ^2 verwenden, da man 0,04 entsprechend als klinisch kleine Heterogenität ansehen könnte. <u>Vorgeschlagene Änderung:</u> Wir schlagen vor, τ^2 als Heterogenitätskriterium für Metaanalysen aufzunehmen.
9.3.10 (S. 190)	<u>Anmerkung:</u> Im Abschnitt „Subgruppenanalysen“ wird ausgeführt: „Bei mehr als 2 Subgruppen werden – wenn sinnvoll – die paarweisen statistischen Tests auf Interaktionen durchgeführt. Paare, die zum Niveau $\alpha = 0,05$ nicht statistisch signifikant sind (bei gleichzeitiger Signifikanz der restlichen Paare), werden zu einer Gruppe zusammengefasst. Die Ergebnisse der verbleibenden Gruppen werden dann getrennt berichtet und es werden getrennte Nutzensaussagen für diese Gruppen abgeleitet. Finden sich keine Paare, die zum Niveau $\alpha = 0,05$ nicht statistisch signifikant sind, werden keine Paare gebildet, sondern es wird eine separate Aussage für jede Subgruppe getroffen.“

	Diese Regel ist nicht ganz eindeutig: Liegen zum Beispiel 3 Subgruppen A, B und C vor, für den Vergleich AC gibt es eine signifikante Interaktion, für die Vergleiche AB und BC jedoch nicht, dann ist nicht klar, ob A mit B oder B mit C zusammengefasst werden soll, oder ob in diesem Fall eine separate Aussage für jede Subgruppe getroffen werden soll. <u>Vorgeschlagene Änderung:</u> Eine eindeutige Beschreibung dieser Regel zur Zusammenfassbarkeit wäre hilfreich.
9.3.11 (S. 191)	<u>Anmerkung:</u> Im Abschnitt zum Umgang mit nicht oder nicht vollständig publizierten Daten ist dargelegt: „Beim weitgehenden oder vollständigen Fehlen relevanter Informationen kann es vorkommen, dass eine Publikation nicht bewertet bzw. eine Studie nicht für die Analyse herangezogen werden kann. In diesen Fällen wird lediglich dargestellt, dass es zu einem bestimmten Thema weitere Daten gibt, die aber für eine Qualitätsbewertung bzw. für die Analyse nicht zur Verfügung stehen.“ In diesen Fällen ist die Wahrscheinlichkeit für einen Publication Bias hoch. Der obige Satz legt jedoch nahe, dass dies nicht in der Bewertung berücksichtigt wird. <u>Vorgeschlagene Änderung:</u> Wünschenswert ist eine Klarstellung, dass auch in diesen Fällen die Studie berücksichtigt wird. Sinnvolle Maßnahmen zur Berücksichtigung wären Sensitivitätsanalysen mit Annahmen über mögliche Studienergebnisse und/oder die Herabstufung der Ergebnissicherheit.
9.3 allgemein (S. 173 ff.)	<u>Anmerkung:</u> Im Entwurf der Allgemeinen Methoden 6.0 konnten wir keine Ausführungen zum Umgang mit bzw. zur Bewertung von Crossover-Studien identifizieren. <u>Vorgeschlagene Änderung:</u> Ein neues Unterkapitel über die Auswertung von Crossover-Studien mit den wichtigsten diesbezüglichen Aspekten wäre wünschenswert, da hier die Gefahr inadäquater Auswertungen besonders groß ist.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

- [1] Graham, R., Mancher, M., Miller-Wolman, D., Greenfield, S., Steinberg, E. Clinical Practice Guidelines we can trust. National Academy Press; letzter Zugriff: 31.01.2020. <https://www.nap.edu/catalog/13058/clinical-practice-guidelines-we-can-trust>. 2011
- [2] Hoffmann-Eßer, W., Siering, U., Neugebauer, E.A.M., Brockhaus, A.C., Lampert, U., et al. Guideline appraisal with AGREE II: Systematic review of the current evidence on how users handle the 2 overall assessments. PLoS One, 2017; 12 (3): e0174831-e0174831
- [3] Hoffmann-Eßer, W., Siering, U., Neugebauer, E.A.M., Brockhaus, A.C., McGauran, N., et al. Guideline appraisal with AGREE II: online survey of the potential influence of AGREE II items on overall assessment of guideline quality and recommendation for use. BMC Health Serv Res, 2018; 18 (1): 143-143
- [4] Hoffmann-Eßer, W., Siering, U., Neugebauer, E.A.M., Lampert, U., Eikermann, M. Systematic review of current guideline appraisals performed with the Appraisal of Guidelines for Research & Evaluation II instrument-a third of AGREE II users apply a cut-off for guideline quality. J Clin Epidemiol, 2018; 95: 120-127
- [5] Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Leitliniensynopse für die Aktualisierung des DMP Brustkrebs. Berichtsplan, Version 1.0; letzter Zugriff: 31.01.2020. https://www.iqwig.de/download/V19-06_DMP-Brustkrebs_Berichtsplan_V1-0.pdf 2020
- [6] Rücker, G., Schwarzer, G., Carpenter, J.R., Schumacher, M. Undue reliance on I(2) in assessing heterogeneity may mislead. BMC Med Res Methodol, 2008; 8 79-79

A.1.28 – Merck Serono GmbH

Autorinnen und Autoren

- Osowski, Ulrike

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden

Bitte nutzen Sie pro Person 1 Zeile.

Dr. Ulrike Osowski

Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)

☒ im Namen folgender Institution / Organisation: Merck Serono GmbH

☐ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.
./.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung
Abschnitt 9.3.3 „Beurteilung klinischer Relevanz“, S. 177/178	<u>Anmerkung:</u> Im Abschnitt 9.3.3 schlägt das IQWiG neue Kriterien zur Anwendung von Responderanalysen vor zur Beurteilung klinischer Relevanz vor: 1) Falls in einer Studie Responderanalysen unter Verwendung einer MID präspezifiziert sind und das Responsekriterium mindestens 15 % der Skalenspannweite des verwendeten Erhebungsinstruments entspricht, werden diese Responderanalysen ohne weitere Prüfung des Responsekriteriums für die Bewertung herangezogen. 2) Falls präspezifiziert Responsekriterien im Sinne einer MID unterhalb von 15 % der Skalenspannweite liegen, werden diese regelhaft nicht herangezogen. [...] Entweder können die Analysen der kontinuierlichen Daten herangezogen werden, für die Relevanzbewertung wird dabei auf ein allgemeines statistisches Maß in Form von standardisierten Mittelwertdifferenzen (SMDs, in Form von Hedges' g) zurückgegriffen. Dabei wird eine Irrelevanzschwelle von 0,2 verwendet: Liegt das zum Effektschätzer korrespondierende Konfidenzintervall vollständig oberhalb dieser Irrelevanzschwelle, wird davon ausgegangen, dass die Effektstärke nicht in einem sicher irrelevanten Bereich liegt. Dies soll gewährleisten, dass der Effekt hinreichend sicher mindestens als klein angesehen werden kann [233]. Alternativ können post hoc spezifizierte Analysen mit einem Responsekriterium von genau 15 % des Skalenrangs berücksichtigt werden.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	3) <i>Liegen sowohl geeignete Responderanalysen (Responsekriterium präspezifiziert mindestens 15 % des Skalenrangs oder post hoc genau 15 % des Skalenrangs) als auch Analysen stetiger Daten vor, werden die Responderanalysen herangezogen.“</i> Die Anwendung einer generischen MID-Schwelle von 15 % der theoretisch möglichen Skalenspannweite über Erhebungsinstrumente der verschiedensten Indikationen hinweg, ist wissenschaftlich nicht ausreichend begründet und kann daher nicht unterstützt werden. Die Herleitung dieser These („<i>eine Sichtung von systematischen Übersichtsarbeiten</i>“) wird im Methodenpapier nicht dargestellt und bleibt daher nicht nachvollziehbar. Die vom IQWiG zitierten Referenzen (Nordin 2016, St-Pierre 2016) thematisieren einzelne Symptome wie „<i>fatigue</i>“ und „<i>symptoms and functional limitations in individuals with rotator cuff disorders</i>“. Eine Übertragbarkeit auf krankheitsspezifische wie generische Erhebungsinstrumente bleibt fraglich. Die Evidenz der Aussage, dass „<i>sich MIDs in vielen Fällen zwischen 10 % und 20 % der Spannweite der Scores eines Erhebungsinstruments bewegen</i>“, bleibt unerklärt. Es ist nicht nachzuvollziehen, welche Kriterien bei der Recherche zugrunde lagen und was tatsächlich unter „<i>vielen Fällen</i>“ zu verstehen ist. Auf Basis dieser nicht näher bezeichneten Evidenz soll eine Schwelle für alle MID in Höhe von 15 % abgeleitet werden, die sich gemäß der Beschreibung des IQWiG aus den Grenzwerten 10 % und 20 % herleitet. Ob die Wahl auf den Mittelwert einer konkreten Begründung folgt, bleibt unklar. Ungeachtet der Herleitung der Schwelle für MID ist es aus Sicht von Merck zu berücksichtigen, dass die Validierung einer MID keineswegs pauschalisiert werden kann. Aus der Arbeit von Revicki et al.: <i>„The MID for a PRO instrument is not an immutable characteristic, but may vary by population and context, and no one MID may be valid for all study applications. MID estimates should be based on multiple approaches and triangulation of methods. Anchor-based methods applying various relevant patient-rated, clinician-rated, and disease-specific variables provide primary and meaningful estimates of an instrument’s MID.“ⁱⁱ</i> Die Autoren konkludieren:

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	„We recommend that the MID is based primarily on relevant patient-based and clinical anchors, with clinical trial experience used to further inform understanding of MID.” (Revicki, D. et al., 2008) Merck befürwortet diese Empfehlung, patientenberichtete Anker in der Validierung zu nutzen, um die klinische Relevanz einer bestimmten MID zu gewährleisten. Diese Sichtweise wird unter anderem auch von den Zulassungsbehörden geteilt (EMA, 2016). Eine von der AG Biostatistik des vfa vorgenommene Auswertung auf Basis von bisher in AMNOG-Verfahren akzeptierten MIDs zeigt zudem, dass die vorgeschlagene 15 %-Schwelle in ca. 90 % der Fälle zu einer Erhöhung der bisher validierten und in der Anwendung akzeptierten MIDs führen würde. So würde z.B. für das in der Onkologie sehr häufig verwendete Erhebungsinstrument EORTC QLQ-C30 eine sehr deutliche Erhöhung der MID um 50 % resultieren (von bisher 10 auf 15). In diesem speziellen Fall und in allen anderen ähnlich gelagerten Fällen, wäre es nicht nachvollziehbar, warum etablierte MIDs, deren klinische Relevanz bereits belegt ist, im Rahmen der Nutzenbewertung nicht verwendet werden sollten. Die EORTC beispielsweise hat es sich zum Ziel gemacht, MIDs für alle QLQ-C30 Skalen zu validieren (Musoro et al., 2019), erste Ergebnisse wurden bereits publiziert. Die Ergebnisse solcher wissenschaftlich fundierten Arbeiten sollten Berücksichtigung finden. Die Einführung einer generischen Schwelle von 15% Punkten der Skalenbreite würde zu einer regelhaften Erhöhung der Schwellenwerte führen, die zuvor als ausreichend galten, um den Zusatznutzen patienten- bzw. klinisch relevant zu evaluieren. Dies stünde einer gerechten Bewertung im Rahmen der frühen Nutzenbewertung entgegen. Des Weiteren ist der Schwellenwert von 15% Punkten infrage zu stellen, da die Skalierung des HRQoL Instruments ausschlaggebend sein kann für das Erreichen dieser generischen Schwelle. Wie in Tabelle 1 gezeigt, bestimmt die Wahl einer 5- bzw. einer 11-Punkt-Skala die Chance für das Erreichen dieser MID, bedingt durch eine Veränderung um 1 bzw. 2 Score Punkte. Aus dem Beispiel zeigt sich, dass bei einer 11-stufigen Skala eine Intensitätsveränderung von „überwiegend“ auf „einigermaßen“ den vom IQWiG geforderten Schwellenwert erfüllte, wohingegen bei einer 5-stufigen Skala eine patientenrelevante Veränderung in einem entsprechenden Bereich nicht festgestellt würde

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>																																								
	(„mittelmäßig“). Eine generische 15% Punkteskala kann nicht als eine adäquate Methode zur Bestimmung der MID herangezogen werden. <table><tr><th colspan="10">Intensität der Schmerzen</th></tr><tr><th>Keine</th><th colspan="3">Schwach</th><th colspan="3">Mittelmäßig</th><th colspan="2">Stark</th><th>Sehr stark</th></tr><tr><td>Keine</td><td>Sehr schwach</td><td>Schwach</td><td>Ziemlich schwach</td><td>Erträglich</td><td>Mittelmäßig</td><td>Überwiegend</td><td>Relativ stark</td><td>Stark</td><td>Sehr stark</td></tr><tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>Kaum zu ertragen</td></tr></table> <i>Tabelle 1: Beispiel für unterschiedliche Skalen</i> Vorgeschlagene Änderung: Die vorgeschlagenen Änderungen zur MID-Bewertungsmethodik sollten gänzlich entfallen. Die Etablierung von MIDs sollte wissenschaftlich-methodisch hergeleitet werden und sich auch auf bisherige Erfahrungen/Praxis stützen. Neue Standards zur MID-Bewertung sollten einem konsensbasierten, konstruktiven Dialog zwischen IQWiG, G-BA, Akademia und vfa folgen und schnellstmöglich gemeinsam erarbeitet werden. Das Ziel sollte sein, einen wissenschaftlich fundierten Kriterienkatalog zu etablieren, der für MIDs im Rahmen von Nutzenbewertungen regelhaft anzuwenden ist. Hierbei sollten u.a. auch die Ergebnisse der europäischen Innovative Medicines Initiative (IMI) Topic: „Establishing international standards in the analysis of patient reported outcomes and health-related quality of life data in cancer clinical trials“ Berücksichtigung finden. (IMI, 2019)	Intensität der Schmerzen										Keine	Schwach			Mittelmäßig			Stark		Sehr stark	Keine	Sehr schwach	Schwach	Ziemlich schwach	Erträglich	Mittelmäßig	Überwiegend	Relativ stark	Stark	Sehr stark										Kaum zu ertragen
Intensität der Schmerzen																																									
Keine	Schwach			Mittelmäßig			Stark		Sehr stark																																
Keine	Sehr schwach	Schwach	Ziemlich schwach	Erträglich	Mittelmäßig	Überwiegend	Relativ stark	Stark	Sehr stark																																
									Kaum zu ertragen																																
Abschnitt 3.3.3 „Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V“, S. 67	Anmerkung: <i>Aus dem Vorschlag des IQWiG Methodenpapiers: „In Anlehnung an das Verfahren für binäre Daten nach Skipka et al. [646] lassen sich auch Schwellenwerte für das Effektmaß SMD (Cohen's d bzw. Hedges' g) ableiten. Man kann zeigen, dass die für ein bestimmtes Ausmaß wünschenswerte Effektstärke ungefähr halbiert werden muss, um den entsprechenden Schwellenwert abzuleiten. Als Orientierung für die theoretisch wünschenswerten Effektstärken bei schwerwiegenden Symptomen können die Mittelwerte der Bereiche für die übliche Einteilung von Cohen's in kleine (SMD zwischen 0,2 und 0,5) mittlere (SMD zwischen 0,5 und 0,8) und große Effekte (SMD > 0,8) dienen. Tabelle 6 gibt einen Überblick über die verwendeten Schwellenwerte. Im Gegensatz zu den Schwellenwerten für die relativen Maße ist bei der SMD zu beachten, dass das Konfidenz-intervall vollständig oberhalb des entsprechenden Schwellenwerts liegen muss und der Nulleffekt hier nicht durch die Zahl 1 sondern durch die Zahl 0 gegeben ist.“</i>																																								

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	Die vorgeschlagenen Schwellenwerte für das Effektmaß SMD sind nach Ansicht der Merck Serono GmbH nicht ausreichend begründet, wissenschaftlich nicht nachvollziehbar oder nachprüfbar. Für die Werte der Zielkategorie schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen und gesundheitsbezogene Lebensqualität fehlen Referenzen, insbesondere auch eine wissenschaftlich fundierte Begründung, warum das Konfidenzintervall von Cohen's d bzw. Hedges' g vollständig oberhalb des entsprechenden Schwellenwerts liegen muss. <u>Vorgeschlagene Änderung:</u> Die allgemein akzeptierte Einteilung nach Cohen's soll für den Punktschätzer gelten und nicht für das Konfidenzintervall. Bei Abweichung von der Standardmethode sollte es zu einem wissenschaftlichen Austausch zwischen IQWiG, G-BA, Akademie und vfa kommen, um wissenschaftlich fundierte Rationale für Themen wie die Quantifizierung des Zusatznutzens von stetigen und quasistetigen Zielgrößen zu erörtern und festlegen zu können, und damit die Qualität und Relevanz der Nutzenbewertung sicherzustellen.
Abschnitt 9.3.12 „Umgang mit unvollständigen Daten“, S. 191-192	<u>Anmerkung:</u> <i>Aus dem Vorschlag des IQWiG Methodenpapiers:</i> <i>„Fehlende Daten erhöhen nicht nur die statistische Unsicherheit der Effektschätzung, sondern führen auch zu einem Verzerrungspotenzial, [...]. Ergebnisse fließen daher in der Regel nicht in die Nutzenbewertung ein, wenn diese auf weniger als 70 % der in die Auswertung einzuschließenden Patientinnen und Patienten basieren, [...].“</i> <i>Die Ergebnisse werden ferner auch dann nicht in die Nutzenbewertung einbezogen, wenn der Unterschied der Anteile nicht berücksichtigter Patientinnen und Patienten zwischen den Gruppen größer als 15 Prozentpunkte ist.“</i> In Leitlinien der Zulassungsbehörden für die Nutzen-Risiko-Bewertung von Medikamenten vor Marktzulassung werden keine entsprechenden Richtwerte zur Einschätzung der Anteile fehlender Werte gefordert (EMA, 2010; EMA 2017). Die Herleitung der Nichtakzeptanz von Studienergebnissen ab einem Grenzwert von 15 Prozentpunkten Unterschied fehlender Daten zwischen den Gruppen bleibt unreferenziert und daher nicht nachvollziehbar.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<u>Vorgeschlagene Änderung:</u> Die vorgegebenen Schwellenwerte für die Nichtberücksichtigung von Studienergebnissen sollten mit Literaturangaben untermauert werden. Für die Anwendung adäquater Ersetzungsstrategien und statistischer Auswertungsmethoden sollten Empfehlungen benannt werden. Merck schlägt einen regelmäßigen Austausch vor, um gemeinsam mit IQWiG, G-BA, Akademia und vfa wissenschaftlich fundierte Rationale für Themen wie den Umgang mit unvollständigen Daten erörtern und festlegen zu können.

