

**Dokumentation und Würdigung
der Anhörung zum Entwurf der
Allgemeinen Methoden 5.0**

Version 1.0 vom 10.07.2017

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
Inhaltsverzeichnis	ii
Abkürzungsverzeichnis	vi
1 Dokumentation der Anhörung	1
2 Würdigung der Anhörung	2
2.1 Würdigung allgemeiner Stellungnahmen	2
2.2 Würdigung der Stellungnahmen zu Abschnitt 1.2 „Evidenzbasierte Medizin“	2
2.3 Würdigung der Stellungnahmen zu Abschnitt 1.3 „Gesundheitsökonomie“	3
2.4 Würdigung der Stellungnahmen zu Abschnitt 1.3.3 „Methodische Standards in der Gesundheitsökonomie“	3
2.5 Würdigung der Stellungnahmen zu Abschnitt 2.1 „Produktspezifische Verfahrensabläufe“	4
2.6 Würdigung der Stellungnahmen zu Abschnitt 2.1.5 „Potenzialbewertungen“	5
2.7 Würdigung der Stellungnahmen zu Abschnitt 2.1.6 „Bewertung gemäß § 137h SGB V“	5
2.8 Würdigung der Stellungnahmen zu Abschnitt 2.1.7 „Addendum“	6
2.9 Würdigung der Stellungnahmen zu Abschnitt 2.1.8 „Gesundheitsinformationen“	6
2.10 Würdigung der Stellungnahmen zu Abschnitt „2.1.10 HTA-Bericht“	8
2.11 Würdigung der Stellungnahmen zu Abschnitt 2.2.1 „Einbindung von Betroffenen in die Erstellung von IQWiG-Produkten“	9
2.12 Würdigung der Stellungnahmen zu Abschnitt 2.2.2 „Auswahl externer Sachverständiger“	10
2.13 Würdigung der Stellungnahmen zu Abschnitt 2.2.3 „Gewährleistung der fachlichen Unabhängigkeit“	10
2.14 Würdigung der Stellungnahmen zu Kapitel 3 „Nutzenbewertung medizinischer Interventionen“	11
2.15 Würdigung der Stellungnahmen zu Abschnitt 3.1.3 „Ermittlung des Schadens medizinischer Interventionen“	11
2.16 Würdigung der Stellungnahmen zu Abschnitt 3.1.4 „Endpunktbezogene Bewertung“	11
2.17 Würdigung der Stellungnahmen zu Abschnitt 3.2.4 „Patientenberichtete Endpunkte“	12
2.18 Würdigung der Stellungnahmen zu Abschnitt 3.2.5 „Nutzen und Schaden in kleinen Populationen“	12

2.19	Würdigung der Stellungnahmen zu Abschnitt 3.3.2 „Studien zur Nutzenbewertung von Arzneimitteln“	12
2.20	Würdigung der Stellungnahmen zu Abschnitt 3.3.3 „Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V“	13
2.21	Würdigung der Stellungnahmen zu Abschnitt 3.4 „Nicht medikamentöse therapeutische Interventionen“	15
2.22	Würdigung der Stellungnahmen zu Abschnitt 3.5 „Diagnostische Verfahren“	16
2.23	Würdigung der Stellungnahmen zu Abschnitt 3.6 „Früherkennung und Screening“ und Abschnitt 3.7 „Prävention“	18
2.24	Würdigung der Stellungnahmen zu Abschnitt 3.8 „Potenzialbewertung“	19
2.25	Würdigung der Stellungnahmen zu Abschnitt 4.1 „Einführung“ des Kapitels 4 „Kosten-Nutzen-Bewertung medizinischer Interventionen“	20
2.26	Würdigung der Stellungnahmen zu Abschnitt 4.1.3 „Zeithorizont“	21
2.27	Würdigung der Stellungnahmen zu Abschnitt 4.1.7 „Spezifische Datenquellen der Gesundheitsökonomie“	21
2.28	Würdigung der Stellungnahmen zu Abschnitt 4.3.3 „Maß des Gesamtnutzens“	21
2.29	Würdigung der Stellungnahmen zu Abschnitt 4.3.4 „Datengrundlage“	22
2.30	Würdigung der Stellungnahmen zu Abschnitt 4.4.2 „Abgrenzung der Kosten“	22
2.31	Würdigung der Stellungnahmen zu Abschnitt 4.6.4 „Sonderkonstellationen“ ..	22
2.32	Würdigung der Stellungnahmen zu Abschnitt 4.9.4 „Interquartilsregion als Maß der Streuung für Preisverhandlungen“	23
2.33	Würdigung der Stellungnahmen zu Kapitel 5 „Leitliniensynopsen und Versorgungsanalyse“	23
2.34	Würdigung der Stellungnahmen zu Abschnitt 5.2.4 „Bewertung der methodischen Leitlinienqualität“	25
2.35	Würdigung der Stellungnahmen zu Abschnitt 5.2.2 „Übertragbarkeit auf das deutsche Gesundheitssystem“	25
2.36	Würdigung der Stellungnahmen zu Abschnitt 5.4 „Versorgungsanalyse“	26
2.37	Würdigung der Stellungnahmen zu Abschnitt 6.1 „Hintergrund und Ziel“ des Kapitels 6 „HTA-Berichte“	26
2.38	Würdigung der Stellungnahmen zu Abschnitt 6.3.1 „Auswahlkriterien“	26
2.39	Würdigung der Stellungnahmen zu Abschnitt 6.3.2 „Prüfung und Gewichtung der Themen“	27
2.40	Würdigung der Stellungnahmen zu Abschnitt 6.3.4 „2. Stufe des Auswahlverfahrens: Auswahl von Themen, zu denen HTA-Berichte erstellt werden“	27
2.41	Würdigung der Stellungnahmen zu Abschnitt 6.5 „Bearbeitung der Themen (HTA-Berichte)“	27
2.42	Würdigung der Stellungnahmen zu Abschnitt 6.5.2 „Gesundheitsökonomie“ ..	28

2.43	Würdigung der Stellungnahmen zu Abschnitt 6.5.6 „Organisatorische Aspekte“	28
2.44	Würdigung der Stellungnahmen zu Kapitel 7 „Evidenzbasierte Gesundheitsinformationen für Bürgerinnen und Bürger“ und zu Abschnitt 7.1 „Hintergrund und Ziele“	29
2.45	Würdigung der Stellungnahmen zu Abschnitt 7.2.1 „Themenkatalog gemäß dem Generalauftrag“	29
2.46	Würdigung der Stellungnahmen zu Abschnitt 7.2.2 „Identifizierung der Informationsbedürfnisse / Aufbereitung von Krankheitserfahrungen“	30
2.47	Würdigung der Stellungnahmen zu Abschnitt 7.3 „Informationsbeschaffung zur Erstellung von Gesundheitsinformationen“	31
2.48	Würdigung der Stellungnahmen zu den Abschnitten 7.4 „Auswahl der Evidenz“, 7.5 „Auswahl der dargestellten Ergebnisse“, 7.6 „Wahl und Darstellung von Vergleichen“, 7.7 „Umgang mit Zahlen und Risikoangaben“ und 7.8 „Berücksichtigung von Alters- und Geschlechtsunterschieden“	31
2.49	Würdigung der Stellungnahmen zu den Abschnitten 7.9 „Anpassung an die Zielgruppe“ und 7.9.1 „Beteiligung von Betroffenen“	33
2.50	Würdigung der Stellungnahmen zu Abschnitt 7.13 „Transparenz über Verfasser und Herausgeber“	33
2.51	Würdigung der Stellungnahmen zu Kapitel 8 „Informationsbeschaffung“	34
2.52	Würdigung der Stellungnahmen zu Abschnitt 8.1 „Umfassende Informationsbeschaffung“	34
2.53	Würdigung der Stellungnahmen zu Abschnitt 8.1.1 „Suche in bibliografischen Datenbanken“	35
2.54	Würdigung der Stellungnahmen zu Abschnitt 8.1.2 „Suche in Studienregistern“	35
2.55	Würdigung der Stellungnahmen zu Abschnitt 8.2 „Fokussierte Informationsbeschaffung“	35
2.56	Würdigung der Stellungnahmen zu Abschnitt 8.3 „Orientierende Recherchen“	37
2.57	Würdigung der Stellungnahmen zu Abschnitt 9.1.3 „Rangordnung verschiedener Studienarten / Evidenzgrade“	37
2.58	Würdigung der Stellungnahmen zu Abschnitt 9.3.3 „Beurteilung klinischer Relevanz“	37
2.59	Würdigung der Stellungnahmen zu Abschnitt 9.3.4 „Bewertung subjektiver Endpunkte bei offenem Studiendesign“	38
2.60	Würdigung der Stellungnahmen zu Abschnitt 9.3.7 „Adjustierung und multifaktorielle Verfahren“	38
2.61	Würdigung der Stellungnahmen zu Abschnitt 9.3.8 „Metaanalysen“	38
2.62	Würdigung der Stellungnahmen zu Abschnitt 9.3.9 „Indirekte Vergleiche“	40
2.63	Würdigung der Stellungnahmen zu Abschnitt 9.3.10 „Subgruppenanalysen“	41

2.64	Würdigung der Stellungnahmen zu Abschnitt 9.3.13 „Auswertung abhängiger Daten“	44
2.65	Literatur.....	45
3	Offenlegung potenzieller Interessenkonflikte	48
3.1	Potenzielle Interessenkonflikte von Stellungnehmenden aus Organisationen, Institutionen und Firmen	48
3.2	Potenzielle Interessenkonflikte von stellungnehmenden Privatpersonen.....	50
3.3	Potenzielle Interessenkonflikte von weiteren Teilnehmern der wissenschaftlichen Erörterung (wissenschaftlicher Beirat)	50
4	Dokumentation der wissenschaftlichen Erörterung – Teilnehmerliste, Tagesordnung und Protokoll.....	54
4.1	Teilnehmerliste der wissenschaftlichen Erörterung.....	54
4.2	Liste der Stellungnahmen, zu denen kein Vertreter an der wissenschaftlichen Erörterung teilgenommen hat.....	55
4.3	Tagesordnung der wissenschaftlichen Erörterung.....	56
4.4	Protokoll der wissenschaftlichen Erörterung.....	56
	Anhang A – Dokumentation der Stellungnahmen	100

Abkürzungsverzeichnis

Abkürzung	Bedeutung
ABDA	Bundesvereinigung Deutscher Apothekerverbände
AGREE	Appraisal of Guidelines for Research & Evaluation
AM-NutzenV	Arzneimittel-Nutzenbewertungsverordnung
AMVSG	Arzneimittelversorgungsstärkungsgesetz
AWMF	Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften
BMG	Bundesministerium für Gesundheit
DMP	Disease-Management-Programm
EbM	evidenzbasierte Medizin
G-BA	Gemeinsamer Bundesausschuss
GKV	gesetzliche Krankenversicherung
GoR	Grade of Recommendation (Empfehlungsgrad)
HTA	Health Technology Assessment
IQR	Interquartilsregion
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
KNB	Kosten-Nutzen-Bewertung
LoE	Level of Evidence (Evidenzlevel)
NVL	Nationale VersorgungsLeitlinie
OECD	Organisation for Economic Cooperation and Development
RCT	Randomized controlled Trial (randomisierte kontrollierte Studie)
SGB	Sozialgesetzbuch
SGB V	Sozialgesetzbuch V. Buch
WHO	World Health Organization
WiDO	Wissenschaftliches Institut der Allgemeinen Ortskrankenkasse

1 Dokumentation der Anhörung

Am 07.12.2016 wurde der Entwurf der Allgemeinen Methoden Version 5.0 veröffentlicht und zur Anhörung gestellt. Bis zum 31.01.2017 konnten schriftliche Stellungnahmen eingereicht werden. Insgesamt wurden 34 Stellungnahmen form- und fristgerecht abgegeben. Diese Stellungnahmen sind im Anhang abgebildet.

Unklare Aspekte in den schriftlichen Stellungnahmen wurden in einer wissenschaftlichen Erörterung am 23.03.2017 im Institut diskutiert. Das Wortprotokoll der Erörterung befindet sich in Kapitel 4.

2 Würdigung der Anhörung

Im Stellungnahmeverfahren zum Dokument „Allgemeine Methoden: Entwurf für Version 5.0“ vom 07.12.2016 gingen bis zum 31.01.2017 Stellungnahmen von 34 Organisationen, Institutionen, Firmen und Einzelpersonen ein. Alle Stellungnahmen wurden bei der Überarbeitung des Methodenpapiers berücksichtigt. Sofern sich aus den Stellungnahmen Änderungen in den jeweiligen Abschnitten der Allgemeinen Methoden ergaben, ist dies im vorliegenden Dokument explizit beschrieben.

2.1 Würdigung allgemeiner Stellungnahmen

In 1 Stellungnahme wird angeregt, Resamplingverfahren in das Methodenspektrum aufzunehmen sowie einen Abschnitt zur grafischen Präsentation von Ergebnissen. Die Visualisierung der bewerteten Informationen sei für die Kommunikation der Ergebnisse eines Nutzenbewertungsprozesses von großer Bedeutung.

Das Methodenpapier stellt kein umfassendes Lehrbuch dar und es werden nicht alle infrage kommenden statistischen oder grafischen Methoden im Einzelnen beschrieben.

Studien, die den Nutzen einer ergänzenden grafischen Aufbereitung numerischer Darstellungen untersuchen, zeigen zu den Endpunkten Wissen, Verständlichkeit, Risikowahrnehmung und Verständlichkeit kein konsistentes Ergebnis. Die evidenzbasierte Leitlinie „Evidenzbasierte Gesundheitsinformation“ kommt daher in Bezug auf diese Frage auf Basis einer niedrigen Evidenzqualität zu einer Kann-Empfehlung [1]. Die Nutzenbewertungen des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) beinhalten bereits jetzt verschiedene Visualisierungen, beispielsweise ein PRISMA-Flussdiagramm oder eine Landkarte zur Beleglage, sodass unklar ist, wo hier weiterer Bedarf an grafischen Verdeutlichungen besteht. In den Gesundheitsinformationen des IQWiG werden Grafiken vor allem zur ergänzenden Darstellung komplexer Sachverhalte eingesetzt, beispielsweise regelhaft in den Entscheidungshilfen zu Krebsfrüherkennungs-Untersuchungen. Abschnitt 7.7 „Umgang mit Zahlen und Risikoangaben“ wird um eine entsprechende Ausführung ergänzt. Ein weiterer Änderungsbedarf im Methodenpapier wird nicht gesehen.

2.2 Würdigung der Stellungnahmen zu Abschnitt 1.2 „Evidenzbasierte Medizin“

Stellungnehmende regen an, die Abgrenzung der Evidence-based Medicine im individuellen Behandlungsverhältnis von der Evidence-based Health Care bei der abstrakt-generellen Versorgungsgestaltung klarer herauszuarbeiten. Es bedürfe bei der Evidence-based Health Care der Reflexion, welche Auswirkungen Entscheidungen zur abstrakt-generellen Versorgung haben, um in diesem Kontext die Studienlage zu analysieren.

Der Abschnitt beschreibt die Bedeutung der evidenzbasierten Medizin allgemein und für die Arbeit des Instituts. Er stellt in der vorliegenden Fassung klar, dass das Institut selbst keine Entscheidungen zur Gestaltung der Versorgung trifft. Seine Hauptaufgabe besteht vielmehr

darin, vorgegebene Fragen seiner Auftraggeber wie des Gemeinsamen Bundesausschusses (G-BA), ob ein Nutzen oder Schaden durch eine Maßnahme nachzuweisen ist, möglichst zuverlässig zu beantworten. Aspekte zum Beispiel der Patientenversorgung, die außerhalb der Nutzenbewertung liegen, sollen dann vom G-BA bei seinen Entscheidungen berücksichtigt werden. Einzelheiten dazu sind in der im Abschnitt zitierten Verfahrensordnung des G-BA geregelt.

Stellungnehmende weisen darauf hin, dass die Evidenz für einige Themen / Krankheiten – insbesondere seltene Erkrankungen – lückenhaft sei. Das Institut solle darlegen, ob es in solchen Situationen auf Daten der Kostenträger (GKV) zurückgreifen könne.

Daten der Kostenträger sind grundsätzlich nicht dazu geeignet, Fragen der Nutzenbewertung zu beantworten [2]. Für welche anderen Fragestellungen Daten der Kostenträger genutzt werden können, ist insbesondere in Abschnitt 5.4 „Versorgungsanalyse“ beschrieben.

Stellungnehmende fordern mit Verweis z. B. auf § 5 der Arzneimittel-Nutzenbewertungsverordnung, das Methodenpapier generell auf die bestverfügbare Evidenz abzustellen.

Das Institut beschreibt in Abschnitt 1.2.4 allgemein, dass die Anforderungen an die Ergebnissicherheit einer Studie der zu beantwortenden Frage und der zu treffenden Entscheidung angemessen sein müssen. Welche Evidenz für welche Frage recherchiert und bewertet wird, ist in den vertiefenden Abschnitten zum Beispiel zur Bewertung von Arzneimitteln und zu nicht medikamentösen Verfahren detailliert beschrieben.

2.3 Würdigung der Stellungnahmen zu Abschnitt 1.3 „Gesundheitsökonomie“

In einigen Stellungnahmen wird angeregt, anstelle der Formulierung „Gesundheitsökonomie im engen Sinn“ den Begriff „gesundheitsökonomische Evaluation“ zu verwenden.

In den Allgemeinen Methoden wurden Formulierungen an den betreffenden Stellen angepasst.

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

In 1 Stellungnahme wird gefordert, die Aussage mit Quellen zu belegen, dass „die überwiegende Mehrheit der Länder, in denen Kosten-Nutzen-Bewertungen (KNBs) zur Entscheidungsfindung beitragen, keinen (festen) Schwellenwert“ habe.

Entsprechende Quellen wurden in Abschnitt 1.3.3 der Allgemeinen Methoden ergänzt.

2.5 Würdigung der Stellungnahmen zu Abschnitt 2.1 „Produktspezifische Verfahrensabläufe“

1 Stellungnahme fordert Kriterien, die es erlaubten nachzuvollziehen, weswegen auf eine Anhörung zum Berichtsplan verzichtet werden könne.

Sofern eine Anhörung nach den Methoden des Instituts vorgesehen ist, auf diese aber ggf. verzichtet wird, wird dies im jeweiligen Bericht kurz begründet. Überdies wird im Methodenpapier bereits erläutert, dass ein Verzicht insbesondere bei Berichten im Zusammenhang mit der Erstellung von Informationsmaterialien zu bestehenden Richtlinieninhalten sowie bereits zur Anhörung gestellter Methodik infrage kommt. Es ergibt sich kein Änderungsbedarf für die Allgemeinen Methoden.

Für die Auswahl externer Sachverständiger bei Dossierbewertungen, Kosten-Nutzen-Bewertungen nach § 35b SGB V, Potenzialbewertungen, Bewertungen gemäß § 137h SGB V und bei der Erstellung von Gesundheitsinformationen sollen die Auswahlkriterien konkret benannt werden.

Auf der Website des Instituts findet sich unter den FAQs für Sachverständige, die sich beteiligen möchten, als Antwort zu der Frage „Wie werden Sachverständige aus der Datenbank ausgewählt?“ eine Auflistung der Auswahlkriterien für externe Sachverständige (<https://www.iqwig.de/de/sich-beteiligen/fachwissen-einbringen/faq-fur-sachverstandige.3247.html>). Es ergibt sich kein Änderungsbedarf für die Allgemeinen Methoden.

1 Stellungnahme bezweifelt die Vollständigkeit der auf Seite 37 aufgeführten Institutsprodukte. Es fehlten Berichte zu Fragestellungen nach §§ 135 und 137c SGB V.

Die entsprechende Aufstellung beschreibt nur diejenigen Produkte, für die aufgrund von zeitlichen Restriktionen eine Auswahl über die genannte Datenbank erfolgt. Wie den unmittelbar vorausgehenden Ausführungen zu entnehmen ist, werden selbstverständlich auch in die anderen Produkte des Instituts externe Sachverständige eingebunden. Es ergibt sich kein Änderungsbedarf für die Allgemeinen Methoden.

Im gleichen Zusammenhang wird ein proaktives Zugehen auf die medizinische Fachwelt zur Einbindung von Sachverständigen vorgeschlagen.

Sowohl die Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) als auch eine Reihe von Fachgesellschaften sind in der Hinsicht angesprochen worden, dass sie ihre Mitglieder motivieren mögen, sich als Sachverständige zu registrieren oder sich an entsprechenden Ausschreibungen zu beteiligen. Das Institut wählt nämlich Sachverständige über solche, für jeden zugängliche Verfahren aus. Eine Benennung durch Fachgesellschaften kommt nicht in Betracht. Darüber hinaus ist anzumerken, dass die AWMF

in der Vergangenheit in den Fällen, in denen kein Sachverständiger aus der Datenbank rekrutiert werden konnte, konkret angesprochen wurde, dies jedoch in allen Fällen ohne Erfolg geblieben ist. Es ergibt sich kein Änderungsbedarf für die Allgemeinen Methoden.

2.6 Würdigung der Stellungnahmen zu Abschnitt 2.1.5 „Potenzialbewertungen“

In Stellungnahmen wird die systematische Beurteilung von Potenzialbewertungen durch die Fachgesellschaften empfohlen.

Sowohl das Gesetz als auch die Verfahrensordnung des G-BA sieht für Potenzialbewertungen im Rahmen der Erprobungsanträge gemäß § 137e SGB V weitreichende Vertraulichkeitsregelungen vor. Stellungnahmen Externer sind daher nicht möglich. Erst dann, wenn der G-BA eine Erprobungsrichtlinie gemäß § 137e SGB V beschließt, erhalten die themenspezifisch jeweils einschlägigen Fachgesellschaften Gelegenheit, sich hierzu zu äußern. Es ergibt sich kein Änderungsbedarf für die Allgemeinen Methoden.

Im Zusammenhang mit Potenzialbewertungen und Bewertungen gemäß § 137h SGB V sieht die Methodik optionale eigene Recherchen zur Unterstützung der Bewertung vor. Die Voraussetzungen für deren Durchführung sollen gemäß einer Stellungnahme detailliert dargestellt werden.

Hierzu wird klargestellt, dass die Grundlage der Bewertung in den benannten Fällen der an das Institut übermittelte Antrag darstellt. Eigene Recherchen dienen der Unterstützung der Bewertung, z. B. hinsichtlich der Einarbeitung in das medizinische Thema, der Bewertung von Endpunkten oder der Planung einer Erprobungsstudie. Es ergibt sich kein Änderungsbedarf für die Allgemeinen Methoden.

2.7 Würdigung der Stellungnahmen zu Abschnitt 2.1.6 „Bewertung gemäß § 137h SGB V“

1 stellungnehmende Organisation regt an, die Bewertung von neuen wissenschaftlichen Methoden zu Medizinprodukten hoher Risikoklasse an den IDEAL-Empfehlungen [3] auszurichten.

Entsprechend den gesetzlichen Vorgaben beschreibt das Methodenpapier 2 Stufen, die von neuen Behandlungsmethoden erreicht werden können, nämlich Erwartbarkeit (d. h. Potenzial) und Nachweis von Nutzen. Diesen beiden Stufen entsprechen die Phasen 2b („Exploration“) und 3 („Assessment“) gemäß den IDEAL-Empfehlungen. Hierbei werden von der IDEAL Collaboration in Phase 2b vergleichende, möglichst prospektive Studien zu patientenrelevanten Endpunkten gefordert, was gemäß dem Methodenpapier für den Nachweis eines Potenzials ausreicht. In Phase 3 sind nach den IDEAL-Empfehlungen primär randomisierte kontrollierte Studien (RCTs) durchzuführen, was auch im Methodenpapier für den Nutzenachweis regelhaft gefordert wird. Damit entsprechen die Evaluationsphasen und die Evidenzanforderungen, die von der IDEAL-Gruppe empfohlen werden, weitgehend den

beiden Stufen (Potenzial und Nutzen), die im Methodenpapier definiert werden. Ein Änderungsbedarf für das Methodenpapier ergibt sich nicht.

1 Stellungnahme weist darauf hin, dass das vom IQWiG neu aufgenommene Kapitel fälschlicherweise eine Bewertung von Medizinprodukten hoher Risikoklasse suggeriere.

Da das IQWiG allenfalls Behandlungsmethoden (mit oder ohne maßgebliches Medizinprodukt), jedoch keine einzelnen Medizinprodukte bewertet, wäre es in der Tat falsch, wenn eine Bewertung von Medizinprodukten im Methodenpapier genannt würde. Jedoch findet sich an keiner Stelle im Methodenpapier diese Formulierung, sodass hier kein Änderungsbedarf besteht.

2.8 Würdigung der Stellungnahmen zu Abschnitt 2.1.7 „Addendum“

1 Stellungnahme bemängelt die für Addenda beschriebene Methodik als der Bedeutung von Addenda nicht gerecht werdend.

Hierzu ist anzumerken, dass ein Addendum grundsätzlich der gleichen Methodik wie das zugrunde liegende Institutsprodukt folgt. Es ergibt sich kein Änderungsbedarf für die Allgemeinen Methoden.

2.9 Würdigung der Stellungnahmen zu Abschnitt 2.1.8 „Gesundheitsinformationen“

In 1 Stellungnahme wird angeregt, künftig in den Gesundheitsinformationen zu vermerken, wenn diese in Kooperation mit einer Selbsthilfeorganisation entstanden sind.

In Abschnitt 7.9.1 „Beteiligung von Betroffenen“ wird ausgeführt, dass bei der Priorisierung und Fokussierung der Themen im Rahmen der Konsultation von Selbsthilfegruppen auch Erfahrungen von Betroffenen berücksichtigt werden. Dazu wird bei der Erarbeitung eines neuen Themas regelhaft bei der Patientenvertretung des G-BA angefragt, themenspezifisch entsprechende Selbsthilfegruppen zu benennen. Es erfolgt dann eine Kontaktaufnahme seitens des Instituts. Bei einer positiven Rückmeldung werden mit den jeweiligen Organisationen Gespräche geführt, die Ergebnisse per Protokoll festgehalten und bei der Entwicklung der Informationen berücksichtigt. Außerdem erhalten die Selbsthilfegruppen die Möglichkeit, 1 Stellungnahme zu den Textentwürfen abzugeben.

Das Institut wird prüfen, ob und unter welchen Voraussetzungen eine allgemeine Danksagung, wie sie aktuell bereits für die externen Sachverständigen auf der Website des Instituts veröffentlicht wird, auch für die beteiligten Selbsthilfegruppen möglich – und von diesen erwünscht – ist. Ein Änderungsbedarf für die Allgemeinen Methoden ergibt sich nicht.

1 Stellungnehmender empfiehlt die tabellarische Aufstellung der bisherigen Themen und der vorgesehenen Themen.

In Abschnitt 7.2.1 „Themenkatalog gemäß dem Generalauftrag“ wird die Themenauswahl transparent beschrieben. Die Veröffentlichung einer umfangreichen Liste mit allen bereits bearbeiteten und zukünftigen Themen ist nicht Gegenstand eines Methodenpapiers. Zudem kann sich jeder und jede Interessierte auf unserer Website www.gesundheitsinformation.de einen Überblick über die bereits veröffentlichten Themen verschaffen. Es ergibt sich kein Änderungsbedarf für die Allgemeinen Methoden.

In 1 Stellungnahme wird angemerkt, dass im Rahmen der Aktualisierung von Gesundheitsinformationen die Einbindung von externen Gutachterinnen und Gutachtern bei umfassenden, grundlegenden inhaltlichen Änderungen nicht optional, sondern regelhaft stattfinden sollte.

Dies ist bei Aktualisierungen regelhaft Praxis. Das Institut wird den entsprechenden Absatz umformulieren.

In 1 Stellungnahme wird kritisiert, dass der Kreis der Organisationen, Organe und Institutionen, die im Rahmen des nicht öffentlichen Stellungnahmeverfahrens zu Gesundheitsinformationen zur Abgabe einer Stellungnahme aufgefordert würden, beschränkt sei. Im Kuratorium seien lediglich die Organisationen der Leistungserbringer und die Selbstverwaltungsorgane der Trägerorganisationen des Gemeinsamen Bundesausschusses, inklusive Patientenvertretern, repräsentiert. Somit bleibe unter anderem Personen und Institutionen der Wissenschaft sowie Herstellern von pharmazeutischen Produkten und Medizinprodukten diese Möglichkeit verwehrt.

In § 8 Abs. 4 der Stiftungssatzung ist die Möglichkeit zur Anhörung insbesondere bei Gesundheitsinformationen wie folgt geregelt: „Vor Abgabe von Empfehlungen des Instituts zu Angelegenheiten, welche unmittelbar die Patientenbelange tangieren, insbesondere zur Bereitstellung verständlicher allgemeiner Informationen zur Qualität und Effizienz in der Gesundheitsversorgung, ist den im Kuratorium vertretenen Repräsentanten der Patientinnen und Patienten sowie den Beauftragten der Bundesregierung für die Belange der Patientinnen und Patienten Gelegenheit zur Stellungnahme im Rahmen einer Anhörung zu geben.“

Die Möglichkeit zur Anhörung wurde vonseiten des IQWiG aus Interesse an der Beteiligung aller Mitglieder auf das gesamte Kuratorium erweitert. Im Kuratorium sind nicht nur, wie der Stellungnehmende fälschlicherweise beschreibt, die Organisationen der Leistungserbringer und die Selbstverwaltungsorgane der Trägerorganisationen des Gemeinsamen Bundesausschusses, inklusive Patientenvertretern, repräsentiert, sondern auch maßgebliche, nicht im G-BA vertretene Organisationen von Sozialpartnern und sonstige für das Gesundheitswesen relevante Organisationen, darunter auch die Verbände der Hersteller pharmazeutischer Produkte, die Bundesvereinigung Deutscher Apothekerverbände (ABDA) und die AWMF. Es liegt in der Verantwortung der Vertreterinnen und Vertreter, die Stellungnahme der von ihnen

vertretenen Firmen, Organisationen oder Institutionen zu organisieren oder diese zur Abgabe einer Stellungnahme anzuregen. So organisiert beispielsweise die AWMF die Anhörung der medizinischen Fachgesellschaften.

Es ergibt sich kein Änderungsbedarf für die Allgemeinen Methoden.

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

In 1 Stellungnahme wird die Frage gestellt, wer unter welchen Voraussetzungen Stellungnahmen zum vorläufigen Basisbericht abgeben kann.

Für die Health-Technology-Assessment(HTA)-Berichte gelten für die Abgabe von Stellungnahmen die gleichen Voraussetzungen wie für alle anderen Produkte des IQWiG, bei denen Stellungnahmeverfahren durchgeführt werden. Diese sind in Abschnitt 2.2.5 „Stellungnahmeverfahren“ des Methodenpapiers beschrieben; ein Änderungsbedarf für die Allgemeinen Methoden ergibt sich nicht.

In 1 Stellungnahme wird gefordert, dass ein Stellungnahmeverfahren nicht nur zum vorläufigen Basisbericht, sondern auch zum Herausgeberkommentar des IQWiG erfolgen sollte.

Die HTA-Berichte bestehen aus einem Basisbericht und aus einem Herausgeberkommentar. Der Basisbericht beinhaltet die Fragestellung des HTA-Berichts, eine Beschreibung des methodischen Vorgehens und die Antworten auf die Fragestellung des HTA-Berichts. Er wird durch externe Sachverständige erstellt. Zu einer vorläufigen Version des Basisberichts können Stellungnahmen abgegeben werden. Im Herausgeberkommentar erfolgt dann eine Einordnung der Arbeitsergebnisse durch das IQWiG.

Stellungnahmen zu IQWiG-Produkten können sich, sofern vorgesehen, grundsätzlich auf die projektspezifische Methodik eines Berichtes und/oder die Berücksichtigung qualitativ angemessener Unterlagen in den Berichten beziehen. Folgerichtig können Stellungnahmen zum vorläufigen Basisbericht, nicht aber zum Herausgeberkommentar abgegeben werden. Es ergibt sich somit kein Änderungsbedarf für die Allgemeinen Methoden.

In 1 Stellungnahme wird angeführt, dass ein Stellungnahmeverfahren zum Berichtsprotokoll wünschenswert sei.

Stellungnahmen können zum vorläufigen Basisbericht abgegeben werden. Hier können Hinweise zur projektspezifischen Methodik und/oder zur Berücksichtigung qualitativ angemessener Unterlagen abgegeben werden. Ein Änderungsbedarf für die Allgemeinen Methoden ergibt sich nicht.

2.11 Würdigung der Stellungnahmen zu Abschnitt 2.2.1 „Einbindung von Betroffenen in die Erstellung von IQWiG-Produkten“

In mehreren Stellungnahmen wird darauf hingewiesen, dass Betroffene in die Erstellung der HTA-Berichte eingebunden werden sollten.

Patientinnen und Patienten sowie Bürgerinnen und Bürger werden umfassend in die HTA-Berichte eingebunden. So stammen die Themenvorschläge für die HTA-Berichte von Versicherten und interessierten Einzelpersonen. Ebenso sind Patienten- und Bürgervertreterinnen und -vertreter in die Auswahl der Themen für die HTA-Berichte eingebunden.

Darüber hinaus wird der Vorschlag der Stellungnehmenden aufgegriffen und in Abschnitt 2.2.1 „Einbindung von Betroffenen in die Erstellung von IQWiG-Produkten“ des Methodenpapiers klargestellt, dass auch bei der Erstellung der HTA-Berichte eine mündliche Konsultation Betroffener erfolgt.

In mehreren Stellungnahmen wird, neben der Bezugnahme auf die HTA-Berichte, ganz allgemein gefordert, Betroffene grundsätzlich in alle Institutsprodukte einzubinden und deren Aussagen transparent darzustellen. Für die Einbindung wird eine stringente, evidenzbasierte Methodik gefordert, um eine adäquate Bestimmung der Patientenrelevanz von Endpunkten zu ermöglichen.

Mit Ausnahme von Potenzialbewertungen, Bewertungen gemäß § 137h SGB V und Addenda (siehe Tabelle 2 der Allgemeinen Methoden) werden Betroffene in die Produkte regelhaft eingebunden. Die Ausnahmen leiten sich aus den gesetzlich vorgegebenen und mit dem G-BA vereinbarten kurzen Bearbeitungsfristen bzw. Vertraulichkeitsregelungen ab. Addenda beziehen sich immer auf ein zugrunde liegendes Institutsprodukt, in das Betroffene regelhaft eingebunden werden, sofern es sich nicht um die hier begründeten Ausnahmeprodukte handelt.

Der Austausch mit Betroffenen veranschaulicht infrage kommende Endpunkte und ermöglicht ein tieferes Verständnis bzw. eine bessere Zuordnung zu den oben genannten Kriterien, ohne die Patientenrelevanz im Grundsatz zu definieren. Für dieses Ziel wird das beschriebene Vorgehen unter Abwägung des Aufwands und Ertrags als angemessen betrachtet.

In 1 Stellungnahme wird befürchtet, dass die Fokussierung auf Betroffene zum gänzlichen Fehlen der Patientenperspektive bzw. der Patientenvertreterperspektive führen könnte.

Das Institut hat bisher sehr gute Erfahrungen mit der Gewinnung Betroffener gemacht und teilt daher die Befürchtung nicht. Es wird aber die Entwicklung der Beteiligung von Betroffenen prüfen. Für die Allgemeinen Methoden ergibt sich kein Änderungsbedarf.

1 Stellungnahme sieht in der Findung von Betroffenen über die regelhafte Anfrage bei der Patientenvertretung des G-BA eine Lücke bezüglich Personen, die Routinemaßnahmen,

Präventionsleistungen oder geburtshilfliche Leistungen in Anspruch nehmen und somit keine Patientinnen oder Patienten im eigentlichen Sinne sind.

Die Beschreibung des Vorgehens in den Methoden als „regelmäßig“ impliziert die Möglichkeit eines anderen Vorgehens in solchen Fällen. Es ergibt sich kein Änderungsbedarf für die Allgemeinen Methoden.

An der Auswahl von Betroffenen zur Einbindung in die Erstellung von IQWiG-Produkten wurde bemängelt, dass im Falle von Kleinkindern oder Demenzerkrankten diese nicht selbst befragt werden könnten und hier ersatzweise nahe Angehörige hinzugezogen werden sollten.

Tatsächlich bezieht der in den Methoden in Abschnitt 2.2.1 beispielhaft beschriebene Personenkreis, der als betroffen zu verstehen ist, Angehörige explizit als mögliche Vertreterinnen oder Vertreter mit ein. Es ergibt sich kein Änderungsbedarf für die Allgemeinen Methoden.

1 Stellungnahme regt an, die verschiedenen Dokumente auf der Institutswebsite zum Thema „Patientensicht einbringen“ zusammenzuführen und verständlicher zu benennen.

Das Institut wird prüfen, inwiefern die Verständlichkeit der Institutswebsite zum Thema „Patientensicht einbringen“ verbessert werden kann. Für die Allgemeinen Methoden ergibt sich kein Änderungsbedarf.

2.12 Würdigung der Stellungnahmen zu Abschnitt 2.2.2 „Auswahl externer Sachverständiger“

In 1 Stellungnahme wird im Rahmen der Einbindung externer Sachverständiger ein Vorschlagsrecht einer Fachgesellschaft im Hinblick auf deren Evidenzbasierte-Medizin(EbM)-Kompetenz gefordert.

Alle sich berufen fühlenden Personen, auch die Mitglieder von Fachgesellschaften, können sich als externe Sachverständige in der Sachverständigendatenbank auf der Website des Instituts registrieren oder sich an Ausschreibungen beteiligen. Es ergibt sich kein Änderungsbedarf für die Allgemeinen Methoden.

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

1 Stellungnahme hinterfragt die Einbindung externer Sachverständiger in die Produkt-erstellung und fordert eine Konkretisierung der Sachverhalte, bezüglich derer externe Sachverständige beteiligt werden müssten oder könnten bzw. nicht beteiligt würden.

Dieser Punkt wurde bereits im Rahmen der Anhörung zu den Allgemeinen Methoden Version 4.2 gewürdigt (siehe das Dokument auf der Institutswebsite „Dokumentation und Würdigung

der Anhörung zum Entwurf der Allgemeinen Methoden 4.2, Version 1.0 vom 22.04.2015“, Abschnitt 2.2 „Würdigung der Stellungnahmen zu Abschnitt 2.1 Produktspezifische Verfahrensabläufe“). Es ergibt sich kein Änderungsbedarf für die Allgemeinen Methoden.

2.14 Würdigung der Stellungnahmen zu Kapitel 3 „Nutzenbewertung medizinischer Interventionen“

In 1 Stellungnahme wird gefordert, die Überschrift von Kapitel 3 in „Nutzen- und Schadenbewertung medizinischer Interventionen“ zu ändern, da der Schaden nicht nur die Abwesenheit eines Nutzens oder einen geringen Nutzen darstelle.

In Abschnitt 3.1.1 werden Schadensaspekte im Detail behandelt und es wird sehr deutlich herausgestellt, dass der Begriff Nutzenbewertung diese Schadensaspekte umfasst. Da der Begriff Nutzenbewertung zudem seit Langem etabliert ist, wird diesem Vorschlag nicht gefolgt.

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

In 1 Stellungnahme wird gefragt, wie entschieden werde, welche unerwünschten Wirkungen relevant seien, und ob es nicht wichtig sei, Patientenvertreterinnen und Patientenvertreter hierzu einzubeziehen.

Für diese Fragen wird auf Abschnitt 2.2.1 („Einbindung von Betroffenen in die Erstellung von IQWiG-Produkten“) der Allgemeinen Methoden verwiesen. Dort ist unter anderem beschrieben, dass und wie Betroffene insbesondere zur Festlegung patientenrelevanter Endpunkte eingebunden werden.

2.16 Würdigung der Stellungnahmen zu Abschnitt 3.1.4 „Endpunktbezogene Bewertung“

In mehreren Stellungnahmen wird die Hinzunahme des neuen Kriteriums, dass zur Ableitung eines Belegs aus lediglich 1 Studie mindestens 1000 Patientinnen oder Patienten pro Studienarm erforderlich sind, kritisiert. Zum einen sei diese Zahl viel zu hoch und zum anderen nicht begründet.

Regelhaft wird mehr als 1 Studie für die Ableitung eines Belegs gefordert. Der Beleg aus 1 Studie soll einen Ausnahmefall darstellen. Die Erfahrung der letzten Jahre zeigt, dass die pharmazeutischen Unternehmer in ihren Dossiers häufig Belege ableiten, obwohl nur 1 Studie vorliegt. Daher besteht die Notwendigkeit klarer und einfach zu prüfender Kriterien dafür, wann ein solcher Ausnahmefall vorliegt. Ein wichtiges Kriterium ist die Konsistenz der Ergebnisse aus 1 Studie über verschiedene Teilpopulationen hinweg. Dafür bietet sich die Betrachtung der einzelnen Studienzentren an. Zur Beurteilung der Konsistenz sind präzise Schätzungen innerhalb der Zentren notwendig. Für eine ausreichende Präzision werden 10 Ereignisse pro Arm und Zentrum angesetzt. Bei einer Ereignisrate von 10 % und

100 Patientinnen oder Patienten pro Arm und Zentrum sind etwa 10 Ereignisse zu erwarten. Geht man darüber hinaus von 10 Zentren aus, ergibt sich die angesetzte minimale Fallzahl von 1000 Patientinnen oder Patienten pro Behandlungsarm. In den Stellungnahmen und auch in der mündlichen Erörterung wurden von den Stellungnehmenden keine alternativen Kriterien oder Fallzahlen genannt, die eine ausreichend präzise Konsistenzprüfung ermöglichen.

In der mündlichen Erörterung wurde von einzelnen Stellungnehmenden vorgeschlagen, die formulierten Kriterien als orientierend aufzufassen und das im Dokument auch so zu benennen. Dem Vorschlag wird gefolgt und die Formulierung in Abschnitt 3.1.4 der Allgemeinen Methoden entsprechend angepasst.

2.17 Würdigung der Stellungnahmen zu Abschnitt 3.2.4 „Patientenberichtete Endpunkte“

In 1 Stellungnehmer äußert, es sei ihm unklar, was unter dem in diesem Abschnitt gebrauchten Ausdruck „offene Studien“ zu verstehen sei.

Der Begriff „offene Studie“ (englisch: „open-label study“) ist in der evidenzbasierten Medizin eine gängige Bezeichnung für eine nicht verblindete Studie und auch in Abschnitt 3.2.4 „Patientenberichtete Endpunkte“ so zu verstehen. Dies wird in den Allgemeinen Methoden präzisiert.

2.18 Würdigung der Stellungnahmen zu Abschnitt 3.2.5 „Nutzen und Schaden in kleinen Populationen“

In 1 Stellungnahme wird angeregt, in den Allgemeinen Methoden eine Definition für seltene Erkrankungen anzugeben.

In dem betreffenden Abschnitt 3.2.5 wird beschrieben, dass es keinen Grund dafür gibt, grundsätzlich andere Maßstäbe an die Nutzenbewertung in kleinen Populationen – so auch bei seltenen Erkrankungen – anzusetzen. Deshalb und auch vor dem Hintergrund, dass es keine einheitliche Definition für seltene Erkrankungen gibt, erscheint es wenig zielführend, in den Allgemeinen Methoden eine genaue Definition dafür anzugeben. Daraus ergibt sich insgesamt kein Änderungsbedarf für die Allgemeinen Methoden.

2.19 Würdigung der Stellungnahmen zu Abschnitt 3.3.2 „Studien zur Nutzenbewertung von Arzneimitteln“

In Abschnitt 3.3.2 der Allgemeinen Methoden werden beispielhaft mögliche Gründe für die Nichtdurchführbarkeit von RCTs aufgeführt. Einer der genannten Gründe ist die „starke Präferenz einer bestimmten Therapiealternative von Therapeuten und / oder Patienten“. In 1 Stellungnahme wird darauf hingewiesen, dass die Präferenz kein akzeptabler Grund für die fehlende Durchführbarkeit einer RCT sein könne, weil sie nicht überprüfbar und darüber

hinaus auch stark beeinflussbar sei. Weiterhin wird in der Stellungnahme angeführt, dass bei Unwissen, welche der Therapiealternativen hinsichtlich des Therapieziels die bessere sei, die Präferenz keine Rolle spielen dürfe. Liege dieses Unwissen nicht vor, sollte diese Studie tatsächlich undurchführbar sein, aber auch überflüssig, weil das Wissen über die Verhältnisse der Effekte der geplanten Alternativen schon existiere. Der Stellungnehmende fordert, diese Begründung zu streichen.

Der Argumentation des Stellungnehmenden wird gefolgt. Der Passus in den Allgemeinen Methoden wird entsprechend angepasst.

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

In 1 Stellungnahme wird kritisiert, dass in dem Abschnitt „Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V“ der Allgemeinen Methoden nahezu nicht auf den Schaden der bewerteten Interventionen eingegangen werde. Es werde ausschließlich von Nutzenbewertung und Zusatznutzen gesprochen.

Der betreffende Abschnitt beschreibt die Besonderheiten des Verfahrens und der Inhalte der frühen Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V, wie beispielsweise die grundsätzlichen Anforderungen an das Dossier des pharmazeutischen Unternehmers, und erläutert die Begriffe Wahrscheinlichkeit und Ausmaß des Zusatznutzens. Darin wird unter anderem Bezug genommen auf die Arzneimittel-Nutzenbewertungsverordnung (AM-NutzenV) als gesetzliche Grundlage [4] sowie auf die Dossievorlagen, die Teil der Verfahrensordnung des G-BA sind [5], und den Begriff Schaden im Zusammenhang mit der frühen Nutzenbewertung nicht verwenden. Stattdessen wird in der AM-NutzenV der Begriff Nebenwirkungen verwendet, ohne dass sich hieraus ein inhaltlicher Unterschied zwischen der frühen Nutzenbewertung und sonstigen Arzneimittelbewertungen ergibt. Für die frühe Nutzenbewertung von Arzneimitteln gelten daher auch die übrigen Abschnitte der Allgemeinen Methoden, wie beispielsweise Abschnitt 3.1.1 „Patientenrelevanter Nutzen und Schaden“, in dem die Begriffe Nutzen und Schaden voneinander abgegrenzt werden. Insgesamt ergibt sich daraus kein Änderungsbedarf für die Allgemeinen Methoden.

In 1 Stellungnahme wird vorgeschlagen, Abschnitt 3.3.3 der Allgemeinen Methoden um mehrere Unterabschnitte zu erweitern, in denen jeweils der Umgang des IQWiG mit verschiedenen Punkten im Zuge der frühen Nutzenbewertung beschrieben werden sollte, z. B. die Prüfung der Zweckmäßigkeit der Vergleichstherapie, der Umgang mit den Angaben zu zugelassenen Anwendungsgebieten oder die Bewertung des medizinischen Nutzens.

In Abschnitt 3.3.3 sind zum einen die Anforderungen an das Dossier des pharmazeutischen Unternehmers sowie zum anderen die Grundzüge und methodischen Besonderheiten der frühen Nutzenbewertung beschrieben. Die vom Stellungnehmenden vorgeschlagenen Abschnitte beziehen sich entweder auf Punkte, die nicht Bestandteil der Dossierbewertung sind

(wie die Bewertung des medizinischen Nutzens), oder auf Punkte, die an anderer Stelle bereits hinreichend beschrieben sind.

Beispielsweise ist in Abschnitt 3.3.3 beschrieben, dass das IQWiG überprüft, ob die vom pharmazeutischen Unternehmer gewählte Vergleichstherapie als zweckmäßig im Sinne des § 35a SGB V und der AM-NutzenV gelten kann. Damit zieht das IQWiG die in den gesetzlichen Grundlagen [4] und der Verfahrensordnung des G-BA [5] beschriebenen Kriterien für die zweckmäßige Vergleichstherapie heran. Diese Kriterien sind in der AM-NutzenV und der Dossievorlage (Modul 3) hinreichend beschrieben.

Insgesamt ergibt sich kein Änderungsbedarf für die Allgemeinen Methoden.

In 1 Stellungnahme wird gefordert, eine Methodik dafür zu formulieren, inwieweit die für eine Population verfügbare Evidenz auf andere Populationen übertragen werden kann. Der Stellungnehmende bezieht sich dabei auf das Arzneimittelversorgungsstärkungsgesetz (AMVSG), das eine Änderung der AM-NutzenV vorsieht, nach der der G-BA prüfen soll, ob für pädiatrische Patientengruppen, die in der Studienpopulation nicht oder nicht hinreichend vertreten sind und für die die Zulassung aufgrund der Übertragung von Evidenz ausgesprochen wurde, ein Zusatznutzen anerkannt werden kann.

Für die Übertragung von Evidenz gibt es keine allgemein anerkannten Anforderungen. Dies beschreibt auch der Gesetzgeber in der Begründung des AMVSG [6]. In einer Nutzenbewertung sind die konkreten Daten dahin gehend zu prüfen, ob auf deren Basis eine Extrapolation möglich ist oder nicht. Laut AMVSG ist eine Übertragung der Evidenz nur dann zulässig, wenn dies nach dem Stand der wissenschaftlichen Erkenntnisse auch im Hinblick auf die Nutzenbewertung zulässig und begründet ist. In der Regel wird es sich demnach um eine Einzelfallentscheidung handeln. Darüber hinaus ist der G-BA laut Gesetz aufgefordert, Einzelheiten zur Übertragung von Evidenz in seiner Verfahrensordnung zu regeln. Eine solche Beschreibung liegt noch nicht vor. Insgesamt ergibt sich kein Änderungsbedarf für die Allgemeinen Methoden.

In 1 Stellungnahme wird gefordert, dass in Abschnitt 3.3.3 weitere Unterabschnitte zu den vom IQWiG angewandten Methoden in der Dossierbewertung eingefügt werden sollen. Dabei soll das IQWiG sein Vorgehen zur Bewertung der Epidemiologie (z. B. zur Bewertung unterschiedlicher Quellen und der Bedeutung früherer Beschlüsse) darstellen. Auch ein Unterabschnitt für die Darstellung der Überprüfung der Angaben zu den Kosten für die GKV soll eingefügt werden (z. B. welche Packungen für eine Kostenkalkulation herangezogen werden, welche GKV-Zusatzleistungen berücksichtigt werden oder welche Preisabschläge kalkuliert werden). Der Stellungnehmende nennt als Grund dafür Abweichungen von den G-BA-Beschlüssen und Inkonsistenzen in der Bewertung des IQWiG.

In jeder Dossierbewertung werden zunächst die Herleitung der vom pharmazeutischen Unternehmer angegebenen Anzahl der GKV-Patientinnen und -Patienten in der Zielpopulation sowie die Herleitung der 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 pharmazeutischen Unternehmers transparent dargestellt und jeweils begründet. Die methodischen Grundlagen für die Bewertung sind in der Dossievorlage (Abschnitt 3.3.3 und 3.3.4) dargestellt.

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. Damit müssen die Bewertungsergebnisse des IQWiG und des G-BA nicht notwendigerweise gleich ausfallen.

Bei Änderungen der Datengrundlage verschiedener Dossiers zu gleichen Indikationen im Zeitverlauf sind daraus resultierende Abweichungen bei den IQWiG-Bewertungen möglich.

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

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

In 1 Stellungnahme wird angeregt, konkret zu beschreiben, wie Daten aus randomisierten und nicht randomisierten Studien gemeinsam betrachtet werden, denn gemäß § 137h SGB V sei in einer Erprobung einer neuen Medizinproduktmethode neben der RCT für den Nutzenachweis auch parallel eine Beobachtungsstudie vorgeschrieben. Auch in anderen Stellungnahmen wird vorgeschlagen, insbesondere Registerdaten in der Nutzenbewertung mehr Beachtung zu schenken.

In seiner Verfahrensordnung hat der G-BA aus der Begründung zum Versorgungsstärkungsgesetz [7] die Anregung übernommen, dass Krankenhäuser im Rahmen einer Erprobung gemäß § 137h SGB V auch „im Sinne einer Beobachtungsstudie flankierende Daten zu Wirksamkeit und Sicherheit“ erheben. Die Verwendung der Begriffe Wirksamkeit und Sicherheit macht deutlich, dass diese flankierenden Daten regelhaft keine Bedeutung haben für die Bewertung des Nutzens. Nur in den wenigen Fällen, in denen sich aus einer vergleichenden Beobachtungsstudie sehr deutliche Unterschiede ergeben, lassen sich diese Unterschiede hinreichend sicher als Nutzen oder Schaden interpretieren (siehe Abschnitt 3.2.2 im Methodenpapier). Auch international entspricht diese Sichtweise dem wissenschaftlichen Standard, wie beispielsweise eine aktuelle Analyse zu Medizinprodukten in der Kardiologie zeigt [8]. In der methodischen Literatur wird weiterhin davon abgeraten, in die Bewertung von Interventionseffekten nicht randomisierte Evidenz einzubeziehen, wenn randomisierte Evidenz vorhanden ist; und auch von gemeinsamen Metaanalysen randomisierter und nicht randomisierter Evidenz wird abgeraten [9]. Daher besteht weiterhin keine Notwendigkeit, systematisch oder regelhaft nicht randomisierte Evidenz in Nutzenbewertungen einfließen zu lassen oder das Methodenpapier diesbezüglich anzupassen.

In 1 Stellungnahme wird empfohlen, Nutzenbewertungen im nicht medikamentösen Bereich „auf neue Wirkprinzipien zu fokussieren“, hierbei die physikalische Wirkungsweise von

Medizinprodukten stärker zu beachten und anhand „anzuwendender Naturgesetze“ und „der technisch-mathematischen Berechenbarkeit“ von Wirkungen die Bewertung durchzuführen. Zu hohe methodische Anforderungen verhinderten vor allem Schritttinnovationen, so die Sorge der Stellungnehmenden.

Welche Modifikation eines Medizinprodukts dazu führt, dass es sich um eine neue, zu bewertende Methode handelt, wird nicht vom IQWiG, sondern vom G-BA entschieden. Im Kontext des § 137h SGB V hat der Gesetzgeber festgelegt, dass Methoden nur dann zu bewerten sind, wenn sie ein neues theoretisch-wissenschaftliches Konzept darstellen. Daher ist die Sorge des Stellungnehmenden unbegründet, dass medizintechnische „Detailänderungen und Weiterentwicklungen“ durch Nutzenbewertungen blockiert werden könnten. Es wird jedoch nicht möglich sein, aus einem theoretisch-wissenschaftlichen Konzept oder Wirkprinzip direkt einen patientenrelevanten Nutzen oder Schaden abzuleiten. In der Geschichte der Medizin gibt es viele Beispiele, bei denen plausible Wirkprinzipien sich als unwirksam oder schädlich erwiesen haben [10]. Der Hauptgrund, der zu Irrtümern in der Medizin führt, ist ein zu starkes Vertrauen auf pathophysiologische Prinzipien [11]. Daher ergibt sich aus diesem Vorschlag keine Änderung im Methodenpapier.

Eine weitere stellungnehmende Organisation äußert die Hoffnung, dass im nicht medikamentösen Bereich inzwischen doch häufiger aussagekräftige randomisierte Studien durchgeführt werden und damit für Nutzenbewertungen erwartet werden können. Hierbei wird auf das Förderprogramm „Klinische Studien“ des Bundesministeriums für Bildung und Forschung (BMBF) und der Deutschen Forschungsgemeinschaft (DFG) verwiesen.

Den Stellungnehmenden ist insofern zuzustimmen, als die finanzielle Förderung klinischer Studien zu nicht medikamentösen Interventionen sich weltweit verbessert hat – insbesondere im Bereich der Medizinprodukte und Biotechnologien [12]. Die internationalen Erfahrungen in der Bewertung von Behandlungsmethoden mit Medizinprodukten weisen jedoch darauf hin, dass die Evidenzbasis in diesem Bereich weiterhin schwach ist [13]. Aktuelle Untersuchungen aus der Kardiologie, Chirurgie und Orthopädie zeigen, dass für viele neu eingeführte Behandlungsmethoden mit Medizinprodukten keine oder nur wenig aussagekräftige randomisierte Studien vorliegen [14-16]. Inwieweit die im April 2017 vom Europäischen Parlament verabschiedete EU-Medizinprodukte-Verordnung an dieser Situation etwas ändert, muss abgewartet werden. Daher ist es angebracht, unverändert im Methodenpapier darauf hinzuweisen, dass „gerade im nicht medikamentösen Bereich [...] oft aussagekräftige randomisierte Studien“ fehlen.

2.22 Würdigung der Stellungnahmen zu Abschnitt 3.5 „Diagnostische Verfahren“

1 stellungnehmende Organisation stört sich an der „Auffassung des IQWiG, wonach auch der Nutzen von diagnostischen Maßnahmen zunächst von der Existenz und auch dem Nutzen einer Behandlungsmethode abhängt“. Zum Ersten könne man im konkreten Einzelfall vor einer Testdurchführung nicht wissen, ob das Testergebnis zu einer Therapie führe. Zum Zweiten

werde auch bei Nichtvorhandensein einer Therapie den Betroffenen durch die Feststellung einer Diagnose geholfen, weil so „oft jahrelange Diagnoseodysseen“ vermieden werden könnten.

Der Stellungnahme liegen Missverständnisse zugrunde. Zum einen fordert das methodische Vorgehen des IQWiG nicht, dass bei allen Patientinnen und Patienten nach Durchführung eines diagnostischen Tests therapeutische (oder präventive) Maßnahmen erfolgen. Es ist ausreichend, wenn sich bei einem Teil der Patientinnen und Patienten solche Maßnahmen ergeben und zu einem Nutzen führen. Dass im konkreten Einzelfall vor der Testdurchführung unklar ist, ob und welche Folgemaßnahmen sich ergeben, liegt in der Natur der Sache. Zum anderen wäre in der Tat ein möglicher Nutzen eines diagnostischen Tests auch darin zu sehen, dass dieser durch die Feststellung einer Krankheit die Durchführung weiterer Diagnostik überflüssig macht und so die Belastung der Patientin oder des Patienten reduziert. Diese Situation, „dass das zu prüfende diagnostische Verfahren ein anderes, bereits etabliertes diagnostisches Verfahren lediglich ersetzen soll“, wird im Methodenpapier bereits explizit dargestellt, sodass keine Ergänzungen notwendig sind.

1 stellungnehmende Organisation erklärt, dass es in Interaktionsdesigns zu diagnostischen Verfahren nicht erforderlich sei, dass bei einer Stichprobenunvollständigkeit die ausgewertete Stichprobe repräsentativ für die initiale Gesamtstichprobe sei. Wesentlich sei allein, dass die Stichprobencharakteristika genau beschrieben seien und sich nicht zwischen den Interventionsgruppen unterscheiden.

Die Argumentation der Stellungnehmenden ist nachvollziehbar. Auch die beiden genannten Bedingungen (die Strukturgleichheit der Gruppen wird aufrechterhalten und eine Stichprobencharakterisierung erlaubt den Vergleich mit der eigenen Zielpopulation) erscheinen sinnvoll. Die Formulierung im Methodenpapier wurde daher dementsprechend angepasst: Anstelle einer „repräsentativen Stichprobe“ wird nun auf „eine klar charakterisierte und in den Therapiearmen gleichartig selektierte Stichprobe“ verwiesen.

1 stellungnehmende Organisation kritisiert, dass eine separate Bewertung von diagnostischen und therapeutischen Interventionen nur bei den Screeningverfahren (Abschnitt 3.6), nicht aber bei den diagnostischen Verfahren (Abschnitt 3.5) genannt sei. Das Prinzip der Linked Evidence sei im Grundsatz bei allen diagnostischen Verfahren (inklusive Screening) in Betracht zu ziehen.

Die Argumentation der Stellungnehmenden ist in der Sache nachvollziehbar. Tatsächlich aber wird im Methodenpapier das Prinzip der Linked Evidence, also die separate Bewertung der diagnostischen Güte und therapeutischen Konsequenz, auch in Abschnitt 3.5 zu den diagnostischen Methoden aufgeführt. So heißt es an einer Stelle, dass „es nicht immer notwendig [ist], erneut die gesamte diagnostisch-therapeutische Kette zu untersuchen, weil sich die therapeutischen Konsequenzen, die sich aus dem neuen Test ergeben, nicht von

denen des bisherigen Tests unterscheiden“ (mit Zitierung einer zentralen Quelle zu Linked Evidence [17]). Dass Linked Evidence im Bereich von Screeningmaßnahmen stärker betont wird als bei diagnostischen Interventionen, liegt daran, dass diese Methodik im Bereich Screening häufiger einsetzbar ist, weil hier bei asymptomatischen Personen die therapeutischen Konsequenzen eines diagnostischen Tests überschaubarer und damit sicherer abschätzbar sind. Bei Patientinnen und Patienten mit Symptomen, bei denen zahlreiche Differenzialdiagnosen in Betracht kommen und sich dementsprechend auch zahlreiche therapeutische Konsequenzen ergeben können, ist das Prinzip der Linked Evidence erheblich schwerer anwendbar. Es ergibt sich somit kein Bedarf an Änderungen im Methodenpapier.

1 stellungnehmende Organisation rät dazu, diagnostische und prognostische Untersuchungsmethoden separat zu betrachten, „da sich die Anforderungen an die genetische Beratung [...] deutlich unterscheiden“, was dann „Unterschiede bei der Kosten-Nutzen-Analyse erwarten“ lasse.

In der Bewertung diagnostischer oder prognostischer Untersuchungsmethoden prüft das IQWiG auftragsgemäß den Nutzen und Schaden, analysiert jedoch weder rechtliche noch gesundheitsökonomische Aspekte systematisch. Daher ergibt sich auch hier kein Änderungsbedarf.

2.23 Würdigung der Stellungnahmen zu Abschnitt 3.6 „Früherkennung und Screening“ und Abschnitt 3.7 „Prävention“

In 1 Stellungnahme wird die Annahme (in Abschnitt 3.5) hinterfragt, „dass der Wert diagnostischer Verfahren ausschließlich durch sich anschließende therapeutische oder präventive Verfahren definiert wird“. Die Ausführungen zu Überdiagnosen passten hierzu nicht: Wenn Überdiagnosen als patientenrelevanter Endpunkt gesehen würden, müsste dies auch für Diagnosen gelten.

Die Stellungnahme verwendet eine Formulierung, die in dieser Strenge nicht im Methodenpapier enthalten ist. So lautet die vom IQWiG gewählte Formulierung: „Diagnostische Verfahren zeichnen sich dadurch aus, dass ihr gesundheitsbezogener Nutzen (oder Schaden) im Wesentlichen erst dadurch zustande kommt, dass sich ihnen therapeutische oder präventive Verfahren anschließen.“ Die Wortwahl („im Wesentlichen“, nicht aber „ausschließlich“) macht deutlich, dass im Einzelfall durchaus Situationen vorstellbar sind, in denen diagnostische Interventionen direkte Effekte haben. Wenn ein diagnostischer Test 2 sehr unterschiedliche Gesundheitszustände mit hoher Sicherheit unterscheiden kann, kann die jeweilige Diagnose als hinreichend sicheres Surrogat für patientenrelevante Konsequenzen aus dem Testergebnis gelten. Eine solche Situation ist bei Screeningmaßnahmen häufiger anzutreffen als in der Diagnostik symptomatischer Personen. In beiden Bereichen können jedoch schwere neu diagnostizierte Krankheitszustände unter Morbidität als patientenrelevant angesehen werden. Vorstellbar ist hierbei auch, dass sowohl Diagnosen als auch Überdiagnosen auch dann bewertungsrelevant sind, wenn es nicht um ein Screening, sondern um

Diagnostik geht. Weil die von den Stellungnehmenden gesehene Diskrepanz somit nicht vorhanden ist, ergibt sich kein Änderungsbedarf.

In 1 Stellungnahme wird kritisiert, dass in diesen Abschnitten die Bezeichnung „Patientinnen / Patienten“ verwendet werde, obwohl es sich um gesunde Personen handle.

In beiden kritisierten Abschnitten tauchen die Wörter „Patientinnen / Patienten“ allein nicht auf. Stattdessen wird der Begriff „Personen“ verwendet. Dass „patientenrelevante Endpunkte“ genannt werden, ist inhaltlich korrekt, weil im sozialrechtlichen Kontext auch die gesundheitlichen Effekte von Prävention, Früherkennung und Screening sich auf die Endpunkt-kategorien Mortalität, Morbidität und gesundheitsbezogene Lebensqualität (und damit auf Patientinnen und Patienten) beziehen. Ein Bedarf an einer geänderten Wortwahl ergibt sich nicht.

2.24 Würdigung der Stellungnahmen zu Abschnitt 3.8 „Potenzialbewertung“

Im Kontext der neu eingeführten §-137h-Bewertung fragt 1 stellungnehmende Organisation, in welchen Situationen das IQWiG ergänzend Studien recherchiere. Solche Recherchen seien in Abschnitt 2.1.6 genannt, würden aber in Abschnitt 3.8 nicht näher erläutert.

Die Stellungnahme bezieht sich auf den folgenden Satz in Abschnitt 2.1.6: „Optional kann zur Unterstützung der Bewertung eine eigene Recherche des Instituts durchgeführt werden.“ Es ist jedoch falsch, hieraus abzuleiten, dass „das IQWiG ergänzend Studien recherchiert“, um „die Informationen [für die §-137h-Bewertung] zu vervollständigen“. Basis der §-137h-Bewertungen sind die von Krankenhäusern und Medizinprodukteherstellern vorgelegten Unterlagen. Die zur Bewertung der Behandlungsmethode vorgelegte Evidenz wird nicht vom IQWiG vervollständigt. Sofern das IQWiG eigene Recherchen durchführt, dienen diese anderen Zwecken, beispielsweise der Einarbeitung in das medizinische Thema, der Bewertung von Endpunkten oder der Planung einer Erprobungsstudie. Erst wenn nach der initialen Bewertung ein Potenzial feststellbar ist, „wird das IQWiG ergänzend beauftragt, die Evidenz für die jeweilige Methode systematisch zu erfassen und auszuwerten“ [18].

Um diesen wichtigen Sachverhalt unmissverständlich darzustellen, wurde in Abschnitt 3.8 der folgende Satz eingefügt: „Eigene Recherchen können zur Unterstützung der Bewertung durchgeführt werden, es ist jedoch nicht Aufgabe oder Ziel, die vorgelegten Unterlagen zu vervollständigen.“

In 1 Stellungnahme wird kritisiert, dass aus nicht vergleichenden Studien Effekte abgeleitet würden. Ein Potenzial sei das Potenzial eines Effektes. Da nicht vergleichende Studien für sich genommen nur Prä-post-Vergleiche erlaubten, sei es allenfalls über unadjustierte indirekte Vergleiche zwischen nicht vergleichenden Studien möglich, Effekte abzuschätzen.

Die Stellungnahme weist zu Recht darauf hin, dass eine nicht vergleichende Studie allein keine Effektschätzung zulässt. Wie von den Stellungnehmenden richtig vermutet (und vom

IQWiG auch bereits so methodisch angewendet), sind an dieser Stelle unadjustierte indirekte Vergleiche zwischen nicht vergleichenden Studien gemeint. Der Anregung der Stellungnehmenden folgend wird die Formulierung dieses Punktes präzisiert, indem auf den „indirekten Vergleich zu den Ergebnissen anderer Studien (Literaturkontrollen)“ explizit Bezug genommen wird.

In mehreren Stellungnahmen wird die untere Grenze angesprochen, die für die Feststellung eines Potenzials erforderlich ist. In 1 Stellungnahme wird die Festlegung von Grenzwerten für das relative Risiko als zu strikt und kompliziert angesehen. Stattdessen wird vorgeschlagen, die Gleichgerichtetheit von Effekten zu prüfen.

Wie von den Stellungnehmenden treffend erläutert, wurden im Bereich der Potenzialbewertung „durch den Gesetzgeber offene Begriffe eingeführt [...], die den klassischen Kategorisierungen der EbM nicht unbedingt entsprechen“. In der vom Institut vorgeschlagenen Operationalisierung des Potenzialbegriffs sind daher die Grenzwerte für das relative Risiko nur als „ungefähre Grenzen“ benannt worden. Die praktische Anwendung dieser Grenzen wird zeigen, inwieweit dieses Vorgehen angemessen und umsetzbar ist. Der Vorschlag, statt des relativen Risikos zu prüfen, ob die vorgelegten Effekte gleichgerichtet sind, würde die Evidenzanforderungen an ein Potenzial eher erhöhen, weil für eine Gleichgerichtetheit mindestens 2 vergleichende Studien erforderlich sind, während bislang 1 vergleichende Studie (oder 1 indirekter Vergleich zwischen 2 Fallserien) mit einem entsprechend großen Effekt für die Feststellung eines Potenzials ausreicht. Daher erscheint es aktuell nicht angebracht, das Methodenpapier diesbezüglich zu ändern.

Nach dem Wunsch 1 stellungnehmenden Organisation soll im Methodenpapier explizit genannt werden, dass Bewertungen gemäß § 137h SGB V auch Aussagen zu der „flankierenden Beobachtungsstudie“ enthalten, da eine solche Studie gemäß der Verfahrensordnung des G-BA vorzusehen sei.

Angesichts der geringen wissenschaftlichen Bedeutung, die Beobachtungsstudien für den Vergleich von Interventionen haben, erscheint es nicht erforderlich, im Methodenpapier spezifisch darauf hinzuweisen, dass Bewertungen gemäß § 137h SGB V auch Aussagen hierzu enthalten.

2.25 Würdigung der Stellungnahmen zu Abschnitt 4.1 „Einführung“ des Kapitels 4 „Kosten-Nutzen-Bewertung medizinischer Interventionen“

In 1 Stellungnahme wird gefordert, die Aufzählung in Abschnitt 4.1 zu ergänzen. Es sollen die Definitionen der Zielpopulation und der Zielgrößen als weitere Entscheidungsparameter für die Erarbeitung einer Kosten-Nutzen-Bewertung aufgenommen werden.

Zum einen sind die Patientengruppe und die Zielgrößen gemäß § 35b SGB V im Auftrag durch den G-BA festzulegen. Zum anderen handelt es sich nicht um eine abschließende

Aufzählung der Entscheidungen, die für eine Kosten-Nutzen-Bewertung zu treffen sind. In den Allgemeinen Methoden wurde die Auflistung daher nicht ergänzt, der Satz aber entsprechend angepasst, um zu verdeutlichen, dass es sich nicht um eine abschließende Aufzählung handelt.

2.26 Würdigung der Stellungnahmen zu Abschnitt 4.1.3 „Zeithorizont“

In 1 Stellungnahme wird die Einführung des Begriffs der Modellierung in diesem Abschnitt gewünscht. Zudem soll auf die Berücksichtigung langfristiger Lebensqualitätsgewinne eingegangen werden.

In den Allgemeinen Methoden wurde ein einführender Satz zu entscheidungsanalytischen Modellen ergänzt. Die Fokussierung auf langfristige Lebensqualitätsgewinne ist jedoch nicht sinnvoll. Durch eine Modellierung sollen sämtliche langfristige Effekte auf die Kosten und den Nutzen berücksichtigt werden und nicht ausschließlich mögliche Lebensqualitätsgewinne.

2.27 Würdigung der Stellungnahmen zu Abschnitt 4.1.7 „Spezifische Datenquellen der Gesundheitsökonomie“

1 Stellungnehmender regt an, dass im Rahmen der geforderten Beschreibung und Begründung der Kontrolle von Störgrößen für Auswertungen von Sekundärdaten explizit auf die zeitabhängigen Störgrößen eingegangen werden sollte.

Die Formulierung in den Allgemeinen Methoden umfasst alle Arten von Störgrößen, sodass die zeitabhängigen Störgrößen bereits abgebildet werden. Für das Methodenpapier ergibt sich daher kein Änderungsbedarf.

2.28 Würdigung der Stellungnahmen zu Abschnitt 4.3.3 „Maß des Gesamtnutzens“

In 2 Stellungnahmen wird auf die nach Ansicht der Stellungnehmenden missverständliche Formulierung aufmerksam gemacht, dass die Nutzwerte bei an klinischen Studien Teilnehmenden erhoben worden sein sollten. In 1 Stellungnahme wird vorgeschlagen, den Satz um Teilnehmerinnen und Teilnehmer an Lebensqualitätsstudien zu erweitern.

Diesem Vorschlag wird nicht gefolgt, da der Begriff „Lebensqualitätsstudien“ nicht eindeutig definiert ist. Die Formulierung in den Allgemeinen Methoden wurde jedoch angepasst.

In 1 Stellungnahme wird darauf hingewiesen, dass eine Spezifizierung, wie Endpunkte zu gewichten seien, fehle. Im Weiteren wird erläutert, dass eine Gewichtung multipler Endpunkte durch den Entscheidungsträger außerhalb der Expected Utility Theory voraussichtlich in den meisten Fällen zu deutlich mehr Annahmen führe als bei den oben genannten Gesamtnutzenmaßen und zudem weniger explizit und nachvollziehbar sei. Es wird vorgeschlagen, eine Sensitivitätsanalyse z. B. mittels der Berechnung von qualitätsadjustierten Lebensjahren durchzuführen.

Nach § 35b SGB V legt der G-BA in seiner Beauftragung einer KNB unter anderem fest, welches Maß für den Gesamtnutzen bei der Bewertung zu berücksichtigen ist. Für alle Bereiche der KNB inklusive des Maßes des Gesamtnutzens sind Untersuchungen der Unsicherheit vorgesehen. Dass eine Gewichtung multipler Endpunkte außerhalb der Expected Utility Theory weniger explizit und nachvollziehbar ist, ist in der Stellungnahme nicht mit entsprechenden Quellen belegt und bleibt somit eine Mutmaßung. Eine zusätzliche Sensitivitätsanalyse mittels eines anderen als des im Auftrag festgelegten Maßes des Gesamtnutzens wird nicht als weiterführend betrachtet; daher besteht kein Änderungsbedarf.

2.29 Würdigung der Stellungnahmen zu Abschnitt 4.3.4 „Datengrundlage“

1 Stellungnahme fragt nach, warum die systematische Recherche nach Publikationen zum Maß des Gesamtnutzens nicht nach dem in Abschnitt 8.1 und 8.2 beschriebenen Vorgehen durchgeführt werde.

Hier handelt es sich um ein Missverständnis: Die Recherche wird nach dem Vorgehen in Abschnitt 8.2 durchgeführt. Im Methodenpapier wird ein Verweis eingefügt, um dies klarzustellen.

2.30 Würdigung der Stellungnahmen zu Abschnitt 4.4.2 „Abgrenzung der Kosten“

In 1 Stellungnahme wird gewünscht, dass Auswirkungen auf die Freizeit von Betroffenen und Angehörigen sowie der Aufwand für unbezahlte Arbeit (z. B. Hausarbeit) regelhaft in der gesellschaftlichen Perspektive bei einer KNB berücksichtigt werden. Weiterhin wird eine Klärung gewünscht, inwiefern Kosten für Erkrankungen bei einer Lebensverlängerung berücksichtigt werden.

Wie auch in Abschnitt 4.4.2 der Methoden beschrieben, ist die Berücksichtigung dieser Kostenkomponenten nur optional – und nicht regelhaft – vorgesehen, da diese nur selten in empirischen Studien erhoben werden. Liegen jedoch repräsentative und valide Informationsquellen zum Freizeitverlust bzw. zu den Kosten für unbezahlte Arbeit (z. B. Hausarbeit) vor, können diese z. B. je nach Relevanz der finanziellen Auswirkungen berücksichtigt werden. In Abschnitt 4.2.2 ist weiterhin auch die Berücksichtigung von Kosten bei einer Lebensverlängerung erläutert. So werden interventionsassoziierte zukünftige Kosten berücksichtigt. Nicht interventionsassoziierte Kosten können in separaten Sensitivitätsanalysen berücksichtigt werden. Somit ergibt sich kein Änderungsbedarf in den Allgemeinen Methoden.

2.31 Würdigung der Stellungnahmen zu Abschnitt 4.6.4 „Sonderkonstellationen“

In 1 Stellungnahme wird argumentiert, dass bei beiden beschriebenen Sonderkonstellationen, in denen sich nicht unmittelbar eine Empfehlung anhand der Effizienzgrenze ableiten lasse, die aktuell zu bewertende Intervention als kosteneffektiv anzusehen sei und erst die nachfolgende Bewertung von den genannten Einschränkungen betroffen sei. Zudem stelle eine

Intervention mit sehr niedrigen inkrementellen Kosten und sehr großem inkrementellem Nutzen eine weitere Sonderkonstellation dar. Überdies wird angeregt, den folgenden Satz zu streichen: „Hier kann die Ausgaben-Einfluss-Analyse weitere Daten liefern, indem die Auswirkungen auf das Budget dargestellt werden.“ Dies wird damit begründet, dass es nicht sinnvoll sei, die Kosten-Nutzen-Bewertung durch eine Ausgaben-Einfluss-Analyse zu ersetzen.

Die zu bewertende Intervention bestimmt nicht die Effizienzgrenze. Die beiden in den Allgemeinen Methoden dargestellten Sonderkonstellationen werden durch die Lage der letzten Intervention auf der Effizienzgrenze (Komparator) verursacht. Da in diesen Fällen keine typische Effizienzgrenze gezeichnet werden kann, lässt sich für die Prüflintervention nicht unmittelbar eine Empfehlung ableiten.

Die Verfügbarkeit einer Intervention mit einem sehr günstigen Kosten-Nutzen-Verhältnis begründet keine Abweichung vom Vorgehen, die Kosteneffektivität nachfolgender Interventionen anhand der Effizienzgrenze zu bewerten.

Die Ausgaben-Einfluss-Analyse soll im Fall der beschriebenen Sonderkonstellationen die Effizienzgrenze nicht ersetzen, sondern zusätzliche Informationen liefern. Es ergibt sich daher kein Änderungsbedarf.

2.32 Würdigung der Stellungnahmen zu Abschnitt 4.9.4 „Interquartilsregion als Maß der Streuung für Preisverhandlungen“

1 Stellungnehmender kritisiert, dass die Wahl der Interquartilsregion (IQR) als Maß der Streuung ohne nutzentheoretische Fundierung sei.

Die Abbildung der IQR ist ein gängiges Maß für die Streuung in gesundheitsökonomischen Analysen. Wie beschrieben, können aber auch andere Maße der Streuung zur Anwendung kommen. Es ergibt sich daher kein Änderungsbedarf in den Allgemeinen Methoden.

2.33 Würdigung der Stellungnahmen zu Kapitel 5 „Leitliniensynopsen und Versorgungsanalyse“

1 Stellungnehmende schlägt vor, den neu eingeführten Begriff „normativ“ wieder aus dem Abschnitt zu entfernen, da dies zu fremdartigen Termini wie „normative Standards“ (Tautologie) und „normative Empfehlungen“ (Oxymoron) führe.

Auch wenn Leitlinien lediglich Handlungsempfehlungen zu sein beanspruchen, ist mittlerweile davon auszugehen, dass sie normative Kraft entwickeln, d. h. sie prägen stark die Erwartungsbildung von Ärztinnen / Ärzten, Patientinnen / Patienten, Kostenträgern und Juristinnen / Juristen. Um dies deutlicher herauszustellen, wurden geringfügige Änderungen im Methodenpapier vorgenommen („normativ im Sinne der Erwartungsbildung“).

Mehrere Stellungnehmende merken an, dass entsprechend dem Methodenpapier Version 5.0 die inhaltliche Korrektheit von Leitlinien hinter der formalen Korrektheit zurückstehe und die

Probleme der derzeitigen Leitlinienentwicklung (Aktualität, Zeitaufwand, Umfang, Umsetzung im Versorgungsalltag und Einflussnahme von Interessengruppen) nicht adressiert würden.

Eine inhaltliche Prüfung von Leitlinien durch das IQWiG würde einerseits die Arbeit der Leitliniengruppen hinterfragen und wäre andererseits mit einem erheblichen Zeitaufwand verbunden, da damit die Überprüfung der Evidenzgrundlage jeder einzelnen Leitlinienempfehlung verbunden wäre. Konkret würde dies eine detaillierte Überprüfung – von den Literaturrecherchen über die Metaanalysen bis hin zur Empfehlungsformulierung – implizieren. Dies ist im Rahmen einer Leitliniensynopse weder zielführend noch praktisch umsetzbar.

Die angesprochenen Aspekte der Leitlinienentwicklung werden in den Leitliniensynopsen adressiert. So wird beispielsweise auf Empfehlungen, die aufgrund nicht hoher Grades of Recommendation (GoRs), aber nach Einschätzung von Klinikern eine hohe praktische Relevanz im Versorgungskontext Deutschland haben, hingewiesen.