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

1. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008; 61(2): 102-109
2. European Medicines Agency Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man, EMA/CHMP/292464/2014 (2016) [Zugriff online 30.01.2020] URL: https://www.ema.europa.eu/en/documents/other/appendix-2-guideline-evaluation-anticancer-medicinal-products-man_en.pdf
3. Musoro JZ, Coens C, Fiteni F et al., EORTC Breast and Quality of Life Groups, Minimally Important Differences for Interpreting EORTC QLQ-C30 Scores in Patients With Advanced Breast Cancer, JNCI Cancer Spectrum (2019) 3(3): pkz037
4. Innovative Medicines Initiative Topic: Establishing international standards in the analysis of patient reported outcomes and health-related quality of life data in cancer clinical trials (2019) [Zugriff online 30.01.2020] URL: https://www.imi.europa.eu/sites/default/files/uploads/documents/apply-for-funding/future-topics/IndicativeTopic_CancerPROs.pdf
5. European Medicines Agency Guideline on Missing Data in Confirmatory Clinical Trials, EMA/CPMP/EWP/1776/99 Rev. 1 (2010) [Zugriff online 30.01.2020] URL: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-missing-data-confirmatory-clinical-trials_en.pdf
6. European Medicines Agency ICH E9 (R1) addendum on estimands and sensitivity5 analysis in clinical trials to the guideline on statistical 6 principles for clinical trials (2017) [Zugriff online 30.01.2020] URL: https://www.ema.europa.eu/en/documents/scientific-guideline/draft-ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical_en.pdf

A.1.29 – MSD Sharp & Dohme GmbH

Autorinnen und Autoren

- Simang, Michael
- Scherr, Maximilian
- Tuebben, Stefanie
- Wiefarn, Stefan

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Tuebben Stefanie
Scherr Maximilian
Dr. Simang Michael
Wiefarn Stefan
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
☒ im Namen folgender Institution / Organisation: MSD Sharp & Dohme GmbH
☐ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

<i>Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.</i>

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
3.3.3 Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V D) Stetige oder quasistetige Zielgrößen mit jeweils vorliegenden SMDs (S.67)	<u>Anmerkung:</u> In dem Entwurf 6.0 des IQWiG Methodenpapiers wird im Abschnitt 3.3.3 eine Methodik zur Quantifizierung des Zusatznutzens bei stetigen oder quasistetigen Zielgrößen mit jeweils vorliegender Standardisierter Mittelwertdifferenz (SMD) vorgeschlagen. „D) Stetige oder quasistetige Zielgrößen mit jeweils vorliegenden SMDs <i>In Anlehnung an das Verfahren für binäre Daten nach Skipka et al. [646] lassen sich auch Schwellenwerte für das Effektmaß SMD (Cohen's d bzw. Hedges' g) ableiten. Man kann zeigen, dass die für ein bestimmtes Ausmaß wünschenswerte Effektstärke ungefähr halbiert werden muss, um den entsprechenden Schwellenwert abzuleiten. Als Orientierung für die theoretisch wünschenswerten Effektstärken bei schwerwiegenden Symptomen können die Mittelwerte der Bereiche für die übliche Einteilung von Cohen's d in kleine (SMD zwischen 0,2 und 0,5) mittlere (SMD zwischen 0,5 und 0,8) und große Effekte (SMD > 0,8) dienen. Tabelle 6 gibt einen Überblick über die verwendeten Schwellenwerte. Im Gegensatz zu den Schwellenwerten für die relativen Maße ist bei der SMD zu beachten, dass das Konfidenz-intervall vollständig oberhalb des entsprechenden Schwellenwerts liegen muss und der Nulleffekt hier nicht durch die Zahl 1 sondern durch die Zahl 0 gegeben ist.“</i>

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>																
	Tabelle 6: Schwellenwerte zur Ausmaßbestimmung für die SMD <table><tr><th colspan="2" rowspan="2"></th><th colspan="2">Zielgrößenkategorie</th></tr><tr><th>schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen und gesundheitsbezogene Lebensqualität^a</th><th>nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folge- komplikationen) und Nebenwirkungen</th></tr><tr><td rowspan="3">Ausmaß- kategorie</td><td>erheblich</td><td>0,6</td><td>nicht besetzt</td></tr><tr><td>beträchtlich</td><td>0,3</td><td>0,5</td></tr><tr><td>gering</td><td>0,2</td><td>0,2</td></tr></table> a: Voraussetzung ist – wie für alle patientenberichteten Endpunkte – die Verwendung eines validierten bzw. etablierten Instruments. Generell begrüßt MSD einen Vorschlag für die Ableitung des Zusatznutzens für stetige oder quasistetige Zielgrößen. Jedoch erscheint die vorgeschlagene Neuerung insbesondere für die Grenzwerte bei „schwerwiegenden (bzw. schweren) Symptomen (bzw. Folgekomplikationen) und Nebenwirkungen und gesundheitsbezogenen Lebensqualität“ nicht nachvollziehbar. So lassen sich lediglich die Schwellenwerte bei der Zielgrößenkategorie „nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen“ anhand der Einteilung nach Cohen nachvollziehen. Dazu ist zu bemerken, dass Cohen diese Einteilung auf den Punktschätzer und nicht die Konfidenzintervallgrenzen bezieht (1). Dies ist insbesondere vor dem Hintergrund, dass die Größe des Konfidenzintervalls maßgeblich von der Anzahl der Patienten in der jeweiligen Studie abhängt, als problematisch anzusehen. Des Weiteren schlagen wir vor, Hedges'g mit einer Grenze von 0,2 allenfalls als ein Kriterium unter vielen für die Quantifizierung des Zusatznutzens von stetigen und quasistetigen Auswertungen heranzuziehen. Besonders bei Auswertungen zu patientenberichteten Endpunkten sollte aus Sicht von MSD auch immer eine Einzelfallabwägung stattfinden. <u>Vorgeschlagene Änderung:</u> Die Herleitung von Schwellenwerten für die Ableitung eines Zusatznutzens bei stetigen bzw. quasistetigen Zielgrößen sollte anhand von wissenschaftlich etablierten Kriterien durchgeführt werden. Das Heranziehen von Schwellenwerten gemäß Cohen stellt sicherlich eine Möglichkeit dar, jedoch erlaubt dies aus Sicht von MSD lediglich eine Orientierung für den Punktschätzer, nicht jedoch als Schwellenwert für das Konfidenzintervall. In jedem Fall sollte jedoch eine Lösung im wissenschaftlichen Austausch und Konsens erarbeitet			Zielgrößenkategorie		schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen und gesundheitsbezogene Lebensqualität ^a	nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folge- komplikationen) und Nebenwirkungen	Ausmaß- kategorie	erheblich	0,6	nicht besetzt	beträchtlich	0,3	0,5	gering	0,2	0,2
				Zielgrößenkategorie													
		schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen und gesundheitsbezogene Lebensqualität ^a	nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folge- komplikationen) und Nebenwirkungen														
Ausmaß- kategorie	erheblich	0,6	nicht besetzt														
	beträchtlich	0,3	0,5														
	gering	0,2	0,2														

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	werden, die auch die Besonderheiten bei patientenberichteten Endpunkten berücksichtigt. Vor dem Hintergrund der besonderen Bedeutung von patientenberichteten Endpunkten sollte hier in der Regel eine Einzelfallabwägung bevorzugt werden.
9.3.3 Beurteilung klinischer Relevanz (S.177)	<u>Anmerkung:</u> So sehr eine Diskussion von Relevanzschwellen (MIDs) für patientenberichtete Endpunkte zu begrüßen ist, so sehr fehlen aus Sicht von MSD ausreichende Belege für einheitliche Schwellenwerte. Die vom IQWiG angeführte Evidenz setzt sich aus systematischen Übersichtsarbeiten (2, 3) zu MIDs von Erhebungsinstrumenten patientenberichteter Endpunkte in der Indikation Schultergelenkserkrankungen und einer indikationsübergreifenden Übersichtsarbeit mit Beschränkung auf das Symptom Erschöpfung zusammen. In beiden Übersichtsarbeiten weisen die Autoren auf erhebliche Limitationen hin, welche die Übertragbarkeit der Ergebnisse bereits im indikationsspezifischen Kontext als unzureichend erscheinen lässt. <u>Vorgeschlagene Änderung:</u> Wie in beiden durch das IQWiG angeführten Publikationen hervorgeht, empfiehlt sich eine indikations- und studienspezifische Abwägung der Erhebungsinstrumente und der herangeführten verwendeten Relevanzschwellen. Bereits in der Praxis etablierte und validierte Schwellenwerte zur Beurteilung einer klinischen Relevanz, welche zudem in der Nutzenbewertung bereits akzeptiert worden sind, sollten beibehalten werden. Künftige Lösungsvorschläge sollten im wissenschaftlichen Austausch und Konsens erarbeitet werden.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

1. Cohen J. Statistical Power Analysis For The Behavioral Sciences. 2nd Edition. Second Edition ed: Lawrence Erlbaum Associates; 1988.
2. Nordin Å, Taft C, Lundgren-Nilsson Å, Dencker A. Minimal important differences for fatigue patient reported outcome measures—a systematic review. BMC medical research methodology. 2016;16(1):62.
3. St-Pierre C, Desmeules F, Dionne CE, Frémont P, MacDermid JC, Roy J-S. Psychometric properties of self-reported questionnaires for the evaluation of symptoms and functional limitations in individuals with rotator cuff disorders: a systematic review. Disability and rehabilitation. 2016;38(2):103-22.

A.1.30 – Novartis Pharma GmbH

Autorinnen und Autoren

- Diwischek, Florian
- Finkes, Monika
- Gartner-Freyer, Daniela
- Hentschke, Christian
- Klebs, Sven
- Marx, Almuth
- Melzer, Nima
- Stemmer, Volker
- Wasmuth, Timo

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Diwischek, Florian; Dr.
Finkes, Monika
Gartner-Freyer, Daniela; Dr.
Hentschke, Christian
Klebs, Sven; Dr.
Marx, Almuth; Dr.
Melzer, Nima; Dr.
Stemmer, Volker; Dr.
Wasmuth, Timo; Dr.
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
☒ im Namen folgender Institution / Organisation: Novartis Pharma GmbH
☐ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
Abschnitt 3.2.1 (S. 55-57)	<u>Anmerkung:</u> In Abschnitt 3.2.1 zur Auswirkung nicht publizierter Studienergebnisse auf das Fazit beschreibt das IQWiG die Konsequenzen einer unvollständigen Datenübermittlung auf die Nutzenbewertung. Im Vergleich zur vorherigen Fassung des Methodenpapiers werden nun sieben Szenarien der Daten(un)vollständigkeit und die resultierenden Auswirkungen auf das Ausmaß des Nutzens bzw. Schadens detailliert dargestellt. Hierzu möchte die Novartis Pharma GmbH zu zwei Aspekten Stellung nehmen: Der erste Aspekt betrifft allgemein den Einflussbereich des pharmazeutischen Unternehmers (pU). Selbst bei vollständiger Datenübermittlung von Seiten des pU, besteht die Möglichkeit, dass dennoch Daten vorhanden sind, auf die der pU keinen Zugriff hat (Szenarien 2 und 3 in Tabelle 4). Gerade bei Substanzen, die sich schon länger im Markt befinden, ist die Wahrscheinlichkeit hoch, dass Studiendaten existieren, die sich dem Einflussbereich des pU entziehen. Für fehlende Daten aus solchen Studien, sollte nach Ansicht der Novartis Pharma GmbH der pU nicht haftbar gemacht werden, im Sinne, dass ihm daraus ein erheblicher Nachteil bei der Nutzenbewertung entsteht. Zudem ist eine detaillierte Operationalisierung zur Einschätzung des Ausmaßes der fehlenden Daten vorgegeben (vernachlässigbar, mäßig, erheblich, siehe Tabelle 4). In den meisten Fällen wird jedoch das Ausmaß schwierig bis unmöglich zu bestimmen sein: Die Relevanz der Daten und die durch das

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	Fehlen entstandene Verzerrung kann ohne Vorliegen der Daten kaum valide eingeschätzt werden. Der zweite Aspekt bezieht sich speziell auf die Vorgehensweise in Szenario 4: Falls die Datenübermittlung unvollständig ist und der pU nicht sicher nachweisen kann, dass das Ausmaß der fehlenden Daten „sicher irrelevant“ ist, resultiert eine „unklare“ Datenlage stets in der Konsequenz, dass kein Beleg (und kein Hinweis oder Anhaltspunkt) für einen Nutzen bzw. Schaden beansprucht werden kann. „Unklar“ ist die Datenlage in diesem Zusammenhang, wenn der Verdacht, dass insgesamt Daten fehlen, nicht eindeutig widerlegt werden kann. Auch hier können die Relevanz einer möglichen Datenlücke und die daraus resultierende Verzerrung kaum valide eingeschätzt werden. <u>Vorgeschlagene Änderung:</u> Sofern Hinweise darauf vorliegen, dass möglicherweise relevante Studien existieren, deren Ergebnisse unvollständig oder gar nicht für die Nutzenbewertung vorliegen, kann das bei der Bewertung in Form einer Einschränkung der Ergebnissicherheit Berücksichtigung finden, darf aber keinesfalls zwangsläufig mit der Bewertung „kein Beleg (und kein Hinweis oder Anhaltspunkt)“ (Szenarien 3 und 4) enden.
Abschnitt 3.3.3 (S. 67)	<u>Anmerkung:</u> Im vorliegenden Entwurf zum Methodenpapier 6.0 erweitert das IQWiG den Abschnitt 3.3.3 um Schwellenwerte zur Ausmaßbestimmung für SMDs (standardized mean difference) bei stetigen und quasistetigen Variablen. Das vorgeschlagene Vorgehen bei der Ausmaßbestimmung eines potenziellen Zusatznutzens für SMDs ist nicht wissenschaftlich begründet und in sich selbst nicht konsistent. Analog zum Vorgehen bei binären Zielgrößen erfolgt eine Abstufung der zu erreichenden Schwellenwerte für das jeweilige Ausmaß des potenziellen Zusatznutzens unterschieden nach schwerwiegenden (Kategorie 1) und nicht schwerwiegenden (Kategorie 2) Symptomen. Während in der Zielgrößenkategorie 2 die Schwellenwerte für die Kategorien „geringer“ und „beträchtlicher Zusatznutzen“ genau der von Cohen vorgeschlagenen Einteilung entsprechen, ist das Vorgehen in der Zielgrößenkategorie 1 weder in sich konsistent noch nachvollziehbar. Das IQWiG folgert aus entsprechenden Überlegungen für binäre Zielgrößen, dass die von Cohen genannten Werte ungefähr halbiert werden müssten. Dies wurde für die Kategorien

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	„beträchtlicher“ und „erheblicher Zusatznutzen“ auch so umgesetzt. Jedoch wird für einen „geringen Zusatznutzen“ abweichend von dieser Methodik gefordert, dass Hedges' g inklusive des zugehörigen Konfidenzintervalls vollständig oberhalb von 0.2 liegt. Dieser Wert wurde also im Widerspruch zu der vom IQWiG zitierten Herleitung nicht halbiert. Gründe hierfür werden nicht genannt und die Wahl des Schwellenwerts erscheint willkürlich. Dieses Vorgehen ist nach Ansicht der Novartis Pharma GmbH weder wissenschaftlich fundiert noch entspricht es den üblichen Standards der evidenzbasierten Medizin. <u>Vorgeschlagene Änderung:</u> Die Abstufung der Schwellenwerte für stetige Zielgrößen auf Basis der SMD sollte analog zur Abstufung der Schwellenwerte für die Quantifizierung des Zusatznutzens bei binären Endpunkten wissenschaftlich begründet und nach Möglichkeit auch validiert werden.
Abschnitt 5.1.4 (S. 118)	<u>Anmerkung:</u> Auch im vorliegenden Entwurf zum Methodenpapier 6.0 ist bei der Leitliniensynopse keine inhaltliche Prüfung vorgesehen. Zudem lässt das beschriebene Vorgehen keine Möglichkeit offen, neben Leitlinien bestverfügbare Evidenz beispielsweise aus Studien bei den Empfehlungen für Disease-Management-Programme (DMP) einzubeziehen. Nach Ansicht der Novartis Pharma GmbH besteht somit die Gefahr, neue hochwertige Therapien zu übersehen sowie bei komplexen Krankheitsbildern möglicherweise gar keine Empfehlungen zur Behandlung im DMP zu Verfügung stellen zu können. <u>Vorgeschlagene Änderung:</u> Neben der methodischen Prüfung der Leitlinien sollte auch eine inhaltliche erfolgen. Außerdem sollte in die Empfehlungen für DMPs auch die bestverfügbare Evidenz einbezogen werden.
Abschnitt 8.2.5 (S. 159)	<u>Anmerkung:</u> In Abschnitt 8.2.5 listet das IQWiG die Studienregister auf, die im Rahmen einer Dossierbewertung für die Recherche nach relevanten Studien heranzuziehen sind. Hier wird das Register PharmNet.Bund aufgelistet, das jedoch in der aktuellen Dossievorlage für Modul 4 in Abschnitt 4.2.3.3 zur Suche in

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	Studienregistern nicht mehr aufgeführt ist. Die Novartis Pharma GmbH sieht durch dieses Studienregister keinen zusätzlichen Nutzen für die Studiensuche. <u>Vorgeschlagene Änderung:</u> Aus Sicht der Novartis Pharma GmbH sollte die Nennung des Studienregisters PharmNet.Bund in Abschnitt 8.2.5 entfallen.
Abschnitt 9.3.3 (S. 177-178)	<u>Anmerkung:</u> Mit den neuen Regelungen zur Beurteilung klinischer Relevanz im Abschnitt 9.3.3 verfolgt das IQWiG folgende Ziele: Zum einen geht es darum „sicherzustellen, dass in Responderanalysen im Rahmen der Nutzenbewertung geeignete Responseschwellen eingesetzt werden“. Insbesondere soll erreicht werden, „dass eine Responseschwelle hinreichend sicher für die Patientinnen und Patienten spürbare Veränderungen abbildet“. Zum anderen soll „eine ergebnisgesteuerte Berichterstattung [...] durch eine beliebige Auswahl einer von vielen möglichen MIDs“ (minimal important difference) minimiert werden. Als geeignete Responseschwelle wird vom IQWiG regelhaft ein Responsekriterium definiert, das mindestens 15 % der Gesamtskalenspannweite des Erhebungsinstruments beträgt. Zudem sollen künftig MIDs nicht mehr herangezogen werden, auch wenn sie validiert, akzeptiert und präspezifiziert sind, falls sie kleiner sind als 15 % der Spannweite des betreffenden Erhebungsinstruments. Gegen die Einführung der 15 %-Grenze sprechen aus Sicht der Novartis Pharma GmbH die folgenden Gründe: Mit dem vorgeschlagenen Responsekriterium von 15 % bei MIDs besteht die Gefahr, dass zugunsten einer höheren Ergebnissicherheit patientenrelevante Veränderungen möglicherweise übersehen werden. Laut der vom IQWiG zitierten Übersichtsarbeiten lägen die meisten MIDs „in vielen Fällen“ zwischen 10 % und 20 % der Spannweite von Erhebungsinstrumenten. Diese Aussage ist nicht nachvollziehbar und das Vorgehen ist nicht sachgerecht. Die beiden beispielhaft genannten Übersichtsarbeiten betreffen zumindest lediglich eine Symptomatik (Fatigue) bzw. Patienten mit einer sehr speziellen Erkrankung (rotator cuff disorder) und können aus Sicht der Novartis Pharma GmbH nicht als Beleg für die angegebene Spanne betrachtet werden. Unklar bleibt zudem, wie repräsentativ die Grenzen von 10 und 20 % tatsächlich sind,

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	aus denen scheinbar willkürlich als Mittelwert und ohne weitere Begründung die 15 % abgeleitet werden. Außerdem wurden viele Instrumente mit dem Ziel entwickelt, auch sehr seltene Krankheitsmanifestationen oder extreme Ausprägungen abzudecken. Die Vielzahl extremer Ausprägungen erhöht die Spanne des Instruments deutlich, verändert jedoch die mittlere Ausprägung einer realen Patientenpopulation kaum. Das bedeutet, dass die meisten betroffenen Patienten bestimmte Randwerte nie oder nur in sehr seltenen Fällen erreichen. Ein gutes Beispiel hierfür ist der Psoriasis Area and Severity Index (PASI). Der PASI erfasst das Ausmaß der Rötung, Schuppung und Verhärtung der Haut bei Patienten mit Psoriasis auf einer Skala von 0-72 Punkten. 72 Punkte repräsentieren eine maximal ausgeprägte Schuppung, Rötung und Verhärtung an ausnahmslos allen Teilen der Haut aller Körperareale. Diese extreme Ausprägung kommt äußerst selten vor. Bereits ein PASI von > 10 gilt laut Leitlinien als Kriterium für eine mittelschwer bis schwer ausgeprägte Psoriasis (1). Folgt man der Regelung des IQWiG, so müsste eine MID für den PASI mindestens 10,8 Punkte ($= 15 \% \cdot 72 \text{ Punkte}$) betragen – mehr als ein Patient mit beispielsweise 10,2 Punkten sich überhaupt verbessern kann. Folgt man dem Vorschlag des IQWiG, kann also nicht jeder Patient, der eine mittelschwere bis schwere Psoriasis hat, eine klinisch relevante Verbesserung erfahren. Selbst wenn sein PASI von anfänglich 10,2 auf 0 sinkt, beträgt die Besserung nur 14 % der Skalenspannweite und liegt somit unter der geforderten Schwelle von mindestens 15 %. Offensichtlich hat das IQWiG mit dem Ziel „dass eine Responseschwelle hinreichend sicher für die Patientinnen und Patienten spürbare Veränderungen abbildet“ eine so extreme Regelung vorgeschlagen, dass etablierte und validierte Responseschwellen nicht mehr als patientenrelevant gelten. Dieses Beispiel zeigt, dass die Herleitung der 15 %-Grenze nicht pauschal für alle Skalen geeignet ist: Die in den bisherigen AMNOG-Verfahren akzeptierten MIDs liegen mehrheitlich unterhalb der 10 %-Grenze. Zudem sind patientenrelevante Verbesserungen für verschiedene Skalen, Indikationen, Schweregrade usw. durchaus unterschiedlich und können nicht regelhaft mit einer Pauschallösung abgehandelt werden. Ein Beispiel hierfür ist der 6-Minuten-Gehtest (6MWT, six-minute-walk test), ein standardisiertes Verfahren zur objektiven Beurteilung einer Belastungseinschränkung, die auf kardiopulmonale Ursachen zurückzuführen ist. Der Test ermittelt die Gehstrecke, die ein