1 Stellungnehmende merkt an, dass eine Begründung fehle, warum sich Leitliniensynopsen besser als Grundlage für Disease-Management-Programm(DMP)-Empfehlungen eignen als eine einzelne, multiprofessionell erstellte Nationale VersorgungsLeitlinie (NVL). Die Stellungnehmende fordert in diesem Zusammenhang eine unabhängige Kosten-Nutzen-Analyse zum Verfahren der Leitliniensynopsen des IQWiG.

Leitliniensynopsen fassen die Empfehlungen von systematisch recherchierten, methodisch hochwertigen Leitlinien zu einem (oder mehreren) speziellen Gesundheitsproblem(en) zusammen und helfen so, einen Versorgungsstandard zu beschreiben. Nicht immer entsprechen die deutschen Leitlinien dem aktuellen Stand, da ihr Veröffentlichungsdatum z. B. länger zurückliegt als das international anerkannter Leitlinien. Auch zeigt die Prüfung der Literatur, dass Leitliniengruppen sehr unterschiedliche Studien als Grundlage ihrer Empfehlungen heranziehen. Ergänzend ist anzuführen, dass die NVL bisher nicht zu allen DMP-relevanten Krankheitsbildern Leitlinien herausgegeben hat. Dies macht es erforderlich, sich nicht nur auf deutsche Leitlinien zu begrenzen, sondern im Sinne der Übertragbarkeit internationale Leitlinien ohne Hierarchisierung zu berücksichtigen. Die Verfügbarkeit von Therapiemöglichkeiten für Deutschland wird vom IQWiG zudem regelhaft geprüft und Empfehlungen werden entsprechend mit Hinweisen ergänzt.

Die vom IQWiG erstellten Leitliniensynopsen dienen als Grundlage der Abstimmungen über Richtlinienausführungen im G-BA zu den bestehenden und möglichen weiteren DMPs (siehe die entsprechende gesetzliche Grundlage).

1 Stellungnehmende merkt an, dass mit der Dichotomisierung durch das IQWiG zur Vereinheitlichung der unterschiedlichen Systeme zur Graduierung von Empfehlungen und Evidenz ein Informationsverlust und die Gefahr von Fehlinterpretationen einhergingen.

Da es international bisher keinen Konsens über eine Vereinheitlichung von Evidenz- und Empfehlungsgraduierungssystemen gibt, für die Leitliniensynopsen jedoch eine Vereinheitlichung der Systeme notwendig ist, wurde eine Dichotomisierung für die GoRs und Levels of Evidence (LoEs) vorgenommen. Der hiermit einhergehende Informationsverlust muss in Kauf genommen werden. Um die Gefahr von Fehlinterpretationen zu minimieren, wurde in der Überarbeitung des Methodenpapiers Version 5.0 die Dichotomisierung über hoch und niedrig verändert in hoch und nicht hoch, um mittlere Empfehlungen nicht mit der Kategorie niedrig abzustufen.

2.34 Würdigung der Stellungnahmen zu Abschnitt 5.2.4 „Bewertung der methodischen Leitlinienqualität“

1 Stellungnehmende kann keinen Mehrwert in der Appraisal-of-Guidelines-for-Research-&-Evaluation(AGREE)-Bewertung erkennen.

Die Ergebnisse der AGREE-Bewertung sind einerseits Teil der Aufträge durch den G-BA, andererseits helfen sie den Arbeitsgruppen im G-BA zu entscheiden, ob eine Einschätzung, die einzig auf einer methodisch schwachen Leitlinie beruht, im neuen oder zu aktualisierenden DMP berücksichtigt werden sollte. Ergänzend ist anzumerken, dass eine methodisch weniger gute Leitlinie nicht automatisch inhaltlich schwach ist.

2.35 Würdigung der Stellungnahmen zu Abschnitt 5.2.2 „Übertragbarkeit auf das deutsche Gesundheitssystem“

1 Stellungnehmende hinterfragt die Entscheidung des IQWiG im Sinne der Übertragbarkeit, nur Leitlinien aus Organisation-for-Economic-Cooperation-and-Development(OECD)-Staaten den Leitliniensynopsen zugrunde zu legen, und schlägt eine Hierarchisierung des Kriteriums „Übertragbarkeit“ vor (deutsche, europäische und internationale Leitlinien).

Unbestritten ist, dass die OECD-Länder untereinander in vielerlei Hinsicht Unterschiede zeigen, nicht allein im Hinblick auf die Gesundheitssysteme. Hieraus ergibt sich, dass mit besonderem Augenmerk die Übertragbarkeit von Leitlinienempfehlungen auf das deutsche Gesundheitssystem geprüft werden muss. Andererseits können Leitlinien aus diesen Ländern, insbesondere wenn sie sich durch eine wissenschaftliche Fundierung der Leitlinienmethodik auszeichnen, für die Entscheidungen des G-BA hinsichtlich der Ausgestaltung der DMPs in Deutschland bedeutsam sein.

Leitliniensynopsen zeichnen sich dadurch aus, dass sie den international höchsten wissenschaftlichen und medizinischen Versorgungsstandard zusammenfassen. Sie fassen hierbei Leitlinien der letzten 3 bis 5 Jahre zusammen, um so diesem Anspruch gerecht zu werden. Nicht immer entsprechen die deutschen Leitlinien dem aktuellen Stand, da ihr Veröffentlichungsdatum z. B. länger zurückliegt als das international anerkannter Leitlinien. Auch zeigt die Prüfung der Literatur, dass Leitliniengruppen sehr unterschiedliche Studien als Grundlage ihrer Empfehlungen heranziehen. Dies macht es erforderlich, sich nicht nur auf deutsche

Leitlinien zu begrenzen, sondern im Sinne der Übertragbarkeit internationale Leitlinien ohne Hierarchisierung zu berücksichtigen. Die Verfügbarkeit von Therapiemöglichkeiten für Deutschland wird vom IQWiG zudem regelhaft geprüft und Empfehlungen werden entsprechend mit Hinweisen ergänzt.

2.36 Würdigung der Stellungnahmen zu Abschnitt 5.4 „Versorgungsanalyse“

1 Stellungnehmende hinterfragt, wieso das Thema komplexe Interventionen im Zusammenhang mit der Versorgungsanalyse im Methodenpapier nicht adressiert wird, und vermisst eine Reflexion der Wissensbestände in Bezug auf die Identifikation, Analyse, Synthese und Interpretation komplexer Interventionen.

Das Aufgabenfeld Versorgungsanalyse wird aktuell neu bzw. wieder aufgebaut. Inwieweit das Thema komplexe Interventionen hier zukünftig eine Rolle spielt, ist noch offen und wird ggf. in der nächsten Version des Methodenpapiers aufgegriffen.

1 Stellungnehmende fragt im Zusammenhang mit der Versorgungsanalyse, wie klar Sollzustände definiert sind und welche Kriterien herangezogen werden.

Geplant ist es, die Sollzustände insbesondere auf der Basis der Empfehlungen von Leitlinien bzw. Leitliniensynopsen festzulegen. Die Istzustände sollen anhand von Routinedaten erhoben werden; hierzu ist geplant, unterschiedliche Quellen einzubeziehen.

2.37 Würdigung der Stellungnahmen zu Abschnitt 6.1 „Hintergrund und Ziel“ des Kapitels 6 „HTA-Berichte“

In 2 Stellungnahmen wird gefragt, ob nur nicht medikamentöse Leistungen, die im Rahmen der gesetzlichen Krankenversicherung erbracht werden, Gegenstand der HTA-Berichte sein können.

Die Themen der HTA-Berichte sind nicht auf Leistungen beschränkt, die im Rahmen der gesetzlichen Krankenversicherung erbracht werden können. Um dies zu verdeutlichen, wurde die Formulierung in Abschnitt 6.1 „Hintergrund und Ziel“ präzisiert. Medikamentöse Verfahren als primärer Bewertungsgegenstand sind jedoch als Thema für die HTA-Berichte ausgenommen.

2.38 Würdigung der Stellungnahmen zu Abschnitt 6.3.1 „Auswahlkriterien“

In 1 Stellungnahme wird die Frage gestellt, ob HTA-Fragestellungen zu seltenen Erkrankungen von einer Bearbeitung ausgeschlossen sind.

In Abschnitt 6.3.1 „Auswahlkriterien“ des Methodenpapiers werden exemplarisch verschiedene Kriterien für die Auswahl von Themen für HTA-Berichte genannt. Neben der Anzahl der betroffenen Personen sind dies unter anderem die Krankheitslast oder der

Schweregrad der Erkrankung, die zur Fragestellung vorhandene Evidenz und die Kosten der Intervention. Ziel ist es, unter Berücksichtigung aller für einen Themenvorschlag relevanten Kriterien eine qualitative Gesamtbeurteilung der Fragestellung zu erreichen. Bei der Gesamtbeurteilung kann sowohl eine hohe Zahl als auch eine geringe Zahl Betroffener für die Eignung eines Themas sprechen. Es besteht kein Ergänzungs- oder Änderungsbedarf.

2.39 Würdigung der Stellungnahmen zu Abschnitt 6.3.2 „Prüfung und Gewichtung der Themen“

In mehreren Stellungnahmen wird die in Abschnitt 6.3 „Auswahl der Themen für die HTA-Berichte“ beschriebene Gewichtung der HTA-Fragestellungen durch das IQWiG angesprochen. Es wird die Frage gestellt, ob diese Gewichtung durch das IQWiG vor der ersten Auswahlstufe, in der ein mit Patientinnen und Patienten sowie Bürgerinnen und Bürgern besetzter Auswahlbeirat Themen für HTA-Berichte nominiert, bereits eine Vorselektion von Themen darstellt.

Es erfolgt keine Vorselektion der aufbereiteten HTA-Fragestellungen durch das IQWiG vor der ersten Auswahlstufe. Um mögliche Missverständnisse zu vermeiden, wurde die Darstellung in Abbildung 17 „Schematische Darstellung des mehrstufigen Auswahlverfahrens“ des Methodenpapiers angepasst. Ebenso wurde Abschnitt 6.3.2 „Prüfung und Gewichtung der Themen“ neu benannt und die textliche Darstellung in diesem Absatz konkretisiert.

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

In 1 Stellungnahme wird darauf hingewiesen, dass im Methodenpapier das Gremium, mit dem sich die Institutsleitung bei der Auswahl der Themen ins Benehmen setzt, in Abbildung 17 „Schematische Darstellung des mehrstufigen Auswahlverfahrens“ sowie in Abschnitt 6.3.4 „Zweite Stufe des Auswahlverfahrens: Auswahl von Themen, zu denen HTA-Berichte erstellt werden“ unterschiedlich benannt wird.

Die Benennung des Gremiums, mit dem sich die Institutsleitung bei der Auswahl der Themen ins Benehmen setzt, wurde in Abbildung 17 des Methodenpapiers sowie in Abschnitt 6.3.4 „Auswahl der Themen für die HTA-Berichte“ des Methodenpapiers vereinheitlicht.

2.41 Würdigung der Stellungnahmen zu Abschnitt 6.5 „Bearbeitung der Themen (HTA-Berichte)“

In 1 Stellungnahme wird gefragt, ob für ausgewählte Aspekte der HTA-Berichte primäre Forschungsmethoden im Sinne eigener Datenerhebungen zum Einsatz kommen sollen.

In der Literatur werden sowohl eigene Datenerhebungen als auch die Nutzung vorhandener Informationsquellen als Möglichkeit der Informationsbeschaffung bei der Erstellung eines

HTA-Berichts genannt [19]. Auch für die HTA-Berichte des IQWiG ist die Durchführung eigener Erhebungen nicht ausgeschlossen. Ob für die Erstellung eines HTA-Berichts eigene Datenerhebungen notwendig sind, ist aber immer abhängig von der Fragestellung des HTA-Berichts. Es besteht kein Ergänzungs- oder Änderungsbedarf.

2.42 Würdigung der Stellungnahmen zu Abschnitt 6.5.2 „Gesundheitsökonomie“

In 1 Stellungnahme wird festgestellt, dass unklar sei, wie in den HTA-Berichten die Kosteneffektivität beurteilt werden solle, da in Deutschland kein expliziter Schwellenwert existiere. Es wird angeregt, eine Erläuterung zu ergänzen, die den Autorinnen und Autoren der HTA-Berichte dabei helfen soll, sich diesem Problem anzunähern. Als Beispiele werden der Vergleich mit inkrementellen Kosteneffektivitätsverhältnissen für andere bereits erstattete Leistungen und die Kriterien der World Health Organization (WHO) genannt.

Das Ziel einer Kosten-Nutzen-Bewertung in den HTA-Berichten ist eine vergleichende Beurteilung der Kosten-Nutzen-Verhältnisse der betrachteten Interventionen und nicht notwendigerweise eine allgemeine Aussage zur Kosteneffektivität der Prüflintervention. Die Ergebnisse einer eigenen Kosten-Nutzen-Bewertung sind grafisch in einem Koordinatensystem darzustellen, wodurch ein direkter Vergleich der Interventionen ermöglicht wird. Die Empfehlungen der WHO sehen vor, eine Intervention als kosteneffektiv zu bewerten, sofern sie das Dreifache des Pro-Kopf-Bruttoinlandsprodukts pro Einheit des betrachteten Nutzenparameters nicht überschreitet. Dadurch würde jedoch ein indikationsübergreifender Schwellenwert etabliert, der – wie auch von den Stellungnehmenden eingeräumt – zurzeit in Deutschland nicht existiert.

Die Beurteilung der Kosteneffektivität in den HTA-Berichten war auch Thema der wissenschaftlichen Erörterung am 22.03.2017. Aus der Erörterung ergaben sich keine weiterführenden konkreten Vorschläge. Es ergibt sich daher kein Änderungsbedarf.

2.43 Würdigung der Stellungnahmen zu Abschnitt 6.5.6 „Organisatorische Aspekte“

In 1 Stellungnahme wird angemerkt, dass in Abschnitt 6.5.6 „Organisatorische Aspekte“ auch zu fragen sei, welche Auswirkungen die organisatorischen Rahmenbedingungen einer Untersuchungs- bzw. Behandlungsmethode auf Patientinnen und Patienten haben.

Im Abschnitt „Organisatorische Aspekte“ der HTA-Berichte liegt der Schwerpunkt zum einen auf der Darstellung der Wechselwirkungen, die durch eine Untersuchungs- bzw. Behandlungsmethode für die Organisation der Versorgung entstehen. Zum anderen werden Faktoren, die die Implementierung einer Untersuchungs- bzw. Behandlungsmethode beeinflussen, untersucht. Sofern sich hier Auswirkungen organisatorischer Rahmenbedingungen auf Betroffene identifizieren lassen, ist eine Darstellung in diesem Abschnitt der HTA-Berichte nicht ausgeschlossen. Dabei ist darauf zu achten, dass inhaltliche Redundanzen in den ethischen, sozialen und rechtlichen Aspekten einer Untersuchungs- bzw. Behand-

lungsmethode möglichst gering gehalten werden. Es besteht kein Ergänzungs- oder Änderungsbedarf.

2.44 Würdigung der Stellungnahmen zu Kapitel 7 „Evidenzbasierte Gesundheitsinformationen für Bürgerinnen und Bürger“ und zu Abschnitt 7.1 „Hintergrund und Ziele“

1 Stellungnehmender hält es im Sinne der Transparenz und Vollständigkeit für erstrebenswert, die Begleitinformationen zu den im Rahmen der frühen Nutzenbewertung von Arzneimitteln erstellten Dossierbewertungen nach der finalen Entscheidung des G-BA über den Zusatznutzen um das Ergebnis dieser Entscheidung zu ergänzen.

1 weiterer Stellungnehmender sieht einen Widerspruch zwischen den aufgeführten Zielen bei der Erstellung von Gesundheitsinformationen und der verkürzten, aus dem Zusammenhang gelösten Ergebnisdarstellung in den Begleitinformationen zu den im Rahmen der frühen Nutzenbewertung von Arzneimitteln erstellten Dossierbewertungen. Er schlägt einen Verzicht auf die Erstellung dieser Begleitinformationen vor, da die enthaltenen Informationen nicht zu einer informierten Entscheidungsfindung beitragen könnten.

Aus Sicht unserer betroffenen Nutzer besteht gerade zu einem so relativ frühen Zeitpunkt nach der Markteinführung ein großes Interesse an möglichen Vor- und Nachteilen der im Rahmen der frühen Nutzenbewertung von Arzneimitteln vom IQWiG bewerteten Wirkstoffe. Diese Informationen werden in den begleitenden Gesundheitsinformationen allgemein verständlich aufbereitet. Die Frage des Ausmaßes des Zusatznutzens spielt für unsere Nutzer vermutlich eher eine nachgeordnete Rolle; sie dient eher als Grundlage für die nachfolgenden Preisverhandlungen. Daher verzichtet das Institut bewusst auf diese Angaben und verlinkt diesbezüglich spezifisch auf die jeweilige Internetseite des G-BA. Unabhängig davon stellt das Institut, wenn während der Stellungnahme beim G-BA neue valide Daten zum Nutzen und Schaden eingeführt werden, diese dar respektive aktualisiert seine Informationen auf Basis beauftragter Addenda. Es besteht kein Ergänzungs- oder Änderungsbedarf.

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

Ein Stellungnehmer wünscht sich eine Auflistung mit dem Abgleich der Erkrankungen mit mindestens 1 % Prävalenz in der Bevölkerung und den Themen der bisher erstellten Gesundheitsinformationen.

In Abschnitt 7.2.1 „Themenkatalog gemäß dem Generalauftrag“ wird die Themenauswahl transparent beschrieben. Die Veröffentlichung einer umfangreichen Liste mit allen bereits bearbeiteten und zukünftigen Themen ist nicht Gegenstand eines Methodenpapiers. Zudem kann sich jede und jeder Interessierte auf unserer Website www.gesundheitsinformation.de einen Überblick über die bereits veröffentlichten Themen verschaffen. Es besteht kein Ergänzungs- oder Änderungsbedarf.

1 Stellungnehmender hält die Themenauswahl des IQWiG für intransparent und postuliert, dass den Themen auf gesundheitsinformation.de offenkundig andere Kriterien zugrunde lägen als die Prävalenz.

Die Aussage des Stellungnehmenden zu den „offenkundig anderen Kriterien“ bleibt substanzlos und wird weder näher erläutert noch belegt. Es sei an dieser Stelle zudem noch einmal darauf hingewiesen, dass laut dem Methodenpapier die Prävalenz nicht das einzige Auswahlkriterium ist. Laut Abschnitt 7.2.1 kann der auf Basis des Versorgungsreports des Wissenschaftlichen Instituts der Allgemeinen Ortskrankenkasse (WIdO) erstellte Themenkatalog erweitert werden, unter anderem in Abhängigkeit von Auftragsthemen des IQWiG oder um Erkrankungen, deren Bedeutung und Krankheitslast über die Prävalenz / Inzidenz nur unzureichend erfasst werden. Weitere Themen können, wie in Abschnitt 7.2 beschrieben, auch aus Direktaufträgen des G-BA oder des Bundesministeriums für Gesundheit (BMG) resultieren. Es besteht kein Ergänzungs- oder Änderungsbedarf.

Derselbe Stellungnehmender merkt an, dass der der Themenauswahl zugrunde liegende Versorgungsreport des WIdO berücksichtige nur einen Ausschnitt der GKV-Population, und hinterfragt dessen Eignung. Er schlägt die Verwendung von Daten vor, die die Gesamtheit der GKV-Population umfassen, und nennt beispielhaft die nach § 267 SGB V für den Risikostrukturausgleich erhobenen Daten.

Es ist richtig, dass die Daten des WIdO nicht die gesamte GKV-Population umfassen. Mit dem Versorgungsreport liegen aber in bereits aufbereiteter und einfach zugänglicher Form die Daten von über 25 Millionen Versicherten vor – das ist immerhin etwa ein Drittel der deutschen Wohnbevölkerung. Zudem sind die Daten des Versorgungsreports bezüglich der Faktoren Alter und Geschlecht im Hinblick auf die deutsche Wohnbevölkerung adjustiert. Eine entsprechende Aufbereitung der Daten des Risikostrukturausgleichs würde einen ungleich höheren finanziellen und personellen Ressourcenaufwand erfordern und ist daher mit gleichzeitigem Blick auf die möglichen Unterschiede im kassenspezifischen Morbiditätsrisiko und den Zweck der Datenerhebung aus der Sicht des Instituts unangemessen. Nach dieser Abwägung hält das Institut die Daten des Versorgungsreports für ausreichend repräsentativ, um auf dieser Basis seine Themenauswahl zu treffen. Es besteht kein Ergänzungs- oder Änderungsbedarf.

2.46 Würdigung der Stellungnahmen zu Abschnitt 7.2.2 „Identifizierung der Informationsbedürfnisse / Aufbereitung von Krankheitserfahrungen“

1 Stellungnehmender kritisiert die Formulierung, dass themenspezifisch Selbsthilfeorganisationen hinsichtlich des Informationsbedarfs von Betroffenen einbezogen werden können. Die Einbindung von Selbsthilfeorganisationen erscheine regelhaft sinnvoll, um die Patientensicht besser zu erfassen.

Die Formulierung wurde gewählt, da es nicht sinnvoll schien, eine Formulierung zu wählen, die besagt, dass bei jedem Thema regelhaft die Einbeziehung von Selbsthilfeorganisationen erfolgt. Das hat schlicht den Grund, dass es nicht zu jedem Thema – beispielsweise bei vielen akuten Erkrankungen – eine Selbsthilfeorganisation gibt. Das Institut hat den Abschnitt umformuliert, um dieses Missverständnis zu klären und weiteren Fehlinterpretationen vorzubeugen.

2.47 Würdigung der Stellungnahmen zu Abschnitt 7.3 „Informationsbeschaffung zur Erstellung von Gesundheitsinformationen“

In 1 Stellungnahme wird kritisiert, dass in Abschnitt 7.3 weder die Befragung Betroffener noch die Einbindung von Selbsthilfeorganisationen angesprochen werde, obwohl dies seit Langem gängige Praxis sei.

Diese Kritik beruht vermutlich auf einem Missverständnis. In Abschnitt 7.3 „Informationsbeschaffung zur Erstellung von Gesundheitsinformationen“ wird im engeren Sinne die in Abschnitt 8.2 und insbesondere in den Abschnitten 8.2.1 und 8.2.2 des Kapitels 8 „Informationsbeschaffung“ ausgeführte fokussierte Informationsbeschaffung bei der Erstellung von Gesundheitsinformationen beschrieben. Das heißt, dass es hier um die Recherche in bibliografischen Datenbanken zur Identifikation der relevanten systematischen Übersichten und qualitativen Publikationen geht. Die bei der Erstellung von Gesundheitsinformationen erfolgende regelhafte Einbeziehung von Betroffenen wird dagegen in den Abschnitten 7.2.2 und 7.9 beschrieben. Es besteht kein Ergänzungs- oder Änderungsbedarf.

2.48 Würdigung der Stellungnahmen zu den Abschnitten 7.4 „Auswahl der Evidenz“, 7.5 „Auswahl der dargestellten Ergebnisse“, 7.6 „Wahl und Darstellung von Vergleichen“, 7.7 „Umgang mit Zahlen und Risikoangaben“ und 7.8 „Berücksichtigung von Alters- und Geschlechtsunterschieden“

1 Stellungnehmender merkt an, dass in den genannten Abschnitten eine statistisch-methodisch getriebene Bewertung unter nicht ausreichender Berücksichtigung alltäglicher und klinischer Aspekte stattfindet und allgemein im gesamten Feld „Gesundheitsinformationen“ auch eine Einbeziehung klinischer Expertise erforderlich sei.

Der Stellungnehmende hat vermutlich die Inhalte der dem Abschnitt 7.4 vorangehenden Abschnitte 7.2 und 7.3 übersehen oder fehlinterpretiert. Diese Abschnitte beschreiben ausführlich die verschiedensten Schritte, die bei der Erstellung der Gesundheitsinformationen erfolgen, um gerade die Perspektive der Betroffenen und deren alltäglichen Probleme, Erfahrungen und Fragen zu erfassen. Abschnitt 7.4 beschreibt den Umgang mit der Evidenz, die zu diesen Aspekten identifiziert wurde. Dass die Bewertung und Auswahl der Evidenz transparenten und eindeutigen methodischen Vorgaben folgen, ist ein Grundanspruch von evidenzbasierten Gesundheitsinformationen. Zudem wird unter anderem in Abschnitt 7.10 beschrieben, dass zur Sicherstellung einer sachlich angemessenen Darstellung regelhaft eine

externe fachliche Begutachtung erfolgt. Zudem haben im Rahmen der nicht öffentlichen Stellungnahme auch die Fachgesellschaften die Möglichkeit, ihre klinische Expertise einzubringen. Es besteht kein Ergänzungs- oder Änderungsbedarf.

Derselbe Stellungnehmende führt weiter aus, dass in Abschnitt 7.5 nur kurz kommentiert werde, dass Begleitumstände der Behandlung für die Patientinnen und Patienten eine Rolle spielen könnten, zu dieser Thematik jedoch keine detaillierten Angaben zum methodischen Vorgehen gemacht würden. Stattdessen werde auf die Methodik in Kapitel 3 „Nutzenbewertung medizinischer Interventionen“ abgestellt, ohne dass eine Begründung erfolge, warum für unterschiedliche Fragestellungen die gleiche Methodik herangezogen werde. Zudem sollten auch weiterführende Endpunkte wie z. B. die Patientenzufriedenheit oder klinisch relevante Surrogatendpunkte (LDLC) unbedingt berichtet werden.

Diese Darstellung scheint ein wenig missverständlich. Bei genauer Betrachtung des Abschnitts 7.5 sollte klar werden, dass sich der Verweis auf die Abschnitte 3.1 und 3.2 der Methoden auf die Darstellung von Informationen zu Ergebnissen (Nutzen und Schaden) zu patientenrelevanten Endpunkten – insbesondere zur Mortalität, Morbidität und gesundheitsbezogenen Lebensqualität – beziehen. In diesen Abschnitten wird zudem der Umgang mit Surrogaten patientenrelevanter Endpunkte beschrieben. Erst dann folgt die Aussage, dass auch Begleitumstände der Behandlung (Zeitaufwand, körperliche, seelische, finanzielle und soziale Belastung) eine Rolle spielen können. Es wird nicht, wie der Stellungnehmende impliziert, an diese rein beschreibende Darstellung solcher Behandlungsumstände die gleiche Methodik angelegt wie an die Aussagen zum Nutzen und Schaden einer medizinischen Intervention. Es besteht kein Ergänzungs- oder Änderungsbedarf.

Weiterhin bemängelt der Stellungnehmende, dass das IQWiG an die Gesundheitsinformationen nicht dieselben Maßstäbe anlege wie an die Nutzenbewertung. Beispielsweise sei die Bewertung auf systematische Reviews hoher Qualität beschränkt (unabhängig von der Qualität der Einzelstudien) und die Evidenz, die älter sei als 3 Jahre, werde unabhängig vom Anwendungsgebiet ausgeschlossen.

Da es sich bei den evidenzbasierten Gesundheitsinformationen eben nicht um Nutzenbewertungen handelt, werden diese auch nach einer eigenen, der Aufgabe und dem Zweck angemessenen Methodik erstellt. Wie in den Allgemeinen Methoden des IQWiG ausgeführt, sind im Hinblick auf die kurze Halbwertszeit des medizinischen Wissens systematische Übersichten mit einer älteren Suche, abhängig vom Thema und von der Forschungsaktivität auf diesem Gebiet, möglicherweise häufiger veraltet. Daher werden primär erst einmal systematische Reviews herangezogen, deren Suche nicht älter als 3 Jahre ist, was ja im Übrigen nicht heißt, dass die eingeschlossenen Primärstudien nicht älter als 3 Jahre sein können. Es wird in den Allgemeinen Methoden aber auch ausgeführt, dass dieser Zeitraum begründet auch verändert werden kann. Es besteht kein Ergänzungs- oder Änderungsbedarf.

Der Stellungnehmende führt weiter aus, dass es sinnvoll wäre, die Aussagen aktueller deutscher Leitlinien explizit als relevante Informationsquelle einzubeziehen.

Die nationalen Leitlinien werden nicht als regelhafte Informationsquellen aufgeführt, weil es zu vielen Themen noch keine evidenzbasierten Leitlinien gibt, sondern beispielsweise nur Leitlinien der Klassifikation S1 oder S2k veröffentlicht wurden. Es besteht kein Ergänzungs- oder Änderungsbedarf.

2.49 Würdigung der Stellungnahmen zu den Abschnitten 7.9 „Anpassung an die Zielgruppe“ und 7.9.1 „Beteiligung von Betroffenen“

1 Stellungnehmender bemängelt, dass bei diesen beiden Schritten bei der Erstellung von Gesundheitsinformationen eine klinische Expertise nicht vorgesehen zu sein scheine, obwohl das Ziel explizit nicht nur die Verständlichkeit, sondern auch die wissenschaftliche Präzision sei. Darüber hinaus sei davon auszugehen, dass Medizinerinnen und Mediziner, die im Umgang mit betroffenen Patientinnen und Patienten erfahren seien, auch über Expertise in Bezug auf die Einschätzung der Verständlichkeit von Gesundheitsinformationen verfügten.

An dieser Stelle geht es um die Einbeziehung der Betroffenen und nicht um die sie behandelnden Personen. Letztere werden im nachfolgenden Abschnitt 7.10 beschrieben, in der Form, dass zur Sicherstellung einer sachlich angemessenen Darstellung regelhaft eine fachliche Begutachtung durch eine externe Sachverständige / einen externen Sachverständigen erfolgt. Im Rahmen dieser Begutachtung wird die Gutachterin / der Gutachter dezidiert darum gebeten, ihre / seine Sicht einzubringen, den Inhalt der Gesundheitsinformation mit der von ihr / ihm erlebten Versorgungsrealität abzugleichen und gegebenenfalls kritisch zu kommentieren. Das Institut hält es für selbstverständlich, dass eine Gutachterin / ein Gutachter es auch anmerkt, wenn sie / er den begutachteten Text für nicht ausreichend verständlich hält. Abgesehen von der regelhaften Begutachtung durch eine externe klinische Expertin oder einen externen klinischen Experten haben im Rahmen der nicht öffentlichen Stellungnahme auch die Fachgesellschaften die Möglichkeit, ihre entsprechende Expertise einzubringen. Es besteht kein Ergänzungs- oder Änderungsbedarf.

2.50 Würdigung der Stellungnahmen zu Abschnitt 7.13 „Transparenz über Verfasser und Herausgeber“

1 Stellungnehmender weist darauf hin, dass im Sinne der Transparenz nicht nur das IQWiG, sondern auch andere Institutionen und Verbände, die an den erstellten und veröffentlichten Informationen mitgewirkt haben, genannt werden sollten.

Die an der Erstellung der Gesundheitsinformationen beteiligten externen Sachverständigen werden im Rahmen einer Danksagung auf der Website genannt. Diese Nennung setzt aber das Einverständnis der jeweiligen Person voraus. Zurzeit überprüft das Institut auch die Möglichkeit einer freiwilligen Nennung der beteiligten Selbsthilfeorganisationen. Eine namentliche Nennung der Betroffenen ist aus Gründen des Datenschutzes nicht vorgesehen.

2.51 Würdigung der Stellungnahmen zu Kapitel 8 „Informationsbeschaffung“

1 Stellungnahme weist darauf hin, dass Studienergebnisse auch auf Kongressen und / oder in verschiedenen Medien veröffentlicht würden. Solche Daten sollten vom IQWiG in die Informationsbeschaffung integriert und ggf. auch bewertet werden.

Die Allgemeinen Methoden Version 5.0 enthalten in Abschnitt 8.1.4 (Teil B: „Ausgewählte Fachzeitschriften und Kongressbände“) eine Begründung, weshalb in der Regel auf eine Suche in Kongressbeiträgen verzichtet wird. Das Ziel einer Suche in Kongressabstracts ist in der Regel die Identifikation von unpublizierten Studien. Jedoch haben Studienregister bei der Suche nach unpublizierten Studien in den letzten Jahren stark an Bedeutung gewonnen. Dies liegt daran, dass die gesetzliche Verpflichtung zur Studienregistrierung (sowie zur Veröffentlichung der Ergebnisse) sich international weiter verbreitet. Deshalb werden für eine umfassende Informationsbeschaffung standardmäßig Studienregister durchsucht, Kongressabstracts jedoch in der Regel nicht berücksichtigt. Inhaltlich ergibt sich deshalb kein Änderungsbedarf.

1 Stellungnahme schlägt leichte sprachliche Änderungen („Suche“ statt „Recherche“, „Informationsquellen“ statt „Suchquellen“) vor.

Die Empfehlungen wurden geprüft und ggf. Begrifflichkeiten an den entsprechenden Stellen im Methodenpapier vereinheitlicht bzw. überarbeitet.

In 1 Stellungnahme wird kritisiert, dass unaufgefordert übermittelte Studiendaten inhaltlich nicht mehr berücksichtigt würden. Aus Sicht des Stellungnehmenden widerspreche das den Grundzügen der evidenzbasierten Medizin und es bestehe die Gefahr, dass relevante Informationen sowie neueste Evidenz nicht mehr berücksichtigt würden, was dann umgekehrt ebenfalls eine Verzerrung des Berichtsergebnisses nach sich ziehe.

Die betreffende Passage stellt, anders als vom Stellungnehmenden angenommen, keine neue Regelung dar, sondern war bereits Teil der Allgemeinen Methoden Version 4.2. Die Argumente des Stellungnehmenden entsprechen im Wesentlichen denen, die bereits in vorangegangenen Stellungnahmen zu diesem Passus vorgebracht wurden. Für die inhaltliche Auseinandersetzung mit den Argumenten wird deshalb auf die Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden 4.2 verwiesen [20]. Insgesamt ergibt sich kein Änderungsbedarf für die Allgemeinen Methoden.

2.52 Würdigung der Stellungnahmen zu Abschnitt 8.1 „Umfassende Informationsbeschaffung“

1 Stellungnahme sieht die Definition der Suchtechniken Snowballing und Similar-Articles-Recherchen als unzureichend an.

Für eine detailliertere Definition und Erläuterung wird auf die EuNetHTA Guideline [21] verwiesen. Für das Methodenpapier ist keine weitergehende Erläuterung notwendig.

1 Stellungnahme kritisiert an mehreren Stellen die Darstellung des Vorgehens bei der Erstellung von Suchstrategien in bibliografischen Datenbanken (Konkretisierung des Planungsschritts, Bedarf an Fachexpertise, objektiver Ansatz nach Hausner et al. etc.).

Die wesentlichen Punkte der Strategieerstellung werden erläutert. In den entsprechenden Zitaten finden sich zusätzliche Informationen und es wird auf mögliche Limitierungen eingegangen. Es besteht kein Bedarf, im Methodenpapier weitere Details zu nennen.

2.53 Würdigung der Stellungnahmen zu Abschnitt 8.1.1 „Suche in bibliografischen Datenbanken“

1 Stellungnahme regt an, im Methodenpapier klarzustellen, dass eine zusätzliche Suche in PubMed nach nicht indextierten Referenzen obsolet sei und auch andere Oberflächen wie Ovid, die die entsprechende Aktualität aufwiesen, alternativ zur Suche in Medline herangezogen werden könnten.

PubMed enthält auch weiterhin Referenzen, die nicht über andere Suchoberflächen (z. B. Ovid) zur Verfügung stehen, wie etwa Publikationen des Segmentes „Books“ (z. B. AHRQ-Berichte). Auch können immer wieder zeitliche Verzögerungen bei der Übermittlung von Daten aus PubMed festgestellt werden. Da sich diese Unterschiede zwischen den Oberflächen aber immer weiter verringern, ist dem Stellungnehmenden insofern recht zu geben, als auch eine Suche in Medline über Ovid oder über eine andere Oberfläche zu einem vollständigen Ergebnis führen kann. Im Methodenpapier wird dies daher als „Kann-Möglichkeit“ konkretisiert.

2.54 Würdigung der Stellungnahmen zu Abschnitt 8.1.2 „Suche in Studienregistern“

1 Stellungnahme merkt an, dass eine Suche in Studienregistern mit wirkstoff- und indikationsbezogenen Suchbegriffen analog zu der bibliografischen Literaturrecherche durchgeführt werden könne.

Die im Methodenpapier zitierte Literatur weist darauf hin, dass eine Suche in Studienregistern eben nicht analog zur bibliografischen Recherche durchgeführt werden kann. Der Stellungnahme ist keine Literatur zu entnehmen, die ihre Ansichten stützt. Es wird daher kein Änderungsbedarf im Methodenpapier gesehen.

2.55 Würdigung der Stellungnahmen zu Abschnitt 8.2 „Fokussierte Informationsbeschaffung“

1 Stellungnahme kritisiert die gewählten Begrifflichkeiten „umfassende Informationsbeschaffung“ und „fokussierte Informationsbeschaffung“.

Es ist richtig, dass der Begriff „fokussierte Informationsbeschaffung“ im Deutschen nicht geläufig ist. Im englischsprachigen Raum wird dieser allerdings gelegentlich verwendet, z. B. bei Golder [22]. Da in der Stellungnahme kein alternativer deutscher Begriff genannt wird, wird kein Änderungsbedarf für das Methodenpapier gesehen.

Laut einer Stellungnahme passt die Referenz von Glanville [23] nicht, um darzustellen, was „ausgewogene Sensitivität und Präzision“ meint.

Die Referenz von Glanville [23] wurde angeführt, um einen Cut-off-Wert der Sensitivität bei möglichst präzisen Recherchen festzulegen. Glanville [23] sieht bei der Bewertung von Studienfiltern für gesundheitsökonomische Evaluationen eine Sensitivität von 80 % als ausreichend an, wenn das Ziel besteht, eine maximale Präzision bei der Recherche zu erzielen. Es ergibt sich keine Änderung.

1 Stellungnahme bezieht sich auf den Fall, dass systematische Übersichten als Quelle für Primärstudien dienen. Bei einem solchen Vorgehen wird die Informationsbeschaffung für den Zeitraum, der nicht durch die systematische Übersicht abgedeckt wird, aktualisiert. Laut der Stellungnahme muss dann die ganze Recherche wiederholt werden. Dies sei aber in der Praxis aufgrund der unvollständigen Dokumentation kaum möglich.

Die Aktualisierung der Informationsbeschaffung erfolgt nach der Methodik des IQWiG für den Zeitraum, der nicht durch die systematische Übersicht abgedeckt wird. Um das Vorgehen klarzustellen, wird dies im Methodenpapier ergänzt.

1 Stellungnahme regt an, den Terminus „fokussierter Informationsbedarf“ im Text näher zu erläutern.

Im Text werden Situationen, in denen eine fokussierte Informationsbeschaffung angewendet wird, sowie praktische Beispiele beschrieben. Es wird daher kein Ergänzungsbedarf im Methodenpapier gesehen.

1 Stellungnahme gibt an, dass es grundsätzlich akzeptabel sei, bestimmte Publikationstypen (z. B. Kongressabstracts) durch eine NOT-Suche zu eliminieren. Die Stellungnahme regt an, dies im Methodenpapier zu erwähnen.

Wir stimmen mit den Stellungnehmenden darin überein, dass eine NOT-Verknüpfung für den Ausschluss bestimmter Publikationstypen möglich ist. Eine diesbezügliche Ergänzung im Methodenpapier ist nicht vorgesehen, da eine solche Spezifizierung zu detailliert ist und zudem auch kurzfristige Änderungen unterliegen kann.

2.56 Würdigung der Stellungnahmen zu Abschnitt 8.3 „Orientierende Recherchen“

1 Stellungnahme gibt an, dass es validierte Studienfilter als solche nicht gebe und dass fast alle sogenannten Standardfilter für eine adäquate Sensitivität und Präzision themenspezifisch angepasst werden müssten. Es wird kritisiert, dass dies vom IQWiG ignoriert werde.

In der Literatur existieren sogenannte validierte Studienfilter, wie dies beispielsweise in dem Grundsatzartikel von Jenkins [24] dargestellt wird. Zudem existieren in der Literatur Checklisten, mit denen geprüft werden kann, ob der Einsatz eines validierten Filters adäquat ist [25]. Es wird kein Änderungs- oder Ergänzungsbedarf im Methodenpapier gesehen.

1 Stellungnahme weist darauf hin, dass die Aussage zur Fehleranfälligkeit von Suchstrategien nicht Gegenstand des referenzierten Aufsatzes von Sampson sei.

Das Zitat wird geändert und die Formulierung zur Klarstellung im Methodenpapier angepasst.

1 Stellungnahme gibt an, dass das bei der Erläuterung der weiteren Suchtechniken genannte Zitat Dorée 2016 eine „Level of Evidence 5“-Quelle sei, die nicht qualitätsgeprüft und deswegen auch nicht zu zitieren sei.

Das Methodenkapitel von Dorée et al. [26] wurde auf der Plattform SuRe Info (Summarized Research in Information Retrieval for HTA) für Expertinnen und Experten veröffentlicht. Bei dieser Plattform handelt es sich nicht um Expertenmeinungen, sondern um die systematische und transparent dargestellte Zusammenfassung der Evidenz zum Thema „Information Retrieval“. Es wird kein Änderungsbedarf im Methodenpapier gesehen.

2.57 Würdigung der Stellungnahmen zu Abschnitt 9.1.3 „Rangordnung verschiedener Studienarten / Evidenzgrade“

In 1 Stellungnahme wird gefordert, indirekte Vergleiche und Netzwerk-Metaanalysen in die allgemeine Evidenzhierarchie einzuordnen.

In Abschnitt 9.3.9 der Allgemeinen Methoden wird erläutert, dass indirekte Vergleiche und Netzwerk-Metaanalysen in der Regel eine geringere Ergebnissicherheit aufweisen als adäquate direkt vergleichende Studien. Eine exakte Einordnung in eine starre Evidenzhierarchie ist nicht sinnvoll; daher besteht hier kein Änderungsbedarf.

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

In 1 Stellungnahme wird angeregt, sich bei der Relevanzbewertung von Skalenergebnissen auf die 2. der 3-stufigen Hierarchie zu beschränken. Eine feste Grenze zur Prüfung eines relevanten Effekts (Stufe 1) könne es gar nicht geben, da sie indikationsspezifisch sei und nur im Zusammenhang mit Effekten bei anderen Endpunkten beurteilt werden könne.

Die Stellungnehmenden weisen zu Recht darauf hin, dass etablierte Irrelevanzschwellen für die 1. Stufe der Relevanzbeurteilung in der Praxis nicht existieren und somit keine Rolle spielen. In der Praxis werden in der Tat ausschließlich die Stufen 2 (Responderanalysen) und 3 (Beurteilung über die standardisierte Mittelwertdifferenz) angewendet. Die Darstellung der 1. Stufe dient vielmehr der Darstellung des Ideals einer Prüfung auf relevante Effekte und soll dem Verständnis der Relevanzprüfung dienen. Responderanalysen liegen für die Bewertung nicht immer vor und können, da die individuellen Patientendaten nicht verfügbar sind, auch nicht regelhaft berechnet werden. Daher stellt die 3. Stufe oft die einzige Möglichkeit dar, überhaupt Aussagen zur Relevanz zu generieren, was die Stellungnehmenden auch konstatieren. Die Notwendigkeit einer Änderung dieses Abschnitts wird nicht gesehen.

2.59 Würdigung der Stellungnahmen zu Abschnitt 9.3.4 „Bewertung subjektiver Endpunkte bei offenem Studiendesign“

In 1 Stellungnahme wird angeregt, auf Methoden zu verweisen, die in Metaanalysen eine Biaskorrektur für fehlende Designqualität ermöglichen (Welton et al. [27]).

Die Methode nach Welton et al. [27] stellt ein Biasmodell dar, das sehr sensitiv ist für die Art und Weise der Modellierung, sodass die Welton et al. [27] selbst vor einer routinemäßigen Anwendung warnen. Dieses Verfahren wird nicht regelhaft verwendet und es besteht daher kein Änderungsbedarf.

2.60 Würdigung der Stellungnahmen zu Abschnitt 9.3.7 „Adjustierung und multifaktorielle Verfahren“

In 1 Stellungnahme wird gefordert, Methoden für Kausalmodelle zur Kontrolle von zeitabhängigem Confounding bei Beobachtungsstudien sowie Verfahren zum Umgang mit Behandlungswechseln zu ergänzen.

Aus Beobachtungsstudien werden regelhaft keine Nutzenbelege abgeleitet, sodass die Methoden für Kausalmodelle zur Kontrolle von zeitabhängigem Confounding keine zentrale Rolle spielen. Die Aufnahme von statistischen Verfahren zum Umgang mit Behandlungswechseln wird für zukünftige Versionen des Methodenpapiers in Erwägung gezogen.

2.61 Würdigung der Stellungnahmen zu Abschnitt 9.3.8 „Metaanalysen“

In 1 Stellungnahme wird zur begrifflichen Klarstellung angeregt, Fixed-Effects-Modelle und Random-Effects-Modelle voneinander abzugrenzen.

Eine solche begriffliche Abgrenzung ist in Abschnitt 9.3.8 vorhanden. Zudem wird auf zahlreiche Literaturstellen verwiesen, die weitere Erläuterungen beinhalten; daher besteht hier kein Änderungsbedarf.

In 1 Stellungnahme wird gefragt, ob die Grenze von 0,05 auch beim Test auf Heterogenität Verwendung finden soll.

Die Aussage, dass dem so ist, steht eindeutig in Abschnitt 9.3.8 (Teil B); daher besteht hier kein Änderungsbedarf.

In 1 Stellungnahme wird angemerkt, dass eine unterschiedliche Mindestzahl von Studien genannt wird, um Metaanalysen mit zufälligen Effekten darzustellen.

Es ist hier zu unterscheiden, ob eine Metaanalyse mit zufälligen Effekten verwendet wird, um einen gepoolten Effekt zu schätzen (hier gilt die Mindestzahl von 5 Studien), oder ob in Situationen, in denen es aufgrund hoher Heterogenität nicht sinnvoll ist, einen gepoolten Effekt darzustellen, Prädiktionsintervalle verwendet werden, um die Gleichgerichtetheit der Effekte zu prüfen (hierfür gilt die Mindestzahl von 4 Studien). Die entsprechenden Passagen wurden überarbeitet, um dies deutlicher darzustellen.

Es gibt mehrere Stellungnahmen zur Verwendung des Signifikanzniveaus von 5 % beim Heterogenitätstest in Metaanalysen. Diese Änderung wird in den meisten Stellungnahmen begrüßt, aber teilweise auch kritisiert. So wird in 1 Stellungnahme dafür plädiert, die Grauzone eines p-Werts zwischen 5 % und 20 % zu erhalten. In einigen Stellungnahmen wird gefordert, die Wahl des Signifikanzniveaus ausführlich zu begründen, und es wird eine Überprüfung der Metaanalysen auf Basis der neuen Regelung durch das IQWiG gefordert, da die Auswirkungen auf die Nutzenbewertung möglicherweise erheblich seien.

Das vorherige Vorgehen, regelhaft nicht zu poolen, falls der Heterogenitätstest zum Niveau 20 % signifikant ist, wurde vorher häufig kritisiert und die vorgeschlagene Änderung ist eine Konsequenz aus dieser Kritik. Das Institut hat sich vorab die Auswirkungen der Änderung des Vorgehens zum Heterogenitätstest angesehen und festgestellt, dass die Auswirkungen auf die Nutzenbewertung als gering einzustufen sind. Eine empirisch begründbare, eindeutig „richtige“ Schwelle für den p-Wert eines Signifikanztests gibt es nicht, sodass hier arbiträre Festlegungen unvermeidbar sind. Es besteht daher hier kein Änderungsbedarf.

Die Anwendung der Knapp-Hartung-Methode für Metaanalysen mit zufälligen Effekten wird in einigen Stellungnahmen begrüßt. In mehreren Stellungnahmen wird die Anwendung der Knapp-Hartung-Methode aber auch hinterfragt. Es wird angeregt, dass das IQWiG die Vorteile gegenüber dem bisherigen Vorgehen schildert und eine Einschätzung zu den Abweichungen der Ergebnisse der Knapp-Hartung-Methode und des bisherigen Vorgehens voneinander angibt.

Die Vorteile der Knapp-Hartung-Methode sind seit Langem bekannt und die Anwendung dieser Methode wurde vom IQWiG mehrfach auf nationalen und internationalen Kongressen angekündigt und ausführlich diskutiert. Weitere Details sind der zitierten Literatur zu entnehmen; daher besteht hier kein Änderungsbedarf.

In mehreren Stellungnahmen wird kritisiert, dass das Vorgehen bei Metaanalysen im Fall sehr weniger Studien unklar sei.

Zur Durchführung von Metaanalysen im Fall sehr weniger Studien wurde vom IQWiG eine Expertenrunde ins Leben gerufen mit dem Resultat, dass es keine eindeutig beste Methode gibt. Es kommen mehrere Methoden infrage, von denen je nach Datensituation eine oder mehrere ausgewählt werden müssen. Die entsprechenden Ergebnisse wurden auf nationalen und internationalen Kongressen vorgestellt und ausführlich diskutiert. Weitere Details sind der zitierten Literatur zu entnehmen; daher besteht hier kein Änderungsbedarf.

Mehrere Stellungnahmen äußern sich – zum Teil kritisch, zum Teil positiv – zur einschränkenden Verwendung der Peto-Odds-Ratio-Methode bei einer geringen Zahl von Ereignissen. Zum Teil werden klare Kriterien und eine Validierung dieser Kriterien gefordert.

Die Limitationen der Peto-Odds-Ratio-Methode sind seit Langem bekannt. Die im Entwurf neu angegebene Literatur liefert gerade verbesserte Anwendungskriterien. Es wurde eine weitere Literaturstelle ergänzt; weiteren Änderungsbedarf gibt es nicht.

2.62 Würdigung der Stellungnahmen zu Abschnitt 9.3.9 „Indirekte Vergleiche“

In mehreren Stellungnahmen wird gefordert, klare quantitative Kriterien vorzugeben für eine adäquate Begründung zur Durchführung indirekter Vergleiche, für adäquate Brückenkomparatoren und für eine ausreichende Ähnlichkeit von Studien.

Es ist offensichtlich, dass es für qualitative wissenschaftliche Begründungen keine klaren quantitativen Kriterien gibt. Weitere Details sind der zitierten Literatur zu entnehmen; daher besteht hier kein Änderungsbedarf.

In mehreren Stellungnahmen wird darauf hingewiesen, dass im Entwurf des Methodenpapiers die Methode von Signorovitch et al. [28] fälschlicherweise als Verfahren für unverbundene Netzwerke dargestellt werde, es sich stattdessen aber um einen adjustierten indirekten Vergleich mit gemeinsamen Brückenkomparatoren handle, der aber zusätzlich versuche, Unterschiede in den Populationscharakteristika der berücksichtigten Studien zu korrigieren.

Es ist richtig, dass die Methode von Signorovitch et al. [28] nicht ausschließlich für unverbundene Netzwerke geeignet ist. Dies wird aber auch nicht behauptet. In ca. 40 % der Fälle wird die Methode von Signorovitch et al. [28] zur Evidenzsynthese von Einzelarmstudien oder in Situationen ohne Brückenkomparator angewendet (Phillippo et al. [29]). Das Zitat ist somit korrekt und es besteht hier kein Änderungsbedarf.

In mehreren Stellungnahmen wird gefordert, neben adjustierten indirekten Vergleichen auch Netzwerk-Metaanalysen ohne adäquate Brückenkomparatoren sowie Ergebnisse aus einzelnen Armen verschiedener Studien bei Nutzenbewertungen zu akzeptieren, zumindest

wenn auf Basis dieser Studien eine Zulassung erfolgte. Es wird hierbei auf Verfahren für Beobachtungsstudien wie das Propensity Score Matching verwiesen. Aufgrund der erhöhten Unsicherheit seien auf Basis solcher Analysen zwar keine Belege, aber wenigstens Hinweise auf einen Zusatznutzen möglich.

Es ist richtig, dass Netzwerk-Metaanalysen ohne adäquate Brückenkomparatoren sowie Vergleiche auf Basis einzelner Arme aus verschiedenen Studien mit erhöhter Unsicherheit behaftet sind. Die Stellungnehmenden verkennen aber das Ausmaß dieser Unsicherheit. Bereits adjustierte indirekte Vergleiche mit adäquaten Brückenkomparatoren sind mit erhöhter Unsicherheit behaftet und erlauben – wenn überhaupt – regelhaft nur Anhaltspunkte für einen Zusatznutzen. Netzwerk-Metaanalysen ohne adäquate Brückenkomparatoren sowie Vergleiche auf Basis einzelner Arme aus verschiedenen Studien besitzen einen Evidenzgrad noch unterhalb von guten Kohortenstudien, in denen die zu vergleichenden Gruppen wenigstens noch zeitlich parallel beobachtet werden. Netzwerk-Metaanalysen ohne adäquate Brückenkomparatoren sowie Vergleiche auf Basis einzelner Arme aus verschiedenen Studien kommen somit als Grundlage für Nutzenbewertungen nicht infrage; daher besteht hier kein Änderungsbedarf.

In mehreren Stellungnahmen wird darauf hingewiesen, dass es unklar sei, wann Netzwerk-Metaanalysen ohne adäquate Brückenkomparatoren sowie Vergleiche auf Basis einzelner Arme aus verschiedenen Studien akzeptiert würden, da dies in Einzelfällen aufgrund dramatischer Effekte doch möglich sei.

Dramatische Effekte stellen eine Ausnahme dar. Der Umgang mit dramatischen Effekten wird in Abschnitt 3.2.2 des Methodenpapiers beschrieben. Es wird nicht in allen anderen Abschnitten des Methodenpapiers jedes Mal wieder neu auf eine solche Ausnahme hingewiesen; daher besteht hier kein Änderungsbedarf.

2.63 Würdigung der Stellungnahmen zu Abschnitt 9.3.10 „Subgruppenanalysen“

In mehreren Stellungnahmen wird gefordert, bei der Bewertung der Verlässlichkeit von Subgruppenanalysen auch indirekte Evidenz wie die biologische Rationale, Laboruntersuchungen oder Tierstudien einzubeziehen. Die Stellungnehmenden verweisen dabei auf die in den Arbeiten von Sun et al. aus den Jahren 2010 [30] und 2012 [31] genannten Kriterien.

Die Bewertung der genannten indirekten Evidenz wird weder als machbar noch als zielführend angesehen. Würde man dem Vorschlag der Stellungnehmenden folgen, wäre es sicherlich nicht ausreichend, selektiv einzelne Arbeiten zu sichten, sondern dies würde vielmehr eine systematische Aufarbeitung dieser Evidenz notwendig machen. Darüber hinaus ist unklar, welchen Stellenwert indirekte Evidenz bei der Bewertung der Verlässlichkeit von Subgruppenanalysen haben kann. Von den Stellungnehmenden wurden weder in den Stellungnahmen noch in der Erörterung konkrete Vorschläge gemacht, wie die praktische

Umsetzung der Berücksichtigung indirekter Evidenz aussehen könnte. Es ergibt sich daher kein Änderungsbedarf für die Allgemeinen Methoden.

Mehrere Stellungnahmen nehmen Bezug auf das Kriterium, dass Subgruppenanalysen bei binären Daten regelhaft nur dann durchgeführt werden, falls mindestens 10 Ereignisse in einer der Subgruppen aufgetreten sind. Diese Eingrenzung wird grundsätzlich begrüßt. Zum Teil wird jedoch eine Begründung für die Zahl 10 gefordert. Zum Teil wird vorgeschlagen, eine Mindestanzahl von Ereignissen pro Subgruppe (z. B. 10) oder sogar pro Behandlungsarm (z. B. 5) anzusetzen.

In den Stellungnahmen und auch in der mündlichen Erörterung wurden von den Stellungnehmenden keine Arbeiten oder Ansätze genannt, aus denen ein Kriterium für die Frage, wann Subgruppenanalysen durchgeführt werden sollten, abgeleitet werden könnte. Die Forderung von 10 Ereignissen in mindestens einer der Subgruppen kann sicher als eine niedrige Schwelle dafür angesehen werden, überhaupt eine Interaktion zwischen einer Behandlung und einem Subgruppenfaktor aufzudecken. Eine Mindestanzahl von Ereignissen für jede Subgruppe anzusetzen, wird kritisch gesehen. Fälle, in denen Ereignisse (fast) ausschließlich in einer Subgruppe auftreten – und dort vornehmlich in einem der Behandlungsarme –, würden so nicht entdeckt. In der mündlichen Erörterung wurde vorgeschlagen, eine Mindestanzahl von Patientinnen / Patienten pro Subgruppe, z. B. 10 Patientinnen / Patienten, zu fordern, da eine Suche nach Interaktionen bei sehr kleinen Subgruppen auch kaum möglich sei. Dieser Vorschlag wurde aufgenommen und im entsprechenden Abschnitt wurde ergänzt, dass Subgruppenanalysen nur dann durchgeführt werden, wenn mindestens 10 Patientinnen / Patienten pro Subgruppe vorliegen.

In 1 Stellungnahme wurde gefordert, aus Symmetriegründen neben einer Mindestanzahl von 10 Ereignissen auch eine Mindestanzahl von 10 Patientinnen / Patienten ohne Ereignis zu nennen.

Der Beweggrund für die Einführung einer Mindestanzahl von Ereignissen ist die Erfahrung aus vergangenen Bewertungen, dass oft Endpunkte mit wenigen bis sehr wenigen Ereignissen vorliegen. In dieser Situation sind Subgruppenanalysen zwecklos, da sie keine Trennschärfe besitzen. Der umgekehrte Fall, dass fast alle Patientinnen und Patienten ein Ereignis erfahren, ist in der Praxis von untergeordneter Relevanz und bedarf daher keines Einbezugs in das Kriterium.

In mehreren Stellungnahmen wurde die Streichung der Problematik des multiplen Testens kritisiert.

Das Problem des multiplen Testens besteht bei zahlreichen Subgruppenanalysen, wie von den Stellungnehmenden angemerkt, nach wie vor und wurde an entsprechender Stelle wieder ergänzt.

In den Stellungnahmen wurde teilweise kritisiert, dass im einleitenden Text zur Auflistung der Probleme bei Subgruppenanalysen anstelle der Formulierung „werden in der methodischen Literatur sehr kritisch diskutiert“ nun die Formulierung „werden in der methodischen Literatur teilweise kritisch diskutiert“ gewählt und der Stellenwert der Kritikpunkte somit abgeschwächt worden sei.

Diese Änderung soll die Probleme bei Subgruppenanalysen nicht abschwächen. Es soll vielmehr deutlicher werden, dass Subgruppenanalysen trotz ihrer Problematik sinnvoll und auch notwendig sind zur Bewertung von medizinischen Interventionen

Des Weiteren wird in den Stellungnahmen zum Teil die Aufnahme von 2 Punkten in die Auflistung der Probleme bei Subgruppenanalysen kritisiert:

- 1) gefährdete Vergleichbarkeit der Behandlungsgruppen innerhalb der Subgruppen bei fehlender stratifizierter Randomisierung*
- 2) Interpretationsproblematik bei Wechselwirkungen höherer Ordnung*

Es wird angemerkt, dass eine Randomisierung, die nach vielen Faktoren stratifiziert werde, sowie Subgruppenanalysen für Wechselwirkungen höherer Ordnung in der Praxis nicht oder kaum sinnvoll durchführbar seien.

Es mag der Eindruck entstanden sein, dass für zukünftige Nutzenbewertungen umfangreichere Stratifizierungen bei der Randomisierung und bei Wechselwirkungen höherer Ordnung gefordert werden. Das ist nicht der Fall. Beide Punkte finden sich lediglich in der Auflistung der Probleme bei Subgruppenanalysen.

In 1 Stellungnahme wurde darum gebeten, die Passage, die die Situation eines Vorliegens von mehr als 2 Subgruppen behandelt, umzuformulieren. Das Wort „andernfalls“ in der folgenden Passage sei unklar: „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. Andernfalls werden keine Paare gebildet, sondern es wird eine separate Aussage für jede Subgruppe getroffen.“

Die Passage wurde geändert. Der Bezug des Wortes „andernfalls“ wurde klarer beschrieben („andernfalls“ wurde ersetzt durch „Finden sich keine Paare, die zum Niveau $\alpha = 0,05$ nicht statistisch signifikant sind ...“).

In 1 Stellungnahme wurde darauf hingewiesen, dass ein Textabschnitt zu Subgruppenanalysen mit Merkmalen, die nach der Randomisierung erhoben werden, an einer falschen Stelle im Textverlauf stehe.

Der Textabschnitt zu Subgruppenanalysen mit Merkmalen, die nach der Randomisierung erhoben werden, wurde an eine passende Stelle verschoben.

2.64 Würdigung der Stellungnahmen zu Abschnitt 9.3.13 „Auswertung abhängiger Daten“

Die Aufnahme des neuen Abschnitts zur Auswertung bei abhängigen Daten wird in mehreren Stellungnahmen begrüßt. In 1 Stellungnahme wird auf die Nichtallgemeingültigkeit der Aussage hingewiesen, dass die Anwendung der einfachen Standardverfahren für unabhängige Beobachtungseinheiten bei abhängigen Daten zu einer Unterschätzung der Varianz und möglicherweise zu falschen Signifikanzaussagen führe.

Dieser Hinweis ist korrekt. Die Aussage zur Unterschätzung der Varianz in Abschnitt 9.3.13 bezieht sich ausschließlich auf abhängige Daten innerhalb eines Behandlungsarms, wofür die Aussage wiederum richtig ist. Der entsprechende Text wurde umformuliert, um dies zu verdeutlichen. Es wurde zudem ergänzt, dass auch Abhängigkeiten zwischen den Behandlungsarmen durch geeignete statistische Verfahren zu berücksichtigen sind.

2.65 Literatur

1. Lühnen J, Albrecht M, Mühlhauser I, Steckelberg A. Leitlinie evidenzbasierte Gesundheitsinformation [online]. 20.02.2017 [Zugriff: 20.04.2017].
URL: <http://www.leitlinie-gesundheitsinformation.de/wp-content/uploads/2014/05/Leitlinie-evidenzbasierte-Gesundheitsinformation.pdf>.
2. Windeler J, Lauterberg J, Wieseler B, Sauerland S, Lange S. Patientenregister für die Nutzenbewertung: kein Ersatz für randomisierte Studien. Dtsch Arztebl 2017; 114(16): 783-786.
3. McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, Marshall JC et al. No surgical innovation without evaluation: the IDEAL recommendations. Lancet 2009; 374(9695): 1105-1112.
4. Bundesministerium für Gesundheit. Verordnung über die Nutzenbewertung von Arzneimitteln nach § 35a Absatz 1 SGB V für Erstattungsvereinbarungen nach § 130b SGB V (Arzneimittel-Nutzenbewertungsverordnung - AM-NutzenV) [online]. 27.03.2014 [Zugriff: 03.04.2017]. URL: <http://www.gesetze-im-internet.de/bundesrecht/am-nutzenv/gesamt.pdf>.
5. Gemeinsamer Bundesausschuss. Verfahrensordnung des Gemeinsamen Bundesausschusses [online]. URL: <https://www.g-ba.de/informationen/richtlinien/42/>.
6. Bundesrat. Gesetz zur Stärkung der Arzneimittelversorgung in der GKV (GKV-Arzneimittelversorgungsstärkungsgesetz - AMVSG): Gesetzesbeschluss des Deutschen Bundestages; Bundesrat Drucksache 195/17 [online]. 10.03.2017 [Zugriff: 20.04.2017]. URL: https://www.bundesrat.de/SharedDocs/drucksachen/2017/0101-0200/195-17.pdf?__blob=publicationFile&v=5.
7. Deutscher Bundestag. Entwurf eines Gesetzes zur Stärkung der Versorgung in der gesetzlichen Krankenversicherung (GKV-Versorgungsstärkungsgesetz – GKV-VSG): Gesetzentwurf der Bundesregierung; Bundestag Drucksache 18/4095 [online]. 25.02.2015 [Zugriff: 20.04.2017]. URL: <http://dip21.bundestag.de/dip21/btd/18/040/1804095.pdf>.
8. Resnic FS, Majithia A, Marinac-Dabic D, Robbins S, Ssemaganda H, Hewitt K et al. Registry-based prospective, active surveillance of medical-device safety. N Engl J Med 2017; 376(6): 526-535.
9. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016; 355: i4919.
10. Skrabanek P, McCormack J. Follies and fallacies in medicine. Whithorn: Tarragon Press; 1998.
11. Prasad V, Gall V, Cifu A. The frequency of medical reversal. Arch Intern Med 2011; 171(18): 1675-1676.

12. Moses H 3rd, Matheson DH, Cairns-Smith S, George BP, Palisch C, Dorsey ER. The anatomy of medical research: US and international comparisons. *JAMA* 2015; 313(2): 174-189.
13. Fuchs S, Olberg B, Panteli D, Perleth M, Busse R. HTA of medical devices: challenges and ideas for the future from a European perspective. *Health Policy* 2017; 121(3): 215-229.
14. Chang L, Dhruva SS, Chu J, Bero LA, Redberg RF. Selective reporting in trials of high risk cardiovascular devices: cross sectional comparison between premarket approval summaries and published reports. *BMJ* 2015; 350: h2613.
15. Adie S, Harris IA, Naylor JM, Mittal R. Are outcomes reported in surgical randomized trials patient-important? A systematic review and meta-analysis. *Can J Surg* 2017; 60(2): 86-93.
16. Nieuwenhuijse MJ, Nelissen RG, Schoones JW, Sedrakyan A. Appraisal of evidence base for introduction of new implants in hip and knee replacement: a systematic review of five widely used device technologies. *BMJ* 2014; 349: g5133.
17. Merlin T, Lehman S, Hiller JE, Ryan P. The "linked evidence approach" to assess medical tests: a critical analysis. *Int J Technol Assess Health Care* 2013; 29(3): 343-350.
18. Gemeinsamer Bundesausschuss. Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses: Bewertung des Nutzens und Potenzials neuer Untersuchungs- und Behandlungsmethoden mit Medizinprodukten hoher Risikoklasse nach § 137h des Fünften Buches Sozialgesetzbuch [online]. 21.07.2016 [Zugriff: 20.04.2017]. URL: https://www.g-ba.de/downloads/40-268-3902/2016-07-21_137h-Regelbeauftragung-IQWiG_TrG.pdf.
19. Perleth M, Busse R, Gerhardus A, Gibis B, Lühmann D, Zentner A (Ed). *Health Technology Assessment: Konzepte, Methoden, Praxis für Wissenschaft und Entscheidungsfindung*. Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft; 2014.
20. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden 4.2 [online]. 22.04.2015 [Zugriff: 20.04.2017]. URL: https://www.iqwig.de/download/DWA_Methoden_4-2.pdf.
21. European Network for Health Technology Assessment. Process of information retrieval for systematic reviews and health technology assessments on clinical effectiveness: version 1.1 [online]. 12.2016 [Zugriff: 20.04.2017]. URL: http://www.eunetha.eu/sites/default/files/Guideline_Information_Retrieval_V1-1.pdf.
22. Golder S, Glanville J, Ginnelly L. Populating decision-analytic models: the feasibility and efficiency of database searching for individual parameters. *Int J Technol Assess Health Care* 2005; 21(3): 305-311.
23. Glanville J, Kaunelis D, Mensinkai S. How well do search filters perform in identifying economic evaluations in MEDLINE and EMBASE. *Int J Technol Assess Health Care* 2009; 25(4): 522-529.

24. Jenkins M. Evaluation of methodological search filters: a review. *Health Info Libr J* 2004; 21(3): 148-163.
25. Bak G, Mierzwinski-Urban M, Fitzsimmons H, Morrison A, Maden-Jenkins M. A pragmatic critical appraisal instrument for search filters: introducing the CADTH CAI. *Health Info Libr J* 2009; 26(3): 211-219.
26. Dorée C, Hausner E, Mathisen M, Waffenschmidt S. Value of using different search approaches [online]. 17.10.2016 [Zugriff: 20.04.2017].
URL: <http://vortal.htai.org/?q=node/993>.
27. Welton NJ, Ades AE, Carlin JB, Altman DG, Sterne JAC. Models for potentially biased evidence in meta-analysis using empirically based priors. *J R Stat Soc Ser A Stat Soc* 2009; 172(1): 119-136.
28. Signorovitch JE, Wu EQ, Yu AP, Gerrits CM, Kantor E, Bao Y et al. Comparative effectiveness without head-to-head trials: a method for matching-adjusted indirect comparisons applied to psoriasis treatment with adalimumab or etanercept. *Pharmacoeconomics* 2010; 28(10): 935-945.
29. Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. Methods for population-adjusted indirect comparisons in submission to NICE [online]. 12.2016 [Zugriff: 20.04.2017]. (NICE DSU Technical Support Documents; Band 18).
URL: <http://www.nicedsu.org.uk/TSD18%20Population%20adjustment%20TSD%20-%20FINAL.pdf>.
30. Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ* 2010; 340: c117.
31. Sun X, Briel M, Busse JW, You JJ, Akl EA, Mejza F et al. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. *BMJ* 2012; 344: e1553.

3 Offenlegung potenzieller Interessenkonflikte

Im Folgenden sind die potenziellen Interessenkonflikte der Stellungnehmenden sowie weiterer Teilnehmer an der wissenschaftlichen Erörterung zusammenfassend dargestellt. Alle Informationen beruhen auf Selbstangabe der einzelnen Personen anhand des „Formblatts zur Offenlegung potenzieller Interessenkonflikte“. Das Formblatt ist unter www.iqwig.de abrufbar. Die in diesem Formblatt aufgeführten Fragen finden sich im Anschluss an diese Zusammenfassung.

3.1 Potenzielle Interessenkonflikte von Stellungnehmenden aus Organisationen, Institutionen und Firmen

Organisation / Institution	Name ¹	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6	Frage 7
Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF)	Nothacker, Monika	ja	ja	ja	ja	nein	nein	ja
AstraZeneca GmbH	Fimm, Désirée	ja	nein	nein	nein	nein	nein	nein
	Lützelberger, Uwe	ja	ja	nein	nein	nein	ja	nein
Bayer Vital GmbH	Dintsios, Charalabos-Markos	ja	ja	nein	nein	nein	nein	nein
Boehringer Ingelheim Pharma GmbH & Co. KG	Henschel, Andreas	ja	nein	nein	nein	nein	nein	nein
Bristol-Myers Squibb GmbH & Co. KGaA	Kucka, Annett	ja	ja	nein	nein	nein	nein	nein
	Kupas, Katrin	ja	nein	nein	nein	nein	ja	nein
Bundesverband Medizintechnologie e. V. (BVMed)	Behle, Barbara	ja	nein	ja	nein	nein	nein	ja
Deutsche Gesellschaft für Human-genetik e. V. (GfH)	Heidemann, Simone	ja	nein	ja	nein	ja	nein	ja
Deutsche Gesellschaft für Innere Medizin e. V. (DGIM)	Sauerbruch, Tilman	ja	nein	ja	nein	ja	nein	ja
Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e. V. (GMDS)	Mansmann, Ulrich	nein	ja	nein	nein	nein	nein	ja

¹ Version 11/2016

Organisation / Institution	Name ¹	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6	Frage 7
Deutsche Krankenhausesellschaft e. V. (DKG)	Neumeyer-Gromen, Angela	ja	nein	nein	nein	nein	nein	nein
Deutsches Netzwerk für evidenzbasierte Medizin e. V. (DNEbM)	Lühmann, Dagmar	ja	ja	ja	ja	ja	nein	ja
Deutsche Region der Internationalen Biometrischen Gesellschaft (IBS-DR)	Friede, Tim	ja	ja	ja	nein	nein	nein	nein
Ecker + Ecker GmbH	Ecker, Christof	ja	ja	ja	nein	nein	nein	nein
	Moos, Malte	ja	ja	ja	nein	nein	nein	nein
GlaxoSmithKline GmbH & Co. KG (GSK)	Figel, Bernadette	ja	nein	nein	nein	nein	nein	nein
	Hennig, Michael	ja	nein	nein	nein	nein	ja	nein
Janssen-Cilag GmbH	Eisele, Lewin	ja	nein	nein	nein	nein	nein	nein
	Leisten, Marie-Kristin	ja	nein	nein	nein	nein	ja	nein
Kassenzahnärztliche Bundesvereinigung K. d. ö. R. (KZBV)	Beck, Jörg	ja	nein	nein	nein	nein	nein	nein
Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen e. V. (MDS)	Eyding, Dirk	nein						
	Janatzek, Sandra	nein						
MSD SHARP & DOHME GmbH	Auracher, Veronika	ja	nein	nein	nein	nein	nein	nein
	Rämsch, Christiane	ja	nein	nein	nein	nein	nein	nein
Novartis Pharma GmbH	Hentschke, Christian	ja	nein	nein	ja	nein	nein	nein
	Klebs, Sven	ja	nein	nein	nein	nein	ja	nein
Pfizer Deutschland GmbH	Kühnast, Sarah	ja	nein	nein	nein	nein	nein	nein
	Leverkus, Friedhelm	ja	nein	nein	nein	nein	ja	nein
Roche Diabetes Care GmbH	Arras-Friederich, Christiane	ja	nein	nein	nein	nein	ja	nein
	Berndt, Karsten	ja	nein	nein	nein	nein	nein	nein
Roche Pharma AG	Knoerzer, Dietrich	ja	nein	nein	nein	nein	ja	nein
	Künzel, Thomas	ja	nein	nein	nein	ja	nein	nein
Sanofi-Aventis Deutschland GmbH	Kienitz, Carsten	ja	nein	nein	nein	nein	nein	nein
	Theobald, Karlheinz	ja	nein	nein	nein	nein	ja	nein
SmartStep Consulting GmbH	Ehlert, Andree	ja	nein	nein	nein	nein	nein	nein

Organisation / Institution	Name ¹	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6	Frage 7
SPECTARIS - Deutscher Industrieverband für optische, medizinische und mechatronische Technologien e. V.	Rapp, Cordula	ja	nein	ja	nein	ja	nein	nein
Verband forschender Arzneimittelhersteller e. V. (vfa)	Rasch, Andrej	ja	nein	nein	nein	nein	nein	nein
	Werner, Sebastian	ja	nein	nein	nein	nein	nein	nein
ZVEI – Zentralverband Elektrotechnik- und Elektronikindustrie e. V.	Bätzel, Andreas	ja	nein	nein	nein	nein	nein	nein

3.2 Potenzielle Interessenkonflikte von stellungnehmenden Privatpersonen

Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6	Frage 7
Dintsios, Charalabos-Markos	ja	ja	nein	nein	nein	nein	nein

3.3 Potenzielle Interessenkonflikte von weiteren Teilnehmern der wissenschaftlichen Erörterung (wissenschaftlicher Beirat)

Name ²	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6
Eisert, Albrecht	nein	nein	ja	ja	nein	nein

² Version 11/2013

Im „Formblatt zur Offenlegung potenzieller Interessenkonflikte“ wurden folgende 6 Fragen gestellt (Version 11/2013):

Frage 1: Sind oder waren Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor angestellt bei einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere bei einem pharmazeutischen Unternehmen, einem Hersteller von Medizinprodukten oder einem industriellen Interessenverband?

Frage 2: Beraten Sie oder haben Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor ein Unternehmen, eine Institution oder einen Interessenverband im Gesundheitswesen, insbesondere ein pharmazeutisches Unternehmen, einen Hersteller von Medizinprodukten oder einen industriellen Interessenverband direkt oder indirekt beraten?

Frage 3: Haben Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor direkt oder indirekt von einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere einem pharmazeutischem Unternehmen, einem Hersteller von Medizinprodukten oder einem industriellen Interessenverband Honorare erhalten für Vorträge, Stellungnahmen oder Artikel?

Frage 4: Haben Sie und / oder hat die Einrichtung³, für die Sie tätig sind, abseits einer Anstellung oder Beratungstätigkeit innerhalb des laufenden Jahres und der 3 Kalenderjahre davor von einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere einem pharmazeutischem Unternehmen, einem Hersteller von Medizinprodukten oder einem industriellen Interessenverband finanzielle Unterstützung für Forschungsaktivitäten, andere wissenschaftliche Leistungen oder Patentanmeldungen erhalten?

Frage 5: Haben Sie und / oder hat die Einrichtung, für die Sie tätig sind, innerhalb des laufenden Jahres und der 3 Kalenderjahre davor sonstige finanzielle oder geldwerte Zuwendungen (z. B. Ausrüstung, Personal, Unterstützung bei der Ausrichtung einer Veranstaltung, Übernahme von Reisekosten oder Teilnahmegebühren ohne wissenschaftliche Gegenleistung) erhalten von einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere von einem pharmazeutischen Unternehmen, einem Hersteller von Medizinprodukten oder einem industriellen Interessenverband?

Frage 6: Besitzen Sie Aktien, Optionsscheine oder sonstige Geschäftsanteile eines Unternehmens oder einer anderweitigen Institution, insbesondere von einem pharmazeutischen Unternehmen oder einem Hersteller von Medizinprodukten? Besitzen Sie Anteile eines „Branchenfonds“, der auf pharmazeutische Unternehmen oder Hersteller von Medizinprodukten ausgerichtet ist?

³ Sofern Sie in einer ausgedehnten Institution tätig sind, genügen Angaben zu Ihrer Arbeitseinheit, zum Beispiel Klinikabteilung, Forschungsgruppe etc.

Im „Formblatt zur Offenlegung potenzieller Interessenkonflikte“ (Version 11/2016) wurden folgende 7 Fragen gestellt:

Frage 1: Sind oder waren Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor bei einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere bei einem pharmazeutischen Unternehmen, Hersteller von Medizinprodukten oder einem industriellen Interessenverband angestellt, für diese selbständig oder ehrenamtlich tätig bzw. sind oder waren Sie freiberuflich in eigener Praxis tätig? (Zu den oben genannten Einrichtungen zählen beispielsweise auch Kliniken, Einrichtungen der Selbstverwaltung, Fachgesellschaften, Auftragsinstitute)

Frage 2: Beraten Sie oder haben Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor ein Unternehmen, eine Institution oder einen Interessenverband im Gesundheitswesen, insbesondere ein pharmazeutisches Unternehmen, einen Hersteller von Medizinprodukten oder einen industriellen Interessenverband direkt oder indirekt beraten (z. B. als Gutachter, Sachverständiger, Mitglied eines Advisory Boards, Mitglied eines Data Safety Monitoring Boards (DSMB) oder Steering Committees)?

Frage 3: Haben Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor direkt oder indirekt von einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere einem pharmazeutischem Unternehmen, einem Hersteller von Medizinprodukten oder einem industriellen Interessenverband Honorare erhalten (z. B. für Vorträge, Schulungstätigkeiten, Stellungnahmen oder Artikel)?

Frage 4: Haben Sie oder haben die von Ihnen unter Frage 1 genannten Einrichtungen innerhalb des laufenden Jahres und der 3 Kalenderjahre davor von einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere einem pharmazeutischem Unternehmen, einem Hersteller von Medizinprodukten oder einem industriellen Interessenverband finanzielle Unterstützung z. B. für Forschungsaktivitäten, die Durchführung klinischer Studien, andere wissenschaftliche Leistungen oder Patentanmeldungen erhalten? (Sofern Sie in einer ausgedehnten Institution tätig sind, genügen Angaben zu Ihrer Arbeitseinheit, zum Beispiel Klinikabteilung, Forschungsgruppe etc.)

Frage 5: Haben Sie oder haben die von Ihnen unter Frage 1 genannten Einrichtungen innerhalb des laufenden Jahres und der 3 Kalenderjahre davor sonstige finanzielle oder geldwerte Zuwendungen (z. B. Ausrüstung, Personal, Unterstützung bei der Ausrichtung einer Veranstaltung, Übernahme von Reisekosten oder Teilnahmegebühren für Fortbildungen / Kongresse) erhalten von einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere von einem pharmazeutischen Unternehmen, einem Hersteller von Medizinprodukten oder einem industriellen Interessenverband? (Sofern Sie in einer ausgedehnten Institution tätig sind, genügen Angaben zu Ihrer Arbeitseinheit, zum Beispiel Klinikabteilung, Forschungsgruppe etc.)