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	Patient innerhalb von 6 Minuten ebenerdig zurücklegen kann. In Validierungsstudien ergaben sich unterschiedliche MIDs für unterschiedliche Patientengruppen: Bei Patienten mit chronischer obstruktiver Lungenerkrankung (COPD) (n=2.112) stellten 30 m einen relevanten Unterschied dar (2), bei Patienten mit idiopathischer Lungenfibrose (n=822) zeigte sich eine MID von 45 m (3), bei Patienten mit koronarer Herzkrankheit nach akutem Koronarsyndrom (n=81) betrug die ermittelte MID lediglich 25 m (4). Ein aktueller systematischer Reviewartikel über den 6MWT bei unterschiedlichen Patientenkollektiven (COPD, Lungenkrebs, koronare Herzkrankheit, diffuse parenchymale Lungenerkrankung, Non-CF-Bronchiektasen und Erwachsene mit Sturzangst) berichtet eine MID-Spannweite von 14-31 m (5). Die maximal zurückgelegte Strecke wird in dem Reviewartikel mit 551 m berichtet. Legt man diese Maximalstrecke, die immer noch unter dem Durchschnittsergebnis von Gesunden für den 6MWT liegt (6), als Skalenspannweite für den 6MWT zugrunde, ergibt sich ein 15 %-Responsekriterium von 83 m ($= 15 \% \cdot 551 \text{ m}$). Bereits dieses Kriterium liegt deutlich über allen zuvor ermittelten MIDs, obwohl die tatsächliche Maximalstrecke, die in 6 Minuten zurückgelegt werden kann, vermutlich noch größer ist. Dieses Beispiel illustriert, weshalb die Übertragbarkeit eines einzigen festgelegten 15 %-Responsekriteriums für die Beurteilung der patientenrelevanten Verbesserung als fragwürdig zu beurteilen ist. Eine ergebnisgesteuerte Berichterstattung kann nach Ansicht der Novartis Pharma GmbH durch andere Methoden als eine pauschale Anhebung des Responsekriteriums sachgerecht und valide verhindert werden. Die Novartis Pharma GmbH ist der Auffassung, dass im Methodenpapier festgehalten werden sollte, dass bereits vom G-BA anerkannte Response-Kategorien (darunter sind MIDs nur eine Kategorie), künftig grundsätzlich anerkannt werden, sofern nicht neue Evidenz eindeutig zeigt, dass sie nicht mehr dem aktuellen Stand der Erkenntnisse entsprechen. Hierfür wäre ein durch den G-BA bereitgestelltes und regelmäßig aktualisiertes Verzeichnis der anerkannten MIDs hilfreich. Diese Anerkennung etablierter MIDs sollte unabhängig davon erfolgen, ob die Analysen präspezifiziert sind, da hier keine Gefahr der ergebnisgesteuerten Berichterstattung besteht. Außerdem ist eine Vergleichbarkeit der Methodik zwischen verschiedenen Nutzenbewertungsverfahren ein weiteres wichtiges Prinzip, an dem sich das Methodenpapier orientieren sollte und dem so genüge getan werden kann.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	<u>Vorgeschlagene Änderung:</u> Die Novartis Pharma GmbH plädiert für ein folgendes Vorgehen zur Akzeptanz und Anwendung von MIDs: Bislang durch den G-BA akzeptierte MIDs sollten in künftigen Nutzenbewertungen weiterhin anerkannt werden. Ferner sollte für Instrumente, für die es bisher keine vom G-BA anerkannten Response-Kriterien gibt, eine individuelle Recherche durch das IQWiG im Rahmen der Nutzenbewertung erfolgen. Falls mehrere MIDs für solche Instrumente validiert wurden, sollte jede dieser MIDs anerkannt werden, sofern diese präspezifiziert wurde. Dies verhindert angemessen eine ergebnisgesteuerte Berichterstattung. Außerdem ist sichergestellt, dass die beste verfügbare Evidenz zur für dieses Instrument gültigen MID herangezogen wird. Diese ist trotz methodischer Einschränkungen einer pauschalen und unbegründeten Festlegung einer willkürlichen Schwelle im Sinne der evidenzbasierten Bewertung vorzuziehen. Sofern nur eine MID validiert wurde, sollte diese anerkannt werden, auch wenn sie nicht präspezifiziert wurde. Hier ist nicht mit einer ergebnisgesteuerten Berichterstattung zu rechnen. Falls mehrere MIDs validiert wurden und keine präspezifiziert wurde, sollte diejenige anerkannt werden, deren Validierungsstudie der klinischen Fragestellung am nächsten kommt.
Abschnitt 9.3.4 (S. 178)	<u>Anmerkung:</u> Im Abschnitt 9.3.4 beschreibt das IQWiG das Vorgehen zur Bewertung von Ergebnissen aus subjektiven Endpunkten bei offenen Studiendesigns. In der Regel wird in diesem Fall aufgrund der potenziell hohen Verzerrung bei statistisch signifikanten Ergebnissen kein Beleg für einen Nutzen oder Schaden abgeleitet. Dass Studienergebnisse aus subjektiven Endpunkten in randomisierten kontrollierten Studien mit offenem Design potenziell verzerrt sind, steht außer Zweifel. Dennoch kann die vorgeschlagene Vorgehensweise, in solchen Fällen regelhaft keinen Zusatznutzen (egal welchen Ausmaßes) abzuleiten, nicht zielführend sein. Patientenberichtete Endpunkte, wie sie beispielsweise zur Erfassung der Morbidität und gesundheitsbezogenen Lebensqualität eingesetzt werden, sind per se subjektiv. In Settings, die keine Verblindung erlauben (z. B. bei Interventionen wie Dialyse), könnten aufgrund der vorgeschlagenen Änderung in Zukunft keine patientenberichteten Endpunkte, die zum Beispiel Lebensqualität abbilden, mehr in die

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	Bewertung des Zusatznutzens einbezogen werden. Aus Sicht der Novartis Pharma GmbH kann es nicht dem Anspruch genügen, einen patientenrelevanten Zusatznutzen festzustellen, wenn systematisch alle patientenberichteten Endpunkte bei bestimmten Studiendesigns nicht herangezogen werden. <u>Vorgeschlagene Änderung:</u> Ein Zusatznutzen sollte auch aus subjektiven Endpunkten in offenen Studien abgeleitet werden können. Aus der potenziellen Verzerrung aufgrund des Designs sollte lediglich eine Einschränkung der Ergebnissicherheit resultieren.
Abschnitt 9.3.13 (S. 192f.)	<u>Anmerkung:</u> Im Abschnitt 9.3.13 wird der Umgang mit unterschiedlichen Beobachtungsdauern in den zu vergleichenden Gruppen klinischer Studien diskutiert. Unterschiedliche Beobachtungsdauern treten insbesondere bei der Erfassung unerwünschter Ereignisse in onkologischen Studien auf, da diese dort häufig nur bis zu einem definierten Zeitraum nach Behandlungsabbruch oder -wechsel erhoben werden. Um dieser Problematik Rechnung zu tragen, werden in der Nutzenbewertung gemäß § 35a SGB V bei Gruppenvergleichen mit unterschiedlichen mittleren Beobachtungsdauern regelhaft Verfahren für die Analyse von Überlebenszeiten angewendet. Ist die Ursache für die unterschiedlichen mittleren Beobachtungsdauern eine Zensierung bei Behandlungsabbruch oder -wechsel, so entsteht aus Sicht des IQWiGs jedoch auch bei Anwendung dieser Verfahren das Problem informativer Zensierungen, woraus ggf. ein hohes Verzerrungspotenzial folgt. Um solche Probleme zu vermeiden, sollte aus Sicht des Instituts deshalb eine vollständige Datenerfassung der unerwünschten Ereignisse erfolgen, und zwar auch nach einem Behandlungsabbruch oder -wechsel. In einem Verfahren der Nutzenbewertung gemäß § 35a SGB V wird im Rahmen der Erhebung von Schadenendpunkten das Schadenspotenzial eines bestimmten Arzneimittels bzw. einer bestimmten Arzneimittelkombination untersucht. Dazu ist es zwingend erforderlich, die beobachteten unerwünschten Ereignisse eindeutig dem Arzneimittel bzw. der Arzneimittelkombination zuordnen zu können. Eine Datenerfassung der unerwünschten Ereignisse deutlich über einen Behandlungsabbruch oder -wechsel hinaus, würde dazu führen, dass die unter der Folgetherapie auftretenden

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	unerwünschten Ereignisse ebenfalls in die Bewertung eingingen. Es bliebe unklar, ob die unerwünschten Ereignisse durch das verfahrensgegenständliche Arzneimittel oder durch die Folgetherapie verursacht werden; eine Bewertung des Schadenspotenzials des Arzneimittels würde damit unmöglich gemacht. Zudem bestünde ein hohes Verzerrungspotenzial infolge unterschiedlicher Folgetherapien innerhalb und zwischen den Behandlungsgruppen. <u>Vorgeschlagene Änderung:</u> Streichung der betreffenden Sätze („Ist die Ursache für die unterschiedlichen mittleren Beobachtungsdauern eine unvollständige Datenerfassung (Zensierung bei Behandlungsabbruch oder -wechsel), so entsteht allerdings auch bei Anwendung von Verfahren für Überlebenszeiten das Problem informativer Zensierungen, woraus ggf. ein hohes Verzerrungspotenzial folgt. Um solche Probleme zu vermeiden, sollte eine vollständige Datenerfassung der unerwünschten Ereignisse erfolgen, und zwar auch nach einem Behandlungsabbruch oder -wechsel [43]“.) in Abschnitt 9.3.13.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

1. Nast A, Amelunxen L, Augustin M, Boehncke W-H, Dressler C, M G, et al. S3-Leitlinie zur Therapie der Psoriasis vulgaris. AWMF online; 2017.
2. Polkey MI, Spruit MA, Edwards LD, Watkins ML, Pinto-Plata V, Vestbo J, et al. Six-minute-walk test in chronic obstructive pulmonary disease: minimal clinically important difference for death or hospitalization. Am J Respir Crit Care Med. 2013;187(4):382-6.
3. du Bois RM, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A, et al. Six-minute-walk test in idiopathic pulmonary fibrosis: test validation and minimal clinically important difference. Am J Respir Crit Care Med. 2011;183(9):1231-7.

4. Gremeaux V, Troisgros O, Benaim S, Hannequin A, Laurent Y, Casillas JM, et al. Determining the minimal clinically important difference for the six-minute walk test and the 200-meter fast-walk test during cardiac rehabilitation program in coronary artery disease patients after acute coronary syndrome. *Arch Phys Med Rehabil.* 2011;92(4):611-9.
5. Bohannon RW, Crouch R. Minimal clinically important difference for change in 6-minute walk test distance of adults with pathology: a systematic review. *J Eval Clin Pract.* 2017;23(2):377-81.
6. Chetta A, Zanini A, Pisi G, Aiello M, Tzani P, Neri M, et al. Reference values for the 6-min walk test in healthy subjects 20-50 years old. *Respir Med.* 2006;100(9):1573-8.

Abkürzungsverzeichnis

6MWT. six-minute-walk test

COPD. Chronic Obstructive Pulmonary Disease

DMP. disease management program

MID. minimal important difference

PASI. Psoriasis Area and Severity Index

pU. pharmazeutischer Unternehmer

SGB. Sozialgesetzbuch

SMD. standardized mean difference

A.1.31 – Novo Nordisk Pharma GmbH

Autorinnen und Autoren

- Bauer, Robert
- Kiencke, Peter

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Bauer, Robert; Dr.
Kiencke, Peter; Dr.
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
☒ im Namen folgender Institution / Organisation: Novo Nordisk Pharma GmbH
☐ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
Abschnitt 9.3.3	<u>Anmerkung:</u> Novo Nordisk begrüßt sehr, dass sich das IQWiG mit der wichtigen Thematik befasst, welche Methodik zur Ableitung einer klinisch relevanten MID skalenbasierter Erhebungsinstrumente angewendet werden sollte. Insbesondere wird begrüßt, dass die Präspezifikation von MIDs als relevantes Kriterium definiert wird. Das vom IQWiG vorgeschlagene zweite Kriterium einer pauschalen MID-Schwelle von 15 % der theoretisch möglichen Skalenspannweite, ist aus unserer Sicht allerdings nicht nachvollziehbar. Zur Herleitung dieser Schwelle bezieht sich das IQWiG auf „eine Sichtung von systematischen Übersichtsarbeiten“. Hier bleibt unklar, ob diese Sichtung vollständig ist, da sich die beiden einzigen vom IQWiG exemplarisch angegebenen Quellen (Nordin 2016, St-Pierre 2016) lediglich auf ausgewählte und sehr spezifische Symptome beziehen. In einem nächsten Schritt wird argumentiert, dass „sich MIDs in vielen Fällen zwischen 10 % und 20 % der Spannweite der Scores eines Erhebungsinstruments bewegen“. Auch hier bleibt unklar, was unter „vielen Fällen“ zu verstehen ist und wie repräsentativ die angeführten Grenzen von 10 % bzw. 20 % wirklich sind. Anschließend wird die 15 %-Schwelle für alle MIDs abgeleitet. Dieser Wert ergibt sich wohl als Mittelwert der o.g. Grenzen von

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	10 % und 20 % - eine konkrete Begründung für die Wahl von 15 % liegt allerdings nicht vor. Neben der Willkürlichkeit dieser Setzung stellt sich die Frage, ob eine verteilungsbasierte und pauschale Schwelle für die MIDs aller Erhebungsinstrumente überhaupt sinnvoll sein kann. In der auch vom IQWiG zitierten Arbeit von Revicki et al. 2008 steht: „The MID for a PRO instrument is not an immutable characteristic, but may vary by population and context, and no one MID may be valid for all study applications. MID estimates should be based on multiple approaches and triangulation of methods. Anchor-based methods applying various relevant patient-rated, clinician-rated, and disease-specific variables provide primary and meaningful estimates of an instrument's MID.“ Die Autoren kommen in ihrer Arbeit daher zu folgender Empfehlung: „We recommend that the MID is based primarily on relevant patient-based and clinical anchors, with clinical trial experience used to further inform understanding of MID.“ Novo Nordisk schließt sich dieser Empfehlung an, da nur durch die Berücksichtigung von Patienten- und Arzt-basierten Ankern sichergestellt werden kann, dass die abgeleitete MID auch klinisch relevant ist. <u>Vorgeschlagene Änderung:</u> Novo Nordisk schlägt vor, von der Etablierung einer allgemeingültigen Schwelle von 15% abzusehen. Responderanalysen unter Verwendung einer <u>präspezifizierten</u> MID sollten grundsätzlich für die Nutzenbewertung herangezogen werden. Zur Beurteilung der Stärke der Ergebnis(un)sicherheit sollte dann in einem zweiten Schritt die Validität der verwendeten MID beurteilt werden. Als Analysen höchster Aussagesicherheit ("Beleg") könnten dabei in Anlehnung an Revicki et al. 2008 Responderanalysen eingestuft werden, welche MIDs verwenden, die auf patienten-basierten und klinischen Ankern basieren.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008; 61(2): 102-109.

A.1.32 – Pfizer Deutschland GmbH

Autorinnen und Autoren

- Genet, Astrid
- Kauffmann, Stephanie
- Kullack, Max
- Leverkus, Friedhelm
- Trescher, Saskia

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Friedhelm Leverkus
Astrid Genet
Stephanie Kauffmann
Dr. Saskia Trescher
Max Kullack
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
☒ im Namen folgender Institution / Organisation: Pfizer Deutschland GmbH
☐ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.

Die Pfizer Deutschland GmbH (nachfolgend: Pfizer) möchte positiv hervorheben, dass sich das IQWiG dazu entschieden hat, den Entwurf für Version 6.0 seiner Allgemeinen Methoden und die darin enthaltenen Änderungen gegenüber der Version 5.0 zur wissenschaftlichen Diskussion im Rahmen eines Stellungnahmeprozesses zu stellen. Pfizer beteiligt sich an der Diskussion und reicht die folgende Stellungnahme ein.

Im Rahmen der Stellungnahme werden die folgenden Punkte kommentiert:

- Umgang mit informativen Zensierungen bei unerwünschten Ereignissen bei Behandlungsabbruch oder Behandlungswechsel
- Neue Skala für stetige oder quasistetige Zielgrößen mit jeweils vorliegenden standardized mean differences (SMDs) für die Feststellung des Ausmaßes des Zusatznutzens auf Endpunktebene
- Bestimmung von **minimal important differences** (MIDs) für die Beurteilung der klinischen Relevanz von Studienergebnissen

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
9.3.13 (S.192-193)	<u>Anmerkung:</u> Das IQWiG beschreibt in Abschnitt 9.3.13 „Umgang mit unterschiedlichen mittleren Beobachtungsdauern“ das Problem informativer Zensierungen bei Anwendung von Verfahren für Überlebenszeiten bei unerwünschten Ereignissen (UE), welches auftritt, wenn bei Behandlungsabbruch oder Behandlungswechsel die entsprechenden Beobachtungen zensiert werden und dies zu unterschiedlichen Beobachtungsdauern in den zu vergleichenden Gruppen führt. Pfizer stimmt mit dem IQWiG damit überein, dass die Anwendung einfacher Methoden basierend auf relativen Häufigkeiten oder Inzidenzdichten in solchen Fällen nicht sinnvoll ist.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	Pfizer stimmt aber nicht zu, dass eine weitere Erfassung der unerwünschten Ereignisse nach einem Behandlungsabbruch oder -wechsel in der Regel immer wünschenswert sei. Die Erhebung der UE bis zum Studienabbruch erfolgt ähnlich in beiden Studienarmen gemäß ITT-Prinzip, das kausale Rückschlüsse erlaubt. Ein vollständiges follow-up der Patienten, die die Studie Protokollgemäß beendet haben, bis zum geplanten Studienende ist somit nicht notwendig (1). Bei onkologischen Indikationen enthält eine leitlinienkonforme Behandlung eine Sequenztherapie. In diesem Fall gehören das Ende der Therapie und der Behandlungswechsel zum Therapieplan. Eine weitere Erhebung der UE wäre in diesem Fall nicht zielführend - entscheidungsrelevant für die Fragestellung der Nutzenbewertung ist die Beurteilung der Patienten in dem Teil der Sequenz, in dem die neuen Wirkstoffe eingesetzt ist. Auch wenn die weitere Erhebung der UE nach Therapieabbruch in manchen Indikationen und Fragestellungen theoretisch relevante Informationen liefern könnte, halten wir diesen Ansatz in der Praxis meistens für nicht machbar, und dies aus mehreren praktischen Gründen. Erstens sind Patienten nach einem Abbruch nicht mehr unter der Behandlung von Interesse und auch nicht mehr verblindet. Dies könnte die UE-Meldung nach dem Abbruch beeinflussen und dazu führen, dass UE, die vor und nach Therapieabbruch erhoben wären, unterschiedlich und vom daher nicht mehr vergleichbar wären. Zweitens könnten die Patienten auch auf eine andere Studie wechseln. Dies führt zu einer weiteren Limitation, denn um an einer anderen Studie teilnehmen zu können, müssen die Patienten ihre Einwilligung für die erste Studie zurückziehen. In diesen Fall ist eine weitere Erhebung der Sicherheitsdaten noch schwieriger. Wir denken, dass die Fragestellung der Nutzenbewertung bei onkologischen Präparaten, die in Sequenztherapien erfolgen, durch die Erfassung der UE nach Behandlungsabbruch oder -wechsel nicht immer richtig beantwortet wäre. <u>Vorgeschlagene Änderung:</u> In der Regel reichen die Betrachtung der bis zum Behandlungsende. Sollen Aussagen über mehrere Sequenzen gemacht werden, dann kann eine Messung der AEs über Behandlungsende sinnvoll sein.
3.3.3 (S. 67)	Anmerkung:

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	Das IQWiG beschreibt in Abschnitt 3.3 „Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V“ des Entwurfs für Version 6.0 seiner Allgemeinen Methoden wie vom pU nachzuweisen ist, mit welcher Wahrscheinlichkeit und in welchem Ausmaß ein Zusatznutzen des zu bewertenden Arzneimittels im Vergleich zur zweckmäßigen Vergleichstherapie vorliegt. Für die Feststellung des Ausmaßes auf Endpunktebene schlägt das IQWiG unter D) Stetige oder quasistetige Zielgrößen mit jeweils vorliegenden SMDs eine neue Skala für stetige oder quasistetige Zielgrößen mit jeweils vorliegenden SMDs vor und leitet in einer neuen Matrix Schwellenwerte für das Effektmaß SMD ab. Pfizer begrüßt grundsätzlich die Möglichkeit, einen quantifizierbaren Zusatznutzens auf Basis von stetigen Zielgrößen abzuleiten zu können. Die konkreten Schwellenwerte für die SMD in Tabelle 6 sind aus Sicht von Pfizer aber fraglich. Das IQWiG liefert keine Erläuterung und verweist auf keine Quellen für die verwendete Methodologie zur Herleitung der Schwellenwerte für die Zielgröße SMD. In "Statistical Power Analysis for the Behavioral Sciences" (1988) sind von Cohen Effektgrößen für SMD wie folgt zugeordnet worden: • Kleine Effektgröße: $d=0,2$ • Mittlere Effektgröße: $d=0,5$ • Große Effektgröße: $d=0,8$ Diese Schwellenwerte nach Cohen gelten aber für den Punktschätzer selbst und nicht für die untere Grenze eines Konfidenzintervalls wie in dem Vorschlag des IQWiG. Vor diesem Hintergrund erscheint die Herangehensweise des IQWiG sehr konservativ. Zusätzlich weist Cohen den Anwender darauf hin, dass "these proposed conventions were set forth throughout with much diffidence, qualifications, and invitations not to employ them if possible. The values chosen had no more reliable a basis than my own intuition" (2). <u>Vorgeschlagene Änderung:</u> Generell sollte die Kategorisierung des Ausmaßes des Zusatznutzens für stetige oder quasistetige Zielgrößen beibehalten werden. Die Wahl der Schwellenwerte sollte jedoch begründet bzw. mit Quellen belegt werden. Die Schwellenwerte sollten nachvollziehbar sein, wissenschaftlich begründet werden können und den international anerkannten Standards und Kriterien der evidenzbasierten Medizin genügen. Deshalb sollte die übliche und allgemein akzeptierte Einteilung für Cohen' d angewendet und nicht modifiziert werden. Zudem sollte berücksichtigt werden, dass Schwellenwerte nach Cohen für den Punktschätzer gelten und nicht für das Konfidenzintervall intendiert sind.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	Es sollte wie im Anhang A des Methodenpapiers für Responder Analyse, eine detaillierte methodische Rationale für die Festlegung der Schwellenwerte dargestellt werden. Es wäre wichtig zu zeigen, dass (1) die verschobene Hypothesengrenze (Schwellenwerte) dieselbe Power hat, die für die Testung der üblichen Hypothesen festgelegt wurde und (2) ähnliche Endpunkte, die unter einer Stetigen- und einem Responder-Ansatz analysiert wurden, zu der gleichen Kategorisierung führen.
9.3.3 (S. 178)	<u>Anmerkung:</u> Das IQWiG beschreibt in Abschnitt 9.3.3 „Beurteilung klinischer Relevanz“, dass für die Bestimmung von MIDs grundsätzlich empirische Verfahren präferiert werden, die eine MID mithilfe patientenberichteter Anker ermitteln. Jedoch werden für einzelne Instrumente teilweise mehrere MIDs veröffentlicht. Weiterhin lägen derzeit keine akzeptierten Standards vor, mit denen die Qualität der Validierungsstudien von MIDs beurteilt werden könne. Daher schlägt das IQWiG vor, Responderanalysen nur dann zu berücksichtigen, wenn das Responsekriterium mindestens 15 % der Skalenspannweite des verwendeten Erhebungsinstruments entspricht. Hierdurch soll nach Aussage des IQWiG erreicht werden, dass eine Responseschwelle hinreichend sicher eine für die Patienten spürbare Veränderung abbildet und das Risiko einer ergebnisgesteuerten Berichterstattung durch eine beliebige Auswahl einer von vielen möglichen MIDs minimiert wird. Nach Ansicht Pfizers sollten validierte, allgemein akzeptierte MIDs weiterhin Verwendung finden, z. B. eine MID von 10 Punkten für den EORTC QLQ C30 und eine MID von 10 Punkten für den FACT-P. Selbst wenn bezüglich der Qualität der Validierungsstudie der MID aufgrund einzelner Kriterien Unsicherheiten bestehen, sollte eine Diskussion über die Aussagekraft der ermittelten MID ermöglicht und nicht von vornherein ausgeschlossen werden. Um eventuell bestehende Unsicherheiten hinsichtlich der Qualität von Validierungsstudien zu MIDs beseitigen zu können, sollten allgemein Kriterien für die Beurteilung der Validierungsstudien formuliert werden. Hierzu finden sich in der Literatur bereits Ansätze [(3), (4), (5)]. Die Bestimmung eines Schwellenwerts für die klinische Relevanz einer Veränderung sollte generell mit Hilfe statistischer Methoden gerechtfertigt werden (z.B. durch ein ankerbasiertes Verfahren). Auch wenn für ein Instrument mehrere validierte MIDs vorliegen, sollten nicht alle MIDs kategorisch abgelehnt werden. In dieser Situation sollte das Risiko einer möglichen ergebnisgesteuerten Berichterstattung durch eine beliebige Auswahl einer MID vielmehr dadurch minimiert werden, dass eine konkrete Methode genannt

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	wird, wie das Responsekriterium basierend auf den validierten MIDs zu wählen ist. Beispielsweise könnte die MID anhand der Ähnlichkeit zur untersuchten Indikation gewählt werden und Sensitivitätsanalysen mit weiteren publizierten MIDs durchgeführt werden. Der gewählte Schwellenwert in Höhe von 15% der Skalenspannweite wird dadurch begründet, dass eine Sichtung von systematischen Übersichtsarbeiten zu MIDs ergeben hat, dass sich MIDs in vielen Fällen zwischen 10 % und 20 % der Spannweite der Scores eines Erhebungsinstruments bewegen. Hierzu werden zwei Referenzen genannt. Die erste Referenz (6) nennt Ergebnisse zu publizierten MIDs für verschiedene Fatigue-Skalen (sowohl multidimensionale Skalen als auch einzelne Items), die zweite Referenz (7) untersucht Instrumente zur Erfassung von Symptomen und funktionalen Einschränkungen bei Erkrankungen der Rotatorenmanschette. Es ist sehr fraglich, ob die Ergebnisse zu zwei sehr spezifischen Symptomen bzw. Erkrankungen auf alle Erhebungsinstrumente übertragen werden können. Der gewählte Schwellenwert in Höhe von 15% der Skalenspannweite würde bei vielen häufig verwendeten Instrumenten, für die akzeptierte validierte MIDs vorliegen, zu deutlich höheren Responseschwellen führen. <u>Vorgeschlagene Änderung:</u> Die vorgeschlagenen Änderungen zur MID-Bewertungsmethodik sind wissenschaftlich nicht hinreichend begründet und sollten gänzlich entfallen. Eine mögliche Änderung der Methodik darf nicht zur Ablehnung von Responseschwellen führen, die bisher als validiert bzw. etabliert galten. Aus Sicht vom Pfizer sollten mögliche neue Standards zur MID-Bewertung nur nach intensivem Austausch mit Vertretern aus Wissenschaft, Praxis und pharmazeutischer Industrie entwickelt werden. Pfizer ist hierbei offen für einen Dialog. Der Austausch könnte im Rahmen einer Serie von Kolloquien und Workshops erfolgen, in denen ein wissenschaftlicher Konsens angestrebt werden sollte. Erst danach sollte eine Änderung des Methodenpapiers in der Sache angestrebt werden.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

1. Leuchs, A. K., Brandt, A., Zinserling, J. & Benda, N. Disentangling estimands and the intention-to-treat principle. *Pharmaceutical Statistics* 2017;16(1):12-19.
2. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. New York: Lawrence Erlbaum Associates; 1988.
3. Coon D. C. & Cook, K. F. Moving from significance to real-world meaning: methods for interpreting change in clinical outcome assessment scores. *Quality of Life Research*. 2018(27):33-40.
4. Musoro Z. J., Hamel, J.-F., Ediebah, D. E., et al. Establishing anchor-based minimally important differences (MID) with the EORTC quality-of-life measures: a meta-analysis protocol. *BMJ Open*. 2018;8:e019117.
5. U.S. Department of Health and Human Services Food and Drug Administration. *Guidance for Industry, Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*. 2009.
6. Nordin Å, Taft, C., Lundgren-Nilsson, Å, et al. Minimal important differences for fatigue patient reported outcome measures—a systematic review. *BMC Medical Research Methodology*. 2016;16(62).
7. St-Pierre C, Desmeules, F., Dionne, C. E., et al. Psychometric properties of self-reported questionnaires for the evaluation of symptoms and functional limitations in individuals with rotator cuff disorders: a systematic review. *Disability and Rehabilitation*. 2016;38(2):103-22.

A.1.33 – Roche Pharma AG

Autorinnen und Autoren

- Knoerzer, Dietrich

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Dietrich Knoerzer
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
☒ im Namen folgender Institution / Organisation: Roche Pharma AG, Emil-Barell-Str. 1, 79639 Grenzach
☐ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.

Der Stellungnahme sind lediglich die Publikationen beigelegt, die nicht bereits im IQWiG-Methodenpapier referenziert sind.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
Abschnitt 3.3.3, S. 67	<u>Anmerkung:</u> Das IQWiG schlägt im Abschnitt 3.3.3 ein Vorgehen zur Quantifizierung des Zusatznutzens von stetigen und quasistetigen Zielgrößen vor. Darin heisst es: „In Anlehnung an das Verfahren für binäre Daten nach Skipka et al. [646] lassen sich auch Schwellenwerte für das Effektmaß SMD (Cohen's <i>d</i> bzw. Hedges' <i>g</i>) ableiten. Man kann zeigen, dass die für ein bestimmtes Ausmaß wünschenswerte Effektstärke ungefähr halbiert werden muss, um den entsprechenden Schwellenwert abzuleiten. Als Orientierung für die theoretisch wünschenswerten Effektstärken bei schwerwiegenden Symptomen können die Mittelwerte der Bereiche für die übliche Einteilung von Cohen's in kleine (SMD zwischen 0,2 und 0,5) mittlere (SMD zwischen 0,5 und 0,8) und große Effekte (SMD > 0,8) dienen. Tabelle 6 gibt einen Überblick über die verwendeten Schwellenwerte. Im Gegensatz zu den Schwellenwerten für die relativen Maße ist bei der SMD zu beachten, dass das Konfidenzintervall vollständig oberhalb des entsprechenden Schwellenwerts liegen muss und der Nulleffekt hier nicht durch die Zahl 1 sondern durch die Zahl 0 gegeben ist.“ Roche begrüßt den Versuch des IQWiG, eine Grundlage für die Quantifizierung des Ausmaßes von Effektgrößen anhand von SMD zu schaffen. Die dabei vorgeschlagenen Schwellenwerte in Tabelle 6 des Entwurfs zum allgemeinen Methodenpapier lassen sich jedoch aus Sicht von Roche nicht vollständig herleiten und sind damit nicht allesamt nachvollziehbar. Damit fehlt aber ein konsistentes Kriterium für die Festlegung von Grenzen. Ferner

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	sollte die wissenschaftliche Rationale, auf denen die Festlegung basiert, erläutert werden. Eine Überprüfung der Konsistenz zu den derzeit geltenden Schwellenwerte für binäre Daten steht ebenfalls aus. <u>Vorgeschlagene Änderung:</u> Aus Sicht von Roche sollten die vorgeschlagenen Schwellenwerte nachvollziehbar hergeleitet und hinsichtlich der Konsistenz zu anderen Schwellenwerten überprüft werden. Diese Voraussetzungen müssen erfüllt sein, um eine weitere methodische Diskussion zu ermöglichen, die von Roche in diesem wie auch in weiteren Punkten sehr befürwortet wird. Roche ist gerne bereit auch inhaltlich mitzuarbeiten.
Abschnitt 9.3.3, S. 175 - 178	<u>Anmerkung:</u> Das IQWiG schlägt im Abschnitt 9.3.3 ein Vorgehen zur Ableitung einer klinisch relevanten MID von skalenbasierten Erhebungsinstrumenten. Darin heisst es: <i>„Eine Sichtung von systematischen Übersichtsarbeiten zu MIDs [z. B. [523,655]] hat ergeben, dass sich MIDs in vielen Fällen zwischen 10 % und 20 % der Spannweite der Scores eines Erhebungsinstruments bewegen. Daraus ergibt sich regelhaft folgendes Vorgehen:</i> 1) <i>Falls in einer Studie Responderanalysen unter Verwendung einer MID präspezifiziert sind und das Responsekriterium mindestens 15 % der Skalenspannweite des verwendeten Erhebungsinstruments entspricht, werden diese Responderanalysen ohne weitere Prüfung des Responsekriteriums für die Bewertung herangezogen.</i> “ Dies aus Sicht von Roche ein wichtiges Thema und auch belegt durch zahlreiche aktuelle internationale Initiativen, die sich mit der Vereinheitlichung von Kriterien für aussagkräftige und wissenschaftlich valide MIDs beschäftigen (z. B. [1]). Das nun vom IQWiG vorgeschlagene Vorgehen, eine pauschale MID-Schwelle von 15 % der theoretisch möglichen Skalenspannweite für alle Erhebungsinstrumente anzuwenden, ist aus Sicht von Roche wissenschaftlich nicht begründet und daher auch nicht nachvollziehbar. Eine Initiative der AG Biostatistik des vfa hat sich bereits mit dem IQWiG in Verbindung gesetzt, um eine nachvollziehbare Methodik in einem konstruktiven Dialog mit allen Beteiligten und Experten zu erörtern und um letztendlich zu einem verbindlichen Kriterienkatalog zu kommen, der bei der Ableitung bzw. Validierung von MIDs regelhaft angewendet werden sollte.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	Zur Herleitung der 15%- Schwelle verwendet das IQWiG mehrere Schritte: - Literaturrecherche, als „eine Sichtung von systematischen Übersichtsarbeiten“. Hier ist nicht ersichtlich, ob diese Sichtung vollständig bezüglich der Übersichtsarbeiten und der untersuchten Instrumente ist, da sich die beiden einzigen vom IQWiG exemplarisch angegebenen Quellen ([2, 3]) lediglich auf ausgewählte und sehr spezifische Symptome wie „<i>fatigue</i>“ und „<i>symptoms and functional limitations in individuals with rotator cuff disorders</i>“ beziehen. - In der Beschreibung, dass „<i>sich MIDs in vielen Fällen zwischen 10 % und 20 % der Spannweite der Scores eines Erhebungsinstruments bewegen</i>“ bleibt unklar, was unter „<i>vielen Fällen</i>“ zu verstehen ist und wie repräsentativ die angeführten Grenzen von 10 % bzw. 20 % wirklich sind. Auf Basis dieser unklaren Evidenz wird ohne Angabe der Herleitung, die 15 %-Schwelle für alle MIDs abgeleitet. Dieser Wert ergibt sich wohl als Mittelwert der o.g. Grenzen von 10 % und 20 % - eine konkrete Begründung für die Wahl von 15 % liegt allerdings nicht vor. Neben der Willkürlichkeit dieser Setzung stellt sich die Frage, ob eine verteilungsbasierte und pauschale Schwelle für die MIDs aller Erhebungsinstrumente überhaupt sinnvoll sein kann. In der auch vom IQWiG zitierten Arbeit [4] heißt es dazu: <i>„The MID for a PRO instrument is not an immutable characteristic, but may vary by population and context, and no one MID may be valid for all study applications. MID estimates should be based on multiple approaches and triangulation of methods. Anchor-based methods applying various relevant patient-rated, clinician-rated, and disease-specific variables provide primary and meaningful estimates of an instrument’s MID.“</i> Die Autoren kommen in ihrer Arbeit zu folgender Empfehlung <i>„We recommend that the MID is based primarily on relevant patient-based and clinical anchors, with clinical trial experience used to further inform understanding of MID.“</i> Roche teilt diese Sicht der Dinge, da nur durch die Berücksichtigung von Patienten- und Arzt-basierten Ankern sichergestellt werden kann, dass die abgeleitete MID auch klinisch relevant ist. Diese Sichtweise wird unter anderem auch von der EMA geteilt ([5]).

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	<u>Vorgeschlagene Änderung:</u> Roche hält zur Lösung von diesen und ähnlich gelagerten Fragestellungen ein Gremium für unabdingbar, dass aus IQWiG, G-BA, Akademie, und Industrievertretern besteht und schon verschiedentlich von Roche vorgeschlagen wurde. Im konkreten Fall könnte das Gremium den von der AG Biostatistik bereits angeregten konstruktiven Dialog schnellstmöglich beginnen – mit dem Ziel, einen internationalen und wissenschaftlich fundierten Kriterienkatalog zu etablieren, der für MIDs im Rahmen von Nutzenbewertungen regelhaft anzuwenden ist.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

1. Musoro ZJ, Hamel JF, Ediebah DE, Cocks K, King MT, Groenvold M, Sprangers MA, Brandberg Y, Velikova G, Maringwa J, Flechtner HH. Establishing anchor-based minimally important differences (MID) with the EORTC quality-of-life measures: a meta-analysis protocol. *Bmj Open*. 2018; 8(1).
2. Nordin A, Taft C, Lundgren-Nilsson A, Dencker A. Minimal important differences for fatigue patient reported outcome measures: a systematic review. *BMC Med Res Methodol* 2016; 16: 62.
3. St-Pierre C, Desmeules F, Dionne CE, Fremont P, MacDermid JC, Roy JS. Psychometric properties of self-reported questionnaires for the evaluation of symptoms and functional limitations in individuals with rotator cuff disorders: a systematic review. *Disabil Rehabil* 2016; 38(2): 103-122.
4. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining

responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008; 61(2): 102-109.

5. EMA. Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man. The use of patient-reported outcome (PRO) measures in oncology studies. 2016.

A.1.34 – Sanofi-Aventis Deutschland GmbH

Autorinnen und Autoren

- Garbe, Janine
- Kienitz, Carsten
- Schinzel, Stefan
- Wolff, Alexandra

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Garbe, Janine; Dr.
Kienitz, Carsten
Schinzal, Stefan
Wolff, Alexandra
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
☒ im Namen folgender Institution / Organisation: Sanofi-Aventis Deutschland GmbH
☐ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

<i>Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.</i>

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
3.3.3 Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V (S.67)	<u>Anmerkung:</u> Der Abschnitt 3.3.3 wurde unter „D)“ um das Vorgehen zur Quantifizierung des Zusatznutzens bei stetigen oder quasistetigen Zielgrößen mit jeweils vorliegenden standardisierten Mittelwertdifferenzen ergänzt. Dazu heißt es: „In Anlehnung an das Verfahren für binäre Daten nach Skipka et al. [646] lassen sich auch Schwellenwerte für das Effektmaß SMD (Cohen’s d bzw. Hedges’ g) ableiten.“ Sanofi begrüßt grundsätzlich die Möglichkeit, auf Basis von stetigen oder quasistetigen Endpunkten mit jeweils vorliegenden standardisierten Mittelwertdifferenzen das Ausmaß des Zusatznutzens zu quantifizieren. Weiter schreibt das IQWiG: „Man kann zeigen, dass die für ein bestimmtes Ausmaß wünschenswerte Effektstärke ungefähr halbiert werden muss, um den entsprechenden Schwellenwert abzuleiten. Als Orientierung für die theoretisch wünschenswerten Effektstärken bei schwerwiegenden Symptomen können die Mittelwerte der Bereiche für die übliche Einteilung von Cohen’s d in kleine (SMD zwischen 0,2 und 0,5) mittlere (SMD zwischen 0,5 und 0,8) und große Effekte (SMD > 0,8) dienen. Tabelle 6 gibt einen Überblick über die verwendeten Schwellenwerte. Im Gegensatz zu den Schwellenwerten für die relativen Maße ist bei der SMD zu beachten, dass das Konfidenzintervall vollständig oberhalb des

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>																
	<i>entsprechenden Schwellenwerts liegen muss und der Nulleffekt hier nicht durch die Zahl 1 sondern durch die Zahl 0 gegeben ist.“</i> Ursprünglich wurden die Werte 0,2, 0,5 und 0,8 von Cohen 1988 vorgeschlagen, um in Versuchen zur Verhaltensforschung, in denen keine oder nur spärliche Vorinformationen vorliegen, eine erste grobe Orientierung für die Fallzahlplanung zu geben [1]. Die Eignung dieser Werte als Anker für die Beurteilung der klinischen Relevanz von Effekten in der Arzneimittelentwicklung ist zu hinterfragen bzw. zu evaluieren. Weiter ist anzumerken, dass sich die von Cohen 1988 vorgeschlagenen Schwellenwerte auf Punktschätzer bezogen und nicht als untere Konfidenzintervallschranken zu verstehen waren. Tabelle 6: Schwellenwerte zur Ausmaßbestimmung für die SMD <table><tr><th colspan="2" rowspan="2"></th><th colspan="2">Zielgrößenkategorie</th></tr><tr><th>schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen und gesundheitsbezogene Lebensqualität^a</th><th>nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folge- komplikationen) und Nebenwirkungen</th></tr><tr><td rowspan="3">Ausmaß- kategorie</td><td>erheblich</td><td>0,6</td><td>nicht besetzt</td></tr><tr><td>beträchtlich</td><td>0,3</td><td>0,5</td></tr><tr><td>gering</td><td>0,2</td><td>0,2</td></tr></table> a: Voraussetzung ist – wie für alle patientenberichteten Endpunkte – die Verwendung eines validierten bzw. etablierten Instruments. Darüber hinaus werden in den in Tabelle 6 definierten Zielgrößenkategorien Nebenwirkungen aufgeführt. Es ist unklar wie das Effektmaß standardisierte Mittelwertdifferenzen bei Nebenwirkungen Anwendung findet. Auch ist die Herleitung nicht für alle der in Tabelle 6 aufgeführten Schwellenwerte nachvollziehbar. <u>Vorgeschlagene Änderung:</u> Eine transparente wissenschaftliche Herleitung geeigneter Schwellenwerte sollte sich - in Anlehnung an das Verfahren bei binären Endpunkten - im „Anhang A – Rationale der Methodik zur Feststellung des Ausmaßes des Zusatznutzens“ wiederfinden.			Zielgrößenkategorie		schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen und gesundheitsbezogene Lebensqualität ^a	nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folge- komplikationen) und Nebenwirkungen	Ausmaß- kategorie	erheblich	0,6	nicht besetzt	beträchtlich	0,3	0,5	gering	0,2	0,2
				Zielgrößenkategorie													
		schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen und gesundheitsbezogene Lebensqualität ^a	nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folge- komplikationen) und Nebenwirkungen														
Ausmaß- kategorie	erheblich	0,6	nicht besetzt														
	beträchtlich	0,3	0,5														
	gering	0,2	0,2														
9.3 Spezielle biometrische Aspekte (S.173 ff.)	<u>Anmerkung:</u> Am 20.06.2018 wurde der IQWiG-Bericht Nr. 638 „Untersuchung der statistischen Eigenschaften von Verfahren zur Übertragbarkeit von Studienergebnissen auf Teilpopulationen“ mit umfangreichen Simulationen zur empirisch bestimmten statistischen Power der																