Frage 6: Besitzen Sie Aktien, Optionsscheine oder sonstige Geschäftsanteile eines Unternehmens oder einer anderweitigen Institution im Gesundheitswesen, insbesondere von einem pharmazeutischen Unternehmen oder einem Hersteller von Medizinprodukten? Besitzen Sie Anteile eines „Branchenfonds“, der auf pharmazeutische Unternehmen oder Hersteller von Medizinprodukten ausgerichtet ist? Besitzen Sie Patente für ein pharmazeutisches Erzeugnis oder ein Medizinprodukt oder eine medizinische Methode oder Gebrauchsmuster für ein pharmazeutisches Erzeugnis oder ein Medizinprodukt?

Frage 7: Sind oder waren Sie jemals an der Erstellung einer Leitlinie oder Studie beteiligt, die eine mit diesem Projekt vergleichbare Thematik behandelt/e? Gibt es sonstige Umstände, die aus Sicht eines unvoreingenommenen Betrachters als Interessenkonflikt bewertet werden können (z. B. Aktivitäten in gesundheitsbezogenen Interessengruppierungen bzw. Selbsthilfegruppen, politische, akademische, wissenschaftliche oder persönliche Interessen)?

4 Dokumentation der wissenschaftlichen Erörterung – Teilnehmerliste, Tagesordnung und Protokoll

4.1 Teilnehmerliste der wissenschaftlichen Erörterung

Name	Organisation / Institution / Firma / privat
Altenhofen, Lutz	IQWiG
Arras-Friedrich, Christiane	Roche Diabetes Care GmbH
Auracher, Veronika	MSD SHARP & DOHME GmbH
Bätzel, Andreas	ZVEI - Zentralverband Elektrotechnik- und Elektronikindustrie e. V.
Beck, Jörg	Kassenzahnärztliche Bundesvereinigung K. d. ö. R. (KZBV)
Behle, Barbara	Bundesverband Medizintechnologie e. V. (BVMed)
Bender, Ralf	IQWiG
Berndt, Karsten	Roche Diabetes Care GmbH
Breuer, Susanne	IQWiG
Dintsios, Ch.-Markos	Bayer Vital GmbH / privat
Ehlert, Andree	SmartStep Consulting GmbH
Eisele, Lewin	Janssen-Cilag GmbH
Eisert, Albrecht	Wissenschaftlicher Beirat
Eyding, Dirk	Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen e. V. (MDS)
Figel, Bernadette	GlaxoSmithKline GmbH & Co. KG (GSK)
Fimm, Désirée	AstraZeneca GmbH
Friede, Tim	Deutsche Region der Internationalen Biometrischen Gesellschaft e. V. (IBS-DR)
Großelfinger, Robert	IQWiG
Hausner, Elke	IQWiG
Heidemann, Simone	Deutsche Gesellschaft für Humangenetik e. V. (GfH)
Hennig, Michael	GlaxoSmithKline GmbH & Co. KG (GSK)
Henschel, Andreas	Boehringer Ingelheim Pharma GmbH & Co. KG
Hentschke, Christian	Novartis Pharma GmbH
Janatzek, Sandra	Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen e. V. (MDS)
Kienitz, Carsten	Sanofi-Aventis Deutschland GmbH
Klebs, Sven	Novartis Pharma GmbH
Knoerzer, Dietrich	Roche Pharma AG
Kucka, Annett	Bristol-Myers Squibb GmbH & Co. KGaA
Kühnast, Sarah	Pfizer Deutschland GmbH
Künzel, Thomas	Roche Pharma AG
Kupas, Katrin	Bristol-Myers Squibb GmbH & Co. KGaA
Lange, Stefan	IQWiG
Leisten, Marie-Kristin	Janssen-Cilag GmbH

Name	Organisation / Institution / Firma / privat
Leverkus, Friedhelm	Pfizer Deutschland GmbH
Lühmann, Dagmar	Deutsches Netzwerk für evidenzbasierte Medizin e. V. (DNEbM)
Lützelberger, Uwe	AstraZeneca GmbH
Mansmann, Ulrich	Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e. V. (GMDS)
Moos, Malte	Ecker + Ecker GmbH
Neumeyer-Gromen, Angela	Deutsche Krankenhausgesellschaft e. V. (DKG)
Nothacker, Monika	Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF)
Rämsch, Christiane	MSD SHARP & DOHME GmbH
Rapp, Cordula	SPECTARIS – Deutscher Industrieverband für optische, medizinische und mechatronische Technologien e. V.
Rasch, Andrej	Verband Forschender Arzneimittelhersteller e. V. (vfa)
Rörtgen, Thilo	Protokollant (Sitzungsprotokollarischer Dienst des Landtags NRW)
Sauerbruch, Tilman	Deutsche Gesellschaft für Innere Medizin e. V. (DGIM)
Sauerland, Stefan	IQWiG
Schwalm, Anja	IQWiG
Siering, Ulrich	IQWiG
Skipka, Guido	IQWiG
Theobald, Karlheinz	Sanofi-Aventis Deutschland GmbH
Vervölgyi, Volker	IQWiG
Waffenschmidt, Siw	IQWiG
Waltering, Andreas	IQWiG
Werner, Sebastian	Verband Forschender Arzneimittelhersteller e. V. (vfa)
Windeler, Jürgen	IQWiG / Moderation

4.2 Liste der Stellungnahmen, zu denen kein Vertreter an der wissenschaftlichen Erörterung teilgenommen hat

In der folgenden Tabelle werden Stellungnahmen genannt, zu denen trotz Einladung kein Stellungnehmender oder Vertreter zur wissenschaftlichen Erörterung erschienen ist.

Organisation / Institution / Firma / Privatperson
Bundesarbeitsgemeinschaft SELBSTHILFE von Menschen mit Behinderung und chronischer Erkrankung und ihren Angehörigen e.V. (BAG SELBSTHILFE)
Bundesverband der Arzneimittel-Hersteller e. V. (BAH)
Bundesverband der Pharmazeutischen Industrie e. V. (BPI)
Deutsche Diabetes Gesellschaft e. V. (DDG) / Deutsche Gesellschaft für Chirurgie e. V. (DGCH) / Deutsche Gesellschaft für Epileptologie e. V. / Deutsche Gesellschaft für Gynäkologie und Geburtshilfe e. V. (DGGG) / Deutsche Gesellschaft für Hämatologie und Medizinische Onkologie e. V. (DGHO) / Deutsche Gesellschaft für

Kardiologie – Herz- und Kreislaufforschung e. V. (DGK) / Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e. V. (GMDS) / Deutsche Gesellschaft für Neurologie e. V. (DGN) / Deutsche Gesellschaft für Pneumologie und Beatmungsmedizin e. V. (DGP) / Deutsche Gesellschaft für Rheumatologie e. V. (DGRh) / Deutsche Region der Internationalen Biometrischen Gesellschaft (IBS-DR) / (gemeinsame Stellungnahme)
Deutsche Gesellschaft für Hebammenwissenschaft e. V. (DGHWi)
Deutsche Gesellschaft für Orthopädie und Unfallchirurgie e. V. (DGOU)
Deutsche Gesellschaft für Perinatale Medizin (DGPM)
Ecker + Ecker GmbH
Institut für Gesundheits- und Pharmaökonomie (IfGPh)

4.3 Tagesordnung der wissenschaftlichen Erörterung

	Begrüßung
	Einleitung
TOP 1	Fokussierte Informationsbeschaffung
TOP 2	Kosteneffektivitätsvergleiche in HTA-Berichten
TOP 3	Kriterien für einen Beleg aus einer Studie
TOP 4	Metaanalysen
TOP 5	Subgruppenanalysen
TOP 6	Beteiligung von Betroffenen an der Entwicklung von HTA-Berichten
TOP 7	Verschiedenes

4.4 Protokoll der wissenschaftlichen Erörterung

Datum: 22. März 2017; 10:30 bis ca. 13:30 Uhr

Ort: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG),
Im Mediapark 8, 50670 Köln

Moderation: Jürgen Windeler

Stefan Lange: Schönen guten Tag! Wir fangen jetzt schon mal an. Herr Windeler müsste aber jeden Moment eintreffen. Es gab gerade eine kleine Terminkollision.

Ich stelle mich kurz vor: Mein Name ist Stefan Lange. Ich bin hier der stellvertretende Leiter. Insofern vertrete ich jetzt Herrn Windeler an dieser Stelle.

Herzlich willkommen zur wissenschaftlichen Erörterung der Stellungnahmen zum Entwurf der Allgemeinen Methoden, Version 5.0. Ich kann mich nicht erinnern, dass wir jemals so zahlreich hier zusammengesessen haben. Das ist sehr erfreulich, weil es zeigt, dass das, was

wir hier so machen, auf ein gewisses Interesse stößt und wir durchaus interessante Vorschläge zu diskutieren haben.

Wie alle Erörterungen hier im Hause – das werden vermutlich die meisten von Ihnen wissen – wird auch diese aufgezeichnet, also sowohl elektronisch als auch stenografisch. Das wurde Ihnen mit der Einladung mitgeteilt. Wer jetzt nicht damit einverstanden ist aus datenschutzrechtlichen Gründen, persönlichen Gründen, der müsste das jetzt kundtun und den Raum verlassen. - Das scheint bei Niemandem der Fall zu sein. Dann freue ich mich, denn dann bleiben wir so zahlreich. Das ist der erste formale Punkt.

Der zweite Punkt ist, dass wir Ihnen eine Tagesordnung zugesendet haben, wie wir das auch immer machen, mit den Punkten, wo wir aus Ihren Stellungnahmen heraus Diskussionsbedarf bzw. Erläuterungsbedarf sehen, wo wir Rückfragen haben, was nicht bedeutet, dass alle anderen Punkte, die Sie uns mitgeteilt haben, in der Versenkung verschwinden, sondern die waren für uns klar und führen entweder zu einer entsprechenden Änderung im Methodenpapier oder zumindest zu einer, wie wir das nennen, Würdigung, wo wir Ihre Argumente aufgreifen und uns dazu positionieren und darlegen, warum wir nicht glauben, dass sie geeignet sind, den Entwurf in der entsprechenden Weise zu ändern.

Nichtsdestotrotz haben Sie die Gelegenheit, unter Punkt „Verschiedenes“ irgendetwas anzubringen, was Sie unter den ersten sechs TOPs nicht finden. Wir sind gerne bereit, uns das anzuhören.

Was nicht notwendig ist, ist, die eigenen Stellungnahmen noch einmal ausführlich zu verlesen. Die kennen wir. Manche werden ja vorher auch schon veröffentlicht. Insofern besteht wahrscheinlich auch kein Bedarf.

Der letzte Punkt, den wir an dieser Stelle immer erwähnen, ist, dass es heute nicht darum geht, die gesetzlichen Grundlagen zu diskutieren, die im SGB V verankert sind. Wir werden uns bemühen, wenn solche Punkte aufkommen, sie relativ kurz abzuhandeln und zu beenden.

Jetzt freue ich mich sehr, dass Jürgen Windeler da ist, weil ich nun die schwere Aufgabe der Moderation abgeben kann.

Sollte es Fragen geben, können Sie die natürlich im Vorfeld noch stellen. Ansonsten werden wir nach der kurzen Begrüßung durch Herrn Windeler in die Tagesordnung einsteigen.

Moderator Jürgen Windeler: Schönen guten Morgen! Ich bitte um Verständnis oder um Nachsicht, dass sich noch ein Termin etwas verlängert hat, der leider nicht nur aus Höflichkeitsgründen nicht abubrechen war, aber mir war völlig klar, dass hier eine kompetente Vertretung meinen Job übernehmen würde. Insofern versuche ich nun, da einzusteigen.

Ich begrüße Sie zu dieser wissenschaftlichen Erörterung.

Ich möchte vor Eintritt in die Tagesordnung etwas sagen, was aber nichts mit Formalia zu tun hat. Ich möchte gerne einige Punkte ansprechen, die nicht auf der Tagesordnung stehen. Ich

möchte auch kurz ansprechen, dass und warum sie nicht auf der Tagesordnung stehen. Es sind drei Aspekte.

Der erste Aspekt ist der, dass wir in vielen Stellungnahmen die Forderungen – „Wünsche“ ist ein höflicher Begriff – gelesen haben, die man eigentlich besser nach Berlin hätte schicken sollen. Wegelystraße 8 ist der Gemeinsame Bundesausschuss, Friedrichstraße 108 ist das BMG. Und wenn es um die Frage geht, dass gesetzliche Regelungen oder aber Verfahrensordnungsregelungen des G-BA adressiert werden, dann möge man bitte nicht Forderungen an das IQWiG stellen. Ich war auch ein bisschen überrascht, weil man ja einige Dinge im Gesetz nachlesen kann, also etwa Vertraulichkeitsregelungen im Zusammenhang mit Dossierbewertungen, Vertraulichkeitsregelungen im Zusammenhang mit 137e und h. Das steht alles im Gesetz. Dann Forderungen zu lesen, das IQWiG möge doch die Unterlagen zur Verfügung stellen, hat mich ein bisschen überrascht. Wir werden uns auch mit diesen Dingen in der Würdigung nicht näher auseinandersetzen, sondern nur auf gesetzliche Regelungen verweisen.

Barbara Behle: Ich verstehe Ihren Punkt völlig, Herr Prof. Windeler. Es geht nur um Formalia. Beispielsweise bitten wir, dass die Bewertung von Medizinprodukten hoher Risikoklassen, so wie Sie es häufig formuliert haben, doch bitte konform dem vom Gesetzgeber vorgegebenen Formulierungsvorschlag „Bewertung der neuen Untersuchungs- und Behandlungsmethoden mit Medizinprodukten hoher Risikoklassen“ ... Also da geht es formal eher um solche Aspekte, was wir angemerkt haben.

Moderator Jürgen Windeler: Ich bin ja in meiner Äußerung nicht auf bestimmte einzelne Stellungnahmen eingegangen. Hier ging es primär nicht nur um Formulierungsänderungen und -wünsche, sondern um weiterreichende Beteiligungsmöglichkeiten und -wünsche. Ob das jetzt der BVMed so aufgeschrieben hat, weiß ich nicht, aber in vielen anderen Stellungnahmen stand das so drin. Solche Wünsche sind einfach nicht zu erfüllen.

Der zweite Punkt, der mit dem ersten etwas zu tun hat, ist, dass in sehr vielen Stellungnahmen Beteiligungen an den Prozessen des IQWiG thematisiert worden sind, Beteiligungen vor Bewertungen, während der Bewertungen und nach den Bewertungen in vielfältiger Form. Ich möchte an dieser Stelle zum Ausdruck bringen, dass das IQWiG solche Beteiligungen im Sinne von Mitarbeit Sachverständiger als auch Stellungnahmeverfahren sehr ernst nimmt. Ich erlaube mir auch, darauf hinzuweisen, dass wir Diskussionen, die wir auch hausintern hatten zu der Frage, wie wir Prozesse im Institut verbessern können, so ausgestalten können, dass möglicherweise Fristen, die uns gesetzt werden, besser eingehalten werden können, immer wieder geführt haben und führen mussten, ob wir in diesem Zusammenhang auch Stellungnahmeverfahren, Stellungnahmemöglichkeiten beschneiden, ob wir Fristen verkürzen – wohlgemerkt: verkürzen, nicht verlängern –. Wir haben uns in allen Fällen dagegen entschieden, uns Stellungnahmemöglichkeiten zu nehmen, und haben die belassen. Wir werden nur Stellungnahmemöglichkeiten, Beteiligungsmöglichkeiten, abgesehen von gesetzlichen Regelungen, nicht ausbauen können. Wir sehen auch keine Notwendigkeit dazu. Denn das, was wir im Moment in den Stellungnahmeverfahren mitbekommen, thematisieren, auch heute thematisieren, erscheint uns ausreichend. Ich will ergänzen, dass ich mich immer etwas wundere, wenn die Beteiligungsmöglichkeiten beim IQWiG auf der einen Seite eingefordert werden, auf der anderen Seite dann die Politik in Berlin angesprochen wird mit der Klage, dass alles so lange dauert, und es teilweise die gleichen Beteiligten sind. Da ist

mein Verständnis für den Wunsch, sich zu beteiligen, etwas limitiert. Damit werden wir uns also auch nicht weiter auseinandersetzen.

Der dritte Punkt ist der, dass wir in einigen neuen Kapiteln, wo wir im Grunde so etwas wie einen Platzhalter gesetzt haben, um zu sagen, das machen wir, für diesen Platzhalter sehr ausführliche, teilweise seitenlange Ausführungen dazu bekommen haben, dass das doch alles noch viel genauer beschrieben werden muss, in Details, Gewichtigungen und ich weiß nicht in welchen Priorisierungen noch alles. Ich will an dieser Stelle klar sagen: Wir halten diese Beschreibung für ausreichend, werden zu der Frage, wie wir internationale Projekte – das ist einer dieser Punkte – im Detail ausgestalten und auswählen, im Methodenpapier nicht weiter ausführen. Wenn wir hier irgendjemandem Rechenschaft ablegen müssen, auch was zum Beispiel Mittelverwendung angeht, dann werden wir das gegenüber unseren Vorständen, Aufsichtsgremien machen. Wir sehen im Moment keine Notwendigkeit, das im Methodenpapier näher zu spezifizieren.

Diese Punkte wollte ich ansprechen, damit deutlich ist, warum sie nicht auf dieser Tagesordnung auftauchen, und um anzukündigen, dass sie auch in der schriftlichen Stellungnahme keinen großen Raum einnehmen werden.

Unter „Verschiedenes“ besteht Gelegenheit, über die Tagesordnung hinaus Punkte anzusprechen.

Es gibt eine Aufzeichnung, einen Protokollanten und ein Band. Beide wären sehr glücklich darüber, wenn Sie bitte bei jeder Wortmeldung, insbesondere weil es so viele sind, Ihren Namen sagen, nicht nur beim ersten Mal, wie es sonst üblich ist, damit wirklich alle Äußerungen, die getätigt werden, den richtigen Personen im Protokoll zugeordnet werden können.

Wenn es jetzt keine Wortmeldungen gibt, dann würde ich jetzt in die Tagesordnung einsteigen.

Tagesordnungspunkt 1:

Fokussierte Informationsbeschaffung

Siw Waffenschmidt: Guten Morgen! Es geht um das Kapitel 8 „Informationsbeschaffung“. Das haben wir weitreichend überarbeitet in dieser Fassung. Eine der größten Änderungen ist, dass wir die Recherchen in „umfassende Informationsbeschaffung“ und „fokussierte Informationsbeschaffung“ aufgeteilt haben und dann eine „orientierende Recherche“. Dazu kamen einige Stellungnahmen. Zu zwei Stellungnahmen haben wir eine Rückfrage. Eine Frage richtet sich an die KZBV, die darstellt, dass sie gerne eine weitere inhaltliche Differenzierung haben will. Unsere Frage an die KZBV ist, welche Informationen denn genau fehlen, um die Unterschiede zwischen den Recherchen oder der Informationsbeschaffung deutlicher zu machen.

Jörg Beck: Ich kann darauf direkt antworten. Wir hatten registriert, dass Sie das neu eingeführt haben, haben ja auch den Quellenverweis dort platziert. Für uns ist beim Lesen im

Methodenpapier im ersten Moment nicht ganz deutlich geworden, wann welches Verfahren zum Einsatz kommen soll. Das war der primäre Grund, was wir angemerkt haben.

Siw Waffenschmidt: Wir haben dort Beispiele aufgeführt. Wir haben ja zum einen gesagt, dass es auf jeden Fall eine systematische Recherche ist und dass es eben bestimmte Gründe gibt, warum man dann eine fokussierte Recherche machen sollte. Bei Nutzenbewertungen werden in der Regel natürlich umfassende Informationsbeschaffungen gemacht. Wir haben drei Beispiele aufgeführt, bei denen wir eine fokussierte Recherche jetzt schon durchführen. Es kann aber natürlich in Zukunft sein, dass das noch für andere Punkte der Fall sein wird. Von daher ist es jetzt schwierig, das genauer aufzuführen, wenn man das nicht jedes Jahr aktualisieren will, wo man das macht und wo man das nicht macht.

Jörg Beck: Vielen Dank für die Erläuterung. Das wird sich, wie Sie schon sagten, im Verfahren dann zeigen.

Siw Waffenschmidt: Gut, dann ist diese Frage erledigt.

Wir haben noch eine zweite Frage, und zwar an Novartis. Da ist uns nicht ganz klar, ob das gegebenenfalls ein Missverständnis ist. Für die Suche nach dem Maß des Gesamtnutzens schlagen wir eine fokussierte Informationsbeschaffung durch. Wir sind uns nicht ganz klar, ob Novartis da mit uns übereinstimmt oder nicht, ob das ein Missverständnis war. Können Sie das kurz sagen? Falls es für Sie kein Missverständnis ist, falls Sie also nicht mit uns übereinstimmen, dass eine fokussierte Informationsbeschaffung da ausreichend ist, dann lautet unsere Frage: Warum?

Sven Klebs: Ich habe Ihre Frage, ehrlich gesagt, nicht genau verstanden, weil sich unser Punkt im Grunde nur auf einen formalen Aspekt beschränkt, nämlich ob man nicht klarer darstellen könnte, dass es unter gewissen Bedingungen ausreichend ist, zum Beispiel durch eine NOT-Suche Kongress-Abstracts aus der Suche von vornherein herauszunehmen, weil die Quelle, die Sie auch zitieren, zeigt, dass das in den meisten Fällen keinen besonderen Nutzen bringt. Es geht dann eher darum, dass es nicht zu einer Kritik kommt, dass man unvollständig gesucht hätte, weil man sich eben von vornherein darauf beschränkt hat, diese herauszunehmen. Das war im Prinzip unsere wesentliche Anmerkung.

Siw Waffenschmidt: Wir reden über zwei verschiedene Stellungnahmen. Deswegen kläre ich das. Die Stellungnahme ist im Abschnitt 4.3.4 „Maß des Gesamtnutzens“. Wir geben da an, dass es eine fokussierte Recherche sein soll. Ich glaube, dass Sie da nicht mit uns übereinstimmen. Der Verweis von 4.3.4 - das ist das, wie eine Kosten-Nutzen-Bewertung durchgeführt wird - bezieht sich dann natürlich auf Kapitel 8. Deswegen habe ich das jetzt genannt.

Sven Klebs: Unser Verständnis war, dass Sie bei der Kosten-Nutzen-Bewertung eine fokussierte Recherche durchführen würden. Da die Kosten-Nutzen-Bewertung ähnlich strukturiert ist von der Evidenzanforderung wie die normale Nutzenbewertung, war unsere Frage, warum man dann, wie Sie eben gesagt haben, nicht auch eine breit angelegte Suche, sondern eine fokussierte wählt. Sie haben ja eben als Beispiel aufgeführt, dass bei der frühen Nutzenbewertung eine breite Suche und keine fokussierte Suche durchgeführt wird.

Siw Waffenschmidt: Dann ist es klar. Für die Nutzenbewertung wird natürlich eine umfassende Informationsbeschaffung gemacht. Es geht hier nur um eine Teilfragestellung innerhalb von Kosten-Nutzen-Bewertungen. Da müssen ja sehr viele Recherchen durchgeführt werden. Für diese eine Teilfragestellung würden wir es als ausreichend ansehen, wenn eine fokussierte Informationsbeschaffung durchgeführt wird, weil wir das nach Studien zum Maß des Gesamtnutzens für ausreichend halten.

Sven Klebs: Da fehlt mir jetzt eine Vorstellung, wie so ein Fall aussehen könnte, ob da wirklich eine fokussierte Suche ausreichen würde, wo die Unterschiede wirklich entstehen könnten, ob man relevante Unterschiede generieren würde mit 2 Suchstrategien. Dazu kann ich jetzt nicht zu sagen, ob es tatsächlich einen Unterschied machen würde, weil mir ein praktisches Beispiel fehlt.

Siw Waffenschmidt: Gut. Dann ist unsere Frage beantwortet. Das ist für uns klar.

Moderator Jürgen Windeler: Gibt es weitere Wortmeldungen zu dem Thema? - Das sehe ich nicht. Frau Waffenschmidt hatte schon gesagt, dass die Frage beantwortet ist, Unklarheiten beseitigt sind.

Wir kommen zu

Tagesordnungspunkt 2:

Kosteneffektivitätsvergleiche in HTA-Berichten

Anja Schwalm: Ich komme aus dem Bereich der Gesundheitsökonomie. Die Frage bezieht sich auf das Kapitel 6, auf die ökonomische Bewertung in den HTA-Berichten beim Themencheck Medizin.

Eine kurze Erläuterung: Für diese ökonomische Bewertung beim Themencheck Medizin gibt es grundsätzlich zwei Möglichkeiten. Das wäre einmal die Erstellung einer systematischen Übersicht oder die Durchführung einer eigenen gesundheitsökonomischen Analyse. Auf das letztere bezieht sich die Frage, die wir haben.

Wir haben für diese eigene gesundheitsökonomische Analyse auf die entsprechenden Kapitel im Abschnitt zur Kosten-Nutzen-Bewertung verwiesen, haben aber explizit den Abschnitt zur Effizienzgrenze herausgenommen, weil nach unserer Ansicht das Ziel der HTA-Berichte beim Themencheck Medizin nicht die Empfehlung eines Erstattungsbetrages ist. Die GMDS hat in ihrer Stellungnahme darauf hingewiesen, dass somit unklar ist, wie die Kosteneffektivität beurteilt werden kann, da es ja in Deutschland keinen expliziten Schwellenwert gibt. Sie schlagen unter anderem eine Orientierung an den WHO-Kriterien vor. Die WHO-Kriterien sehen ja vor, dass eine Intervention als kosteneffektiv beurteilt wird, sofern das inkrementelle Kosteneffektivitätsverhältnis das Dreifache des Pro-Kopf-Bruttoinlandprodukts nicht übersteigt. Damit würde ja de facto ein Schwellenwert etabliert, den, wie die GMDS in ihrer Stellungnahme ausgeführt hat, es ja nicht gibt. Unsere Frage richtet sich also konkret an die GMDS: Welche Möglichkeiten schlagen Sie vor, um Aussagen zur Kosteneffektivität abzuleiten, ohne dabei auf einen expliziten indikationsübergreifenden Schwellenwert zu rekurren?

Ulrich Mansmann: Ich muss diese Frage an Uwe Siebert weitergeben, der in der Kommission das Thema bearbeitet.

Moderator Jürgen Windeler: Der aber nicht da ist, wenn ich das richtig verstehe.

Ulrich Mansmann: Ja.

Moderator Jürgen Windeler: Wir können jetzt gerne zur Kenntnis nehmen, dass die GMDS diese Frage nicht beantworten kann. Aber eigentlich ist die mündliche Erörterung dazu da, die Dinge hier zu thematisieren.

Ulrich Mansmann: Im Moment habe ich darauf keine schlüssige Antwort.

Tim Friede: Vielleicht darf ich kurz einhaken. Wir haben gemeinsam eine Stellungnahme abgegeben. Es war also nicht nur die Stellungnahme der GMDS, sondern gemeinsam mit der IBS-DR. Ich kann diese Frage hier aber auch nicht beantworten, muss aber dazu sagen, dass ich ursprünglich nachgefragt habe, ob es möglich wäre, dass pro Fachgesellschaft zwei kommen dürfen. Dann hätten wir uns natürlich etwas breiter aufstellen können.

Moderator Jürgen Windeler: Gibt es jemand anderen im Raum in der relativ großen Runde, der zu dieser Frage einen Beitrag leisten kann?

Simone Heidemann: Ich habe noch einen anderen Aspekt zu diesem Tagesordnungspunkt.

Moderator Jürgen Windeler: Okay. Wir bleiben bei diesem Punkt. Es scheint also nicht der Fall zu sein, Frau Schwalm. Dann muss diese Frage offenbleiben, und wir müssen versuchen, in der Stellungnahme damit umzugehen.

Monika Nothacker: Die Antwort kann doch sicher nachgereicht werden. Das wäre ja der allgemeinen Auseinandersetzung förderlich.

Moderator Jürgen Windeler: Das ist grundsätzlich richtig, aber noch schöner wäre es, wenn wir die mündlichen Erörterungen machen und die Dinge diskutieren könnten. Mit nachgereichten Unterlagen ist eine Diskussion natürlich nicht möglich. Wir werden auch keine langen Schriftwechsel machen. Wir können uns gerne darauf verständigen, dass die Stellungsnehmenden innerhalb kurzer Zeit, diese Woche, uns zu dieser Frage von Frau Schwalm eine Rückmeldung geben, aber wir werden diese Rückmeldung nehmen und uns nicht weiter mit Rückfragen damit auseinandersetzen. Das ist schade, muss man klar sagen.

Stefan Lange: Vielleicht darf ich noch ergänzen, Frau Nothacker: Das schadet auch der Transparenz. Es wird ja immer hochgehalten, wie wichtig Transparenz ist, und dem Institut in Stellungnahmen immer gerne unterstellt, es sei sehr intransparent. Das ist sicher nicht so furchtbar transparent.

Monika Nothacker: Vielen Dank. Ich denke, es ist ja jetzt hier transparent diskutiert und beschlossen worden und kann dann entsprechend veröffentlicht werden.

Moderator Jürgen Windeler: Letzte Bemerkung von mir, dass die Tagesordnung vor einer Woche verschickt worden ist, wie ich mich gerade noch mal vergewissert habe, und vielleicht gewisse Möglichkeiten bestanden haben, das vorzubereiten. Noch mal: Die mündliche Erörterung ist dazu da, die Dinge hier zu diskutieren und im hin und her zu erörtern, nachzufragen. Wir machen es in diesem Fall so, dass Sie das nachreichen können. Eigentlich ist das aber nicht Sinn der Sache.

Ulrich Mansmann: Eigentlich liegt hier vielleicht auch ein Missverständnis vor. Wir haben in unserer Stellungnahme auf dieses Problem hingewiesen. Uns war nicht klar, dass wir es lösen müssen. Aber ich werde dies gerne nachliefern.

Moderator Jürgen Windeler: Okay. Das lasse ich mal so stehen.

Ch.-Markos Dintsios: Es gibt ja hinlänglich Auseinandersetzungen in der internationalen Literatur zur Bestimmung von Schwellenwerten, vor allem auf indikationsübergreifende aggregierte Endpunkte, wie dem QALY-Ansatz sozusagen. Der eine Vorschlag ist der Vorschlag mit dem dreifachen Bruttoinlandsprodukt per capita von der WHO. Bei Drittweltländern sieht das ein bisschen anders aus. Da ist er adaptiert. Da gibt es aus der USA Literatur, aus England, wo es auch angewendet wird. Ich denke, wie Sie es geschildert haben, Sie haben nur die Problematik aufgegriffen. Eine Antwort darauf wollten Sie nicht oder konnten Sie nicht geben. Vielleicht kann sich das IQWiG mit der Materie trotzdem befassen, weil es muss ja irgendwann, wenn es vor dieser Situation steht, eine Aussage treffen, ob eine Technologie im Vergleich zu einer anderen kosteneffektiv ist. Dann kann es, wenn es Glück hat, auf Basis von entsprechenden Recherchen Literatur dazu finden, aber vielleicht muss es ja den Weg gehen, den Frau Schwalm eben beschrieben hat, nämlich eine eigene Bewertung zu machen. Und dann braucht es für sich selber einen Referenzpunkt. Das ist, glaube ich, eher ein Auftrag zurück an das IQWiG, vielleicht unter Berücksichtigung der internationalen Literatur.

Simone Heidemann: Wir haben Ihnen eine kurze Stellungnahme zukommen lassen, die ich am ehesten diesem Tagesordnungspunkt zuordnen würde, Kosteneffektivitätsvergleiche. Wir haben Sie darauf hingewiesen, dass in Ihrem Methodenpapier das Gendiagnostikgesetz überhaupt keine Berücksichtigung findet. Wir finden das aber wichtig, weil dieses Gesetz, das seit 2010 in Kraft ist, für genetische Untersuchungen erhebliche Zusatzkosten verursachen kann, für verpflichtende genetische Beratung, für Vernichtungspflichten usw. Nun mögen Sie vielleicht geneigt sein, zu sagen, Sie können sich immer noch damit beschäftigen, wenn Sie sich mal mit genetischen Untersuchungen beschäftigen sollten. Es ist allerdings so, dass dieses Gesetz eine Legaldefinition enthält, die genetische Untersuchungen sehr umfangreich definiert, viel umfangreicher als Sie und ich das vorher jemals für möglich gehalten hätten. Zum Beispiel werden Genproduktanalysen als genetische Untersuchungen definiert. Oder auch vorgeburtliche Untersuchungen, Ultraschall, Blutuntersuchungen am Blut der Schwangeren fallen unter Umständen in den Anwendungsbereich. Wenn man darauf nicht gefasst ist, könnte es natürlich passieren, dass Kosten-Nutzen-Analysen nicht korrekt sind, wenn man diese Beratungspflichten und Vernichtungspflichten nicht in die Kostenanalyse aufnimmt.

Moderator Jürgen Windeler: Ich bin Ihnen für den letzten Satz dankbar, der den Fokus Ihrer Stellungnahme, was diesen konkreten Tagesordnungspunkt angeht, unterstreicht. Ansonsten hat Herr Lange vermutlich in der Einführung gesagt, dass die Punkte, die wir jetzt hier adressieren, nicht bedeutet, dass wir die anderen nicht zur Kenntnis nehmen, sondern nur bedeutet, dass wir die anderen alle zur Kenntnis genommen haben, meinen, verstanden zu haben und auch – das betrifft auch das Gendiagnostikgesetz - in angemessener Weise im Methodenpapier abbilden werden. Aber der Hinweis auf diesen konkreten Punkt ist völlig nachvollziehbar.

Tilman Sauerbruch: Es ist vielleicht eine Frage aus Unkenntnis heraus: Tauchen diese Kosteneffektivitätsvergleiche regelhaft in den AMNOG-Dossiers vom IQWiG auf?

Moderator Jürgen Windeler: Nein. Die Kosteneffektivitätsvergleiche, die jetzt hier angesprochen sind, sind im Moment relevant für die HTA-Berichte im Zusammenhang mit dem Themencheck Medizin. Dossierbewertungen enthalten keine Kosteneffektivitätsvergleiche, weder in den Dossiers noch in unseren Dossierbewertungen, wie auch alle anderen Produkte im Moment nicht.

Tilman Sauerbruch: Ist das geplant?

Moderator Jürgen Windeler: Es steht im Methodenpapier, dass das möglich ist. Es steht im Gesetz, dass das möglich ist. Aber damit das möglich ist, muss irgendjemand das IQWiG beauftragen. Da es bisher nicht passiert ist und nach meiner Information auch nicht geplant ist, ist es im Moment jedenfalls nicht kurzfristig absehbar, dass solche Dinge in die Produkte des IQWiG einfließen.

Gibt es weitere Wortmeldungen? – Das sehe ich nicht.

Dann kommen wir zum nächsten Punkt, einen Punkt, zu dem relativ intensiv Stellung genommen wurde.

Tagesordnungspunkt 3:

Kriterien für einen Beleg aus einer Studie

Guido Skipka: Ich komme aus der Biometrie im IQWiG. Ich beziehe mich auf den Abschnitt 3.1.4 „Endpunktbezogene Bewertung“. Da gibt es eine Passage zur Frage, wann man aus einer Studie einen Beleg ableiten kann. Es ist so, dass wir regelhaft mehr als eine Studie für einen Beleg fordern und nur im Ausnahmefall aus einer Studie einen solchen Beleg ableiten. Das sieht in den Dossiers, die wir bekommen, nach unserer Erfahrung anders aus. Da wird regelhaft aus einer Studie ein Beleg abgeleitet. Wir hätten gerne konkrete Kriterien für diesen aus unserer Sicht Sonderfall. Ein wichtiges Kriterium wäre die Konsistenz der Ergebnisse verschiedener Teilpopulationen. Da bieten sich Betrachtungen einzelner Zentren an. Wir haben ja auch das Kriterium, dass wir mindestens zehn Studienzentren fordern.

Für die Beurteilung der Konsistenz halten wir präzise Schätzungen innerhalb der Zentren für notwendig. Unser Vorschlag in dem Entwurf 5.0 war, dass wir mindestens 1.000 Patienten pro Studienarm vorliegen haben. Dazu gab es in einigen Stellungnahmen Kritik, die Zahl sei

zu hoch, außerdem nicht begründet. Unseres Wissens gibt es keine konkreten Vorschläge in der Literatur, wie man diese Konsistenzprüfungen machen sollte, mit welchen Zahlen. Unsere Überlegungen waren folgende: Wenn wir 1.000 Patienten und zehn Zentren haben, dann haben wir eine Gruppengröße pro Zentrum von etwa 100 Patienten. Wenn man jetzt von einer Ereignisrate von 10 % ausgeht, würden sich in jedem Arm in jedem Zentrum etwa zehn Ereignisse realisieren. Das halten wir für erforderlich, um eine ausreichende Präzision für die Konsistenzprüfung zu haben. Es gibt andere Stellungnahmen, die erachten überhaupt das Kriterium auf Basis der Fallzahl für nicht sinnvoll.

Unsere konkrete Frage an Sie lautet: Wie viele Patienten erachten Sie denn als notwendig, damit man ausreichend präzise eine Konsistenzprüfung machen kann, bzw. welche Kriterien würden Sie sonst vorschlagen?

Dietrich Knoerzer: Wir hatten in unserer Stellungnahme versucht, diese CGMP-Points-to-consider-guidance populär zu machen - die waren im Entwurf 4.2 auch mit drin - und sich anhand dieser Kriterien entlang zu hangeln, also interne Validität, wie gerade angesprochen, externe Validität, was natürlich schwierig ist - Herr Windeler, Sie haben es ja schon extra publiziert -, und die ganzen anderen Kriterien, die damit durchgegangen werden.

Warum wir die Fallzahl per se für kritisch halten, ist: Für eine gutgemachte Studie gibt es ja idealerweise eine Fallzahlplanung. Wenn diese Fallzahlplanung durch einen großen Effekt oder sonst was eine kleinere Fallzahl ergibt als diese 1.000, dann ist diese Studie ja trotzdem gut gemacht und auch ethisch vertretbar. Deswegen schienen uns die 1.000 statistisch wenig begründet.

Guido Skipka: Eine direkte Antwort darauf: Natürlich hat jede vernünftige Studie eine Fallzahlplanung. Und da kommen zum Teil natürlich deutlich geringere Zahlen raus. Jetzt muss man sagen, dass sich die Fallzahlplanung immer nur auf den primären Endpunkt bezieht. Jetzt steht auch im Dokument, das Sie angesprochen haben, dass man, wenn man aus einer Studie so etwas wie einen Beleg ableitet, da auch Konsistenz über Endpunkte hinweg haben muss. Von daher kann man sich aus unserer Sicht nicht nur auf die Fallzahl für den primären Endpunkt beschränken.

Monika Nothacker: Das ist richtig. Wir hatten in unserer Stellungnahme von einer angemessenen Fallzahlberechnung gesprochen, was natürlich in der Regel nicht der Fall ist. Aber trotzdem haben Sie gerade selber von den Ereignisraten gesprochen und eine willkürlich angenommen. Die könnte ja unterschiedlich sein. Aus unserer Sicht könnte das nur ein Anhaltspunkt sein. Was man dann überprüfen müsste, wären tatsächlich Effektstärken und Ereignisraten. Das würde man aufnehmen, selbst wenn die nicht primär berechnet worden sind.

Stefan Lange: Um es noch mal deutlich zu machen – möglicherweise gibt es da ein Missverständnis –: Grundsätzlich haben wir in der Wissenschaft das Replikationserfordernis. Das ist eigentlich unumstritten, behaupte ich jetzt einfach mal so, es sei denn, jemand sagt, das ist nicht richtig, wie es heutzutage ja gerne mal gemacht wird. Das heißt, wenn man auf dieses Replikationserfordernis verzichtet, was es ja bedeutet, wenn man nur eine Studie hat, dann wird sie nicht repliziert. Dann muss man doch Ersatzkriterien haben. Ein Ersatzkriterium ist die Konsistenz, wie Guido Skipka gerade ausgeführt hat. Die Frage der Konsistenz kann

sich doch in der ursprünglichen Fallzahlplanung überhaupt nicht widerspiegeln. Insofern passt doch dieses Argument überhaupt nicht. Wir wissen doch alle, wie gering die statistische Power von solchen Konsistenzbetrachtungen ist. Die geht gegen null. Wenn natürlich in der ursprünglichen Fallzahlplanung genau das mitberücksichtigt worden wäre, zu sagen, ja, wir wollen Konsistenz über die Zentren zeigen, dann kann man darüber reden, aber ich kenne, ehrlich gesagt, keinen einzigen Fall, Herr Knoerzer. Wenn Sie uns so etwas schildern könnten, dann wären wir wahnsinnig dankbar, und dann werden wir das natürlich aufgreifen.

Frau Nothacker, das ist so ein bisschen wie mit der Frage eben. Wir kriegen immer gerne in Stellungnahmen Probleme aufgezeigt, die wir lösen sollen, wie Herr Dintsios eben auch gesagt hat. Wenn wir dann mal Vorschläge machen, versuchen, Probleme zu lösen, dann heißt es: Die sind doch alle gar nicht wohlbegründet. - Wir wären wirklich irrsinnig dankbar, sehr dankbar, wenn wir für Lösungsvorschläge, die wir unterbreiten, die dann aber mangels irgendwelcher Fundierung kritisiert werden, Alternativvorschläge bekommen und nicht so allgemeine, wie sie in der CHMP-Guideline drinstehen, wir sollen mal die interne Validität, die wir sowieso prüfen, die eh klar ist, und die externe Validität, die eh klar ist, heranziehen. Das hilft wirklich nicht, Herr Knoerzer.

Dietrich Knoerzer: Es sind jetzt mehrere Punkte gewesen. Der erste Punkt war, ob ich eine Studie kenne, die eine Fallzahlplanung gemacht hat für Konsistenzprüfungen. Die Antwort ist: natürlich nicht. Ich glaube auch – so ehrlich muss man halt auch sein –, dass Sie so eine Studie durch keine Ethikkommission dieser Welt durchkriegen. Die wäre ja dann für die wichtigsten Endpunkte massiv overpower. Und dann würde jeder sagen: Es ist schwer, so eine Studie durchzudrücken. Das ist der eine Punkt.

Der andere Punkt ist: Jetzt kann man sich natürlich gegen diese CHMP-Guidance stellen oder sagen, die ist nicht genug. Es war ein Versuch - wir haben das auch schon mal in Dossiers probiert –, zu sagen, das ist eine Absprungbasis, da sind Kriterien genannt, die man ein bisschen operationalisieren könnte. Natürlich ist das ein bisschen generell. Das haben ja die Guidances so an sich. Aber man könnte ja gemeinsam darüber reden, was man darunter versteht. Vorschläge haben wir in Dossiers, nicht in der Stellungnahme schon gemacht.

Ulrich Mansmann: Ich meine, wir haben das auch als Problem identifiziert in unserer Stellungnahme, haben empfohlen, dass man das vielleicht über Simulationsstudienmodelle angehen könnte. Vielleicht die generelle Empfehlung an Formulierungen dieser Art, dass man die vielleicht nicht so apodiktisch formuliert, sondern da noch einen gewissen Ermessensspielraum in die Formulierung einbringt.

Michael Hennig: Ich möchte genau diesen Punkt aufgreifen, weil ich finde, dass eine absolute Setzung von einer naturgemäß willkürlichen Zahl wie 1.000 in diesem Fall nicht zielführend ist, weil man auch besondere Konstellationen berücksichtigen muss. Wenn wir über seltene Indikationen reden, wenn wir über onkologische Entwicklungen reden, macht eine absolute Fallzahl weniger Sinn, sodass ich mir wünschen würde, dass man das jetzt nicht so kategorisch sieht, wie das in dem aktuellen Entwurf genannt wurde mit einer absoluten Fallzahl von 1.000.

Désirée Fimm: Unser Problem mit den 1.000 Patienten war: Ich bin aus der Onkologie, und da ist es zum Beispiel fast unmöglich, Studien aufzusetzen, die pro Arm 1.000 Patienten

haben. Es gibt ja auch noch andere Indikationen außer der Onkologie, wo das auch eher unmöglich wäre. Dann wäre es einfach nicht möglich, einen Beleg für diese Indikation herzuleiten. Das ist unser Problem. Das müsste man vielleicht relativieren oder absenken, je nachdem, wie die Möglichkeit ist, Patienten zu rekrutieren.

Moderator Jürgen Windeler: Ich möchte noch mal daran erinnern, dass wir über eine Ausnahmesituation sprechen. Wenn es nicht möglich ist, aus einer Studie, die aus welchen Gründen auch immer nur 200 Patienten umfasst, einen Beleg abzuleiten, dann ist das so. Dann werden wir diese Unsicherheit beibehalten, möglicherweise immer beibehalten, weil eine zweite Studie, die mehr Patienten umfasst und sie irgendein Kriterium reißt, einfach nicht da ist. Dann wird man mit der Unsicherheit einfach leben müssen. Ich glaube, dann werden alle damit leben müssen, dass es einen Beleg für diese Therapie nicht gibt. Das ist der eine Punkt. Noch mal: Wir reden über Ausnahmesituationen. Das ist mir wichtig.

Der zweite Punkt ist: Ich habe noch nicht verstanden – die Frage geht an Herrn Hennig oder auch an Ulrich Mansmann -: Wenn wir jetzt keine apodiktische oder klare Grenze schreiben, wo soll denn dann die Grenze sein? Wir können natürlich reinschreiben: Es hängt von der Situation ab. - Ich warte schon auf die nächsten Stellungnahmen, die sagen: Könnt ihr das nicht ein bisschen spezifizieren? Wo müsste denn zum Beispiel eine Untergrenze liegen?

Ulrich Mansmann: Ich denke, man kann das Wort „Orientierung“ hineinbringen. Dann ist eigentlich eine Richtung gesetzt und sind Möglichkeiten offen. Diese spezifische Beantwortung dieser Frage bedeutet viel Arbeit. Ich hoffe, dass wir das in der GMDS mit unseren Wissenschaftlern im Laufe des nächsten Jahres durch Simulationsstudien leisten können.

Stefan Lange: Das finde ich prima. Wir würden gerne dieses Angebot aufnehmen und darauf zurückkommen.

Ich habe eine Rückfrage an Herrn Knoerzer. Ich muss ganz offen gestehen, dass ich nicht verstehe, was Sie gerade gesagt haben. Sie sagen, es sei ethisch nicht vertretbar, eine Studie dahingehend zu konzipieren, dass man aus ihr sichere Ergebnisse ableiten kann. Das haben Sie quasi gerade gesagt. Das finde ich, ehrlich gesagt, bemerkenswert. Vielleicht können Sie das einmal präzisieren.

Dietrich Knoerzer: So habe ich es natürlich nicht gemeint. Das wäre ja fast schon böswillig interpretiert. Was ich versucht habe, zu sagen, ist, dass diese Studien angelegt werden. Die werden ja mit Ethikbehörden und mit Zulassungsbehörden besprochen - das ist ja kein Geheimnis -, indem man für primäre Fragestellungen – deswegen hat man ja eine Hierarchisierung der Fragestellungen - versucht, einen schlüssigen Nachweis zu führen. An dem orientiert sich die Fallzahlplanung. Das finde ich im Grunde alles sehr schlüssig. Üblicherweise sind es eben nicht die Konsistenzkriterien, weil die natürlich dazu führen würden, dass sich die Fallzahlen für den primären Vergleich massiv erhöhen würden. Dann wiederum haben Sie sehr wohl unter Umständen die Situation, dass die Leute sagen: Jetzt exponieren wir aber ein bisschen viele Patienten ohne Not, ohne diesen schlüssigen Nachweis bereits zu haben, um weiter voranzuschreiten. Darüber können wir zwei uns jetzt ewig unterhalten, das ist aber die Realität, mit der wir in Unternehmen konfrontiert sind.

Moderator Jürgen Windeler: Noch mal die Rückmeldung: Ich kann das Argument, Studien so zu planen, nachvollziehen, nur ich würde es nicht als ethisches Argument betrachten. Denn, wie Herr Lange schon gesagt hat, das ethische Argument an der Stelle bedeutet, mit der Fallzahlplanung fokussiert auf einen primären Endpunkt in Kauf zu nehmen, dass für viele andere Endpunkte und für viele andere Fragen, die für die Patienten ethisch von großer Bedeutung wären, die Power nicht ausreicht. Das ist ja die Konsequenz. Ich kann aus vielen Gründen verstehen, dass man Studien so plant, aber aus Patientenschutzgründen und damit aus ethischen Gründen ist das, glaube ich, ein schwaches Argument.

Dietrich Knoerzer: Was ich mit „ethisches Argument“ gemeint habe, ist, dass Sie mit diesen Studien natürlich bei einer Ethikkommission vorsprechen. Das hatte ich damit gemeint. Ich hatte nicht das ethische Argument vorgeschoben, sondern ich habe gemeint, Sie sprechen bei einer Ethikkommission vor, und da werden entsprechende Kommentare kommen.

Moderator Jürgen Windeler: Die Frage würde also eher an die Ethikkommission gehen als an Sie. Sie haben recht.

Wir würden jetzt im Moment die Frage als orientierend, als beantwortet, jedenfalls nicht als geklärt - das Angebot der Fachgesellschaften nehmen wir gerne auf, weiter zur Klärung beizutragen, ansonsten noch mal über die Formulierung nachzudenken – ansehen, wobei man natürlich überlegen muss, dass Kriterien entweder Kriterien oder wirklich gute Orientierung sein müssen oder besser nicht aufgeschrieben werden sollten, was ja bedeuten kann, dass wir auch schreiben: Beleg aus einer Studie ist ausnahmsweise möglich. - Orientierung muss schon gewisse Grenzen haben. Eine Orientierung, es soll angemessen gemacht werden, hilft, glaube ich, im Methodenpapier speziell Niemandem weiter und würde vermutlich auch allen Ihren Erwartungen nicht entsprechen.

Tim Friede: Guido Skipka hatte ja schon in seiner Einleitung zu dieser Problematik weitere Informationen gegeben, also eine Rationale erläutert, die wir ja so nicht kannten. Ich kann mir schon vorstellen, dass es dann, wenn man hier weitere Details aufnimmt oder es etwas unterfüttert und dann klar ist, dass vielleicht in anderen Zusammenhängen eine Gleichwertigkeit vielleicht anders erreicht werden kann, glaube ich, vielleicht einen Ansatz für eine Lösung wäre.

Moderator Jürgen Windeler: Gut. Gibt es weitere Wortmeldungen? – Das sehe ich auch nach sorgfältigem Gucken nicht.

Wir kommen zum

Tagesordnungspunkt 4:

Metaanalysen

Ralf Bender: Ich komme ebenfalls aus dem Bereich Biometrie im IQWiG: Wir haben ja in dem neuen Entwurf für den Abschnitt 9.3.8 zu Metaanalysen überarbeitet. Das beinhaltet mehrere Aspekte. Einmal haben wir das erforderliche Signifikanzniveau für Heterogenitätstests angepasst. Dieser Punkt kommt gleich beim TOP „Subgruppen“ noch mal. Dann haben wir regelhaft einen Wechsel von der DerSimonian-Laird-Methode zu Knapp-

Hartung vollzogen. Es geht um die Zahl von Studien, die man braucht, um Random-Effects-Metaanalysen durchzuführen. Und es geht um die Wahl der Methode bei sehr wenigen Studien.

Generell wurden diese Änderungen von der großen Mehrheit der Stellungnahmen begrüßt. Es gab aber auch einige Hinterfragungen. Ich habe zwei Fragen. Die erste ist – das taucht sowohl in der gemeinsamen Stellungnahme von GMDS und IBS-DR als auch in der von GSK auf –: Da wird gefragt, ob wir eine systematische Literaturrecherche für diese Änderungen gemacht hätten. Meine Frage ist, ob die Stellungnehmenden es für erforderlich halten, eine systematische Literaturrecherche vorher zu machen und zu dokumentieren, um von DerSimonian-Laird-Methode auf Knapp-Hartung zu wechseln.

Michael Hennig: Sie haben uns direkt angesprochen. Grundsätzlich begrüßen wir es sehr, dass methodische Anpassungen stattfinden. Wir wollen auch die Knapp-Hartung-Methode als solche überhaupt nicht infrage stellen. Die Frage, die wir in unserer Stellungnahme geschrieben haben, ging eher in die grundsätzliche Richtung, auf welcher Basis denn methodische Anpassungen vom IQWiG vorgenommen werden. Wir halten sehr wohl eine gewisse Systematik, die es sicherlich in Ihrem Hause auch geben wird, für geboten, um sicherzustellen, dass neuere Entwicklungen entsprechend berücksichtigt werden.

Ralf Bender: Das kann ich beantworten: Wir versuchen natürlich, ganz allgemein up to date zu bleiben, was die wissenschaftliche Methodik betrifft. Die Biometrie erstellt jedes Jahr einen sogenannten Literaturscreeningbericht, worin wichtige neue Methoden enthalten sind. Auf dieser Basis glauben wir, dass wir bei der Methodik up to date sind. Und sobald wir etwas finden, was wichtig ist, findet es Eingang in die Methodik. Gerade dieser Wechsel von DerSimonian-Laird nach Knapp-Hartung ist längst überfällig. Das ist eine Konsequenz davon.

Ulrich Mansmann: Sie haben in Ihrem Methodenpapier gesagt, dass Sie die Bayes'schen Verfahren mittlerweile als eine Methodik ansehen, die in Ihren Argumenten angebracht werden kann. In den Bayes'schen Verfahren sind eigentlich die Art und Weise, Metaanalysen mit wenigen Studien zu machen, straight forward. Deshalb frage ich mich, ob man in diesem Punkt nicht so etwas hineinschreiben könnte, dass Bayes'sche Verfahren da eine Möglichkeit bieten, sowohl in Zufalls- wie auch in Fixed-Effect-Aspekten.

Andree Ehlert: Ich möchte auch auf das Argument mit der systematischen Literaturrecherche zurückkommen. Uns interessiert gar nicht so sehr die Begründung dafür, dass Sie jetzt zum Beispiel das Niveau 5 % gewählt haben, sondern ganz einfach die Begründung für den Wechsel, und zwar aus dem Argument, die alten Nutzenbewertungen zu schützen. Dies kann ja schon einen Bruch in der Methodik darstellen und möglicherweise erhebliche Konsequenzen haben. Hat man das einmal durchgespielt und untersucht?

Ralf Bender: Ja, haben wir gemacht. Die Konsequenz war genau nicht so, wie Sie beschrieben haben, sodass wir da eigentlich ein gutes Gefühl hatten, die Sachen zu vereinfachen und das Niveau hier anzupassen. Es gibt erstaunlich wenige Auswirkungen.

Moderator Jürgen Windeler: So gut wie gar keine, um es etwas pointierter zu sagen.

Andree Ehlert: Dann wäre es doch schon gelöst und super, wenn Sie dazu noch einen Halbsatz aufnehmen würden.

Guido Skipka: Zur Empirie kann ich noch ein, zwei Sätze sagen. Wir haben uns in der Tat alle Metaanalysen, die wir in unseren Berichten in den Jahren 2014 und 2015 gemacht haben, noch einmal angesehen, haben uns die Metaanalysen, wo wir eine Heterogenität im Bereich p-Wert 5 bis 20 % haben - das war, wo wir bis dato nicht gepoolt haben in der Metaanalyse -, noch mal extra angeschaut. Das war so in 20 % der Metaanalysen der Fall. Wenn man jetzt durch Absenkung der Schwelle auf 5 % doch gepoolt hätte und hätte dann geguckt, was sich am Fazit ändert, dann gab es, glaube ich, von insgesamt 150 Metaanalysen zwei Metaanalysen, wo es zu einer leichten Änderung kam. Einmal wurde bei einem Dossier von einem nicht quantifizierbaren Zusatznutzen auf einen beträchtlichen Zusatznutzen angehoben. In einem anderen Fall wurde ein Anhaltspunkt weggenommen. Diese Auswirkungen hielten wir für äußerst gering.

Ralf Bender: Zu dem Hinweis auf Bayes'sche Methoden: Dass wir Bayes'sche Methoden grundsätzlich als Option ansehen, steht ja im Methodenpapier. Dass das allerdings so ganz straight forward ist, kann ich nicht so ganz nachvollziehen. Man muss ja immerhin dann noch A-priori-Verteilungen angeben, und die muss man erst mal haben. Gerade in den problematischen Fällen, wo man noch argumentiert, man braucht möglicherweise A-priori-Verteilungen mit gewissen Vorinformationen, die nicht uninformativ sind, dann ist das alles andere straight forward. Dann muss man sich nämlich erst einmal einigen und einen Konsens herstellen, was die richtige A-priori-Verteilung ist. Also so ganz straight forward ist das meiner Ansicht nach nicht.

Ulrich Mansmann: Zu den A-priori-Geschichten: Wenn ich eine andere Analyse mache, mache ich auch A-priori-Verteilungen. Wir nehmen die natürlich irgendwie als gottgegeben hin, aber auch als Frequentist habe ich meine A-Prioris laufen, mit denen ich arbeite, nur die sind eben gottgegeben.

Stefan Lange: Trotzdem noch einmal eine Anregung an Ulrich Mansmann und Tim Friede: Es wäre natürlich hilfreich – wir haben an anderer Stelle schon mal darüber gesprochen –, wenn man sich – in Anführungsstrichen – ein Stück weit einigen könnte auf bestimmte Basisparameter, was vernünftiger A-priori-Verteilungen sein könnten. Das würde möglicherweise solche Bayes'schen Analysen durchaus attraktiver und nachvollziehbarer machen. Vielleicht kann man sich in der Gesellschaft ein bisschen engagieren.

Tim Friede: Ich möchte noch einmal betonen, dass wir als Fachgesellschaft begrüßt haben, dass das IQWiG die Initiative übernommen und erkannt hat, dass es hier ein methodisches Problem gibt, das man angehen muss. Wir beteiligen uns natürlich gerne an der Lösungsfindung. Im Moment gehen, was technische Details angeht, ein bisschen in die Meinungen auseinander, was straight forward ist und was man da sinnvoll als A-priori-Verteilung für bestimmte Parameter annimmt. Aber ich bin eigentlich guter Dinge, dass man da zu einem Konsens kommen kann, an dem wir uns gerne beteiligen.

Das Grundproblem ist ja, dass wir bei den Metaanalysen, die hier im IQWiG durchgeführt werden, in der Regel keine fünf Studien haben und dass hier sozusagen das Knapp-Hartung-Verfahren, was jetzt als Regel für Metaanalysen mit mehr als fünf Studien angesetzt wird, in

der Praxis nicht groß etwas ändern wird, weil eben ganz viele Metaanalysen mit zwei, drei, vier Studien durchgeführt werden, und das in einer Situation, wo man einfach von der Statistik her mit Standardverfahren die Heterogenität, um die es ja zwischen den Studien geht, so nicht ohne Weiteres schätzen kann, und als weiteres Problem eben hat, dass man in den Verfahren, die man verwendet, diese Unsicherheit in der Schätzung der Heterogenität mit einbeziehen muss. Ich denke, da gibt es, wie es schon gesagt wurde, natürlich Verfahren, die das im Prinzip leisten können. Da gibt es bestimmte technische Details, die wir da weiter erörtern müssen und wo wir dann vielleicht in so eine Guideline Empfehlungen, in so ein Methodenpapier Empfehlung für solche Auswertungen mit einbeziehen kann. Das unterstützen wir natürlich gerne.

Wir sehen diesen Vorschlag vom IQWiG als einen ersten Schritt in die richtige Richtung, ich glaube, weg von Heterogenitätstests, weil in diesen Situationen, die für das IQWiG wichtig sind mit zwei, drei, vier Studien, bringen die absolut nichts, keinen Informationsgewinn, sondern man muss sehen, dass eine Heterogenität zwischen Studien irgendwo was Natürliches ist. Die muss einbezogen werden in adäquater Weise in die Auswertung. Ich glaube, es geht eigentlich nur um das Feintuning bestimmter Methoden, die im Prinzip verfügbar sind. Ich bin auch da guter Dinge, dass wir da vorankommen.

Ralf Bender: Das leitet perfekt über zu meiner zweiten Frage. Offen geblieben ist ja dann im Methodenpapier, wie man es genau macht, wenn eben weniger als fünf Studien da sind. Darauf wird auch in der gemeinsamen Stellungnahme von GMDS und IBS-DR hingewiesen. Es wird zum Teil aber auch gefordert, eine klare Methodik anzugeben. Das durchzieht sich praktisch durch alle Stellungnahmen, die vonseiten der Industrie kommen. Meine Frage an die Stellungnehmenden ist: Gibt es denn zurzeit eine klare Standardmethode für diesen schwierigen Fall mit sehr wenigen Studien? Aus unserer Sicht eben nicht. Deswegen haben wir auch keine klare, universal anwendbare Methode da hineingeschrieben.

Friedhelm Leverkus: Ich gebe zu bedenken, dass wir in dem Setting von der Zulassung häufig solche Replikationen haben, die gefordert worden sind. Und wenn wir solche Replikation haben, sind Modelle wie Random-Effects-Modelle, da wo ich Heterogenität halt schätzen will, eher fraglich, sondern da habe ich im Prinzip eigentlich nicht zwei Studien, die ich nach dem Fixed-Effects-Modell halt poolen kann. Von daher wäre von dem Modell her, was ich habe, also von dem Modell, wie die Daten generiert werden, von der Voreinstellung ein Fixed-Effects-Modell vorhanden. Dann gibt es natürlich noch Situationen, wo ich wirklich Heterogenität habe, wo ich unterschiedliche Designs habe, wo solche Methoden, die Tim Friede vorgeschlagen hat, bedenkenswert wären und wo man vielleicht schauen muss, inwieweit das für solche Situationen angebracht wäre.

Guido Skipka: Das mit den Fixed-Effects-Modellen kann ich nachvollziehen für den Fall, den du angesprochen hast, dass man sozusagen Zwillingstudien plant und durchführt, die sehr ähnlich designed und durchgeführt sind. Da würde ich in der Tat auch sagen, da ist das Fixed-Effects-Modell völlig unkritisch. Nur, das ist nicht der typische Fall bei uns. Das ist doch eher selten. Wir sehen auch bei zwei, drei Studien, dass die Ergebnisse zum Teil deutlich divergieren. Da täte ich mich schwer, einfach ein Fixed-Effects-Modell laufen zu lassen.

Friedhelm Leverkus: Auch beim Fixed-Effects-Modell können Ergebnisse differieren und schwanken. Die Frage ist nur: Wie unterschiedlich ist designed? Ich gebe dir recht, Guido, wenn man jetzt unterschiedliche Indikationen hat, wo man die Replikation macht, dann ist sicherlich fraglich, ob man da den Fixed Effects anwenden kann. Die Frage danach, wie gehe ich in solchen Situationen um, finde ich berechtigt.

Tim Friede: Ich möchte kurz Guido Skipka unterstützen in seiner Einschätzung, was die Wahl der Modelle Fixed-Effects gegen Random-Effects angeht. Ich würde auch sagen, es muss ganz klare Gründe geben, mit denen man dieser Fixed-Effects-Annahme begründen kann. Ansonsten würde ich eigentlich regelhaft – das sage ich als Wissenschaftler – davon ausgehen, dass es eine Heterogenität gibt, und ein Verfahren anwenden, was das im Prinzip berücksichtigen kann.

Ich - vielleicht spreche ich jetzt mehr als Wissenschaftler und nicht als Präsident meiner Fachgesellschaft - würde sagen: Ja, wir haben Lösungen für die Random-Effects-Metaanalyse mit wenigen Studien. Die haben wir in den letzten Monaten dazu publiziert. Da kann man sicherlich jetzt nicht erwarten, dass das jetzt schon im Methodenpapier regelhaft integriert ist. Aber wir halten die Bayes'sche Metaanalyse für eine Möglichkeit, für einen Weg vorwärts in diesen Situationen. Und wir haben ja auch ganz konkrete Vorschläge gemacht, was die A-priori-Verteilung für den Heterogenitätsparameter angeht. Um den geht es ja nur. Ich denke, dass wir da im Dialog weiterkommen. Ich würde mir wünschen, dass man spätestens in einer weiteren Version von diesen Methodenpapier, wenn nicht sogar in der Revision jetzt hier, konkrete Dinge in dem Bezug auch noch aufnehmen kann.

Ulrich Mansmann: Als Ergänzung: Wenn ich diese Bayes'sche Analyse mache, verstehe ich schon den Zusammenhang zwischen Heterogenitätsannahmen und Effektschätzungen. Und ich kann dann auch prüfen, vielleicht aus sachlichen Gründen, welche Heterogenitätsannahmen plausibel sind oder nicht plausibel sind. Das ist ein Punkt, der ist vielleicht in der IQWiG-Logik nicht stringent genug, aber ich glaube schon, dass man da zu einem Schluss kommen kann, der mit einer Evidenz getragen wird.

Ch.-Markos Dintsios: Die Frage nach der Methode und der Verortung im Methodenpapier hat ja für uns als Industrie mit Planungssicherheit zu tun. Wir kommen ja weniger aus der wissenschaftlichen Schiene, sondern eher aus der pragmatischen Umsetzung der Erfordernisse, die im Methodenpapier des IQWiG an uns gestellt werden. Deswegen zurück ans IQWiG. Wir haben es ja schon ein paar Mal erlebt, dass Methoden bei der Imputation fehlender Daten - also schon angefangen vor einigen Jahren –, die dann angewendet wurden, vom IQWiG kritisiert wurden ob ihrer Nachteile. Da haben wir uns mal geeinigt, dass alle diese Methoden Vor- und Nachteile haben. Letztendlich sind wir dann so auseinandergeschieden. Bis heute wissen wir nicht, wie man es festlegen soll. Die Frage geht auch in diesem Fall zurück an das IQWiG: Könnte man denn zumindest dann - ich verwende den Begriff, den Stefan Lange nicht gerne hört – Methodenpluralismus tolerieren im Sinne von: „Wir machen das und wir machen auch etwas anderes und schauen, was rauskommt, aber zumindest haben wir eine Orientierung als Industrie“?

Andree Ehlert: Ich unterstütze voll und ganz den kritischen Umgang und die kritische Diskussion der fixen Effekte bzw. der Modelle, möchte einfach noch mal darauf hinweisen,

dass in der jetzigen Formulierung das aber ganz gewiss zu Missverständnissen führen kann, weil ich sehe, dass der unbedarfte Anwender möglicherweise hier eine Hintertür für die regelhafte Anwendung von fixen Effekten rausliest, weil einfach diese Grenze mit den fünf Studien da drinsteht.

Ralf Bender: Das habe ich überhaupt nicht verstanden.

Andree Ehlert: Es steht jetzt der Satz drin:

„Gerade im Fall sehr weniger Studien sollte daher die Anwendung des Modells mit festem Effekt ... in Erwägung gezogen werden.“

Ich könnte mir vorstellen, dass das sehr großzügig angenommen wird. Nur als Hinweis.

Ralf Bender: Das wird von uns nicht großzügig angenommen, sondern es ist so gemeint, wie es da steht. Wir gucken, ob ein Modell mit festem Effekt hier in einer Situation gerechtfertigt ist. Den anderen Stellungnehmenden war diese Formulierung ja eher zu schwach. Es war ja der Vorschlag von Herrn Leverkus, dass er gesagt hat, er hätte das gerne als Standardmodell von Anfang an. Das ist ja viel stärker, als das jetzt da steht.

Moderator Jürgen Windeler: Jetzt steht, um das Wort aufzugreifen, ein gewisser Methodenpluralismus da drin.

Ralf Bender: Genau das wäre auch meine Antwort zu Marcos gewesen. So, wie du das formuliert hast, steht es im Moment genau da drin. Wir beschreiben einige Methoden, die für diesen schwierigen Fall infrage kommen, weil es zurzeit keine klare Standardmethode gibt. Gut, es ist etwas in der Entwicklung. Wir sind ja auch da im guten Kontakt. Vielleicht als Information für alle anderen: Es gibt eine Art Expertenrunde von Metaanalysen, woran auch Tim Friede beteiligt ist. Wir schreiben auch ein gemeinsames Paper. Vielleicht kommen wir in Zukunft da weiter; das muss man sehen. Im Moment sehen wir jedoch keine klare, eindeutig beste Standardmethodik für den Fall von sehr wenigen Studien. Genau das ist im Moment im Methodenpapier abgebildet.

Stefan Lange: Ich habe noch eine Frage an Tim Friede und Ulrich Mansmann. Ich habe den Eindruck, dass sich eure beiden Aussagen ein Stück weit widersprochen haben. Ich habe Tim Friede so verstanden, dass er sich vorstellen kann und es auch publiziert sei, dass wir bestimmte Standardvorgaben für A-priori-Verteilungen zum Beispiel vom τ^2 festschreiben können sollten - das würde die Sache sehr vereinfachen -, Ulrich hat aber, glaube ich, eben gesagt, man sollte es eher von Fall zu Fall entscheiden. Oder habe ich das falsch verstanden?

Ulrich Mansmann: Ich habe gesagt, dass im Rahmen dieser Bayes'schen Analyse die Bedingungen an die Verteilung von τ^2 modifiziert werden können. Ich kann dann auch verstehen, wie Effektschätzungen und Heterogenitätsannahmen miteinander korrespondieren. Ich kann verstehen, dass A-priori-Annahmen die Effektschätzung ausmachen. Und ich kann dann auch wiederum argumentativ aus dem Studiendesign vorgehen und sagen, in welche Richtung Heterogenität plausibel sein wird oder nicht. Aber lassen wir das. Das ist zu kompliziert.

Sandra Janatzek: Ich möchte gerne auf die Absenkung des Signifikanzniveaus zurückkommen für die Heterogenitätstests. Passt das jetzt, oder erst diesen Strang zu Ende führen?

Moderator Jürgen Windeler: Bitte.

Sandra Janatzek: Ich finde es sehr erfreulich, dass Sie diese Empirie betrieben haben. In der Tat haben wir uns genau die Frage gestellt: Hätte es denn in den vergangenen Berichten zu Änderungen geführt? Grundsätzlich kann man sich ja schon Änderungen vorstellen in der qualitativen Ergebnissicherheit. Das ist klar. Dadurch, dass jetzt häufiger metaanalytische Auswertungen gemacht werden, wird man vielleicht grundsätzlich häufiger zu präziseren Schätzungen kommen können. Es können sich aber auch – das hat ja auch das Beispiel von dir, Guido, gezeigt – Änderungen in der Aussagesicherheit ergeben. Ganz klar: Der Algorithmus zur Generierung der Aussagesicherheit – Beleg, Hinweis, Anhaltspunkt – sieht ja eine Differenzierung zwischen Situationen mit homogenen und heterogenen Ergebnissen vor. Beispielsweise könnte man sich ja in der Situation grundsätzlich vorstellen, auch wenn das jetzt in 2014 und 2015 nicht passiert ist, dass etwa in der Situation, in der wir nur eine niedrigverzerrte Studie mit einem signifikanten Vorteil haben, das ja zunächst nur zu einem Hinweis führen würde, aber – so sieht es ja Ihr Methodenpapier vor - eine Replikation quasi durch dann noch vorhandene hochverzerrte Studien gegeben sein kann. Dann – das meine schon – kann es einen Unterschied machen, ob ich eine metaanalytische Zusammenfassung aller Studien in der Situation machen kann und darüber vielleicht zu einem signifikanten Effekt kommen kann oder ob ich sie nicht machen kann. Mit der Absenkung des Signifikanzniveaus jetzt beim Heterogenitätstest würde man ja erwarten, dass grundsätzlich häufiger solche metaanalytischen Zusammenfassungen gemacht werden. Das ist das eine. Ich schätze also sehr, dass Sie diese Empirie betrieben haben, aber ich tue mich trotzdem etwas schwer damit, weil natürlich übermorgen der Fall kommen könnte, in dem es doch anders aussieht. Wenn Sie dazu noch etwas sagen könnten, wäre das für uns sehr hilfreich.

Zum anderen habe ich eine Verständnisfrage. Ich verstehe das neue Vorgehen so, dass Sie auch die Suche nach potentiellen Ursachen von Heterogenität erst dann einsetzen, wenn der p-Wert des Heterogenitätstest unter der neuen 5%-Schwelle liegt. Vergibt man sich da nicht etwas an inhaltlicher Beschäftigung mit der Fragestellung, wenn man sehr viel zurückhaltender damit einsetzt, über Effektmodifikationen nachzudenken, über spezifische Aussagen nachzudenken, die vielleicht abzuleiten wären und der Sache vielleicht angemessener wären, wenn Sie erst sehr viel später mit diesem Prozess einsetzen? Das kann ich mir noch nicht ganz erklären. Ich will noch einmal grundsätzlich dafür werben: Wir wollen ja nicht nur formale Algorithmen abarbeiten, sondern die sollen dazu dienen, dass wir sachgerechte und dem Inhalt angemessene Aussagen schaffen können.

Moderator Jürgen Windeler: In der Tat: Das ist das Ziel.

Guido Skipka: Das waren jetzt viele Aspekte. Du hast gesagt, wir haben jetzt eine gewisse Änderung der Methodik, indem wir die Schwelle abgesetzt haben, was im Einzelfall natürlich zu einer ganz gravierenden Änderung der Aussagen führen kann. Ich denke, das ist richtig, kann uns passieren, ist aber mit jeder Änderung der Operationalisierung irgendwie inhärent.

Du hattest da – das war mir nicht ganz klar – die Aussagesicherheit mit hineingebracht. Das habe ich noch nicht ganz verstanden. Erst einmal: Diese Schwelle ist für uns relevant gewesen für die Frage: Betrachten wir das gepoolte Ergebnis und leiten daraus ab, oder schauen wir uns die Einzelergebnisse sozusagen als Punktwolke an? Das hat mit Ergebnissicherheit erst einmal nichts zu tun. Das kommt als zweiter Aspekt hinzu. Was man natürlich sagen kann: Nur weil wir Ergebnisse nicht gepoolt haben, heißt es ja nicht, dass wir nichts ableiten. Wir haben ja das Konzept der gleichgerichteten Effekte. Das steht ja auch im Methodenpapier.

Jetzt habe ich den zweiten Punkt vergessen.

Moderator Jürgen Windeler: Das war die Frage, ob wir nicht Dinge übersehen, wenn wir den Level hochsetzen.

Guido Skipka: Es ist vollkommen klar: Durch Absenkung der Grenze von 0,2 auf 0,05 schauen wir jetzt häufiger auf gepoolte Ergebnisse. Wir gestatten uns eine etwas größere Heterogenität in den Daten und ignorieren die dann ein Stück weit, weil wir uns auf das gepoolte Ergebnis beschränken. Unsere Erfahrung in den letzten Jahren war halt, durch die Schwelle 0,2 hatten wir sehr viele Situationen, wo wir eben nicht gepoolt haben, obwohl vom Gefühl her oder optisch gesehen im Forest Plot, es gar nicht so schlimm aussah. Wir hatten also aus unserer Sicht viele falschpositive Signale. Das wollten wir jetzt dadurch etwas reduzieren, dass wir sagen: Wir gestatten uns jetzt etwas mehr Heterogenität. Dadurch machen wir uns das Leben vielleicht etwas einfacher und klarer.

Ralf Bender: Zur Ergänzung: Wir diskutieren ja unsere Ergebnisse auf wissenschaftlichen Kongressen und haben einfach die Erfahrung gemacht, dass wir an sehr vielen Beispielen, wo wir gesagt haben, aufgrund der Grenze 0,2 haben wir nicht gepoolt ... Es kamen dann ganz viele Leute zu mir, zum Beispiel Herr Schumacher aus Freiburg, der sagte, er hätte in der Situation völlig problemlos gepoolt. Er meint, wir wären da zu streng gewesen. Okay, dann senken wir das jetzt ab.

Sandra Janatzek: Ich stimme da vollkommen zu. Auch uns ging es beim Lesen Ihrer Berichte immer mal wieder so, dass wir gesagt haben, hier ist zwar das formale Heterogenitätskriterium zum 20%-Niveau erfüllt im Sinne von „es liegt Heterogenität vor“, aber eigentlich sehen wir auch die Studien als metaanalytisch kombinierbar an. Da stimmen wir vollkommen zu. Uns geht es aber darum, dass es eben auch andere Situationen geben kann und potentiell auch geben wird. Deshalb hatten wir in unserer Stellungnahme nachgefragt, ob auch ein Weg denkbar wäre, der eine gewisse Grauzone zulässt, also dass man eben den Einwand von Wissenschaftlern, bei denen Sie jetzt Ihre Ergebnisse vorgestellt haben, auffängt, indem man sagt, metaanalytische Zusammenfassungen machen wir ab einem Signifikanzniveau von 5 % des Heterogenitätstests, aber wir sehen sie grundsätzlich als verwertbar an oberhalb eines p-Wertes von 20 %, und in den restlichen Fällen entscheiden wir das inhaltlich, sprich mit Blick auf den Konfidenzintervallplot. Das lässt natürlich weniger greifbares Vorgehen zu, mehr Interpretationsspielraum, aber vielleicht ist der auch der Sache angemessen. Das wäre unsere Frage. Wäre das nicht auch eine Möglichkeit, mit den Einwendungen der Wissenschaftler bei den Kongressen umzugehen?

Dirk Eyding: Ergänzend dazu: Es ist ja ein Unterschied, ob man viele oder wenige Studien hat. Die Frage, die sich uns dann noch gestellt hat, ist, ob man nicht doch wieder zu dem

deskriptiven Heterogenitätsmaß zurückkommt, so etwas wie i^2 oder so, weil die Fälle, die gerade beschrieben worden sind, bei einem p von kleiner 0,2, bei vielen Studien dann doch eine Metaanalyse durchführen könnte, ist ja vielleicht mit einem i^2 verbunden, das nicht so kritisch anzusehen ist, wo man konventionellerweise, was ja auch empfohlen wird, auch durch Cochrane, dann auch eine Metaanalyse hätte machen können, aber bei Fällen mit weniger Studien kann ja p kleiner 0,2 schon eine bedeutsame Heterogenität bedeuten, sodass das i^2 50 % oder mehr umfassen würde, sodass man sich da deutliche Unterschiede zwischen den Studien zusammenpoolt, obwohl da möglicherweise tatsächlich Unterschiede existieren. Insofern ist das vielleicht ein Punkt, worüber man nachdenken sollte.

Stefan Lange: Als allererstes vielleicht: TOP 4 und TOP 5 gehören ein Stück weit zusammen. Insofern bin ich auch mit dem Wort „Absenken“ nicht so ganz glücklich. Wir haben immer schon bei den 5 % gesagt: Das ist für uns ein klares Signal. Da ist Heterogenität. - Die 20 % waren für uns immer so ein bisschen: Naja, da könnte was sein. - Wir bleiben ja bei den 5 %.

Sandra Janatzek: Das Absenken bezog sich auf die Metaanalyse.

Stefan Lange: Ist schon klar. Irgendwie gehört das ja zusammen, jedenfalls nach meinem Verständnis. Das ist das eine.

Das Zweite ist: Der eine Grund, insbesondere was die Metaanalysen angeht – das hat gerade auch Ralf gesagt –, ist die häufige Situation, die man antrifft, dass auch eine Heterogenitätsstatistik von 10 % oder 12 %, also vom p -Wert her, nicht so furchtbar dramatisch aussieht.

Ein anderer Grund ist auch ganz klar – Jürgen Windeler hat es in der Einleitung gesagt –: Wir müssen gucken, wie wir Ressourcen schonen. Wenn wir Heterogenitätsbetrachtungen machen – da gehören wiederum TOP 4 und TOP 5 ein Stück weit zusammen -, die hinterher keine Konsequenz haben, und zwar in 98 % der Fälle keine Konsequenz haben, dann ist es einfach ein Stück weit unnötiger Ressourcenverbrauch. Wir kriegen ständig vom Gesetzgeber und auch von anderen, die das alles unterstützen, gesagt: kürzere Fristen, kürzere Fristen, kürzere Fristen. Das hört irgendwann auf. Das geht irgendwann nicht mehr. Dafür ist das natürlich auch eine gewisse Unterstützung, einfach Ressourcen zu schonen.

Die Methoden, die im Methodenpapier stehen - das wissen alle –, beschreiben den Regelfall. Es gibt ja das Sprichwort, dass Ausnahmen die Regel bestätigen. Sonst gibt es eben keine Regel. Jetzt kann man das natürlich so machen, wie du, Sandra, es vorgeschlagen hast, und sagen, es gibt einen Graubereich. Ja klar, aber den gibt es sowieso. Ob wir den nun reinschreiben oder nicht, den gibt es. Wir sind natürlich im Spannungsfeld zwischen dem, was wir jetzt immer wieder diskutieren, entweder ist das IQWiG zu vage - dann können wir uns nicht daran orientieren, dann haben wir keine Planungssicherheit – oder es ist zu strikt - dann ist auch alles doof -.

Das ist eben unser Dilemma, was wir haben. Ich muss ganz offen gestehen, so, wie du deine Einführung begonnen hast: „Aber es könnte ja morgen sein, dass ...“ - Ja, sicher, aber das haben wir bei allem. Wir können auch mit den 5 % arbeiten, und hinterher stellen wir fest, eigentlich wären 4,9 % sachgerecht gewesen. Diesen einen Fall können wir auch nicht ausschließen. Das ist völlig klar. Aber ich glaube, wenn man schon Empirie hat, die einem

zeigt, in 98, 99 % der Fälle ist das eigentlich überflüssig - das geht in ähnlicher Weise bei TOP 5 weiter -, dann kann man das ja vielleicht auch ernst nehmen.

Moderator Jürgen Windeler: Ich möchte eine ergänzende Bemerkung machen. Du hast ja vorhin beschrieben, dass es natürlich jetzt passieren wird, dass wir weniger Anlässe schaffen, uns mit Heterogenität zu befassen.

Sandra Janatzek: Ich habe das gefragt.

Moderator Jürgen Windeler: Die Absenkung der Schwelle legt das ja nahe. Irgendwas muss ja dazwischen passieren, was wir jetzt nicht mehr genauso betrachten, wie wir es bisher betrachtet haben. Das - du hast es ja auch so angedeutet – kann man ja erst mal als einen Mangel verstehen. Es ist sogar ein Mangel. Dem steht aber ein massiver Vorteil entgegen, indem wir aufhören, uns mit uninteressanten oder vielleicht auch nur zufälligen Signalen zu beschäftigen und zu versuchen, die aufzuklären. Nach unserer Auffassung ist es sehr gut vertretbar, die Beschäftigung mit diesen spurious Signalen sozusagen herunterzufahren, damit auch zu riskieren, dass wir mit echter Heterogenität, jedenfalls wenn sie sich nur im Signifikanzniveau abbildet, ein bisschen restriktiver sind, unter empirischen Gesichtspunkten vertretbar, unter Ressourcengesichtspunkten umso mehr vertretbar. Ich darf vielleicht auch darauf aufmerksam machen, dass der G-BA bei den Signalen, die wir dann dem G-BA mitgeteilt haben auf den bisherigen Analysen, so gut wie nie eine Differenzierung nach Subgruppen gemacht hat, sodass wir uns auch da die Frage stellen müssen, ob das Sinn macht. Dieses alles zusammen liefert eine gute Basis, um das Niveau in dieser Weise zu verändern und dabei selbstverständlich – das darf ich an dieser Stelle allerdings auch sagen – nicht jegliche Beschäftigung mit inhaltlich naheliegender Heterogenität aufzugeben, nur weil der p-Wert oberhalb von 0,05 liegt.

Dirk Eydung: Ich wollte nur sagen: Konsequenzen sind Konsequenzen. Dass, was Sie angesprochen haben, der G-BA das häufig nicht umsetzt, ist sicher wahr. Andererseits ist es auch so, dass natürlich schon viele Dossiers dazu geführt haben, dass Subgruppeneffekte vom IQWiG nahegelegt worden sind, die der G-BA nicht umgesetzt hat. Das heißt, das gibt es schon.

Es stimmt auch, dass meistens irgendwelche Belege, also p kleiner 0,05 bei Interaktionstests für Subgruppenbildung oder auch bei Metaanalysen, da mit eingeflossen sind. Aber ich glaube, dass man die Fälle mit 0,2 dabei nicht unterschätzen darf. Man kommt auf eine Bewertung, ob das ein Zufallstreffer ist oder nicht. Darum kommt man ja auch bei einer 0,05-Regelung nicht herum. Man muss sich jetzt angucken, ob das wirklich ein realer Effekt ist. Wir haben ganz viele Tests, die wir da machen. Also da spielt jetzt auch schon das Subgruppenthema mit hinein. Es ist jetzt nicht nur das Metaanalysenthema.

Zur Beurteilung, ob das ein konsistenter Effekt ist, der sich bei dem Subgruppenmerkmal zum Beispiel an mehreren Endpunkten zeigt, und einmal ein p-Wert von 0,03 für den Interaktionstest aufgetreten ist und mal von 0,1, spielt das schon eine große Rolle, um tatsächlich zu einem Urteil zu kommen, das eine Subgruppenbewertung nach sich ziehen sollte. Dass der G-BA das nicht immer nachvollzogen hat, ändert daran aus meiner Sicht überhaupt nichts, dass das berechtigt ist. Das sollte man nicht als Maßstab nehmen.

Stefan Lange: Um es ganz deutlich zu sagen, nicht immer, sondern nie.

Dirk Eyding: Das stimmt nicht.

Stefan Lange: Da sind auch Vertreter sozusagen von eurer Muttergesellschaft dabei, die regelmäßig sagen: Nein, das interessiert uns nicht. - Vielleicht ist „nie“ auch falsch, aber ich sage: nahezu nie.

Moderator Jürgen Windeler: In Ausnahmefällen ist es umgesetzt und berücksichtigt worden.

Guido Skipka: Dirk, gerade bei Subgruppen ist es so: Wir sind verpflichtet, uns Subgruppen anzuschauen. Wir betrachten viele Endpunkte, viele Vergleiche. Das heißt, wir haben zunehmend mit sehr vielen Subgruppen zu tun. Die Grenze 0,2 bedeutet natürlich auch, dass wir sehr oft Signale sehen. Das macht uns das Leben extrem schwer, obwohl hinten dabei nichts herauskommt. Das darf man nicht unterschätzen, was diese ganze Subgruppenthematik für eine Arbeit ist.

Der andere Punkt ist: Du hast gesagt, wenn ich dich richtig verstanden habe, ob man sich nicht eher am I^2 orientiert anstatt am p-Wert. Ich meine, das kann man umrechnen. Sowohl der p-Wert für Heterogenität als auch das I^2 sind beides Funktionen der Q-Statistik und der Zahl der Studien. Das heißt, ich kann diese 0,05 sofort, wenn ich die Zahl der Studien kenne, umrechnen, was das für das I^2 bedeutet. Mathematisch ist das dasselbe.

Moderator Jürgen Windeler: Weitere Wortmeldungen sehe ich nicht. Wir bleiben ja in der Sphäre des Themas.

Wir kommen zum

Tagesordnungspunkt 5:

Subgruppenanalysen

Guido Skipka: Es bietet sich sogar inhaltlich an, mit diesem Thema weiterzumachen. Es geht im Abschnitt zu den Subgruppenanalysen konkret um die Frage: Wann betrachten wir eigentlich Subgruppenanalysen? Unser Vorschlag in den Methoden 5.0 ist, dass wir nur dann regelhaft in Subgruppenanalysen einsteigen, wenn wir mindestens zehn Ereignisse in einer der Subgruppen vorliegen haben, unabhängig davon, wie sich die Ereignisse auf die beiden Behandlungsgruppen verteilen oder wie viele Ereignisse in der anderen Subgruppe vorliegen.

Diese Eingrenzung, dieser Vorschlag wurde grundsätzlich positiv von Ihnen aufgenommen. So interpretiere ich jedenfalls die Stellungnahmen. Es wurde allerdings, wie immer, wenn wir eine Grenze festlegen, eine Begründung dafür gefordert. Es gab auch zwei konkrete Vorschläge, wie man das ändern könnte. Zum einen wurde vorgeschlagen, fünf Ereignisse in jeder Subgruppe und jedem Behandlungsarm zu fordern, also sozusagen erst einmal so eine Mindestereigniszahl in allen Armen, oder auch, dass in allen Subgruppen zehn Ereignisse vorliegen sollten. Wie schon beim Thema zuvor, ist uns aus der Literatur kein Kriterium bekannt, wann man in Subgruppen einsteigen sollte und wann nicht. Es ist so ähnlich wie mit

unseren 1.000 Patienten und den zehn Zentren. Wir halten zehn Ereignisse erst einmal für erforderlich, damit man eine nötige Präzision hat.

Unsere konkrete Frage an Sie lautet: Kennen Sie Arbeiten, an denen wir uns orientieren können, also wissenschaftliche Arbeiten, bzw. welche Kriterien würden Sie vorschlagen?

Désirée Fimm: Ich kann dazu etwas kurz erklären: Von uns kam der Vorschlag mit den jeweils fünf. Nach Ihrem Vorschlag wäre es ja auch möglich, dass wir in einem Arm elf Ereignisse hätten und in dem anderen zwei oder drei. Das fanden wir ein bisschen willkürlich gesetzt. Dann kann man sagen, dass die zwei oder drei auch nicht aussagekräftig sein könnten. Deswegen würden wir auf jeden Fall sagen, dass in beiden Armen mindestens fünf vorhanden sein sollten, was bei uns bei AstraZeneca global auch so gemacht wird. Die Frage werde halt einfach, ob es Sinn ergeben würde, mit unter zehn in einem Arm zu arbeiten? Können Sie das erklären?

Guido Skipka: Ich versuche es. Ich habe folgendes Beispiel im Kopf: Wir haben zwei Subgruppen, sagen wir mal Männer und Frauen. Bei den Frauen tritt das Ereignis, was wir uns anschauen, nicht oder so gut wie nicht auf, vielleicht ein, zwei, drei Ereignisse, bei den Männern haben wir vielleicht in der Interventionsgruppe zwölf,15 Ereignisse, in der Kontrollgruppe auch wieder keine bis sehr wenige. Da würde ich schon meinen, dass man da reinschauen sollte, was das jetzt begründet, warum bei den Männern unter Intervention auf einmal Ereignisse in einer gewissen Deutlichkeit auftreten, die sonst nirgendwo auftreten. Diesen Fall würden wir verpassen, wenn wir in allen Subgruppen Mindestereignisse fordern. Das ist unsere Begründung gewesen, warum wir für eine Subgruppe eine Mindestzahl fordern, aber nicht für alle.

Katrin Kupas: Das Beispiel ist mir völlig einleuchtend. Vielleicht kann man sich darauf einigen, dass man sagt, man guckt nur Subgruppen mit mindestens zehn Patienten an. Wir haben häufig bei Regionseffekten Gruppen, wo nur fünf, sechs Patienten drin sind. Da kann man eigentlich gar keine Aussage mehr treffen. Ich schlage vor, dass man es darauf bezieht und dann die zehn Ereignisse in einer der Subgruppen zusätzlich fordert.

Guido Skipka: Da muss ich nachfragen: Wenn wir zehn Patienten und zehn Ereignisse fordern, dann müsste jeder ein Ereignis haben. Oder habe ich Sie falsch verstanden?

Katrin Kupas: Nein, die erste Forderung wäre, dass alle Subgruppenkategorien mindestens zehn Patienten haben. Ich habe häufig gerade bei Regionen ganz viele Weiße, vielleicht zwei, drei Schwarze, fünf, sechs Asiaten. Daraus kann ich eigentlich keine Aussagen ziehen, und da stellen wir regelhaft keine Punktschätzer dar. Das Zweite wäre, um halt solche Fälle, wie Sie das dargestellt haben, abbilden zu können, zu sagen, in einer der Ausprägungen brauche ich diese zehn Ereignisse, aber halt nicht in allen.

Moderator Jürgen Windeler: Ich gebe zu Protokoll: Guido Skipka nickt. - „Verständnisvoll“ würde ich das nennen.

Ulrich Mansmann: Ich habe noch eine Bemerkung zur Symmetrie. Wenn in dem Protokoll Ereignis als das und das interpretiert ist, aber nur zwei Ereignisse auftreten, aber acht Nichtereignisse, dass man aufgrund der binären Endpunkte der Symmetrie das auch

umgekehrt sehen kann. Das steht jetzt nicht so im Text dezidiert drin. Das war unser Kommentar.

Guido Skipka: Den Kommentar habe ich gelesen. Das werden wir uns entsprechend überlegen.

Jürgen Windeler: Weitere Punkte, weitere Wortmeldungen sehe ich nicht. Doch, Herr Vervölgyi.

Volker Vervölgyi: Das betrifft einen anderen Punkt. Wenn wir mit diesem Thema fertig sind, dann möchte ich eine weitere Frage zum Thema „Subgruppenanalysen“ stellen. Es geht um die Verlässlichkeit der Aussagen aus Subgruppenanalysen. Wir zitieren im Methodenpapier die Arbeit von Sun et al. aus 2010 aus dem BMJ. Da sind verschiedene Kriterien genannt, wann man von Verlässlichkeit von Subgruppenanalysen ausgeht. Wir nennen da nur beispielhaft, dass die Subgruppenmerkmale a priori festgelegt worden sein sollten. In den Stellungnahmen insbesondere von GSK und von Roche wird auf die weiteren Kriterien referenziert, insbesondere dass man indirekte Evidenz mit einbeziehen sollte. Indirekte Evidenz wird hier beschrieben als „biologisch rationale Laboruntersuchungen“ und „Tierversuchen“. Das ist erst einmal aufgefallen und nicht ganz klar, wie der praktische Umgang sein soll. Wenn es so sein sollte – das ist ja heute schon mehrfach genannt worden, eine systematische Betrachtung des Ganzen –, dann müsste man eigentlich überlegen: Will ich systematisch Tierversuche, Laboruntersuchungen usw. aufarbeiten? Aus unserer Sicht ist das erst mal nicht sinnvoll aus verschiedenen Gründen. Erst einmal hat es keinen patientenrelevanten Bezug, dass man sich fragen muss, welchen Erkenntnisgewinn man daraus zieht. Das andere ist natürlich wieder ein Ressourcenaspekt. Unsere Frage ist, ob Sie das in dem Sinne gemeint haben, in dem Sinne einer systematischen Aufarbeitung des Ganzen? Die andere Frage ist: Wäre es nicht aus Ihrer Sicht sinnvoll, das in die Dossievorlagen aufzunehmen?

Michael Hennig: Es geht nicht um Tierversuche. Es geht nicht um Laborparameter. Wir haben in unserer Stellungnahme eigentlich nur über die biologische Rationale gesprochen. Wir haben auch Bezug genommen auf die Arbeit, die Sie ja auch hergenommen haben, von Sun im British Medical Journal 2012 veröffentlicht. Wir finden die Berücksichtigung jeglicher biologischer Rationale bei der Beurteilung der Glaubwürdigkeit für Subgruppen wissenschaftlich sehr sinnvoll, etabliert und zu berücksichtigen. Das war der Punkt, den wir machen wollten, den wir in dem jetzigen Methodenpapier vermisst haben, weil man sich ja auch im Klaren darüber sein muss, dass es sehr wohl zufällige, biologisch sinnfreie Beobachtungen geben kann. Deswegen wollten wir noch mal darauf hinweisen, dass die Berücksichtigung der biologischen Rationale sehr wichtig ist.

Dietrich Knoerzer: Ich kann das nur bestätigen, was Herr Hennig gerade gesagt hat. Der Punkt steht so hier drin, weil er in der Übersetzung direkt aus dem Paper ist.

Was ich noch als Ergänzung sagen könnte, ist: Es ist ja schon nicht so, dass wir mit den pivotalen Studien das erste Mal mit Informationen um dieses Produkt konfrontiert sind, sondern wir haben ja auch Informationen idealerweise aus Phase 1 und Phase 2. Wir hatten jetzt ein bisschen mehr gemeint, dass man solche Informationen auch mit heranzieht, um eine biologische Rationale für dieses Produkt mit zu gewinnen.

Volker Vervölgyi: Zum Punkt der biologischen Rationale: Das ist aus meiner Sicht sicherlich schwierig, weil man eine biologische Rationale in jedem Fall finden kann, sowohl in die eine als auch in die andere Richtung. Da gibt es sicherlich immer Argumente dafür. Aus Ihren Stellungnahmen geht es tatsächlich nicht direkt hervor, dass es sich nur darauf bezieht, sondern auch auf die andere indirekte Evidenz, die ich gerade genannt habe. Das war uns einfach nur nicht ganz klar. Also, Ihnen geht es maßgeblich tatsächlich darum, sich das biologisch plausibel zu erklären, ob diese Subgruppenanalyse irgendwie Sinn macht oder nicht. Es gibt ja andere Aspekte, die auch in dem Paper genannt werden, zum Beispiel: Ist es konsistent, über mehrere Endpunkte usw. ... Das sind sicherlich alle Aspekte, die wir uns sowieso anschauen auf Basis der Studien, die wir einschließen. Da frage ich mich tatsächlich, ob es darüber hinaus sinnvoll ist und ob es einen großen Erkenntnisgewinn bringt, noch die biologische Rationale heranzuziehen.

Michael Hennig: Wie gesagt, die biologische Rationale ist ein wichtiger Aspekt und sollte auch berücksichtigt werden. Wir haben darüber hinaus auch in unserer Stellungnahme auch noch auf andere Aspekte, die in dem Sun-Paper genannt werden, Bezug genommen. Da geht es ganz konkret darum, dass hier auch gefordert wird, dass die Anzahl der Subgruppenanalyse überschaubar sein sollte. Die schlagen konkret eine Größenordnung von fünf vor. Das haben wir noch einmal in den Ring geworfen, um es mal so auszudrücken. Es ging um die Frage der Konsistenz mit Evidenz von anderen Studien. Und es ging um die Konsistenz über verschiedene Endpunkte. All das sind unserer Auffassung nach wichtige Aspekte, die auch in der wissenschaftlichen Literatur so beschrieben werden, um letztendlich die Glaubwürdigkeit von Subgruppenanalysen zu untersuchen. Diese sollten auch berücksichtigt werden. Wir konnten es in der Klarheit in Ihrem Methodenpapier nicht finden. Deswegen haben wir noch mal darauf hingewiesen, dass diese Aspekte auch Berücksichtigung finden sollten.

Stefan Lange: Ich habe zwei Rückfragen. Wir sprechen ja jetzt im Wesentlichen über Dossierbewertungen. Da war ja auch die Frage von Herrn Vervölgyi: Sollen wir die Anforderungen an Dossierbewertungen, an die Dossiers reinschreiben, die Templates, dass die Hersteller sich freundlicherweise immer dann zur biologischen Rationale von durchgeführten und nicht durchgeführten Subgruppenanalysen äußern? Ich bin nicht sicher, ob Sie das wirklich so ernst meinen.

Die zweite Frage, die ich habe, würde mich wirklich interessieren. Das meine ich es ganz ernst. Kennen Sie eine Definition von „biologischer Rationale“, die für uns hilfreich operationalisiert werden kann? Was ist „biologische Rationale“?

Michael Hennig: Grundsätzlich möchte ich sagen: Alles, was ich sage, meine ich sehr wohl ernst. Ich denke, das trifft bei Ihnen auch zu.

Ich kenne keine Definition von „biologischer Rationale“, aber es gehört für mich zur guten wissenschaftlichen Praxis, wenn man denn Effekte beobachtet, wenn man zum Beispiel Gendereffekte beobachtet, dass man sich natürlich anschaut, hat der Wirkmechanismus meiner Substanz, gibt es entsprechende Voruntersuchungen, die zeigen, dass es wirklich solche Effekte geben kann oder dass es Effekte geben kann zwischen unterschiedlichen Ethnizitäten. Das sind sehr wohl anerkannte Formen, die biologische Rationale zu untersuchen. Dementsprechend würde ich es auch begrüßen. Wir machen es ohnehin. Wenn

wir solche auf den ersten Blick vielleicht fragwürdig erscheinende Merkmale beobachten, dann stellen wir uns natürlich die Frage: Woher kommen die? Haben die eine Rationale? Wenn ja, kann man die mit irgendeiner Form bestätigen? Deswegen ist das für uns ohnehin gelebte Praxis, unabhängig davon, was in der Dossievorlage steht.

Dietrich Knoerzer: Was ich noch sagen wollte, Herr Vervölgyi: Wir haben versucht, zu schreiben: das Vorhandensein indirekter Evidenz, zum Beispiel einer biologischen Rationale. - Und was ich versucht habe, zu sagen, ist: Man kann auch mal gucken, die indirekte Evidenz aus den Studien, die in der Studienentwicklung davorlagen, noch mal zur Kontrolle zu verwenden, um aus dieser Schwierigkeit, die ich auch sehe, eine biologische Rationale zu generieren, ein bisschen herauszukommen. Nichtsdestotrotz bin ich bei Herrn Hennig. Auch wir würden immer, wenn wir so einen Punkt sehen, versuchen, zu verstehen: Ist das ein Signal, oder ist das ein Multiplizitätsproblem?

Moderator Jürgen Windeler: Die beiden Fragen, die Stefan Lange gestellt hat, sind schon sehr relevant. Die erste Frage ist: Das, was jetzt hier zur Diskussion steht, kann unter Umständen – ich vernehme Sie so, es soll - Auswirkungen auf die Inhalte der Dossiers haben. Ich möchte noch mal darauf aufmerksam machen, dass ich die Diskussion im Moment etwas schräg finde. Wir reden natürlich dann nicht – warum auch? – über eine biologische Rationale nur für die auffälligen Effekte - wir haben gerade über die schlechte Power von Heterogenitätstests gesprochen -, sondern wir reden dann bitte auch über eine biologische Rationale für alle unauffälligen Effekte. Die Diskussionen, die ich aus dem G-BA mitbekomme, die sich um die Geschlechtsdifferenzierungen drehen, die wir nicht so selten in Dossiers finden, zeigen sehr deutlich, dass die Diskussion über die biologische Rationale, glaube ich, ein - da gebe ich Ihnen recht - relevanter Teilaspekt der gesamten Diskussion ist. Ob sie ein besonders wichtiger ist, wage ich heftig zu bezweifeln.

Andrej Rasch: Ich möchte kurz etwas dazu anmerken. Das war ja auch Gegenstand unserer Stellungnahme, in der wir darauf verwiesen haben, dass es diese weit bekannten Kriterien in der wissenschaftlichen Literatur bei der Überprüfung der Aussagekraft von Subgruppenanalysen gibt. Dazu gehören nach unserer Auffassung auch die Kriterien der biologischen Rationale, auch der Konsistenz der Ergebnisse aus ähnlichen Studien. Das sind im Übrigen auch Kriterien, die vom G-BA explizit in den tragenden Gründen herangezogen werden, wenn es um Befunde zu Subgruppenergebnissen geht, die vom IQWiG identifiziert werden, die auch diskutiert werden, die dann aber zum Beispiel von den medizinischen Fachgesellschaften noch mal kritisch erläutert werden, ob aus Sicht der Fachvertreter diese medizinische Rationale gegeben ist oder nicht. Wenn das ein wichtiger Diskussionsbestandteil beim G-BA ist, warum sollte das nicht auch ein wichtiger Diskussionsbestandteil beim IQWiG sein?

Ch.-Markos Dintsios: Ich persönlich würde gar nicht von biologischer Rationale, sondern von biologischer Irrationale sprechen wollen in den Fällen, wo es sozusagen vom Signal her überhaupt nicht interpretierbar ist im Vergleich zu dem Vorwissen, was wir haben, dem naturwissenschaftlichen Vorwissen. Ich würde sogar bis in die Präklinik hineingehen, nicht nur in die klinischen Studien. Sie haben den Wirkmechanismus angesprochen, Herr Kollege Hennig. Der ist ja dann auch vermutlich dynamisch sozusagen, untersucht am Tierexperiment vielleicht an der Zelllinie in vitro. Wieso er doch eine Relevanz hat, was das Bewertungsverfahren anbelangt, vielleicht nicht im vorgeschalteten Schritt, Herr Windeler,

aber im Nachgang, weil ganz einfach die Subgruppenproblematik uns wiederum bei dem Erstattungsbetrag einholt. Das heißt, ja, aus Ihrer Sicht vielleicht reduzierte Relevanz oder nicht so ausschlaggebende, aus unserer Sicht für das, was nach der Beschlussfassung des G-BA anbelangt, eminent wichtig.

Tilman Sauerbruch: Vielleicht ist das eine Wiederholung. Die ganze Medizin lebt natürlich davon, dass sie mit biologischer Rationale arbeitet. Das ist ja immer von den Klinikern oder auch von den Medizinerinnen die Betrachtung, wenn sie auf das IQWiG schauen. Sonst verfangen wir uns im Dschungel der statistischen Methoden, ohne dass Dinge Sinn machen. Ob Sie das jetzt in Ihr Papier hineinschreiben, oder ob das in der Gesamtbetrachtung dann beim G-BA stattfindet, das ist vielleicht nicht so entscheidend. Aber es ist natürlich etwas, was entscheidend ist für die Gesamtbetrachtung des Problems.

Friedhelm Leverkus: Ich wollte auch noch mal darauf hinweisen, dass es einfach darauf ankommt, wenn man ein positives Signal sieht, dass man das auch inhaltlich interpretiert. Da, glaube ich, ist es vielleicht eine Lösung, dass die Interpretation beim G-BA stattfinden kann. Das ist sicherlich nachvollziehbar, wenn die noch mal diese inhaltlichen medizinischen Sachen machen müssen, dass es ein großer Aufwand ist.

Dann möchte ich auch noch mal auf die Sun-Arbeit zurückkommen, eine wunderbare Arbeit. Aber man muss im Prinzip auch den Kontext sehen. Der Kontext ist: eine klinische Prüfung. Wie interpretiere ich Subgruppen in einer klinischen Prüfung? Wir haben es im Prinzip ja mit einem anderen Verfahren zu tun. Wir haben es im Prinzip eigentlich mit einer posteriori-Analyse von klinischen Prüfungen zu tun. Da macht der Begriff „a priori“ oder „posteriori“ meines Erachtens nur eingeschränkt Sinn oder ist anders zu interpretieren als in der Sun-Arbeit. Es kann sein, dass in den eigentlichen Studienprotokollen niemals geplant worden ist, die Geschlechtssubgruppe zu untersuchen. Aber ist das jetzt posteriori oder a priori? Wenn diese Studie nach 2011 aufgesetzt worden ist, dann steht im AMNOG-Gesetz ja schon drin, dass eine Subgruppenanalyse nach dem Geschlecht durchgeführt werden soll. Es ist im Prinzip im Studienprotokoll nicht geplant worden, ist aber im Prinzip im Gesetz schon vorgeschrieben. Von daher gehört diese Begrifflichkeit „a priori“, „posteriori“ im Prinzip in eine klinische Studie hinein. Wir müssen sicherlich die Dinge, die wir in einer klinischen Studie als Subgruppen schon definiert haben, auch im Dossier darstellen. Aber es ist im Prinzip noch kein Qualitätsmerkmal. Es kann durchaus auch sein, dass in der klinischen Studie Subgruppen spezifiziert worden sind aufgrund des primären Endpunkts, aber dass in dem Dossier, wo wir andere Endpunkte haben, wo wir die Lebensqualität sehr stark untersuchen müssen, wo wir die unerwünschten Ereignisse sehr stark untersuchen müssen, noch andere Subgruppen fachgerecht oder sachgerecht sind. Von daher verwirrt im Bereich unserer Dossierbewertungen der Begriff „a priori“ und „posteriori“ nur.

Dirk Eyding: Ich wollte noch mal darauf zurückkommen, was wir vorhin schon diskutiert hatten, nämlich den p-Wert, jetzt im Zusammenhang mit der für mich bei Subgruppenanalysen ganz wichtigen Konsistenzprüfung. Im Gegensatz zu den Metaanalysen ist es ja so, dass wir die Subgruppenergebnisse eigentlich nur durch die Vorlage des Dossiers bekommen. Die sind sonst nirgendwo verfügbar, wahrscheinlich nicht mal im Studienbericht. Das heißt, um eine Konsistenzprüfung vorzunehmen, ist es aus meiner Sicht viel schwieriger, wenn man nur die p-kleiner-0,05-Subgruppenergebnisse im Dossier geliefert bekommt, weil das viel weniger sind und eben eine Konsistenzprüfung nicht möglich ist. Mein Punkt vorhin

war ja auch bei den Subgruppenanalysen, dass aus meiner Sicht die Schwelle bei 0,2 erhalten bleiben sollte, unter anderem weil die Konsistenz über das vorgebliche Subgruppenmerkmal eben viel besser geprüft werden kann, wenn man diese Ergebnisse auch hat. Das heißt, zur Prüfung sollte man diese Daten auf jeden Fall haben. Aus den Dossievorlagen sollte das nicht verschwinden, damit man diese Prüfung noch vornehmen kann, denn ansonsten sind diese Daten einfach nicht verfügbar.

Stefan Lange: Ich muss gestehen, dass ich ein bisschen den Eindruck habe, dass das jetzt schon fast etwas grotesk wird. Ich hatte eben gefragt: Was ist „biologische Rationale“? Das wurde mir nicht beantwortet.

(Michael Hennig: Ich hatte Beispiele genannt!)

- Ich hatte Sie nach einer Definition gefragt. Sie haben gesagt, eine solche kennen Sie auch nicht. Beispiele sind keine Definitionen.

Herr Sauerbruch, die Medizin ist voll von Irrtümern, die auf der biologischen Plausibilität beruhen, die hinterher widerlegt worden sind. Wir wissen es doch alle: Mediziner, wir alle, sind ja Meister darin, post hoc die Ergebnisse zu erklären.

Der Vorteil, lieber Herr Leverkus, an „a priori“, „a posteriori“ ist schon, dass es einigermaßen gut definiert ist im Gegensatz zu dem, was wir gerade besprochen haben. Aber wenn wir jetzt noch anfangen, das, was wir gar nicht so gut beschreiben können – ich jedenfalls nicht –, auszuweiten, jetzt bei der Schwelle 0,2 sind, also jetzt auch noch versuchen wollen, wahnsinnig viele Signale biologisch zu erklären, warum sie da auftreten und da nicht, ich glaube, dann wird es ein bisschen kompliziert. Nehmen wir einmal die Lebensqualität. Wo fangen wir denn da an, wenn wir da Subgruppenauffälligkeiten haben, mit der biologischen Rationale zu argumentieren? Ich meine das wirklich ernst. Ich sehe da erhebliche Schwierigkeiten, das sauber darstellen zu können.

Dirk Eyding: Ich habe das mit der Konsistenzprüfung nicht gemeint, um die biologische Rationale zu begründen, sondern ich meinte den Punkt der Konsistenz eines Subgruppeneffektes einschätzen zu können, wird, wenn ich beim Schwellenwert von 0,05 bekomme ich vielleicht zwei sogenannte Belege für eine Interaktion und sehe sie nicht mehr und kriege sie auch nie mehr, dass ich für das gleiche Subgruppenmerkmal eben auch ein paar - bei anderen Endpunkten - Hinweise habe mit einem p von 0,06. Die sehe ich nicht mehr. Für die Konsistenzprüfung ist das für mich ein ganz wichtiges Merkmal. Das ist, glaube ich, in der Argumentation im G-BA immer eine wichtige Stellgröße gewesen, aufgrund derer man dann in den seltenen Fällen – das gebe ich zu, das sind seltene Fälle, das liegt nicht an uns – tatsächlich ein Subgruppenmerkmal beschrieben worden ist. Es ist aber ein wichtiges Merkmal, um zu prüfen, ob das ein Zufallsbefund ist oder nicht. Es geht mir dabei überhaupt nicht um die biologische Rationale, sondern um die Konsistenz des Subgruppeneffekts.

Friedhelm Leverkus: Herr Lange, ich gebe Ihnen natürlich recht: „posteriori“ und „a priori“ in der klinischen Prüfung machen Unterschiede in der Aussagekraft. Mir geht es im Prinzip nur darum: Wenn wir über Nutzenbewertung sprechen, macht es keinen Sinn. Denn es ist auch so: Wenn man sich Lebensqualitätsparameter anschaut und sieht Subgruppen, dann fragt man sich natürlich auch: Macht das jetzt Sinn, was wir da sehen, oder inwieweit ist das ein

Zufallsbefund? Es ist vielleicht keine biologische Rationalität, sondern es ist dann vielleicht eine soziologische oder psychologische Rationalität, die man sich dann natürlich erfragt.

Michael Hennig: Noch einmal zurück zu unserem Lieblingsbegriff heute, der biologischen Rationalen oder - ich nehme das gerne von Herrn Dintsios auf - der biologischen Irrationalen. Sie haben ja gefragt, wie man das überhaupt sauber machen kann. Ich glaube, hier geht es gar nicht um die Frage, wie man das sauber in Form einer Checkliste abarbeiten kann, sondern es geht einfach darum, wie man das gemäß guter wissenschaftlicher Praxis grundsätzlich berücksichtigen sollte. Ich möchte gerne wiederholen, was ich gerade gesagt habe, weil ich das für so wichtig empfinde, dass einfach diese Berücksichtigung dieser Rationalen oder Irrationalen einfach dazugehört. Die kann man nicht einfach ausblenden, auch nicht aus dem Hinweis, dass man das vielleicht nicht sauber machen kann. Dass man es machen sollte, steht für mich außer Frage.

Volker Vervölgyi: Das wäre noch mal meine Frage gewesen. Es sind ja in dieser Arbeit mehrere Kriterien genannt worden. Angenommen, ich habe jetzt zwei Studien, die ich in meine Fragestellung einschließe, und es zeigt sich ein Geschlechtseffekt, also ein Subgruppeneffekt für das Geschlecht, konsistent in beiden Studien über mehrere Endpunkte, und ich kann noch so lange nachdenken, mir fällt keine biologische Rationale dazu ein. Wäre das dann ein Grund für Sie, diese Subgruppe nicht zu betrachten, obwohl ich eigentlich tatsächlich konsistente Effekte sehe?

Michael Hennig: Das habe ich so nicht gesagt. Ich habe nur gesagt, man sollte die biologische Rationale berücksichtigen. Es mag Fälle geben, wie Sie sie geschildert haben, dass es keine biologische Rationale gibt, aber ich habe mir zumindest Gedanken darüber gemacht. Das war mein Punkt, dass man das einfach berücksichtigen sollte, wohlwissend, dass es in manchen Fällen dies nicht gibt.

Moderator Jürgen Windeler: Ich lasse das jetzt mal so stehen. Diese Äußerung widerspricht einfach Anforderungen - das muss ich klar sagen -, die wir im nächsten Stellungnahmeverfahren kriegen werden, wo wir nämlich dann reinschreiben: Wir werden das berücksichtigen. - Dann wird der Stellungnehmende fragen: Wie denn? - Und dann werden wir sagen: Das wissen wir auch nicht. Herr Hennig weiß es auch nicht, wie es genau passieren soll. - Der Punkt ist, dass wir in der Tat für diese biologische Rationale keine ... Ich glaube, dass wir alle uns vorstellen können, was das ungefähr bedeutet. Es wäre auch unproblematisch, das als Kriterium für die Validität von Subgruppenanalysen irgendwo aufschreiben zu können, dass wir dann immer noch nicht genau wissen, was das bedeutet, und es dann sozusagen eines von mehreren Ermessenskriterien ist, die man da reinschreibt. Wenn wir solche Ermessenskriterien aufschreiben und uns jetzt einig sind unter 55 Teilnehmern, dass wir dann zu diesem Punkt nächstes Mal keine Stellungnahmen bekommen werden, das näher zu spezifizieren, dann können wir uns, glaube ich, in diese Richtung verständigen. Aber noch mal: Entweder wir gehen in die Richtung einer von Ihnen teilweise geforderten Konkretisierung von Operationalisierung, oder wir belassen es bei der Aufzählung von Kriterien. Wohlgemerkt: Ich kann damit leben, ich kann damit gut leben, aber das entspricht nicht immer Ihren Vorstellungen.

Michael Hennig: Ich habe nicht gefordert, dass das konkretisiert, dass es operationalisiert wird. Ich habe darauf hingewiesen, es sollte berücksichtigt werden. Nur für das Protokoll.

Moderator Jürgen Windeler: Ich habe das wohl verstanden, aber ich habe nur darauf hingewiesen, dass andere, wenn wir schreiben: „Das soll berücksichtigt werden“, vermutlich nach einer Operationalisierung fragen, wir vorher also auch.

Tilman Sauerbruch: Ich will es abschließen. Ich glaube, Sie als Institut sind damit überfordert, eine biologische Rationale da reinzuschreiben. Das betrifft auch Ihre Methodik. Aber Sie können eine Subgruppe identifizieren. Dafür gibt es eine biologische Erklärung. Das kann ja sehr wohl vorkommen. Aber das muss auf anderen Ebenen diskutiert werden. Insofern möchte ich Ihnen recht geben.

Moderator Jürgen Windeler: Ich fühle mir nicht recht gegeben. Denn ich habe ausdrücklich nicht gesagt, dass wir überfordert wären. Ich habe gesagt, das geht nicht. Ich wäre sehr interessiert daran, wenn wir da jetzt reinschreiben, es ist zu berücksichtigen, dann die Operationalisierung von anderen Fachgesellschaften, zum Beispiel der AWMF, zu lesen, wie man das konkret machen soll. Darauf bin ich sehr neugierig.

Ulrich Mansmann: Ich habe noch etwas zu einem anderen Punkt. Wenn noch Kommentare im Raum sind, dann würde ich gerne zurückgestellt werden.

Ch.-Markos Dintsios: Ich glaube, ein pragmatischer Vorschlag wäre, dort, wo ein Signal erkannt wird, aber der Klinik vorher in all den Studien, die wir hatten, nicht sichtbar war, zumindest die Möglichkeit auch im Rahmen eines Freitextes in dem Template zu geben, auf eine potentielle biologische Irrationale - deswegen habe ich es „Irrationale“ genannt, also ein bisschen pointiert - ... Also da, wo wider Erwarten etwas plötzlich kommt, was vorher nicht da war, dass uns zumindest - wir erstellen Sie ja auch selber in dem Sinne - diese Möglichkeit gegeben wird.

Und ein Beispiel auf die Frage von Stefan Lange zur Lebensqualität: Natürlich gibt es da keine biologische Rationale. Es gibt aber oft Kontextfaktoren. Ein typisches Beispiel Subsidiaritätslebensfragebogen in drei angelsächsischen Ländern: Wie begehen Sie Selbstmord? - Kanada anders, Amerika anders, England anders, einmal Schwert, einmal Pistole, einmal Jagdgewehr. Wenn Sie das Instrument querbeet über die Angelsachsen angewandt hätten, hätten Sie auf einmal Subgruppenprobleme gehabt zwischen Kanadiern, Anglikanern, US-Amerikanern und Briten. Da gibt es also auch Fälle, wo es eine Irrationale geben kann. Die ist nicht biologisch bedingt, die ist, wie Friedhelm Leverkus gesagt hat, eher sozialfaktorenmäßig angelegt

Moderator Jürgen Windeler: Wenn man jetzt diesem Konsistenzfrageargument folgt, würde ich aber noch mal darauf hinweisen, dass wir dann auch die Konsistenzfrage stellen müssen, wenn wir etwas nicht mehr sehen als Unterschied. Das geht immer so ein bisschen unter. Wir reden immer über Signale, denen man dann nachgehen muss. Wir reden nicht nur darüber, dass man Signalen nachgehen muss. Wenn, dann reden wir über biologische Rationale, um den Begriff jetzt ein letztes Mal zu gebrauchen, jedenfalls von mir, über Erklärungsmöglichkeiten von Ergebnissen generell, nicht bloß von Signalen. Nicht nur die Signale sind die auffälligen.

Jetzt Tim Friede. Geht es um das Thema?

Tim Friede: Es geht genau um das Thema, und auch in Richtung Operationalisierung. Ich glaube, das ist ja schon deutlich geworden. Es gab ja schon einen Vorschlag, der in die Richtung Operationalisierung ging. Ich wollte dem hinzufügen, dass man auch nicht die Vergleichstherapie ganz vergessen darf. Wenn ich in meinem Entwicklungsprogramm bisher keine Interaktion mit Geschlecht beobachtet habe und jetzt auf einmal eine beobachte, kann es natürlich sein, dass die Vergleichstherapie irgendwie sensitiv ist. Aber wenn ich natürlich da auch einen Pool von Studien habe, wo nie eine Interaktion mit Geschlecht beobachtet wurde, dann halte ich es doch für unwahrscheinlicher, dass das Signal, das ich jetzt sehe, das statistisch empirische Signal wirklich etwas mit einem biologischen Mechanismus zu tun hat.

Moderator Jürgen Windeler: Es liegt mir auf der Zunge: Was ich da mache, ist, ich stütze mich auf alte Empirie, die sehr zuverlässig sein mag, nicht auf eine biologische Rationale. Das ist etwas anderes.

Tim Friede: Von mir kommt ja der Begriff „biologische Rationale“ nicht. Was wir jetzt versucht haben, ansatzweise zu operationalisieren, war ja - das klingt ja auch vorhin schon an -, dass man zunächst einmal in den klinischen Daten - das ist ja wohl auch das Vorgehen - des Entwicklungsprogramms guckt, ob es Anhaltspunkte gibt. Dann guckt man in Studien zur Vergleichstherapie, ob es dort Anhaltspunkte gibt. Und wenn man dann meint, man müsste noch weitergehen, dann zieht man nichtklinische Daten heran. Aber ich glaube schon, dass man da auch eine Kaskade oder eine Operationalisierung durchaus formulieren kann, wir vielleicht nicht hier im Moment, aber, ich denke, es ist schon möglich.

Ulrich Mansmann: Ich möchte zu dem zweiten Subgruppenpunkt etwas sagen, der auf Seite 194 steht, nämlich die Subgruppen, die dann postrandomisiert entstehen. Das ist ja schon eine Sache, die extrem wichtig wird. Zum Beispiel in der Onkologie hat eine frühe Veränderung eines Biomarkes eine gewisse Aussagekraft für die Wirksamkeit einer gewissen Therapie. Ich schlage vor, dass man in der Stellungnahme diese Subgruppen extra behandelt und nicht mit dem Wort hinstellt, dass sie grundsätzlich unzuverlässig sind. Es gibt ja Gewichtungssachen, also diese Methoden, die letztendlich Überlebenssachen und longitudinale Daten zusammenbringen. Es gibt ja schon Verfahren, wo man hier sauber arbeiten kann. Man sollte das vielleicht auch so darstellen im Methodenpapier, dass das eine andere Klasse ist von Subgruppenanalysen mit spezieller Methodik.

Sandra Janatzek: Ich möchte gerne auf das Signifikanzniveau für den Interaktionstest zurückkommen und Sie vom IQWiG gerne fragen, wie Sie sich zu dem Punkt stellen, den wir eben schon erläutert haben. Aus unserer Sicht gibt es halt einen Unterschied zwischen dem Absenken des Signifikanzniveaus bei den Heterogenitätstests in den Metaanalysen und jetzt diesem klaren 5%-Signifikanzniveau für die Interaktionstests in den Subgruppenanalysen, jedenfalls dann, wenn wir Subgruppenanalysen im Kontext AMNOG meinen. Das hat Herr Eyding ja eben schon dargestellt. Während wir in den Metaanalysen und insbesondere Sie in jedem Fall, egal ob man jetzt das Signifikanzniveau abgesenkt oder es wie bisher belässt, alle Daten sehen und auch der Leser Ihrer Berichte alle Daten sieht, zum Beispiel im Konfidenzintervalleplot, in den dazugehörigen Ergebnistabellen, ist es jetzt ja hier in der Subgruppenproblematik ein bisschen anders. Auf gut Deutsch: Bei den Metaanalysen können Sie sich selbstverständlich überlegen, dass Sie abweichend von dem Signifikanzniveau 5 % vielleicht trotzdem mal bei einem größeren p-Wert sagen: Eigentlich sieht die Systematik vor,

wir machen jetzt eine metaanalytische Zusammenfassung der Studien, aber wir halten die für ungeeignet, weil ... und bilden folgende Subgruppen von Studien. - Dazu liegen alle Daten vor. Der Unterschied hier scheint mir doch zu sein, jedenfalls dann, wenn wir im Zusammenhang sind von AMNOG oder überhaupt vom Vorliegen von Studienberichten, dass jedenfalls dann, wenn das klare 5%-Signifikanzniveau in den Dossievorlagen abgebildet wird, Sie eben nicht mehr die Möglichkeit haben wie bei den Metaanalysen, sich einen Eindruck zu verschaffen, ob auch mit einem p-Wert des Interaktionstests oberhalb von 5 % vielleicht eigentlich tatsächlich mal ein gewisses Signal zu sehen ist, dass wir eigentlich würdigen wollen. Das Signal ist dann unsichtbar, weil jedenfalls für die patientenrelevanten Endpunkte nicht zwingend diese Subgruppenanalyse im Studienbericht vorzufinden ist, soweit ich das im Moment überblicke.

Guido Skipka: Ich verstehe dein Argument. Wenn man die Studienpopulation als Ganzes betrachtet, dann hat man alle Informationen, und in den Subgruppen kappen wir das irgendwie. Aber man muss ja fragen, wie das praktisch aussehen soll. Wir haben bisher, wo wir auch Interaktionen bis 0,2 sehen wollten, Dossiers mit mehreren 10.000 Seiten, wovon die Hälfte Subgruppenanalysen waren. Denn wir haben mehrere Fragestellungen, mehrere Endpunkte und dann noch vier, fünf Subgruppenfaktoren. Das multipliziert sich halt jedes Mal. Das ist schlicht und einfach aus unserer Sicht nicht handelbar für das, was im Endeffekt dabei herauskommt. Von daher müssen wir irgendeinen Filter setzen, wann wir uns die Dinge anschauen. Wir hatten den bis jetzt bei 0,2. Jetzt haben wir das auf 0,05 abgesenkt. Du, Dirk, hast gesagt, mich interessiert auch 6- und 7%-Niveau. Nichtsdestotrotz würde ich die Frage zurückstellen: Wo ist denn bei euch die Grenze, wo ihr nicht mehr reinschauen würdet?

Dirk Eyding: Ich würde es so lassen, wie es war. 20 % finde ich super. Die Frage kann man zurückstellen: Wie macht ihr eine Konsistenzprüfung, ohne diese Daten zu haben, die ihr nicht haben werdet? Wenn das nicht in den Dossievorlagen steht, wird es diese Subgruppenanalyse nicht geben. Es gibt nur die Belege. Die 0,2-Effekte werden alle nicht mehr auftauchen. Das wird das Dossier verkleinern. Ein 10.000-Seiten-Dossier ist ja wirklich der Extremfall, den du uns gerade geschildert hast. Es ist nicht der Regelfall, dass 10.000-Seiten-Dossiers eingereicht werden. Selbst wenn – es sind dann auch mehrere Fragestellungen –, würde ich es für sinnvoll halten, die Konsistenzprüfung so durchzuführen. Also ich frage zurück: Wie macht man Konsistenzprüfungen, wenn nur die Belege da sind?

Stefan Lange: Da müssen wir ein bisschen aufpassen. Vielleicht habe ich es selber mit induziert. Wenn wir über Dossievorlagen sprechen, sprechen wir über ein Verfahren beim Gemeinsamen Bundesausschuss, nicht über unsere Dossierbewertungen. Ich glaube, es ist dem Gemeinsamen Bundesausschuss unbenommen, in Dossievorlagen als Anforderung ich weiß nicht was hineinzuschreiben, aber ich hätte jetzt einen Vorschlag, was man hineinschreiben könnte. Was wir ja auf jeden Fall kriegen, sind Geschlecht, Alter, Schweregrad. Ich glaube, das ist a priori, lieber Herr Leverkus, so festgelegt. Jetzt könnte man aber noch reinschreiben, dass in Abhängigkeit von der biologischen Rationale erklärt werden muss, welche weiteren Subgruppenanalysen gemacht worden sind und nicht gemacht worden sind.

Dirk Eyding: Das beantwortet nicht die Frage nach der Konsistenzprüfung.

Volker Vervölgyi: In Modul 5, also im nicht veröffentlichten Teil, werden die natürlich trotzdem vorhanden sein. Das heißt, unsere Konsistenzprüfung ist dadurch erst einmal nicht beeinträchtigt. Das einzige, was sich wahrscheinlich ändern wird, ist, dass man im Modul 4 nicht mehr 1.000 Seiten Subgruppenanalysen hat, sondern vielleicht nur noch 200.

Dirk Eyding: Ich bezweifle, dass in den Modul-5-Unterlagen für die relevanten Endpunkte alle Subgruppenanalysen für alle Endpunkte, die uns interessieren, vorhanden sind. Das werden sie nicht sein. Das werden wesentliche Subgruppenanalysen sein, vielleicht sogar nur für den primären Endpunkt, der uns möglicherweise in den Studien überhaupt nicht interessiert. Ich bin jetzt nicht ganz im Film drin, was genau in den Studienberichten gemacht wird, aber definitiv nicht das, was immer gebraucht würde.

Moderator Jürgen Windeler: Ich sehe keine weiteren Wortmeldungen. Dann möchte ich diesen Tagesordnungspunkt abschließen.

Ich möchte aber noch einen Punkt ansprechen, der mir im Moment ein bisschen Unbehagen macht, nämlich zu dem Wort, das wir jetzt mehrfach benutzt haben mit den zwei Teilen, das ich nicht noch mal verwenden wollte: Es gibt ja die Diskussionen im Gemeinsamen Bundesausschuss, an denen wissenschaftliche Fachgesellschaften teilnehmen, die Trägerorganisationen teilnehmen, wo die zur Überzeugung kommen, dass für so gut wie keine Subgruppenanalyse, für kein Signal es eine ausreichende Rationale gibt, um daraus einen Zusatznutzen zu differenzieren, sodass das eigentlich ein Grund für die meisten dieser Signale oder für fast alle ist, die sogenannte biologische Rationale zu verneinen. Wenn das aber das Ende der Strecke ist und wir uns unter anderem auch um Effizienzfragen kümmern müssen, dann müssen wir auch zur Kenntnis nehmen, dann müssen auch alle Beteiligten, wir auch, sagen: Okay, dann werden wir uns eben in diese Form von Konsistenzprüfung, uns die Signale anzugucken, die wahrscheinlich alle keine Rationale haben, nicht so tief begeben wie bisher. Alle Beteiligten, insbesondere auch die Argumentation im G-BA und in den Erörterungen, die ich teilweise mitgeteilt bekomme, könnten ja ihr Verhalten an dieser Stelle ändern. Aber ansonsten müssen wir ein Stück weit, um uns nicht auch ganz anders gearteten Vorwürfen auszusetzen, darauf reagieren und sagen: Gut, wenn es so ist, dass die Signale – ich überspitze jetzt ausdrücklich – alle nichts bedeuten, dann kann es nur bedeuten, dass wir uns so intensiv wie bisher um diese Signale nicht kümmern müssen. Ich glaube nicht, dass sie so wenig bedeuten, um das gleich dazu zu sagen, aber das ist eine andere Diskussion.

Das Problem haben wir, glaube ich, noch nicht gelöst, jedenfalls noch nicht für alle befriedigend gelöst, aber, ich glaube, wir können jetzt diesen Tagesordnungspunkt verlassen, es sei denn, ich sehe irgendwelche Widerworte. Diese sehe ich nicht.

Wir kommen zum

Tagesordnungspunkt 6:

Beteiligung von Betroffenen an der Entwicklung von HTA-Berichten

Ulrich Siering: Die Frage unter diesem TOP bezieht sich auf die neu in das Methodenpapier aufgenommenen Abschnitte zum Themencheck Medizin. Hier wurde in mehreren Stellungnahmen der Einbezug von Betroffenen nicht nur in die Themenauswahl, sondern auch

in der Stellung der HTA-Berichte angesprochen. Insbesondere in der Stellungnahme der BAG SELBSTHILFE wurde auch mit Blick auf die HTA-Berichte gefordert, dass eine methodisch abgesicherte Einbindung der Patienten in die Berichterstellung erfolgen sollte. Wir haben uns gefragt, was sich hinter dieser methodisch abgesicherten Einbindung versteckt, und wollten daher fragen, ob Sie uns die abgesicherten Methoden für die Einbindung von Betroffenen in die Erstellung der Berichte nennen können, die Sie hier im Blick hatten.

Moderator Jürgen Windeler: Nun hätten wir gerne die BAG SELBSTHILFE selber gefragt, aber wie ich der Teilnehmerliste entnehme, ist die BAG SELBSTHILFE nicht da. Das stelle ich erst einmal so fest. - Frau Lühmann. Sie sprechen aber nicht für die BAG SELBSTHILFE?

Dagmar Lühmann: Ich spreche nicht für die BAG SELBSTHILFE, aber wir haben in unserer Stellungnahme etwas ganz Ähnliches thematisiert. Deswegen kann ich ja mal versuchen, zu erläutern, was zumindest wir gemeint haben.

Was relativ gut im Methodenpapier abgehandelt ist, obwohl man sich vielleicht ein bisschen mehr Transparenz wünschen könnte, sind die Schritte, wie man zu einem Thema kommt, also wie die Themenwahl stattfinden soll. Dann wird darauf verwiesen, dass die Bewertung quasi dem Konzept von Health Technology Assessment folgt und mehrdimensional angelegt ist. Es wird also gesagt, es muss in jedem Fall was zum Nutzen gesagt werden. Die Gesundheitsökonomie, Ethik, soziale Aspekte usw. müssen angesprochen werden. Was jetzt hier fehlt, ist der Schritt, die Gewichtung vorzunehmen, welcher Aspekt bei dem betreffenden Thema den Schwerpunkt bekommen soll. Ich könnte mir vorstellen, dass die BAG SELBSTHILFE ... Ich weiß es nicht, aber zumindest wir haben gemeint, dass in diesen Prozess, für den es Vorschläge für formalisierte Verfahren gibt, also sozusagen Scoping-Verfahren, Scoping-Workshops, in der Tat Patienten einbezogen werden. Es gibt Vorbilder. Ich glaube, das NICE machte das in Großbritannien. Es gibt aus kanadischen HTA-Institutionen dort Vorbilder. Ein sehr komplexes Modell ist jetzt gerade in einem von der EU geförderten Forschungsprojekt, dem INTEGRATE-HTA-Projekt, behandelt worden, was unter der Federführung der Uni Bremen entstanden ist. Dort ist eine Reihe von Methodenpapieren herausgekommen, die dieses durchaus thematisieren. Das ist der eine Aspekt, den ich hierzu anbringen wollte.

Der zweite Aspekt, wie man sich weiterhin ein formalisiertes Einbringen vorstellen könnte, ist, dass in diesen komplexen Technologiebewertungen durchaus Dimensionen drin sind, was zum Beispiel die Ethik anbelangt, die man relativ schlecht auf Metaebene, Sekundärdatenebene bearbeiten kann. Das heißt, dort kann die Notwendigkeit entstehen, tatsächlich zu primären Datenerhebungsmethoden zu greifen. Das hört sich jetzt so gewaltig an, aber im Prinzip denken wir da eher an qualitative Verfahren, zum Beispiel im Sinne von Fokusgruppen, Verfahren oder qualitativen Interviews zum Beispiel mit Betroffenen, die natürlich nicht der Dimension Nutzenbewertung zuzuordnen sind, sondern der Eruiierung des Kontextes bzw. Einordnung der Ergebnisse, ethische Einordnung usw.

Also, diese beiden Aspekte wären es, Scoping und qualitative Methoden.

Moderator Jürgen Windeler: Ich könnte mir vorstellen, dass die BAG SELBSTHILFE, jedenfalls was die Nennung von Methoden angeht, ähnlich geantwortet hätte, um das mal so

zu formulieren. Aber Sie stellen sich vor, dass das auch die BAG SELBSTHILFE gemeint haben könnte.

Dagmar Lühmann: Das weiß ich nicht. Wir haben das gemeint.

Moderator Jürgen Windeler: Ich verstehe. Ich denke, das klärt oder sortiert unsere Frage ein Stück weit. Gibt es weitere Wortmeldungen zu dem Punkt? – Sehe ich auch nach kurzer Denkpause nicht.

Dann kommen wir zum

Tagesordnungspunkt 7:

Verschiedenes

Gibt es Wortmeldungen Ihrerseits?

Ulrich Mansmann: Ich möchte etwas zum Abschnitt 6.3.1 etwas fragen. Da geht es praktisch um Präventionsmaßnahmen. Ein Gesetzestext ist genannt. Dort steht, dass Versicherte und sonstige interessierte Einzelpersonen beim Institut Bewertungen nach § 139a Abs. 3 Nr. 1 und 2 zu medizinischen Verfahren und Technologien vorschlagen können. Das Institut soll die für die Versorgung von Patientinnen und Patienten besonders bedeutsamen Vorschläge auswählen und bearbeiten. - Das ist der Satz.

Uns war nicht klar, ob damit nicht logischerweise Primärpräventionsmaßnahmen, Lebenszieländerung, Supplementierung von Lebensmitteln, implizit von Evaluation ausgeschlossen sind, weil da ja diese Patienten Aspekte nicht so im Vordergrund seien. Ich wollte einfach nur mal nachfragen, wie das zu lesen ist, ob wir da richtig liegen, wenn wir das so verstehen, dass Primärpräventionsmaßnahmen nicht evaluiert werden.

Ulrich Siering: Wir teilen diese Einschätzung nicht, dass durch diese Formulierung entsprechende Maßnahmen ausgenommen wären.

Dagmar Lühmann: Jetzt sind es plötzlich zwei Aspekte geworden, weil ich daran kurz anknüpfen möchte. Wie eng muss man das sehen mit dem Bezug auf das SGB V? Ich weiß, da sind Sie nicht die richtigen Adressaten, daran etwas zu ändern, aber Sie können ja zumindest eine Auskunft geben. Der ganze Bereich Rehabilitation fällt in ein anderes SGB. Das Thema „Prävention“ sprachen Sie an. Das sind ja möglicherweise Leistungen, die überhaupt nicht dem Krankenkassenrecht oder so unterliegen, sondern die selbstinitiierte Dinge sind. Wie eng muss man diesen Fokus sehen? Das ist meine Frage.

Der zweite Aspekt, den ich ansprechen möchte, ist Folgender: Ich möchte gerne noch mal den Vorschlag „Methodik von Bewertung von komplexen Interventionen“ in den Ring werfen, gerade im Zusammenhang mit den HTA-Bewertungen, mit der Multidimensionalität. Da gibt es mindestens seit dem Jahr 2000 wenn nicht schon länger Überlegungen, wie komplexe Interventionen, die aus mehreren Komponenten bestehen, wo die Komponenten interagieren, wo es stark auf die Implementierung ankommt, unterschiedliche Outcomes fokussiert werden ... Da gibt es eine relativ stringente Methodik, vom MAC-Council in Großbritannien

entworfen, wie man solche Verfahren bewerten kann, ohne auf die klassische Systematic Review plus Metaanalysemethodik zurückgeworfen zu sein. Da drin kommen Methoden vor, natürlich auch systematisch bewertet, wie man zum Beispiel qualitativ Ergebnisse zusammenfassen kann und dadurch deutlich mehr Informationsgewinn einfahren kann, als wenn man das nur reduziert auf das alleinige betrachten eines erzielten Effektes.

Désirée Fimm: Ich habe eine Frage zu den indirekten Vergleichen. Es geht darum, dass Sie jetzt ausgeschlossen haben, dass einarmige Studien akzeptiert werden. Es gibt aber Produkte, die von der EMA zugelassen werden aufgrund dieser einarmigen Vergleiche. Dann müssen wir die ja streng genommen auch zum Dossier mit einreichen. Wie würden Sie dann da vorgehen? Was ist der Lösungsansatz, wenn wir nur die einarmigen Studien haben, die wir dann zusammensetzen müssen mittels verschiedener Dossierungsmethoden, die dann bewertet werden müssten, denn es ist ja meistens der Fall, dass sie einfach nicht bewertet werden? Was sagen Sie dazu? Was sollte man ansonsten tun?

Monika Nothacker: Ich habe eine kleine Nachfrage an Herrn Siering. Wenn Sie sozusagen Primärpräventionsmaßnahmen oder - die Hebammen haben uns geschrieben - von Schwangeren nicht ausschließen, bedeutet das dann, dass Sie „Patientinnen“ und „Patienten“ durch andere Ausdrücke in diesem Abschnitt des Methodenpapiers ersetzen?

Ulrich Siering: Der erste Punkt bezog sich auf die Frage, inwieweit durch die Formulierung, die dort gewählt wird, und den Bezug auf die entsprechenden Paragraphen bestimmte Interventionen ausgeschlossen werden, eine ähnliche Fragerichtung wie die Frage vorher. Wir interpretieren die Formulierung im Gesetz nicht dahingehend, dass entsprechende Interventionen aus der Bearbeitung ausgeschlossen sind, sondern dass die Formulierung so weit gefasst ist, dass das möglich ist. Was tatsächlich ausgenommen ist, sind Arzneimittelbewertungen.

Worauf bezog sich noch einmal deine Frage?

Monika Nothacker: Es ist wirklich nur eine kleine redaktionelle Sache. Es sind ja nicht alles Patienten oder Patientinnen, sondern eben dann Betroffene oder andere.

Ulrich Siering: Wir werden prüfen – ich bitte um Entschuldigung, dass wir jetzt keine dezidierten Zusagen machen können –, ob und wie bestimmte Formulierungen geändert werden. Wir werden uns das auf jeden Fall noch einmal ganz genau anschauen.

Stefan Sauerland: Einen Punkt noch.

Vorneweg: Natürlich ist das mit dem Benennen der Teilnehmer an Präventionsmaßnahmen ein Punkt, den wir auch in anderen Berichten und auch in anderen Teilen des Methodenpapiers schon umgesetzt haben. Es gibt ja ein eigenes Kapitel zu Präventionsmaßnahmen. Da ist allenfalls von Versicherten oder Teilnehmern die Rede. Da ist nicht von Patienten die Rede.

Zu den komplexen Interventionen, die Frau Lühmann angesprochen hat: Das ist immer ein Punkt, der einerseits bei den eigenen Nutzenbewertungen des Institutes eine Rolle spielt, der aber jetzt natürlich in besonderer Weise bei den Themencheckbearbeitungen eine Rolle

spielen wird. Bei den Bewertungen, die zu den nichtmedikamentösen Interventionen erfolgen, hatten wir bislang nach meiner Einschätzung eher wenig komplexe Interventionen. Es gibt Interventionen, die tatsächlich mal einen höheren Komplexitätsgrad aufweisen – so etwas wie Telekardiologie –, aber, ich glaube, wir sind uns einig, dass alleine die Definition dessen, was eine komplexe Intervention ist, dann schon wiederum sehr schwierig wird. Wir sind natürlich mit den Kollegen, die sich in diesem Bereich akademisch betätigen, also gerade dieses INTEGRATE-HTA-Projekt, im Austausch. Ich war selber 2015 bei dem Workshop, wo die Abschlussergebnisse vorgestellt wurden. Und auch die britischen Ergebnisse sind uns natürlich wohlbekannt. Ich gehe davon aus, dass auch die externen Sachverständigen, mit denen wir, egal in welchem Bereich, zusammenarbeiten, daraus auch Anleihen nehmen werden.

Moderator Jürgen Windeler: Jetzt kommt Ralf Bender zu dem indirekten Vergleich.

Ralf Bender: Genau. Da war die Frage eben, was wir machen, wenn es auf Basis einarmige Studien eine Zulassung gibt. Dann gibt es regelhaft keinen Nutzenbeleg. Wir brauchen den kontrollierten Vergleich. Wenn der nicht vorliegt, dann gibt es regelhaft daraus keinen Nutzenbeleg. Ausnahmen gibt es natürlich: dramatischer Effekt. Wenn sich so etwas abzeichnet, dann ist natürlich alles möglich. Aber regelhaft werden wir aus einarmigen Studien keinen Nutzenbeleg ableiten.

Désirée Fimm: Aber es gibt ja auch noch die EMA, die ja schon positiv bewertet hat. Das widerspricht sich ja dann. Wir verstehen nicht ganz, warum das Votum der EMA in dem Sinne übergangen wird.

Ralf Bender: Das ist ja kein Widerspruch. Wir machen ja nicht das Gleiche wie die EMA. Die EMA macht die Zulassung, und wir machen die Nutzenbewertung. Sonst wären wir ja überflüssig. Wenn aus der Zulassung gleich der Nutzenbeleg erwüchse, bräuchte man das IQWiG ja gar nicht.

Moderator Jürgen Windeler: Und andere HTA-Institutionen auch nicht. Wenn wir uns im Bereich von, was ja meistens von besonderem Interesse ist, AMNOG bewegen und weiteren Zusatznutzen bewegen, würden wir sowieso über völlig andere Dinge reden als die Zulassung bei der EMA. Und in der Situation von Nutzenbewertungen sind wir zwar auch, aber bei den Arzneimitteln relativ wenig.

Aber noch mal: Es gibt keinen Widerspruch zwischen Beleg des Zusatznutzens und der Zulassung der EMA, und es gibt auch keinen Widerspruch zwischen Nutzenbewertungen, die wir machen, und der Zulassung bei der EMA. Beides ist möglich.

Désirée Fimm: Das heißt, Sie würden jetzt sagen, dass sämtliche Methodenadjustierungen nicht ausreichen, um quasi eine RCT nachzubilden. Man kann ja auch einfach verschiedene Szenarien durchgehen, um zu gucken, ob die Annahmen in dem Modell, ob daraus die gleichen Ergebnisse resultieren. Und wenn sie es halt tun, dann scheint das Modell ja relativ robust zu sein. Man muss ja auch gemäß § 5 diese Studien einreichen. Die EMA macht ja trotzdem indirekt so eine Evaluierung in der Hinsicht. Wenn wir davon ausgehen, dass Nutzen auf jeden Fall da ist für die entsprechende Patientenpopulation, dann finden wir, dass

das ein bisschen mehr angeglichen werden müsste und Hintertüren offengehalten werden sollten.

Ralf Bender: Natürlich sind Verfahren, wo eine Kontrolle für Confounder noch mit gewissen Adjustierungsverfahren ermöglicht wird, besser, als wenn man rein naiv einarmige Studien miteinander vergleicht. Aber da man eben in der Praxis nie sicherstellen kann, dass man alle Confounder und alle Effektmodifikatoren kennt und auch gemessen hat, bleibt es dabei, dass regelhaft daraus kein Nutzenbeleg erwächst. Wenn wir eine grobe Evidenzhierarchie machen wollen, wären wir noch unterhalb von einer prospektiven Kohortenstudie, wo wenigstens die beiden Gruppen parallel untersucht werden. Wir brauchen kontrollierte Vergleiche für unsere Nutzenbelege regelhaft.

Friedhelm Leverkus: Das verstehen wir. Der Punkt ist natürlich, dass die EMA normalerweise ja auch immer randomisierte vergleichende Studien fordert. Es gibt Ausnahmesituationen, die zum Beispiel zu einem Conditional Approval führen, wo die EMA auf randomisierte kontrollierte Studien verzichtet, weil das Medikament so ein Durchbruch oder was weiß ich halt ist. Das ist eigentlich schade, dass solche Situationen gar nicht im Nutzenbewertungsprozess abgebildet werden können. Und diese MAIC hätten die Möglichkeit, die abzubilden, und haben natürlich sicherlich nicht so eine Aussagekraft – vollkommen klar – wie ein randomisiertes clinical trial.

Stefan Lange: Wichtig ist, weil das ja scheinbar durcheinandergeht: Die EMA macht was anderes als wir. Deswegen hilft es nicht, zu sagen, aber die EMA macht doch. Wir haben eine ganz andere Fragestellung und eine ganz andere Aufgabe als die europäische Zulassungsbehörde. Ob die EMA für die Frage der Nutzen-Risiko-Abwägung – da steht im Vordergrund: Kann man es verantworten, eine solche Substanz in die Versorgung zu bringen, und zwar im Wesentlichen aus Sicherheitsgründen? – ... Wenn die EMA der Ansicht ist, dass da gegebenenfalls einarmige Studien ausreichen, dann nehmen wir das so zur Kenntnis. Unsere Frage ist – wir reden jetzt über den Beleg, wir reden noch nicht einmal irgendwie über die höchste Kategorie der Sicherheit... -: Reicht es aus für die höchste Kategorie der Sicherheit eine Beobachtung von so 100 Patienten, dass es einen Zusatznutzenbeleg gibt? - Da muss ich ganz offen gestehen: Nein. Das ist auch eine ganz andere Frage. Jetzt kommen wir zur Hintertür. Das ist interessant. Wir sollen uns immer Hintertürchen offenlassen. Das müssen wir auch einmal ins Methodenpapier hineinschreiben. Natürlich gibt es eine solche Hintertür. Das hat Ralf Bender gerade gesagt. Wenn der Unterschied genügend groß ist, dann können selbst wir den nicht ausblenden. Aber da haben wir schon gewisse Anforderungen. Die kennen Sie aber auch, was wir darunter verstehen. Insofern: Theoretisch denkbar ist es schon. Wir wissen doch alle, was das jetzt für Studien sind im Conditional Approval, lieber Herr Leverkus, und welche Unterschiede wir dort sehen gegebenenfalls zu dem, was man vielleicht erwartet. Das ist überschaubar.

Friedhelm Leverkus: Herr Lange, ich möchte Ihnen zustimmen, das ist sicherlich kein Beleg. Da ist eine größere Unsicherheit mit bei. Das ist vollkommen klar. Aber es gibt natürlich Situationen, wo ... Die EMA macht wieder etwas anderes. Von der Entscheidungstheorie sagen die, wir machen ein Bridging oder solche Geschichten. Wir machen ja im Prinzip die Orphan Drugs, sind ja irgendwo halt abgebildet in so einer Situation. Und für die Medikamente, wo wir ein Conditional Approval haben, haben wir im Prinzip keine Abbildung. Ich meine, das ist jetzt nicht Ihre Aufgabe, Gesetze zu schreiben, die

das im Prinzip halt machen, aber, ich denke, wenn man so etwas machen will, wenn man nachverfolgt und sagt: „Okay, das ist jetzt noch mal was Besonderes“, dann wären solche Methoden sicherlich eine Möglichkeit, da eine Aussage zu schaffen, die noch besser ist als die nichtadjustieren Vergleiche.

Ralf Bender: Wenn wir eine schöne Netzwerkmetaanalyse haben mit adäquaten Brückenkomparatoren oder - eben auch genannt - adjustierter indirekter Vergleich, landen wir regelhaft bei einem Anhaltspunkt, also der dritten Stufe. In wenigen Ausnahmefällen, die in der Praxis aber noch nie vorgekommen sind, kann man auch bei einem Hinweis landen, aber in aller Regel landen wir bei einem Anhaltspunkt, der niedrigsten Stufe. Wenn wir darunter gehen, landen wir unter einem Anhaltspunkt. So ist das regelhaft.

Moderator Jürgen Windeler: Ich möchte ergänzen, weil Sie die Verbindung zu dem Conditional Approval hergestellt haben. Solche Situationen wie Anhaltspunkt, möglicherweise noch mit ein bisschen schwächlichen Daten, könnten beim G-BA dazu führen, dass er Beschlüsse befristet. Insofern ist es nicht sofern von dem Conditional Approval mit dem einzigen Unterschied, dass der G-BA bedauerlicherweise keine Sanktionsmöglichkeiten hat und keine Durchgreifsmöglichkeiten, die Conditional wirklich zu überprüfen, sondern regelhaft mit der Situation konfrontiert ist, dass nach Ende der Befristung keine neuen Daten vorliegen. Das ist auch ein bisschen so, wo man natürlich über die Jahre über die 150 oder 200 Bewertungen sich einen Eindruck verschafft, was mit dem, was man tut, passiert und wo man möglicherweise auf das, was am Ende herauskommt, dann am Anfang reagiert. Wenn wir schon wissen – und es ist gute Evidenz dafür –, dass es keine neuen Daten geben wird nach der Entscheidung des G-BA, sodass wir auf diese auch nicht bauen können, dass wir auch nicht Unsicherheit akzeptieren können, weil neue Daten kommen, dann ist klar, dass wir – da haben wir unsere Methodik jetzt gar nicht geändert – dann in der Frage des Bestehens auf guten Daten am Anfang noch etwas sorgfältiger gucken werden.

Katrin Kupas: Vielleicht ganz generell: Das ist ja im Moment ganz kategorisch ausgeschlossen, dass überhaupt irgendwelche Simulationen gemacht werden können. Im Moment gibt es ganz viele Bestrebungen, dass diese Verfahren weiter untersucht werden und dann auch sichergestellt wird, sie halten die Power, sie halten vor allem den Alphafehler, und das vielleicht ein bisschen offener gelassen wird. Das werden wir dokumentieren über eine Publikation, dass so ein Verfahren geeignet ist, einen historischen Vergleich vernünftig abzubilden, das es dann schon akzeptiert werden würde. Da sind wir natürlich in der Pflicht, dann entsprechend Untersuchungen zu machen und zu zeigen, der Alphafehler ist nicht verletzt und die Methode funktioniert und justiert auch vernünftig.

Ralf Bender: Wenn man eine Situation macht, also sich eine Wahrheit für eine Simulation denkt, und dort einen wichtigen Effektmodifikator einbaut, dann aber so tut, als ob man ihn nicht kennt, und ihn dann weglässt und nicht berücksichtigt, dann brauchen wir nicht viel Simulation, um zu zeigen, dass da der Typ-1-Fehler auf jeden Fall verletzt ist. Man wird keine Simulationsstudie hinkriegen, wo man belegen kann, dass man Power und Niveau einhalten kann auch ohne relevante Effektmodifikatoren.

Ch.-Markos Dintsios: Ein Kommentar zu Ihrer Aussage, Herr Windeler: Wir hatten auch Befristungen beim G-BA mit der Auflage der Durchführung von klinischen Studien, die sich

auf ein Zeitfenster von 12 bis 18 Monaten beliefen, wo am Anfang der G-BA selber nicht mal einsichtig war, aber im Nachgang verstanden hat, dass die Durchführung einer klinischen Studie innerhalb dieses Zeitfensters nicht machbar ist. Und siehe da, in den letzten zwei, drei Jahren sehen wir auch den G-BA selber dabei, dass er eine Befristung verlängert. Das ist ein lernendes System. Also hier kann man das Argument durchaus aufgreifen. Aber Sie haben es etwas monolithisch dargestellt, also das Daten nicht geliefert werden in allen Fällen oder dass der G-BA nicht die Sanktionsmöglichkeit einer EMA oder einer nationalen Zulassungsbehörde hat. Manchmal waren halt auch die Auflagen des G-BA vielleicht unrealistisch, zu erwartungsvoll und liefen an der Realität vorbei.

Moderator Jürgen Windeler: Wobei meine beiden Aussagen, erstens, dass der G-BA keine Sanktionsmöglichkeiten hat, und zweitens, dass nach bisherigen Erfahrungen fast nie, sehr selten neue Daten - ich habe nicht einmal von neuen Studien gesprochen – nach den Befristungen vorgelegt worden sind - vorsichtig ausgedrückt –, so falsch nicht waren. Dass die Fristen, die der G-BA gesetzt hat, für Studien, mit denen man jetzt längerfristige Endpunkte betrachten will, kurz waren, gebe ich zu. Teilweise sind sie befristet worden, weil Studien bereits liefen. Das will ich aber nicht thematisieren. Ich will nur sagen: Der G-BA hat Befristungen gesetzt, und neue Daten gab es nicht, nicht nie, aber ziemlich selten, und Sanktionsmöglichkeiten hat er nicht. Ob die EMA die praktisch hätte, ist eine ganz andere Frage, aber der G-BA hat sie jedenfalls nicht.

Andreas Bätzel: Medizintechnik kam bisher relativ wenig vor. Ich hatte in der Einladung vernommen, dass über das diskutiert wird, was tatsächlich jetzt verändert wurde in der Version 5.0. Nicht verändert wurden einige Aspekte, die eben nichtmedikamentöse Verfahren einschließen. Daher eine eher offene Frage in Richtung IQWiG, ob es Ihrerseits den Gedanken, möglicherweise Bestrebungen gibt, all das, was an eben nichtmedikamentösen Eingriffen nutzenbewertet wird, potentialbewertet wird, auch noch mal eingehender zu spezifizieren, das heißt, auf Einwände, die mehr oder weniger aus dem Bereich der Medizintechnik kommen, in einem Dialog einzugehen und zu gucken, wenn ihr das irgendwie nicht hinkriegt, liegt es daran, dass ihr das nicht versteht, oder geht es nicht?

Stefan Sauerland: Den letzten Punkt habe ich nicht ganz verstanden, was da nicht gehen soll oder was nicht hinzukriegen ist. Vom Grundsatz her stehen wir auf dem Standpunkt, dass medizintechnische Interventionen in sehr ähnlicher oder fast gleicher Weise bewertet werden können wie alle anderen Interventionen. Wir haben ja in dem Kapitel zu nichtmedikamentösen Interventionen ein ganzes Bündel von möglichen Maßnahmen. Das eine ist die Medizintechnik – sicherlich –, dass andere ist aber dann Chirurgie oder Zahnheilkunde, oder wir haben jetzt Psychotherapieevaluation. Das sind alles Dinge, die man bewerten kann und wo wir vor allen Dingen sehen, dass hier keine spezifische Bewertungsmethodik in irgendeiner Weise abgeändert werden muss. Es gibt natürlich immer wieder diese Punkte mit den Lernkurveneffekten bei Interventionen, wo der Therapeut ein wesentlicher Teil der Intervention ist, und es gibt auch immer das Problem mit der Verblindung. Das ist alles bekannt. Ich finde es immer schwierig, wenn – das war auch kürzlich in der Presse, von BVMed zum Beispiel – die Forderung aufgestellt wird, man bräuchte irgendeine ganz neue Methodik oder man muss grundsätzlich Tabula rasa machen und dann überlegen, wie man Medizintechnik bewerten kann. Das ist aus unserer Sicht nicht notwendig.

Andreas Bätzel: Noch einmal ergänzend zu den Punkten, die Sie eben nannten, mit Lernkurveneffekten: Hinzu kommt halt etwas, was wir schon als grundlegend unterschiedlich sehen zu der Anwendung Ihrer Methodik auf den Arzneimittelbereich, nämlich die zwei Aspekte, dass Medizintechnik als Produkt immer Teil einer Methode ist und insofern auch Teil von einem Bündel von gleichzeitig angewendeten Medizintechniken oder anderen Verfahren unter Einflussnahme von Anwendern ist als Gesamtkomplex. Da sehen wir einen methodischen Unterschied, der sich da ergeben müsste. Zum anderen ist die Zweckorientierung von Medizintechnik nicht unbedingt der patentrelevante Endpunkt. Von daher sehen wir da Schwierigkeiten – so erzählen es uns die betroffenen Firmen –, das entsprechend herunterzubrechen über Surrogate oder wie auch immer.

Barbara Behle: Ich möchte das, was der Kollege von ZVEI sagt, noch einmal aufgreifen. Uns geht es ja nicht darum, dass wir sagen, wir wollen keine Nutzenbewertung. Uns geht es nur darum, dass die Besonderheit von Medizinprodukten in der Nutzenbewertung gesehen wird. Wir haben ganz andere Innovationszyklen. Wir haben eine ganz andere Regidität. Darüber hinaus sind es, was auch der Kollege von ZVEI schon sagte, Methoden. Da kommt es immer darauf an: Ich verwende ein Medizinprodukt innerhalb einer Methode – deshalb auch mein Eingangsstatement mit der Bitte, das wirklich gemäß der Terminologie des Gesetzgebers zu verwenden. Da spielen zum Beispiel Erfahrungen eines Chirurgen etc. pp. ... Das sind ganz entschiedene Erfolgsfaktoren. Da können Sie nicht für einen Blister von der Pille sagen: „Das habe ich jetzt. Die muss jetzt so angewendet werden, dreimal am Tag oder einmal am Tag“, sondern das sind halt sehr komplexe Vorgänge. Insofern möchte ich das nur unterstreichen, was mein Vorredner gesagt hat.

Moderator Jürgen Windeler: Ich würde jetzt gerne eine Diskussion insbesondere mit den Industrieverbänden unterbrechen, die in die Richtung läuft: Machen wir hier Nutzenbewertung oder nicht, und wie machen wir die? Das können wir an anderen Stellen diskutieren. Sie wissen auch, dass das IQWiG dazu eine etwas andere Position hat.

Den Umstand, den Sie beide beschreiben, nämlich dass die Besonderheiten der Anwendung von Medizinprodukten innerhalb von Methoden bei der Bewertung berücksichtigt werden sollten, den berücksichtigen wir auch, den berücksichtigen wir übrigens bei Arzneimitteln ganz genauso.

Auf andere Argumente will ich jetzt hier nicht eingehen. Das kann ich an anderer Stelle machen. Das ist hier nicht der richtige Ort.