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	Anhebungsregel (AHR₁₅) veröffentlicht [2]. Es bleibt unklar, ob dieses Verfahren zwischenzeitlich weiter evaluiert wurde und warum es nicht Eingang in den Entwurf zur Version 6 der Allgemeinen Methoden gefunden hat. <u>Vorgeschlagene Änderung:</u> Wir schlagen die Aufnahme der Anhebungsregel als Verfahren zur Übertragbarkeit von Studienergebnissen auf Teilpopulationen im Methodenpapier vor.
9.3.3 Beurteilung klinischer Relevanz (S.175 ff.)	<u>Anmerkung:</u> Im Abschnitt 9.3.3 heißt es: „<i>Es gibt derzeit keine akzeptierten Standards, mit denen die Qualität dieser Studien bewertet und die Aussagekraft der ermittelten MIDs abgeschätzt werden kann [117,176,405]. Es wird zwar an der Entwicklung eines solchen Standards gearbeitet [175,405], erste Ergebnisse dieser Bemühungen lassen aber weiterhin Fragen offen [176].</i>“ Sanofi teilt diese Einschätzung und sieht Handlungsbedarf bzgl. der Weiterentwicklung solcher Standards in einem konstruktiven Dialog mit allen Beteiligten und Experten. Weiter schreibt das IQWiG: „<i>Eine Sichtung von systematischen Übersichtsarbeiten zu MIDs [z. B. [523,655]] hat ergeben, dass sich MIDs in vielen Fällen zwischen 10 % und 20 % der Spannweite der Scores eines Erhebungsinstruments bewegen.</i>“ Die beispielhaft aufgeführten Publikationen zur Sichtung der systematischen Übersichtsarbeiten beziehen sich auf ausgewählte und sehr spezifische Symptome wie „fatigue“ und „symptoms and functional limitations in individuals with rotator cuff disorders“. Eine vollständige Sichtung, transparente Publikation und Diskussion der Ergebnisse ist wünschenswert, um die aufgeführten Schwellenwerte nachvollziehen zu können. Schließlich wird das Allround-Responsekriterium präsentiert: „1) Falls in einer Studie Responderanalysen unter Verwendung einer MID präspezifiziert sind und das Responsekriterium mindestens 15 % der Skalenspannweite des verwendeten Erhebungsinstruments entspricht, werden diese Responderanalysen ohne weitere Prüfung des Responsekriteriums für die Bewertung herangezogen. 2) Falls präspezifiziert Responsekriterien im Sinne einer MID unterhalb von 15 % der Skalenspannweite liegen, werden diese regelhaft nicht herangezogen.“

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	Die Herleitung des Schwellenwertes ist intransparent. Falls es sich um eine verteilungsbasierte Herleitung auf Basis der gesichteten systematischen Übersichtsarbeiten handelt, scheint dies in diesem Kontext fragwürdig. Jegliche krankheits- und populationsspezifischen Einflüsse werden ebenso wie Skalencharakteristika ignoriert. Zu beachten ist, dass es sich bei Verwendung des Allround-Schwellenwertes für Patient Reported Outcomes mit etablierten MIDs, die - wie das IQWiG ganz richtig schrieb - in vielen Fällen bei 10 % liegen, um eine Erhöhung der MID um 50 Prozent handeln würde. Ein Beispiel dafür stellt der in der Onkologie häufig verwendete Fragebogen EORTC QLQ-C30 dar. <u>Vorgeschlagene Änderung:</u> Von der Einführung eines Allround-Responderschwellenwertes sollte abgesehen werden. Etablierte MIDs, deren klinische Relevanz bereits belegt ist, sollten unter Berücksichtigung neuer wissenschaftlicher Erkenntnisse, im Rahmen der Nutzenbewertung weiterhin auf Akzeptanz stoßen. Insbesondere sollten krankheits- und populationsspezifische Einflüsse ebenso wie Skalencharakteristika nicht ignoriert werden. Der konstruktive Dialog zur Entwicklung akzeptierter Standards sollte geführt werden mit dem Ziel, einen internationalen und wissenschaftlich fundierten Kriterienkatalog zu etablieren, der auch für MIDs im Rahmen von Nutzenbewertungen regelhaft anzuwenden ist.
9.3.12 Umgang mit unvollständigen Daten (S.191 f.)	<u>Anmerkung:</u> Der Abschnitt 9.3.12 zum Umgang mit unvollständigen Daten wurde ergänzt: <i>„Das Verzerrungspotenzial steigt mit steigendem Anteil fehlender Werte. Ergebnisse fließen daher in der Regel nicht in die Nutzenbewertung ein, wenn diese auf weniger als 70 % der in die Auswertung einzuschließenden Patientinnen und Patienten basieren, das heißt, wenn der Anteil der Patientinnen und Patienten, die nicht in der Auswertung berücksichtigt werden, größer als 30 % ist. In der Literatur werden zum Teil bereits Auswertungen, in denen 20 % der Patientinnen und Patienten nicht berücksichtigt werden, als nicht mehr aussagekräftig betrachtet [618].</i> <i>Die Ergebnisse werden ferner auch dann nicht in die Nutzenbewertung einbezogen, wenn der Unterschied der Anteile nicht berücksichtigter Patientinnen und Patienten zwischen den Gruppen größer als 15 Prozentpunkte ist.“</i>

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	Eine systematische Literaturrecherche, auf deren Basis eine Herleitung der Schwellenwerte nachvollziehbar ist, wird vermisst. <u>Vorgeschlagene Änderung:</u> Eine transparente wissenschaftliche Herleitung der Schwellenwerte sollte ergänzt werden.
9.3.13 Umgang mit unterschiedlichen mittleren Beobachtungsdauern (S.192 f.)	<u>Anmerkung:</u> Der Abschnitt 9.3.13 zum Umgang mit unterschiedlichen mittleren Beobachtungsdauern wurde ergänzt. Darin heißt es: „Ist die Ursache für die unterschiedlichen mittleren Beobachtungsdauern eine unvollständige Datenerfassung (Zensierung bei Behandlungsabbruch oder -wechsel), so entsteht allerdings auch bei Anwendung von Verfahren für Überlebenszeiten das Problem informativer Zensierungen, woraus ggf. ein hohes Verzerrungspotenzial folgt. Um solche Probleme zu vermeiden, sollte eine vollständige Datenerfassung der unerwünschten Ereignisse erfolgen, und zwar auch nach einem Behandlungsabbruch oder -wechsel [43].“ Falls anstelle der Auswertung unerwünschter Ereignisse (UE) als binäre Daten Ereigniszeitanalysen durchgeführt werden sollen, ist zu spezifizieren wie mit dem Ereignis Tod zu verfahren ist (Zensierung oder Auswertung als UE-freies Überleben). <u>Vorgeschlagene Änderung:</u> Eine Empfehlung zum Umgang mit konkurrierenden Ereignissen bei Ereigniszeitanalysen unerwünschter Ereignisse ist wünschenswert.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

[1] Cohen J. Statistical Power Analysis in the Behavioral Sciences (2nd edition). Hillsdale (NJ): Lawrence Erlbaum Associates, Inc., 1988

[2] IQWiG Berichte – Nr. 638. Untersuchung der statistischen Eigenschaften von Verfahren zur Übertragbarkeit von Studienergebnissen auf Teilpopulationen, 2018

A.1.35 – SmartStep Consulting GmbH

Autorinnen und Autoren

- Loske, Sonja

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden

Bitte nutzen Sie pro Person 1 Zeile.

Loske, Sonja

Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)

- ☒ **im Namen folgender Institution / Organisation: SmartStep Consulting GmbH**
- ☐ **als Privatperson(en)**

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.

Nicht zutreffend.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
Abschnitt 3.1.3 „Ermittlung des Schadens medizinischer Interventionen“, S. 47 – 50	<u>Anmerkung:</u> Dieser Abschnitt wurde derart umgestaltet, dass bezüglich der Auswertung der möglichen Schadensaspekte die Kausalität zwischen dem aufgetretenen Ereignis und der medizinischen Intervention in den Vordergrund gerückt wird. Während in den Allgemeinen Methoden Version 5.0 hier von unerwünschten Ereignissen und Wirkungen gesprochen wird, die es bei der Ermittlung des Schadens zu berücksichtigen gilt, sind es im Entwurf der Allgemeinen Methoden Version 6.0 spezifisch die unerwünschten Wirkungen. Der Unterschied zwischen unerwünschten Ereignissen und unerwünschten Wirkungen liegt darin begründet, dass bei letzteren ein kausaler Zusammenhang mit der Intervention angenommen wird, während eine derartige Bewertung bei unerwünschten Ereignissen nicht relevant ist. Es wird ferner ausgeführt, dass die Aufarbeitung und Darstellung der unerwünschten Wirkungen auf diejenigen mit besonderer Relevanz begrenzt werden sollte. Die genannten Beispiele umfassen beispielsweise unerwünschte Wirkungen mit möglicher Dosis-Wirkungs-Beziehung oder diejenigen, welche von Patientinnen und Patienten als besonders wichtig angesehen werden. Im Fall der frühen Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V bilden hierfür die unerwünschten Ereignisse, die schwerwiegenden unerwünschten Ereignisse und ggf. die schweren unerwünschten Ereignisse, die gemäß den Anforderungen der Verfahrensordnung (VerfO) des Gemeinsamen Bundesausschusses (G-BA) im Dossier des pharmazeutischen Unternehmers übermittelt werden, die Grundlage. Weitere Kriterien für die Auswahl sollen ferner die Patientenrelevanz der Ereignisse und die Unterschiede zwischen den Behandlungsarmen sein.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	Zusätzlich gibt es Anforderungen an die Mindesthäufigkeit der Ereignisse im Fall der nicht schwerwiegenden bzw. nicht schweren unerwünschten Ereignisse, die im Rahmen der Änderung der Dossievorlage für Modul 4 eingeführt wurden und ab dem 01.04.2020 verbindlich umzusetzen sind. Die oben aufgeführten Eckpunkte des Abschnittes 3.1.3 im Entwurf der Allgemeinen Methoden Version 6.0 werfen aus Sicht der SmartStep Consulting GmbH an mehreren Stellen Fragen auf, so dass die Angaben hier präzisiert werden sollten. Während eingangs betont wird, dass die unerwünschten Wirkungen und damit die Kausalität zwischen Intervention und Ereignis von Relevanz sind, werden wiederum in den Erläuterungen zum Vorgehen für Verfahren der frühen Nutzenbewertung gemäß § 35a SGB V die unerwünschten Ereignisse als Grundlage herangezogen. Als Kriterien für die Auswahl werden hier nur die Patientenrelevanz sowie die Unterschiede zwischen den Behandlungsarmen aufgeführt. Jedoch ist alleine unter Berücksichtigung dieser Kriterien die Herstellung einer Kausalität zwischen den unerwünschten Ereignissen und der Intervention nicht möglich. Auch werden explizit die Mindesthäufigkeiten der Ereignisse für nicht schwerwiegende bzw. nicht schwere unerwünschte Ereignisse angeführt, während dies jedoch gemäß VerO des G-BA beispielsweise auch für schwerwiegende und schwere unerwünschte Ereignisse der Fall ist. Außerdem sollte ergänzt werden wie im Falle von älteren Arzneimitteln, die sich zum ersten Mal der frühen Nutzenbewertung gemäß § 35a SGB V unterziehen müssen (beispielsweise Arzneimittel, die vor dem 01.01.2011 bereits auf dem deutschen Markt verfügbar und bisher nicht zu Lasten der gesetzlichen Krankenversicherung [GKV] ordnungsfähig waren), vorgegangen werden sollte, da es hier ja ein bereits etabliertes Sicherheitsprofil gibt. Hier wäre es beispielsweise denkbar, dass auf Grundlage der Periodic Safety Update Reports (PSUR) entsprechende Auswertungen bezüglich neuer Sicherheitssignale durchgeführt werden könnten. <u>Vorgeschlagene Änderung:</u> Präzisierung der Angaben zur Auswahl der relevanten spezifischen unerwünschten Wirkungen im Rahmen der frühen Nutzenbewertung gemäß § 35a SGB V

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
Abschnitt 9.3.3 „Änderungen zur Beurteilung der klinischen Relevanz“, S. 177 f.	<u>Anmerkung:</u> Mit dem Entwurf der Allgemeinen Methoden Version 6.0 legt das IQWiG pauschale Grenzen für die Minimal Important Difference (MID) im Rahmen von Responderanalysen fest. Diese soll mindestens 15 % der Skalenspannweite des verwendeten Erhebungsinstrumentes entsprechen, wenn die Responderanalysen in der Studie präspezifiziert sind; eine MID von genau 15 % der Skalenspannweite wird im Fall von post-hoc durchgeführten Responderanalysen festgelegt. Das IQWiG leitet die Skalenspannweite aus einer Analyse von systematischen Übersichtsarbeiten zu MIDs ab, die in vielen Fällen zeigen, dass sich diese zwischen 10 und 20 % der Spannweite der Scores eines Erhebungsinstruments bewegen. Aus Sicht der SmartStep Consulting GmbH ist ein derartiger Ansatz im Hinblick auf Responderanalysen unter Verwendung einer MID basierend auf der Skalenspannweite nicht sachgerecht und zu generalisiert. Im Fall des Vorliegens von Validierungsstudien für die relevanten Erhebungsinstrumente einschließlich der Festlegung einer MID liegt eine entsprechende wissenschaftliche Bewertung vor, die spezifisch für das jeweilige Erhebungsinstrument und die untersuchte Population bzw. Indikation ist. Derartige pauschale Grenzen könnten für Erhebungsinstrumente relevant sein, welche (noch) nicht über entsprechende Validierungsstudien verfügen, aber in Fällen, in denen diese vorliegen, stellt eine entsprechende pauschale Grenze einen methodischen Rückschritt dar. Eine generalisierte MID kann nicht eine mittels einer Studie in der relevanten Population bzw. Indikation und spezifisch für das Erhebungsinstrument ermittelten MID ersetzen. Insbesondere stellt sich hier die Frage, inwiefern eine MID von $\geq 15\%$ der Skalenspannweite einer entsprechenden klinischen Relevanz bzw. Patientenrelevanz gleichzusetzen ist. <u>Vorgeschlagene Änderung:</u> Berücksichtigung der für ein spezifisches Erhebungsinstrument im Rahmen einer Validierungsstudie ermittelten MID und Nutzung einer generalisierten MID für Erhebungsinstrumente ohne Vorliegen entsprechender Validierungsstudien

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

Es wurde keine Literatur zitiert, daher entfallen Angaben zum Literaturverzeichnis.

A.1.36 – Takeda Pharma Vertrieb GmbH & Co. KG

Autorinnen und Autoren

- Fanter, Lena
- Hahn, Andreas
- Ratsch, Boris

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Hahn, Andreas; Dr.
Fanter, Lena; Dr.
Ratsch, Boris; Dr.
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
☒ im Namen folgender Institution / Organisation: Takeda Pharma Vertrieb GmbH & Co. KG
☐ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

<i>Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.</i>

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung																
3.3.3 (S. 67 f.)	<i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i> <u>Anmerkung:</u> Zur Quantifizierung eines Zusatznutzens für stetige und quasistetige Zielgrößen definiert das IQWiG Schwellenwerte für SMD wie Hedges' g oder Cohens d in Anlehnung an die Schwellenwerte für binäre Zielgrößen. Tabelle 6: Schwellenwerte zur Ausmaßbestimmung für die SMD <table><tr><th colspan="2" rowspan="2"></th><th colspan="2">Zielgrößenkategorie</th></tr><tr><th>schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen und gesundheitsbezogene Lebensqualität^a</th><th>nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen</th></tr><tr><th rowspan="3">Ausmaß- kategorie</th><th>erheblich</th><td>0,6</td><td>nicht besetzt</td></tr><tr><th>beträchtlich</th><td>0,3</td><td>0,5</td></tr><tr><th>gering</th><td>0,2</td><td>0,2</td></tr></table> a: Voraussetzung ist – wie für alle patientenberichteten Endpunkte – die Verwendung eines validierten bzw. etablierten Instruments. Entsprechend sind die Schwellenwerte für schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen und gesundheitsbezogene Lebensqualität sowie für nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen getrennt ausgewiesen. Diese Schwellenwerte beziehen sich wie auch bei den binären Zielgrößen auf die untere Grenze des 0,95-Konfidenzintervalls und dürfen in jenem nicht enthalten sein. Zur Herleitung der Schwellenwerte zitiert das IQWiG das Verfahren für binäre Daten nach Skipka et al. [646]. Hieraus lassen sich jedoch nicht die genannten Schwellenwerte für schwerwiegende (bzw. schwere)			Zielgrößenkategorie		schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen und gesundheitsbezogene Lebensqualität ^a	nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen	Ausmaß- kategorie	erheblich	0,6	nicht besetzt	beträchtlich	0,3	0,5	gering	0,2	0,2
				Zielgrößenkategorie													
		schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen und gesundheitsbezogene Lebensqualität ^a	nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen														
Ausmaß- kategorie	erheblich	0,6	nicht besetzt														
	beträchtlich	0,3	0,5														
	gering	0,2	0,2														

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	Symptome (bzw. Folgekomplikationen) und Nebenwirkungen und gesundheitsbezogene Lebensqualität ableiten. Die Herleitung ist somit insgesamt unvollständig und zumindest in Teilen willkürlich. <u>Vorgeschlagene Änderung:</u> Die Herleitung der Schwellenwerte für eine Quantifizierung des Zusatznutzens sollte vollständig auf geeigneter wissenschaftlicher Literatur beruhen.
<u>9.3.3</u> (S. 175 f.)	<u>Anmerkung:</u> Das IQWiG definiert eine neue Vorgehensweise vor dem Hintergrund, dass es derzeit keine allgemein akzeptierte Methodik zur Validierung von MIDs gäbe. Es definiert deshalb ein pauschales Responderkriterium von mindestens 15% der Skalenspannweite. Liegen präspezifizierte Responsekriterien im Sinne einer MID unterhalb dieser Grenze vor, sollen diese regelhaft nicht herangezogen werden. Werde ein Responsekriterium post hoc definiert, so müsse dieses genau 15% der Skalenspannweite betragen. Lügen sowohl Responseanalysen als auch Analysen der stetigen Daten vor, so würden die Responseanalysen herangezogen, so sie den definierten Kriterien entsprächen. Zur Herleitung der angeführten Schwelle verweist das IQWiG auf „eine Sichtung von systematischen Übersichtsarbeiten“ wobei unklar bleibt ob diese einer vorab definierten Systematik oder einer ergebnisgesteuerten Intention folgte. Die einzigen vom IQWiG angegebenen Quellen (Nordin 2016, St-Pierre 2016) nehmen zudem nur auf ausgewählte und sehr spezifische Symptome wie „fatigue“ und „symptoms and functional limitations in individuals with rotator cuff disorders“ Bezug und sind deshalb kaum zu verallgemeinern. Das IQWiG argumentiert, dass „sich MIDs in vielen Fällen zwischen 10 % und 20 % der Spannweite der Scores eines Erhebungsinstruments bewegen“ würden ohne dies konkret zu belegen. Auch die nachfolgend festgelegte Schwelle von 15% der Skalenspannweite wird nicht begründet, sodass deren Rationale unklar bleibt. Die willkürliche Festlegung von MIDs ist deshalb zu kritisieren, da diese Vorgehensweise insbesondere die Bedürfnisse der betroffenen Patienten nicht berücksichtigt. <u>Vorgeschlagene Änderung:</u>

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	In Anerkennung der Situation, dass es derzeit keine allgemein akzeptierte Methodik zur Validierung von MIDs gibt, sollte in Abstimmung zwischen IQWiG, Akademie und pharmazeutischer Industrie ein wissenschaftlich fundierter und verbindlicher Kriterienkatalog entworfen werden, welcher dann im Rahmen der Nutzenbewertung Anwendung finden sollte.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