Cordula Rapp: Es geht jetzt nicht um das gleiche Thema, was Sie eben abgewürgt haben, sondern ich habe eine ganz konkrete Nachfrage zu der Auswahl der Methoden nach 139b Abs. 5 SGB V, nicht durch den Prozess der Auswahl, sondern da werden ausdrücklich die Arzneimittel ausgeschlossen. Jetzt frage ich mich gerade: Haben Sie die Formulierung so gewählt, weil Sie spezifisch einzelne Medizinprodukte bewerten wollen innerhalb dieses Verfahrens? Denn das sind ja auch immer Teile von Verfahren oder von Präventionsmaßnahmen. Wir hatten ja eben darüber geredet, dass auch diese Präventionsmaßnahmen darunter fallen können, also die patientenbetroffeneninitiierten Methoden, die Sie bewerten.

Stefan Lange: Das berührt ein bisschen die Problematik, die wir schon eben hatten. Hier ist im Grunde genommen der Gesetzestext relativ klar, weil die Arzneimittel in einem eigenen Aufzählungspunkt im 139b Abs. 5 referenzierten, also auf die in 139a Abs. 3 referenzierten Punkte genau abzielt. Da ist explizit die Arzneimittelbewertung nicht aufgeführt. Das lässt die Interpretation zu, und sie ist uns auch vom Gesetzgeber bestätigt worden, dass das nicht aus Versehen weggelassen worden ist, sondern mit Absicht. Es soll eben in erster Linie keine Arzneimittelbewertung – so ist das Gesetz zu verstehen; das können wir, glaube ich, hier so sagen; das ist unsere Auskunftsmöglichkeit – sein. Das heißt aber nicht, dass es eine Bewertung von Medizinprodukten oder Medizintechnik sein soll, sondern es ist relativ offen, was dort in den anderen Punkten steht, auf die referenziert wird, aber explizit in erster Linie keine Arzneimittelbewertung.

Moderator Jürgen Windeler: Wenn man die Gesetzesexegese noch ein kleines Stück weitertreibt, dann heißt es eigentlich dort, dass die Nutzenbewertung von Arzneimitteln nicht gemacht werden soll im Rahmen der Themencheckbewertung. Das bedeutet nicht, dass Arzneimittel zum Beispiel als sehr relevante Vergleichstherapien bei irgendwelchen Indikationen nicht in die Bewertung einbezogen werden können. Das werden wir tun. Wir sollen – so sagt es das Gesetz – keine expliziten oder vielleicht auch isolierten Nutzenbewertungen von Arzneimitteln machen. Dazu gibt es naheliegenderweise andere Verfahren.

Cordula Rapp: Die Erläuterung war sehr wichtig für mich, auch als Person, um das besser zu verstehen. Aber die Frage ist eher gewesen, warum das ausgeschlossen wurde. Das es so ist, war mir klar. Mir geht es darum: Soll der Umkehrschluss daraus sein, dass andere Maßnahmen aufgesplittet werden, eben nicht komplex als Maßnahme, Methode in diesem Fall bewertet werden, also Teilschritte bewertet werden im Rahmen dieser Vorschläge?

Moderator Jürgen Windeler: Also noch mal: Warum das ausgeschlossen worden ist, verweise ich an Berlin, Friedrichstraße 108. Wir haben das nicht ausgeschlossen, der Gesetzgeber hat das ausgeschlossen.

Cordula Rapp: Das war wohl ein Missverständnis. Das Gesetz habe ich gelesen, habe ich verstanden. Es ging eher darum: Die Formulierung kam mir so vor, als wäre auf etwas anderes abgezielt worden. Dann entschuldige ich mich. Dann ist es vielleicht ein Versehen gewesen, oder ich habe es missinterpretiert.

Moderator Jürgen Windeler: Das Gesetz hat eine bestimmte Formulierung und verweist auf Aufgaben 139a. Die Abstimmung mit dem Gesetzgeber hat ergeben, dass damit gemeint ist: Nutzenbewertungen von Arzneimitteln sollen in 139b nicht gemacht werden, weil – naheliegend – es dafür viele andere Verfahren im SGB V gibt. Das bedeutet aber nicht – auch so die Interpretation des Gesetzgebers –, dass wir nicht in die Themencheckbewertungen Arzneimittel einbeziehen können. Eine andere Intention von Arzneimitteln weg auf einen anderen Fokus habe ich dem Gesetz und auch der Interpretation des Gesetzes nicht entnommen. Frage damit beantwortet?

Cordula Rapp: Ja.

Moderator Jürgen Windeler: Und ich habe die Diskussion am Anfang nicht abgewürgt – ich bitte um Verständnis –, sondern ich habe sie auf ein anderes Forum verlegt. Wir sitzen auf allen möglichen Podiumsdiskussionen miteinander und unterhalten uns über die Sinnhaftigkeit von Innovationszyklen und Nutzenbewertungen. Das müssen wir hier nicht auch noch machen.

Gibt es weitere Wortmeldungen? – Das sehe ich nicht. Dann bedanke mich bei allen, dass Sie da waren. Ich bedanke mich für Ihre vielen konstruktiven, lebendigen, kritischen Beiträge. Ich lade Sie zu einem abschließenden oder auftaktigen Imbiss ein und wünsche Ihnen einen guten Heimweg. Vielen Dank.

Anhang A – Dokumentation der Stellungnahmen

Inhaltsverzeichnis

	Seite
A.1 – Stellungnahmen von Organisationen, Institutionen und Firmen	A 3
A.1.1 – Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF)	A 3
A.1.2 – AstraZeneca GmbH.....	A 8
A.1.3 – Bayer Vital GmbH.....	A 17
A.1.4 – Boehringer Ingelheim Pharma GmbH & Co. KG.....	A 24
A.1.5 – Bristol-Myers Squibb GmbH & Co. KGaA	A 29
A.1.6 – Bundesarbeitsgemeinschaft SELBSTHILFE von Menschen mit Behinderung und chronischer Erkrankung und ihren Angehörigen e. V. (BAG SELBSTHILFE).....	A 35
A.1.7 – Bundesverband der Arzneimittel-Hersteller e. V. (BAH).....	A 41
A.1.8 – Bundesverband der Pharmazeutischen Industrie e. V. (BPI)	A 46
A.1.9 – Bundesverband Medizintechnologie e. V. (BVMed)	A 52
A.1.10 – Deutsche Diabetes Gesellschaft e. V. (DDG) / Deutsche Gesellschaft für Chirurgie e. V. (DGCH) / Deutsche Gesellschaft für Epileptologie e. V. / Deutsche Gesellschaft für Gynäkologie und Geburtshilfe e. V. (DGGG) / Deutsche Gesellschaft für Hämatologie und Medizinische Onkologie e. V. (DGHO) / Deutsche Gesellschaft für Kardiologie – Herz- und Kreislaufforschung e. V. (DGK) / Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e. V. (GMDS) / Deutsche Gesellschaft für Neurologie e. V. (DGN) / Deutsche Gesellschaft für Pneumologie und Beatmungsmedizin e. V. (DGP) / Deutsche Gesellschaft für Rheumatologie e. V. (DGRh) / Deutsche Region der Internationalen Biometrischen Gesellschaft (IBS-DR) / (gemeinsame Stellungnahme).....	A 57
A.1.11 – Deutsche Gesellschaft für Hebammenwissenschaft e. V. (DGHWi)	A 62
A.1.12 – Deutsche Gesellschaft für Humangenetik e. V. (GfH)	A 68
A.1.13 – Deutsche Gesellschaft für Innere Medizin e. V. (DGIM).....	A 70
A.1.14 – Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e. V. (GMDS) / Deutsche Region der Internationalen Biometrischen Gesellschaft (IBS-DR) (gemeinsame Stellungnahme)	A 78
A.1.15 – Deutsche Gesellschaft für Orthopädie und Unfallchirurgie e. V. (DGOU) A	92
A.1.16 – Deutsche Gesellschaft für Perinatale Medizin (DGPM)	A 94
A.1.17 – Deutsche Krankenhausgesellschaft e. V. (DKG)	A 96
A.1.18 – Deutsches Netzwerk für evidenzbasierte Medizin e. V. (DNEbM)	A 101
A.1.19 – Ecker + Ecker GmbH.....	A 105

A.1.20 – GlaxoSmithKline GmbH & Co. KG (GSK)	A 117
A.1.21 – Institut für Gesundheits- und Pharmaökonomie (IfGPh)	A 126
A.1.22 – Jansen-Cilag GmbH	A 133
A.1.23 – Kassenzahnärztliche Bundesvereinigung K. d. ö. R. (KZBV).....	A 138
A.1.24 – Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen e. V. (MDS)	A 146
A.1.25 – MSD SHARP & DOHME GmbH	A 155
A.1.26 – Novartis Pharma GmbH.....	A 164
A.1.27 – Pfizer Deutschland GmbH.....	A 172
A.1.28 – Roche Diabetes Care GmbH.....	A 176
A.1.29 – Roche Pharma AG.....	A 179
A.1.30 – Sanofi-Aventis Deutschland GmbH.....	A 183
A.1.31 – SmartStep Consulting GmbH.....	A 197
A.1.32 – SPECTARIS – Deutscher Industrieverband für optische, medizinische und mechatronische Technologien e. V. / ZVEI – Zentralverband Elektrotechnik- und Elektronikindustrie e. V. (gemeinsame Stellungnahme)	A 201
A.1.33 – Verband forschender Arzneimittelhersteller e. V. (vfa)	A 205
A.2 – Stellungnahmen von Privatpersonen.....	A 222
A.2.1 – Dintsios, Charalabos-Markos	A 222

A.1 – Stellungnahmen von Organisationen, Institutionen und Firmen

A.1.1 – Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF)

Autoren:

- Kopp, Ina
- Nothacker, Monika
- Selbmann, Hans Konrad

Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V.	<i>Association of the Scientific Medical Societies in Germany</i>
---	---



**Stellungnahme
der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften
(AWMF)**

zum Entwurf des IQWiG–Methodenpapiers:
„Allgemeine Methoden - Entwurf für Version 5.0“ vom 07.12.2016

Die AWMF wurde am 07.12.2016 über die Möglichkeit zur Stellungnahme zum oben genannten Entwurf gebeten. Die AWMF hat ihrerseits ihre Mitgliedsfachgesellschaften gebeten, bei gegebenem Bedarf eine eigene Stellungnahme zu verfassen. Die bis einschließlich 27.01.2017 bei der AWMF eingegangenen fünf Stellungnahmen von insgesamt 15 Fachgesellschaften (eine gemeinsame Stellungnahme von 11 Fachgesellschaften, die alle in der ad hoc Kommission „Frühe Nutzenbewertung“ der AWMF vertreten sind) sind dieser Stellungnahme der AWMF als Anlage beigefügt. In diesen Stellungnahmen finden sich auch Kommentare zu Aspekten, die im Folgenden nicht adressiert sind und die wir zu berücksichtigen bitten.

Wir bitten besonders auch jene Kommentare zu bedenken, die sich auf nicht vom IQWiG überarbeitete Abschnitte beziehen. Wir regen an, eine komplette Überarbeitung der Allgemeinen Methoden zu prüfen. In diesem Zusammenhang weisen wir auch auf zum Teil veraltete Zitate zu evidenzbasierter Medizin bzw. deren Strategien hin.

1. Allgemeine Anmerkungen

Die AWMF begrüßt die detaillierte und weitgehend allgemeinverständliche Darlegung der Methoden des IQWiG. Wir halten neben methodischer Transparenz das Einbringen von wissenschaftlicher Expertise für ein hohes Gut zur Bewertung von Interventionen. Wir verweisen hierzu auch auf unsere Stellungnahme von 2014 zum Entwurf der „Allgemeinen Methoden Version 4.2“ vom 18. 06. 2014 (siehe: <http://www.awmf.org/die-awmf/awmf-stellungnahmen.html>).

Erstmals wurde in dieser Version der Begriff „normativ“ eingeführt, was zu fremdartigen Termini wie „normative Standards“ (Tautologie) und „normative Empfehlungen“ (Oxymoron) führte. Wir schlagen vor, diese Einführung generell rückgängig zu machen.

2. Einbeziehen des wissenschaftlichen-medizinischen Sachverstands

In Bezug auf die Möglichkeiten des Einbringens wissenschaftlicher Expertise sehen wir weiterhin wichtigen Verbesserungsbedarf. Die AWMF hat nach wie vor Sorge, dass relevante wissenschaftliche Inhalte nicht oder erst sehr spät in die Diskussion aufgenommen werden können. So erhält das IQWiG zum Beispiel zu Gesundheitsinformationen regelmäßig zahlreiche Kommentare, die auch regelhaft zu inhaltlichen Änderungen führen.

Umsomehr verwundert es uns, dass im neuen Abschnitt 2.1.1 die Option zum Verzicht auf eine Anhörung zum Berichtsplan „in bestimmten Fällen“ eingeführt wurde.

Wenn keine Änderung des geplanten methodischen Vorgehens im Berichtsplan nötig war, sollten zudem die Stellungnehmenden über diesen Sachverhalt informiert werden.

3. Ergänzung der Ausführungen zur Bewertung gemäß § 137h SGB V

Das IQWiG führt in Abschnitt 2.1.6 aus, die Bewertung von neuen wissenschaftlichen Methoden zu Medizinprodukten hoher Risikoklassen entsprechend des Vorgehens einer Potenzialbewertung vorzunehmen. Wir empfehlen diesbezüglich, die Vorschläge einer internationalen Expertengruppe zur Evaluation von Medizinprodukten mit unterschiedlich „sicheren“ Studiendesigns auf die Anwendbarkeit für die erforderlichen Bewertungen zu prüfen (IDEAL-Statement¹)

Im Rahmen der Potentialbewertung wurde bisher erst eine vorgeschlagene Studie begonnen zu realisieren und ein groß angelegter RCT zum Vergleich verschiedener interventioneller Verfahren beim Prostatakarzinom musste trotz methodischer Beteiligung des IQWiG mangels Rekrutierung beendet werden.

4. Ergänzungen zu Ausführungen zu Nutzenbewertungen medizinischer Interventionen zu Endpunktbezogener Bewertung (3.1.4)

Wenngleich wir dem IQWiG zustimmen, dass eine ausreichende Fallzahl gegeben sein muss, um Unterschiede in Subgruppen sicher identifizieren zu können, ist die vorgeschlagene Festlegung auf mindestens 1000 arbiträr. Stattdessen sollten angemessene Fallzahlberechnungen durchgeführt werden.

5. Überarbeitung des Kapitels 5 Leitliniensynopsen und Versorgungsanalysen

In Kap. 5.1. Hintergrund wird ausgeführt, dass „Leitlinien, wenn sie in einem systematischen Verfahren erstellt werden und der Erstellungsprozess transparent dokumentiert ist, normative Standards für alle Bereiche der Versorgungskette enthalten [können] [297].“ Es wird weiterhin von normativen Empfehlungen gesprochen. Dieser Aussage stimmt die AWMF nicht zu. Leitlinienempfehlungen sind nicht normativ, sondern als Entscheidungshilfen zu verstehen, deren Anwendbarkeit in der individuellen Situation geprüft werden muss. Wir verweisen hier auch auf die Stellungnahme der AWMF zu Normierungsvorhaben des Europäischen Komitees für Normung².

Das Kapitel 5 handelt indes nicht von Leitlinien, sondern – gemäß der geänderten Überschrift – von Leitliniensynopsen. Es wird in keiner Weise dargelegt, warum Leitliniensynopsen sich besser als Grundlage für DMP Empfehlungen eignen als eine einzelne, multiprofessionell erstellte Nationale VersorgungsLeitlinie, die den Besonderheiten des deutschen Gesundheitssystems gerecht wird. Aus Sicht der AWMF sollte dringend eine unabhängige Kosten-Nutzen-Analyse zum Verfahren der Leitliniensynopsen des IQWiG durchgeführt werden, auch in Bezug auf die erforderliche Bearbeitungszeit und den darauf resultierenden Verlust an Aktualität.

¹ McCulloch P, Altman DG, Campbell WB, et al. No surgical innovation without evaluation: the IDEAL recommendations. Lancet. 2009;374(9695):1105-1112.

² Stellungnahme der AWMF zu Normierungsvorhaben des Europäischen Komitee für Normung (Comité Européen de Normalisation - CEN) vom 15. 5. 2014. Verfügbar: <http://www.awmf.org/die-awmf/awmf-stellungnahmen.html>

Wir verstehen den Wunsch des IQWiG, die unterschiedlichen Systeme zur Graduierung von Empfehlungen und Evidenz zu vereinheitlichen. Die gewählte Form der Dichotomisierung in „hoch“ und „niedrig“ ist aber nicht begründet und kann die Intention von Leitlinienautoren verzerren. Für die Nutzer resultieren Informationsverluste und die Gefahr von Fehlinterpretationen.

Die Festlegung der Relevanzkriterien für ein DMP, nur bezogen auf die Empfehlungsstärke und die inhaltliche Konsistenz von Leitlinienempfehlungen verschiedener Organisationen und ohne Berücksichtigung der vorab mit hohem Aufwand vorgenommenen AGREE-Bewertung können wir nicht nachvollziehen. Eine Begründung für dieses Vorgehen ist im Methodenpapier nicht hinterlegt.

Zum Abschnitt „Versorgungsanalyse“ ist grundsätzlich die Relevanz dieses Produkts zu hinterfragen. Unserer Kenntnis nach erstellt das IQWiG in praxi keine Versorgungsanalysen. Die Aufarbeitung von Versorgungsdaten bzw. die Entwicklung von Messgrößen zur Beurteilung der Versorgungsqualität fällt in den Aufgabenbereich des IQTIG (Institut für Qualität und Transparenz im Gesundheitswesen).

5. Ergänzung des Kapitels 6 über HTA-Berichte gemäß § 139b Absatz 5 SGB V

HTA Themen können laut des Entwurfs von interessierten Einzelpersonen vorgeschlagen werden. Die AWMF regt an, auch Fachgesellschaften - oder anderen Organisationen im Gesundheitswesen - ein Vorschlagsrecht einzuräumen.

Für die Auswahl der jährlich eingegangenen Themenvorschläge werden anerkannte Kriterien benannt, die laut Verfahren zunächst vom IQWiG beurteilt werden sollen.

Die weitere Auswahl von zu bearbeitenden HTA Themen ist allerdings unvollständig beschrieben in Hinblick auf die in den nächsten beiden Schritten anzulegenden Auswahlkriterien und nicht nachvollziehbar in Hinblick auf das Bewertungs- und Entscheidungsgremium in Auswahlstufe 2.

Wir gehen davon aus, dass in der Auswahlstufe 1 (Abb. 17) die im einleitenden Abschnitt des Kapitels 6.3. avisierte Einbeziehung der wissenschaftlichen Perspektive durch einen Vertreter der AWMF im Auswahlbeirat realisiert wird.

Die AWMF hat Zweifel, dass die ethischen und sozialen Aspekte, die entsprechend Kap. 6.5.3 und 6.5.4 regelhaft in HTA-Projekten adressiert werden sollen, ausschließlich fragebogengestützt bearbeitet werden können. Hier muss sicher in einer Erprobungsphase zunächst Erfahrung gesammelt werden.

Für Rückfragen stehen wir gern zur Verfügung

Dr. med. Monika Nothacker, MPH; [REDACTED]
 Prof. Dr. med Ina Kopp; [REDACTED]
 Prof. Dr. Hans-Konrad Selbmann, [REDACTED]

Anhang:

- Gemeinsame Stellungnahme von Fachgesellschaften in der Ad hoc Kommission „Nutzenbewertung“ der AWMF: DGHO mit DDG, DGCH, DGGG, DGK, DGN, DGP, DGRh, DGVS, GMDS/IBS
- Stellungnahme der DGCH und DGBMT (Ass. Mitglied)
- Stellungnahme der DGHWI
- Stellungnahme der DGIM
- Stellungnahme der DGOU

A.1.2 – AstraZeneca GmbH

Autoren:

- Merens, Chantal

Schriftliche Stellungnahme zu IQWiG Allgemeine Methoden, Entwurf für Version 5.0 vom 21.12.2016

AstraZeneca nimmt Stellung zu folgenden Abschnitten der Allgemeinen Methoden des IQWiG:

2.2.3B Gewährleistung der Unabhängigkeit externer Sachverständiger

2.2.7 Wissenschaftliche Beratung

3.1.4 Endpunktbezogene Bewertung

9.3.8 Meta-Analysen

9.3.9 Indirekte Vergleiche

9.3.10 Subgruppenanalysen

1. Transparenz bei der Beratung durch externe Sachverständige

IQWiG Allgemeine Methoden

S. 37 / 2.2.3 B Gewährleistung der Unabhängigkeit externer Sachverständiger

Die Namen der externen Sachverständigen, die an der Erstellung der Produkte des Instituts beteiligt waren, werden in der Regel in den entsprechenden Produkten veröffentlicht. Diese Veröffentlichungen sind grundsätzlich über die Website frei zugänglich.

Position AstraZeneca

Im Sinne der Transparenz und Nachvollziehbarkeit ist es erforderlich im Methodenpapier diejenigen Sachverhalte genau zu konkretisieren, an denen externe Sachverständige beteiligt werden müssen, können bzw. nicht beteiligt werden. Unter Berücksichtigung der bereits im zurückliegenden Stellungnahmeverfahren zum Entwurf der Version 4.2 der Allgemeinen Methoden geführten Diskussionen, sind zumindest der Fragenkatalog respektive die Thematik der Beratung, unter Wahrung der Vertraulichkeit gewisser Informationen und Dokumente, zu veröffentlichen.

Vorgeschlagene Änderung

Die vorangehend dargelegte methodische Darstellung bzw. Veröffentlichung der Thematik der externen Beratung durch Sachverständige sollte sowohl bei den Produkten Kosten-Nutzen-Bewertung und Potenzialbewertung als auch bei den Institutsprodukten Berichte, Rapid Report und Dossierbewertung umgesetzt werden.

2. Wissenschaftliche Beratung

IQWiG Allgemeine Methoden

S.41 / 2.2.7 Wissenschaftliche Beratung

Das Institut beteiligt sich in speziellen Fällen an der wissenschaftlichen Beratung (Scientific Advice) von Studiensponsoren, z. B. im Rahmen der Mitarbeit im europäischen Netzwerk EUnetHTA. Primäres Ziel ist dabei die Unterstützung des Designs von Studien, die aussagekräftige Daten für die Nutzenbewertung liefern. Um die Unabhängigkeit der Bewertung auch in den Fällen zu gewährleisten, in denen die Nutzenbewertung Studien enthält, zu denen das Institut eine Beratung durchgeführt hat, wird eine angemessene organisatorische Trennung von Beratung und Bewertung sichergestellt.

Position AstraZeneca

Frühe Beratungen im Rahmen der Planung von klinischen Studien sind aus Sicht von AstraZeneca eine sehr wichtige und seit Längerem bereits auch regelmäßig (beim G-BA) in Anspruch genommene Möglichkeit. Das IQWiG bestimmt als ein bevorzugter Partner des G-BA die wissenschaftlich-methodischen Standards für die Nutzenbewertung maßgeblich mit und verfügt daher über eine entsprechend hohe Kompetenz, um sich künftig auch selbst in wissenschaftlichen Beratungen einzubringen.

AstraZeneca begrüßt deshalb ausdrücklich dieses Vorhaben des IQWiG. Gleichzeitig weisen wir jedoch darauf hin, dass im Methodenpapier, Abschnitt 2.2.7, die möglichen Beratungsinhalte weiter konkretisiert werden sollten, zumal bestimmte (wichtige) Elemente des Studiendesigns von den Anforderungen und Vorgaben des G-BA abhängig sind und eben nicht vom IQWiG festgelegt werden (z.B. zweckmäßige Vergleichstherapie). Insofern sollte im Methodenpapier des Weiteren auch klargelegt werden, dass das IQWiG solche wissenschaftlichen Beratungen, sofern diese relevant mit Blick auf spätere Nutzenbewertungen nach §35a SGB V sind, grundsätzlich im Benehmen mit dem G-BA durchführt und diesem auch die Beratungsniederschrift zur Kenntnis weiterleitet - denn für den betroffenen Hersteller wird am Ende entscheidend sein, inwieweit der G-BA die betreffende Studie tatsächlich für die Nutzenbewertung heranzieht. Dies ist auch deshalb von Bedeutung, weil das IQWiG auch heute schon teilweise zu anderen Einschätzungen gelangt als der G-BA, einschließlich auch eigener wissenschaftlicher Vorgehensweisen, die vom G-BA nicht übernommen wurden (z.B. die Methodik des IQWiG zur endpunktbezogenen Bestimmung des Ausmaßes des Zusatznutzens).

Im Übrigen könnte im Methodenpapier zusätzlich auch der Anspruch dokumentiert werden, solche wissenschaftlichen Beratungen möglichst immer im Beisein bzw. unter Beteiligung von Vertretern der Zulassungsbehörden durchzuführen. Gerade in wissenschaftlich-methodischen Fragen des Studiendesigns (z.B. Endpunkt-Operationalisierung, Beobachtungsdauer, Erhebung von Lebensqualität etc.) sind aus Herstellersicht - mittelfristig - möglichst weitgehend harmonisierte Anforderungen an Studien wünschenswert; gemeinsame Beratungen zu konkreten Fragen können das gegenseitige Problemverständnis hierfür fördern. Die allgemeinen Rahmenbedingungen stützen diesen Vorschlag ebenfalls: Der G-BA arbeitet bereits regelhaft mit den deutschen Zulassungsbehörden zusammen (insbesondere auch hinsichtlich Beratungen), und auf EU-Ebene ist die Vertiefung der Kooperation zwischen EMA und EUnetHTA ebenfalls ein Schwerpunkt der Joint Action 3 und des HTA Network.

3. Zusätzliche Anforderung an die Ableitung eines Belegs aus einer Studie

IQWiG Allgemeine Methoden

S. 51 / 3.1.4 Endpunktbezogene Bewertung

Aus lediglich 1 Studie kann im Ausnahmefall für einen Endpunkt ein Beleg für eine spezifische (Teil-)Population abgeleitet werden, wenn ein Studienbericht gemäß 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 mindestens 1000 Patienten eingeschlossen und es gibt mindestens 10 Zentren.*
- *Der beobachtete Effektschätzer hat einen sehr kleinen zugehörigen p-Wert ($p < 0,001$).*
- *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 an 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.*

Position AstraZeneca

Die Ergänzung der Anforderungen an die Ableitung eines Belegs basierend auf einer Studie um das Kriterium „mindestens 1000 Patienten pro Studienarm“ ist nicht nachvollziehbar. Zum einen besteht bereits das Kriterium, dass der beobachtete Effektschätzer einen sehr kleinen zugehörigen p-Wert ($p < 0,001$) haben muss, wobei die statistische Signifikanz durch die Gruppengröße und das Ausmaß des Effektes direkt beeinflusst wird. Dieser geforderte p-Wert von $< 0,001$ ist bereits deutlich konservativer als der von der EMA im Rahmen von Zulassungen basierend auf einer Studie geforderte und etablierte p-Wert von $< 0,05$. (CPMP, 2001) Der geforderte p-Wert ist somit ein bereits hinreichendes Kriterium zur Bewertung der Evidenz. Zum anderen wird nicht begründet, wieso gerade 1000 Patienten pro Arm als adäquat angenommen und gefordert werden. Die zusätzliche Forderung von 1000 Patienten pro Studienarm schließt gewisse Indikationen schlichtweg a priori aus, in denen Studien mit mindestens 2000 Patienten (2-armige RCT) aufgrund der Seltenheit der Zielerkrankung nicht realisierbar sind.

Stattdessen wäre es mit Blick auf indikationsspezifische Unterschiede von Mehrwert für die Nutzenbewertung, wenn eine Einordnung der vorgelegten Studie in die bisherige Gesamt-Evidenz zu dem betreffenden Anwendungsgebiet vorgenommen und damit ihr relativer Informationswert entsprechend gewürdigt würde.

Vorgeschlagene Änderung

Es soll von der zusätzlichen Anforderung „mindestens 1000 Patienten pro Studienarm“ abgesehen werden.

4. Inkonsistenz in Bezug auf Mindestzahl an Studien

IQWiG Allgemeine Methoden

S. 188 & 190 / 9.3.8 Meta-Analysen

*Die Anwendung von Meta-Analysen mit zufälligen Effekten stößt jedoch im Fall sehr weniger Studien (**weniger als 5**) an ihre Grenzen. Da im Fall sehr weniger Studien die Heterogenität nicht verlässlich schätzbar ist [316], kann die Anwendung von Meta-Analysen mit zufälligen Effekten zu sehr breiten Konfidenzintervallen führen, die möglicherweise keine Aussagen zur Evidenzlage mehr zulassen.*

...

*Die Anwendung von Meta-Analysen mit zufälligen Effekten und zugehörigen Prädiktionsintervallen im Fall von sehr wenigen Studien (**z. B. weniger als 5**) wird in der Literatur kritisch gesehen, da eine mögliche Heterogenität nur sehr unpräzise geschätzt werden kann [316]. Das Institut stellt Prädiktionsintervalle in Forest Plots von Meta-Analysen mit zufälligen Effekten in der Regel dar, wenn **mindestens 4 Studien** vorhanden sind und wenn die grafische Darstellung der Heterogenität wichtig ist. Dies ist insbesondere dann der Fall, wenn aufgrund von zu großer Heterogenität gar kein gepoolter Effekt präsentiert wird.*

Position AstraZeneca

Das Institut nennt eine unterschiedliche Mindestzahl an Studien um Meta-Analysen mit zufälligen Effekten darzustellen. Die betreffenden Abschnitte sollten dahingehend überarbeitet und die Mindestzahl an Studien harmonisiert werden.

5. Berücksichtigung indirekter Vergleiche basierend auf einarmigen Studien

IQWiG Allgemeine Methoden

S.193/ 9.3.9 Indirekte Vergleiche

Die Anwendung nicht adjustierter indirekter Vergleiche (d. h. die Verwendung einzelner Arme aus verschiedenen Studien) wird abgelehnt [45]. Dies gilt auch für Methoden für indirekte Vergleiche, in denen über Modellierungen mit starken Annahmen über die unbekanntene Effekte [107] oder mithilfe von Verfahren aus dem Bereich der Kausalmodelle für Beobachtungsstudien mit untestbaren Annahmen [596] versucht wird, Effektschätzungen trotz fehlender Brückenkomparatoren zu ermöglichen. Es werden ausschließlich adjustierte indirekte Vergleiche über adäquate Brückenkomparatoren akzeptiert.

Position AstraZeneca

Sofern die Zulassung eines Produktes durch die EMA auf einarmigen Studien basiert, ist es nach § 5 Abs. 3 AM-NutzenV unumgänglich, auf Basis dieser Studien ein Dossier zur Bewertung des medizinischen Zusatznutzens einzureichen. Der prinzipielle Ausschluss einer positiven Nutzenbewertung durch generelle Nicht-Berücksichtigung dieser Studien ist nicht zielführend. Es sollte zumindest in diesem Fall möglich sein, die bestverfügbare Evidenz durch geeignete Methoden der Adjustierung in einem indirekten Vergleich zu generieren und z.B. mittels Sensitivitätsanalysen zu stärken. Sind die Populationen der zu vergleichenden Studienarme hinsichtlich ihrer Charakteristika im Wesentlichen vergleichbar, bzw. kann insbesondere durch statistische Methoden (z. B. durch Adjustierung für mögliche Störgrößen

mittels Propensity-Score-Matching) eine weitreichende Annäherung an eine Variablenverteilung erreicht werden, die eine Randomisierung adäquat nachbildet, sollte die Evidenz für eine Nutzenbewertung grundsätzlich herangezogen werden. Dies würde einerseits der Forderung des §5 Abs. 3 AM-NutzenV entsprechen, nach der für „die erstmalige Bewertung [...] grundsätzlich die Zulassungsstudien zugrunde zu legen“ sind. Andererseits würde eine Berücksichtigung unter Würdigung der vorhandenen Unsicherheit dem positiven Votum der EMA nicht entgegenstehen. Grundsätzlich stimmt AstraZeneca der Auffassung des IQWiG zu, dass RCT den Goldstandard darstellt. In der hier angesprochenen Situation einer Zulassung auf Grundlage einer einarmigen Studie jedoch scheint die auch von Vertretern der EMA geäußerte Betrachtungsweise zielführender, bei der Abwägung der Qualität und des Umfangs der Evidenz von einem Kontinuum auszugehen, anstatt vereinfachend dichotome Kriterien (RCT vs. SAT) heranzuziehen (Eichler, Adaptive licensing: taking the next step in the evolution of drug approval, 2012; Eichler, From adaptive licensing to adaptive pathways: Delivering a flexible life-span approach to bring new drugs to patients, 2015) Auf diese Weise wird die Betrachtung der bereits zu einem frühen Zeitpunkt verfügbaren Evidenz ermöglicht, anstatt, nach einem binären Paradigma, alle Evidenz pauschal abzulehnen, die nicht einem gewissen Standard genügt.

Vorgeschlagene Änderung

Sofern die Zulassung für ein Arzneimittel auf Basis einarmiger Studien erteilt wurde, können grundsätzlich i.S. der Berücksichtigung der bestverfügbaren Evidenz auch indirekte Vergleiche einzelner Arme aus verschiedenen Studien Berücksichtigung finden, sofern zum Zeitpunkt der Zulassung nur einarmige Studien zur Verfügung standen und die im indirekten Vergleich berücksichtigten Populationen und Studien hinreichend ähnlich sind, bzw. durch geeignete statistische Verfahren (z.B. Propensity Score-basierte statistische Verfahren (Stürmer, 2010)) eine hinreichende Vergleichbarkeit erreicht wurde. Es muss eine adäquate Diskussion und Würdigung der vorhandenen Unsicherheit erfolgen.

Aufgrund der erhöhten Ergebnisunsicherheit können aus solchen Analysen maximal Hinweise auf einen Zusatznutzen abgeleitet werden.

Kommentar

Auf S. 193 wird die Arbeit von Signorovitch et al. 2010 als Verfahren zitiert, mit dem „versucht wird, Effektschätzungen trotz fehlender Brückenkomparatoren zu ermöglichen.“ Dieses ist unzutreffend, da es sich bei der von Signorovitch et al. vorgeschlagenen Methode um einen adjustierten indirekten Vergleich mit gemeinsamen Brückenkomparator handelt, der (im Gegensatz zu der Methode von Bucher et al.) aber zusätzlich versucht, für Unterschiede in den Charakteristika der beiden Studienpopulationen der berücksichtigten RCTs zu korrigieren. Der betreffende Abschnitt sollte überarbeitet werden.

6. Stratifizierte Randomisierung

IQWiG Allgemeine Methoden

S.194/ 9.3.10 Subgruppenanalysen

Vergleichbarkeit zwischen Behandlungsgruppen innerhalb der Subgruppen: Erfolgte die Randomisierung nicht nach dem Subgruppenmerkmal stratifiziert, so könnten sich die Behandlungsgruppen bei kleinen Fallzahlen innerhalb der Subgruppen in prognostischen

Faktoren unterscheiden [133,627]. In diesem Fall ist die Vergleichbarkeit der Behandlungsgruppen innerhalb der Subgruppen gefährdet, sodass (nicht) vorhandene Unterschiede zwischen Subgruppen allein durch diese Imbalance bedingt sein können.

Position AstraZeneca

Randomisierung an sich dient der zufälligen Zuordnung zu den Behandlungsarmen von bekannten und unbekanntem Merkmalen der Versuchspersonen, die in einer Studie als Störvariablen wirksam werden könnten, und verhindert somit Bias aufgrund von unbekanntem Störgrößen. Die stratifizierte Randomisierung (mit einer Einschränkungsmethode, z.B. Blockbildung) kann zusätzlich zu einer angemessenen Balance bezüglich der stratifizierten Faktoren und möglicherweise auch zu mehr statistischer Power führen. (Schulz, 2007) Allerdings sollte nicht außer Acht gelassen werden, dass die Anzahl an Stratifizierungsfaktoren sinnvollerweise begrenzt ist. Des Weiteren kann sich auch während des Zulassungsverfahrens respektive während der Nutzenbewertung die Notwendigkeit von weiteren, a priori nicht absehbaren Subgruppen ergeben.

7. Ableitung der Nutzaussagen von Subgruppen

IQWiG Allgemeine Methoden

S.195/ 9.3.10 Subgruppenanalysen

Ist das Ergebnis eines Heterogenitätstests- oder Interaktionstests zwischen wichtigen Subgruppen zum Niveau $\alpha = 0,05$ signifikant, liegt eine Effektmodifikation, das heißt zwischen den Subgruppen unterschiedliche Effekte, vor. In diesem Fall werden die Ergebnisse der Subgruppen nicht zu einem gemeinsamen Effektschätzer gepoolt. [...], und es werden getrennte Nutzaussagen für diese Gruppen abgeleitet.

Position AstraZeneca

Die regelmäßige Ableitung der Nutzaussagen für Subgruppen im Fall von zum Niveau $\alpha = 0,05$ signifikanten Interaktionstests ist, insbesondere im Hinblick auf das Problem des multiplen Testens, nicht zielführend.

Vorgeschlagene Änderung

Liegt ein zum Niveau $\alpha = 0,05$ signifikantes Ergebnis eines Interaktionstests zwischen Subgruppen vor, sollten zu diesem Ergebnis noch weitere Hinweise auf Kausalität vorliegen (z.B. biologische Plausibilität, Konsistenz über weitere Endpunkte, klinisch relevanter Unterschied, Beobachtung desselben Effekts in anderen Studien (Replizierbarkeit), ...), um daraus getrennte Nutzaussagen für diese Gruppen abzuleiten.

8. Restriktion der Durchführung von Subgruppenanalysen

IQWiG Allgemeine Methoden

S.195/ 9.3.10 Subgruppenanalysen

Damit überhaupt eine gewisse Wahrscheinlichkeit für einen solchen Nachweis vorliegt, werden Subgruppenanalysen bei binären Daten regelhaft nur dann durchgeführt, falls mindestens 10 Ereignisse in einer der Subgruppen aufgetreten sind.

Position AstraZeneca

AstraZeneca begrüßt grundsätzlich eine entsprechende Restriktion für die Analyse von Subgruppen, rät darüber hinaus jedoch zu einer noch differenzierteren Betrachtungs- und Vorgehensweis. Es ist nicht adäquat, auf Basis des Stratums mit der größten Anzahl an Ereignissen zu entscheiden, ob ein Interaktionstest zwischen bestimmten Subgruppen durchgeführt wird oder nicht. Die Varianz der Schätzung im Rahmen von Interaktionstests wird hauptsächlich durch das Stratum mit den wenigsten Ereignissen beeinflusst. Daher sollte die Restriktion auf Basis des Stratums mit der geringsten Anzahl an Ereignissen definiert werden. Die Vorgabe des IQWiG basiert darauf, dass ein Stratum, und somit das Stratum mit den meisten Ereignissen, mindestens 10 Ereignisse aufweisen muss.

Die Festlegung der Anzahl an Ereignissen bleibt ungeachtet der vorangegangenen Ausführungen willkürlich, wobei eine kleine Anzahl von z.B. 5 Ereignissen erfahrungsgemäß praktikabel ist und die Anzahl an falsch positiven Interaktionstests, welche bei vielen durchgeführten Subgruppenanalysen auftreten können, reduziert.

Vorgeschlagene Änderung

Subgruppenanalysen bei binären Daten werden regelhaft nur durchgeführt wenn folgend Bedingung erfüllt ist:

≥ 5 Ereignisse pro Behandlungsarm in jeder Subgruppen Kategorie (z.B. mindestens 5 Ereignisse bei den männlichen Studienteilnehmer im Verum-Arm).

Hiermit ist gewährleistet, dass eine Subgruppenanalyse nur durchgeführt wird, wenn in jedem Stratum innerhalb eines jeden Behandlungsarms mindestens 5 Ereignisse aufgetreten sind.

Literaturverzeichnis

- CPMP. (2001). Points to consider on application with 1. Meta-Analysis; 2. One pivotal study. (CPMP/EWP/2330/99).
- Eichler, H. (2012). Adaptive licensing: taking the next step in the evolution of drug approval. *Clin Pharmacol Ther* 91(3), S. 426-37.
- Eichler, H. (2015). From adaptive licensing to adaptive pathways: Delivering a flexible life-span approach to bring new drugs to patients. *Clin Pharmacol Ther* 97, S. 234–246.
- Schulz, K. F. (2007). Reihe Epidemiologi 6: Generierung von Randomisierungslisten in randomisierten Studien: Zufall, nicht Auswahl. *Zeitschrift für ärztliche Fortbildung und Qualität im Gesundheitswesen Volume 101, Issue 6*, S. 419–426.
- Stürmer, T. (2010). Treatment Effects in the Presence of Unmeasured Confounding: Dealing With Observations in the Tails of the Propensity Score Distribution—A Simulation Study. *Am J Epidemiol.* 172(7), S. 843–854.

A.1.3 – Bayer Vital GmbH

Autoren:

- Appelrath, Meike
- Dintsios, Charalabos-Markos



mittel den Patienten im Versorgungsalltag überhaupt dauerhaft zur Verfügung stehen. Somit haben die vermeintlich technischen Bewertungen des IQWiG auch Einfluss auf die Versorgungssituation der Versicherten in Deutschland.

Dies wird immer dann bedenklich, wenn das IQWiG seinen Ermessensspielraum – der trotz aller festgelegten Methoden an mancher Stelle entstehen kann – einseitig nutzt und im Rahmen der frühen Nutzenbewertung im Zweifel „gegen“ ein Arzneimittel entscheidet – mit dem nicht selten allzu schnellen Verweis auf die strengen Kriterien der evidenz-basierten Medizin. Dabei sind allerdings immer wieder auch Widersprüchlichkeiten festzustellen, die mitunter auch ein gewisses Maß an Willkür vermuten lassen.

Bayer als größtes forschendes pharmazeutisches Unternehmen in Deutschland ist grundsätzlich bereit, seine Arzneimittelinnovationen einer fairen Bewertung unterziehen zu lassen – wir scheuen dies schon deshalb nicht, da wir von deren Mehrwert überzeugt sind. Wir begrüßen, dass Studiendaten transparent gemacht und Vergleiche zu anderen Therapien durchgeführt werden. Dabei dürfen aber keine methodisch unüberwindbaren Hürden etabliert werden, die den Mehrwert eines Arzneimittels aus methodischer Rigorosität marginalisieren. In Deutschland ist mit der frühen Nutzenbewertung ein Weg eingeschlagen worden, der Spezifika mit sich bringt, die die pharmazeutischen Unternehmen vor mitunter unlösbare methodische Dilemmata stellen. Wir appellieren an das IQWiG, sich beim Umgang mit den Dilemmata konstruktiver einzubringen.

Wir sind uns bewusst, dass es eine große Herausforderung darstellt, einerseits nach replizierbaren Methoden immer „gleich“ zu bewerten und gleichzeitig einen pragmatischen Ansatz zu finden, der den spezifischen Besonderheiten von Erkrankungen und Therapieoptionen gerecht wird. Die im Entwurf der Version 5.0 vorgeschlagenen Änderungen sind teilweise zwar zu begrüßen, doch bleiben die aus unserer Sicht wesentlichen Probleme in den Bewertungsverfahren bestehen, so zum Beispiel:

- die mangelnde Akzeptanz von Endpunkten wie z.B. Metastasen-freies Überleben als patientenrelevanter Endpunkt in frühen onkologischen Therapielinien,
- die mangelnde Akzeptanz von Surrogatendpunkten, die in einigen chronischen Indikationen mitunter unumgänglich sind,
- die grundsätzliche Ablehnung von Studiendaten im Falle geringfügiger Abweichungen zwischen Fachinformation und Studiendesign,
- die mangelnde Bereitschaft zur Akzeptanz plausibler Analogieschlüsse, z.B. bei der Übertragung von Erkenntnissen auf in den Studien nicht vertretene Patientengruppen,
- die vom IQWiG entwickelte Methodik zur Bestimmung des Ausmaßes des Zusatznutzens (u.a. die dabei vorgenommene Hierarchisierung von Zielgrößen, die hyperkonservative ausschließliche Betrachtung der oberen Konfidenzintervallgrenze, die arbiträr gewählten Schwellenwerte),
- die semi-quantitative und damit willkürliche Saldierung von Nutzen und Schaden.



Dies führt in Summe dazu, dass die bestverfügbare Evidenz häufig gänzlich ignoriert wird, nur weil es sich nicht um die – in Theorie – bestmögliche Evidenz handelt. Aufgrund des Zeitpunktes der frühen Nutzenbewertung unmittelbar nach Zulassung eines Arzneimittels liegen regelhaft keine anderen als die Zulassungsstudien vor. Die Anforderungen des IQWiG an Studien widersprechen aber regelmäßig den Anforderungen der Zulassungsbehörden, mit denen die Studiendesigns abgestimmt wurden. Hier besteht aus Sicht von Bayer weiter dringender Diskussions- und Anpassungsbedarf.

Auf diese Problematik wird in der hier vorliegenden Stellungnahme aber im Folgenden nicht näher eingegangen. Die weitere Kommentierung beschränkt sich auf die aktuell beabsichtigten Änderungen der „Allgemeinen Methoden“, also auf den Entwurf für Version 5.0 vom 07.12.2016 und hier konkret auf folgende Abschnitte:

- 2.2.1 Einbindung von Betroffenen in die Erstellung von IQWiG-Produkten
- 3.1.4 Endpunktbezogene Bewertung
- 9.3.8 Meta-Analysen
- 9.3.10 Subgruppenanalysen

2.2.1 Einbindung von Betroffenen in die Erstellung von IQWiG-Produkten

Bayer begrüßt die Ergänzung des Abschnitts 2.2.1 über die **Einbindung von Betroffenen bei der Erstellung von IQWiG-Produkten**. Es ist unerlässlich, die Sicht von Betroffenen stärker zu berücksichtigen, z.B. bezüglich der Frage von patientenrelevanten Endpunkten und relevanten Patientensubgruppen.

Allerdings ist die methodische Beschreibung, wie diese Einbindung erfolgen soll, unterschiedlich konkret und lässt an einigen Stellen insbesondere im Vergleich zur Beschreibung anderer Vorgehensweise einen nicht unerheblichen Ermessensspielraum seitens des IQWiG. Während beispielsweise relativ genau beschrieben wird, warum Betroffene und nicht Patientenvertreter bzw. -repräsentanten eingebunden werden sollen, finden sich keine Hinweise darauf, was genau in welcher Form und welcher Detailtiefe erhoben und in welcher Form verwendet und veröffentlicht werden soll.

Aus Sicht von Bayer sollte die Beteiligung von Betroffenen **grundsätzlich bei allen IQWiG-Produkten erfolgen** und nicht von Zeitfaktoren oder der Vertraulichkeit des Berichtsthemas abhängig gemacht werden, da dies Steuerungsmöglichkeiten durch das Auftragsinstitut zulässt. Auch für Betroffene können Vertraulichkeitsvereinbarungen getroffen werden, die somit deren Beteiligung an Produkten, die einem vertraulichen Berichtsthema unterliegen, ermöglicht.

Darüber hinaus wäre es nicht nur wünschenswert, sondern im Sinne der Transparenz auch erforderlich, dass die **Aussagen der Betroffenen transparent** gemacht werden. Für die interessierte Öffentlichkeit sollte nachvollziehbar dargestellt werden, welche Interessen, Einstellungen und Präferenzen die Betroffenen haben und inwieweit diese in die Bewertung durch das IQWiG eingeflossen sind. Ansonsten bleibt unklar, ob bzw. inwieweit die Aussagen



der Patienten tatsächlich Berücksichtigung gefunden haben. Auch die Unterschiede in der Einschätzung von Betroffenen aufgrund unterschiedlicher persönlicher, krankheits-spezifischer oder sozio-ökonomischer Situationen sowie individueller Einstellungen sollten deutlich gemacht werden. Um den Datenschutz zu gewährleisten, können die Betroffenen anonym bleiben und müssen nicht namentlich genannt werden.

Hinsichtlich der Vorgehensweise zur Einbindung von Betroffenen in die Erstellung von IQWiG-Produkten, bleibt für uns offen, ob die Ergebnisse aus hierfür vom IQWiG initiierten relevanten Analysen bereits in die Methoden Eingang gefunden haben, z.B. die Ergebnisse der Projekte „Nutzertestung von Gesundheitsinformationen des IQWiG durch sozial benachteiligte Personen“ (http://www.iqwig.de/download/GA13-01_Arbeitspapier_Nutzertestung_durch_sozial_Benachteiligte.pdf) und „Nutzertestung von Gesundheitsinformationen des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen“ der Medizinischen Hochschule Hannover (http://www.mh-hannover.de/fileadmin/institute/epidemiologie/public_health/downloads/KursbeschrMar2011/NutzertestungAbschlussberichtIQWiG_16_12.pdf). Diese sollten aus unserer Sicht in die Beschreibung der methodischen Vorgehensweise eingearbeitet werden.

Bayer regt zudem an, die unterschiedlichen Dokumente im Bereich „Patientensicht einbringen“ (<http://www.iqwig.de/de/sich-beteiligen/patientensicht-einbringen.3070.html>) ggf. zusammenzuführen oder zumindest noch verständlicher zu benennen. Für Patienten ist die Unterteilung in „Beteiligung von Betroffenen bei der Dossierbewertung“ und „Beteiligung von Betroffenen bei der Erstellung von Berichten zur Nutzenbewertung“ nicht unbedingt nachvollziehbar.

3.1.4 Endpunktbezogene Bewertung

Das IQWiG formuliert Bedingungen, wann im Ausnahmefall aus einer Studie für einen Endpunkt ein „**Beleg**“ für einen **Zusatznutzen** abgeleitet werden kann. Bislang galt u.a., dass die Studie multizentrisch an mindestens 10 Zentren durchgeführt wurde. Es wurde nun neu ergänzt, dass in der Studie **in jedem Studienarm mindestens 1.000 Patienten** eingeschlossen sein müssen.

Es ist für Bayer nicht nachvollziehbar, auf Basis welcher Erkenntnisse der Schwellenwert von 1.000 Patienten festgelegt wurde. Diese Festlegung erscheint willkürlich, da das IQWiG keine tiefergehende Rationale hierfür liefert und auch nicht auf eine entsprechende Literaturstelle in der Methodenliteratur verweist. Bereits die Anzahl von mindestens 10 Zentren war willkürlich in der Vorgängerversion gewählt. Mit der neuen Bedingung wird z.B. auch eine Studien-Konstellation mit mehr als 2 Armen und summarisch mehr als 2.000 Patienten (bspw. dreiarmlige Studie mit jeweils 900 Patienten) durch eine Herabstufung der Aussagesicherheit sanktioniert, obwohl die Evidenzbasis hier ggf. noch breiter ist.

Durch diese zusätzliche Bedingung wird das Erreichen einer hohen Aussagesicherheit weiter erschwert. Auch besteht hier erneut die Gefahr, dass seltenere Erkrankungen systematisch im Bewertungsprozess benachteiligt werden, da es in manchen Indikationen nicht möglich ist, Studien mit derart hohen Fallzahlen durchzuführen.



Das IQWiG verweist zu Beginn des hier kommentierten Abschnittes 3.1.4 zwar darauf, dass man sich in der Nutzenbewertung bei der Einschätzung der Stärke der Ergebnis(un)sicherheit an „internationalen Standards der evidenzbasierten Medizin“ orientiere. Es ist für Bayer aber nicht ersichtlich, welche internationalen Standards eine Mindestanzahl von 1.000 Patienten je Studienarm fordern, um von einer hohen Ergebnissicherheit ausgehen zu können. Im Allgemeinen sollten Forderungen zu Belegen für einen Endpunkt mit denjenigen der Zulassungsstellen konform gehen.

9.3.8 Meta-Analysen

Bei der Durchführung von Meta-Analysen beabsichtigt das IQWiG mehrere methodische Änderungen/ Anpassungen, die Bayer insgesamt für sinnvoll erachtet.

Bezüglich der Durchführung von **Meta-Analysen mit zufälligen Effekten** weist das IQWiG darauf hin, dass „gemäß neueren Empfehlungen“ **regelmäßig die Knapp-Hartung-Methode** verwendet werden solle. Es wäre wünschenswert, dass das IQWiG die Vorteile gegenüber dem bisherigen Vorgehen schildert und eine Einschätzung zu den Abweichungen der Ergebnisse zwischen der Knapp-Hartung-Methode und dem bisherigen Vorgehen angibt, damit für Hersteller eine Quantifizierung dieser Abweichung ersichtlich wird.

Zu begrüßen ist die Absicht des IQWiG zukünftig Studienergebnisse auch dann zusammenzufassen, wenn der **Heterogenitätstest einen p-Wert $\geq 0,05$** liefert. Bislang durfte der p-Wert die Schwelle von 0,2 nicht unterschreiten. Somit wird zukünftig ein hyperkonservatives Vorgehen vermieden, das die Durchführung von Meta-Analysen bisher erschwert hat. Nichtsdestotrotz stellt sich aus Sicht von Bayer die Frage nach der Auswirkung dieser Änderung im Methodenentwurf auf bereits durchgeführte Meta-Analysen, zumindest im Rahmen der frühen Nutzenbewertung von Arzneimitteln. Wir schlagen daher vor, in alldenjenigen Fällen, in welchen die alte Regelung zur Anwendung kam, eine erneute Überprüfung der Meta-Analysen auf Basis der neuen Regelung durch das IQWiG zu veranlassen.

Auch die **Anwendung eines Modells mit fixen Effekten** bei Vorliegen von **weniger als 5 Studien** ist aus Sicht von Bayer sinnvoll – zumal in der frühen Nutzenbewertung regelmäßig weniger als 5 Studien vorliegen, die zu einer Meta-Analyse zusammengefasst werden können. Das vorgeschlagene Vorgehen ist aus Sicht von Bayer aber eher vage gehalten. Die Kann-Formulierung zu der Breite der Konfidenzintervalle beinhaltet keine Orientierungsgrößen. Ferner wird nur auf die Anzahl der Studien und nicht auf deren Größe abgestellt. Dies ermöglicht eine ergebnisoffene kontextbezogene Einschätzung der individuellen Eignung der Methoden durch das IQWiG. Aus Sicht von Bayer sollte regelmäßig das Modell mit fixen Effekten herangezogen werden, um Planungssicherheit für die Hersteller zu gewährleisten.



9.3.10 Subgruppenanalysen

Ähnlich wie bei der Durchführung von Meta-Analysen will das IQWiG zukünftig auch bei **Subgruppenanalysen** Abstand nehmen von einem hyperkonservativen Vorgehen und das **Signifikanzniveau** für die Ergebnisse von Heterogenitäts- oder Interaktionstests **von 0,2 auf 0,05 absenken**. Diese Änderung ist insofern zu begrüßen, als dass falsch-positive Ergebnisse damit reduziert werden, die in der Vergangenheit zu einer ungerechtfertigten Unterteilung in Subgruppen geführt haben, die bekanntlich mit verschiedenen methodischen Herausforderungen einhergeht – z.B. durch die Verkleinerung der Patientenpopulationen i.d.R. mit einem Verlust der Power und damit einhergehenden schwieriger zu erreichenden statistischen Signifikanz.

Gleichzeitig stellt sich auch hier die Frage, wie mit den bereits in der Vergangenheit durchgeführten Bewertungen umzugehen ist. Es sollte geprüft werden, ob es dort durch die Absenkung des Schwellenwertes zu anderen Subgruppenbewertungen gekommen wäre.

A.1.4 – Boehringer Ingelheim Pharma GmbH & Co. KG

Autoren:

- Henschel, Andreas
- Peil, Barbara

Boehringer Ingelheim Pharma GmbH & Co. KG · 55216 Ingelheim am Rhein

**IQWiG Institut für Qualität und
Wirtschaftlichkeit
Im Mediapark 8
50670 Köln**

**Boehringer Ingelheim
Pharma GmbH & Co. KG**

PM Germany
Market Access & Health Care
Affairs

27. Januar 2017

Schriftliche Stellungnahme zum Methodenentwurf

Sehr geehrte Damen und Herren,

anliegend erhalten Sie die Stellungnahme der Boehringer Ingelheim
Pharma GmbH & Co KG zum Entwurf der Version 5.0 der Allgemeinen
Methoden.

Die zitierte Literatur ist im Volltext inklusive eines standardisierten und
vollständigen Literatur- bzw. Anlagenverzeichnisses enthalten. Die zitierten
Literaturstellen sind in einer zusätzlichen Datei im RIS-Format aufbereitet.

Wir wären Ihnen dankbar, wenn Sie uns eine Eingangsbestätigung
zukommen lassen könnten.

Mit freundlichen Grüßen

i.A. 
Dr. Andreas Henschel

Dr. Andreas Henschel

Telefon 
Telefax 

Binger Straße 173
55216 Ingelheim am Rhein
Telefon 06132 77-0
Telefax 06132 72-0
www.boehringer-ingelheim.de

Sitz Ingelheim am Rhein
Registergericht Mainz
HR A 22206


Komplementär
Boehringer Ingelheim
Deutschland GmbH

Geschäftsführung
Stefan Rinn
(Vorsitzender)
Ursula Fuggis-Hahn
Ralf Gorniak
Dr. Douglas Khoo
Andreas Krüger

Vorsitzender des Aufsichtsrates
Dr. Wolfgang Baiker

Sitz Ingelheim am Rhein
Registergericht Mainz
HR B 23260

Stellungnahme Boehringer Ingelheim zum Entwurf der „Allgemeinen Methoden“ Version 5.0 des IQWiG

Am 07. Dezember 2016 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 der allgemeinen Methoden und nimmt im folgenden Stellung.

Der Entwurf zum Methodenpapier 5.0 des IQWiG enthält eine Reihe relevanter Änderungen für die frühe Nutzenbewertung nach § 35a SGB V. In der Gesamtschau kann die neue Version der allgemeinen Methoden aus Sicht von BI an vielen Stellen zur Transparenz der Bewertungen des Instituts beitragen. Auf die Änderungen in den Abschnitten *3.1.4 Endpunktbezogene Bewertung* sowie *9.3.10 Subgruppenanalysen* wird nachfolgend näher eingegangen.

3.1.4 Endpunktbezogene Bewertung

Bislang war es für die Ableitung eines Belegs auf Basis einer Studie notwendig eine multizentrische Studie mit mindestens 10 Studienzentren vorzuweisen. Im Entwurf zum Methodenpapier 5.0 wird ergänzt, dass die Studien zukünftig mindestens 1.000 Teilnehmer pro Studienarm umfassen sollen.

Eine Präzisierung der Anforderungen ist aus Sicht von BI zu begrüßen. Die Festlegung der Anzahl der Studienteilnehmer sollte jedoch grundsätzlich über eine Fallzahlschätzung erfolgen (Röhrig et al., 2010; ICH Expert Working Group, 1998). So hängt die jeweilig notwendige Fallzahl von der zugrundeliegenden Fragestellung ab. Sowohl eine zu kleine als auch zu große Fallzahl können aus medizinischen, statistischen und ethischen Überlegungen ungerechtfertigt sein (Röhrig et al., 2010; Altman, 1980). Darüber hinaus würde eine Festlegung auf eine konkrete Anzahl von Patienten dazu führen, dass es bei Erkrankungen mit geringer Prävalenz nicht möglich sein würde, einen Beleg mit einer Studie zu zeigen. Um zu gewährleisten, dass die Ableitung eines Belegs auf Basis einer multizentrischen Studie erkrankungsunabhängig erfüllt werden kann, sollte die Fallzahl daher am klinisch relevanten Effekt ausgerichtet werden.

Ergänzend ist anzumerken, dass in Studien mit Interventionen, die große Effekte zeigen, beispielsweise auch mit beträchtlich weniger als 1.000

Patienten ein deutlich kleinerer p-Wert als 0,05 erreicht werden könnte. Zusätzlich sollte zur Bewertung immer auch die Lage des Effektschätzers und des Konfidenzintervall mit einbezogen werden, und nicht alleine auf den p-Wert und die Fallzahl konzentriert werden.

9.3.10 Subgruppenanalysen

Im Methodenpapier 4.2 verwendete das Institut für Interaktionstests im Rahmen von Subgruppenanalysen sowohl die Beleg-Schwelle mit einem Signifikanzniveau $\alpha = 0,05$ als auch die Hinweis-Schwelle mit einem Signifikanzniveau $\alpha = 0,2$. Wie aus dem Entwurf der allgemeinen Methoden 5.0 hervorgeht, soll die Hinweis-Schwelle künftig nicht mehr angewendet werden. BI wertet diese Entscheidung als positiv, da der Wegfall die Wahrscheinlichkeit falsch-positiver Testungen reduziert.

Das Problem der Multiplizität, das in der vorherigen Version noch diskutiert wurde, ist allerdings auch bei Anwendung der Beleg-Schwelle weiterhin vorhanden. Die daraus resultierenden Limitierungen durch die zu erwartende Anzahl falsch-positiver Subgruppenanalysen sollten daher in der Interpretation beachtet werden (vgl. Higgins und Green, 2011). Aus diesem Grund wäre es aus Sicht von BI wichtig, den Umgang mit Multiplizität erneut in das Methodenpapier aufzunehmen und zu diskutieren.

Literatur:

- [1] Altman, D. G. 1980. Statistics and ethics in medical research: III How large a sample? *British Medical Journal*, 281(6251), 1336-38.
- [2] Higgins, J. & Green, S. 2011. Cochrane Handbook for Systematic Reviews of Interventions. Verfügbar: <http://handbook.cochrane.org/> [Aufgerufen am 2017-01-09].
- [3] ICH Expert Working Group. 1998. ICH harmonized tripartite guideline: statistical principles for clinical trials, E9 (Current Step 4 version, dated 5 February 1998). Verfügbar: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf [Aufgerufen am 2017-01-23].
- [4] Röhrig, B., du Prel, J. B., Wachtlin, D., et al. 2010. Fallzahlplanung in klinischen Studien: Teil 13 der Serie zur Bewertung wissenschaftlicher Publikationen. *Dtsch Arztebl Int*, 107(31-32), 552-6.

A.1.5 – Bristol-Myers Squibb GmbH & Co. KGaA

Autoren:

- Kupas, Katrin

Stellungnahme zum Entwurf der IQWiG „Allgemeinen Methoden“ Version 5.0 (ENTWURF)

Bristol-Myers Squibb
GmbH & Co. KGaA
Arnulfstraße 29
D-80636 München
Briefanschr.:
D-80632 München

Telefon (089) 121 42-0
[REDACTED]
<http://www.b-ms.de>
Registerger. München HRB 154472

[REDACTED]
[REDACTED]
[REDACTED]
Vorsitzender des Aufsichtsrates:
François Duplaix

Persönlich haftender Gesellschafter:
Bristol-Myers Squibb Verwaltungs GmbH
mit Sitz in München
Registerger. München HRB 149502
Geschäftsführer: Han Steutel

Aktualisierung des Abschnitts 9.3.9 zu indirekten Vergleichen

Folgende Passagen wurden im Methodenentwurf für Version 5.0 ergänzt bzw. geändert, zu der BMS Stellung nimmt:

„Dies gilt auch für Methoden für indirekte Vergleiche, in denen über Modellierungen mit starken Annahmen über die unbekanntes Effekte [107] oder mithilfe von Verfahren aus dem Bereich der Kausalmodelle für Beobachtungsstudien mit untestbaren Annahmen [596] versucht wird, Effektschätzungen trotz fehlender Brückenkomparatoren zu ermöglichen. Es werden ausschließlich adjustierte indirekte Vergleiche über adäquate Brückenkomparatoren akzeptiert. Hierzu zählen insbesondere das Verfahren von Bucher et al. [88] sowie die oben erwähnten Methoden der Netzwerk-Meta-Analysen.“

Grundsätzlich stimmt BMS zu, dass indirekte Vergleiche über einen Brückenkomparator solchen ohne Brückenkomparator vorzuziehen sind. Dennoch ist es nicht immer möglich, einen indirekten Vergleich über einen Brückenkomparator durchzuführen. Zur Erreichung der besten verfügbaren Evidenz können dann nur historische Vergleiche oder Vergleiche mit modellierenden Verfahren herangezogen werden. Daher ist es für BMS nicht nachvollziehbar, dass diese kategorisch abgelehnt werden.

Die Verfahren, die einen Vergleich ohne Brückenkomparator über eine Modellierung ermöglichen, werden stetig weiterentwickelt und systematisch auf ihre statistische Eignung, auch im Vergleich zur akzeptierten Bucher-Methode, untersucht [1]. Bei einer dokumentierten Eignung der Verfahren sollte die Möglichkeit bestehen, dass solche Verfahren auch für die Nutzenbewertung herangezogen werden können. Insbesondere bei Ergebnissen mit einer bekannten Verzerrung zu Ungunsten des zu untersuchenden Arzneimittels oder bei großen Effekten sollten diese Verfahren berücksichtigt werden. Die Größenordnung der Effekte sollte abhängig vom Fehler 1. und 2. Art der Verfahren gewählt werden und deutlich unter der Schwelle des sogenannten dramatischen Effektes liegen, da diese Verfahren im Gegenteil zum historischen Vergleich Verzerrungsaspekte teilweise kontrollieren. So sollte z.B. bei einem Verfahren mit einem dokumentierten Fehler 1. Art von <5% ein zum 5%-Niveau signifikantes Ergebnis ausreichend sein.

Abschnitt 3.1.4 zu Belegen aus einer Studie

Folgende Passagen wurden im Methodenentwurf für Version 5.0 ergänzt bzw. geändert, zu der BMS Stellung nimmt:

„Die Studie ist multizentrisch, in jeden Studienarm wurden mindestens 1000 Patienten eingeschlossen und es gibt mindestens 10 Zentren.“

Die Möglichkeit für einen Endpunkt ein Beleg abzuleiten wurde um eine geforderte Bedingung erweitert. Der Methodenentwurf für Version 5.0 definiert eine Fallzahl von mindestens 1000 eingeschlossenen Patienten pro Studienarm. Dies ist aus Sicht von BMS nicht nachvollziehbar. Jeder klinischen Studie liegt eine Fallzahlplanung zugrunde. Diese Fallzahlplanung ist eine Notwendigkeit in klinischen Studien, da weder zu kleine noch zu große Fallzahlen klinisch, methodisch oder ethisch zu rechtfertigen sind. Die Bewertung der Ergebnissicherheit sollte somit nicht an die Fallzahl der zu untersuchenden Studie geknüpft werden, sondern vielmehr sollte in Anlehnung an die Bewertung durch die EMA die Größe des Effekts und die somit erreichte statistische Sicherheit ausschlaggebend sein [2].

Aus Sicht von BMS wäre zudem eine Erläuterung über die rationale Begründung für die Fallzahl von mindestens 1000 eingeschlossenen Patienten erforderlich.

„Zusammenführung und Aktualisierung der vorherigen Abschnitte zu Subgruppenanalysen zum neuen Abschnitt 9.3.10“

Im Methodenentwurf für Version 5.0 wurden im neu verfassten Abschnitt einige Aspekte ergänzt bzw. geändert, zu denen BMS Stellung nimmt:

- *Stellenwert von Subgruppenanalysen*
- *Effektmodifikation durch mehr als ein Subgruppenmerkmal*
- *Mindestzahl von Ereignissen für Subgruppenanalysen bei binären Daten*
- *Schwelle für p-Wert des Interaktionstests*

Im vorliegenden Methodenentwurf relativiert das IQWiG die in der zitierten Literatur [3,4] vorgenommene kritische Diskussion von Subgruppenanalysen mit der Änderung der Formulierung von „sehr kritisch diskutiert“ auf „teilweise kritisch diskutiert“. Nach Auffassung von BMS werden Subgruppenanalysen und deren Bewertbarkeit weiterhin sehr kritisch diskutiert [5]. Die ursprüngliche Formulierung sollte daher beibehalten werden.

Positiv sieht BMS die Streichung der Hinweis-Schwelle von $p < 0,2$ für eine Effektmodifikation, da die Wahrscheinlichkeit falsch-positiver Interaktionstests dadurch reduziert wird.

Auch begrüßt BMS die Einführung eines fixen Schwellenwerts von mindestens 10 aufgetretenen Ereignissen in einer Subgruppe: *„Damit überhaupt eine gewisse Wahrscheinlichkeit für einen solchen Nachweis vorliegt, werden Subgruppenanalysen bei binären Daten regelhaft nur dann durchgeführt, falls mindestens 10 Ereignisse **in einer der Subgruppen** aufgetreten sind.“*

BMS bittet jedoch um eine Rationale dieses Schwellenwertes von 10 Patienten. Zudem bleiben, trotz Einführung dieses fixen Schwellenwerts, auch weiterhin Analysen von Subgruppen mit einer großen Unsicherheit behaftet. So muss bspw. berücksichtigt werden, dass es durchaus häufig Subgruppenanalysen gibt, bei denen nur in einzelnen Subgruppen mehr als 10 Ereignisse auftreten, wie z.B. bei der Ethnie. Hier kann sich eine Aussage in einer Subgruppenausprägung, die auf einer sehr geringen Anzahl an Ereignissen basiert, durch ein Ereignis mehr oder weniger grundlegend ändern und damit u.U. zu einem positiven Interaktionstest führen. Auch gibt es, insbesondere bei kleinen Subgruppen, das Problem, dass die Power deutlich reduziert ist, und so vorhandene Effekte nicht gezeigt werden können. Das Powerproblem wird im neuen Abschnitt zu Subgruppenanalysen zwar diskutiert, jedoch gibt es keinen Lösungsvorschlag zum Umgang mit diesem Problem. Nach Ansicht von BMS sollten Subgruppen auch nach Einführung einer solchen Schwelle nur unter Berücksichtigung aller Limitationen und sehr vorsichtig interpretiert werden. Zudem würde BMS es begrüßen, wenn die Voraussetzungen für Subgruppenanalysen dahingehend geändert würden, dass in allen Ausprägungen des Subgruppenmerkmals anstatt nur in einer der Subgruppen mehr als 10 Ereignisse aufgetreten bzw. mehr als 10 Patienten enthalten sein müssen.

Ein neuer Aspekt ist die Beschreibung der Problematik der Effektmodifikation durch mehr als ein Subgruppenmerkmal (sog. Wechselwirkung höherer Ordnung). BMS stimmt zu, dass

diese theoretisch möglich sind. Jedoch ist die Interpretierbarkeit unter anderem durch die noch kleiner werdenden Fallzahlen noch weiter limitiert.

BMS bittet das IQWiG, diese Limitationen entsprechend zu berücksichtigen und in die Version 5.0 mit aufzunehmen.

Grundsätzlich begrüßt BMS die Zusammenführung und Aktualisierung der vorherigen Abschnitte zu Subgruppenanalysen zum neuen Abschnitt 9.3.10. Jedoch sind aus Sicht von BMS bei der Interpretation von Subgruppenanalysen auch bei signifikanten Interaktionstest die Limitationen und die Zuverlässigkeit von Subgruppenanalysen zu berücksichtigen.

Referenzen

- [1] Siebold M., Buchner H., Kupas, K: Statistical properties of the Bayes-Benchmarking-Analysis: A simulation study; Master's thesis, Institute for Medical Information Processing, Biometrics and Epidemiology, Ludwig-Maximilians-Universität München, to be published
- [2] CPMP (2001) Points to consider on application with 1. Meta-Analysis; 2. One pivotal study (CPMP/EWP/2330/99)
- [3] Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet* 2000; 355(9209): 1064-1069.
- [4] Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med* 1992; 116(1): 78-84
- [5] Sun et al. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. *BMJ* 2012;344: e1553

A.1.6 – Bundesarbeitsgemeinschaft SELBSTHILFE von Menschen mit Behinderung und chronischer Erkrankung und ihren Angehörigen e. V. (BAG SELBSTHILFE)

Autoren:

- Danner, Martin

Stellungnahme der

Bundesarbeitsgemeinschaft SELBSTHILFE von Menschen mit Behinderung und chronischer Erkrankung und ihren Angehörigen (BAG SELBSTHILFE) e. V.

zum

Entwurf des Methodenpapiers 5.0 des IQWiG

Hintergrund

Als Dachverband von 120 Bundesverbänden der Selbsthilfe chronisch kranker und behinderter Menschen sowie von 13 Landesarbeitsgemeinschaften misst die BAG SELBSTHILFE der Neufassung des Methodenpapiers große Bedeutung bei. Wir nehmen zu folgenden Punkten Stellung:

1.2 Evidenzbasierte Medizin

Die Darstellungen zur evidenzbasierten Medizin sind ebenso zutreffend wie begründenswert. Allerdings wird die Abgrenzung zwischen Evidenz-based-Medicine im individuellen Behandlungsverhältnis und Evidence-based-Health-Care bei der abstrakt-generellen Versorgungsgestaltung nicht klar herausgearbeitet.

Ebenso wie es bei der Evidenz-based-Medicine im individuellen Behandlungsverhältnis nicht nur auf die abstrakte Studienlage, sondern auf die individuelle Krankheitsgeschichte und Werthaltungen des einzelnen Patienten ankommt, bedarf es bei der Evidence-based-Health-Care der Reflektion, welche Auswirkungen Entscheidungen zur abstrakt-generellen Versorgung haben, um in diesem Kontext die Studienlage zu analysieren.

Insbesondere der Abschnitt 1.4 zur evidenzbasierten Entscheidungsfindung im Gesundheitswesen bleibt daher seltsam blass, obwohl die meisten Produkte des Instituts doch die abstrakt-generelle Versorgungsgestaltung betreffen. Auch die Überlegungen in 1.3 zur Gesundheitsökonomie beziehen sich ersichtlich auf eine abstrakt-generelle Betrachtung, deren Zielparameter allerdings unklar bleiben.

2.2.1 Einbindung von Betroffenen in die Erstellung von IQWiG-Produkten

Der Begriff der Erkrankung wird in der Medizinteorie seit langem sehr intensiv mit den unterschiedlichsten Definitionsversuchen diskutiert. Allgemein anerkannt ist dabei jedoch, dass mit dem Auftreten einer Erkrankung die Lebensumstände der Menschen teilweise gravierend verändert werden, dass sich nicht nur die persönliche Lebensplanung, das Verhältnis zum sozialen Umfeld, aber auch die Werthaltungen erheblich ändern können. Umgekehrt können auch medizinische Interventionen ebenso gravierende Auswirkungen haben.

Im Abschnitt 7.2.3 „Patientenwege“ werden diese Dimensionen durchaus im Methodenpapier auch im Einzelnen angesprochen.

Das Wissen zu all diesen Umständen und Veränderungen ist nicht nur für das praktische Handlungsgeschehen im Einzelfall essentiell, sondern auch für die abstrakt-generelle Bewertung von Behandlungsoptionen im HTA-Verfahren und für die Erstellung von Gesundheitsinformationen. Die Patientenrelevanz von Endpunkten kann bspw. gar nicht adäquat bestimmt werden, wenn nicht die Patientenperspektive umfassend in all ihrer Komplexität analysiert wird.

Ohne die Berücksichtigung der Patientensicht können all diese IQWiG-Produkte nicht adäquat erstellt werden.

Diese in methodischer Sicht wichtige Feststellung fehlt im Abschnitt 2.2.1, in dem lediglich deskriptiv vermerkt wird, wie die Einbindung Betroffener stattfinden kann, bzw. stattfindet.

Dies ist deshalb besonders problematisch, weil ohne eine methodisch abgesicherte Einbindung der Patientenperspektive die Gefahr besteht, dass Studienbewertungen vom Bearbeiter/von der Bearbeiterin auf der Basis sog. Alltagstheorien zur jeweiligen Erkrankung erfolgen. Man stellt sich dann vor, was für einen Betroffenen wohl wichtig ist oder welche Bedeutung eine bestimmte Einschränkung hat. Ein solches Vorgehen jedoch ist schlicht unwissenschaftlich bzw. genügt nicht dem Erfordernis, dass sich Wissenschaftler stets kritisch mit sog. Vorverständnissen ihres Tuns auseinandersetzen haben.

Zwar wird an der einen oder anderen Stelle im Methodenpapier angesprochen, dass zur Klärung dieser Frage qualitative Studien herangezogen werden können oder Betroffene befragt werden können. Es fehlt aber insoweit an einer klaren methodischen Definition der Herangehensweise.

Es wäre wünschenswert, wenn das Methodenpapier hier eine Präzisierung erführe.

Auch die Möglichkeit, Patientenerfahrungen über den Austausch in Selbsthilfegruppen und -organisationen zu aggregieren, wird im Methodenpapier nicht hinreichend erfasst. Dies zeigt sich vor allem im 3. Absatz des Abschnitts, der sachlich unrichtig, zumindest aber irreführend ist. Der Absatz muss vielmehr lauten:

„Bei der Auswahl der Teilnehmenden wird der Fokus auf tatsächlich Betroffene gerichtet. Wo diese nicht selbst befragt werden können (z.B. bei Kleinkindern oder schwer Demenzkranken), sind nahe Angehörige zu befragen. Menschen, die nicht selbst Patienten sind, können nicht in vergleichbarer Weise beurteilen, wie Betroffenen Symptome empfinden, ihre Funktionen und Aktivitäten wahrnehmen können oder in ihrer Lebensqualität beeinträchtigt sind.“

Wünschenswert wäre es schließlich auch, wenn die im 5. Absatz angesprochenen Konsultationsgespräche auf der Basis einer methodisch abgesicherten Grundlage stattfinden würden. Zielsetzungen, Fragestellungen, Ergebnisverlauf und Ergebnissicherung sollten auf der Basis eines ex ante- und ex post-Sachverhaltes analysiert werden. Nur so kann der Erkenntnisgewinn, der durch die Konsultationen erfolgt, objektiv erhoben werden und nur so kann vom Bearbeiter auch konzentriert reflektiert werden, wo seine Alltagstheorien zur Erkrankung und ihren Folgen bestätigt und wo sie widerlegt wurden. Denkbar ist auch, dass nach dem Konsultationsgespräch Unklarheiten fortbestehen, so dass weitere Erhebungen notwendig sind.

Ausdrücklich begrüßt wird jedoch, dass weder die Namen der Betroffenen noch deren Aussagen im Einzelnen veröffentlicht werden, wenn diese dies nicht wünschen. Hier steht das Recht auf informationelle Selbstbestimmung der Betroffenen über dem Transparenzinteresse der (Fach-)Öffentlichkeit.

3. Nutzenbewertung medizinischer Interventionen

Aus Sicht der BAG SELBSTHILFE ist es elementar, dass sich der Nutzen von Behandlungsmethoden anhand patientenrelevanter Endpunkte bemisst. Diese Auffassung gilt jedoch für den Nutzen von diagnostischen Methoden in einer spezifischen Art und Weise: Hier liegt der Nutzen aus Sicht der Patientenvertretung in der besseren Diagnostik, also dem besseren Erkennen der Krankheit und der für das Ausmaß bedeutsamen Einstufung. Die gegenteilige Auffassung des IQWiG, wonach auch der Nutzen von diagnostischen Maßnahmen zunächst von der Existenz und auch dem Nutzen einer Behandlungsmethode abhängt ist und danach auch noch die patientenrelevanten Endpunkte in Abhängigkeit vom Testergebnis zu bestimmen sind (S. 44), wird aus zwei Gründen abgelehnt: Zum einen wäre bei dieser Auffassung die Erstattung von Diagnostika bei seltenen Erkrankungen, für die es noch keine Behandlung gibt, ausgeschlossen. Damit mutet man den Betroffenen nicht nur oft jahrelange Diagnoseodysseen auf behandelbare Erkrankungen zu, sondern will ihnen dann, wenn sich der Verdacht auf eine bestimmte Erkrankung verdichtet, die entsprechende Diagnostik verweigern bzw. ihnen auferlegen, diese aus eigener Tasche zu bezahlen. Dies ist nicht nachvollziehbar.

Zum anderen hat diese Auffassung zur Folge, dass es auch bei behandelbaren Krankheiten statistisch keinen Nutzen gibt, wenn sich die Behandlung aufgrund der Diagnostik nur in Einzelfällen ändert. Denn dann wird es regelmäßig keine statis-

tisch signifikante Verbesserung der patientenrelevanten Endpunkte geben; für den einzelnen Patienten ist die Diagnostik jedoch dennoch oft lebensentscheidend, da eben nicht von vorneherein ausgeschlossen werden kann, dass er in die Ausnahme-gruppe fällt, in der sich die Behandlung aufgrund des Testergebnisses ändert.

3.8 Potentialbewertung

In Kapitel 3.8 sind die spezifischen Evidenzstufen (gering, sehr gering, minimal) zur Potenzialbewertung und deren Bedeutung nach § 137e nach Antragstellung durch externe Anbieter durch das IQWiG beschrieben.

So heißt es auf Seite 76 wie folgt:

„Daher ergibt sich ein Potenzial insbesondere dann, wenn Studien geringer Ergebnissicherheit mindestens kleine Effekte zeigen, wenn Studien sehr geringer Ergebnissicherheit mindestens mittlere Effekte zeigen oder wenn Studien minimaler Ergebnissicherheit mindestens große Effekte zeigen. Als ungefähre Grenzen zwischen kleinen, mittleren und großen Effekten können für das relative Risiko Werte von 0,8 und 0,5 dienen [147,419].“

Aus Sicht der BAG SELBSTHILFE stellt diese Differenzierung nach dem relativen Risiko eine zu große Hürde dar, wenn man diesen Maßstab ganz allgemein anlegt. Daher wird vorgeschlagen, dass auch Studien wie Fallserien nach dem Gesichtspunkt der Gleichgerichtetheit (beispielsweise 2 Fallserien mit statistisch signifikanten Ergebnissen in gleicher Richtung) als ausreichend bewertet werden können. Aus Erfahrungen in der Patientenvertretung sehen wir hier durchaus Diskussionsbedarf. Aus Sicht der BAG SELBSTHILFE ist zu betonen, dass hier durch den Gesetzgeber offene Begriffe eingeführt worden sind, die den klassischen Kategorisierungen der EBM nicht unbedingt entsprechen, die aber die Grundlage zur Bildung von Kriterien liefern sollen, wann Erprobungsverfahren durchgeführt werden sollen.

Aus Sicht der BAG SELBSTHILFE sollte dieser Prozess der Kriterienbildung behutsam erfolgen, so dass die Festlegung konkreter Grenzwerte noch als verfrüht erscheint.

7.3 Informationsbeschaffung zur Erstellung von Gesundheitsinformationen

Wie bereits zu Abschnitt 2.2.1 ausgeführt wurde, können insbesondere die Gesundheitsinformationen des IQWiG nur dann ihren Zweck adäquat erfüllen, wenn die Sicht der Betroffenen, insbesondere deren Informationsbedarfe, Ausgangspunkt der Überlegungen ist.

Auch hier geht es darum, nicht abgesicherte Vorverständnisse der Autorinnen und Autoren, d.h. deren Alltagstheorien zu jeweiligen Erkrankung, deren Folgen und den unterstellten Informationsbedarfen der Betroffenen, kritisch zu reflektieren und ggf. zu falsifizieren. Ohne diese Problematik methodisch aufzubereiten, wird dies im Abschnitt 7.3 durchaus anerkannt, indem ausgeführt wird, dass es für die Autorinnen und Autoren darum geht, „sich so gut wie möglich in die Situation Be-

troffener zu versetzen“. Auch hier wird das Fehlen einer stringenten Methodik deutlich:

Richtig wird zwar ausgeführt, dass qualitative Literatur, systematische Übersichten oder Primärstudien hier weiter helfen können.

Kurioserweise wird jedoch im Abschnitt 7.3 weder die Befragung Betroffener noch die Einbindung von Selbsthilfeorganisationen angesprochen, obwohl dies seit langem gängige Praxis ist.

Dies muss ergänzt und in den zu 2.2.1 angesprochenen methodischen Gesamtrahmen eingebettet werden.

Ein evidenzbasiertes Vorgehen erfordert, die Alltagstheorien der Autorinnen und Autoren anhand aller verfügbaren Erkenntnisquellen zu überprüfen. Hierzu zählen neben der Literaturanalyse, der Auswertung von Übersichtsarbeiten und Einzelstudien selbstverständlich auch die Konsultation von Betroffenen und die Einbindung von Selbsthilfeorganisationen.

Außerdem bittet die BAG SELBSTHILFE darum, dass künftig in den Gesundheitsinformationen des IQWiG zumindest vermerkt wird, wenn diese in Kooperation mit einer Selbsthilfeorganisation entstanden ist. Dies diene nicht nur der Transparenz, sondern böte den Organisationen auch die Möglichkeit, sich zu einer evidenzbasierten Erstellung von Gesundheitsinformationen zu bekennen.

Dr. Martin Danner

Düsseldorf, 31.01.2017

A.1.7 – Bundesverband der Arzneimittel-Hersteller e. V. (BAH)

Autoren:

- Poß, Rudolf

Stellungnahme des

Bundesverbandes der Arzneimittel-Hersteller e.V. (BAH)

im Rahmen des Stellungnahmeverfahrens

Allgemeine Methoden: Entwurf für Version 5.0

des Instituts für Qualität und Wirtschaftlichkeit im
Gesundheitswesen

Einleitung

Der BAH bedankt sich für die Möglichkeit, im Rahmen dieses Verfahrens Stellung nehmen zu dürfen. Er begrüßt, dass das IQWiG die Aktualisierung seiner „Allgemeinen Methoden“ einem öffentlichen Stellungnahmeverfahren unterzieht und damit transparent handelt. Bewertungen des Instituts, besonders im Rahmen der Nutzenbewertung nach §35a SGB V, sind von großem Einfluss auf Entscheidungen der Selbstverwaltung unseres Gesundheitssystems. Die Stellungnahme beinhaltet die Zuarbeit von Mitgliedsunternehmen des BAH. Zu folgenden Punkten des Methodenpapiers nehmen wir Stellung:

2.2.1 Einbindung von Betroffenen in die Erstellung von IQWiG-Produkten

Der BAH begrüßt die Ergänzung des Abschnitts 2.2.1 über die Einbindung von Betroffenen bei der Erstellung von IQWiG-Dokumenten. Es ist wichtig, die Patientensicht stärker zu berücksichtigen.

Allerdings ist die methodische Beschreibung nicht immer hinreichend konkret und lässt einen deutlichen Ermessensspielraum. Während beispielsweise relativ genau beschrieben wird, warum Betroffene und nicht Patientenvertreter eingebunden werden sollen, finden sich keine Hinweise zu Form und Detaillierungsgrad der Erhebung und was veröffentlicht werden soll.

Aus Sicht des BAH sollte die Beteiligung von Betroffenen grundsätzlich bei allen IQWiG-Produkten erfolgen. Vertraulichkeit und Datenschutz der Betroffenen können durch geeignete Maßnahmen sichergestellt werden.

Darüber hinaus wäre es nicht nur wünschenswert, sondern im Sinne der Transparenz auch erforderlich, dass die Aussagen der Betroffenen dargestellt werden, wenigstens in aggregierter Form auf Ebene der Präferenzkriterien. Für die interessierte Öffentlichkeit sollte nachvollziehbar sein, welche Interessen, Einstellungen und Präferenzen die Betroffenen tatsächlich haben und inwieweit diese in die Bewertung durch das Institut eingeflossen sind. In der Darstellung sollten alle erhobenen Kriterien berücksichtigt werden, unabhängig von der Beurteilung ihrer Relevanz durch das IQWiG.

3.1.4 Endpunktbezogene Bewertung

Das IQWiG formuliert Bedingungen, wann ausnahmsweise aus einer Studie für einen Endpunkt ein „Beleg“ für einen Zusatznutzen abgeleitet werden kann. Bislang galt (Version 4.2), dass die Studie multizentrisch an mindestens 10 Zentren durchgeführt wurde. Es wurde nun ein zusätzliches Kriterium eingeführt, nämlich dass in der Studie in jedem Studienarm mindestens 1.000 Patienten eingeschlossen sein müssen.

Es ist für den BAH unklar, auf Basis welcher Erkenntnisse der Schwellenwert von 1.000 Patienten gesetzt wurde. Das IQWiG liefert keine Rationale hierfür und verweist auch nicht auf eine entsprechende Literaturstelle. Bereits die Anzahl von mindestens 10 Zentren wurde ohne Rationale festgelegt.

Durch diese zusätzliche Bedingung wird das Erreichen einer hohen Aussagesicherheit weiter erschwert. Auch besteht hier erneut die Gefahr, dass seltenere Erkrankungen und spezielle

Therapiesituationen systematisch im Bewertungsprozess benachteiligt werden, da es in manchen Indikationen nicht möglich ist, Studien mit solch hohen Fallzahlen aufzusetzen.

Die Effektgröße stellt das maßgebliche Kriterium für die Ergebnissicherheit dar und sollte primär Berücksichtigung finden. Auch die Studienqualität (Verzerrungspotential und Design) sollte mitberücksichtigt werden.

9.3.8 Meta-Analysen

Bei der Durchführung von Meta-Analysen nimmt das IQWiG methodische Anpassungen vor, die der BAH grundsätzlich begrüßt.

So beabsichtigt das IQWiG, zukünftig Studienergebnisse auch dann zusammenzufassen, wenn der Heterogenitätstest einen p-Wert $\geq 0,05$ liefert. Damit wird zukünftig ein hyperkonservatives Vorgehen vermieden, das die Durchführung von Meta-Analysen bisher erschwert hat. Das ist grundsätzlich zu begrüßen. Es stellt sich jedoch die Frage nach der Auswirkung dieser Änderung im Methodenentwurf auf bereits durchgeführte Meta-Analysen, vor allem im Rahmen der frühen Nutzenbewertung von Arzneimitteln. Eine Überprüfung der Meta-Analysen auf Basis der neuen Regelung durch das IQWiG erscheint hier angemessen.

Auch die Anwendung eines Modells mit fixen Effekten bei Vorliegen von weniger als 5 Studien ist aus Sicht des BAH sinnvoll – zumal in der frühen Nutzenbewertung regelhaft weniger als 5 Studien vorliegen, die zu einer Meta-Analyse zusammengefasst werden können. Das Vorgehen ist jedoch unklar formuliert, da „kontextabhängig“ Alternativen vorgeschlagen werden. Im Sinne der Verfahrensklarheit macht eine Festlegung auf das bewährte Modell mit fixen Effekten Sinn.

9.3.9 Indirekte Vergleiche

Es ist nachvollziehbar, dass indirekte Vergleiche über einen Brückenkomparator solchen ohne Brückenkomparator vorzuziehen sind. Dennoch ist es nicht immer möglich, einen indirekten Vergleich über einen Brückenkomparator durchzuführen. Die Verfahren, die einen Vergleich ohne Brückenkomparator über eine Modellierung ermöglichen, werden kontinuierlich weiterentwickelt. Daher ist es nicht nachvollziehbar, dass diese kategorisch abgelehnt werden. Vielmehr sollten solche Verfahren mit Begründung herangezogen werden können, falls ein Vergleich über einen Brückenkomparator nicht möglich ist. Dies macht vor allem konservativer Schätzung (Verzerrung zu Ungunsten der zu überprüfenden Intervention) Sinn.

9.3.10 Subgruppenanalysen

Der berechtigte Hinweis auf die multiple Testthematik in der Version 4.2 („Bei der Analyse zahlreicher Subgruppen besteht eine manchmal recht hohe Wahrscheinlichkeit, dass die Ergebnisse irgendeiner Subgruppe statistische Signifikanz erreichen, obwohl es sich in Wahrheit um ein zufälliges Ergebnis handelt“) taucht in der Version 5.0 nicht mehr auf. Diese

Problematik sollte weiterhin adressiert werden und die alte Formulierung beibehalten werden.

Weiterhin wurde folgende Passage zur Aussagekraft von Subgruppenanalysen in der Version 5.0 ergänzt:

„Vergleichbarkeit zwischen Behandlungsgruppen innerhalb der Subgruppen: Erfolgte die Randomisierung nicht nach dem Subgruppenmerkmal stratifiziert, so könnten sich die Behandlungsgruppen bei kleinen Fallzahlen innerhalb der Subgruppen in prognostischen Faktoren unterscheiden [133,627]. In diesem Fall ist die Vergleichbarkeit der Behandlungsgruppen innerhalb der Subgruppen gefährdet, sodass (nicht) vorhandene Unterschiede zwischen Subgruppen allein durch diese Imbalance bedingt sein können“

Eine stratifizierte Randomisierung ist in der Praxis nicht immer für alle Subgruppen logistisch durchführbar. Gerade eine Verwendung von zu vielen Stratifizierungsmerkmalen erhöht das Risiko von Imbalancen innerhalb der einzelnen Merkmalskombinationen. Darüber hinaus, stellt eine Stratifizierung nach einem Subgruppenmerkmal nicht notwendigerweise eine Gleichverteilung anderer prognostischer Faktoren sicher. Ferner können Subgruppen auch auf regulatorische Anforderungen zurückgehen, die eine a-priori Stratifizierung unmöglich machen. Grundsätzlich sollte eine sorgfältige Abwägung stattfinden, ob eine Schwächung der Aussagekraft infolge Subgruppenstratifizierung durch eine entsprechende medizinische Rationale gerechtfertigt ist.

Ähnlich wie bei der Durchführung von Meta-Analysen will das IQWiG zukünftig auch bei Subgruppenanalysen Abstand nehmen von einem hyperkonservativen Vorgehen und das Signifikanzniveau für die Ergebnisse von Heterogenitäts- oder Interaktionstests von 0,2 auf 0,05 absenken. Diese Änderung ist zu begrüßen, da sie die Problematik des multiplen Testens adäquater adressiert und die Menge falsch-positiver Ergebnisse reduziert, die in der Vergangenheit zu einer ungerechtfertigten Unterteilung in Subgruppen geführt haben. Auch hier erachtet der BAH den unter „9.3.8 Meta-Analysen“ gemachten Vorschlag, bereits durchgeführte Subgruppenanalysen auf Basis der neuen Regelung erneut zu rechnen, als sinnvoll.

Berlin, 27. Januar 2017

gez. Rudolf Poß
Bundesverband der Arzneimittel-Hersteller e. V. (BAH)

A.1.8 – Bundesverband der Pharmazeutischen Industrie e. V. (BPI)

Autoren:

- Gerbsch, Norbert
- Peters, Jens
- Wilken, Matthias

Stellungnahme

Stellungnahme

zum Entwurf des IQWiG

„Allgemeine Methoden“ Version 5.0

Stand: 31.01.2017

Stellungnahme

Das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) hat seine „Allgemeinen Methoden“ inhaltlich aktualisiert und ergänzt.

Zum Entwurf der Version 5.0 der „Allgemeinen Methoden“ möchte der Bundesverband der Pharmazeutischen Industrie (BPI) Stellung nehmen:

Einbindung von Betroffenen (Abschnitt 2.2.1)

Die vorgesehene Einbeziehung von Betroffenen bei der Erstellung von Institutsprodukten ist sinnvoll und zu begrüßen. Ebenfalls ist zu begrüßen, dass das Institut anerkennt, dass die Sichtweisen von tatsächlich Betroffenen sich maßgeblich von der Sichtweise von Repräsentanten aus dem Bereich der Selbsthilfe oder gar aus dem Bereich des Verbraucherschutzes unterscheiden können, wenn diese keine Patienten und daher nicht selbst betroffen sind.

Es ist ebenfalls sinnvoll, den Kreis der Betroffenen nicht zu eng zu fassen, sondern bspw. auch Eltern und Angehörige zu diesem Kreis hinzuzunehmen.

Grundsätzlich ist aus Sicht des BPI eine Einbindung mehrerer Betroffener wünschenswert, die Vorgaben bleiben hierzu zu vage.

Auch wenn nachvollziehbar ist, dass die Namen der in eine Konsultation eingebundenen Betroffenen grundsätzlich nicht veröffentlicht werden, sind die vorgesehenen Vorgaben zur Verfahrenstransparenz aus Sicht des BPI insgesamt nicht ausreichend. Die an die Betroffenen gestellten Fragen und Antworten, Dissens wie Konsens sollten nachvollziehbar sein und in den Institutsprodukten transparent dargestellt werden. Die Würdigung und der Einfluss auf das Bewertungsergebnis sollten ersichtlich sein.

Bzgl. der Auswahl der Betroffenen soll einerseits auf die Patientenvertretung im G-BA zurückgegriffen werden. Andererseits kann das IQWiG auch eigenständig Betroffene bestimmen. Kriterien für den Auswahlprozess sollten ergänzend dargestellt werden. Die Schritte des Auswahlverfahrens sollten transparent dargestellt werden.

Es ist zu begrüßen, dass Betroffene mit Interessenkonflikten von Konsultationen nicht ausgeschlossen werden.

Die generell getroffene Aussage, dass eine (erneute) Einbindung von Betroffenen verzichtbar sei, wenn ein Berichtsthema bereits im Rahmen einer früheren Konsultation bearbeitet wurde, ist nachdrücklich zu hinterfragen. Dies wird sehr stark davon abhängen, ob sich bspw. in einem Therapiegebiet maßgebliche Veränderungen ergeben haben, die Einfluss auf die Beurteilung von Sachverhalten durch Betroffene haben können. Auch das zeitliche Delta zwischen einer ersten und einer Folgeeinbindung wäre zu berücksichtigen. Ob Betroffene auf eine Einbindung verzichten wollen, sollten diese vornehmlich selbst einschätzen.

Wissenschaftliche Beratung (Abschnitt 2.2.7)

Das IQWiG gibt an, sich an der wissenschaftlichen Beratung (Scientific Advice) von Studiensponsoren zu beteiligen, z. B. im Rahmen der Mitarbeit im europäischen Netzwerk EUnetHTA.

Stellungnahme

Das Methodenpapier bleibt an dieser Stelle sehr vage. Es fehlen konkrete Erläuterungen, wie diese Beteiligung ausgestaltet ist und was mit „speziellen Fällen“ genau gemeint ist. Die Beteiligung des IQWiG an den Beratungsprozessen sollte detaillierter beschreiben werden. Ebenso bleibt unklar, wie sich die Beteiligung des IQWiG im Verhältnis zum Gemeinsamen Bundesausschuss verhält, der ebenfalls im Rahmen des europäischen Netzwerks EUnetHTA mitarbeitet.

Die angesprochene organisatorische Trennung zwischen Beratung und Bewertung erscheint künstlich und wenig hilfreich. Beratung und Bewertung bzw. Bewertung und Beratung beeinflussen einander und hängen voneinander ab. Zum einen ist sicherzustellen, dass Erkenntnisse aus der Bewertung in zukünftige Beratungen einfließen, zum anderen ist es wichtig, dass Ergebnisse von Beratungen auch im späteren Bewertungsprozess anerkannt und nicht erneut hinterfragt werden.

Endpunktbezogene Bewertung (Abschnitt 3.1.4)

Der Entwurf sieht mit Bezug auf die Ableitung eines Belegs aus einer Studie u. a. vor, dass in jeden Studienarm 1000 Patienten eingeschlossen worden sind.

Die fixe Schwelle von 1000 Patienten ist starr und nicht belegt. Die Einführung dieses fixen Schwellenwerts würde für einige Indikationsgebiete wie bspw. im Bereich der Onkologie die Ableitungen eines Belegs aus einer Studie bis zur Unmöglichkeit erschweren.

Die erforderliche Patientenzahl sollte daher nicht fixiert werden, sondern abhängig von der Indikation, der Seltenheit einer Erkrankung und den zu erwartenden Effekten bestimmt werden. Hierbei sind auch ethische Fragen zu berücksichtigen, die sich auf die Größe der Studienpopulation auswirken können. Die Anforderungen an die Ableitung eines Belegs aus einer Studie sollten daher die indikationsspezifischen Besonderheiten der Studiendurchführung und die Aspekte der zugrundeliegenden Fallzahlplanung adäquat berücksichtigen.

Identifizierung der Informationsbedürfnisse / Aufbereitung von Krankheitserfahrungen (Abschnitt 7.2.2)

Der Entwurf sieht vor, dass themenspezifisch Selbsthilfeorganisationen zum Informationsbedarf von Betroffenen einbezogen werden können. Die Einbindung von Selbsthilfeorganisationen erscheint regelhaft sinnvoll, um die Patientensicht besser zu erfassen.

Auswahl der dargestellten Ergebnisse (Endpunkte) (Abschnitt 7.5)

Neben den Endpunkten zur Mortalität, Morbidität und gesundheitsbezogenen Lebensqualität sind auch weitere Begleitumstände der Behandlung als für Patienten wichtige Informationen genannt.

Stellungnahme

Unberücksichtigt bleiben die Patientcompliance und –adherence. Da diese mitentscheidend für die Therapiequalität sind, sollten beide Begriffe an dieser Stelle ergänzend aufgeführt werden.

Meta-Analysen (Abschnitt 9.3.8)

Die Diskussion neuer methodischen Entwicklungen im Bereich der Durchführung von Meta-Analysen ist grundsätzlich zu begrüßen. Die Berücksichtigung der Knapp-Hartung-Methode im Hinblick auf die Durchführung von Meta-Analysen mit zufälligen Effekten ist adäquat.

Es wäre in der Folge wichtig, dass die Dossievorlagen des G-BA für die frühe Nutzenbewertung von Arzneimittel angepasst werden, da diese derzeit noch die DerSimonian and Laird-Methode empfehlen.

Im Bereich der frühen Nutzenbewertung von Arzneimitteln liegt in der Regel eine begrenzte Anzahl von Studien vor. Es ist vor dem Hintergrund der Limitationen von Meta-Analysen mit zufälligen Effekten sinnvoll und positiv, dass das IQWiG in der überarbeiteten Version des Methodenpapiers alternativ ein Modell mit festem Effekt oder eine qualitative Zusammenfassung positiv würdigt. Die Formulierung, dass diese Alternativen „in Erwägung gezogen werden“ sollten, ist allerdings sehr schwach. Es wäre wünschenswert, wenn deutlicher gemacht wird, dass bei Vorliegen weniger Studien ein Modell mit fixen Effekten angewendet und nur in Ausnahmefällen auf ein Modell mit zufälligen Effekten ausgewichen werden soll.

Das IQWiG wird gebeten, zusammen mit dem G-BA prüfen, inwieweit sich aus den Änderungen im Entwurf des Methodenpapiers Anpassungsbedarf im Hinblick auf die Dossievorlage des G-BA für Verfahren der frühen Nutzenbewertung von Arzneimitteln ergibt.

Indirekte Vergleiche (Abschnitt 9.3.9)

Das IQWiG nimmt zu einem aus seiner Sicht adäquaten Vorgehen bei indirekten Vergleichen Aktualisierungen vor, die Aussagen bleiben im Ergebnis aber weiterhin unkonkret.

Es fehlen Hinweise, was aus Sicht des Instituts „adäquate Brückenkomparatoren“ sein können. Auch bleibt unklar, was eine „adäquate Begründung“ sein könnte, um auf Methoden für indirekte Vergleiche zurückgreifen zu können.

Es ist zuzugestehen, dass vielfach ein anerkannter methodischer Standard in diesem Bereich noch nicht etabliert ist. Dennoch wäre es zu begrüßen, wenn das IQWiG hier aus seiner Erfahrung mit indirekten Vergleichen in einer Vielzahl von Verfahren der frühen Nutzenbewertung konkreter werden könnte.

Viele indirekte Vergleiche im Bereich der frühen Nutzenbewertung von Arzneimitteln scheitern beim IQWiG aufgrund einer nicht ausreichenden Ähnlichkeit der einzubeziehenden Studien. Leider bleibt das Methodenpapier auch zu dieser Frage unkonkret.

Die Aussagen in diesem Abschnitt bleiben im Hinblick auf planbare und verlässliche Rahmenbedingungen für die Verwendung von indirekten Vergleichen bspw. in Verfahren der frühen Nutzenbewertung von Arzneimitteln unzureichend. Dem IQWiG ist bekannt, dass in einer Vielzahl von Verfahren im Bereich der frühen Nutzenbewertung zum Teil mit großem

Stellungnahme

Ressourcenaufwand indirekte Vergleiche erstellt und vorgelegt worden sind. Konkrete Vorgaben für ein adäquates methodisches Vorgehen bei indirekten Vergleichen sind essentiell. Eine verstärkte Beteiligung des Instituts bei der Entwicklung angemessener Antworten auf die offenen methodischen Fragen bleibt aus Sicht des Verbandes prioritär.

Die im Entwurf des Methodenpapiers weiterhin erkennbare Ablehnung nicht adjustierter indirekter Vergleiche ist unter Berücksichtigung der Bewertungspraxis des Instituts nicht nachvollziehbar, da in Frühbewertungsverfahren von Arzneimitteln (bspw. Ledipasvir/Sofosbuvir) solche indirekten Vergleiche unter bestimmten Voraussetzungen bereits akzeptiert wurden. Der Zusatznutzenbewertung ist bereits zu entnehmen, dass ein nicht adjustierter indirekter Vergleich bspw. akzeptiert werden kann, wenn der beobachtete Effekt eine Größenordnung hat, nach der ausgeschlossen werden kann, dass er allein auf systematischer Verzerrung beruht. Bereits akzeptierte Kriterien für nicht adjustierte indirekte Vergleiche sollten auch Eingang in das aktualisierte Methodenpapier finden.

Subgruppenanalysen (Abschnitt 9.3.10)

Das IQWiG setzt sich in diesem Abschnitt intensiv mit den Limitationen von Subgruppenanalysen auseinander, insbesondere wenn diese post-hoc durchgeführt werden. Dennoch wird die Notwendigkeit solcher Analysen behauptet: sie seien „für manche Fragestellungen die auf absehbare Zeit beste wissenschaftliche Evidenz“.

Beim Lesen des Abschnitts entsteht der Eindruck, das IQWiG wolle die bekannten Limitationen von Subgruppenanalysen relativieren. Während es in der aktuellen Version 4.2 des Methodenpapiers bspw. noch heißt, dass Subgruppenanalysen in der Literatur „sehr kritisch diskutiert“ werden, heißt es im Entwurf der Methoden 5.0 nun unter Verweis auf die identischen Quellen, dass Subgruppenanalysen „teilweise kritisch diskutiert“ werden.

Im Vergleich zu den „Allgemeinen Methoden 4.2“ wird auch das Problem des multiplen Testens bei Subgruppenanalysen im aktuellen Entwurf nicht mehr thematisiert.

Im Ergebnis entsteht der Eindruck, dass die bekannten und vielfach diskutierten methodischen Limitationen bei Subgruppenanalysen unter bestimmten Bedingungen „keine Geltung haben“ könnten. Das IQWiG als wissenschaftliches Institut sollte diesen Eindruck, der sich insbesondere im Vergleich zu den eigenen Äußerungen in den „Allgemeinen Methoden 4.2“ aufdrängt, unbedingt vermeiden.

A.1.9 – Bundesverband Medizintechnologie e. V. (BVMed)

- 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 2017

Stellungnahme

zum

Entwurf des allgemeinen Methodenpapierses Version 5.0 des IQWiG

Die vorliegende Stellungnahme gibt die Auffassung der Mitglieder des Fachbereiches Market Access der AG Methodenbewertung des Bundesverbands Medizintechnologie e. V. (BVMed) wieder.

Die vorliegende Stellungnahme nimmt Bezug auf das vom IQWiG neu aufgenommene Kapitel „Bewertung von Medizinprodukten hoher Risikoklassen“. Wir bitten jedoch, zukünftig die korrekte Formulierung des § 137h SGB V „Bewertung neuer Untersuchungs- und Behandlungsmethoden mit Medizinprodukten hoher Risikoklasse“ anzuwenden.

I. Grundsätzliche Anmerkung

Anzahl von Bewertungen von Untersuchungs- und Behandlungsmethoden mit Medizinprodukten steigt

Die Einführung der Bewertung neuer Untersuchungs- und Behandlungsmethoden mit Medizinprodukten hoher Risikoklassen und die ggf. Erprobung dieser Methoden wird die Anzahl der Bewertungen nicht-pharmakologischer neuer Untersuchungs- und Behandlungsmethoden mit Medizinprodukten hoher Risikoklassen durch das IQWiG deutlich erhöhen. Die Bearbeitung muss durch das IQWiG zeitnah erfolgen.

BVMed bietet Expertise für das Kuratorium an

Die Besonderheiten der Bewertung neuer Untersuchungs- und Behandlungsmethoden mit Medizinprodukten hoher Risikoklassen sollen, im Gegensatz zu den Methoden der Nutzenbewertung von Arzneimitteln, ausreichend in der Arbeitsweise des IQWiG berücksichtigt werden. Der BVMed sollte zukünftig für die Medizinprodukteindustrie aktiv beratend im Kuratorium der Stiftung für Qualität und Wirtschaftlichkeit im Gesundheitswesen vertreten sein, um die Expertise der Medizinprodukteindustrie einzubringen.

II. Zu den einzelnen Kapiteln

Zu 2.1.6

Bewertung gem. § 137 h SGB V

Die Veröffentlichung des Berichts des Institutes sollte zeitgleich mit der Übermittlung an den G-BA erfolgen und nicht erst sechs Wochen später. Alle Beteiligten im anschließendem Stellungnahmeverfahren des G-BA sollen ausreichend Zeit haben, die Inhalte des Berichts angemessen zu berücksichtigen. Für die Potentialbewertung gem. § 137h SGB V ist kein öffentliches Stellungnahmeverfahren vorgesehen. Grundsätzlich ist daher bei den unterschiedlichen IQWiG-Produkten ein öffentliches Stellungnahmeverfahren vorzusehen.

Zu 2.2.7

Wissenschaftliche Beratung

Das IQWiG ist Mitglied im Europäischen Netzwerk für Health Technology Assessment (EUnetHTA). Dem BVMed stellt sich nun die Frage, ob das IQWiG in diesem Zusammenhang am o. g. Beratungsprozess beteiligt war/ist. Es ist zu klären, ob die Beratungsleistung des EUnetHTA geeignet ist, europäische Studienanforderungen, inklusive die des G-BA bzw. des IQWiG, verbindlich für Methoden mit Medizinprodukten zu beantworten. Grundsätzlich bleibt die Frage nach dem Einklang der vom Institut eingeführten Methodik und den internationalen Standards. Der Gesetzesauftrag gemäß §139a SGB V fordert eindeutig, „... dass die Bewertung des medizinischen Nutzens nach den international anerkannten Standards der evidenzbasierten Medizin ... erfolgt.“

In diesem Zusammenhang bitten wir um Beantwortung folgender Fragen:

- > Plant das EUnetHTA mit aktiver Unterstützung des IQWiG auch für die Hersteller von Medizinprodukten eine Beratung zu Studienanforderungen an Methoden mit Medizinprodukten?
- > Wäre ein Beratungsgespräch des EUnetHTA für die EU-Mitgliedsstaaten als verbindlich anzusehen?
- > Wird eine derartige Anfrage aktiv auch vom G-BA bzw. vom IQWiG begleitet? Beinhaltet diese eine verbindliche Aussage zu einem möglichen Studiendesign?
- > Ist damit eine G-BA-Beratungsanfrage zu einer Methodenbewertung gem. § 137e SGB V und § 137h SGB V bei Vorliegen eines EUnetHTA-Beratungsergebnis redundant?
- > Wie kann eine aktive Beteiligung der Hersteller an der wissenschaftlichen Beratung im Rahmen der Mitarbeit im europäischen Netzwerk EUnetHTA gesichert werden? Wie bringt sich das IQWiG in dieser Fragestellung konkret ein?

Zu 3.4

Nichtmedikamentöse therapeutische Interventionen

Es ist nicht zielführend, an die Bewertung neuer Untersuchungs- und Behandlungsmethoden mit Medizinprodukten hoher Risikoklassen die gleichen Maßstäbe anzulegen wie an die Bewertung von Arzneimitteln. Die Besonderheiten der Methoden mit Medizinprodukten müssen bei einer Methodenbewertung berücksichtigt werden. Da dies nur indirekt Aufgabe des IQWiG ist, benennen wir hier nur exemplarisch und in Ergänzung zu den Ausführungen im Methodenpapier wichtige Aspekte:

> **Schnelle Innovationszyklen:**

Die Durchführung einer RCT ist zeitintensiv und mit einer langen Studiendauer verbunden. Das birgt das Risiko, dass eine Studie noch durchgeführt wird, obwohl bereits Neuentwicklungen oder Weiterentwicklungen der eingesetzten Medizinprodukte zur Verfügung stehen.

> **Rigidität:**

Es besteht kaum die Möglichkeit, im Rahmen einer herkömmlichen randomisierten kontrollierten Studie auf Veränderungen zu reagieren oder neue Erkenntnisse bei der Durchführung einfließen zu lassen. Eine Anpassung des Studiendesigns ist schwer umsetzbar.

> **Lernkurveneffekte und Anwenderqualifikation:**

Die klinischen Ergebnisse von Methoden mit Medizinprodukten hängen nicht ausschließlich von den Medizinprodukten selbst, sondern oft auch von der Ausbildung, Kompetenz und Erfahrung der Endverbraucher bzw. Anwender ab. Die Lernkurveneffekte von z. B. Chirurgen, Operationsteam und Krankenhaus haben potenziell einen Einfluss auf die Behandlungsergebnisse.

In der Verfahrensordnung des G-BA sind verschiedene Formen der Beteiligung an der Erprobung gemäß § 137h SGB V zur Bewertung neuer Untersuchungs- und Behandlungsmethoden mit Medizinprodukten hoher Risikoklassen vorgesehen. Neben einer randomisierten kontrollierten Studie, falls diese durchführbar ist, sind im Sinne einer Beobachtungsstudie flankierende Daten zur Wirksamkeit und Sicherheit zu erheben. In diesem Abschnitt sollte erwähnt und in einem eigenen Kapitel thematisiert werden, wie diese Ergebnisse bei der Bewertung neuer Untersuchungs- und Behandlungsmethoden mit Medizinprodukten hoher Risikoklassen zu berücksichtigen sind.

Zu 3.8

Potentialbewertung

In der Verfahrensordnung des G-BA sind verschiedene Formen der Beteiligung an der Erprobung gemäß § 137h SGB V zur Bewertung neuer Untersuchungs- und Behandlungsmethoden mit Medizinprodukten hoher Risikoklassen vorgesehen. Für die dort aufgeführten Beobachtungsstudien zur Erhebung flankierender Daten zur Wirksamkeit und Sicherheit sind entsprechende Hinweise an geeigneter Stelle hinzuzufügen.

Formulierungsvorschlag des BVMed:

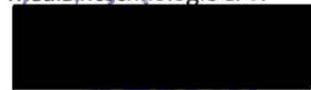
Die Planbarkeit einer Studie stellt bei Bewertungen gemäß § 137h SGB V kein entscheidungsrelevantes Kriterium für eine Potenzialaussage dar. Es werden in den eingereichten Unterlagen enthaltene Vorschläge für eine Erprobungsstudie und die in der Verfahrensordnung des G-BA vorgesehenen Beobachtungsstudien geprüft bzw. Eckpunkte einer möglichen Erprobung und Beobachtung konkretisiert.

III. Abschließende Anmerkung

Die Arbeitsweise des IQWiG muss sich an die Dynamik und Besonderheiten der Technologien auch in Zusammenhang mit der Bewertung neuer Untersuchungs- und Behandlungsmethoden mit Medizinprodukten hoher Risikoklassen anpassen. Die derzeitige Ausgestaltung des Methodenpapiers ist nicht ausreichend. Die Besonderheiten der Bewertung neuer Untersuchungs- und Behandlungsmethoden mit Medizinprodukten hoher Risikoklassen sind in der Arbeitsweise des IQWiG zu berücksichtigen.

Mit freundlichen Grüßen

BVMed – Bundesverband
Medizintechnologie e. V.



Joachim M. Schmitt
Geschäftsführer
Mitglied des Vorstands



Olaf Winkler
Leiter Referat Gesundheitssystem

A.1.10 – Deutsche Diabetes Gesellschaft e. V. (DDG) / Deutsche Gesellschaft für Chirurgie e. V. (DGCH) / Deutsche Gesellschaft für Epileptologie e. V. / Deutsche Gesellschaft für Gynäkologie und Geburtshilfe e. V. (DGGG) / Deutsche Gesellschaft für Hämatologie und Medizinische Onkologie e. V. (DGHO) / Deutsche Gesellschaft für Kardiologie – Herz- und Kreislaufforschung e. V. (DGK) / Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e. V. (GMDS) / Deutsche Gesellschaft für Neurologie e. V. (DGN) / Deutsche Gesellschaft für Pneumologie und Beatmungsmedizin e. V. (DGP) / Deutsche Gesellschaft für Rheumatologie e. V. (DGRh) / Deutsche Region der Internationalen Biometrischen Gesellschaft (IBS-DR) / (gemeinsame Stellungnahme)

Autoren:

- Wörmann, Bernhard

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

24. Januar 2017

**Gemeinsame Stellungnahme zu
IQWiG Allgemeine Methoden
Entwurf Version 5.0**



Deutsche
Gesellschaft
für Chirurgie



Deutsche Gesellschaft für
Epileptologie e.V.
Deutsche Sektion der Internationalen Liga gegen Epilepsie



Deutsche Gesellschaft
für Gynäkologie und
Geburtshilfe e.V.



DGK.
Deutsche Gesellschaft für Kardiologie
– Herz- und Kreislaufforschung e.V.



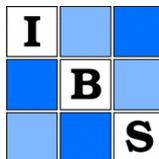
Deutsche Gesellschaft für Medizinische Informatik,
Biometrie und Epidemiologie e.V.



DGN
Deutsche Gesellschaft
für Neurologie



Deutsche Gesellschaft
für Rheumatologie e.V.



Internationale Biometrische
Gesellschaft

Inhaltsverzeichnis

- Einleitung
- Einbindung von Betroffenen (Kapitel 2.2.1.)
- Endpunktbezogene Bewertung (Kapitel 3.1.4.)
- Früherkennung und Screening (Kapitel 3.6.)
- Leitlinien und Versorgungsanalysen (Kapitel 5)
- Gesundheitsinformationen (Kapitel 7)
- Informationsbeschaffung (Kapitel 8)
- Subgruppenanalysen (Kapitel 9.3.10)
- Was fehlt im Methodenpapier 5.0?
 - Umgang mit Addenda
 - Einbindung von Experten
 - Diagnostische Verfahren
 - Bewertung von Medizinprodukten
 - Rationale der Methodik zur Feststellung des Ausmaßes des Zusatznutzens

Einleitung

Das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) legt nach der Veröffentlichung seines Methodenpapiers 4.2 vom April 2015 jetzt den Entwurf der nächsten, jährlich geplanten Aktualisierung vor. Schwerpunkte sind Leitliniensynopsen und Versorgungsanalyse, Gesundheitsinformationen, Informationsbeschaffung und Subgruppenanalysen. Dazu wurde eine Reihe weiterer, inhaltlicher Änderungen vorgenommen. Unsere Anmerkungen zu den Änderungen sind:

Einbindung von Betroffenen (Kapitel 2.2.1.)

Wir begrüßen die aktuelle Änderung der Einbindung von „tatsächlich Betroffenen“. Es fehlt allerdings eine verbindliche und nachvollziehbare Strategie zur tatsächlichen Einbindung der Patienten. Zahlreiche Berichte der frühen Nutzenbewertung, auch der jüngeren Vergangenheit, wurden ohne Einbindung von Patienten(-vertretern) erstellt. Wie wird sichergestellt, dass die noch höhere Anforderung der Identifikation und Motivation von „tatsächlich Betroffenen“ erfüllt werden kann?

Endpunktbezogene Bewertung (Kapitel 3.1.4.)

Die Änderung in diesem Kapitel bezieht sich auf Bedingungen, die zum Beleg eines Zusatznutzens aus einer einzigen Studie erfüllt werden müssen. Die bisherigen Bedingungen (multizentrisch, mindestens 10 Studienzentren) werden ergänzt durch mindestens „1000 Patienten“. Die Festlegung einer Zahl ist sinnvoll, allerdings fehlt die Begründung für die Festlegung der Grenze bei „1000“.

Früherkennung und Screening (Kapitel 3.6.)

Die Aufnahme der „Überdiagnose“ als ein Element des potenziellen Schadens der Früherkennung ist sinnvoll. Nicht konsequent ist allerdings die Anerkennung einer „Überdiagnose“ als Schaden, wenn gleichzeitig die Diagnose einer Erkrankung an sich vom IQWiG nicht als nützlich betrachtet wird, siehe Kapitel 3.5, Diagnostische Tests. Hier müssen die gleichen Kriterien für Diagnose und Überdiagnose gelten, siehe auch Abschnitt „Was fehlt im Methodenpapier 5.0?“ am Ende dieser Stellungnahme.

Leitlinien und Versorgungsanalysen (Kapitel 5)

Aus dem Kapitel „Leitlinien und Versorgungsanalyse“ ist jetzt ein Kapitel „Leitliniensynopsen und Versorgungsanalyse“ geworden. Die Beschreibung der Methoden ist umfangreich und nachvollziehbar. Die Basierung der Bewertung von Leitlinien auf den Empfehlungen von AGREE und dem AGREE-II-Instrument mit 6 Domänen ist sinnvoll.

Die Probleme der derzeitigen Leitlinienentwicklung (Aktualität, Zeitaufwand, Umfang, Umsetzung im Versorgungsalltag und Einflussnahme von Interessengruppen) werden allerdings in diesem Methodenpapier nicht wahrgenommen. Auch die inhaltliche Korrektheit von Leitlinien steht hinter der formalen Korrektheit zurück.

Gesundheitsinformationen (Kapitel 7)

Das Thema Gesundheitsinformationen stand schon im Zentrum der letzten Aktualisierung. Wir haben nicht den Eindruck, dass die Akzeptanz und der Durchdringungsgrad der vom IQWiG angebotenen Gesundheitsinformationen in den letzten Jahren spürbar gestiegen sind. Vielleicht sind eine grundsätzliche Hinterfragung des Konzeptes und auch der eingesetzten Ressourcen wichtiger als methodische Anpassungen.

Informationsbeschaffung (Kapitel 8)

Die Kriterien der Informationsbeschaffung werden etwas ausführlicher dargestellt.

Ein Herausforderung aus ärztlicher Sicht ist die zunehmend häufige Veröffentlichung relevanter, auch Standard-verändernder Studienergebnisse auf Kongressen und/oder in verschiedenen Medien, auch zum Zweck der Markteinführung von Arzneimitteln bei den Zulassungsbehörden, bevor sie in Peer-Review-Journals

veröffentlicht werden. Diese Veröffentlichungen in Kurzform betreffen auch Daten von Studien, die ihr Ziel nicht erreichten und nur auf Kongressen vorgestellt, aber nie voll publiziert werden. Wünschenswert wäre die Entwicklung einer Strategie und einer Methodik seitens des IQWiG, solche Daten grundsätzlich in die Informationsbeschaffung zu integrieren und ggf. auch zu bewerten.

Subgruppenanalysen (Kapitel 9.3.10)

In die Überarbeitung dieses Kapitels wurden wesentliche Vorschläge aus der Kommission Nutzenbewertung der AWMF aufgenommen, die auch beim gemeinsamen Workshop am 21. September 2016 in Berlin diskutiert worden waren. Als zentrale Punkte der Akzeptanz von Subgruppen wurden deren Festlegung *a priori* und ihre Verwendung als Stratifikationsmerkmal festgeschrieben. Ebenfalls wurde das Niveau für eine Effektmodifikation als das Ergebnis eines Heterogenitäts- oder Interaktionstests zwischen wichtigen Subgruppen auf $\alpha = 0,05$ gesenkt. Damit sinkt die Zahl der Verfahren in der frühen Nutzenbewertung, in denen Subgruppen allein aufgrund der Identifikation einer Effektmodifikation diskutiert werden. Wir bedanken uns für die konstruktive Diskussion und die Aufnahme der Vorschläge.

Was fehlt im Methodenpapier 5.0?

Hier müssen wir leider einen Großteil der schon früher schon geäußerten Kritikpunkte aufgreifen, und noch etwas erweitern:

Umgang mit Addenda (Kapitel 2.1.7.)

In letzter Zeit ist die Zahl der vom G-BA an das IQWiG ergangenen Aufträge zur Erstellung von Addenda im Rahmen der frühen Nutzenbewertung neuer Arzneimittel deutlich angestiegen. 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 5.0, 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.

Einbindung von Experten (Kapitel 2.2.2.)

Das IQWiG hat Probleme, qualifizierte Experten für ihre Berichte zu finden. Wir halten eine grundlegende Überarbeitung des Expertenkonzeptes mit dem Ziel offener Diskussionen anstelle der Einholung von medizinischer Expertise auf der Basis von Fragenkatalogen für erforderlich.

Diagnostische Verfahren (Kapitel 3.5)

Das Methodenpapier geht von der Grundannahme aus, dass der Wert diagnostischer Verfahren ausschließlich durch sich anschließende therapeutische oder präventive Verfahren definiert wird. Im Kapitel „Früherkennung und Screening“ wird bei der „Überdiagnose“ allerdings von dieser Festlegung abgewichen.

Wir halten die „Diagnosesicherung“ ebenfalls für ein wichtiges Ziel der Diagnostik, auch ohne unmittelbar anschließende, therapeutische oder präventive Verfahren. Die Sicherung einer Diagnose hilft dem Patienten beim Umgang mit der Erkrankung, vermeidet die Durchführung weitergehender (überflüssiger) Untersuchungen, ist Voraussetzung für die Kodierung von Erkrankungen nach dem ICD-Code, und hat sozialmedizinische Bedeutung.

Bewertung von Medizinprodukten

Die Bewertung von Medizinprodukten ist eine neue Aufgabe des G-BA. Auch in diesem Bereich können Berichtsaufträge an das IQWiG erteilt werden. Wir halten es für sinnvoll, das Methodenpapier 5.0 zu erweitern und die Bewertungskriterien für Medizinprodukte zu integrieren. Vorschläge für ein abgestuftes Verfahren wurden bereits von der Deutschen Gesellschaft für Chirurgie und der Deutschen Gesellschaft für Biomedizinische Technik erarbeitet, siehe http://www.dgch.de/fileadmin/media/presse/Pressemappe_DGCH_PK_Dez_2015.pdf.

Anhang A (Rationale der Methodik zur Feststellung des Ausmaßes des Zusatznutzens)

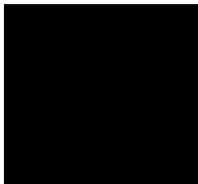
Die methodische Basis der Festsetzung von Operationalisierungskriterien wurde zwischenzeitlich publiziert (Skipka et al., 2016). Es fehlen allerdings weiterhin Ansätze für eine inhaltliche Validierung der festgelegten Grenzwerte.

Es fehlt eine umfassende Auseinandersetzung mit den verschiedenen Methoden zur Berücksichtigung von Postprogressionstherapien einschl. des Switching (Crossover) bei der Berechnung der Überlebenszeit.

Das seit Beginn des AMNOG-Verfahrens ungelöste Problem der fehlenden Publikation und Nachvollziehbarkeit des zugrunde liegenden Algorithmus in der verwendeten Software SAS bleibt bestehen.

Für Rückfragen stehen wir jederzeit gern zur Verfügung.

Mit freundlichen Grüßen



Prof. Dr. med. Bernhard Wörmann
Medizinischer Leiter der DGHO
Alexanderplatz 1
10178 Berlin
Tel.: 030 27876089- 0
Fax: 030 27876089-18



A.1.11 – Deutsche Gesellschaft für Hebammenwissenschaft e. V. (DGHWi)

Autoren:

- Mattern, Elke

Anmerkungen
der Deutschen Gesellschaft für Hebammenwissenschaft e.V.
zum
Entwurf für Version 5.0 vom 07.12.2016
Allgemeine Methoden
des IQWiG

Kapitel 2 Produkte des Instituts

Kapitel 2.2.1 Einbindung von Betroffenen in die Erstellung von IQWiG-Produkten:

S. 35, 1. Absatz „... findet primär zu Beginn der Berichtsbearbeitung im Rahmen der Festlegung patientenrelevanter Endpunkte und relevanter Subgruppen statt.“

→ was ist hier unter primär zu verstehen? Ist eine Einflussnahme der Betroffenen im weiteren Verlauf oder bei der Ergebnisvorstellung noch möglich?

S. 35, 2. Absatz „Betroffene können insbesondere sein: Patientinnen, Patienten, Eltern oder sonstige Angehörige sowie potentielle Teilnehmer an Präventionsverfahren.“

→ Es wird in dem gesamten Dokument von Patientinnen und Patienten gesprochen, dabei ist einzubeziehen, dass nicht alle Personen die Leistungen der GKV in Anspruch nehmen, Patientinnen bzw. Patienten sind, sondern auch Personen die Routinemaßnahmen, Präventionsleistungen oder geburtshilfliche Leistungen in Anspruch nehmen. In diesen Fällen ist der Begriff *Patientin / Patient* nicht zutreffend.

S. 35, 4. Absatz „Zur Findung Betroffener wird regelhaft die Patientenvertretung im G-BA gebeten, selbst oder über ihre Mitgliedsorganisationen Betroffene zu einem Thema zu benennen.“

→ Die Einbeziehung werdender und junger Mütter scheint mir hier nicht gegeben. Die Auswahl über die Patientenvertretungen des G-BA schließt ganz normale Mütter aus.

→ Darüber hinaus ist nicht genau definiert, wie die einzelnen Verfahren der Einbeziehung der Betroffenen gestaltet sind. Was bedeutet mündliche Konsultation oder schriftliche Konsultation und welche Verbindlichkeit haben diese?

Kapitel 3. Nutzenbewertung medizinischer Interventionen

→ Im gesamten Kapitel 3 sollte durchgängig sowohl der Nutzen als auch der Schaden explizit ausgewiesen und diskutiert werden, da der Schaden nicht nur die Abwesenheit eines Nutzens oder einen geringen Nutzen darstellt. Die Überschrift des Kapitels sollte daher lauten "Nutzen- und Schaden-Bewertung medizinischer Interventionen".

→ Insgesamt ist das Kapitel sehr langatmig, da sich die Erklärungen zu Verfahren der Nutzenbewertung wiederholen. Vielleicht wäre es möglich, die Grundzüge der Verfahren und Bewertungskriterien voranzustellen und in den einzelnen Kapiteln darauf zu verweisen?

Kapitel 3.1.1 Definition des patientenrelevanten medizinischen Nutzens bzw. Schadens

auf Seite 42, vorletzter Absatz sollte "(bzw. Schadens)" nicht in Klammern angegeben werden, sondern gleichwertig behandelt werden - wie auf S. 48 im vorletztem Abschnitt.

Kapitel 3.1.3 Ermittlung des Schadens medizinischer Interventionen

Seite 47, 2. Absatz (Mitte): "Gerade zur Erfassung unerwartet auftretender unerwünschter Ereignisse ist jedoch eine generelle Abfrage des Befindens der Patientinnen und Patienten erforderlich".

→ Hier ist unklar, was unter "generelle Abfrage" zu verstehen ist. Sollte hier nicht eher eine "offene Abfrage" vorgesehen werden, in dem Sinne, dass die Patientinnen und Patienten auch unerwünschte Ereignisse angeben können, welche nicht durch vorformulierte "geschlossene Statements" vorgegeben sind? - Dadurch hätten die Patientinnen und Patienten auch die Möglichkeit, unerwünschte Ereignisse anzugeben, welche ihnen besonders wichtig sind oder ihre Lebensqualität beeinträchtigen, die jedoch durch die Forscherinnen und Forscher in der Abfrage nicht vorhergesehen wurden.

Seite 47, 3. Absatz: "Daraus folgend ist es notwendig, die Aufarbeitung und Darstellung auf relevante unerwünschte Wirkungen zu beschränken."

→ Diese Aussage ist nicht nachvollziehbar. Wie wird entschieden, welche unerwünschten Wirkungen relevant sind? Wäre es hier nicht wichtig, Patienten-Vertreterinnen und -Vertreter einzubeziehen, die bei dieser Entscheidung mitwirken?!

→ Außerdem ist nicht nachvollziehbar, warum nicht auch "unerwünschte Ereignisse" mit berücksichtigt werden, die für Patientinnen und Patienten von Bedeutung sind. Diese könnten ja möglicherweise unerwünschte Wirkungen sein, die zunächst nicht als solche erkannt werden, aber in Zukunft ggf. als solche bestätigt werden könnten.

Seite 48, Absatz vor der nächsten Überschrift: *"Diese Zusammenstellung potenziell relevanter unerwünschter Wirkungen bildet die Grundlage für die Bewertung des Schadens auf Basis der in die Nutzenbewertung eingehenden Studien."*

→ Diese Aussage greift zu kurz; hier fehlen die Ergebnisse von qualitativen (interpretativ-hermeneutischen) Studien zum Erleben und zur Lebensqualität von Patientinnen und Patienten! D.h.: es müssten auch (interpretativ-hermeneutische) Studien zur Ermittlung des Nutzens und des Schadens herangezogen werden.

Kapitel 3.2.4 Patienten berichtete Endpunkte

Seite 57, 2. Absatz *"Da Angaben zu PROs aufgrund ihrer Natur subjektiv sind, sind offene Studien in diesem Bereich nur von eingeschränkter Validität."*

→ Es ist unklar, was mit "offene Studien" gemeint ist (interpretativ-hermeneutische Studien? Nicht-verblindete Studien?)

Kapitel 3.3.3 Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V

→ Insgesamt: Es ist auffällig, dass in diesem Kapitel zur Nutzenbewertung von Arzneimitteln die Erhebung und Bewertung des Schadens der pharmazeutischen Interventionen so gut wie keine Rolle spielen. Hier wird ausschließlich von Nutzenbewertung und Zusatznutzen gesprochen, aber nicht vom (potenziellen) Schaden.

→ D.h. die Bewertung des Schadens durch Arzneimittel-Interventionen sollte eigens und ebenbürtig mit der Einschätzung des Nutzens thematisiert werden - auch im Sinne der Transparenz und klareren Unterscheidung von Nutzen und Schaden (d.h. "Schaden" als eigenes Konstrukt und nicht als Nicht-Nutzen oder Nebenwirkungen)!

→ Seite 61 bis 62: Im Dossier des pharmazeutischen Unternehmers sollte bereits eine Auskunft zum Schaden / potenziellen Schaden gefordert werden!

Kapitel 3.5 Diagnostische Verfahren,
Kapitel 3.6 Früherkennung und Screening
sowie Kapitel 3.7 Prävention:

→ Insbesondere in diesen Kapiteln ist der Begriff Patientin/Patient nicht zutreffend, da es sich - insbesondere bei der Primär- und Sekundärprävention - nicht um Patientinnen/Patienten, sondern um diesbezüglich gesunde Personen handelt.

Kapitel 4. Kosten-Nutzen-Bewertung medizinischer Interventionen

→ Allgemeiner Eindruck

Die Methoden werden sehr ausführlich und verständlich dargelegt. Die ausführliche Darstellung kann jedoch auch zu einer Verwässerung kritischer Punkte führen. Zum Teil wirken *die Methoden* rechtfertigend und gehen weit über eine Darstellung und Festsetzung der Methodik hinaus.

Beispiel S. 83, letzte Absatz aus: 4.2.1 Grundlagen

„Gesundheitsökonomische Modelle sind somit, wie mathematisch-formalisierte Modelle, auch eine vereinfachte Abbildung der Realität. Durch eine bewusste Reduktion der Komplexität auf die für das Entscheidungsproblem relevanten Entscheidungsfaktoren und -variablen wird zudem analytische Klarheit geschaffen.“

Kapitel 4.1.3. Zeithorizont

→ Wieso wird nicht über einen längeren Zeitraum modelliert? Und diese Modellierung gesondert ausgewiesen? Siehe auch S. 80, 4.1.6. Modell (evtl. missverständliche Formulierung?)

Kapitel 4.1.5 Versorgungspfad

„Weiterhin sind aktuelle Therapieempfehlungen für Deutschland darzustellen, wobei gültige Leitlinien herangezogen werden sollten.“

→ Prüfung des Einbezuges auch internationaler Leitlinien zur Darstellung des Versorgungspfades, wenn keine deutschen vorhanden sind. Da alltagspraktisch auch auf internationale zurückgegriffen wird.

Kapitel 4.3 Nutzen

→ Die Methoden, mit denen der Nutzen im Rahmen der Nutzenbewertung bestimmt wird, werden in Kapitel 3 beschrieben. In einer Kosten-Nutzen-Bewertung (KNB) kann der studienbelegte Nutzen vom modellierten Nutzen unterschieden werden, wenn einer KNB ein längerer als der studienbelegte Zeithorizont zugrunde gelegt wird.

Kapitel 4.3.3 Maß des Gesamtnutzens

S 90, vorletzte Absatz, 2. Satz *„QALYs sollten nur verwendet werden, wenn die einfließenden Werte zu den Gesundheitszuständen zunächst bei den Betroffenen erhoben worden sind, die aktuell oder in der Vergangenheit diese Gesundheitszustände erlebt haben. Die Daten sollten bei den an klinischen Studien Teilnehmenden erhoben worden sein.“*

→ Eventuell missverständlich: Wieso sollten die Daten ausschließlich bei an klinischen Studien Teilnehmenden erhoben worden sein? Soll beschrieben werden, dass gleichzeitig Daten aus klinischen Studien zum Gesundheitszustand und Daten zur Lebensqualität erhoben werden sollen?

Kapitel 4.4.2 Abgrenzung der Kosten

S. 95, letzte Absatz *„Die meisten empirischen Studien berücksichtigen die Auswirkungen auf die Freizeit von Betroffenen und Angehörigen nicht. Insofern bezieht das Institut nicht regelhaft den Zeitaufwand für Betroffene und ihre Angehörigen in die gesellschaftliche Perspektive ein. Für den Fall, dass dennoch repräsentative und valide Informationsquellen zum Zeitaufwand vorliegen, kann dieser Zeitaufwand in Sensitivitätsanalysen zur Perspektive der Gesellschaft berücksichtigt werden.“*

→ Zuvor wurde festgesetzt, dass das Modell nicht abhängig von der verfügbaren Datenlage gewählt werden sollte, daher erscheint es nun inkonsequent, dass ausgehend von einer mangelnden Datenlage ein *kann* statt ein *wird* oder *soll* festgeschrieben wird.

→ Es ist nicht deutlich, ob und inwiefern *externe Kosten* grundsätzlich berücksichtigt werden, auch wenn zum Teil Kosten in Form des Zeitaufwandes von Angehörigen erwähnt werden.

S. 96, B) Indirekte Kosten, letzte Absatz „International wird diskutiert, auch unbezahlte Arbeit (z. B. Hausarbeit) in einer KNB zu berücksichtigen. Das Institut sieht dies im Regelfall vorerst nicht vor.“

→ Vor allem unter der Berücksichtigung von Genderaspekten und auf Grundlage einer fehlenden Argumentation des Instituts erscheint es nicht sinnvoll., die Kosten für unbezahlte Arbeit nicht zu berücksichtigen.

→ Eine präzisere Beschreibung der indirekten Kosten wäre wünschenswert. Werden zum Beispiel Kosten durch konkurrierende Erkrankungen bei Lebensverlängerung berücksichtigt?

Kapitel 7 Evidenzbasierte Gesundheitsinformation für Bürgerinnen und Bürger

Kapitel 7.2.3 Patientenwege

S. 142, Tabelle 10

→ zweite Zeile: die Reihenfolge der Tätigkeiten im Zusammenhang mit einer Erkrankung sollte überdacht werden. Eher: Selbsthilfe, Informationssuche, Arztbesuch, Einnahme von Medikamenten.

→ **siebte Zeile: in der ersten Spalte müsste es heißen „Kontaktstelle im Gesundheits- und Sozialwesen“**

Bei einer Aufzählung in der zweiten Spalte sind Berufe vergessen. Vielleicht könnte man größere Gruppen zusammenfassen: Therapeuten, Gesundheitsfachberufe.

A.1.12 – Deutsche Gesellschaft für Humangenetik e. V. (GfH)

Autoren:

- Gillessen-Kaesbach, Gabriele

gfh Geschäftsstelle Inselkammerstraße 5 82008 München-Unterhaching

An
Katharina Rommel
Institut für Qualität und Wirtschaftlichkeit
im Gesundheitswesen (IQWiG)
Im Mediapark 8
50670 Köln

per Email an: [REDACTED]

nachrichtlich an:
[REDACTED]

25. Januar 2017

IQWiG - Allgemeine Methoden 5.0

Sehr geehrte Frau Rommel,
vielen Dank für die Möglichkeit zum Methodenpapier des IQWiG Stellung
nehmen zu können.

Im Hinblick auf die für uns bedeutsamen Punkte ist uns Folgendes
aufgefallen:

Das IQWiG fasst bei seinen Analysen diagnostische und prognostische
Untersuchungsmethoden zusammen: *"Im Folgenden werden daher
Verfahren zur Bestimmung diagnostischer oder prognostischer Informationen
gemein-sam als diagnostische Verfahren betrachtet."* (Seite 68).

Da sich die Anforderungen an die genetische Beratung im Kontext
diagnostischer genetischer Untersuchungen/Analysen von denen prädiktiver
deutlich unterscheiden, würden wir demzufolge Unterschiede bei der Kosten-
Nutzen-Analyse erwarten. Wir möchten deshalb anregen, nochmals zu
überdenken, ob die Zusammenfassung von diagnostischen und prädiktiven
Tests bei der Analyse neuerer Diagnostikangebote noch sinnvoll ist und nicht
eher getrennt zu betrachten wäre.

Noch eine kleine Anmerkung am Rande, das GenDG wird in dem
umfangreichen Literaturverzeichnis nicht erwähnt.

Mit freundlichen Grüßen

[Handwritten signature]
[REDACTED]

Prof. Dr. med. Gabriele Gillessen-Kaesbach

Vorsitzende der Deutschen Gesellschaft für Humangenetik

german society of human genetics
www.gfhev.de

Vorsitzende

Prof. Dr. med. Gabriele Gillessen-
Kaesbach, Lübeck

Stellvertretende Vorsitzende

Prof. Dr. med. Markus Nöthen, Bonn
Prof. Dr. rer. nat. Thomas Eggermann,
Aachen

Schatzmeisterin

Prof. Dr. med. Ute Hehr, Regensburg

Schriftführer

Prof. Dr. rer. nat. Uwe Kornak, Berlin

Wissenschaftlicher Beirat

Prof. Dr. med. Dr. Judith Fischer, Freiburg
Prof. Dr. med. Thomas Haaf, Würzburg
Prof. Dr. rer. nat. Eva Klopocki, Würzburg
PD Dr. med. Johannes Schumacher, Bonn
Prof. Dr. med. Bernd Wollnik, Göttingen
Prof. Dr. med. Jörg T. Epplen, Bochum
(Tagungspräsident 2017)
Prof. Dr. Peter Wieacker, Münster
(Tagungspräsident 2018)

Adresse der Vorsitzenden

Institut für Humangenetik
Ratzeburger Allee 160
Haus 72
23538 Lübeck

Geschäftsstelle

Dr. rer. biol. hum. Christine Scholz
Inselkammerstraße 5
82008 München-Unterhaching

[REDACTED]
[REDACTED]
[REDACTED]

[REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]

Vereinsregister München

VR 12341

[REDACTED]
[REDACTED]
[REDACTED]

A.1.13 – Deutsche Gesellschaft für Innere Medizin e. V. (DGIM)

Autoren:

- Fölsch, Ulrich
- Sauerbruch, Tilman

Kommentar der Deutschen Gesellschaft für Innere Medizin zum Methoden-Report (Version 5.0 vom 07.12.2016) des IQWiG

Die Vertreter der DGIM betonen, dass sie mit großem Interesse den sehr sorgfältig erstellten neuerlichen Report zu den Allgemeinen Methoden des IQWiG gelesen haben. Wir möchten ausdrücklich betonen, dass wir es als äußerst positiv sehen, mit welcher Akribie das Institut seine Methoden transparent und detailliert beschreibt, so dass diese für jeden Außenstehenden einsehbar und nachvollziehbar sind. Wir erlauben uns, Ihnen hier einige Notizen/Kommentare zu übermitteln.

1.1. Gesetzliche Aufgaben (S. 3)

Es ist jetzt vorgesehen, dass interessierte Einzelpersonen Bewertungen zu medizinischen Untersuchungs- und Behandlungsmethoden vorschlagen können. Hier schlägt die DGIM vor, dass auch wissenschaftliche Fachgesellschaften solche Vorschläge machen können. Wir gehen davon aus, dass die vorgeschlagenen Themen dann öffentlich gemacht werden.

1.2 Evidenzbasierte Medizin (S. 6)

Es wird darauf eingegangen, dass die Evidenz für einige Themen/Krankheiten lückenhaft ist. Dies trifft naturgemäß auf seltene Erkrankungen besonders zu. Hier bittet die DGIM darzulegen oder zu

klären, ob das Institut auf Daten der Kostenträger (GKV) zurückgreifen kann.

1.2.3 Die Strategie der EbM (S.6)

Ist es dem Institut wirklich möglich, Lebensstiländerungen zu erfassen?

1.2.4 Die Bedeutung der Ergebnissicherheit (S. 8)

Punkt 3: In diesem Abschnitt wird von vornherein davon ausgegangen, dass sichere Empfehlungen nicht möglich sind. Dies kann erst ein mögliches Ergebnis der entsprechenden Prüfung sein. Es erscheint nicht logisch, von vornherein sichere Empfehlungen auszuschließen.

1.2.5 Die Verbindung von Ergebnissicherheit und Qualitätsnähe (S. 8)

Gibt es ein Maß für hohe interne Validität und externe Validität?

Grundlage einer Nutzenbewertung (S. 9)

Punkt 1): Bitte prüfen, ob der Kausalitätsbegriff hier wirklich hergehört und richtig ist.

Real World Trials (S.10)

Wie oft greift das Institut wirklich auf solche Trials zurück?

1.3.2 Die internationalen Standards der Gesundheitsökonomie (S. 13)

Hier wird eine Liste der Vorgaben dargestellt, jedoch geht für den Leser nicht klar hervor, welche Gewichtung das IQWiG unter diesen Faktoren vornimmt.

Produkte des Institutes (S.15)

Hier wäre eine Grafik interessant, wie sich die verschiedenen Produkte prozentual, z. B. für einen Zeitraum von 2 Jahren, verteilen.

2.1.2 Rapid Report (S. 21)

Solche Rapid Reports können gesundheitspolitisch relevant sein.

(Zeitnahe Information zu relevanten Entwicklungen im

Gesundheitswesen). Umso mehr wäre es für die Qualität wichtig, hier (ggf.

auch mit engem zeitlichen Limit) eine externe Beurteilung des Dossiers

durch wissenschaftliche Fachgesellschaften transparent zuzulassen.

2.1.5 Potenzialbewertung (S.26)

Auch hier ist keine systematische Beurteilung durch Fachgesellschaften

vorgesehen. Wir empfehlen diese.

2.1.8 Gesundheitsinformation (S.29)

Hier empfehlen wir eine tabellarische Aufstellung der bisherigen Themen und der vorgesehenen Themen.

3 Nutzenbewertung medizinische Interventionen (S. 43)

Es erscheint sehr schwierig, die gesundheitsbezogene Lebensqualität zu erfassen. Hier sollten noch einmal die harten Kriterien dargelegt werden.

3.2.5 Nutzen und Schaden in kleinen Populationen (S. 57)

Dies ist sicher ein wichtiges und schwer zu lösendes Problem. Dennoch sollte das Institut sich dazu durchringen, eine Definition von seltenen Erkrankungen darzulegen (auch wenn diese international nicht einheitlich ist).

3.4 Nicht-medikamentöse therapeutische Interventionen (S. 67)

Hier wird darauf verwiesen, dass es gerade im nicht-medikamentösen Bereich oft keine aussagekräftigen randomisierten Studien gibt. Ist das wirklich so ohne Weiteres zu belegen? Es werden zunehmend randomisierte Studien publiziert (siehe z. B. auch das Programm klinische Studien der DFG/BMBF).

3.5 Diagnostische Verfahren (Seite 69)

Die Relevanz der Verknüpfung diagnostischer und therapeutischer Interventionen ist sehr plausibel dargelegt. Dennoch wäre zu fragen, ob dies in der bisherigen Berichtspraxis des IQWiG eine große Rolle gespielt hat. Insgesamt wird auf die Problematik sehr detailliert eingegangen, ohne dass der Leser weiß, wie häufig dieses Problem überhaupt bearbeitet wird.

3.6, 3.7, 3.8 Früherkennung, Screening und Potenzialbewertung (S.72-77)

Auch hier wäre es interessant zu erfahren, welche quantitative Rolle diese Produkte des Instituts spielen.

4 Kostennutzen – Bewertung medizinischer Subventionen (S.78 -114)

Diese Bereiche werden in diesem Kapitel nebst Unterkapiteln sehr detailliert ausgeführt. Für den nicht Geschulten ist dieser Text jedoch kaum verständlich. Gleichzeitig sind solche Kostennutzenberechnungen den Unterzeichnern als Produkt des Institutes nie herausragend aufgefallen. Auch hier wären Beispiele interessant. Wird diese Methodik auch für eine Preisfindung im Rahmen des AMNOG-Verfahrens angewandt?

Leitlinien, Synopsen und Versorgungsanalyse (S.115)

Es wird zwar angegeben, dass grundsätzlich alle Leitlinien aus den OECD-Mitgliedsländern in Betracht gezogen werden, in der darauf folgenden Bewertungsmethodik bleibt aber offen, welche nun eigentlich wie herangezogen werden.

5.4. Versorgungsanalyse (S. 121)

Hier wird auf den Abgleich von Soll- und Ist-Zustand abgehoben. Wie klar sind Soll-Zustände definiert und welche Kriterien werden herangezogen?

5.4.5 Mögliche Versorgungsparameter (S. 123)

Hier wird auf Qualitätsindikatoren abgehoben. Sind das Qualitätsindikatoren, die mit dem IQTIG/G-BA abgestimmt sind? Wie werden diese erhoben?

Bestimmung des Versorgungsstandards (S. 127)

Siehe auch oben! Indikatoren für Struktur-, Prozess- und der Ergebnisqualität. Die Kriterien erscheinen nur partiell transparent.

Evidenzbasierte Gesundheitsinformationen für Bürgerinnen und Bürger (S.137)

Hier wird der Grundsatz aufgezeigt, auf direkte Empfehlung zu verzichten, gleichzeitig heißt es unter den Zielen (Seite 138) „Unterstützung einer aktiven und informierten Entscheidungsfindung“. Beide Grundsätze könnten sich widersprechen.

7.2.1 Themenkatalog gem. Generalauftrag (S. 139)

Als Marge werden Erkrankungen mit mindestens 1 % Prävalenz in der Bevölkerung herangezogen. Gibt es eine Auflistung der verschiedenen Krankheiten mit dieser Prävalenz von mindestens 1 % vs. den bisher erstellten Gesundheitsinformationen?



Prof. Dr. Tilman Sauerbruch
Beauftragter der DGIM



Prof. Dr. Ulrich Fölsch
Generalsekretär der DGIM

A.1.14 – Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e. V. (GMDS) / Deutsche Region der Internationalen Biometrischen Gesellschaft (IBS-DR) (gemeinsame Stellungnahme)

Autoren:

- Friede, Tim
- Mansmann, Ulrich



gm^{ds}

Deutsche Gesellschaft für
Medizinische Informatik,
Biometrie und
Epidemiologie e.V.

GMDS e.V. > Industriestraße 154 > D-50996 Köln

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

Per E-Mail: methoden@iqwig.de

GMDS Geschäftsstelle

Industriestraße 154
D-50996 Köln

Telefon: +49 (02236) 33 19 958
Telefax: +49 (02236) 33 19 959

E-Mail: info@gm^{ds}.de
Internet: www.gm^{ds}.de

Geschäftsführung

Beatrix Behrendt

Präsident und Vize-Präsidenten

Prof. Dr. Ulrich Mansmann
(München), Präsident

Prof. Dr. Andreas Stang; MPH
(Essen), 1. Vizepräsident

Prof. Dr. Paul Schmücker
(Mannheim), 2. Vizepräsident

IBS-DR Geschäftsstelle

Heike Krubert
c/o Institut für Biometrie, Epidemiologie und
Informationsverarbeitung
Tierärztliche Hochschule
Hannover
Bünteweg 2
D-30559 Hannover

Telefon: +49 (0) 511 953 79 51
Telefax: +49 (0) 511 953 79 74

E-Mail: biometrische-gesellschaft@tiho-hannover.de

IBS-DR - Präsident und Vizepräsident

Prof. Dr. Tim Friede
(Göttingen), Präsident

Prof. Dr. Andreas Faldum
(Münster), Vizepräsident

31. Januar 2017

Sehr geehrter Herr Professor Windeler,

wir bedanken uns für die Möglichkeit zur Stellungnahme des IQWiG-Papiers
Allgemeine Methoden in der Version 5.0.

Die Deutsche Region der Internationalen Biometrischen Gesellschaft (IBS-DR)
und die Deutsche Gesellschaft für Medizinische Informatik, Biometrie und
Epidemiologie (GMDS) e.V. haben die Stellungnahme im Rahmen der
gemeinsamen Präsidiumskommission IQWiG angefertigt.

Mit freundlichen Grüßen



Prof. Dr. Ulrich Mansmann
- Präsident der GMDS -



Prof. Dr. Tim Friede
- Präsident der IBS-DR -

Anlagen: - Formblatt zur Stellungnahme
- Stellungnahme GMDS / IBS-DR

Stellungnahme zum Entwurf der Version 5.0 des IQWiG-Papiers
Allgemeine Methoden
durch die gemeinsame Präsidiumskommission IQWiG der GMDS und IBS-DR

Das Institut für Qualität und Wirtschaftlichkeit in der Medizin hat am 7.12.2016 einen ersten Entwurf für die neue Version des „Allgemeinen Methodenpapier“ vorgelegt. Darin werden die gesetzlichen und wissenschaftlichen Grundlagen der Arbeit am IQWiG erläutert. In der Version 5.0 wurden im Vergleich zur Version 4.2 der „Allgemeinen Methoden“ vom 24.4.2015 neben redaktionellen Modifikationen auch inhaltliche Änderungen durchgeführt. Grundlegende Überarbeitungen erfolgten in den Kapiteln zu *Leitliniensynopsen und Versorgungsanalysen* (Kapitel 5), *Gesundheitsinformation* (Kapitel 6) und *Informationsbeschaffung* (Kapitel 8). Auch das Kapitel 9 zur *Informationsbewertung* wurde in vielen Abschnitten aktualisiert.

Aus biometrischer Sicht hat sich das Verfahrensspektrum verbreitert. Bayesianische Verfahren oder penalisierten Regressionsverfahren werden nunmehr vom Institut als valide Methodik akzeptiert. Es wurden Verfahren zur Metaanalyse weniger Studien und zum Einsatz von Interaktionstests spezifiziert. Hier sei die Methode von Knapp-Hartung für Meta-Analysen mit zufälligen Effekten erwähnt, die nun regelhaft angewendet werden soll. Wir begrüßen diesen Schritt. Allerdings bleibt die Vorgehensweise in der recht häufigen Situation von Meta-Analysen mit weniger als fünf Studien offen. Zudem ist nicht klar, nach welchen Kriterien diese neuen Methoden selektiert wurden und ob es hierzu systematische Literaturrecherchen und ein Methodenvergleich gab.

Resamplingverfahren sind in das Methodenspektrum noch nicht aufgenommen. Im Abschnitt 9.3.4 werden Verfahren zur Quantifizierung von Bias durch fehlende Designqualität beschrieben (Wood et al, 2008), aber nicht auf Methoden verwiesen, die Metaanalysen (und somit Nutzenwerte) hinsichtlich dieser Fehler korrigieren können (Welton et al, 2009). Es wäre wünschenswert, dem Papier einen weiteren Abschnitt zur graphischen Präsentation von Ergebnissen beizufügen. Visualisierung der bewerteten Informationen wäre für die Kommunikation der Ergebnisse eines Nutzenbewertungsprozesses von großer Bedeutung.

Grundsätzlich begrüßt die Kommission die Änderung des Signifikanzniveaus für die Interaktionstests von 0.2 auf 0.05 (Kapitel 9.3.10).

Neben diesen vertieften methodischen Inhalten fehlen dem Methodenpapier Diskussionen zu einigen grundlegenden Aspekten der Nutzenbewertung, die für die weitere Entwicklung derer Methodik wichtig sind:

1. Betrachtungen zum für den Nutzenbewertungsprozess verwendeten Datenkörper. Dieser wird durch verschiedenste Selektionen (auch aufgrund von Merkmalen nach Randomisierung) aus den Daten randomisierter, kontrollierter Studien abgeleitet. Ob diese Transformationen zu Daten führen, die als gleichwertig zu RCT-Daten betrachtet werden können, ist eine wichtige Frage. Das Methodenpapier nimmt die Äquivalenz grundsätzlich an. Die Problematik sehen wir wie folgt: Mit der finalen Indikationsformulierung durch die Zulassungsbehörden beginnt für ein pharmazeutisches Unternehmen in Deutschland die Nutzenbewertung zu einem innovativen medizinischen Produkt. Der Nutzenbewertungsprozess lässt dabei den zur Zulassung einer Indikation führenden Prozess in der Regel außer

Betracht. Die finale Indikationsformulierung im Zulassungsprozess ist komplex mit möglichen Abweichungen zu den in den Studien betrachteten Populationen und Verabreichungsformen. Die im Zulassungsprozess durchgeführten Subgruppenanalysen führen zu einem neuen Blick auf die Studiendaten (etwa durch Entscheidungen basierend auf post-hoc Analysen, durch Abweichung von Population der Indikation und Studienpopulationen) noch bevor der G-BA weitere Spezifizierungen hinsichtlich der für die Nutzenbewertung relevanten Populationen und zweckmäßige Vergleichstherapie vornimmt. Der Zulassungstext ist von großer Bedeutung, da nur die darin enthaltene Information für die Nutzenbewertung berücksichtigt wird. Durch die Abhängigkeit der Population für die Nutzenbewertung von post-hoc ausgewählten Variablen wird Potential für Verzerrungen geschaffen. Durch die Reduzierung der Studienpopulation auf die Zielpopulation folgt weiterhin ein Powerverlust der statistischen Aussagen zu positiven und (fast wichtiger noch) zu unerwünschten Effekten von Arzneimitteln.

2. Das Methodenpapier betont weiterhin die Relevanz der endpunktbezogenen Nutzenbewertung. Die Zusammenschau der endpunktbezogenen Bewertungen führt zum Ergebnis der Nutzenbewertung. Methoden für eine objektive Zusammenschau werden im Methodenpapier nicht diskutiert. Wichtig wären aber Prozesse, die in Zusammenarbeit mit Medizinischen Fachgesellschaften und Patientenorganisationen eine Priorisierung von Endpunkten ermöglichen und daraus abgeleitet die bewertete Zusammenschau.
3. Beratungsmöglichkeit für pharmazeutische Unternehmen: Ein pharmazeutischer Unternehmer kann schon jetzt nach der Verfahrensordnung des G-BA eine Beratung zur Nutzenbewertung bereits vor Beginn der Zulassungsstudien der Phase III auch unter Beteiligung der deutschen Zulassungsbehörden anfordern. Welche Rolle das IQWiG dabei spielen kann, wird nur kurz und unverbindlich im Abschnitt 2.2.7 erwähnt.
4. Es gibt Beispiele, wo aufgrund der Vorgaben durch den G-BA (Population, zweckmäßige Vergleichstherapie) und der methodischen Arbeitsweise des IQWiG der Nutzenbewertungsprozess nicht durchgeführt werden konnte. Es wäre deshalb von Interesse, die Vorgehensweisen beider Institutionen passgenauer aufeinander abzustimmen. Eine solche Intention sollte in einem allgemeinen Methodenpapier angesprochen werden. Hierzu ein extremes Beispiel: In der Nutzenbewertung zu Belimumab hat die Forderung des G-BA nach einer „optimierten“ Standardtherapie statt (wie in der Indikationsformulierung) einer Standardtherapie zu erheblichen Divergenzen mit dem IQWiG geführt. Zum einen ist das IQWiG davon ausgegangen, dass in den relevanten klinischen Studien BLISS-52 und BLISS-76 keine optimierte Standardtherapie angewandt wurde. Damit waren diese Studien aus der Sicht des IQWiG für die Zusatznutzenbewertung irrelevant mit dem Ergebnis, dass dann auch kein Zusatznutzen belegt war. Währenddessen kam der G-BA zum Schluss, dass in diesen Studien doch eine optimierte Standardtherapie angewandt worden war.
5. Angesichts der zunehmenden Anzahl klinischer Prüfungen mit erlaubtem Behandlungswechsel wären Empfehlungen zu möglichen Analysemethoden, oder ggf. Empfehlungen gegen bestimmte Verfahren (z.B. naive Verfahren), hilfreich. Auf diesen Punkt wird unter Abschnitt 9.3.7 eingegangen.
6. Die Abschnitte zur gesundheitsökonomischen Evaluation und Modellierung sind weiterhin gut beschrieben und haben an Klarheit gewonnen. Insbesondere ist zu begrüßen, dass konkrete Anweisungen zur Verwendung von Gesamtnutzenmaßen wie den qualitätsadjustierten Lebensjahren gegeben werden (z.B. Verwendung deutscher Tarife bei Lebensqualitätsindizes zur QALY-Berechnung). Unklar bleibt im Gegensatz dazu das Vorgehen

zur Gewichtung multipler Effizienzkurven beim Vorliegen multipler Nutzen- und Schadensendpunkte und die entsprechende theoretische Grundlage, wenn kein Gesamtnutzenmaß eingesetzt wird.