A.1.37 – UCB Pharma GmbH

Autorinnen und Autoren

- Andreas, Jens-Otto

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Andreas, Jens-Otto
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
☒ im Namen folgender Institution / Organisation: UCB Pharma GmbH
☐ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
Abschnitt 3.3.3 (Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V) (Seite 67)	<u>Anmerkung:</u> Das IQWiG schreibt in seinem Entwurf zum allgemeinen Methodenpapier im Abschnitt 3.3.3 zur Quantifizierung eines Zusatznutzens bei stetigen oder quasistetigen Zielgrößen mit jeweils vorliegenden Standardized Mean Difference (standardisierte Mittelwertdifferenz, SMD): <i>„In Anlehnung an das Verfahren für binäre Daten nach Skipka et al. [646] lassen sich auch Schwellenwerte für das Effektmaß SMD (Cohen's d bzw. Hedges' g) ableiten. Man kann zeigen, dass die für ein bestimmtes Ausmaß wünschenswerte Effektstärke ungefähr halbiert werden muss, um den entsprechenden Schwellenwert abzuleiten. Als Orientierung für die theoretisch wünschenswerten Effektstärken bei schwerwiegenden Symptomen können die Mittelwerte der Bereiche für die übliche Einteilung von Cohen's in kleine (SMD zwischen 0,2 und 0,5) mittlere (SMD zwischen 0,5 und 0,8) und große Effekte (SMD > 0,8) dienen. Tabelle 6 gibt einen Überblick über die verwendeten Schwellenwerte. Im Gegensatz zu den Schwellenwerten für die relativen Maße ist bei der SMD zu beachten, dass das Konfidenzintervall vollständig oberhalb des entsprechenden Schwellenwerts liegen muss und der Nulleffekt hier nicht durch die Zahl 1 sondern durch die Zahl 0 gegeben ist.“</i> UCB begrüßt die Intention des IQWiGs einer Quantifizierung eines Zusatznutzens für stetige und quasistetige Zielgrößen. Jedoch ist das vorgeschlagene Vorgehen nicht ausreichend wissenschaftlich begründet und nur teilweise nachvollziehbar. So lassen sich lediglich die Schwellenwerte der Zielkategorie für nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	Nebenwirkungen auf Basis des Textes herleiten (0,2 und 0,5). Für die Werte der Zielkategorie schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen und gesundheitsbezogene Lebensqualität fehlt eine Herleitung. Somit sind die beschriebenen Mittelwerte der Bereiche in der Tabelle 6 des Entwurfs des Allgemeinen Methodenpapiers nicht nachvollziehbar. Des Weiteren wird erläutert, dass das Konfidenzintervall für die SMD vollständig oberhalb des entsprechenden Schwellenwerts liegen muss, es wird jedoch kein Konfidenzniveau angegeben. Es ist davon auszugehen, dass hier das 2-seitige 95 %-Konfidenzintervall gemeint ist. Die Berücksichtigung der Konfidenzintervallgrenzen stellt einen weit konservativeren Ansatz dar, als der nach Cohen's vorgeschlagene Vergleich mit dem Punktschätzer, der nach jetzigem wissenschaftlichem Stand auch methodisch belegt ist [1]. <u>Vorgeschlagene Änderung:</u> UCB hält die Grundlage für die vom IQWiG vorgeschlagenen Änderungen für nicht ausreichend, so dass eine Änderung der bestehenden Regelungen in dieser Form nicht erfolgen sollte. Außerdem schlägt UCB einen regelmäßigen Austausch z.B. im Rahmen von Workshops vor, um gemeinsam mit IQWiG, G-BA, universitären Einrichtungen und Industrie wissenschaftlich fundierte Rationale für Themen wie die Quantifizierung des Zusatznutzens von stetigen und quasistetigen Zielgrößen zu erörtern und festlegen zu können. Grundsätzlich sollte die Quantifizierung des Zusatznutzens auf den von Cohen angegebenen Schwellenwerten für Punktschätzer beruhen.
Abschnitt 9.3.3 (Beurteilung klinischer Relevanz) (S.175 - 178)	<u>Anmerkung:</u> UCB begrüßt, dass sich das IQWiG mit der wichtigen Thematik befasst, welche Methodik zur Ableitung einer klinisch relevanten Minimal Important Difference (MID) skalenbasierter Erhebungsinstrumente angewendet werden sollte. Die Bedeutung dieses Themas zeigt sich auch darin, dass es z.Zt. zahlreiche internationale Initiativen gibt, die sich mit der Vereinheitlichung von Kriterien für aussagkräftige und wissenschaftlich valide MIDs beschäftigen [2,3]. Das nun vom IQWiG vorgeschlagene Vorgehen, eine pauschale MID-Schwelle von 15 % der theoretisch möglichen Skalenspannweite für alle Erhebungsinstrumente anzuwenden, ist aus Sicht von UCB wissenschaftlich nicht begründet und für UCB daher nicht nachvollziehbar.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	Zur Herleitung dieser Schwelle bezieht sich das IQWiG auf „[...] eine Sichtung von systematischen Übersichtsarbeiten“ (S. 178). Hier bleibt unklar, ob diese Sichtung vollständig ist, da sich die beiden einzigen vom IQWiG exemplarisch angegebenen Quellen (Nordin et al., 2016 und St-Pierre et al., 2016) lediglich auf ausgewählte und sehr spezifische Symptome wie „fatigue“ und „symptoms and functional limitations in individuals with rotator cuff disorders“ beziehen. In einem nächsten Schritt wird argumentiert, dass „[...] sich MIDs in vielen Fällen zwischen 10 % und 20 % der Spannweite der Scores eines Erhebungsinstruments bewegen“ (S. 178). Auch hier bleibt unklar, was unter „vielen Fällen“ zu verstehen ist und wie repräsentativ die angeführten Grenzen von 10 % bzw. 20 % sind. Anschließend wird auf dieser Basis unklarer Evidenz die 15 %-Schwelle für alle MIDs abgeleitet. Dieser Wert könnte sich als Mittelwert der o.g. Grenzen von 10 % und 20 % ergeben - eine konkrete Begründung für die Wahl von 15 % liegt allerdings nicht vor. Insgesamt stellt sich die Frage, ob eine verteilungsbasierte und pauschale Schwelle für die MIDs aller Erhebungsinstrumente überhaupt sinnvoll sein kann. In der auch vom IQWiG zitierten Arbeit von Revicki et al. 2008 [4] heißt es dazu: <i>„The MID for a PRO instrument is not an immutable characteristic, but may vary by population and context, and no one MID may be valid for all study applications. MID estimates should be based on multiple approaches and triangulation of methods. Anchor-based methods applying various relevant patient-rated, clinician-rated, and disease-specific variables provide primary and meaningful estimates of an instrument’s MID.“</i> Die Autoren kommen in ihrer Arbeit daher zu folgender Empfehlung: <i>„We recommend that the MID is based primarily on relevant patient-based and clinical anchors, with clinical trial experience used to further inform understanding of MID.“</i> Dieser Empfehlung schließt sich UCB an, da nur durch die Berücksichtigung von Patienten- und/oder Arzt-basierten Ankern sichergestellt werden kann, dass die abgeleitete MID klinisch relevant ist. Eine von der vfa Arbeitsgruppe Biostatistik vorgenommene Auswertung auf Basis von bisher in AMNOG-Verfahren akzeptierten MIDs zeigt zudem, dass die vorgeschlagene 15 %-Schwelle in ca. 87 % der Fälle zu einer Erhöhung der bisher validierten und in der Anwendung akzeptierten MIDs führen würde. Auch für weitere Skalen würden bereits breit etablierte und publizierte MIDs durch das IQWiG möglicherweise nicht mehr akzeptiert werden. Im Falle der neurodegenerativen Erkrankung Progressive Supranukleäre Blickparese (PSP) beispielsweise liegt

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	für die geläufige PSP-Rating-Skale (PSP-RS) eine publizierte MID von 5,7 vor [5]. Die durchschnittliche Verschlechterung von Patienten beträgt jährlich ca. 10 Punkte, wobei der Gesamtscore Werte von 0–100 erreichen kann [5]. Die Anwendung einer 15% Schwelle auf die PSP-RS würde demnach zu einer MID von 15 Punkten führen, die mehr als eine Verdopplung der publizierten MID darstellt und zudem die jährliche Verschlechterung im natürlichen Verlauf übersteigt. In diesem speziellen Fall und in allen anderen ähnlich gelagerten Fällen, wäre es daher nicht nachvollziehbar, warum etablierte bzw. publizierte MIDs, deren klinische Relevanz bereits belegt ist, im Rahmen der Nutzenbewertung nicht verwendet werden sollten. <u>Vorgeschlagene Änderung:</u> Da es sich bei der Validierung von MIDs um ein komplexes Problem handelt, dass nicht mit einer pauschalen Herangehensweise gelöst werden kann, sollten aus Sicht von UCB die vom IQWiG vorgeschlagenen Änderungen in dieser Form nicht implementiert werden. Stattdessen schlägt UCB einen konstruktiven Dialog z.B. im Rahmen von Workshops zwischen IQWiG, G-BA, universitären Einrichtungen und Industrie vor. Das Ziel dieses Dialogs sollte sein, einen internationalen und wissenschaftlich fundierten Kriterienkatalog zu etablieren, der für die Akzeptanz von MIDs im Rahmen von Nutzenbewertungen regelhaft anzuwenden ist.
Abschnitt 9.3.8 (Metaanalysen) (Seite 183)	<u>Anmerkung:</u> In Kapitel 9.3.8 zu Metaanalysen wird auf bayessche Verfahren im Falle weniger (weniger als 5) Studien hingewiesen: „Ist die Anwendung des Modells mit festem Effekt nicht vertretbar, so kann eine qualitative Evidenzsynthese erfolgen. Kontextabhängig kommen auch alternative Verfahren in Betracht, wie z. B. bayessche Verfahren [...]“ (S. 183) Bayesschen Verfahren im Rahmen der Nutzenbewertung wurden bereits auf der GMDS-Jahrestagung 2019 in Dortmund mit Teilnehmern des IQWiG diskutiert. Da viele Dossiers nur sehr wenige Studien berücksichtigen können, ist eine Konkretisierung dieses Verfahren wünschenswert. <u>Vorgeschlagene Änderung:</u> Aus Sicht von UCB wäre eine Konkretisierung der Akzeptanz der Anwendung von bayesschen Verfahren wünschenswert. Dieses schließt insbesondere die Herleitung und Nutzung von „Prior-Verteilungen“ ein.
Abschnitt 9.3.12 (Umgang mit	<u>Anmerkung:</u> Das IQWiG schreibt in seinem Entwurf zum allgemeinen Methodenpapier im Abschnitt 9.3.12 zum Umgang mit

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
unvollständigen Daten) (Seite 191)	unvollständigen Daten „Das Verzerrungspotenzial steigt mit steigendem Anteil fehlender Werte. Ergebnisse fließen daher in der Regel nicht in die Nutzenbewertung ein, wenn diese auf weniger als 70 % der in die Auswertung einzuschließenden Patientinnen und Patienten basieren, das heißt, wenn der Anteil der Patientinnen und Patienten, die nicht in der Auswertung berücksichtigt werden, größer als 30 % ist. In der Literatur werden zum Teil bereits Auswertungen, in denen 20 % der Patientinnen und Patienten nicht berücksichtigt werden, als nicht mehr aussagekräftig betrachtet [618]. [...] Die Ergebnisse werden ferner auch dann nicht in die Nutzenbewertung einbezogen, wenn der Unterschied der Anteile nicht berücksichtigter Patientinnen und Patienten zwischen den Gruppen größer als 15 Prozentpunkte ist.“ (S. 192) Sowohl in den Guidelines der EMA wie auch der FDA werden keine Richtwerte zur Einschätzung der Anteile fehlender Werte genannt. Die hier zitierte Quelle stammt aus dem Jahr 2002. Im National Research Council Report von 2010, der das zu diesem Zeitpunkt vorhandene wissenschaftliche Know-How über fehlende Daten zusammenfasst, wird konstatiert: <i>„We note that one cannot be specific as to how to set target and maximally acceptable rates for missing data in all clinical trials. The amount of acceptable missing data will depend on many characteristics of the trial, including whether the assumption that the missing data are missing at random is reasonable, the size of the anticipated effect of the intervention under study, and the likelihood that a sensitivity analysis would render the results of the trial inconclusive. Applied research is needed on this topic, and techniques will need to evolve on the basis of that research.“ [6]</i> Aus Sicht von UCB ist diese Einschätzung auch heute noch gültig. Zwar benennt die Referenz vom IQWiG für den Anteil der fehlenden Werte einen Richtwert von 20 % der einzuschließenden Patientinnen und Patienten, jedoch kommen die Autoren auch zu dem Schluss, dass z.B. eine genaue Beschreibung des Patientenflusses bei der Interpretation hilft. Für den Richtwert des Unterschieds zwischen den Gruppen mit 15 Prozentpunkten nennt das IQWiG keine Referenz. UCB begrüßt die Intention, Richtwerte zur Einschätzung des Verzerrungspotenzials für den Anteil fehlender Werte zu quantifizieren, jedoch sollten diese Aussagen und Schlussfolgerungen wissenschaftlich fundiert und mit Referenzen belegt sein.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	<u>Vorgeschlagene Änderung:</u> UCB hält die Grundlage für die vom IQWiG vorgeschlagenen Änderungen für nicht ausreichend, so dass eine Änderung der bestehenden Regelungen in dieser Form nicht erfolgen sollte. Gleichzeitig schlägt UCB einen regelmäßigen Austausch vor, um gemeinsam mit IQWiG, G-BA, universitären Einrichtungen und Industrie wissenschaftlich fundierte Richtlinien für den Umgang mit unvollständigen Daten zu erörtern und festlegen zu können.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

- [1] Cohen J. Statistical Power Analysis in the Behavioral Sciences (2nd edition). Hillsdale (NJ): Lawrence Erlbaum Associates, Inc., 1988
- [2] Musoro ZJ, Hamel J, Ediebah DE on Behalf of the EORTC Quality of Life Group, et al. Establishing anchor-based minimally important differences (MID) with the EORTC quality-of-life measures: a meta-analysis protocol. BMJ Open 2018;8:e019117
- [3] Coon C, Cook K. Moving from significance to real-world meaning: methods for interpreting change in clinical outcome assessment scores; Qual Life Res 2018; 27:33–40
- [4] Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008; 61(2): 102-109
- [5] Hewer S, Varley S, Boxer AL, Paul E, Williams DR. Minimal clinically important worsening on the progressive supranuclear Palsy Rating Scale. Mov Disord 2016; 31(10):1574-7.
- [6] National Research Council (US) Panel on Handling Missing Data in Clinical Trials. The Prevention and Treatment of Missing Data in Clinical Trials. Washington (DC): National Academies Press (US) 2010; 3, Trial Strategies to Reduce the Frequency of Missing Data. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK209896/>

A.1.38 – Verband Forschender Arzneimittelhersteller e. V. (vfa)

Autorinnen und Autoren

- Rasch, Andrej
- Werner, Sebastian

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Werner, Sebastian; Dr.
Rasch, Andrej; Dr.
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
☒ im Namen folgender Institution / Organisation: Verband Forschender Arzneimittelhersteller e.V. - vfa
☐ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

<i>Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.</i>
In den Änderungen im Entwurf der allgemeinen Methoden 6.0 ergeben sich an mehreren Stellen wichtige Änderungen für die Nutzenbewertung von Arzneimitteln, die für die forschenden Pharmaunternehmen von Bedeutung sind, zu denen im Detail Stellung genommen wird. Der Verband forschender Arzneimittelhersteller (vfa) beobachtet insbesondere die angedachten Änderungen des Umgangs mit MID-Responder-Analysen mit großer Sorge. Sie stellen nichts weniger als eine vollständige Neuordnung des methodischen Vorgehens dar, die einen klaren Bruch mit der bewährten Bewertungspraxis bedeuten und wissenschaftlich kaum nachzuvollziehen sind. Der Vorschlag würde zu einer regelhaften Erhöhung von validierten bzw. etablierten Schwellenwerten führen, die zuvor als ausreichend galten, um die für die Patientinnen und Patienten spürbaren Veränderungen hinreichend sicher abzubilden. Ein solcher methodischer Ansatz birgt Gefahren für die Verfügbarkeit von neuen Arzneimitteln in Deutschland, deren schwerpunktmäßiger Nutznachweis auf der Grundlage von patientenberichteten Endpunkten erbracht wird. Änderungen mit einer solchen Tragweite sollten nur vorgeschlagen werden, wenn ihnen ein intensiver Austausch der betroffenen Stakeholder aus Wissenschaft, Praxis und pharmazeutischer Industrie vorausgegangen ist. Der vfa bedauert sehr, dass ein solcher Austausch nicht stattgefunden hat und kann deshalb den Vorschlag auch nicht unterstützen. Der vfa steht aber mit seinen Experten für einen solchen Dialog mit dem Ziel einer Konsensbildung in Zukunft gern zur Verfügung.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
Abschnitt 3.3.3 „Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V“, S. 67	<u>Anmerkung:</u> <i>Im Abschnitt 3.3.3 wird die Methodik zur Quantifizierung der Effektausmaße um die Vorgehensweise bei stetigen oder quasistetigen Zielgrößen mit standardisierten Mittelwertdifferenzen (SMDs) ergänzt. Das IQWiG definiert hierfür Schwellenwerte zur Ausmaßbestimmung bei SMDs in Abhängigkeit von der Zielgrößenkategorie. Das IQWiG führt aus: „In Anlehnung an das Verfahren für binäre Daten nach Skipka et al. [646] lassen sich auch Schwellenwerte für das Effektmaß SMD (Cohen's d bzw. Hedges' g) ableiten. Man kann zeigen, dass die für ein bestimmtes</i>

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<i>Ausmaß wünschenswerte Effektstärke ungefähr halbiert werden muss, um den entsprechenden Schwellenwert abzuleiten. Als Orientierung für die theoretisch wünschenswerten Effektstärken bei schwerwiegenden Symptomen können die Mittelwerte der Bereiche für die übliche Einteilung von Cohen's d in kleine (SMD zwischen 0,2 und 0,5) mittlere (SMD zwischen 0,5 und 0,8) und große Effekte (SMD > 0,8) dienen. Tabelle 6 gibt einen Überblick über die verwendeten Schwellenwerte. Im Gegensatz zu den Schwellenwerten für die relativen Maße ist bei der SMD zu beachten, dass das Konfidenzintervall vollständig oberhalb des entsprechenden Schwellenwerts liegen muss und der Nulleffekt hier nicht durch die Zahl 1 sondern durch die Zahl 0 gegeben ist.“</i> Es ist grundsätzlich zu begrüßen, dass das IQWiG ein methodisches Vorgehen anstrebt, welches dem Ziel dient, auch bei stetigen Daten eine Ausmaß-Quantifizierung der Effektgrößen zu ermöglichen. Jedoch müssen die festgelegten Schwellenwerte und ihre Herleitung grundsätzlich hinterfragt werden, da diese teilweise nicht nachvollziehbar sind bzw. nicht wissenschaftlich begründet wurden. So lassen sich lediglich die Schwellenwerte der Zielgrößenkategorie für „nicht schwerwiegende Symptome und Nebenwirkungen“ auf Basis der üblichen Einteilung von Cohen's d nachvollziehbar herleiten [1]. Diese sehen einen kleinen Effekt mit „0,2“ (Ausmaßkategorie gering) und für einen mittleren Effekt „0,5“ (Ausmaßkategorie beträchtlich) vor. Im Gegensatz zu den Schwellenwerten der o. g. Zielgrößenkategorie sind die Schwellenwerte in der Zielkategorie „schwerwiegende Symptome, Nebenwirkungen und gesundheitsbezogene Lebensqualität“ nicht nachvollziehbar und nicht wissenschaftlich begründet. Für die Ableitung der entsprechenden Schwellenwerte müssten laut IQWiG die Mittelwerte der üblichen Cohen's d-Bereiche gebildet und danach ungefähr halbiert werden. Für dieses Vorgehen findet sich keine wissenschaftliche Rationale. Aus Sicht des vfa entspricht es damit auch nicht den international anerkannten Kriterien oder Standards der evidenzbasierten Medizin. Zugleich ist nicht nachvollziehbar, warum eine Unterteilung in unterschiedliche Zielgrößenkategorien überhaupt erforderlich sein sollte. Eine Anwendung der üblichen Einteilung nach Cohen wäre für beide Zielgrößenkategorien in gleicher Weise sinnvoll durchführbar, ohne die gängige Einteilung abwandeln zu müssen. Das geschilderte Vorgehen in Verbindung mit den unzureichenden Begründungen erweckt zudem den Eindruck, dass das IQWiG hier mittels einer willkürlich erscheinenden mathematischen Festlegung

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	eine normative Entscheidung getroffen hat, die zu einer impliziten Gewichtung der Zielgrößenkategorien führen soll. Darüber hinaus soll laut IQWiG bei der Ableitung des Ausmaßes des Zusatznutzens gelten, dass das gesamte Konfidenzintervall vollständig oberhalb des entsprechenden Schwellenwerts liegen muss (sog. Verschiebung der Nullhypothese). Diese Berücksichtigung der Konfidenzintervallgrenzen stellt einen weit konservativeren methodischen Ansatz dar als der nach Cohen vorgeschlagene Vergleich mit dem Punktschätzer [1]. Die vom IQWiG vorgeschlagenen Schwellenwerte werden daher faktisch in der Regel weit höher liegen, als die nach Cohen vorgeschlagenen Werte. In der Praxis kann die Orientierung an Konfidenzintervallgrenzen dazu führen, dass die tatsächlich beobachteten Punktschätzer 2- bis 3-mal über den nach Cohen vorgeschlagenen Werten (z. B. „0,2“) liegen können. Damit lägen die Punktschätzer in einem Bereich, der nach Cohen bereits als mittlerer Effekt gilt, jedoch nach IQWiG-Ansatz als geringer Effekt klassifiziert werden müsste (so z. B. bei einem Hedges' g von -0,67 mit einem 95 %-Konfidenzintervall von -1,08 bis -0,25 für den Endpunkt „Schmerz durch Spastik (NRS)“ in der Nutzenbewertung A18-27) [7]. Mit diesem konservativen Ansatz werden so die Hürden, einen relevanten und quantifizierbaren Vorteil zeigen zu können, deutlich erhöht. <u>Vorgeschlagene Änderung:</u> Die Schwellenwerte zur Ausmaßbestimmung bei stetigen Daten sollten nachvollziehbar sein, wissenschaftlich begründet werden können und den international anerkannten Standards und Kriterien der evidenzbasierten Medizin genügen. Deshalb sollte die übliche und allgemein akzeptierte Einteilung für Cohen' d angewendet und nicht modifiziert werden. Zudem sollte berücksichtigt werden, dass Schwellenwerte nach Cohen für den Punktschätzer gelten und nicht für das Konfidenzintervall intendiert sind.
Abschnitt 9.3.3 „Beurteilung klinischer Relevanz“, S. 177/178	<u>Anmerkung:</u> Der Abschnitt 9.3.3 wurde an mehreren Stellen ergänzt: ▪ <i>Bisher (im IQWiG-Methodenpapier 5.0) ist bei der Anwendung von Responderanalysen „ein validiertes bzw. etabliertes Responsekriterium“ (im Sinne einer individuellen Minimal important Difference [MID]) erforderlich, um einen Nachweis eines relevanten Effekts zu erbringen.</i> ▪ <i>Das IQWiG ändert diese Anforderungen und schlägt stattdessen ein Responsekriterium vor, welches „mindestens 15 % der Skalenspannweite des verwendeten</i>