7. Die im Rahmen der gesundheitsökonomischen Evaluation eingesetzten Nutzwerte stammen in der Regel aus spezifischen Nutzwert-Surveys (Lebensqualitätsstudien) zu kurz- und langfristigen Konsequenzen der Krankheit und Intervention, nicht aus klinischen Prüfungen. Dies sollte besser verdeutlicht werden und diesen Studien mit patienten-relevanten Outcomes sollte ein entsprechender Stellenwert zugeschrieben werden.

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

Abschnitt 1.3

In diesem Abschnitt wird der Begriff Gesundheitsökonomie „im engen Sinn“ erklärt. Hier sollte stattdessen der Begriff „gesundheitsökonomische Evaluation“ verwendet werden, der zutreffender ist, insbesondere wenn es um die Kosten-Nutzen-Bewertung (KNB) bzw. die Empfehlung eines Höchstbetrags geht.

Abschnitt 1.3.1

In diesem Abschnitt wird beschrieben, dass der „Höchstbetrag die angemessenen Kosten für den Zusatznutzen eines neuen Arzneimittels im Vergleich zu anderen Arzneimitteln und Behandlungsformen in einer Indikation darstellen sollte“. Der nachfolgende Satz lautet „Bedingung für die Beauftragung einer KNB musste also ein Nachweis des Zusatznutzens für ein neues Arzneimittel sein, der in einer Nutzenbewertung des IQWiG vorliegen musste.“ Im zweiten Satz sollte das „also“ gestrichen werden, da dies impliziert, dass eine KNB notwendigerweise nur bei positivem Zusatznutzen durchgeführt werden kann. Dies ist nicht der Fall, denn eine KNB kann theoretisch auch bei negativem inkrementellem Nutzen und geringeren Kosten sinnvoll sein (z.B. sehr viel günstigeres Arzneimittel mit gleichem oder nur geringfügig geringerem Nutzen als der Komparator). Es sollte an dieser Stelle im Dokument explizit dargestellt werden, warum die IQWiG-Methoden in diesen Fällen keine KNB vorsieht, z.B. durch welchen Teil der gesetzlichen Rahmenbedingungen dies ausgeschlossen ist.

Da es bei der KNB um inkrementelle Kosten und inkrementellen Nutzen einer Intervention im Vergleich zu einem Komparator geht, sollte der Satz „Die Darstellung einer begründbaren Relation zwischen den Kosten und dem Nutzen muss also aus der KNB hervorgehen“ geändert werden in „Die Darstellung einer begründbaren Relation zwischen den Zusatzkosten und dem Zusatznutzen muss also aus der KNB hervorgehen.“

„Auch wenn nirgendwo im Gesetz oder in den nachgeordneten Bestimmungen auf die Frage eingegangen wird, wie die Gesundheitsökonomie (siehe Abschnitt 1.3) zu verstehen ist, ergibt es sich aus der Anwendung, dass es sich um KNB und damit um die Gesundheitsökonomie im engen Sinne handelt.“ Sollte geändert werden in: „[...] ergibt es sich aus der Anwendung, dass es sich um KNB und damit um die gesundheitsökonomische Evaluation handelt.“

Abschnitt 1.3.2

Im Dokument wird darauf hingewiesen, dass es zum Vorgehen bei der KNB internationale Standards gibt. Hier sollte explizit unterschieden werden zwischen drei verschiedenen Typen von Standards bzw. methodischen Guidelines:

- a) Methodische Standards zum Vorgehen und der Durchführung von gesundheitsökonomischen Evaluationen
- b) Methodische Standards zum Vorgehen und der Durchführung von (gesundheitsökonomischen) Modellierungen und
- c) Standards für die Berichterstattung (Reporting) aus gesundheitsökonomischen Evaluationen.

Das im Dokument zitierte CHEERS Statement [Referenz 336] gehört zur Kategorie c) Reporting.

Die folgenden internationalen Empfehlungen gehören zu den Standards im Bereich b) Modellierung und sollten entweder an dieser Stelle oder unter Abschnitt 4.2 erwähnt werden.

- Caro JJ, Briggs AH, Siebert U, Kuntz KM, Force I-SMGRPT. Modeling good research practices--overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. *Med Decis Making* 2012;**32**(5):667-77. *Value Health* 2012;**15**(6):796-803.
- Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M, Force I-SMGRPT. Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2. *Med Decis Making* 2012;**32**(5):678-89. *Value Health* 2012;**15**(6):804-11.
- Siebert U, Alagoz O, Bayoumi AM, Jahn B, Owens DK, Cohen DJ, Kuntz KM. State-transition modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-3. *Med Decis Making* 2012;**32**(5):690-700. *Value Health* 2012;**15**(6):812-20.
- Karnon J, Stahl J, Brennan A, Caro JJ, Mar J, Moller J. Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-4. *Med Decis Making* 2012;**32**(5):701-11. *Value Health* 2012;**15**(6):821-7.
- Pitman R, Fisman D, Zaric GS, Postma M, Kretzschmar M, Edmunds J, Brisson M, Force I-SMGRPT. Dynamic transmission modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-5. *Med Decis Making* 2012;**32**(5):712-21. *Value Health* 2012;**15**(6):828-34.
- Briggs AH, Weinstein MC, Fenwick EA, Karnon J, Sculpher MJ, Paltiel AD, Force I-SMGRPT. Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-6. *Med Decis Making* 2012;**32**(5):722-32. *Value Health* 2012;**15**(6):835-42.
- Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB, Force I-SMGRPT. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. *Med Decis Making* 2012;**32**(5):733-43. *Value Health* 2012;**15**(6):843-50.

Diese ISPOR-SMDM Standards werden inzwischen von vielen internationalen wissenschaftlichen Fachzeitschriften vorausgesetzt.

Im Dokument werden die Schritte für die methodischen Vorgaben aufgezählt. Hier wären zu ergänzen: Definition der Zielpopulation, Definition der Zielgrößen, Indexjahr der Kostenerhebung.

Abschnitt 1.3.3

Die Aussage, dass die ganz große Mehrheit der Länder, in denen KNB zur Entscheidungsfindung beitragen, keinen festen bzw. keinen expliziten Schwellenwert haben, sollte mit Quellen unterlegt werden, (z.B. Schwarzer et al. Systematic overview of cost-effectiveness thresholds in ten countries across four continents. J Comp Eff Res 2015;4(5):485-504).

Abschnitt 3.1.4

In Abschnitt 3.1.4 wird darauf eingegangen, in welchen Fällen ein Beleg für einen Zusatznutzen aus lediglich einer Studie abgeleitet werden kann. Insbesondere wird dort neu formuliert, dass dafür mindestens 1000 Patienten pro Studienarm eingeschlossen sein müssen. Bei den Überlegungen zur Frage der Aussagekraft einer einzigen Studie sollten Entwicklungen im Zulassungsprozess im Auge behalten werden. Hier ist die zur Zeit als Draft vorliegende ICH E17 „*General principals for planning and design of multi-regional clinical trials*“ relevant. In dieser Guideline wird die Durchführung von Multi-regional clinical trials (MRCT) empfohlen, wenn eine globale Zulassung in mehreren Regionen angestrebt wird. Damit soll ausdrücklich die Reduzierung der Anzahl separater Studien erreicht und unnötige Duplizierung von Studien vermieden werden. Die Validierung der Ergebnisse durch Replikation soll dabei ersetzt werden durch die Untersuchung der internen Konsistenz der Resultate innerhalb der Studie. Die Beantragung der Zulassung auf Basis eines MRCT wird ausdrücklich befürwortet, so dass in Zukunft damit zu rechnen ist, dass die Situation der Zulassung auf Basis einer einzigen Studie vermehrt auftreten wird. Dies sollte bei der Weiterentwicklung des Methodenpapiers berücksichtigt werden.

Die Patientenzahl von 1000 Patienten pro Behandlungsarm wird im Methodenpapier nicht motiviert. In aller Regel sind Studien schon aus ethischen Gründen nicht größer als absolut notwendig, um valide Aussagen zum primären Endpunkt treffen zu können. Liegen sehr große Studien vor, so liegt das daran, dass der Effekt sehr klein ist und sehr viele Patienten benötigt werden. Unter Berücksichtigung von E17 ist also in Zukunft davon auszugehen, dass die Kategorie „Beleg“ nur noch in Ausnahmefällen überhaupt erreicht werden kann. Die Grenze von 1000 Patienten pro Gruppe plausibilisiert sich daraus, dass ausreichend viele Daten zur Sicherheit vorliegen sollen. Dazu bedarf es aber einer methodischen Herleitung. Zudem ist nicht klar, warum von den Vorgaben der EMA Guideline abgewichen wird, die unter anderem Kriterien wie interne und externe Validität vorsieht, welche nach EMA-Ansatz gegeben ist, wenn die Studie für die Zulassung als pivotale Studie akzeptiert wird.

Abschnitt 4.1

„Für jede KNB müssen Entscheidungen zur Perspektive ...“. Hier wären zu ergänzen die Definition der Zielpopulation und die Definition der Zielgrößen.

Abschnitt 4.1.3

In diesem Abschnitt wird zum ersten Mal auf die Methode der Modellierung hingewiesen. Hier sollte der Begriff kurz eingeführt werden und auch auf langfristige Lebensqualitätsgewinne eingegangen werden, z.B.: „Bei chronischen Erkrankungen sollte möglichst ein längerer Zeithorizont gewählt werden, insbesondere wenn Lebenszeitgewinne oder langfristige Lebensqualitätsgewinne erwartet werden [76,185,444,647]. Für gesundheitsökonomische Evaluationen über längere Zeithorizonte

werden häufig entscheidungsanalytische Modelle eingesetzt [Referenz: Siebert U. When should decision-analytic modeling be used in the economic evaluation of health care? The European Journal of Health Economics 2003;4(3):143-50]. Kosten und Nutzen sollen immer über denselben Zeithorizont modelliert werden.“

Auch auf der Kostenseite werden häufig Einsparungen durch vermiedene Behandlungen erst bei einem längeren Zeithorizont realisiert (s. KNB Antidepressiva).

Abschnitt 4.1.7

Im Dokument wird darauf eingegangen, dass „die statistischen Methoden und die Kontrolle von Störgrößen transparent beschrieben und begründet werden.“ Da in Sekundärdaten sehr häufig die Situation der zeitabhängigen Störgrößen vorliegt, aber immer noch sehr häufig ausschließlich für Baseline-Störgrößen adjustiert wird, sollte hier explizit auf die zeitabhängigen Störgrößen eingegangen werden, die eine wichtige Voraussetzung für die kausale Interpretation der gesundheitsökonomischen Ergebnisse sind, z.B.: „die statistischen Methoden und die Kontrolle von zeitunabhängigen und zeitabhängigen Störgrößen transparent beschrieben und begründet werden.“

Abschnitt 4.1.9

Zur Verdeutlichung des inkrementellen Kosteneffektivitätsverhältnisses könnte der bestehende Satz spezifiziert werden in: „Das letzte Segment der Effizienzgrenze kann unter der Annahme, dass es den Kehrwert der gegenwärtigen Zahlungsbereitschaft für eine zusätzliche Nutzeneinheit darstellt, linear extrapoliert werden.

Bei der Einführung der Effizienzgrenze sollte nochmals festgehalten werden, ob neue Interventionen mit keinem oder negativem Zusatznutzen (z.B. bei erheblich niedrigeren Kosten) mit der Effizienzgrenze analysiert werden oder nicht.

Im Dokument wird angeführt, dass bei Vorliegen mehrerer Effizienzgrenzen der Entscheidungsträger eine Gewichtung vornehmen kann. An dieser Stelle wäre es hilfreich, darauf hinzuweisen, dass in diesem Falle eine (zusätzliche) Sensitivitätsanalyse mittels Berechnung der qualitätsadjustierten Lebensjahre hilfreich sein kann, um eine grobe Einschätzung des Gesamtnutzens zu ermöglichen. Dies kann – trotz der hierfür notwendigen Annahmen – die Transparenz und Nachvollziehbarkeit des Berichtes erhöhen und ggf. eine auf nicht expliziten Annahmen beruhende ad-hoc Gewichtung der multiplen Endpunkte durch den Entscheidungsträger unterstützen. Alternativ sollte ein Vorgehen zur Gewichtung vorgeschlagen werden, da das Dokument sonst an der entscheidenden Stelle der Methodenbeschreibung seine Bedeutung als Methodenhandbuch verliert.

Abschnitt 4.2.1

Im Dokument werden diverse Gründe für eine Zusammenführung von Daten beschrieben. Die Liste sollte ergänzt werden um einige wichtige und in der Praxis häufig vorkommende Bereiche, wie z.B. „linked evidence“ bei diagnostischen oder Screening-Verfahren, der Abwägung von Nutzen und Schaden in einer gesundheitsökonomischen Evaluation und dem Transfer von Daten aus internationalen Studien auf den deutschen Kontext. Eine systematische Aufstellung findet sich in: Siebert U. When should decision-analytic modeling be used in the economic evaluation of health care? The European Journal of Health Economics 2003;4(3):143-50.

Abschnitt 4.3.3

Es wird begrüßt, dass verschiedene systematische und nachvollziehbare Maße des Gesamtnutzens (QALY, DALY, HYE) vorgeschlagen werden und im Unterkapitel A) auf die QALYs genauer eingegangen wird. Damit wird das Dokument dem Charakter eines Methodenhandbuchs gerecht. Ggf. sollte ergänzt werden, dass alle drei Methoden sowohl Schwächen als auch Stärken aufweisen in Bezug auf Annahmen und Praktikabilität. So macht z.B. das HYE weniger Annahmen als das QALY-Konzept, ist dafür aber praktisch wegen der vielen Freiheitsgrade im Allgemeinen in der Routine nicht umsetzbar.

Im Gegensatz dazu fehlt wie in den früheren Versionen des Dokuments eine Spezifizierung, wie Endpunkte zu gewichten sind, völlig. Eine wie auch immer geartete Gewichtung multipler Endpunkte durch den Entscheidungsträger außerhalb der Expected Utility Theorie würde voraussichtlich in den meisten Fällen zu deutlich mehr Annahmen führen als bei den oben genannten Gesamtnutzenmaßen und wäre zudem weniger explizit und nachvollziehbar. Dies widerspräche dem Prinzip der transparenten und expliziten evidenzbasierten Medizin und erinnert an „expertenbasierte“ ad-hoc Entscheidungen aus dem Zeitalter vor der EbM. In diesem Kontext sollte bei Vorliegen verschiedener Nutzen- und Schadensendpunkte, die abzuwägen sind, entlang der gewohnt guten IQWiG-Praxis zu Sensitivitätsanalysen, Transparenz und Nachvollziehbarkeit zumindest eine Sensitivitätsanalyse mit einem der obengenannten Gesamtnutzenmaße empfohlen werden. Die Ergebnisse dieser Sensitivitätsanalyse kann der Entscheidungsträger dann berücksichtigen oder nicht.

Im Unterkapitel A) wird beschrieben, dass die einfließenden Werte zu den Gesundheitszuständen bei den Betroffenen zu erheben sind, die aktuell oder in der Vergangenheit diese Gesundheitszustände erlebt haben. Dies widerspricht der gemachten Aussage, dass die Erhebung bei Betroffenen vs. der Allgemeinbevölkerung ein Werturteil darstellt, welches nicht a priori gefällt werden kann. Zudem hat sich gezeigt, dass Coping-Effekte die Bewertung bei Betroffenen stark verzerren können.

Im Unterkapitel A) wird auch beschrieben, dass die in die Bewertung einfließenden Nutzwerte bei den an klinischen Studien Teilnehmenden erhoben worden sein sollten. Da Nutzwerte vorwiegend in Nutzwert-Surveys (Lebensqualitätsstudien) zu kurz- und langfristigen Konsequenzen der Krankheit und Intervention erhoben werden sollte die Formulierung erweitert werden auf „Die Daten sollten bei an klinischen oder Lebensqualitäts-Studien Teilnehmenden erhoben worden sein“.

Unterkapitel B) beschreibt Methoden zur Erhebung von Präferenzen zur Erstellung eines Maßes des Gesamtnutzens. Die Beschreibung der Methoden der multikriteriellen Entscheidungsfindung und der verschiedenen Verfahren (AHP, CA) wird begrüßt. Ggf. sollte darauf hingewiesen werden, dass sich diese Verfahren insbesondere dann sehr gut eignen, wenn die klinischen Nutzen- und Schadens-Endpunkte innerhalb eines relativ kurzen Zeithorizonts anfallen, keine Langzeitkonsequenzen zu modellieren sind, und die Kurzzeit-Outcomes direkt in die Bewertung/Befragung bei der AHP oder CA eingehen.

Abschnitt 4.4

Das gesamte Kapitel 4.4 ist sehr klar strukturiert und anwendungsorientiert verfasst. Es erfüllt den Zweck einer anwendungsorientierten Anleitung sehr gut. Bei verschiedenen Optionen werden zudem nachvollziehbare Vorgaben gemacht, welche Option in die Basisfallanalyse eingehen soll und welche in eine Sensitivitätsanalyse.

Bereits in Abschnitt 4.1.9 (Interpretation der Ergebnisse) wird das Konzept der Effizienzgrenze eingeführt und gut erklärt. Interpretiert man den Text so, dass bei der Erstellung der Effizienzgrenze das Konzept der erweiterten Dominanz anzuwenden ist, so verwirrt der theoretische Diskurs auf S. 103-104 und sollte ggf. in einen Appendix verschoben werden, da er keine Handlungsrelevanz für die Durchführenden der gesundheitsökonomischen Evaluation hat.

Abschnitt 4.6.4

Zu den beiden beschriebenen Sonderkonstellationen sollte ergänzt werden, dass in beiden Fällen die aktuell untersuchte Intervention als kosteneffektiv bewertet werden kann. Ein Problem ergibt sich erst bei einer nachfolgenden Bewertung. Dies betrifft im Übrigen auch den Fall einer Intervention, die sehr wenig Zusatzkosten bei einem sehr großen Nutzengewinn erzeugt (z.B. mit einem inkrementellen Kosteneffektivitätsverhältnis von 100 EUR pro vermiedenen Todesfall. Eine konsequente Anwendung der Effizienzkurve ist hier sinnlos, da dies bedeuten würde, dass ein neues Arzneimittel, welches für 200 EUR einen Todesfall vermeiden kann, den Preis reduzieren müsste. In solchen Fällen wird vermutlich das inkrementelle Kosteneffektivitätsverhältnis losgelöst von der extrapolierten Steigung beurteilt werden. Dieser Punkt sollte noch im Dokument erläutert werden, da er die Grenzen der Effizienzkurvenmethode zeigt.

Warum die Ausgaben-Einfluss-Analyse hier die Kosteneffektivitätsanalyse ersetzen soll, wird nicht deutlich gemacht und wäre auch nicht im Sinne der Effizienzoptimierung. Deshalb wird vorgeschlagen, den Satz „Hier kann die Ausgaben-Einfluss-Analyse weitere Daten liefern, indem die Auswirkungen auf das Budget dargestellt werden“ zu streichen.

Abschnitt 4.9.4

Die Interquartilsregion (IQR) wird als Grundlage für die Darstellung der Streuung angegeben. Diese Wahl scheint ad-hoc zu sein, ohne nutzentheoretische Fundierung. Im Dokument wird darauf hingewiesen, dass grundsätzlich auch die Angabe von anderen Regionen mit anderen Maßen sinnvoll sein kann. Damit bleibt das Dokument an diesem Punkt vage.

Abschnitt 6.1

In diesem Abschnitt sollte ein Verweis auf Abschnitt 2.1.10 erfolgen, damit deutlich gemacht wird, dass HTA-Berichte vor allem von externen Experten erstellt werden, im Unterschied zu den übrigen Berichten.

Abschnitt 6.3.1

In diesem Abschnitt wird darauf hingewiesen, dass medizinische Untersuchungs- und Behandlungsmethoden bewertet werden.

Im Gesetz steht: „Versicherte und sonstige interessierte Einzelpersonen können beim Institut Bewertungen nach § 139a Absatz 3 Nummer 1 und 2 zu medizinischen Verfahren und Technologien vorschlagen. Das Institut soll die für die Versorgung von Patientinnen und Patienten besonders bedeutsamen Vorschläge auswählen und bearbeiten.“

Damit sind Primärpräventionsmaßnahmen wie Lebensstiländerungen oder Supplementierung von Lebensmitteln implizit von der Evaluation ausgeschlossen. Sollte diese Interpretation korrekt sein,

betrachten wir die gesetzliche Grundlage als verbesserungswürdig. Informationen zu Primärpräventionsthemen haben für die Bürgerinnen und Bürger ggf. eine höhere Relevanz als die Untersuchungs- und Behandlungsmethoden, deren Verwendung meist durch Ärzte initiiert wird, während bei konsumentenassoziierter Primärprävention die Bürgerinnen und Bürger selbst entscheiden müssen und entsprechenden Informationsbedarf haben.

Die Gewichtung der Themenvorschläge ist im Dokument nicht transparent beschrieben. Es sollte transparent dargestellt werden, wie die Gewichtung vorgenommen wird.

Aufgrund der anhaltenden Diskussion zur Kosteneffektivität von Orphan Drugs sollte beim Kriterium „Wie groß ist die Anzahl der Betroffenen?“ geklärt werden, ob dies zur Folge haben kann, dass Orphan Disease Fragestellungen ausgeschlossen sind.

Abschnitt 6.4

S. 134: Die genannten Templates sollten dargestellt werden (z.B. im Anhang).

Im Dokument wird darauf hingewiesen, dass die Abnahme des Berichtsprotokolls und des vorläufigen Basisberichts nach einer Konformitätsprüfung durch das Institut erfolgt. Hier wäre es hilfreich anzugeben, was bei dieser Prüfung abgeglichen wird.

Abschnitt 6.5.2

Bei der Aufzählung der anzuwendenden Kapitel der Gesundheitsökonomie ist Kapitel 4.6 (Bestimmung der Effizienzgrenze und der Ableitung eines Schwellenwertes) ausgeschlossen, da es in den HTA-Berichten nicht um die Festlegung eines Höchstpreises geht. Damit ist aber unklar, wie die Kosteneffektivität beurteilt werden soll, da in Deutschland kein expliziter Schwellenwert existiert. An dieser Stelle wäre eine Erläuterung hilfreich, wie sich die Autoren dem Problem annähern könnten (z. B. Vergleich mit inkrementellen Kosteneffektivitätsverhältnissen für andere klar akzeptierten bereits erstatteten Leistungen, WHO-Kriterien etc.).

Abschnitt 6.5.6

Bei HTA-Berichten, die explizit für Bürgerinnen und Bürger bzw. Patientinnen und Patienten sein sollen, sollten auch Auswirkungen organisatorischer Rahmenbedingungen auf diese Personen berücksichtigt werden, nicht nur auf Professionelle.

Abschnitt 9.3.7

In Abschnitt 9.3.7 werden multifaktorielle Analysen beschrieben, u.a. zur Kontrolle von Confounding. An dieser (oder anderer geeigneter) Stelle wäre zu ergänzen:

- a) Umgang mit zeitabhängigem Confounding, z.B. Einsatz von Kausalverfahren zur Kontrolle von zeitabhängigem Confounding bei Beobachtungsstudien (inverse probability weighting, g-estimation, parametric g-formula).
- b) Umgang mit Behandlungswechsel, Noncompliance und informativer Zensurierung auch bei randomisierten Studien, wenn die ITT-Analyse keine Antwort auf die Forschungsfrage (kausaler Kontrast) gibt.

Die jüngste Literatur zu Empfehlungen zur Analyse von Behandlungswechseln sollte berücksichtigt werden:

- Latimer NR et al. Treatment Switching: statistical and decision-making challenges and approaches. *J Technol Assess Health Care* 2016;32(3):160–6.
- Henshall C et al. Treatment switching in cancer trials: issues and proposals. *J Technol Assess Health Care* 2016;32(3):167-74.

Abschnitt 9.3.8

In Abschnitt 9.3.10 wird das Signifikanzniveau des Interaktionstests für Subgruppenanalysen diskutiert und nur die Grenze 0.05 beibehalten, das ist sehr zu begrüßen. Auf der anderen Seite wurde in Abschnitt 9.3.8 bei den Meta-Analysen allerdings jedwede Grenze für den Test auf Heterogenität gelöscht. Bisher wurde auf ein Niveau von 0,1 bis 0,2 verwiesen und in der Praxis vom IQWiG die Grenze von 0,2 verwendet. Die Frage ist nun, ob die Grenze von 0,05 auch für den Test auf Heterogenität Verwendung finden kann/soll.

Abschnitt 9.3.9

Im Abschnitt 9.3.9 werden die indirekten Vergleiche diskutiert. Dabei werden die sogenannten historischen Vergleiche, d.h. nicht-adjustierte indirekte Vergleiche ohne Brückenkomparator, grundsätzlich als nicht adäquat abgelehnt. Diese sind allerdings in einer Reihe von Indikationen die einzige Möglichkeit eines Vergleichs, da sich RCT mit einer Vergleichsgruppe verbieten. Daher sollte im Einzelfall geprüft werden, ob sich der historische Vergleich nicht doch qualifiziert. Grundsätzlich hält dies ja auch das IQWiG für möglich, wenn der Effekt dramatisch ist, wie im Methodenpapier an anderer Stelle erwähnt.

Abschnitt 9.3.10

(Seite 194 unten, neuer Abschnitt) „Einerseits erfordern die oben aufgeführten Aspekte die Bewertung der Verlässlichkeit von Subgruppenanalysen; Sun et al. [627] identifizieren hierzu Kriterien. So sind beispielsweise Subgruppenanalysen zu Merkmalen, die nicht vor der Randomisierung, sondern im Studienverlauf erhoben wurden (beispielsweise Patienten mit vs. solche ohne Herzinfarkt unter der untersuchten Therapie), grundsätzlich nicht verlässlich. Andererseits können trotz dieser Einschränkungen Subgruppenanalysen für manche Fragestellungen die auf absehbare Zeit beste wissenschaftliche Evidenz für die Bewertung von Effekten in ebendiesen Subgruppen darstellen [237], da beispielsweise ethische Überlegungen gegen die Verifizierung beobachteter Effekte in weiteren Studien sprechen können.“

Das Beispiel der „Subgruppenanalysen zu Merkmalen, die nicht vor der Randomisierung, sondern im Studienverlauf erhoben wurden (beispielsweise Patienten mit vs. solche ohne Herzinfarkt unter der untersuchten Therapie)“ ist hier nicht glücklich gewählt. Diese Art von Analysen sind nicht nur „grundsätzlich nicht verlässlich“, sondern können zu groben Verzerrungen führen. Der resultierende Bias ist in der Literatur unter verschiedenen Bezeichnungen gut beschrieben (conditioning on the future, immortal-time bias, time-dependent bias, etc.). Deshalb sollte diese Art der Subgruppenanalyse nicht als Beispiel für eine Analyse dienen, die „trotz dieser Einschränkungen ...für manche Fragestellungen die auf absehbare Zeit beste wissenschaftliche Evidenz für die Bewertung von Effekten in ebendiesen Subgruppen“ darstellt.

Zudem qualifizieren sich diese Subgruppen nach Kriterien, die im Studienverlauf erhoben wurden, nicht als Effektmodifikatoren, da vor Beginn der Behandlung nicht nach diesen Merkmalen stratifiziert werden kann, bzw. der Arzt nicht aufgrund dieser Kriterien eine Therapieentscheidung fällen kann. Das aber ist ja der eigentliche Sinn der Subgruppierung im AMNOG-Prozess.

Abschnitt 9.3.10

(Seite 195 Mitte, neuer Abschnitt) „Bei der Interpretation von Subgruppenanalysen ist zu beachten, dass ein statistisch signifikanter Effekt in einer Subgruppe, aber kein Effekt oder ein entgegengesetzter Effekt in einer anderen Subgruppe, für sich allein (inferenzstatistisch) nicht als das Vorliegen von zwischen den Subgruppen unterschiedlichen Effekten interpretiert werden kann. Stattdessen muss zunächst der statistische Nachweis unterschiedlicher Effekte zwischen verschiedenen Subgruppen anhand eines adäquaten Homogenitäts- beziehungsweise Interaktionstests vorgenommen werden. Damit überhaupt eine gewisse Wahrscheinlichkeit für einen solchen Nachweis vorliegt, werden Subgruppenanalysen bei binären Daten regelhaft nur dann durchgeführt, falls mindestens 10 Ereignisse in einer der Subgruppen aufgetreten sind. Mit einem Ereignis ist hier gemeint, dass dieses im Verlauf einer Beobachtung auftritt und zu Baseline noch nicht eingetreten ist (z. B. Erlangen einer Virusfreiheit bei infizierten Personen).“

Wenn man hier für einen binären Endpunkt die Zahl 10 Ereignisse als untere Grenze für die Durchführung von Interaktionstests nennt, sollte man aus Symmetriegründen auch die Zahl 10 Patienten ohne Ereignis als untere Grenze nennen. Liegen im Extremfall insgesamt 11 Ereignisse vor, 10 in Subgruppe 1 und eines in Subgruppe 2 sowie kein Ereignis in den weiteren Subgruppen, so wird ein Interaktionstest schwerfallen und nicht interpretierbar sein. Eine genauere Beschreibung für das Mindestmaß scheint hier angezeigt.

Abschnitt 9.3.10

(Seite 195 unten, neuer Abschnitt) „Paare, die zum Niveau $\alpha = 0,05$ nicht statistisch signifikant sind (bei gleichzeitiger Signifikanz der restlichen Paare), werden zu einer Gruppe zusammengefasst.“

„Andernfalls werden keine Paare gebildet, sondern es wird eine separate Aussage für jede Subgruppe getroffen.“

Diese beiden Sätzen sollten neu formuliert werden, da die Aussagen unklar sind. Im 1. Satz „nicht signifikant VERSCHIEDEN sind“? Im 2. Satz „Andernfalls werden keine GRUPPEN gebildet“?

Referenzen:

Welton NJ, Ades AE, Carlin JB, Altman DG, Sterne JAC (2009) Models for potentially biased evidence in meta-analysis using empirical based priors. *JRSS A*;172:119-136.

Wood and Jonathan A C Sterne Jüni, Douglas G Altman, Christian Gluud, Richard M Martin, Anthony J G Lesley Wood, Matthias Egger, Lise Lotte Gluud, Kenneth F Schulz, Peter (2008) Meta-epidemiological study interventions and outcomes: estimates in controlled trials with different doi:10.1136/bmj.39465.451748.AD *BMJ*;336:601-605.

Schwarzer et al. Systematic overview of cost-effectiveness thresholds in ten countries across four continents. *J Comp Eff Res* 2015;4(5):485-504.

Siebert U. When should decision-analytic modeling be used in the economic evaluation of health care? *The European Journal of Health Economics* 2003;4(3):143-50.

Latimer NR et al. Treatment Switching: statistical and decision-making challenges and approaches. *International Journal of Technology Assessment in Health Care* 2016;32(3):160–6.

Henshall C et al. Treatment switching in cancer trials: issues and proposals. *J Technol Assess Health Care*. 2016;32(3):167-74.

A.1.15 – Deutsche Gesellschaft für Orthopädie und Unfallchirurgie e. V. (DGOU)

Autoren:

- Renkawitz, Tobias



PRÄSIDENT 2017
Prof. Dr. med. Ingo Marzi

DGOU Geschäftsstelle · Str. des 17. Juni 106-108 · 10623 Berlin

Herrn
Dennis Makoschey
AWMF-Geschäftsstelle
Birkenstraße 67
10559 Berlin

DGOU Geschäftsstelle
Straße des 17. Juni 106 – 108
10623 Berlin
Tel. +49 - (0)30 - 340 60 36 00
Fax +49 - (0)30 - 340 60 36 01
office@dgou.de
www.dgou.de

Per E-Mail: [REDACTED]

Berlin, 23.01.2017

Stellungnahme zum Entwurf „Allgemeine Methoden“, Version 5.0

Sehr geehrter Herr Makoschey,

zum Entwurf „Allgemeine Methoden“, Version 5.0 des IQWiG bittet die Deutsche Gesellschaft für Orthopädie und Orthopädische Chirurgie (DGOOC), sowie die Deutsche Gesellschaft für Unfallchirurgie (DGU) und die Deutsche Gesellschaft für Orthopädie und Unfallchirurgie (DGOU) nach Beratung mit Prof. Dr. med. Tobias Renkawitz, Leiter des DGOU-Gremiums „Evidenzbasierte Medizin“ (EbM-Gruppe), um Einarbeitung der folgenden Punkte in die Stellungnahme der AWMF:

1. Die im politischen Entscheidungsprozess oftmals so wichtigen „externen Sachverständigen“ will das IQWiG in einer „Sachverständigendatenbank“ führen. Darin können sich *„Interessentinnen und Interessenten unter Angabe ihres Fachgebiets und ihrer fachlichen Expertise“* registrieren lassen. *„Für die zu vergebenden Projekte wird anhand einer Kriterienliste die oder der jeweils am besten geeignete Bewerberin oder Bewerber des entsprechenden Fachgebiets aus der Sachverständigendatenbank ausgewählt und beauftragt.“* Auf Grundlage der EbM-Kompetenz der Deutschen Gesellschaft für Orthopädie und Unfallchirurgie (DGOU) wäre die Einräumung eines Vorschlagsrechtes für die DGOU folgerichtig. Innerhalb und außerhalb der „DGOU-EbM-Gruppe“ gibt es für die unterschiedlichsten Bereiche unseres Faches kompetente Ansprechpartner.
2. Ebenso (siehe auch Pkt. 1) wäre es ebenso folgerichtig, auf Grundlage der EbM-Kompetenz der Deutschen Gesellschaft für Orthopädie und Unfallchirurgie (DGOU), auch für die „externen review“ (*„2.2.4 Review der Produkte des Instituts“- gemeint sind Berichte, Dossierbewertungen etc*) ein Vorschlagsrechtes für die DGOU einzuräumen. *„Die Identifikation externer Reviewer kann durch eine entsprechende Recherche, durch die Kenntnis der Projektgruppe, durch das Ansprechen von Fachgesellschaften, durch eine Bewerbung im Rahmen der Ausschreibung für die Auftragsbearbeitung usw. erfolgen“.*

Mit freundlichen Grüßen

[REDACTED]
Prof. Dr. med. Ingo Marzi
Präsident der DGOU
Präsident der DGU

[REDACTED]
Prof. Dr. med. Andrea Meurer
Stellv. Präsidentin der DGOU
Präsidentin der DGOOC

A.1.16 – Deutsche Gesellschaft für Perinatale Medizin (DGPM)

Autoren:

- Mildenberger, Eva



Deutsche Gesellschaft für Perinatale Medizin

Präsidentin: Prof. Dr. E. Mildenberger, Neonatologie, Klinikum der Johannes Gutenberg-Universität, Mainz; **Vizepräsident:** Prof. Dr. F. Kainer, Abteilung für Geburtshilfe und Pränatalmedizin, Klinik Hallerwiese, Diakonie Neuendettelsau, Nürnberg; **1. Schriftführer:** Prof. Dr. R. Schlößer, Neonatologie, Klinikum der Johann-Wolfgang-Goethe-Universität Frankfurt am Main; **2. Schriftführer:** Prof. Dr. E. Schleußner, Abteilung Geburtshilfe, Universitätsklinikum Jena; **Schatzmeister:** Prof. Dr. R.F. Maier, Zentrum für Kinder- und Jugendmedizin, Universitätsklinikum Gießen und Marburg GmbH, Marburg; **Vorstandsbeiräte:** Prof. Dr. M. Abou-Dakn, Klinik für Gynäkologie und Geburtshilfe, St. Joseph Krankenhaus Berlin; Prof. Dr. U. Gembruch, Abteilung für Geburtshilfe und Pränatalmedizin, Rheinische Friedrich-Wilhelms-Universität, Bonn; Dr. B. Ramsauer, Geburtsmedizin, Vivantes Klinikum Neukölln, Berlin; Prof. R. Rossi, Klinik für Kinder- und Jugendmedizin, Vivantes Klinikum Neukölln, Berlin; Prof. Dr. M. Rüdiger, Neonatologie und Pädiatrische Intensivmedizin, Medizinische Fakultät Carl Gustav Carus der Technischen Universität Dresden; Prof. Dr. H. Stepan, Abteilung für Geburtsmedizin, Universitätsklinikum Leipzig
Ehrenvorsitzende: Prof. Dr. E. Saling, Gründungspräsident, Institut für Perinatale Medizin, Berlin; Prof. em. Dr. J.W. Dudenhausen, Geburtsmedizin, Charité, Campus Virchow Klinikum, Berlin

An das
IQWiG
Im Mediapark 8
50670 Köln

Mainz, 27.1.2017

Stellungnahme der DGPM zu dem von Ihnen verfassten Dokument Entwurf „Allgemeine Methoden“ Version 7.12.2016

Sehr geehrte Damen und Herren,

Sie baten um Stellungnahme zum sogenannten „Methodenpapier“ in der Entwurfversion vom 7.12.2016. Die DGPM gibt keinen Kommentar ab.

Dies ist darin begründet, dass die Vorstandsmitglieder der DGPM sich nicht in der Lage sehen, ein derart umfangreiches Dokument zu komplexen Inhalten im Rahmen ihrer ehrenamtlichen Tätigkeit in angemessener Weise zu evaluieren. Mithin bedeutet die fehlende Kommentierung keinesfalls eine a priori-Befürwortung.

Mit bestem Gruß,

Univ.-Prof. Dr. Eva Mildenberger
Präsidentin der DGPM

Präsidentin: Univ.-Prof. Dr. E. Mildenberger, Neonatologie, Universitätsmedizin der Johannes Gutenberg-Universität, Mainz, Langenbeckstr. 1, 55131 Mainz, [REDACTED]
Vizepräsident: Prof. Dr. F. Kainer, Abteilung für Geburtshilfe und Pränatalmedizin, Klinik Hallerwiese, St.-Johannis-Mühlgasse 19, 90419 Nürnberg, [REDACTED]
Schatzmeister: Univ.-Prof. Dr. R. F. Maier, Universitätsklinikum Marburg, Baldingerstraße, 35043 Marburg, [REDACTED]

A.1.17 – Deutsche Krankenhausgesellschaft e. V. (DKG)

Autoren:

- Brenske, Michael

Institut für Qualität und
Wirtschaftlichkeit im Gesundheitswesen
Herr Prof. Dr. med. Jürgen Windeler
-Stellungnahme zu den „Allgemeinen
Methoden Version 5.0“ -
Im Mediapark 8 (Köln Turm)
50670 Köln

**Stellungnahme der Deutschen Krankenhausgesellschaft zum Entwurf
„Allgemeine Methoden Version 5.0“ des Instituts für Qualität und
Wirtschaftlichkeit im Gesundheitswesen (IQWiG) vom 07.12.2016**

Sehr geehrter Herr Prof. Windeler,
sehr geehrte Damen und Herren,

für die Möglichkeit, zum Entwurf der „Allgemeinen Methoden Version 5.0“, Stellung nehmen zu können, möchten wir Ihnen herzlich danken. In einem komplexen Feld, wie es durch das Methodenpapier abgedeckt wird, ist die Ermöglichung eines Austausches über deren Inhalte, wie er durch ein Stellungnahmeverfahren gewährleistet wird, aus unserer Sicht unverzichtbar.

Den mit den „Allgemeinen Methoden“ verfolgten Ansatz, eine Darstellung Ihrer Vorgehens- und Arbeitsweisen vorzunehmen, begrüßen wir sehr. Nur so kann den Anforderungen zur wissenschaftlichen Transparenz und Nachvollziehbarkeit Ihrer unterschiedlichen Produkte in differenzierter Weise nachgekommen werden. Dies betrifft insbesondere auch die Ausführlichkeit der Kapitel zur Informationsbeschaffung und zu den Gesundheitsinformationen. Der Umstand, dass es sich mittlerweile um die Version 5.0 handelt, zeigt zudem, dass die „Allgemeinen Methoden“ der notwendigen regelmäßigen Überarbeitung und Weiterentwicklung unterliegen. Die Inhalte unserer Stellungnahme möchten wir daher auch als konstruktive Anregungen in einem gewissermaßen kontinuierlichen Dialog verstanden wissen. Im Einzelnen möchten wir an dieser Stelle zu den nachfolgenden Aspekten Stellung nehmen

Interne Qualitätssicherung

Im Methodenpapier wird an verschiedenen Stellen auf die interne Qualitätssicherung der Instituts-Produkte (wie z.B. auf S. 21, 28, 38) hingewiesen. Leider wird nicht näher erläutert, wie diese wichtigen Prozesse gestaltet sind. Im Sinne der gebotenen Transparenz wäre aus unserer Sicht eine genauere Beschreibung dahingehend zielführend, wie diese internen Clearingverfahren bzw. internen Qualitätssicherungsprozesse konkret gestaltet sind.

Einbeziehung medizinischer Fachexpertise

Die Beschreibung der Einbeziehung medizinischer Fachexpertise in die Erstellung der Instituts-Produkte erfolgt unterschiedlich, ohne dass dabei hinreichend klar wird, nach welchen Grundsätzen dabei vorgegangen wird. Beispielsweise wird bei manchen der Produkte lediglich von einem optionalen externen Review oder optionaler Hinzuziehung von Sachverständigen (wie z.B. auf S. 21, 27, 31, 32) gesprochen. An anderer Stelle wird dagegen auf „Forschungsaufträge“ verwiesen, die vergeben werden. Es wäre u. E. hilfreich, nähere Erläuterungen dahingehend vorzunehmen, in welchen Situationen, nach welchen Kriterien und in welchem Umfang mit welcher Intensität zusätzlich welche medizinische Fachexpertise einbezogen wird. Dies würde auch dem Transparenz-Gedanken entsprechen, da nur so in etwa ermessen werden kann, zu welchen Anteilen die Arbeiten der Sachverständigen und des IQWiG in die Erstellung eines Produktes eingeflossen sind.

Offenkundig scheinen Maßstäbe für den entsprechenden Auswahlprozess zu existieren, da an einer Stelle (S. 37) von einer „Kriterienliste“ gesprochen wird. Neben einem allgemeinen Verweis zum Auswahlverfahren auf der Website des Instituts wäre es daher zielführend, die Kriterien dieser Liste in Ihren „Allgemeinen Methoden“ zu konkretisieren. Auf S. 37 Ihres Methodenpapiers heißt es zudem *„...Gemäß seinem gesetzlichen Auftrag bezieht das Institut externe Sachverständige in seine Arbeit ein. Als externe Sachverständige gelten dabei Personen, an die wissenschaftliche Forschungsaufträge im Rahmen der Erstellung oder Begutachtung von Institutsprodukten vergeben werden oder die das Institut in medizinisch-fachlichen Fragestellungen beraten...“* Hier bleibt es für Außenstehende unklar, was konkret unter „Forschungsaufträgen“ zu verstehen ist und wie die Einbindung von externen Sachverständigen dabei konkret erfolgt.

Mit Blick auf die Beauftragung externer Sachverständiger werden auf S. 37 verschiedene Produkte des Instituts samt deren gesetzlicher Grundlage, für die eine Sachverständigendatenbank herangezogen wird, genannt. In der Aufzählung fehlen allerdings die Berichte zu Fragestellungen nach den §§ 135 und 137c SGB V. Wir gehen davon aus, dass auch bei diesen, oft sehr umfangreichen, Aufträgen, regelmäßig externe medizinische Sachverständige einbezogen werden, sodass die Aufzählung entsprechend ergänzt werden sollte. Wenngleich eine Sachverständigendatenbank als Grundlage genannt wird, bleibt offen, inwieweit daneben auch ein proaktives Zugehen von Seiten des Instituts auf die medizinische Fachwelt zur Einbindung von Sachverständigen vorgesehen ist.

Wissenschaftliche Beratung

Unter 2.2.7 wird in einem kurzen Absatz auf „Wissenschaftliche Beratung“ eingegangen. Dabei ist es offensichtlich auch vorgesehen, Sponsoren von Studien spezifisch bezüglich des Studiendesign zu beraten. Es bleibt dabei allerdings unklar, in welcher Tiefe und Verbindlichkeit solche Beratungen durchgeführt werden. Dabei scheint es auch Fälle zu geben, bei denen das IQWiG diese Studien dann in einer eigenen Bewertung aufgreift. Zur Gewährleistung der Unabhängigkeit des IQWiG wird allgemein auf eine „angemessene organisatorische Trennung“ von Beratung und Bewertung hingewiesen. Dies sollte u. E. einer genaueren Beschreibung unterliegen. Daneben wäre zudem eine Aussage wichtig, wie sich Beratungen des IQWiG zu den gesetzlich vorgesehenen Beratungen des Gemeinsamen Bundesausschusses (G-BA), beispielsweise nach den §§ 35a Abs. 7, 137e Abs. 8 und 137h Abs. 6 SGB V verhalten.

Nutzenbeleg/ Endpunktbezogene Bewertung

Im Kapitel 3.1 (Patientenrelevanter medizinischer Nutzen und Schaden) erfolgt 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 gut nachvollziehen. 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. Dekliniert man nämlich die gesetzlichen Regelungen der §§ 137c, 137e und 137h im Hinblick auf die Situation der Durchführung einer Erprobung (mögliche Feststellung des Potenzials auf Grundlage von Studien, die keine RCTs sind) durch, 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.

Zusammenfassend wäre es aus unserer Sicht daher 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.

Auch in anderen Punkten sehen wir noch einen Überarbeitungsbedarf. Legt man die Maßstäbe für einen einzelnen RCT als Beleg (mindestens 2000 Teilnehmer/innen mindestens 1000 je Studienarm) im Bereich seltener oder sehr seltener Erkrankungen zugrunde, so wird auch deutlich, dass diese schwer realisierbar sein können. Auch erscheint uns beispielweise die Fokussierung auf den p-Wert mit $< 0,001$ 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 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. Es wäre daher 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. 48) 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 späteren Kapiteln abgehandelt werden (z. B. dramatischer Effekt, Umgang mit kleinen Populationen bei seltenen Erkrankungen).

Diagnostische Verfahren

Im Kontext der Bewertung von Diagnostischen Methoden erfolgt auch eine Befassung mit dem Werkzeug der „Linked Evidence“. Für uns etwas überraschend, wird „Linked Evidence“ allerdings nur im Zusammenhang mit der Bewertung von Screening-Verfahren als Möglichkeit angesehen, bei anderen diagnostischen Verfahren aber scheinbar nicht in Betracht gezogen. Da es sich beim Screening ebenfalls um diagnostische Verfahren handelt, ist für uns dieser Ansatz unterschiedlicher Bewertungsmaßstäbe nicht nachvollziehbar. Immerhin betreffen Screeningverfahren einschließlich ihres Schadenspotenzials bei der Früherkennung von Erkrankungen viel größere Bevölkerungsgruppen. Gerade vor diesem Hintergrund stellt sich uns die Frage, warum bei der gezielten klinischen Diagnostik höhere Bewertungsmaßstäbe ohne die grundsätzliche Möglichkeit von Linked Evidence angesetzt werden.

Meta-Analysen

Bei der Durchsicht der Ausführungen zu Meta-Analysen in Kapitel 3 und Kapitel 9 ist uns positiv aufgefallen, dass offenbar einzelne Anregungen aus unseren früheren Stellungnahmen aufgegriffen wurden. Daneben möchten wir zur begrifflichen Klarstellung (z.B. auf S. 188) anregen, Fixed-Effects-Modelle mit homogenen Studienergebnissen, deren Schwankungsbreite dem Zufall unterliegt, von Random-Effects-Modellen, deren Studienergebnisse durch tatsächliche Unterschiede zu erklären sind und nicht primär dem Zufall unterliegen, abzugrenzen. Als Quelle könnte das Cochrane Handbook herangezogen werden.

Wie Ihnen bekannt, haben wir auch schon zu Vorgänger-Versionen der „Allgemeinen Methoden“ Stellungnahmen abgegeben. Deren Inhalte haben – soweit sie nicht schon berücksichtigt wurden – prinzipiell auch weiterhin Bestand und wurden, zur Vermeidung von Redundanzen, hier insofern nicht sämtlich wiederholt. Vor dem Hintergrund des erheblichen Umfangs der „Allgemeinen Methoden“ und der Komplexität der darin abgehandelten Inhalte ist unsere Stellungnahme nicht als abschließend anzusehen. Wir behalten uns daher vor, ggf. weitere Aspekte in zukünftige Diskussionsprozesse einzubringen.

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.1.18 – Deutsches Netzwerk für evidenzbasierte Medizin e. V. (DNEbM)

Autoren:

- Lühmann, Dagmar

Stellungnahme des DNEbM zu „Allgemeine Methoden (Entwurf für Version 5.0 vom 07.12.2016)“ des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG)

Die Stellungnahme bezieht sich auf die Ausführungen zu HTA-Berichten (im Wesentlichen Abschnitt 6) und Versorgungsanalysen (im Wesentlichen Abschnitt 5.4).

Das DNEbM begrüßt es, dass nach der Einstellung des Health Technology Assessment (HTA) Programms beim Deutschen Institut für Medizinische Dokumentation und Information (DIMDI) erneut eine Institution des deutschen Gesundheitswesens den gesetzlichen Auftrag erhalten hat, umfassende Bewertungen von Gesundheitstechnologien (HTA) zu übernehmen. Beklagenswert ist allerdings, dass durch die erneute Verankerung des Auftrages im SGB V weite gesundheitsrelevante Versorgungsbereiche, wie Rehabilitation (SGB IX), Pflege (SGB XI) oder Public Health Interventionen von den Bewertungen damit ausgeschlossen sind. Ausgeschlossen wären ebenfalls Bewertungen von Interventionen, für deren Umsetzung die Verantwortung größtenteils bei den Bürgern selber liegt (wie z. B. Lebensstiländerungen als Präventionsmaßnahme, Pflegekonzepte für Familienangehörige).

Sehr klar wird diese GKV-Fokussierung in den Ausführungen zur Themenfindung für HTA: *„Nach § 139b Absatz 5 SGB V können Versicherte und interessierte Einzelpersonen Bewertungen zu medizinischen Untersuchungs- und Behandlungsmethoden bei ausgewählten Krankheiten sowie zu Fragen der Qualität und Wirtschaftlichkeit der im Rahmen der gesetzlichen Krankenversicherung erbrachten Leistungen vorschlagen.“* Dies ist umso bedauerlicher, da es kein weiteres verlässlich öffentlich gefördertes HTA-Programm in Deutschland gibt. Das DNEbM würde es begrüßen, wenn das IQWiG seinen Auslegungsspielraum nutzt und den unterstrichenen Halbsatz aus dem Absatz zur Themenfindung streicht.

Im Entwurf der „Allgemeinen Methoden Version 5.0“ wird die geplante Umsetzung von Health Technology Assessment am IQWiG dargelegt (Einbettung, Themenwahl, methodische Herangehensweise). Hierauf beziehen sich unsere Kommentare.

1. Im HTA Konzept des IQWiG bleibt die Bewertung von komplexen gesundheitsrelevanten Interventionen (weitgehend) ausgespart. Gerade bei der Bewertung komplexer Interventionen kommt allerdings das Potenzial der per definitionem multidimensional angelegten Technologiebewertung des HTA zum Tragen (vergl. hierzu z. B. die Arbeiten des INTEGRATE-HTA Projektes, www.integrate-hta.eu).

Die Bewertung von komplexen Interventionen wird leider nur cursorisch im Abschnitt 5.3.1 (Inhalte der Versorgungsanalyse) adressiert ohne allerdings in den folgenden Abschnitten konkrete methodische Vorgaben für die Bewertung komplexer Interventionen zu machen.

Warum das IQWiG Methodenhandbuch konsequent den seit mehr als 10 Jahren international geführten methodischen Diskurs über komplexe Interventionen ausspart, ist für das DNEbM nicht nachvollziehbar. Neben dem methodischen Leitfaden des britischen Medical Research Council mitsamt präzisierenden Methodenbeiträgen zum Beispiel zur Prozessevaluation, gibt es bereits viele konkrete publizierte Beispiele der Umsetzung von Synthesen komplexer Interventionen. Das DNEbM hätte spätestens im Methodenpapier Version 5.0 eine Reflexion der Wissensbestände über die Identifikation, Analyse, Synthese und Interpretation komplexer Interventionen erwartet.

2. Das DNEbM begrüßt es grundsätzlich, dass mit dem Portal „ThemenCheck Medizin“ Patienten und Bürger die Möglichkeit erhalten, Themen für eine systematische Bewertung nach dem HTA-Konzept vorzuschlagen. Ein derart „offenes“ Verfahren birgt aus unserer Sicht allerdings auch die Gefahr, dass interessengeleitete Bewertungen vorgeschlagen werden. Das DNEbM wünscht sich daher größtmögliche Transparenz bei der weiteren systematischen Prozessierung der Themenvorschläge. Inwieweit diese realisiert werden soll und kann, ist den Ausführungen in Abschnitt 6.3 der „Allgemeinen Methoden“ nicht zu entnehmen. An dieser Stelle wünscht das DNEbM weitere Klarstellungen (wie z. B. Darlegung der Gewichtung von Auswahlkriterien, Interessenkonfliktprüfungen, Dokumentationspflichten, Einsichtnahme-rechte).

3. Die Erstellung von HTA-Berichten ist per definitionem eine multidisziplinäre Aufgabe, die Technologiebewertungen aus klinischer, ökonomischer, organisationaler, sozialer, rechtlicher und ethischer Perspektive vornimmt (1). Im vorliegenden Methodenpapier ist die Beschreibung der vom IQWiG zur Erstellung von HTA-Berichten vorgesehenen Methodik (Abschnitt 6.5 Bearbeitung der Themen) von großer Heterogenität und Detailliertheit gekennzeichnet. Während für die Dimensionen Nutzenbewertung und Gesundheitsökonomie durch Verweise auf die entsprechenden Abschnitte des Methodenpapiers sehr elaborierte Vorgaben bestehen, sind die Ausführungen und Vorgaben zu den übrigen Dimensionen eher allgemein gehalten. Insbesondere wird nicht klar, ob die Umsetzung von HTA-Projekten im IQWiG sich ausschließlich auf Methoden der Sekundärforschung stützen soll oder ob für ausgewählte Aspekte primäre Forschungsmethoden (= Datenerhebungen) zum Einsatz kommen sollen. Die Aufzählung potenzieller methodischer Herangehensweisen zur Bearbeitung legt letzteres nahe. Das DNEbM regt an, diese Aspekte expliziter zu formulieren.

4. Patientenbeteiligung: an mehreren Stellen im Methodenpapier (z. B. Abschnitt 2.2.1) wird darauf verwiesen, dass die Einbindung von Patienten in die Erstellung von systematischen Reviews und Health Technology Assessments derzeit zum etablierten Standard gehört. Abschnitt 2.2.1 führt weiterhin aus, wie die Einbindung von Patienten in die Erstellung der unterschiedlichen IQWiG-Produkte umgesetzt wird. Für die Erstellung von HTA-Berichten ist keine direkte Beteiligung von Patienten vorgesehen. Dies entspricht zum einen nicht den

internationalen Standards und kann – je nach methodischem Ansatz – für die Bewertung einiger Dimensionen (Ethik, soziale Aspekte) unabdingbar sein.

5. Abschnitt 9.4 befasst sich mit dem Stellenwert qualitativer Studien für die einzelnen IQWiG Produkte. Hier fehlt der Verweis, dass qualitative Studien (oder sogar die Gewinnung von Primärdaten) bei der Bewertung von (komplexen) gesundheitsrelevanten Technologien einen hohen Stellenwert haben können und ein spezifisches methodisches Vorgehen erfordern. Dies bezieht sich insbesondere auch auf umfassende Bewertungen nach dem HTA-Konzept (2,3,4).

Literatur

(1) <http://www.htai.org/htai/what-is-hta.html> (accessed 20170129)

(2) Andrew Booth, Jane Noyes, Kate Flemming, Ansgar Gerhardus, Philip Wahlster, Gert Jan van der Wilt, Kati Mozygemba, Pietro Refolo, Dario Sacchini, Marcia Tummers, Eva Rehfues: Guidance on choosing qualitative evidence synthesis methods for use in health technology assessments of complex interventions. European Union Seventh Framework Programme ([FP7/2007-2013] [FP7/2007-2011]) under Grant Agreement No. 306141. (<http://www.integrate-hta.eu/downloads>, accessed 20170129)

(3) Ring N, Jepson R, Ritchie K. Methods of synthesizing qualitative research studies for health technology assessment. *Int J Technol Assess Health Care*. 2011;27(4):384–90.

(4) DeJean D, Giacomini M, Simeonov D, Smith A. Finding Qualitative Research Evidence for Health Technology Assessment. *Qual Health Res* 2016;26(10):1307–17.

A.1.19 – Ecker + Ecker GmbH

Autoren:

- Pütz, Claudia

**Stellungnahme der Ecker + Ecker GmbH zu
Allgemeinen Methoden – Entwurf 5.0 vom 07.12.2016**

Autoren:

Dr. Christof Ecker

Dr. Thomas Ecker

Dr. Malte Moos

Dr. Claudia Pütz

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 Frühe Nutzenbewertung

1. Zweckmäßige Vergleichstherapie
2. Zugelassene Anwendungsgebiete
3. Medizinischer Nutzen
4. Nichtmedikamentöse therapeutische Interventionen
5. „Frühbewertung von Arzneimitteln“ unter www.gesundheitsinformation.de

Kompetenzbereich Biostatistik

6. Evidenztransfer
7. Indirekte Vergleiche
8. Subgruppenanalysen

Kompetenzbereich Gesundheitsökonomie

9. Kostendarstellung
10. Epidemiologie

**Kompetenzbereich Frühe Nutzenbewertung:
1. Zweckmäßige Vergleichstherapie**

Bezug: Abschnitt 3.3.3 der Allgemeinen Methoden, Entwurf für Version 5.0

Problembeschreibung:

In diesem Abschnitt erläutert das IQWiG, dass im Rahmen der Nutzenbewertung eines Arzneimittels gemäß § 35a SGB V der pharmazeutische Unternehmer u. a. Angaben zum medizinischen Zusatznutzen im Verhältnis zur zweckmäßigen Vergleichstherapie machen muss.

In Abschnitt 3.3.3 finden sich umfangreiche Erläuterungen, wie das Institut die in den vorgelegten Unterlagen beschriebenen Effekte unter Berücksichtigung ihrer Ergebnissicherheit bewertet. Zur zweckmäßigen Vergleichstherapie heißt es nur, dass geprüft wird, ob sie als zweckmäßig im Sinne des § 35a SGB V und der Arzneimittel-Nutzenbewertungsverordnung gelten kann. Wie diese Prüfung erfolgt, wird in den Allgemeinen Methoden nicht erläutert.

Handlungsbedarf:

Im Rahmen der Prüfung der zweckmäßigen Vergleichstherapie beschränkt sich das IQWiG zumeist darauf, die vom Hersteller gewählte und operationalisierte zweckmäßige Vergleichstherapie mit der durch den G-BA im Beratungsgespräch getroffenen Festlegung abzugleichen. Teilweise variiert das IQWiG auch den Wortlaut der zweckmäßigen Vergleichstherapie, ohne dass Gründe hierfür angeführt werden oder die methodischen Implikationen hierfür erläutert werden. Eine Prüfung der Zweckmäßigkeit im Sinne des § 35a SGB V und der Arzneimittel-Nutzenbewertungsverordnung erfolgt jedoch nicht – auch nicht, wenn diese aufgrund des medizinischen Erkenntnisfortschritts offensichtlich nicht mehr dem Stand des medizinischen Wissens entspricht.

Vor diesem Hintergrund wäre es wichtig, in den Allgemeinen Methoden die Prüfkompetenz des IQWiG zu konkretisieren.

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 „3.4.3 Überprüfung der zweckmäßigen Vergleichstherapie“ eingeführt werden. In diesem Abschnitt wäre zu beschreiben, wie das IQWiG auf Basis der Kriterien aus § 6 Abs. 3 des 5. Kapitels der VerfO des G-BA die vom pharmazeutischen Unternehmer gewählte zweckmäßige Vergleichstherapie überprüft.

**Kompetenzbereich Frühe Nutzenbewertung:
2. Zugelassene Anwendungsgebiete**

Bezug: Abschnitt 3.3.3 der Allgemeinen Methoden, Entwurf für Version 5.0

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 zugelassene(n) Anwendungsgebiete(n) machen muss. In Abschnitt 3.3.3 finden sich jedoch keine methodischen Anmerkungen, wie das IQWiG mit diesen Angaben umzugehen plant und diese in seiner Nutzenbewertung gemäß § 35a SGB V berücksichtigt.

Handlungsbedarf:

Im Rahmen der frühen Nutzenbewertung stellt sich das Problem, dass ein Arzneimittel für ein umfassendes Anwendungsgebiet zugelassen ist, auch wenn es praktisch nur in einem Teilbereich des Anwendungsgebietes nach dem bisherigen Stand des medizinischen Wissens angewendet wird. Damit ergibt sich die praktische Frage, welche Anwendungsmöglichkeiten eines zu bewertenden Arzneimittels der Nutzenbewertung zugrunde gelegt werden. So verzichtet das IQWiG beispielsweise in der COPD auf eine Nutzenbewertung in dem Teilanwendungsgebiet „COPD-Schweregrad I“, obwohl die nutzenbewerteten Arzneimittel hierfür zugelassen sind, während es beispielsweise in der Depression im Teilanwendungsgebiet „Leichte Episode einer Major Depression“ eine Nutzenbewertung durchführt. Für beide Teilanwendungsgebiete ist aber entsprechend der jeweiligen S3-Leitlinie eine medikamentöse Therapie nicht indiziert.

Dieses Vorgehen erscheint methodisch inkonsistent.

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 „3.4.1 Zugelassene und zu bewertende Anwendungsgebiete“ eingeführt werden. In diesem Abschnitt wäre zu beschreiben, wie das IQWiG das zugelassene (Teil-) Anwendungsgebiet auf der Grundlage des Standes des medizinischen Wissens methodisch so auslegt, dass eine Nutzenbewertung nur dort erfolgt, wo die Anwendung des zu bewertenden Arzneimittels tatsächlich dem Stand des medizinischen Wissens entspricht.

**Kompetenzbereich Frühe Nutzenbewertung:
3. Medizinischer Nutzen**

Bezug: Abschnitt 3.3.3 der Allgemeinen Methoden, Entwurf für Version 5.0

Problembeschreibung:

In diesem Abschnitt erläutert das IQWiG, dass im Rahmen der Nutzenbewertung eines Arzneimittels gemäß § 35a SGB V der pharmazeutische Unternehmer u. a. Angaben zum medizinischen Nutzen machen muss. In Abschnitt 3.3.3 finden sich jedoch keine methodischen Anmerkungen, wie das IQWiG mit diesen Angaben umzugehen plant und diese in seiner Nutzenbewertung gemäß § 35a SGB V berücksichtigt.

Das hier beschriebene Problem besteht unabhängig von den Ausführungen in „Abschnitt 4.3 Nutzen“, da dieser Abschnitt auf einem anderen Nutzenbegriff und einer anderen methodischen Fragestellung (Kosten-Nutzen-Bewertung) beruht.

Diese Fragestellung ist nicht zu verwechseln mit der Frage, unter welchen Umständen die Forderung nach (parallel) vergleichenden Studien unangemessen sein kann und daher eine geringwertige Evidenz ausreichend ist (siehe Abschnitt 3.2.5). Vielmehr geht es um Konstellationen, in denen keine direkte Evidenz vorliegt bzw. vorliegen kann.

Handlungsbedarf:

Nach § 7 Abs. 1 der AM-NutzenV führt das IQWiG im Auftrag des G-BA die Nutzenbewertung auf Basis des vom pharmazeutischen Unternehmen eingereichten Nutzendossiers durch. Insofern hat das IQWiG auch Aussagen zum medizinischen Nutzen zu bewerten, da dies in den nach § 35a Abs. 1 Satz 3 Nr. 2 SGB V verpflichtenden Angaben des pharmazeutischen Unternehmens enthalten ist.

Das IQWiG verfügt in seinen Allgemeinen Methoden bislang nicht über eine Methodik zur Bewertung dieser Angaben, muss sich also bei seiner Bearbeitung mit wissenschaftlich wenig befriedigenden „ad hoc“-Entscheidungen behelfen, sofern es sich überhaupt mit dieser Frage beschäftigt.

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 „3.4.2 Nutzenbewertung“ eingeführt werden. In diesem Abschnitt wäre zu beschreiben, welche Evidenz das IQWiG seiner Bewertung zugrunde legt:

- Patientenpopulation
- Intervention
- Endpunkte
- Studientypen (u. a. Randomisierung, Dauer)

und wie die Feststellung des Ausmaßes des medizinischen Nutzens operationalisiert wird.

Kompetenzbereich Frühe Nutzenbewertung:**4. Nichtmedikamentöse therapeutische Interventionen**

Bezug: Abschnitt 3.4 der Allgemeinen Methoden, Entwurf für Version 5.0

Problembeschreibung:

Auf Seite 66, Abschnitt 2 schreibt das IQWiG: „Die Anforderungen für einen Nutzen sind jedoch bei Berichten und Rapid Reports im Vergleich zu Bewertungen gemäß § 137h SGB V gleich und erfordern in der Regel das Vorhandensein von RCT-Ergebnissen zu patientenrelevanten Endpunkten“.

Nicht medikamentöse therapeutische Interventionen und insbesondere Medizinprodukte sind nicht per se immer geeignet solche Evidenz zu generieren. Zum Beispiel ist die Verblindung nur in seltenen Fällen möglich, was zu einem Bias führen kann.

Handlungsbedarf:

Solche nicht medikamentöse „therapeutische“ Interventionen sind Eingriffe: „[...] die eine nicht etablierte, besondere ärztliche Tätigkeit darstellen, die spezialisiertes ärztliches Können oder Erfahrung voraussetzen. Dieses Vorgehen ist mit der normalen Verabreichung eines Medikaments in der Erwartung, dass es im Körper die erwünschte Wirkung entfaltet, qualitativ nicht vergleichbar [...]“ (siehe G-BA Beschluss zu C3BS-CQR-1 (C Cure[®]) vom 09.03.16)

Aus diesem Grund besitzt „ärztliches Können oder Erfahrung“ bei nicht medikamentösen therapeutischen Interventionen einen hohen Stellenwert. Dies wurde in Ländern mit einer langjährigen Erfahrung in der Methodenbewertung wie z. B. Frankreich bereits berücksichtigt.

Während an die Bewertung von Nutzen und Schaden von Medizinprodukten bezüglich der Ergebnissicherheit prinzipiell kein anderer Maßstab angelegt wird, werden in Frankreich bei der Bewertung Registerdaten berücksichtigt. Die Berücksichtigung von Registern zusammen mit regelmäßigen Neubewertungen ermöglicht den Zugang zu sicheren Innovationen für Patienten.

Lösungsvorschlag:

Im Rahmen dieser Stellungnahme wird die Einfügung von Kriterien und Methoden vorgeschlagen, die beschreiben, wie der höchste Evidenzgrad mit den inhärenten Charakteristiken der nicht medikamentösen therapeutischen Interventionen erreicht werden kann. Beim Fehlen von randomisierten kontrollierten Studien sollte eine angemessene Berücksichtigung von Daten aus alternativen Studienformen stattfinden.

Kompetenzbereich Frühe Nutzenbewertung:**5. „Frühbewertung von Arzneimitteln“ unter www.gesundheitsinformation.de**

Bezug: Abschnitt 7.1 der Allgemeinen Methoden, Entwurf für Version 5.0

Problembeschreibung:

Der Abschnitt „Evidenzbasierte Gesundheitsinformation für Bürgerinnen und Bürger“ wurde neu aufgenommen. In diesem Abschnitt erläutert das IQWiG sowohl die Zielsetzung als auch die Themenauswahl und das Vorgehen beim Erstellen von Gesundheitsinformationen.

Es ist grundsätzlich zu begrüßen, dass das IQWiG die Zielsetzung bei der Erstellung von Gesundheitsinformationen in seinem Methodenpapier angibt. Es fällt jedoch die Diskrepanz zu den tatsächlichen Inhalten von www.gesundheitsinformation.de auf. Nach Angaben des Instituts ist das primäre Ziel „die verständliche Vermittlung von entscheidungsrelevantem Wissen zu gesundheitlichen Fragen. Das Wissen soll Nutzerinnen und Nutzern in ihrer Autonomie und Kompetenz unterstützen, informiert unter verschiedenen Optionen auszuwählen.“ Mit diesem Ziel ist die verkürzte Darstellung der Ergebnisse der Bewertungsberichte des IQWiG zur frühen Nutzenbewertung nach § 35a SGB V nicht vereinbar. Die Zielsetzung der frühen Nutzenbewertung ist ganz explizit die Schaffung einer Basis für die Festlegung eines Erstattungsbetrags, nicht aber Vor- und Nachteile einer Therapieoption in einem therapeutischen Gesamtkontext darzustellen. In der vereinfachten und verkürzten Darstellung herausgelöst aus dem korrekten Gesamtkontext, kann dies zu einer Fehlinformation von Nutzerinnen und Nutzern führen – was der Zielsetzung der Gesundheitsinformationen entgegensteht.

Handlungsbedarf:

Die tatsächlichen Inhalte der Gesundheitsinformationen des IQWiG sind dahingehend zu prüfen, ob sie mit der Zielsetzung und dem gesetzlichen Auftrag des Instituts übereinstimmen.

Lösungsvorschlag:

Im Rahmen dieser Stellungnahme wird vorgeschlagen, die „Frühbewertung von Arzneimitteln“ unter www.gesundheitsinformation.de zu entfernen, da die dort enthaltenen Informationen die Nutzerinnen und Nutzer nicht bei einer informierten Entscheidung unterstützen können.

Kompetenzbereich Biostatistik:**6. Evidenztransfer**

Bezug: Abschnitt 3.3.3 der Allgemeinen Methoden, Entwurf für Version 5.0

Problembeschreibung: In Abschnitt 3.3.3 finden sich umfangreiche Erläuterungen, wie das Institut die in den zur frühen Nutzenbewertung vorgelegten Unterlagen beschriebenen Effekte unter Berücksichtigung ihrer Ergebnissicherheit bewertet.

Dabei stellt sich die Frage, inwieweit die für eine Population verfügbare Evidenz auf andere Populationen übertragen werden kann (Evidenztransfer).

Handlungsbedarf:

Insbesondere in Konstellationen, in denen die Durchführung von Studien aus epidemiologischen, ethischen oder anderen praktischen Gründen nicht realisierbar scheint, stellt sich die Frage, inwieweit ein Evidenztransfer zulässig oder nicht sogar geboten ist. Die EMA hat hierzu für Zulassungsentscheidungen schon einen ersten Bewertungsrahmen vorgelegt, sowohl zu grundsätzlichen Fragen der Extrapolation wie auch zu speziellen Fragen der Extrapolation in pädiatrischen Populationen (Concept paper „Extrapolation of efficacy and safety in medicine development“ und „Extrapolation of efficacy and safety in paediatric medicine development“).

Das IQWiG verfügt derzeit nicht über einen vergleichbaren, allgemeinen Bewertungsrahmen, muss sich also bei in der Nutzenbewertung vorgelegten Evidenztransfers mit wissenschaftlich wenig befriedigenden „ad hoc“-Entscheidungen“ behelfen.

Das deckt nicht die Intention des Gesetzgebers, der mit dem Arzneimittelversorgungsstärkungsgesetz vorsieht, dass bei der Bewertung von Kinderarzneimitteln ein Zusatznutzen anerkannt werden kann, wenn die Zulassung aufgrund eines Evidenztransfers ausgesprochen wurde (§ 5 AM-NutzenV).

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 „3.4.4 Evidenztransfer“ eingeführt werden. In diesem Abschnitt wäre u. a. zu beschreiben:

- In welchen Fällen ist eine Extrapolation grundsätzlich notwendig bzw. sinnvoll?
- Welche methodischen Anforderungen werden an eine Extrapolation gestellt?
- Wie sind die Ergebnisse der Extrapolation im Rahmen des Konzepts der frühen Nutzenbewertung zu interpretieren?

**Kompetenzbereich Biostatistik:
7. Indirekte Vergleiche**

Bezug: Abschnitt 9.3.9 der Allgemeinen Methoden, Entwurf für Version 5.0

Problembeschreibung:

In diesem Abschnitt zu indirekten Vergleichen verlangt das IQWiG eine „adäquate Begründung“, um auf Methoden für indirekte Vergleiche zurückzugreifen.

Des Weiteren werden nach wie vor ausschließlich adjustierte indirekte Vergleiche über „adäquate Brückenkomperatoren“ akzeptiert. Mögliche Methoden sind das Verfahren von Bucher oder das in Abschnitt 9.3.8 beschriebene Verfahren zu Netzwerk-Meta-Analysen. Diese Verfahren setzen die „ausreichende Ähnlichkeit“ von Studien und die Homogenität der paarweisen Meta-Analysen voraus.

Handlungsbedarf:

Gerade vor dem Hintergrund der frühen Nutzenbewertung, in der ca. 30% der Verfahren auf indirekte Vergleiche zurückgreifen, wäre es zu begrüßen, eine geeignete Methodik bzgl. der genannten Punkte festzulegen. Die Bewertungspraxis zeigt, dass es bislang kaum möglich ist, die Anforderungen des IQWiG an einen indirekten Vergleich zu erfüllen. Dies dürfte auch auf die unklaren Anforderungen zurückzuführen sein.

Lösungsvorschlag:

Im Rahmen dieser Stellungnahme wird vorgeschlagen, in Abschnitt 9.3.9 klare Kriterien zu definieren, die festlegen was unter

- einer adäquaten Begründung,
- adäquaten Brückenkomperatoren und
- einer ausreichenden Ähnlichkeit

zu verstehen ist. Eine Quantifizierung der Kriterien ist zu bevorzugen.

**Kompetenzbereich Biostatistik:
8. Subgruppenanalysen**

Bezug: Abschnitt 9.3.10 der Allgemeinen Methoden, Entwurf für Version 5.0

Problembeschreibung:

Bei der Überarbeitung des Methodenpapiers (Zusammenführung der Abschnitte zu Subgruppenanalysen) wurde die Diskussion über die Problematik des multiplen Testens vollständig entfernt. Dies ist unverständlich, da diese Problematik häufig auftaucht und eine hohe Relevanz bei der Interpretation biostatistischer Ergebnisse hat.

Handlungsbedarf:

Aufgrund vielfältiger Subgruppenmerkmale ist es durchaus üblich, auf eine hohe Anzahl von Subgruppenanalysen zu kommen. Es ist leicht möglich, dass die Anzahl der Tests auf Effektmodifikation bis zu 500 erreichen kann. Das heißt, bei einer Irrtumswahrscheinlichkeit von 5%, wäre dann bereits mit 25 „falsch positiven“ Ergebnissen zu rechnen.

Lösungsvorschlag:

Im Rahmen dieser Stellungnahme wird vorgeschlagen, im Abschnitt 9.3.10 das Problem des multiplen Testens, wie es bereits im „alten“ Methodenpapier adressiert war, wieder aufzunehmen.

**Kompetenzbereich Gesundheitsökonomie:
9. Kostendarstellung**

Bezug: Abschnitt 3.3.3 der Allgemeinen Methoden, Entwurf für Version 5.0

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.

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.

Kompetenzbereich Gesundheitsökonomie:**10. Epidemiologie**

Bezug: Abschnitt 3.3.3 der Allgemeinen Methoden, Entwurf für Version 5.0

Problembeschreibung:

In diesem Abschnitt erläutert das IQWiG, dass im Rahmen der Nutzenbewertung eines Arzneimittels gemäß § 35a SGB V der pharmazeutische Unternehmer u. a. Angaben zur Anzahl der Patientinnen und Patienten und Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, machen muss. In Abschnitt 3.3.3 finden sich jedoch keine methodischen Anmerkungen, wie das Institut die Angaben zur Anzahl der Patientinnen und Patienten und Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, überprüft.

Handlungsbedarf:

Die Bewertung der Angaben zur Anzahl der Patientinnen und Patienten und Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, durch das IQWiG weist Inkonsistenzen auf:

- Epidemiologische Darstellungen 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 Epidemiologie“ eingeführt werden. In diesem Abschnitt wäre u. a. zu beschreiben,

- wie das IQWiG die Angaben des pharmazeutischen Unternehmers zur Epidemiologie überprüft,
- in welchem Verhältnis unterschiedliche Quellen stehen,
- welche Bedeutung frühere Beschlüsse in derselben Indikation haben,
- wie Unsicherheiten methodisch ausgewiesen werden.