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<i>Erhebungsinstruments“ entspricht und zudem „präspezifiziert“ sein muss.</i> ▪ <i>Falls präspezifizierte Responsekriterien im Sinne einer MID unterhalb von 15 % der Skalenspannweite liegen, sollen diese nicht herangezogen werden. In diesen Fällen und solchen, in denen gar keine Responsekriterien präspezifiziert wurden, soll entweder auf stetige Daten mit standardisierten Mittelwertdifferenzen (SMDs, in Form von Hedges' g) zurückgegriffen werden, oder alternativ post hoc spezifizierte Analysen mit einem Responsekriterium von genau 15 % des Skalenrangs berücksichtigt werden.</i> ▪ <i>Ausgangspunkt für die Änderung sei, dass es derzeit keine akzeptierten Standards gäbe, mit denen die Qualität von Studien bewertet und die Aussagekraft der ermittelten MIDs abgeschätzt werden könne.</i> ▪ <i>Begründet wird diese Festlegung mit zwei Übersichtsarbeiten zu MIDs, die sich in vielen Fällen zwischen 10 % und 20 % der Spannweite der Scores eines Erhebungsinstruments bewegen würden.</i> Der vfa lehnt diese Änderungen ab. Die vorgeschlagene „präspezifizierte 15%-Mindestresponderschwelle“ erscheint willkürlich festgelegt und kann nicht als allgemein akzeptiert oder etabliert gelten. Die Schwelle ist folglich auch nicht für die Nutzenbewertung von Arzneimitteln geeignet, da hierfür der aktuelle Stand der wissenschaftlichen Erkenntnisse sowie die international anerkannten Kriterien und Standards der evidenzbasierten Medizin maßgeblich sind. Die Einführung der Schwelle würde zu einem klaren Bruch mit der bewährten Bewertungspraxis in der AMNOG-Nutzenbewertung führen und zu einer regelhaften Erhöhung der validierten und etablierten Schwellenwerte führen, die zuvor als ausreichend galten, um die für die Patientinnen und Patienten spürbaren Veränderungen hinreichend sicher abzubilden. Die Schwelle kann daher als deutlich konservativer angesehen werden. Eine vfa-Überprüfung der bisher (bis zum 15.02.2019) in AMNOG-Verfahren als validiert bzw. etabliert akzeptierten MIDs ergibt, dass eine Anwendung der „15%-Responderschwelle“ bei ca. 90 % der MIDs zu einer teils deutlichen Erhöhung des Schwellenwerts führen würde. Darunter sind die MIDs von etablierten Fragebögen, wie z. B. vom SF-36, EORTC-QLQ (-C30, -LC-13, -BR23, -CLL16, -MY20), EQ-5D VAS, FACT-B, FACT-G, FACT-Lym, FACT-P, CAT, TDI, SGRQ, AQLQ, FACIT-F, FKS-DRS u.v.m. Die Erhöhung variiert dabei zwischen Faktor 1,1 und 8,3 sowie im Mittel 2,5. Im Ergebnis würden so fast alle bisher in der AMNOG-Nutzenbewertung als validiert bzw. etabliert geltenden MIDs

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	abgelehnt und die Schwellen der Akzeptanz erhöht. Diese Erhöhung kann den Nachweis eines relevanten Effekts in patientenberichteter Symptomatik oder Lebensqualität deutlich erschweren. Ein solcher methodischer Ansatz birgt damit auch Gefahren für die Verfügbarkeit von neuen Arzneimitteln in Deutschland, deren schwerpunktmäßiger Nutznachweis auf patientenberichteten Endpunkten basiert. Zudem könnte mit einer solchen Schwelle ein Bedeutungsverlust von patientenberichteten Endpunkten einhergehen. Die Schwelle konterkariert damit auch den Anspruch der Nutzenbewertung, bedeutsame Ergebnisse für solche Endpunkte, insbesondere zur Lebensqualität, zu erhalten. Die große Variabilität der beobachteten Erhöhung (zwischen Faktor 1,1 und 8,3) zeigt zudem, dass die vorgeschlagene, generische Responderschwelle „15%“ den bestehenden Besonderheiten unterschiedlicher Therapiegebiete und Skalencharakteristika nicht gerecht werden kann. Eine „one-size-fits-all“-Lösung würde so die vorhandenen Erkenntnisse zu unterschiedlichen Skalenspezifika und krankheitsspezifischen Patientencharakteristika gänzlich ignorieren. Hinsichtlich der vom IQWiG vorgeschlagenen Erforderlichkeit einer Präspezifizierung von MID-Responderanalysen für deren Akzeptanz sollte berücksichtigt werden, dass es Situationen geben kann, in denen Responderanalysen einen relevanten Informationsgewinn liefern können, auch wenn diese nicht präspezifiziert sind. Die Voraussetzung einer Präspezifizierung erscheint daher mit Blick auf die Konsequenzen (d.h. Nicht-Berücksichtigung der Ergebnisse bei fehlender Präspezifizierung) unverhältnismäßig. Auch wird dieses Vorgehen dem gängigen Zusammenspiel von Nutzenbewertung und Zulassung nicht gerecht, da im Rahmen der Zulassung andere Analysetechniken bei patientenberichteten Endpunkten im Vordergrund stehen können. Eine fehlende Präspezifizierung sollte deshalb nicht zu einem generellen Ausschluss der entsprechenden Ergebnisse führen. Abschließend müssen der IQWiG-Ansatz und die Art und Weise des gewählten Vorgehens bei der Festlegung der neuen Bewertungsmethodik hinterfragt werden: So sei der Ausgangspunkt für die Änderung der MID-Bewertungsmethodik nach Aussagen des IQWiG, dass es derzeit keine akzeptierten Standards gäbe, mit denen die Qualität und Aussagekraft der ermittelten MIDs abgeschätzt werden könnten. Zwar gäbe es Bestrebungen der Wissenschaft, solche Standards zu entwickeln, jedoch würden erste Ergebnisse noch einige Fragen offenlassen [2,3]. Davon völlig unabhängig entwickelt das IQWiG dann einen Bewertungsansatz ohne jeglichen Bezug zu den dort diskutieren Kriterien. So weist das IQWiG selbst darauf hin, dass

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	grundsätzlich für die Bestimmung von MIDs empirische Verfahren präferiert werden, die eine MID mithilfe patientenberichteter Anker ermitteln. Auch finden sich weitere mögliche Kriterien zu Studiendesign, Ankerwahl oder psychometrischen Methoden in der wissenschaftlichen Diskussion [4]. Davon unabhängig bestimmt das IQWiG unter Angabe von zwei Übersichtsarbeiten zu MIDs [5,6] eine Schwelle von „15 %“, weil nach Aussagen des IQWiG die Spannweite der Scores in vielen Fällen sich zwischen 10 % und 20 % bewegen würden. Der gewählte Ansatz zur Auswahl der Schwelle „15 %“ (z. B. im Gegensatz zu „10%“) erscheint aus Sicht des vfa willkürlich. Darüber hinaus kann auch angezweifelt werden, dass die Entwicklung der Schwelle mit Blick auf die Generalisierbarkeit der möglichen Therapiegebiete und Skalenspezifika ausreichend ist, da sich die aufgeführten Übersichtsartikel auf eine spezifische Symptomatik (Fatigue) [5] und spezielle Therapiegebiete (rotator cuff disorders) [6] beziehen. Insgesamt ist aus Sicht des vfa der Vorschlag zur MID-Methodik wissenschaftlich nicht hinreichend begründet. Das IQWiG greift bestehende Kriterien aus der wissenschaftlichen Diskussion nicht auf, wählt eine Schwelle auf Grundlage einer arbiträren Festlegung und berücksichtigt dabei die Besonderheiten der Therapiegebiete und Skalenspezifika nicht hinreichend. Eine solche Schwelle kann deshalb aus Sicht des vfa nicht zum international anerkannten Standard der evidenzbasierten Medizin zugerechnet werden. <u>Vorgeschlagene Änderungen:</u> Die vorgeschlagenen Änderungen zur MID-Bewertungsmethodik sind wissenschaftlich nicht hinreichend begründet und sollten gänzlich entfallen. Eine mögliche Änderung der Methodik darf nicht zur Ablehnung von Responseschwellen führen, die bisher als validiert bzw. etabliert galten. Aus Sicht des vfa sollten mögliche neue Standards zur MID-Bewertung nur nach intensivem Austausch mit Vertretern aus Wissenschaft, Praxis und pharmazeutischer Industrie entwickelt werden. Der vfa ist hierbei offen für einen Dialog. Der Austausch könnte im Rahmen einer Serie von Kolloquien und Workshops erfolgen, in denen ein wissenschaftlicher Konsens angestrebt werden sollte. Erst danach sollte eine Änderung des Methodenpapiers in der Sache angestrebt werden.
Abschnitt 9.3.12 „Umgang mit unvollständigen Daten“, S. 191-192	<u>Anmerkung:</u> <i>Im neu verfassten Abschnitt 9.3.12 wird das Vorgehen zum Umgang mit fehlenden Daten aufgrund von Follow-up-Verlusten beschrieben. Demnach werden die Ergebnisse in der Regel nicht in</i>

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<i>die Nutzenbewertung herangezogen, wenn die auf weniger als 70 % der in die Auswertung einzuschließenden Patienten basieren bzw. der Unterschied der Anteile nicht berücksichtigter Patienten zwischen den Gruppen größer als 15 % ist.</i> Es ist grundsätzlich zutreffend, dass das Verzerrungspotenzial mit dem steigenden Anteil der fehlenden Werte zunimmt. Die vom IQWiG vorgegebenen Schwellenwerte für die Nichtberücksichtigung von Studienergebnissen sind jedoch nicht hinreichend mit Publikationen untermauert. Dies gilt insb. für den Schwellenwert von 15 %, für den das IQWiG keine Literaturangaben zitiert. <u>Vorgeschlagene Änderung:</u> Die vorgegebenen Schwellenwerte für die Nichtberücksichtigung von Studienergebnissen sollten mit Literaturangaben untermauert werden.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

- [1]. Cohen J. Statistical Power Analysis in the Behavioral Sciences (2nd edition). Hillsdale (NJ): Lawrence Erlbaum Associates, Inc., 1988
- [2] Devji T, Carrasco-Labra A, Lytvyn L, Johnston B, Ebrahim S, Furukawa T et al. A new tool to measure credibility of studies determining minimally important difference estimates. Cochrane Database Syst Rev 2017; (9 Suppl 1): 58. [IQWiG 175]
- [3] Devji T, Guyatt GH, Lytvyn L, Brignardello-Petersen R, Foroutan F, Sadeghirad B et al. Application of minimal important differences in degenerative knee disease outcomes: a systematic review and case study to inform BMJ Rapid Recommendations. BMJ Open 2017; 7(5): e015587. [IQWiG 176]
- [4] Coon C, Cook K Moving from significance to real-world meaning: methods for interpreting change in clinical outcome assessment scores; Qual Life Res (2018) 27:33–40
- [5] Nordin A, Taft C, Lundgren-Nilsson A, Dencker A. Minimal important differences for fatigue patient reported outcome measures: a systematic review. BMC Med Res Methodol 2016; 16: 62. [IQWiG 523]
- [6] St-Pierre C, Desmeules F, Dionne CE, Fremont P, MacDermid JC, Roy JS. Psychometric properties of self-reported questionnaires for the evaluation of symptoms and functional limitations in individuals with rotator cuff disorders: a systematic review. Disabil Rehabil 2016; 38(2): 103-122 [IQWiG 655]

[7] Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Extrakt aus Cannabis sativa (Spastik aufgrund von multipler Sklerose) – Nutzenbewertung gemäß § 35a SGB V (Ablauf Befristung); A18-27

Abkürzungsverzeichnis:

AQLQ	Asthma Quality of Life Questionnaire
CAT	COPD Assessment Test
EORTC-QLQ-BR23	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire – Breast Cancer 23
EORTC-QLQ-CLL16	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire - Chronic Lymphocytic Leukaemia 16
EORTC-QLQ-C30	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Cancer-30
EORTC-QLQ-LC13	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Lung Cancer 13
EORTC-QLQ-MY20	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Multiple Myeloma 20
EQ-5D VAS	EQ-5D Visual Analog Scale
FACIT-F	Functional Assessment of Chronic Illness Therapy-Fatigue
FACT-B	Functional Assessment of Chronic Illness Therapy - Breast Cancer
FACT-G	Functional Assessment Cancer Therapy-General
FACT-Lym	Functional Assessment Cancer Therapy-Lymphoma
FACT-P	Functional Assessment Cancer Therapy- Prostate
FKSI-DRS	Functional Assessment of Cancer Therapy – Kidney Symptom Index – Disease Related Symptoms
SF-36	Short Form-36
SGRQ	St. George's Respiratory Questionnaire
TDI	Transition Dyspnea Index

A.2 – Stellungnahmen von Privatpersonen

A.2.1 – Kohlmann, Thomas

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Kohlmann, Thomas, Prof. Dr.
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
☐ im Namen folgender Institution / Organisation:
☒ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

<i>Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.</i>

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
9.3.3, insbes. S. 177f	<u>Anmerkung:</u> Im Entwurf für die Version 6.0 vom 05.12.2019 des Papiers über die „Allgemeine Methoden“ wird im Abschnitt 9.3.3 (Beurteilung klinischer Relevanz) für die Betrachtung von Responderanalysen im Rahmen der Nutzenbewertung eine Responseschwelle von 15 % der Skalenspannweite bzw. des Skalenrangs festgelegt. Die Festlegung einer solchen Schwelle erscheint vor dem Hintergrund der im Methodenpapier zu Recht angesprochenen Variabilität der unter der Bezeichnung Minimal Important Difference (MID) publizierten Responseschwellen durchaus gerechtfertigt. Die konkrete Wahl der Responseschwelle von 15 % der Skalenspannweite beruht allerdings auf einer sehr schmalen Evidenzgrundlage. Es werden nur zwei Übersichtsarbeiten mit einem jeweils sehr spezifischen Fokus (Fatigue, Läsionen der Rotatorenmanschette) herangezogen bzw. im Text zitiert. Da die Wahl der Responseschwelle einen entscheidenden Einfluss auf die Ergebnisse einer Responderanalyse haben kann, sollte die empirische Basis dieser Wahl substanziell, am besten im Kontext eines systematischen Reviews, verbreitert werden. Diese Forderung lässt sich u. a. dadurch begründen, dass eine Reihe von in den letzten Jahren publizierten Übersichtsarbeiten zu diesem Thema Spannweiten von MIDs berichten, die niedriger als die im Methodenpapier genannten Spannweite zwischen 10 % und 20 % liegen. Zu diesen Übersichtsarbeiten gehören zum Beispiel: Dabija et al. 2019 (American Journal of Physical Medicine & Rehabilitation) zum Bereich von Erkrankungen der Schulter (auch Läsionen der Rotatorenmanschette), Chung et al. 2017 (Spine) für

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	Erkrankungen der Wirbelsäule, Devji et al. 2017 (BMJ Open) für den Bereich degenerativer Erkrankungen des Knies, Jayadevappa et al. 2017 (Journal of Clinical Epidemiology) zum Bereich der Lebensqualitätsmessung mit krankheitsspezifischen und generischen Instrumenten in diversen Indikationsgebieten oder Ousmen et al. 2018 (Health and Quality of Life Outcomes) zu PROs in der Onkologie. Die in diesen Übersichten publizierten Spannweiten der MIDs rangieren typischerweise zwischen 8 % und 15 % der Skalenspannweite. Die im Methodenpapier gewählte Responseschwelle von 15 % der Skalenspannweite liegt damit an der oberen Grenze der publizierten Werte. Sollte sich in einer systematischen Übersicht der in den genannten Reviews exemplarisch gezeigte Wertebereich bestätigen, wäre es empfehlenswert, die im Kontext der Nutzenbewertung verwendete Responseschwelle entsprechend zu senken.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

A.2.2 – Semrau, Frank

Stellungnahme zum Entwurf der Allgemeinen Methoden Version 6.0

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden Version 6.0“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Semrau, Frank, Dr.
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
☐ im Namen folgender Institution / Organisation:
☒ als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

<i>Sofern Sie allgemeine Anmerkungen zum Entwurf der Allgemeinen Methoden haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.</i>
Alle Anmerkungen und Änderungsvorschläge fußen auf den Grundlagen der „evidence based medicine“ (EbM), hier repräsentiert durch die genannten Arbeiten von Sackett, Glasziou, Lange/Windeler sowie der GRADE-Gruppe (1Guyatt 2011, 2Glasziou 2007, 3Sackett 1996, 4Lange 2018). Wenn einzelne der Arbeiten für einen Kommentar insbesondere relevant sind, so sind diese Arbeiten direkt nach dem Änderungsvorschlag benannt.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
1.2 (S.4), Zeile 4	<u>Anmerkung:</u> Ziel der EbM sind nicht die wissenschaftlich objektiven Methoden der Erhebung, sondern die Robustheit der Belege, die das Vertrauen in die Richtigkeit der Effekte (cause-effect relation) und Effektgrößen begründen. Diese Belege können auch direkt an einem Patienten erfahrbar sein, z.B. dadurch, dass ein großer, reproduzierbarer Effekt unmittelbar oder in sehr zeitnahe Zusammenhang mit der Intervention steht. Direkt beobachtbare Belege begründen eine höhere Qualität der Evidenz als mit Methoden der Statistik ermittelte, potentielle Populationseffekte. Die Mittel der vergleichenden, bestenfalls randomisierten, verblindeten Studien sind insbesondere dann notwendig, wenn die Effekte eher klein sind und/oder erst mit großem zeitlichen Abstand zur Intervention messbar werden. Siehe auch Kapitel 3.2.2, „Dramatischer Effekt“. <u>Vorgeschlagene Änderung:</u> <i>„(...) Belege, die mit möglichst objektiven wissenschaftlichen Methoden erhoben wurden.“ Hierzu gehört auch die reproduzierbare Erfahrbarkeit eines großen Effekts mit direktem zeitlichem Zusammenhang und/oder eindeutiger Kausalität zur Intervention direkt am Patienten (z.B. Insulin bei diabetischer Ketoazidose, Defibrillation bei Kammerflimmern, Epinephrin bei anaphylaktischem Schock)</i> (1Guyatt 2011, 2Glasziou 2007).

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
1.2.3 (S.6.), Zeile 8	<u>Anmerkung:</u> Es muss unterschieden werden, ob „Watchful Waiting“ eine ernsthafte Alternative zur in Frage stehenden Therapie darstellt oder diese Vorgehensweise regelhaft ernste Folgen bis hin zum Tode der Patienten hat. Gibt es keine alternative Therapie um solche ernsten Folgen zu verhindern, so adressiert die untersuchte Therapie einen „unmet medical need“ und sollte prioritär behandelt werden. <u>Vorgeschlagene Änderung:</u> „Zu betonen ist allerdings, dass der Verzicht auf eine Therapie nicht gleichbedeutend mit „nichts tun“ ist (zum Beispiel abwartendes Beobachten, Watchful Waiting [381]).“ Dabei muss unterschieden werden, ob „Watchful Waiting“ eine ernsthafte Alternative zur in Frage stehenden Therapie darstellt oder diese Vorgehensweise regelhaft ernste Folgen bis hin zum Tode der Patienten hat. Gibt es keine alternative Therapie um solche ernsten Folgen zu verhindern, so adressiert die untersuchte Therapie einen „unmet medical need“.
1.2.3 (S.7), Zeile 10	<u>Anmerkung:</u> Der Vollständigkeit halber sollte auch die Möglichkeit der direkten Evidenzgewinnung durch Hypothese, Test und Repetition erwähnt werden. Auch wenn diese Möglichkeit nur auf wenige Interventionen zutrifft, so wäre es umso dramatischer, wenn solche offensichtlich hochwirksamen Therapien nicht in die Patientenversorgung kämen, weil keine RCT existieren oder letztere missinterpretiert werden. <u>Vorgeschlagene Änderung:</u> <i>„Diese Regeln und Werkzeuge, die international als Standard akzeptiert sind und laufend weiterentwickelt werden, sind die methodische Basis der EbM und der Arbeit des Instituts.“</i> Der Gewinnung von Wahrscheinlichkeiten durch vergleichende klinische Studien muss die direkt beobachtbare Erkenntnis zur Seite gestellt werden. In seltenen Fällen kann durch Hypothese, Test und Repetition an einer Reihe von Patienten „direkte“ Evidenz gewonnen werden, die vergleichende Untersuchungen überflüssig macht, insbesondere wenn der Effekt sehr groß ist (dramatischer Effekt) und daher verzerrende Einflüsse ausgeschlossen oder vernachlässigt werden können, eine alternative Behandlungsmethode nicht zur Verfügung steht und/oder die Alternative (z.B. Nicht-Therapie) zu einem hohen Prozentsatz den Tod der Patienten zur Folge hat. So schreibt David Sackett beispielsweise: „However, some questions about therapy do not require randomized trials (successful interventions for otherwise fatal conditions) (...).“ Weitere Erkenntnisse können dann über einarmige Studien mit der Therapie gewonnen werden. Beispiele für solche Therapien sind die mit Hilfe von Insulin bei Ketoazidose oder die

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	Defibrillation bei Kammerflimmern. (3Sackett 1996, 4Lange 2018, 1Guyatt 2011, 2Glasziou 2007)
1.2.3 (S.7), Zeile 15	<u>Anmerkung:</u> Nicht nur bei der Heterogenität der Ergebnisse von Studien desselben Evidenzniveaus, sondern insbesondere auch bei Heterogenität zwischen Studientypen muss gefragt werden, was die Ursache dieser Heterogenität ist. Es ist in solchen Fällen zu kurz gegriffen, die vom Grundsatz her höhere Qualität der Evidenz von RCT als Anlass zu nehmen, andere Studientypen nicht mehr zu beachten. RCT können durch ein fehlerhaftes Design, durch unvollständige Vorabüberlegungen oder durch Fehlen notwendiger Parameter, die in der klinischen Praxis vorhanden sind, ein falsches Bild jenseits bestmöglich reduzierter Verzerrungen zeigen. Bei grundlegend unterschiedlichen Ergebnissen zwischen RCT und Beobachtungsstudien muss daher zwingend nach den Ursachen gesucht werden. <u>Vorgeschlagene Änderung:</u> <i>„Zeigen sich dabei große Unterschiede zwischen den Ergebnissen einzelner Studien oder Studientypen (eine sogenannte Heterogenität), muss versucht werden, diese Unterschiede zu erklären.“</i>
1.2.4 (S.8), Zeile 12	<u>Anmerkung:</u> Es gibt nicht immer Alternativen „die man nicht benachteiligen möchte“. Aus ökonomischen und pragmatischen Gründen sollte der Studientyp herangezogen werden, der entweder die bestmögliche oder, falls nicht vorhanden, eine ausreichende Sicherheit der Antwort auf die Frage erwarten lässt. <u>Vorgeschlagene Änderung:</u> <i>„Es gehört zu den internationalen Standards der evidenzbasierten Medizin, für jede Frage den Typ von Studien als Messinstrument zu benennen, der das Risiko minimiert, dass eine der Alternativen unberechtigterweise benachteiligt wird.“ favorisieren, der eine bestmögliche bis ausreichende Ergebnissicherheit der Antwort auf die Frage erwarten lässt.</i>
1.2.4 (S.8), Zeile 18	<u>Anmerkung:</u> Das Ziel muss stets die möglichst korrekte Darstellung der vollständigen Evidenz sein. Ein Ausschluss weiterer Recherche (z.B. von nicht-RCT) aus Zeit- und Kostengründen ist nicht akzeptabel, sofern nicht überzeugende und unbestrittene Ergebnisse aus mehreren RCT vorliegen und Erkenntnisse aus Beobachtungsstudien in die gleiche Richtung weisen. <u>Vorgeschlagene Änderung:</u> Folgenden Satz streichen: <i>„Zudem wäre eine Untersuchung, die eine Recherche nach Studien unzureichender Ergebnissicherheit einschließt, zeit- und kostenintensiv.“</i>

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
1.2.4 (S.8), Zeile 20	<u>Anmerkung:</u> Dies ist prinzipiell nicht zu beanstanden. Allerdings muss die Frage erlaubt sein, welche Studien die nötige Qualität und Präzision für eine bestimmte Therapie/Methode/Fragestellung innehaben. Hier ist es wichtig darauf hinzuweisen, dass große Effekte mit eher geringen Verzerrungsmöglichkeiten und eindeutigem Endpunkt (z.B. Mortalität) u.U. auch mit einarmigen Studien überzeugend dargestellt werden können. Dies wurde auch weiter oben angesprochen. Insofern ist die hier anklingende Dogmatik bei der Studienausswahl nicht angemessen. <u>Vorgeschlagene Änderung:</u> <i>„Falls sich herausstellt, dass Studien mit der nötigen Qualität und Präzision für die Beurteilung einer spezifischen Intervention oder Fragestellung generell fehlen, ist es die Kernaufgabe des Instituts, den Sachverhalt zu beschreiben mit der Schlussfolgerung, dass auf Basis der „gegenwärtig besten“ Evidenz keine sicheren Empfehlungen möglich sind.“</i>
1.2.5 (S.9), Zeile 23	<u>Anmerkung:</u> Die Randomisierung ist nur ein Ersatz für die Möglichkeit, Therapie und nicht-Therapie an denselben Patienten zu testen. Besteht die Möglichkeit, dies ohne Beeinflussung des für die Therapie maßgeblichen Ausgangsstatus zu tun, ist das ein entscheidender Vorteil gegenüber der Randomisierung. Schließt man nicht-RCT vor einem Assessment aus, so bleiben solche Studien unberücksichtigt. <u>Vorgeschlagene Änderung:</u> <i>„Andere Studientypen als RCTs sind in der Regel für einen Kausalitätsnachweis nicht geeignet. Nur selten ist es möglich, dieselbe Patientengruppe als Verum- und Kontrollgruppe einzusetzen. In nicht randomisierten vergleichenden Studien kann ansonsten grundsätzlich nicht von einer Strukturgleichheit der Gruppen ausgegangen werden. Sie liefern daher immer ein potenziell verzerrtes Ergebnis und können häufig die maßgebliche Frage nicht hinreichend sicher beantworten, ob ein Unterschied des Effektes auf der geprüften Intervention beruht.“</i> (4Lange 2018)
1.2.5 (S.9), Zeile 36	<u>Anmerkung:</u> Die Andersartigkeit der Betreuung kann sich positiv oder nachteilig auf den Effekt auswirken und bei verschiedenen Methoden diametral gegensätzlich wirken. So ist die Compliance von Patienten in pharmazeutischen Studien generell deutlich besser als im Alltag, wogegen bei medizin-technischen Anwendungen gelegentlich herstellerseits Compliance-verbessernde Maßnahmen angeboten werden, die in RCTs zur Reduktion eines Bias aber i.d.R. ausgeschlossen sind. <u>Vorgeschlagene Änderung:</u>

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	„Zudem werden Patientinnen und Patienten in randomisierten Studien oft anders (intensiver und enger) betreut als im Alltag.“ Dies kann eine intensivere Betreuung durch Studienpersonal bedeuten aber auch ein Fehlen externer Betreuung zur Vermeidung von Bias innerhalb einer Studie.
1.2.5 (S.10), Z. 11	<u>Anmerkung:</u> Der Einfluss auf die Therapieadhärenz ist individuell von Studiendesign und Intervention abhängig. Einseitige Beispiele beeinflussen den Leser ungerechtfertigterweise in eine Richtung. <u>Vorgeschlagene Änderung:</u> „Zum Beispiel könnten Unterschiede in der Therapieadhärenz in Studien kleiner ausfallen, in denen Patientinnen und Patienten grundsätzlich sehr eng betreut werden.“ Fehlt dagegen unter Studienbedingungen eine sonst vorhandene systematische Überwachung mit Kontrollmöglichkeit der Adhärenz, so kann dies den Effekt einer Intervention verringern.
1.2.5 (S.10), Z. 26	<u>Anmerkung:</u> Hier konkurrieren m.E. Efficacy (Wirksamkeit unter Idealbedingungen) und Effectiveness (Wirksamkeit unter Alltagsbedingungen) miteinander. Es erscheint sinnvoll, sofern für die betreffende Therapie möglich, diese beiden Fragestellungen getrennt zu beantworten. Einerseits heißt eine Wirksamkeit unter Idealbedingungen noch nicht, dass sich das Konzept unter Realbedingungen realisieren lässt, andererseits kann bei den kondensierten "Idealbedingungen" z.B. innerhalb eines RCT ein Aspekt der Alltagsbedingungen vergessen worden sein, der die Wirksamkeit erst an/in den Patienten bringt. <u>Vorgeschlagene Änderung:</u> „Auch wenn die Kritik mangelnder Alltagsnähe tatsächlich für viele Studien zutreffen mag, wäre nichts dadurch gewonnen, wenn man eine hohe Ergebnissicherheit für eine größere Alltagsnähe aufgeben würde. Man würde einen Mangel durch Inkaufnahme eines zweiten, noch schwerwiegenderen Mangels zu kompensieren versuchen [334].“ Alltagsnähe und mangelnde Alltagsnähe sind kein konkurrierendes Werte-Paar. Es geht eher um den zunächst notwendigen grundsätzlichen Nachweis der Wirksamkeit unter Idealbedingungen (Efficacy), der am ehesten glaubhaft in RCT mit überschaubaren Variablen erbracht werden kann. Die Übertragbarkeit der Wirksamkeit in den klinischen Alltag ist dabei nicht immer zwangsläufig gegeben. Bei der Reduktion der Variablen innerhalb eines RCT kann aber auch eine notwendige, im klinischen Alltag vorhandene Variable (unbeabsichtigt) ausgeschlossen werden. Dann beruht die vermeintlich fehlende Efficacy auf unvollständigen Voraussetzungen und die Bedingungen zur Bestimmung der tatsächlichen Efficacy nähern sich denen der Effectiveness (Wirkung unter Bedingungen im klinischen Alltag) an. Idealerweise werden, wenn möglich, beide Aspekte betrachtet.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
1.2.6 (S.11), Z. 4	<u>Anmerkung:</u> Man muss unterscheiden zwischen Einzelfallberichten (Case Reports) und Interventionen, die regelhaft, d.h. vorhersagbar einen Effekt an (quasi jedem für die Intervention geeigneten) Patienten verursachen. Eine gedankliche Vermischung dieser verschiedenen „Einzelfälle“ ist nicht zielführend. <u>Vorgeschlagene Änderung:</u> „Diese Aussagen betreffen immer Gruppen von Patientinnen und Patienten mit bestimmten Eigenschaften. Aussagen über den Nutzen im Sinne von Erfolgsaussagen für den Einzelfall sind, abgesehen von Ausnahmen (z. B. dramatischer Effekt, sehr kurze Response Zeiten), grundsätzlich nicht möglich. Umgekehrt sind Auf Einzelfällen basierende Erfahrungen sind –abgesehen von Ausnahmen (z. B. dramatischer Effekt)– für eine Nutzenbewertung ungeeignet, da es nicht möglich ist, Ergebnisse im Einzelfall (d. h. ohne Vergleich und ohne Repetition an weiteren Fällen) einer Intervention zuzuschreiben.“
2.1.4 (S.24), Z. 18	<u>Anmerkung:</u> Es ist sehr fraglich, ob nicht-Experten alle offenen und zudem die richtigen Fragen an die Experten innerhalb eines Fragenkatalogs zusammenstellen können. Wenn die nicht-Experten glauben, einen Sachverhalt verstanden zu haben, werden sie die entsprechenden Fragen hierzu nicht stellen - auch wenn sie falsch liegen. Ohne ausführliches Gespräch zum Gesamtkomplex ist ein Gesamtverständnis für die Intervention oft illusorisch. Wenn dann die zu Rate gezogenen Experten selbst keine Erfahrung mit der Intervention haben, sind Missverständnisse zu Intervention und medizinischem Background vorprogrammiert. <u>Vorgeschlagene Änderung:</u> - Handlungsempfehlung: grundsätzliche Intensivierung des Experten-Inputs durch persönliche Gespräche und ggf. Diskussion von Experten mit Erfahrung und solchen ohne Erfahrung mit der betreffenden Intervention. „Die medizinische Expertise wird primär auf Basis eines Fragenkatalogs eingeholt, der zu Beginn der Bewertung an die externen Sachverständigen verschickt wird.“
3.1.3 (S.47), Z. 8	<u>Anmerkung:</u> In Registern dokumentierte, unerwünschte Ereignisse werden i.d.R. als unerwünschte Wirkungen interpretiert. In einer vergleichenden Untersuchung, insbesondere einem RCT, wird oft deutlich, dass ein Teil dieser unerwünschten Ereignisse indikationsspezifisch auftritt und nicht kausal mit der Intervention zusammenhängt. Spätestens nach Feststellung dieser Differenzierung sollten diese Ereignisse nicht mehr (oder entsprechend nur noch anteilig) der Intervention zugerechnet werden. <u>Vorgeschlagene Änderung:</u> - keine -

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	„Für eine Begriffsklärung ist anzumerken, dass bei Verwendung des Begriffs unerwünschte Wirkungen ein kausaler Zusammenhang mit der Intervention angenommen wird,...“
3.1.3 (S.48), Z. 13	<u>Anmerkung:</u> Die oft gehörte Meinung, dass auch unerwünschte Nebenwirkungen am besten in RCT dargestellt werden, ist logisch nicht korrekt. Um die Diskussion darüber zu versachlichen, wäre ein kurzer Exkurs hierzu hilfreich. <u>Vorgeschlagene Änderung:</u> <i>„Gerade für die Erfassung unerwartet auftretender unerwünschter Ereignisse ist jedoch eine generelle Abfrage des Befindens der Patientinnen und Patienten erforderlich. Während sich unerwartete und seltene unerwünschte Ereignisse am ehesten in großen Populationen wie prospektiven Registern erfassen lassen, ist eine Abgrenzung interventionsgenerierter Effekte von gleichartigen indikationsspezifischen Effekten nur durch vergleichende Studien möglich. Zur Ermittlung des Ausmaßes von bekanntermaßen interventionsspezifischen unerwünschten Effekten (z.B. inadäquate Therapieabgaben bei einem Defibrillator) sind dagegen lediglich einarmige Untersuchungen notwendig, da der entsprechende Wert in einer Kontrollgruppe nur „null“ sein kann. Des Weiteren ist Die zielgerichtete Forschung zur Erhebung insbesondere der seltenen schwerwiegenden unerwünschten Ereignisse einschließlich der Darstellung eines kausalen Zusammenhangs mit der medizinischen Intervention ist oftmals erheblich unterrepräsentiert [61,205,387]. Da bei wenig untersuchten Interventionen die unerwünschten Ereignisse zunächst per se der Intervention zugeordnet werden, liegt es u.U. im Interesse des Herstellers durch vergleichende Studien die indikationsspezifischen von den interventionsspezifischen Nebenwirkungen abzugrenzen. Darüber hinaus ist die Qualität der Berichterstattung unerwünschter Ereignisse in Einzelstudien vielfach mangelhaft, was auch zu einer Erweiterung des CONSORT-Statements für randomisierte klinische Studien geführt hat [386].“</i>
3.1.3 (S.49), Z. 14	<u>Anmerkung:</u> Bei Therapieabbrüchen empfiehlt es sich, der Gesamtabbruchrate diejenige der Abbrüche wegen schwerer unerwünschter Nebenwirkungen gegenüberzustellen, da ein wesentlicher Unterschied zwischen „keine laufende Therapie mehr“ und „Schaden am Patienten durch die Therapie, der zum Abbruch führt“ besteht. <u>Vorgeschlagene Änderung:</u> <i>„Grundsätzlich werden die Gesamtraten der schwerwiegenden unerwünschten Ereignisse, ggf. der schweren unerwünschten Ereignisse (z. B. solche mit Common-Terminology-Criteria-for-Adverse-</i>

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<i>Events[CTCAE]-Grad ≥ 3) und der Abbrüche wegen unerwünschter Ereignisse herangezogen.“</i> Bei Therapieabbrüchen empfiehlt es sich, der Gesamtabbruchrate diejenige der Abbrüche wegen schwerer unerwünschter Nebenwirkungen gegenüberzustellen, da ein wesentlicher Unterschied zwischen „keine laufende Therapie mehr“ und „Schaden am Patienten durch die Therapie, der zum Abbruch führt“ besteht.
3.1.4 (S.51), Z. 14	<u>Anmerkung:</u> Die GRADE Gruppe empfiehlt ein Up- bzw. Downgrading ausgehend von der Basiseinordnung des Studientyps. Dies scheint zur Abwägung der Erkenntnissicherheit sinnvoll. In diese Richtung weist auch die Aussage auf Seite 54: „...das Vorliegen schwerwiegender Designmängel bei einer Studie oder auch begründete Zweifel an der Übertragbarkeit auf die Behandlungssituation in Deutschland können z.B. zu einer Verringerung der Aussagesicherheit führen. Auf der anderen Seite können z. B. große Effekte oder eine eindeutige Richtung eines vorhandenen Verzerrungspotenzials eine Erhöhung der Sicherheit begründen.“ <u>Vorgeschlagene Änderung:</u> - Implementierung des Up- und Downgradings von Parametern, Studien bzw. des „body of evidence“ nach GRADE Guidelines: 9 - „Das Institut verwendet die folgenden 3 Kategorien zur Graduierung des Ausmaßes der qualitativen Ergebnissicherheit auf Einzelstudien- und Endpunktebene: ☐☐ <i>hohe qualitative Ergebnissicherheit: Ergebnis bezüglich eines Endpunkts einer randomisierten Studie mit niedrigem Verzerrungspotenzial</i> ☐☐ <i>mäßige qualitative Ergebnissicherheit: Ergebnis bezüglich eines Endpunkts einer randomisierten Studie mit hohem Verzerrungspotenzial</i> ☐☐ <i>geringe qualitative Ergebnissicherheit: Ergebnis bezüglich eines Endpunkts einer nicht randomisierten vergleichenden Studie“</i> Entsprechend den GRADE Guidelines: 9 kann die durch das Studiendesign vorgegebene Basis-Ergebnisqualität ggf. auf- bzw. abgewertet werden. (1Guyatt 2011)
3.1.4 (S.54), Z. 12	<u>Anmerkung:</u> Das erscheint als Dogma nicht sinnvoll, da auch RCT Fehler enthalten können, die sich substanziell auf die Ergebnisse auswirken. Ist auch die Verlässlichkeit der Aussagen eines RCT grundsätzlich höher als bei Beobachtungsstudien, so muss dies im Einzelfall zunächst bestätigt werden, bevor einer Studie oder Studienart der Vorzug gegenüber anderen Studien gegeben wird.

Kapitel/ Abschnitt (Seite) im Entwurf der Allgemeinen Methoden	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<u>Vorgeschlagene Änderung:</u> „Die Aussagen zur Beleglage bei einer Beschränkung auf die höherwertigeren Studien werden durch Hinzunahme der übrigen Studien nicht abgeschwächt, sondern allenfalls aufgewertet;“ sofern keine Heterogenität zwischen den Studientypen Zweifel an der Aussagekraft der höherwertigen Studien aufkommen lässt. Unterscheidet sich die Beleglage zwischen RCTs und Beobachtungsstudien signifikant, so wird unter Zuhilfenahme von Experten nach Gründen für diese Diskrepanz gesucht. Werden keine Hinweise hierzu gefunden, ist es gerechtfertigt, lediglich die höherwertigen Studientypen zur Bestimmung der Beleglage heranzuziehen.
8.1.1 (S.150), Z. 6	<u>Anmerkung:</u> Die Art des Effekts und der Effektgröße hat einen Einfluss auf die relevanten Studien und damit auf die Suchstrategie. Dementsprechend muss bereits während der Vorrecherche geklärt werden, ob es sich um einen dramatischen Effekt handeln könnte. <u>Vorgeschlagene Änderung:</u> „Zu Beginn eines Projekts, vor der Entwicklung der eigentlichen Suchstrategie, wird in der Regel eine Vorabrecherche durchgeführt. Die Vorabrecherche dient der Vorbereitung des Projekts, um potenziell relevante Primärstudien zum Thema zu identifizieren. Hierfür wird nach bereits vorhandenen systematischen Übersichten gesucht [124,252,657]. Dadurch können auch die Ressourcen für die Durchführung des Projekts genauer bestimmt werden. Zudem wird das Ergebnis der Vorabrecherche genutzt, um ein Testset zur Entwicklung von Suchstrategien zu entwickeln.“ Während der Vorabrecherche wird auch gezielt eruiert, ob es sich aufgrund der Effektgröße und der Zeit bis zum Effekteintritt um einen dramatischen Effekt handeln könnte. Hierauf wird dann ggf. die Suchstrategie abgestimmt.
9.1.2 (S.164), Z. 2	<u>Anmerkung:</u> Der Verweis auf das entsprechende Kapitel ist falsch (3.2.1 statt 3.2.2). Zur Verdeutlichung dramatischer Effekte sind zudem unterschiedliche Beispiele hilfreich. <u>Vorgeschlagene Änderung:</u> „Aus einem reinen Vorher-nachher-Vergleich in einem Design mit abhängigen Stichproben ohne Kontrollgruppe lässt sich in der Regel kein Beleg für einen Effekt einer Intervention ableiten. Ausnahmen bilden Krankheitsbilder mit einem deterministischen (oder nahezu deterministischen) Verlauf (z. B. Defibrillation bei Kammerflimmern oder diabetisches ketoazidotisches Koma; siehe Abschnitt 3.2.2).“ (Glasziou 2007)

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

1Guyatt GH et al. (2011) GRADE guidelines: 9. Rating up the quality of evidence. Journal of Clinical Epidemiology 64, 1311e1316

2Glasziou P et al. (2007) When are randomised trials unnecessary? Picking signal from noise, BMJ, Analysis, 334

3Sackett DL et al. (1996) Evidence based medicine: what it is and what it isn't. BMJ 312, 71-72

4Lange S et al. (2018) Klinische Studien und Equipoise, Dt. Ärzteblatt 115, 3, A70-A73