Ecker + Ecker GmbH

27. Januar 2017

A.1.20 – GlaxoSmithKline GmbH & Co. KG (GSK)

Autoren:

- Hennig, Michael

Einleitung

GSK nimmt im Folgenden Stellung zu folgenden Abschnitten des Entwurfs:

- 2.2.1 Einbindung von Betroffenen
- 2.2.7 Wissenschaftliche Beratung
- 3.1.4 Endpunktbezogene Auswertung
- 8 Informationsbeschaffung
- 9.3.8 Meta-Analysen
- 9.3.9 Indirekte Vergleiche
- 9.3.10 Subgruppenanalysen
- 9.3.13 Auswertung abhängiger Daten
- 9.4 Qualitative Methoden

Abschnitt Thema	Stellungnahme
2.2.1 Tabelle 2	<p>Wir begrüßen dieses neue Kapitel, da die dort beschriebene Einbindung von Betroffenen nun transparenter wird.</p> <p>Bei der in Tabelle 2 beschriebenen Form der Einbindung Betroffener für die jeweiligen Institutsprodukte würden wir gerne besser verstehen, warum für zwei Produkte (Potenzialbewertung, HTA-Bericht) keine Einbindung von Betroffenen vorgesehen ist.</p>
2.2.1 „sonstige Wege“	<p>Bei der folgende Passage:</p> <p><i>„Betroffene können insbesondere sein: Patientinnen und Patienten, ihre Eltern oder sonstige Angehörige sowie potenzielle Teilnehmer an Präventionsmaßnahmen. Bei der Auswahl der Teilnehmenden wird der Fokus auf tatsächlich Betroffene gerichtet, denn Patientenvertreterinnen und Patientenvertreter oder Repräsentantinnen und Repräsentanten von Selbsthilfegruppen sind selbst teilweise keine Patienten und können nicht in vergleichbarer Weise beurteilen, wie Betroffene Symptome empfinden, ihre Funktionen und Aktivitäten wahrnehmen können, oder in ihrer Lebensqualität beeinträchtigt sind. Zur Findung Betroffener wird regelhaft die Patientenvertretung im G-BA gebeten, selbst oder über ihre Mitgliedsorganisationen Betroffene zu einem Thema zu benennen. Ergänzend können vom Institut auch über nationale oder lokale Selbsthilfeorganisationen oder -gruppen, über Krankenhäuser oder Arztpraxen, externe Sachverständige oder über sonstige Wege Betroffene gesucht oder angefragt werden.“</i></p> <p>interessiert uns, über welche „sonstige Wege“ Betroffene konkret gesucht oder angefragt werden. Hier würden wir eine transparente Darstellung des gesamten Einbindungsprozesses sehr begrüßen.</p>

Abschnitt Thema	Stellungnahme
2.2.7 Wissenschaftliche Beratung	<p>Wir begrüßen dieses neue Kapitel, da dadurch ein Beitrag zur Optimierung von Studiendesigns geleistet werden kann.</p> <p>Bei der angedeuteten „<i>angemessenen organisatorischen Trennung von Beratung und Bewertung</i>“ interessiert uns, wie diese konkret innerhalb des IQWiG umgesetzt wird. Gibt es z.B. spezifische SOP's, die diesen Aspekt detailliert beschreiben?</p>
3.1.4 Belege aus einer Studie	<p>Die Kriterien für die Ableitung eines Beleges aus einer Studie wurden nochmals verschärft, indem nun zusätzlich gefordert wird, dass in jedem Studienarm mindestens 1000 Patienten eingeschlossen sein sollen.</p> <p>Diese Verschärfung wird nicht begründet und ist für uns nicht nachvollziehbar. Bei diesem pauschalen Schwellenwert bleibt die spezifische Studiensituation ebenso unberücksichtigt, wie die Tatsache, dass bei selteneren Indikationen (z.B. in der Onkologie) solche Fallzahlen weder ethisch noch medizinisch vertretbar sind.</p> <p>Daher sollte die Verschärfung wieder gestrichen werden</p>
8 Informationsbeschaffung	<p>Wir begrüßen die grundlegende Überarbeitung von Kapitel 8 zur Informationsbeschaffung. Hier ist insbesondere die detaillierte Beschreibung der einzelnen Prozessschritte bei der Suche in bibliografischen Datenbanken und Studienregistern zu nennen. Sowohl die detaillierte Darstellung der Qualitätsstandards einer umfassenden Informationsbeschaffung als auch des Vorgehens bei der Prüfung dieser ermöglichen es den Prüfprozess besser nachvollziehen zu können. Desweiteren begrüßen wir, dass Aspekte aus IQWiG Nutzenbewertungen nun auch in diesem Entwurf des Methodenpapiers enthalten sind.</p>

Abschnitt Thema	Stellungnahme
9.3.8 Meta-Analysen mit zufälligen Effekten	<p>Folgende Passagen wurden in der Version 5.0 ergänzt</p> <p><i>„Wenn das Vorhandensein von Heterogenität nicht ausreichend sicher ausgeschlossen werden kann, so sollte somit ein Modell mit zufälligen Effekten gewählt werden. Zur Durchführung von Meta-Analysen mit zufälligen Effekten existieren zahlreiche Methoden [663]. Gemäß neueren Empfehlungen sollte zur Durchführung von Meta-Analysen mit zufälligen Effekten regelhaft die Knapp-Hartung-Methode [300,389] verwendet werden [357,664].“</i></p> <p><i>„Die Anwendung von Meta-Analysen mit zufälligen Effekten stößt jedoch im Fall sehr weniger Studien (weniger als 5) an ihre Grenzen. Da im Fall sehr weniger Studien die Heterogenität nicht verlässlich schätzbar ist [316], kann die Anwendung von Meta-Analysen mit zufälligen Effekten zu sehr breiten Konfidenzintervallen führen, die möglicherweise keine Aussagen zur Evidenzlage mehr zulassen. Gerade im Fall sehr weniger Studien sollte daher die Anwendung des Modells mit festem Effekt oder eine qualitative Zusammenfassung (siehe Abschnitt 3.1.4) in Erwägung gezogen werden. Kontextabhängig kommen auch alternative Verfahren in Betracht, wie z. B. Bayes'sche Verfahren [32,240,607] oder Methoden aus dem Bereich der generalisierten linearen Modelle [408,514]. Des Weiteren wird das Institut, wie im Folgenden beschrieben, nur dann stark heterogene Studienergebnisse meta-analytisch zusammenfassen, wenn plausible Gründe für die Heterogenität ersichtlich sind, die eine Zusammenfassung trotzdem rechtfertigen.“</i></p> <p>Wir begrüßen die Anpassungen grundsätzlich, da hier neuere methodische Entwicklungen auf diesem Gebiet berücksichtigt werden. An dieser Stelle interessiert uns, wie der Prozess der Anpassung an neue methodische Entwicklungen beim IQWiG grundsätzlich gestaltet und dokumentiert ist. Erfolgt z.B. eine systematische Literaturrecherche? Wie erfolgt die Auswahl der neuen Methoden, die Eingang in die aktualisierte Version des Methodenpapiers enthalten.</p>
9.3.8 Geringe Zahl von Ereignissen	<p>Im ersten Abschnitte dieses Kapitels wurde folgender Satz zur Peto-Odds-Ratio-Methode ergänzt:</p> <p><i>„Die Anwendung dieser Methode ist allerdings nur adäquat, wenn die zu schätzenden Effekte nicht zu groß sind und kein unbalanciertes Design vorliegt [71].“</i></p> <p>Wir begrüßen die Anpassungen grundsätzlich, da hier neuere methodische Entwicklungen auf diesem Gebiet berücksichtigt werden. Auch an dieser Stelle interessiert uns, wie der Prozess der Anpassung an neue methodische Entwicklungen beim IQWiG grundsätzlich gestaltet und dokumentiert ist (s.o.).</p>

Abschnitt Thema	Stellungnahme
9.3.9 Indirekte Vergleiche	<p>Folgende Passage wurden in der Version 5.0 ergänzt bzw. geändert:</p> <p><i>„Dies gilt auch für Methoden für indirekte Vergleiche, in denen über Modellierungen mit starken Annahmen über die unbekanntten Effekte [107] oder mithilfe von Verfahren aus dem Bereich der Kausalmodelle für Beobachtungsstudien mit untestbaren Annahmen [596] versucht wird, Effektschätzungen trotz fehlender Brückenkomparatoren zu ermöglichen. Es werden ausschließlich adjustierte indirekte Vergleiche über adäquate Brückenkomparatoren akzeptiert. Hierzu zählen insbesondere das Verfahren von Bucher et al. [88] sowie die oben erwähnten Methoden der Netzwerk-Meta-Analysen.“</i></p> <p>Grundsätzlich ist nachvollziehbar, dass indirekte Vergleiche über einen Brückenkomparator solchen ohne Brückenkomparator vorzuziehen sind. Dennoch ist es nicht immer möglich, einen indirekten Vergleich über einen Brückenkomparator durchzuführen. Die Verfahren, die einen Vergleich ohne Brückenkomparator über eine Modellierung ermöglichen, werden stetig weiterentwickelt und systematisch auf ihre Eignung untersucht. Daher ist es nicht nachvollziehbar, dass diese vom IQWiG grundsätzlich abgelehnt werden. Vielmehr sollte die Möglichkeit bestehen, dass geeignete Verfahren auch für die Nutzenbewertung herangezogen werden können, sollte ein Vergleich über einen Brückenkomparator nicht möglich sein.</p>
9.3.10 Stellenwert von Subgruppenanalysen	<p>Während in der Version 4.2 Subgruppenanalysen wie folgt vom IQWiG charakterisiert werden:</p> <p><i>„Subgruppenanalysen werden in der methodischen Literatur <u>sehr kritisch</u> diskutiert“</i></p> <p>heißt es nun:</p> <p><i>„Auch wenn Subgruppenanalysen für die Optimierung der Behandlung sinnvoll sind, werden sie in der methodischen Literatur <u>teilweise kritisch</u> diskutiert“.</i></p> <p>Es ist für uns nicht nachvollziehbar, wie das IQWiG zu dieser Neueinschätzung kommt. Aus unserer Sicht werden Subgruppenanalysen auch weiterhin sehr kritisch gesehen (s. z.B. Sun, BMJ, 2012, „Credibility of claims of subgroup effects in randomised controlled trials“).</p> <p>Daher bitten wir das IQWiG, bei der ursprünglichen Formulierung zu bleiben.</p>
9.3.10 Multiples Testen	<p>Der berechtigte Hinweis auf die multiple Testthematik in der Version 4.2 (<i>„Bei der Analyse zahlreicher Subgruppen besteht eine manchmal recht hohe Wahrscheinlichkeit, dass die Ergebnisse irgendeiner Subgruppe statistische Signifikanz erreichen, obwohl es sich in Wahrheit um ein zufälliges Ergebnis handelt“</i>) taucht in der Version 5.0 nicht mehr auf.</p> <p>Da sich an dieser Problematik nichts verändert hat, bitten wir das IQWiG, diese Löschung wieder rückgängig zu machen.</p>

Abschnitt Thema	Stellungnahme
9.3.10 Vergleichbarkeit zwischen Behandlungsgruppen innerhalb der Subgruppen	<p>Folgende Passage wurde in der Version 5.0 ergänzt: <i>„Vergleichbarkeit zwischen Behandlungsgruppen innerhalb der Subgruppen: Erfolgte die Randomisierung nicht nach dem Subgruppenmerkmal stratifiziert, so könnten sich die Behandlungsgruppen bei kleinen Fallzahlen innerhalb der Subgruppen in prognostischen Faktoren unterscheiden [133,627]. In diesem Fall ist die Vergleichbarkeit der Behandlungsgruppen innerhalb der Subgruppen gefährdet, sodass (nicht) vorhandene Unterschiede zwischen Subgruppen allein durch diese Imbalance bedingt sein können“.</i></p> <p>Wir begrüßen die Ergänzung dieses Punktes, geben allerdings zu bedenken, dass eine stratifizierte Randomisierung nicht für alle für die Nutzenbewertung herangezogenen Subgruppen logistisch durchführbar ist.</p> <p>Wir bitten das IQWiG darum, diesen praktischen Aspekt mit aufzunehmen.</p>
9.3.10 Effektmodifikation durch mehr als ein Subgruppenmerkmal	<p>Folgende Passage wurde in der Version 5.0 ergänzt: <i>„Effektmodifikation durch mehr als ein Subgruppenmerkmal (Wechselwirkung höherer Ordnung): Liegt für einen Endpunkt beispielsweise ein Unterschied sowohl zwischen 2 Altersgruppen als auch zwischen Männern und Frauen vor, so ist zur Interpretation eine Auswertung separat für jede Altersgruppe getrennt für Männer und Frauen nötig (also eine Analyse zu 4 Subgruppen). Eine solche Auswertung liegt jedoch selten vor.“.</i></p> <p>Wir stimmen zu, dass Wechselwirkungen höherer Ordnung theoretisch möglich sind. Allerdings sollte hierbei auch die Limitationen einer vollständigen Umsetzung aller denkbaren Wechselwirkungen höherer Ordnung vor dem Hintergrund der kleiner werdenden Fallzahlen für die Sub-Subgruppen berücksichtigt werden.</p> <p>Wir bitten das IQWiG darum, diese praktischen Aspekte ebenfalls aufzunehmen.</p>

Abschnitt Thema	Stellungnahme
9.3.10 Post-hoc-Charakter von Subgruppenanalysen	<p>Folgende Passage wurde in der Version 5.0 ergänzt: <i>„Während im Allgemeinen post hoc durchgeführte Subgruppenanalysen auf Studienebene kritisch zu interpretieren sind (auch in Anbetracht der oben genannten methodischen Limitationen), ist man in einer systematischen Übersicht dennoch auf die Verwendung der Ergebnisse solcher Analysen auf Studienebene angewiesen, wenn im Rahmen der systematischen Übersicht genau diese Subgruppen untersucht werden sollen. Solche Subgruppenanalysen sind dann im Sinne der systematischen Übersicht auch nicht als „post hoc“ zu bezeichnen, sondern entsprechen einer in dieser Übersicht zu überprüfenden Hypothese. Insofern ist die Analyse der Heterogenität zwischen den einzelnen Studien und damit gegebenenfalls Subgruppenanalysen eine wissenschaftliche Notwendigkeit.“</i></p> <p>Ob es sich bei einer Subgruppenanalyse gegebenenfalls um eine wissenschaftliche Notwendigkeit handelt, hängt von vielen Faktoren, wie z.B. der biologischen Rationale ab. Dieser wichtige Aspekt kommt bei der o.g. Beschreibung leider zu kurz.</p> <p>Wir bitten das IQWiG darum, die Kriterien für eine wissenschaftliche Notwendigkeit ausgewogen zu diskutieren.</p>

Abschnitt Thema	Stellungnahme
9.3.10 Verlässlichkeit von Subgruppenanalysen	<p>Folgende Passage wurde in der Version 5.0 ergänzt: <i>„Einerseits erfordern die oben aufgeführten Aspekte die Bewertung der Verlässlichkeit von Subgruppenanalysen; Sun et al. [627] identifizieren hierzu Kriterien. So sind beispielsweise Subgruppenanalysen zu Merkmalen, die nicht vor der Randomisierung, sondern im Studienverlauf erhoben wurden (beispielsweise Patienten mit vs. solche ohne Herzinfarkt unter der untersuchten Therapie), grundsätzlich nicht verlässlich. Andererseits können trotz dieser Einschränkungen Subgruppenanalysen für manche Fragestellungen die auf absehbare Zeit beste wissenschaftliche Evidenz für die Bewertung von Effekten in ebendiesen Subgruppen darstellen [237], da beispielsweise ethische Überlegungen gegen die Verifizierung beobachteter Effekte in weiteren Studien sprechen können“.</i></p> <p>Wir begrüßen die Diskussion zur Verlässlichkeit von Subgruppenanalysen auf der Basis der Arbeit von Sun (BMJ 2010). In dieser Arbeit und in der Arbeit des gleichen Autors von 2012 (BMJ, “Credibility of claims of subgroup effects in randomised controlled trials“) werden wichtige Kriterien zur Beurteilung der Glaubwürdigkeit von Subgruppenanalysen genannt, u.a. auch die o.g. Erhebung von Merkmalen vor der Randomisierung. Allerdings werden folgende dort genannte weitere wichtige Aspekte nicht im Methodenpapier 5.0 berücksichtigt:</p> <ul style="list-style-type: none"> - <i>„Was there any indirect evidence to support the apparent subgroup effect-for example, biological rationale, laboratory tests, animal tests?“</i> - <i>„Was the subgroup analysis one of a small number of subgroup hypotheses tested (<=5)?“</i> - <i>„Was the subgroup effect consistent with evidence from previous related studies?“</i> - <i>“Was the subgroup effect consistent across related outcomes?“</i> <p>Wir bitten das IQWiG, diese wichtigen Aspekte ebenfalls zu berücksichtigen und entsprechend darzustellen.</p>

Abschnitt Thema	Stellungnahme
9.3.10 Mindestzahl von Ereignissen für Subgruppenanalysen bei binären Daten	<p>Folgende Passage wurde in der Version 5.0 ergänzt: <i>„Stattdessen muss zunächst der statistische Nachweis unterschiedlicher Effekte zwischen verschiedenen Subgruppen anhand eines adäquaten Homogenitäts- beziehungsweise Interaktionstests vorgenommen werden. Damit überhaupt eine gewisse Wahrscheinlichkeit für einen solchen Nachweis vorliegt, werden Subgruppenanalysen bei binären Daten regelhaft nur dann durchgeführt, falls mindestens 10 Ereignisse in einer der Subgruppen aufgetreten sind. Mit einem Ereignis ist hier gemeint, dass dieses im Verlauf einer Beobachtung auftritt und zu Baseline noch nicht eingetreten ist (z. B. Erlangen einer Virusfreiheit bei infizierten Personen)“.</i></p> <p>Wir begrüßen die Berücksichtigung dieses Kriteriums, welches die praktische Umsetzbarkeit von Subgruppenanalysen adressiert.</p> <p>Hier würden wir gerne besser verstehen, auf welcher Basis der Schwellenwert „10“ vom IQWiG abgeleitet wurde.</p>
9.3.10 Schwelle für p-Wert des Interaktionstests	<p>Während in der Version 4.2 noch folgende Regelung vom IQWiG getroffen wurde:</p> <p><i>„Ein zum Niveau $\alpha = 0,05$ signifikantes Ergebnis wird als Beleg unterschiedlicher Effekte, ein zum Niveau $\alpha = 0,20$ signifikantes Ergebnis wird als Hinweis auf unterschiedliche Effekte in den Gruppen gewertet.“</i></p> <p>wird in der Version 5.0 nunmehr nur noch der Schwellenwerte von $\alpha = 0,05$ benutzt.</p> <p>Wir begrüßen die neue Regelung, da diese die multiple Testthematik besser adressiert.</p>
9.3.13 Auswertung abhängiger Daten	<p>Wir begrüßen dieses neue Kapitel, da der Umgang mit abhängigen Daten relevant ist und sich von den Standardverfahren, basierend auf unabhängigen Beobachtungen, unterscheidet.</p>
9.4 Qualitative Methoden	<p>Wir begrüßen die Überarbeitung und Aktualisierung dieses Kapitels.</p>

A.1.21 – Institut für Gesundheits- und Pharmaökonomie (IfGPh)

Autoren:

- Menzler, Jacob
- Neubauer, Aljoscha

Stellungnahme des
Instituts für Gesundheits- und Pharmaökonomie

zum

Entwurf 5.0 „Allgemeine Methoden“

des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG)

In der Version 5.0 seiner Allgemeinen Methoden stellt das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) Arbeitsweisen, Kriterien und Methoden dar. Das Institut für Gesundheits- und Pharmaökonomie (IfGPh) möchte sich mit folgenden Anmerkungen aktiv an der Weiterentwicklung dieser Methodik beteiligen:

(1) Anwendbarkeit von Leitlinien

„Werden für Leitliniensynopsen neben nationalen auch internationale Leitlinien recherchiert, wird grundsätzlich davon ausgegangen, dass Leitlinien aus Mitgliedsstaaten der Organisation for Economic Co-operation and Development (OECD) von besonderer Relevanz sind.“ (IQWiG Methodenpapier 5.0, Abschnitt 5.2.2, S. 116)

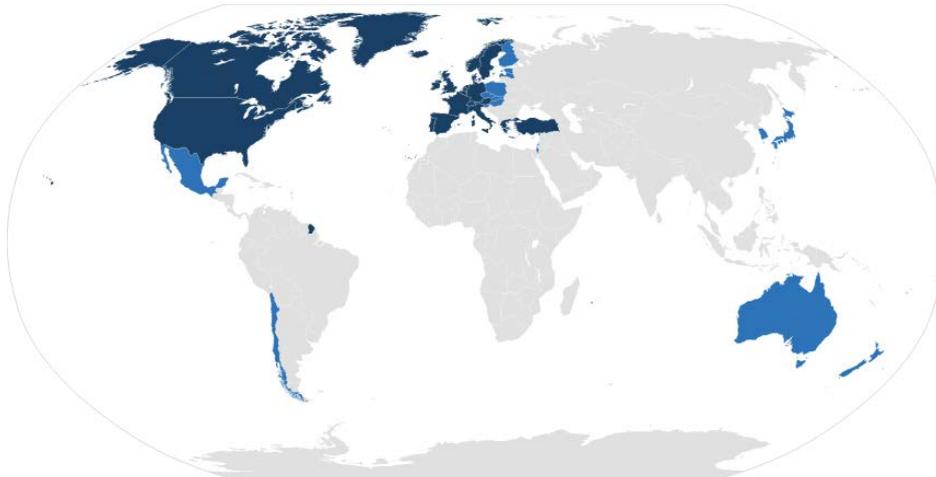


Abb.1 : Geographische Verteilung OECD Mitglieder

Gründungsmitglieder (dunkelblau) und neue Mitglieder (helles Blau)

Quelle:

https://upload.wikimedia.org/wikipedia/commons/4/4c/OECD_member_states_map.svg,

Link überprüft am 25.01.2017

Die besondere Relevanz von OECD-Staaten für die Anwendbarkeit von Leitlinien ist für uns nicht ersichtlich. Zwar sind die Bemühungen der OECD um hochwertige und einheitliche Standards im Gesundheitswesen sehr wichtig, doch darf die Verschiedenheit der OECD-Mitgliedsländer dabei nicht außer Acht gelassen werden.

Die 35 Mitgliedsstaaten (Abbildung 1) verteilen sich über 5 Kontinente und zeigen erhebliche Unterschiede in ihren Gesundheits- und Sozialsystemen, in der Zulassungspraxis für Medikamente und Medizinprodukte, sowie in ihrer demographischen und wirtschaftlichen Entwicklung. Pseudo-quantitative Kriterien wie z.B. die vom IQWiG auch aktuell verwendeten „WHO-Strata“ können diese Problematik nicht auflösen (https://www.iqwig.de/download/V14-02_Abschlussbericht_Version1-1_Leitlinienrecherche-und-bewertung-fuer-ein-DMP-Rheumatoide-Arthritis.pdf; Link überprüft am 25.01.2017), da sie nicht die genannten relevanten Kriterien sondern nur grobe ökonomische und Mortalitäts-Kriterien abbilden.

Vorschlag zur Änderung

Eine Hierarchie der Leitlinien hinsichtlich des Kriteriums „Übertragbarkeit“ wird vorgeschlagen. Diese ist nicht strikt quantitativ, muss aber die Anwendbarkeit für das deutsche Gesundheitssystem sicherstellen:

- Primär sind aus unserer Sicht in Deutschland erstellte, an höchsten wissenschaftlichen und medizinischen Standards ausgerichtete Leitlinien (S3) zu bevorzugen, da sie an die hier herrschenden praktischen Gegebenheiten angepasst sind. Sie berücksichtigen insbesondere die Verfügbarkeit von Therapiemöglichkeiten und die geübte Behandlungspraxis, sowie die relevanten rechtlichen und verwaltungstechnischen Bestimmungen.
- Nachrangig sollten in nächster Stufe europäische Leitlinien berücksichtigt werden. Aufgrund der meist europäischen Zulassungsprozessen für Arzneimittel und Medizinprodukte besteht in der Regel Anwendbarkeit für Deutschland.
- Bestehen auf nationaler und europäischer Ebene keine geeigneten Leitlinienempfehlungen, so kann auch internationale Leitlinien zurückgegriffen werden. Bei deren Auswahl sollte kritisch auf die Anwendbarkeit auf deutsche bzw. europäische Verhältnisse geachtet werden, beispielsweise hinsichtlich Zulassungsstatus und medizinischer Praxis. Hier werden daher in der Regel stärker „europäisch geprägte“ Länder wie USA, Kanada oder Australien häufiger herangezogen werden können.

Das Kriterium „Übertragbarkeit“ von Leitlinien ist unabhängig von deren methodischen Qualität zu sehen. Analog zur Übertragbarkeit von klinischen Studien ist für nicht-deutsche Leitlinien deren Übertragbarkeit zu prüfen, wobei die vorgeschlagene Hierarchie eine erste Voreinteilung erlaubt.

(2) Beleg mit nur einer Studie

„Aus lediglich 1 Studie kann im Ausnahmefall für einen Endpunkt ein Beleg für eine spezifische (Teil-)Population abgeleitet werden, wenn ein Studienbericht gemäß 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 mindestens 1000 Patienten eingeschlossen und es gibt mindestens 10 Zentren.*
- *Der beobachtete Effektschätzer hat einen sehr kleinen zugehörigen p-Wert ($p < 0,001$).“* (IQWiG Methodenpapier 5.0, Abschnitt 3.1.4, S.51)

Wir verstehen das Bemühen des IQWiG durch eindeutige Grenzen Transparenz und klare Bewertungskriterien zu schaffen, die die Durchführung von Bewertungen vereinfachen. Gleichzeitig sollte eine Prüfung auch nach nachvollziehbaren, begründeten Kriterien durchgeführt werden und nicht nach zwar a priori festgelegten, aber unbegründeten Kriterien. Wir können die genannten Zahlen inhaltlich nicht nachvollziehen: neu in den Methoden 5.0 ist nun die Forderung nach mindestens 1000 Patienten. Neben der Frage der Begründung der Grenze von 1000 stellt sich für uns beispielsweise die Problematik der Frage nach der Verteilung von 1000 Patienten auf die geforderten mindestens 10 Zentren: 1000 Patienten mit 991 in Zentrum 1 und jeweils einem Patienten in Zentrum 2 bis 10 erfüllt das Kriterium, während 999 Patienten gleichverteilt über 9 Zentren das Kriterium nicht erfüllt.

Vorschlag zur Änderung

Wir schlagen vor, die genannten 1000 Patienten an 10 Zentren jeweils lediglich als Richtgröße mit beispielsweise 10-20% Toleranz anzugeben. Damit kann eine für die jeweiligen Spezifika einer Studie angemessene Entscheidung getroffen werden.

(3) Medizinprodukte

„Wenngleich die regulatorischen Voraussetzungen für den Marktzugang von Arzneimitteln und nichtmedikamentösen therapeutischen Interventionen unterschiedlich sind, gibt es dennoch keinen Grund, an die Bewertung von Nutzen und Schaden einen bezüglich der Ergebnissicherheit prinzipiell anderen Maßstab anzulegen.“ (IQWiG Methodenpapier 5.0, Abschnitt 3.4, S.66).

Grundsätzlich stimmen wir überein, dass auch die Nutzenbewertung von nicht-medikamentösen Verfahren nach prädefinierten und nachvollziehbaren Kriterien erfolgen muss, um qualitativ hochwertige Produkte sicherzustellen. Allerdings bestehen einige Besonderheiten bei Medizinprodukten:

- 1) Arzneimittel werden in der für den Endverbraucher gültigen Form getestet und -bei erfolgter Zulassung- so auch für viele Jahre benutzt. Medizinprodukte hingegen, auch im Hochrisikosektor, sind demgegenüber derzeit viel schnelleren Innovationszyklen unterworfen. Dabei kann auch das eigentliche Wirkprinzip selbst durch technische Änderungen verbessert werden beispielsweise um mehr Spezifität oder Wirksamkeit

zu erhalten. Oder die technische Umsetzung des Gerätes wird aufgrund technischer Fortschritte oder Regulierungen im Bereich der Elektrotechnik, Software oder Material geändert. Beispielsweise haben Herzschrittmacher seit ihrer Erfindung kaum Änderungen auf Grund eines neuen „Wirkprinzips“ erfahren und sich doch auf Grund besserer Elektroden, Sensoren, Batterien, Software und gewebeverträglichen Materialien sehr stark verändert.

Ein weiteres Beispiel dafür ist das Cochlea Implantat (Hainarosie et al., 2014 [1]), dessen aktuelle Erscheinungsweise und Funktionalität stark von den Entwicklungen der Elektrotechnik abhängt und das von den Fortschritten der Mikrotechnologie überhaupt erst ermöglicht wurde (Abbildung 2). Der technische Fortschritt wirkt sich daher direkt auf das Design des Cochlear Implants aus, während das Wirkprinzip unverändert bleibt.



Abb. 2: Die Entwicklung des Cochlea-Implantats über die Zeit

Die Bilder sind nicht maßstabsgetreu.

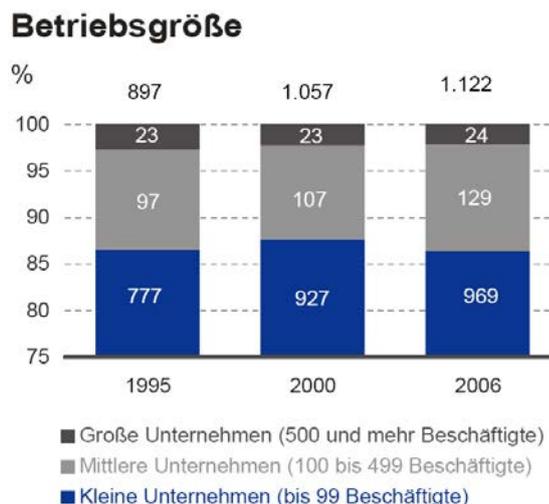
Quelle: <https://twitter.com/CochlearAsiaPac>; Link überprüft 25.01.2017

Design und Funktionsumfang eines Medizinproduktes hängen dabei nicht nur von der Wirkung auf den Patienten, sondern auch von der Funktionalität und dem Bedienkomfort für das medizinische Personal ab. Dass Kategorien wie Funktionalität, Kompatibilität, Ergonomie und Effizienz nicht auf Arzneimittel anwendbar sind, sollte bei einer pauschalen Angleichung der Prüfungsanforderungen für beide Produktkategorien ebenfalls in Betracht gezogen werden.

Bei diagnostischen Tests schlägt das IQWiG bereits vor, lediglich Äquivalenzstudien anzufordern, wenn ein neuer Test weniger invasiv und strahlungsärmer ist, als der bisherige Standardtest (IQWiG Methodenpapier 5.0, Abschnitt 3.5, S.70). Eine solche Regelung ist auch für Medizinprodukte richtig und wünschenswert, wenn diese statt neue Wirkprinzipien zu enthalten, eine verbesserte Umsetzung bekannter Verfahren darstellen.

2) Prinzipiell scheint uns der fundamentale Unterschied im Zulassungsverfahren von Medizinprodukten und Arzneimitteln darauf begründet zu sein, dass die Wirkung, Funktion und der Nutzen eines Medizinproduktes leichter einzusehen sind -und sich eher physikalisch/mathematisch ableiten lassen-, als die komplexer in Organismen wirkenden Arzneimitteln (IQWiG Methodenpapier 5.0: „Medizinprodukte wird abschwächend im Rahmen der Konformitätsbewertung in der aktuellen DIN EN ISO-Norm 14155 (Abschnitt A.6.1 [161]) zwar nicht die randomisierte kontrollierte Studie als Design der Wahl dargestellt, allerdings muss die Wahl des Designs überhaupt begründet werden“). Das Konzept der verblindeten randomisierten, kontrollierten Studie (RCT) ist natürlich meistens dennoch auf Medizinprodukte anwendbar. Das IQWiG weist allerdings auch selbst darauf hin, dass die randomisierte, verblindete Untersuchung von Medizinprodukten an praktische Grenzen stößt. Eine umfangreiche RCT-Testung von neuen Prinzipien ist in Abwägung von Zugewinn an Sicherheit versus Aufwand nicht immer sinnvoll, da hier häufiger als bei Arzneimitteln der „Evidenz-Sonderfall“ anzuwendender Naturgesetze gilt.

3) Aufwand für die Hersteller



Quelle: Statistisches Bundesamt

Abb. 3: Betriebsgrößen in der deutschen Medizintechnikindustrie

Nur 2% der Betriebe in Deutschland haben mehr als 500 Mitarbeiter.

<http://www.medizintechnik-sachsen-anhalt.de/en/downloadengl/category/3-berichte?download=29:lage-und-perspektiven-der-deutschen-medizintechnik&start=20>; Link überprüft am 25.01.17

Neben den inhaltlichen Problemen wollen wir auch auf die Kosten und den organisatorischen Aufwand hinweisen, der auf die Hersteller oder Erfinder zukommt. Medizinprodukte werden derzeit noch überwiegend von kleinen und mittleren Unternehmen entwickelt und hergestellt. So beschäftigen 95 Prozent der MedTech-Unternehmen weniger als 250 Mitarbeiter (<https://www.bvmed.de/download/bvmed-branchenbericht-medtech>; Link überprüft am 25.01.17; Link überprüft am 25.01.17). Viele Ideen und Prototypen gehen auf klinische Anwender zurück, die diese Produkte in Zusammenarbeit mit Partnern aus

der Medizintechnikbranche entwickeln. Kleine Budgets und hohe Eigenleistung sind dabei keine Seltenheit.

Umfangreiche klinische, randomisierte Studien (z.B. 1000 Teilnehmer an 10 Studienzentren) stellen – selbst bei teilweiser externer Förderung – für kleine und mittlere Unternehmen eine unüberwindliche finanzielle Hürde dar. Die Erzeugnisse der Medizintechnikbranche unterliegen zudem sehr kurzen Produktzyklen von etwa 3 Jahren. Bei verpflichtender Bewertung jedes neuen Produktes würden so alle 3 Jahre immense Kosten für Studien entstehen. Diese kurzen Innovationszyklen und auch der häufig begrenzte Patentschutz sind typisch für Medizinprodukte. Die Hersteller bemühen sich daher um rasche Markteinführung und -durchdringung. Mehrjährige klinische Studien sind innerhalb dieser Innovationszyklen nicht realistisch umsetzbar. Durch erhöhte Anforderungen ist eine stärkere (horizontale) Integration auf dem Markt für Medizinprodukte zu erwarten, was in der Folge zu höheren Kosten für Patienten und Krankenkassen und zu weniger Innovation und Wettbewerb führen würde.

Vorschlag zur Änderung:

Das IQWiG sollte auf neue Wirkprinzipien fokussieren, um Detailänderungen und Weiterentwicklungen weiterhin zu ermöglichen. Es sollte auch nicht vergessen werden, dass Ziel und Wirkung des Gerätes vielfach vom Benutzer abhängen (z.B. das invasive Skalpell wird wesentlich vom Chirurgen abhängen), so dass dieser Einfluss des Benutzerumfeldes jeweils geeignet zu berücksichtigen ist.

Die Forderung nach RCTs sollte dabei fallspezifisch erwogen werden, auch vor dem Hintergrund der technisch-mathematischen Berechenbarkeit bzw. Umfangs der biologischen Interaktion. Bestehende oder im Rahmen der Nutzung entstehende Evidenz wie beispielsweise Registerdaten oder Anwenderdaten müssen als Evidenz zulässig bleiben. In vielen Bereichen halten wir –ähnlich wie in der Diagnostik- Äquivalenzstudien für ausreichend. Es sollte keine grundsätzliche RCT-Forderung für Medizinprodukte bestehen, nur weil diese methodisch möglich sind.

München, den 30.01.2017

Prof. Dr. Aljoscha Neubauer

Dr. Jacob Menzler

IfGPh Institut für Gesundheits- und Pharmakoökonomie
Frau-Holle-Str. 9/9A; D-81739 München
Tel. +49-(0)89 – 85 639 8 -41
Fax: +49-(0)89 – 85 639 8 -40
Email:

Literatur:

- [1] HAINAROSIE, M., ZAINEA, V. & HAINAROSIE, R. 2014. The evolution of cochlear implant technology and its clinical relevance. *J Med Life*, 7 Spec No. 2, 1-4.

A.1.22 – Jansen-Cilag GmbH

Autoren:

- Eisele, Lewin
- Erhardt, Wilma
- Leisten, Marie-Kristin

Stellungnahme Janssen-Cilag GmbH

IQWiG Allgemeinen Methoden – Entwurf der Version 5.0

IQWiG Methodenpapier 5.0 Kommentare zu den Änderungen bezüglich: „Ergänzung in Abschnitt 3.1.4 zu Belegen aus einer Studie“

Folgende Passage wurde in Version 5.0 ergänzt:

„Aus lediglich 1 Studie kann im Ausnahmefall für einen Endpunkt ein Beleg für eine spezifische (Teil-) Population abgeleitet werden, wenn ein Studienbericht gemäß 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 mindestens 1000 Patienten eingeschlossen und es gibt mindestens 10 Zentren.“*

Kommentar:

Die Einschätzung, ob eine einzelne Studie zur Ableitung eines Belegs ausreichend ist, sollte eine Überprüfung und Entscheidung für jeden Einzelfall beinhalten und nicht an einer fixen Grenze für die Fallzahl festgemacht werden. Auch Studien hoher Qualität mit geringen Fallzahlen können zur Ableitung eines Belegs geeignet sein. Ebenso können Studien mit hohen Fallzahlen und geringer Qualität nicht geeignet sein. Deshalb sollten über die Fallzahl hinaus die etablierten und üblichen Kriterien zur Beurteilung der Studienqualität, insbesondere das Verzerrungspotenzial und das Studiendesign, zugrunde gelegt werden.

IQWiG Methodenpapier 5.0 Kommentare zu den Änderungen bezüglich: „2.2.1 Einbindung von Betroffenen in die Erstellung von IQWiG-Produkten“

Folgende Passage wurde in Version 5.0 ergänzt:

„Die Einbindung von Betroffenen im Rahmen der Erstellung von systematischen Übersichten und HTA-Berichten ist mittlerweile etablierter internationaler Standard der Nutzenbewertung [128,215,401]. Am IQWiG findet die Einbindung Betroffener primär zu Beginn der Berichtsbearbeitung im Rahmen der Festlegung patientenrelevanter Endpunkte und relevanter Subgruppen statt. Darüber hinaus haben Betroffene im Rahmen der Anhörungsverfahren die Möglichkeit, sich in die Berichterstellung einzubringen. Im Rahmen der Erstellung von Gesundheitsinformationen können Selbsthilfeorganisationen zum Informationsbedarf von Betroffenen und zu den Herausforderungen bei der Bewältigung der Erkrankung befragt werden. Die Entwürfe der Gesundheitsinformationen werden regelhaft einer externen Nutzertesting unterzogen. Darüber hinaus werden themenspezifisch Erfahrungsberichte von Betroffenen erhoben.“

Kommentar:

Zu den Institutsprodukten und Formen der Einbindung Betroffener gehören neben der Berichtsbearbeitung auch die Nutzenbewertung nach § 35a SGB V sowie die Kosten-Nutzen-Bewertung nach § 35b SGB V in Form schriftlicher Konsultation. Auch die PatBeteiligungsV gibt in § 4 „Verfahren der Beteiligung“ Abs. 3 vor: „Die Beteiligung nach § 140f Abs. 4 SGB V muss frühzeitig erfolgen. Dazu werden den in § 2 Abs. 1 genannten und den nach § 3 anerkannten Organisationen die erforderlichen Unterlagen rechtzeitig und vollständig zur Verfügung gestellt und ihnen eine angemessene Frist zur Stellungnahme eingeräumt“. Aus Transparenzgründen sollte eindeutig beschrieben werden, dass und wie regelhaft diese Form der Patienteneinbindung erfolgt.

Folgende Passage wurde in Version 5.0 ergänzt:

„Zur Einbindung von Betroffenen existieren 2 verschiedene Wege: Zum einen kann ein persönliches Konsultationsgespräch geführt werden, bei dem Betroffene und Institutsmitarbeiter in kleinem Kreis über eine Erkrankung sprechen. Zum Zweiten können Sichtweise und Erfahrungen von Betroffenen schriftlich erfragt werden. In beiden Fällen werden potenzielle Interessenkonflikte der Betroffenen erfasst (analog zur Vorgehensweise wie in Abschnitt 2.2.2 beschrieben). Betroffene mit Interessenkonflikten werden jedoch nicht von der Konsultation ausgeschlossen. Die Namen der Betroffenen, die am Konsultationsgespräch teilgenommen oder den Fragebogen zur Konsultation ausgefüllt haben, werden grundsätzlich nicht veröffentlicht, es sei denn, sie willigen explizit in die Namensnennung ein. In welcher Form Betroffene in die Arbeit des Instituts eingebunden werden können, hängt vor allem davon ab, wie viel Zeit für die Berichterstellung zur Verfügung steht und ob das Berichtsthema vertraulich ist. Ferner ist eine (erneute) Einbindung von Betroffenen verzichtbar, wenn ein Berichtsthema bereits im Rahmen einer früheren Konsultation bearbeitet wurde.“

Kommentar:

Die Einbindung von Betroffenen wird von Janssen-Cilag grundsätzlich begrüßt. Gemäß § 140f SGB V erhalten Patienten bzw. deren maßgebliche Organisationen ein Mitberatungsrecht im Gemeinsamen Bundesausschuss. Damit soll die Patientensicht und das Patienteninteresse in der Richtlinienkompetenz des Gemeinsamen Bundesausschuss nach § 92 SGB V ausreichend abgebildet werden. Die im Abschnitt 2.2.1. getroffenen Bestimmungen zur Einbindung von Betroffenen basieren hingegen auf „Kann“-Regelungen. Aus Sicht von Janssen-Cilag wird mit dieser Regelung keine systematische und regelhafte Patienteneinbindung in die Arbeit des IQWiG gewährleistet. Hier bedarf es einer verbindlichen Regelung.

Darüber hinaus sollten aus Transparenzgründen die Ergebnisse der Konsultationsgespräche oder der Fragebögen regelhaft (ggf. anonymisiert) veröffentlicht und die Art der Berücksichtigung der Ergebnisse in der Berichterstellung genannt werden. Dabei sollten bestimmte Typen der Berichterstellung nicht kategorisch ausgeschlossen werden.

Seite 2

IQWiG Methodenpapier 5.0
Kommentare zu den Änderungen bezüglich:
„Aktualisierung des Abschnitts 9.3.9 zu indirekten Vergleichen“

Folgende Passagen wurden in Version 5.0 ergänzt bzw. geändert:

„Dies gilt auch für Methoden für indirekte Vergleiche, in denen über Modellierungen mit starken Annahmen über die unbekanntene Effekte [107] oder mithilfe von Verfahren aus dem Bereich der Kausalmodelle für Beobachtungsstudien mit untestbaren Annahmen [596] versucht wird, Effektschätzungen trotz fehlender Brückenkomparatoren zu ermöglichen. Es werden ausschließlich adjustierte indirekte Vergleiche über adäquate Brückenkomparatoren akzeptiert. Hierzu zählen insbesondere das Verfahren von Bucher et al. [88] sowie die oben erwähnten Methoden der Netzwerk-Meta-Analysen.“

Kommentar:

Grundsätzlich ist nachvollziehbar, dass indirekte Vergleiche über einen Brückenkomparator solchen ohne Brückenkomparator vorzuziehen sind. Dennoch ist es nicht immer möglich, einen indirekten Vergleich über einen Brückenkomparator durchzuführen. Die Verfahren, die einen Vergleich ohne Brückenkomparator über eine Modellierung ermöglichen, werden stetig weiterentwickelt und systematisch auf ihre Eignung untersucht. Daher ist es nicht nachvollziehbar, dass diese kategorisch abgelehnt werden. Vielmehr sollte die Möglichkeit bestehen, dass solche Verfahren bei einer dokumentierten Eignung auch für die Nutzenbewertung herangezogen werden können, sollte ein Vergleich über einen Brückenkomparator nicht möglich sein.

IQWiG Methodenpapier 5.0
Kommentare zu den Änderungen bezüglich:
„Zusammenführung und Aktualisierung der vorherigen Abschnitte zu Subgruppenanalysen zum neuen Abschnitt 9.3.10“

Folgende Passage wurde in Version 5.0 gelöscht:

„Bei der Analyse zahlreicher Subgruppen besteht eine manchmal recht hohe Wahrscheinlichkeit, dass die Ergebnisse irgendeiner Subgruppe statistische Signifikanz erreichen, obwohl es sich in Wahrheit um ein zufälliges Ergebnis handelt.“

Kommentar:

Grundsätzlich besteht beim multiplen Testen die Gefahr falsch-positiver Ergebnisse. Hierbei handelt es sich um eine Thematik, die für Subgruppenanalysen relevant ist, aber auch unabhängig davon Bedeutung hat. Da sich an der Thematik zum multiplen Testen nichts verändert hat, bitten wir das IQWiG, diese Löschung wieder rückgängig zu machen. Darüber hinaus sehen wir die Notwendigkeit diese Thematik in einem allgemeinen Kontext im Methodenpapier adäquat zu berücksichtigen.

Folgende Passage wurde in Version 5.0 ergänzt:

„Effektmodifikation durch mehr als ein Subgruppenmerkmal (Wechselwirkung höherer Ordnung): Liegt für einen Endpunkt beispielsweise ein Unterschied sowohl zwischen 2 Altersgruppen als auch zwischen Männern und Frauen vor, so ist zur Interpretation eine Auswertung separat für jede Altersgruppe getrennt für Männer und Frauen nötig (also eine Analyse zu 4 Subgruppen). Eine solche Auswertung liegt jedoch selten vor.

Kommentar:

Grundsätzlich sind Subgruppenanalysen mit hoher Unsicherheit behaftet. Durch die Berücksichtigung einer Wechselwirkung höherer Ordnung wird diese Unsicherheit noch weiter erhöht. Von daher ist es nicht nachvollziehbar, dass solche Auswertungen zusätzlich gefordert werden. Durch die erhöhte Unsicherheit und schwierige Interpretierbarkeit sehen wir den zusätzlichen Erkenntnisgewinn als äußerst limitiert an. Daher bitten wir das IQWiG, diese Änderung rückgängig zu machen.

Folgende Passage wurde in Version 5.0 ergänzt:

„Vergleichbarkeit zwischen Behandlungsgruppen innerhalb der Subgruppen: Erfolgte die Randomisierung nicht nach dem Subgruppenmerkmal stratifiziert, so könnten sich die Behandlungsgruppen bei kleinen Fallzahlen innerhalb der Subgruppen in prognostischen Faktoren unterscheiden [133,627]. In diesem Fall ist die Vergleichbarkeit der Behandlungsgruppen innerhalb der Subgruppen gefährdet, sodass (nicht) vorhandene Unterschiede zwischen Subgruppen allein durch diese Imbalance bedingt sein können.“

Kommentar:

Wir begrüßen die Ergänzung dieses Punktes, geben allerdings zu bedenken, dass eine stratifizierte Randomisierung nicht für alle für die Nutzenbewertung herangezogenen Subgruppen logistisch durchführbar ist. Gerade eine Verwendung von zu vielen Stratifizierungsmerkmalen erhöht das Risiko von Imbalancen innerhalb der einzelnen Merkmalskombinationen. Darüber hinaus stellt eine Stratifizierung nach einem Subgruppenmerkmal nicht notwendigerweise eine Gleichverteilung anderer prognostischer Faktoren sicher. Zudem könnte durch die Berücksichtigung von zu vielen Stratifizierungsmerkmalen die Randomisierung der Gesamtstudie gefährdet werden. Darüber hinaus können Subgruppen auch regulatorisch (z.B. EMA/FDA) gefordert werden, so dass eine a-priori Stratifizierung nicht immer möglich ist. Wir bitten das IQWiG darum, diesen praktischen Aspekt mit aufzunehmen.

A.1.23 – Kassenzahnärztliche Bundesvereinigung K. d. ö. R. (KZBV)

Autoren:

- Beck, Jörg

KZBV · BEHRENSTRASSE 42 · 10117 BERLIN

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
Stellungnahme zum Entwurf der
Allgemeinen Methoden des IQWiG
Prof. Dr. med. Jürgen Windeler
Im Mediapark 8
50670 Köln

- per E-Mail vorab an methoden@iqwig.de -

Berlin, 31.01.2017

**Stellungnahme der Kassenzahnärztlichen Bundesvereinigung zum
Entwurf der Allgemeinen Methoden des IQWiG, Version 5.0, Stand: 07.12.2016**

Sehr geehrter Herr Professor Windeler,

anbei erhalten Sie die Stellungnahme der Kassenzahnärztlichen Bundesvereinigung zum Entwurf der Allgemeinen Methoden des IQWiG, Version 5.0. Wir bitten um eine Eingangsbestätigung, gern auch an die E-Mail-Adresse [REDACTED]. Vielen Dank.

Mit freundlichen Grüßen


i.A. [REDACTED]

Dr. Jörg Beck MHA
Leiter Abteilung Qualitätsinstitut, Leitlinien

Anlagen:

- Formblatt zur schriftlichen Stellungnahme zum Entwurf
- Stellungnahme der Kassenzahnärztlichen Bundesvereinigung zum Entwurf der Allgemeinen Methoden des IQWiG, Version 5.0

**Stellungnahme der KZBV zum
Entwurf der Allgemeinen Methoden
des IQWiG
Version 5.0, Stand: 07.12.2016**

Die KZBV nimmt zum vorliegenden Entwurf wie folgt Stellung:

Übergreifende Anmerkung

Die in den Allgemeinen Methoden abgebildete Vorgehensweise sieht vor, dass ausschließlich randomisierte kontrollierte Studien (RCTs) als relevante wissenschaftliche Literatur in die Nutzenbewertung einfließen sollen, insofern ein entsprechendes RCT theoretisch realisierbar erscheint. Als Begründung wird angeführt, dass RCTs, sofern sie methodisch adäquat und der jeweiligen Fragestellung angemessen durchgeführt wurden, mit der geringsten Ergebnisunsicherheit behaftet seien. Sie liefern daher die zuverlässigsten Ergebnisse für die Bewertung des Nutzens oder Zusatznutzens einer medizinischen Intervention. Die getroffene allgemeine Aussage negiert, dass neben RCTs nicht auch andere Studien - insbesondere retrospektive Kohortenstudien, das im Kontext nichtmedikamentöser Verfahren maximal zu realisierende Studiendesign - zuverlässige Ergebnisse für die Bewertung des Nutzens abbilden können.

Hierzu wird in Kapitel 3.4 „**Nichtmedikamentöse therapeutische Interventionen**“ ausgeführt:

„In aller Regel werden solche Studien jedoch aufgrund der ihnen immanenten geringeren Ergebnissicherheit allenfalls Anhaltspunkte für einen (Zusatz-)Nutzen oder Schaden liefern können. Die Einbeziehung von Studien eines niedrigeren Evidenzgrades steht im Einklang mit der entsprechenden Regelung im 2. Kapitel § 13 Absatz 2 der Verfahrensordnung des G-BA [248], wo aber dafür eine besondere Begründungspflicht betont wird [...]“

Die dann folgende Ableitung

„D. h., allein die Nichtverfügbarkeit von Studien höchster Evidenz kann im Allgemeinen nicht als ausreichende Begründung für eine Nutzenbewertung auf Basis von Studien niedrigerer Evidenz angesehen werden.“

negiert die Grundsätze der Evidenzbasierten Medizin, die „bestverfügbare“ und eben nicht die „bestmögliche“ Evidenz zur Bewertung heranzuziehen. 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 Gemeinsamen Bundesausschusses (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.

Aus gegebenem Anlass soll auf den aktuell am 24.01.2017 veröffentlichten Vorbericht N15-01 „Systematische Behandlung von Parodontopathien“ hingewiesen werden. Entgegen der vorweg skizzierten Vorgehensweise erfolgt ein kategorischer Ausschluss von retrospektiven Kohortenstudien und es wird die Forderung nach der höchsten Evidenz aufgestellt.

In Kapitel 3.4 „Nichtmedikamentöse therapeutische Interventionen“ auf Seite 66 steht geschrieben, dass

„Studien im nichtmedikamentösen Bereich im Vergleich zu Arzneimittelstudien häufig mit besonderen Herausforderungen und Schwierigkeiten verbunden sind [457]. Beispielsweise wird oft die Verblindung des die Intervention ausführenden Personals unmöglich und die der Patientinnen und Patienten nur schwierig oder ebenfalls nicht zu bewerkstelligen sein.“

Diese Besonderheiten und Schwierigkeiten gelten auch für die Zahnmedizin. Wie sollte der Behandler bei einem Therapieverfahren, bei dem z.B. unter Lokalanästhesie behandelt wird, verblindet werden? Das geht nur, wenn der Behandler Medikamente verteilt, die entweder Verum oder Placebo sind.

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 Abhängigkeit der erhobenen Daten unklar ist. Dies ist nicht nachvollziehbar, da gerade dieses Studiendesign interindividuelle Variabilitäten z.B. durch unterschiedliches Putzverhalten ausschließt und somit Verzerrungen minimiert werden. Bekannte und unbekannt personengebundene Störgrößen werden quasi gleichmäßig auf „Experimental- und Kontrollgruppen“ verteilt.

1. Zu Kapitel 2.2.1 Einbindung von Betroffenen in die Erstellung von IQWiG-Produkten, S. 35 – 36

„Die Einbindung von Betroffenen im Rahmen der Erstellung von systematischen Übersichten und HTA-Berichten ist mittlerweile etablierter internationaler Standard der Nutzenbewertung [128,215,401] [...]“

Stellungnahme der KZBV:

In dem Kapitel 2.2.1 „Einbindung von Betroffenen in die Erstellung von IQWiG Produkten“ wird ausgeführt, in welchen Formen und in welchen Institutsprodukten Betroffene eingebunden werden sollen. Hierzu wird auf S. 35 zu den HTA-Berichten ausgeführt, dass die Einbindung von Betroffenen im Rahmen der Erstellung von systematischen Übersichten und HTA-Berichten mittlerweile etablierter internationaler Standard der Nutzenbewertung sei. Es ist daher nicht nachvollziehbar, dass eine Einbindung Betroffener in die HTA-Berichterstellung nicht vorgesehen ist (Tabelle 2, S. 36). Die KZBV regt an, auch hier zur Beginn der Berichtsbearbeitung Betroffene im Rahmen der Festlegung patientenrelevanter Endpunkte und relevanter Subgruppen einzubinden.

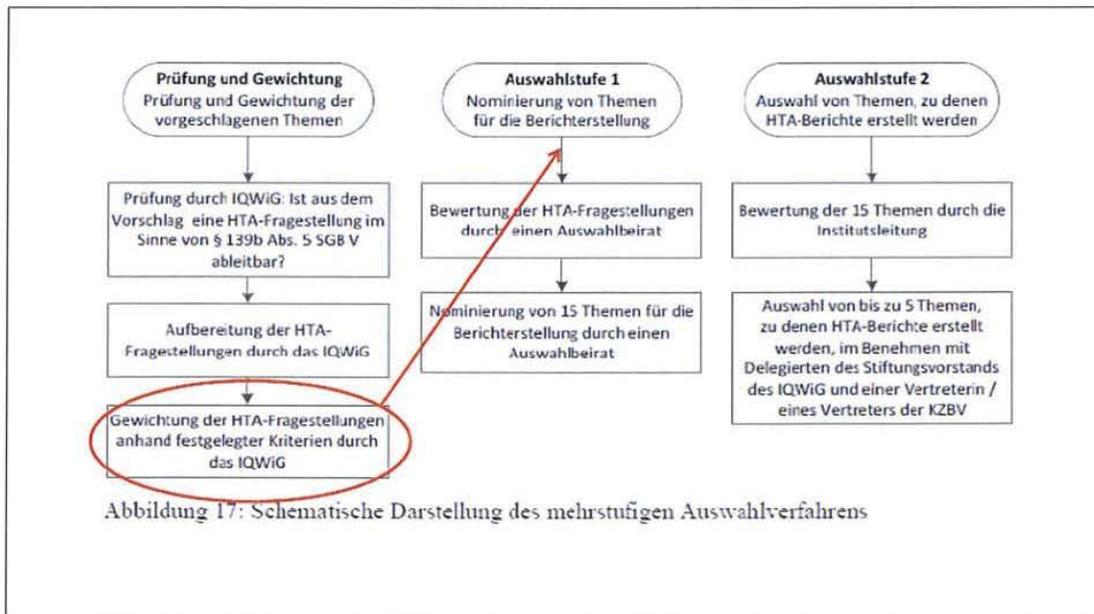
3. Zu Kapitel 5.4.1 Hintergrund, S. 120

„Ärztliches Handeln, das nach medizinisch-wissenschaftlicher Erkenntnis und / oder ärztlicher Erfahrung in der Profession akzeptiert ist, stellt den medizinischen Standard dar [296]. Dieser normative Standard wird hier als Versorgungsstandard bezeichnet. Er kann durch Gesetze, Verordnungen und Richtlinien vorgegeben sein oder in Leitlinien identifiziert werden (siehe auch Abschnitt 5.1). Auch die Referenzwerte von Qualitätsindikatoren können als Versorgungsstandards interpretiert werden [226].“

Stellungnahme der KZBV:

Die KZBV regt an, im Abschnitt 5.4.1. im letzten Absatz unter „B) Versorgungsstandard und Versorgungsrealität“ auf S. 120 den letzten Satz dahingehend zu konkretisieren, was mit „interpretiert“ gemeint ist.

4. Zu Kapitel 6.3 Auswahl der Themen für die HTA-Berichte, S. 132



Stellungnahme der KZBV:

Das IQWiG verfolgt in einer Art „Vorstufe“, selbst eine Prüfung und Gewichtung der Themenvorschläge aufzubereiten und diese als Entscheidungsgrundlage einem Auswahlbeirat zukommen zu lassen. Die KZBV regt diesbezüglich an, im Kapitel 6.3 in der Abbildung 17 auf S. 132 die „Prüfung und Gewichtung“ durch „Formale Prüfung“ zu ersetzen und die Gewichtung (letzter Kasten links) in den Prozessschritt „Auswahlstufe 1“ zu verlagern:

„Gewichtung der HTA-Fragestellungen anhand festgelegter Kriterien durch den Auswahlbeirat.“

Die KZBV verweist zudem auf die bereits vorstehend unter 1. zu Kapitel 2.2.1 gefasste Anregung, auch in die HTA-Berichterstellung eine Einbindung der Betroffenen primär zur Beginn der Berichtsbearbeitung im Rahmen der Festlegung patientenrelevanter Endpunkte und relevanter Subgruppen durchzuführen.

5. Zu Kapitel 6.3.2 Prüfung und Gewichtung der Themen, S. 133

„Die eingereichten Vorschläge werden zeitnah durch das Institut geprüft und gegebenenfalls redaktionell und qualitativ bearbeitet.“

Stellungnahme der KZBV:

Die KZBV regt an, im Abschnitt 6.3.2 auf S. 133 die unspezifische Angabe „zeitnah“ durch eine konkrete Fristangabe zu ersetzen.

6. Zu Kapitel 6.3.3 Erste Stufe des Auswahlverfahrens: Nominierung von Themen, S. 133 - 134

„In der ersten Stufe des Auswahlverfahrens nominiert ein Auswahlbeirat Themen für die HTA-Berichterstellung. Dazu werden dem Auswahlbeirat die aufbereiteten Themenvorschläge und die gewichtete Liste aller HTA-Fragestellungen zur Verfügung gestellt. Der Auswahlbeirat wählt auf dieser Basis 15 Themen für die HTA-Berichterstellung aus. Dabei wird sowohl die Bürger- und Patientensicht als auch die wissenschaftliche Perspektive mit einbezogen.“

Stellungnahme der KZBV:

Die KZBV regt an, die Einbindung des Auswahlbeirats konkreter zu spezifizieren. Ist für die Nominierung der Themen für die HTA-Berichterstellung durch den Auswahlbeirat z.B. ein Sitzungstermin vorgesehen? Auch bleibt der Nominierungsprozess selbst unklar. Soll die Nominierung der Themen in Diskurs oder anonym erfolgen? Erfolgt die Auswahl der Themen auf Grundlage von Mehrheitsbeschlüssen?

7. Zu Kapitel 6.3.4 Zweite Stufe des Auswahlverfahrens: Auswahl von Themen, zu denen HTA-Berichte erstellt werden, S. 134

„Zu den in der ersten Auswahlstufe nominierten Themenvorschlägen wird in der zweiten Stufe des Auswahlverfahrens eine Vorabrecherche (siehe Abschnitt 8.1.1) durchgeführt. Auf dieser Basis wählt die Institutsleitung im Benehmen mit Vertretern der den Stiftungsrat bildenden Organisationen aus den in der ersten Stufe des Auswahlverfahrens bestimmten Themen bis zu 5 aus, für die HTA-Berichte erstellt werden.“

Stellungnahme der KZBV:

Die KZBV regt an, in der Auswahlstufe 2 die Benehmensherstellung durch ein Einvernehmen zu ersetzen.

8. Unter Bezug auf Abbildung 17 auf Seite 132 in Kapitel 6.3

„Auf dieser Basis wählt die Institutsleitung im Benehmen mit Vertretern der den Stiftungsrat bildenden Organisationen aus den in der ersten Stufe des Auswahlverfahrens bestimmten Themen bis zu 5 aus, für die HTA-Berichte erstellt werden.“

Stellungnahme der KZBV:

Die KZBV begrüßt es sehr, in die Auswahlstufe 2 eingebunden zu werden, weist aber auf folgende Inkonsistenz hin: Während im Fließtext in Kapitel 6.3.4 Vertreter des **Stiftungsrats** in die Auswahlstufe 2 eingebunden werden sollen (S. 134), ist in dem korrespondierenden Flussdiagramm die Einbindung von Delegierten des **Stiftungsvorstandes und der KZBV** vorgesehen (S. 132, Abbildung 17). Dies sollte vereinheitlicht werden, auch wenn der identische Organisationskreis gemeint ist.

10. Zu Kapitel 8.2 Fokussierte Informationsbeschaffung, S. 162 - 163

„Nicht für alle Fragestellungen ist es erforderlich oder möglich, eine auf Vollständigkeit ausgelegte Informationsbeschaffung durchzuführen. In solchen Fällen wird eine sogenannte fokussierte Informationsbeschaffung durchgeführt, insbesondere wenn weiterhin der Anspruch an ein systematisches und transparentes Vorgehen besteht [...]“

Stellungnahme der KZBV:

Die KZBV regt an, im Abschnitt 8.2 auf den Seiten 162 f. den Terminus „fokussierter Informationsbedarf“, der zwar unter Bezugnahme auf die Quelle von Glanville [256] eingeführt wird, im Text näher zu erläutern. Die inhaltliche Differenzierung der Begrifflichkeiten wird nicht deutlich. Stellt die „fokussierte Informationsbeschaffung“ z.B. eine Art Zwischenstufe zur „systematischen Literaturrecherche“ und der „orientierenden Recherche“ dar?

Berlin, 31.01.2017



i.A. Dr. Jörg Beck MHA

Leiter Abteilung Qualitätsinstitut, Leitlinien

A.1.24 – Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen e. V. (MDS)

Autoren:

- Eyding, Dirk
- Janatzek, Sandra

MDS | Postfach 10 02 15 | 45002 Essen

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

Vorab per Mail an: methoden@iqwig.de

Servicezeiten:
Montag-Donnerstag 8.30-16.00 Uhr
Freitag 8.30-14.00 Uhr

Datum
31.01.2017

Ihr Ansprechpartner
Dr. Dirk Eyding

Stellungnahme zum Entwurf des Methodenpapiers Version 5.0 vom 07.12.2016

Sehr geehrte Damen und Herren,

hiermit reichen wir folgende Stellungnahme zum Entwurf der Allgemeinen Methoden, Version 5.0, ein. Auch da der MDS bisher noch keine Stellungnahme eingereicht hat, erlauben wir uns auch Punkte zu kommentieren, die unverändert aus Version 4.2 übernommen wurden und bei denen wir eine Aktualisierung vermissen.

Im Einzelnen:

Abschnitt 3.1.4. Belege aus einer Studie (unterer Abschnitt, S. 51)

Hier wird dargelegt, dass unter bestimmten Bedingungen auch aus einer einzigen Studie Aussagen mit der qualitativen Aussagesicherheit „Beleg“ abgeleitet werden können. Wir verstehen insbesondere die Ausführung zum geforderten Signifikanzniveau so, dass die Aussagesicherheit „Beleg“ nur für den jeweiligen Endpunkt, der neben allen anderen Anforderungen auch einen p-Wert von $<0,001$ aufweist, ausgesprochen wird. Andere Endpunkte, die alle anderen Anforderungen erfüllen, aber einen p-Wert von $\geq 0,001$ zeigten, erhalten die Aussagesicherheit „Hinweis“.

Dieser Ansatz ist eine interpretierte und operationalisierte Version der EMA-Guideline zu den Bedingungen für die Zulassung eines Arzneimittels basierend auf einer Studie [1]. Dort wird nicht ganz klar, wie viele und welche Endpunkte und wie viel besser als $p < 0,05$ das Signifikanzniveau sein sollen, um eine Zulassung aussprechen zu können. Es ist aber klar, dass eine Zulassung nicht erst dann ausgesprochen werden kann, wenn *alle* Endpunkte *alle* Anforderungen erfüllen.

Es kann angemessen sein, ein besonders „gutes“ (im Sinne einer geringen Verzerrung und einer hohen externen Validität) Experiment auszuzeichnen, wo zu dieser hohen Qualität auch eine gewisse

– gemessen an konventionellen Vorgaben – Überpowerung gehören kann, um dann aus einem solchen Experiment einen Beleg ableiten zu können.

Unseres Erachtens liegt aber im IQWiG-Vorschlag einer Operationalisierung eine ungünstige Vermischung von qualitativer und quantitativer Ergebnissicherheit vor, für die aufgrund der vom IQWiG gewählten, konfidenzintervallorientierten Methodik der Bestimmung der quantitativen Ergebnissicherheit kein Anlass besteht. Die größere Unsicherheit bzgl. der Effektstärke bei selteneren Ereignissen im Vergleich zu häufigeren (bei angenommen gleichem relativen Effekt) werden durch die Verbreiterung der KIs ausgedrückt. Eine doppelte „Abstrafung“ durch Änderung der qualitativen und quantitativen Ergebnissicherheit erscheint uns überflüssig.

Folge des IQWiG-Vorgehens könnte auch sein, dass bestimmte Klassen von Endpunkten, wir denken hier insbesondere an die selteneren, aber schwerwiegenderen Ereignisse, systematisch abgewertet werden gegenüber denjenigen, für die der Studiendurchführende die Überpowerung vorgesehen hat. Insofern könnte diese Dopplung der Abstrafung sogar verzerrende Konsequenzen zeitigen, indem sich für SUE z.B. eher „Hinweise“ ($p > 0,001$) für einen „geringen“ (95%-KI zu nah an 1) Schaden ergeben, während die selektierten Nutzen-Endpunkte einen „Beleg“ für einen „beträchtlichen“ (Zusatz-)Nutzen erhalten.

Unseres Erachtens besteht kein Anlass, das Durchführen sog. Mega-Trials zu belohnen, indem hier die Chance für Belege besteht. Die Erfahrungen mit Mega-Trials, z.B. den großen Studien zur Thromboseprophylaxe mit den neuen Antikoagulanzen [2-6], zeigen, dass die Kontroversen um ihre Interpretierbarkeit trotz ihrer Größe und des betriebenen Aufwands (oder deshalb) erheblich war, und offene Fragen nun sogar mit nicht-randomisierten Beobachtungsstudien zu beantworten versucht werden [2]. **Insofern würden wir uns dafür aussprechen, diesen Passus ersatzlos zu streichen.**

Wenn man eine solche Belohnung beibehalten möchte, sollte unseres Erachtens prinzipiell, endpunktübergreifend das gesamte Experiment als so hochklassig angesehen werden, dass bei einem Endpunkt nicht ausschließlich durch Verfehlen der $p < 0,001$ -Forderung nur noch ein Hinweis ausgesprochen wird. Denkbar wäre etwa, dass in dem wichtigsten oder einigen wenigen wichtigen Endpunkten das $p < 0,001$ -Kriterium erfüllt sein muss, um prinzipiell für die Studie insgesamt einen Beleg ableiten zu können (unbenommen der endpunktbezogenen Verzerrungspotenzialbewertung, die weiterhin eine begründete Abwertung ermöglicht).

9.3.8. Meta-Analysen und 9.3.10. Subgruppen: Umgang mit Heterogenität

Hier liegt im Verhältnis zur Version 4.2 eine bedeutsame, mit potenziell weit reichenden Folgen verbundene Änderung vor, indem die Hinweise auf Heterogenität bzw. auf Interaktionen mit Subgruppenmerkmalen bei einem p-Wert von $< 0,2$ aber $> 0,05$ komplett entfallen sind. Damit ist eine in Version 4.2 leicht unterschiedliche Behandlung von Testergebnissen mit $0,05 < p < 0,2$ bei Heterogenitätstests in Meta-Analysen bzw. bei Interaktionstests im Rahmen von Subgruppenanalysen ebenfalls weggefallen. (Bei vorgegebenen Subgruppenmerkmalen wurde in diesem Fall ein Gesamteffekt und ein subgruppenspezifischer Effekt dargestellt, während bei

Heterogenität kein Gesamteffekt mehr dargestellt wurde.) In Version 5.0 werden nur noch bei „Belegen“ (d.h. $p < 0,05$) einer Interaktion bzw. von Heterogenität keine Meta-Analysen mehr durchgeführt. Unklar bleibt, ob der „Nachweis der Übertragbarkeit“ (eigentlich ja „fehlender Nachweis der Nichtübertragbarkeit“) in der zulassungsüberschreitenden Situation (Kapitel 3.3.1) häufiger gegeben sein wird als bisher.

Interessant ist, dass sowohl in 9.3.8 als auch in 9.3.10 unbegründet bleibt, weshalb das IQWiG hier eine Änderung vornimmt. Nach wie vor wird jedoch dargestellt, dass die entsprechenden Tests nur eine geringe Power hätten, sodass eigentlich nahe läge, größere p-Werte als 0,05 als Schwelle festzulegen (zum Preis von mehr zufallsbedingt falsch positiven Testergebnissen). Auch wird in 9.3.8 nun darauf hingewiesen, dass bei wenigen Studien in einer Meta-Analyse (2-4) die Schätzung der Heterogenität nicht verlässlich möglich sei, sodass von REM Meta-Analysen eher abzusehen sei, bzw. keine Meta-Analyse überhaupt mehr durchzuführen sei. Heißt das, dass in dieser Situation nur FEM Meta-Analysen durchgeführt werden, weil man von abwesender Heterogenität aufgrund der mangelhaften Schätzbarkeit ausgeht (was vermutlich wenig sachgerecht wäre)? Oder heißt das, dass regelhaft keine Meta-Analysen mehr durchgeführt werden (was bei geringer Heterogenität die Chancen der Powererhöhung durch Meta-Analysen vergibt)? Oder soll – im Gegenteil – bei wenigen Studien regelhaft eine Meta-Analyse basierend auf den alternativ vorgestellten Verfahren (Bayes'sche oder GLM) gerechnet werden?

Zur Illustration der Konsequenzen dieser Änderung haben wir einige wenige, überschlägige Rechnungen durchgeführt. Unser Ergebnis: Selbst beim Vorhandensein von 5 Studien sind Konstellationen wahrscheinlich und möglich, die mit dem $p \geq 0,05$ -Kriterium ein Zusammenfassen qualitativ klar unterschiedlicher Studien (z.B. Studien mit signifikantem Effekt und Studien mit gegenläufigem, aber nicht signifikantem Unterschied) „erzwingen“, obwohl die I-Quadrat-Werte mit $>50\%$ im Bereich „substanzieller“ Heterogenität liegen. Bei weniger als 5 Studien sind die I-Quadratwerte in der $p \geq 0,05$ -Situation bei derselben Effektkonstellation entsprechend noch größer.

Unseres Erachtens steht die relativ fixe Festlegung auf eine $p = 0,05$ Grenze und auf eine (in 9.3.10) entsprechende Sprache im Widerspruch zu den dargestellten Problemen. Da insbesondere in der Frühen Nutzenbewertung häufig nur wenige Studien eingeschlossen und bewertet werden, scheint es uns wahrscheinlich, dass diese Vorfestlegung zur Durchführung einer Meta-Analyse bzw. gemeinsamen Darstellung eines Effektes in Subgruppenanalysen bei $p \geq 0,05$ häufiger zu erkennbar problematischen Schlussfolgerungen führen könnte.

Gerade in Situationen, in denen die Heterogenitätstests nur eine geringe Power haben, **sollte unseres Erachtens die „Grauzone“ $0,05 < p < 0,2$ erhalten bleiben** (geschaffen werden bei Heterogenitätstestergebnissen mit $p < 0,2$), **indem für diese Situation beide Varianten (gemeinsamer Schätzer und getrennte bzw. kein Schätzer) dargestellt werden. Es sollte dann einer inhaltlichen Bewertung obliegen, ob und für welche Variante man sich entscheidet**, und es sollten die tragenden Gründe für diese Entscheidung dargelegt werden.

Grund für das Verwerfen von Heterogenität trotz $0,05 < p < 0,2$ wird in erster Linie die Einschätzung als Zufallsergebnis sein, mit dem man bei einem $p < 0,2$ -Kriterium natürlich rechnen muss. Indizien für ein Zufallsergebnis können z.B. die fehlende Plausibilität des subgruppenbildenden Merkmals oder die

Singularität des Befundes sein, Indizien für einen tatsächlichen Effektunterschied z.B. eine gegebene Plausibilität oder die Konsistenz der Richtung mit anderen Endpunkten.

Neu hinzugekommen für die Durchführung von Subgruppenanalysen (9.3.10) ist, dass sie regelhaft überhaupt nur durchgeführt werden, falls mindestens 10 Ereignisse in einer der Subgruppen aufgetreten sind.

Zum einen ist unklar, was genau gemeint ist (10 Ereignisse pro Zelle, also pro Arm und Subgruppe oder nur pro Subgruppe? Bei mehreren Studien: 10 über alle Studien hinweg oder je Studie?), zum anderen fragt sich, ob ein solches Kriterium – wenn man es denn anwenden möchte – nicht analog auch für Heterogenitätstests gelten sollte. Abhängig von der Antwort wäre zu prüfen, inwiefern diese Regelung ggf. zu häufig zu eigentlich eher als nicht sachgemäß einzuschätzendem Zusammenfassen von Daten bzw. Nichttrennen in Subgruppen führt.

3.3.2. Undurchführbarkeit von RCT bei Arzneimittelbewertungen

Nicht-RCT-Evidenz wird in Arzneimittelbewertungen nur in Ausnahmefällen anerkannt, wenn RCT als undurchführbar eingestuft werden. Als Begründung für die Undurchführbarkeit wird u.a. die „starke Präferenz“ für eine der Therapiealternativen seitens Arzt/Patient angeführt.

Unseres Erachtens ist dies ein nicht akzeptables Kriterium, weil es letztlich außerhalb der Überprüfbarkeit steht und zudem durch begründete oder auch nicht begründete Hoffnungen stark beeinflussbar ist. Beispielhaft sei der Streit um die Gründe für das Scheitern der PREFERE-Studie genannt [7]. Auch das häufige Scheitern von regulatorisch eigentlich geforderten Postmarketingstudien in der Onkologie, die einen Effekt auf eigentlich relevante Endpunkte wie das Gesamtüberleben zeigen sollen bei Zulassungen, die aufgrund von Effekten auf Surrogate ausgesprochen wurden [8], kann (auch) mit „starker Präferenz“ erklärt werden, wenn die erzielten Surrogatvorteile vorschnell als hinreichender Nachweis eines Nutzens akzeptiert werden.

Bei „eigentlichem“ Equipoise, d.h. Unwissen welche der Therapiealternativen hinsichtlich des Therapieziels die Bessere ist, sollte die Präferenz keine Rolle spielen dürfen. Bei tatsächlich fehlender Equipoise sollte diese Studie tatsächlich undurchführbar sein, aber auch überflüssig, weil das Wissen über die Verhältnisse der Effekte der geplanten Alternativen schon existiert.

Über ein unkritisches Akzeptieren einer „starken Präferenz“ kann das IQWiG sich selbst an vielen Stellen binden, Nicht-RCT-Evidenz zu akzeptieren, wenn die „starke Präferenz“ ins Feld geführt wird.

Unseres Erachtens sollte dieses Kriterium ersatzlos gestrichen werden. In solchen Situationen wäre es angemessener, auf das (immer noch) vorhandene Equipoise

hinzuweisen und die entsprechenden Studien (ggf. unter Hinweis auf besondere Anforderungen an eine neutrale, strukturierte Aufklärung der Patienten) zu fordern.

Außerdem ist uns unklar geblieben, weshalb sich die angesprochenen Aspekte in 3.3.2. nicht auch auf Nutzenbewertungen für Nicht-Arzneimittel beziehen.

9.3.3 Relevanzbewertung, insbesondere von Skalen

Bei der Relevanzbewertung von Skalen wird folgende Hierarchie vorgestellt: 1. Prüfung, ob eine Irrelevanzschwelle für einen Mittelwertsunterschied von Gruppen vorliegt, falls das nicht: 2. Prüfung, ob es ein sinnvolles Responsekriterium, z.B. eine individuelle MID gibt, die eine Responderanalyse gestattet, wenn das nicht: 3. Bewertung auf Basis standardisierter Mittelwertsdifferenzen.

Unseres Erachtens weist diese Hierarchie eine wichtige Schwäche auf, und sie wird – möglicherweise deswegen – in der Praxis noch nicht wirklich gelebt; zumindest ist uns nicht bekannt, dass auf Basis etablierter Irrelevanzmittelwertsdifferenzen schon einmal Effekte bewertet wurden, noch dass es für irgendeine Skala einen etablierten Irrelevanzunterschied gebe.

Die von uns gesehene Schwäche liegt darin, dass nicht klar zwischen (ir)relevantem *Effekt* und (ir)relevanter *Änderung* unterschieden wird. In der Hierarchie selbst kann man feststellen, dass sich die Punkte 1. und 3. auf eine Relevanzeffektschwelle beziehen, während Punkt 2. über eine Relevanzschwelle für Effekte gar nicht spricht, sondern lediglich über eine relevante Mindeständerung auf individueller Ebene. Es wird dann sogar expliziert, dass jeder hinsichtlich des über diese relevante Mindeständerung definierten Responsekriteriums statistisch signifikante Effekt relevant sein kann.

Wir meinen, dass es eine feste Grenze eines relevanten Effektes für einen Endpunkt gar nicht geben kann. Die Relevanz eines Effektes bei einem Endpunkt (wie einem skalenbasierten) kann nur im Zusammenhang mit Effekten bei anderen Endpunkten vor dem Hintergrund der Erkrankung beurteilt werden, selbst wenn man konzidiert, dass skalenbasierte Endpunkte eher eine etwas globalere Beurteilung des Gesundheitszustandes implizieren, und – wenn sie PROs sind – in Richtung „subjektiver Nettonutzen“ interpretiert werden können. Dennoch fließen andere Endpunkte in die Bewertung mit ein und auch scheinbar so klar relevante Endpunkte wie das Überleben können in bestimmten Situationen in den Hintergrund treten – und damit auch die Bedeutung von Effekten bzgl. dieses Endpunktes.

Sofern bei binären Endpunkten das Eintreten des Ereignisses für den Patienten spürbar und von einer angenommenen Bedeutung ist, wie Tod, Infektion, Knochenbruch, Exazerbation etc. ist die Relevanz des Endpunktes schon sichergestellt.

Bei einer Skala hingegen kommt als Aufgabe hinzu, sozusagen das „Ereignis“ noch zu definieren, selbst wenn die Domänen der Skala für den Patienten relevante Konzepte erfassen. Die Dichotomisierung eines skalenbasierten Endpunktes z.B. mittels einer individuellen MID leistet genau das.

Unseres Erachtens sollte daher die Bewertung von Effekten bei skalenbasierten Endpunkten immer und ausschließlich auf Responderanalysen beruhen. Um der Gefahr zu entgehen, dass eine bestimmte Responderchwelle ergebnisgesteuert ausgewählt wurde, sollten immer Sensitivitätsanalysen präsentiert werden, in denen andere Schwellen angewendet wurden (idealerweise als Ogive-Plot mit allen Schwellenwerten). Bei Skalen, bei denen (noch) Unklarheit besteht, wie die MID (bzw. ein geeignetes Responsekriterium) aussieht, könnte bei klar jenseits von MIDs bei anderen, ähnlichen Skalen gewählten Schwellen eine Bewertung auch einer solchen Skala möglich sein, denn ein Effekt hinsichtlich eines Responsekriteriums „klar relevante Änderungen“ wäre genauso wichtig wie ein Effekt hinsichtlich eines Responsekriteriums „minimal spürbare Änderungen“. In Fällen, in denen unklar ist, in welche Richtung sich die Krankheit entwickelt, kann es auch angemessen sein, sowohl Verbesserungen als auch Verschlechterungen zu präsentieren.

Insofern könnten aus unserer Sicht die Punkte 1. und 3. im Prinzip entfallen, während zu 2. noch Ergänzungen in obigem Sinne zu fordern wären.

Da in vielen Nutzenbewertungen kein Zugriff auf die Rohdaten besteht, ist die Forderung nach vielen Responderanalysen dort nicht umsetzbar, wenn sie nicht zufällig schon so publiziert sind (was derzeit unwahrscheinlich ist). Insofern wäre aus diesen Überlegungen heraus das Beibehalten einer SMD-basierten Bewertung als ungenaue Näherung für das eigentliche Ziel zu vertreten.

3.8 Potenzialbewertung: Unklarheit hinsichtlich des Effektbegriffs

Im Abschnitt zur Ergebnissicherheit von Studien, die für eine Potenzialbewertung herangezogen werden können, werden in der Kategorie „minimale qualitative Ergebnissicherheit“ Studientypen genannt, die für sich nach dem vom IQWiG zugrunde gelegten kontrafaktischen Effektbegriff [9] nicht für die Ableitung von Effekten geeignet sind: Nicht-vergleichende Studien.

Entweder muss diese Kategorie komplett entfallen, denn auch ein Potenzial ist das Potenzial eines Effektes. Besser wäre es hingegen klarzustellen, dass im Falle unkontrollierter Evidenz historische Kontrollen als unadjustierte indirekte Vergleiche die Grundlage der Potenzialfeststellung sein müssen.

3.5 Diagnostische Verfahren: Anforderungen an Interaktionsdesigns

Im Abschnitt zu den Interaktionsdesigns wird dargelegt, dass es bei allen Studien im Interaktionsdesign wichtig sei, dass „[...] die Informationen (z.B. Gewebeproben) zum interessierenden Merkmal für alle Studienteilnehmer oder zumindest für eine repräsentative Stichprobe vollständig vorliegen [...]“. **Nach unserer Einschätzung ist die Forderung einer repräsentativen Stichprobe nicht zwingend erforderlich**, um Nutzensaussagen ableiten zu können. Für den Fall, dass Informationen zum interessierenden Merkmal nicht für alle Studienteilnehmer, sondern nur für eine Stichprobe vorliegen, ist es aus unserer Sicht ausreichend, wenn diese beiden Anforderungen erfüllt sind:

- 1.) Innerhalb dieser Stichprobe ist die Strukturgleichheit der beiden Therapiegruppen gewährleistet, d.h. die Randomisation ist innerhalb der Stichprobe erhalten.
- 2.) Die wesentlichen Merkmale dieser Stichprobe sind beschrieben, so dass die Population, für die Aussagen anhand der Stichprobe abgeleitet werden können, konkretisiert/formuliert werden kann.

Auf diese Weise wäre es beispielsweise innerhalb von prospektiv-retrospektiven Designs bei onkologischen Erkrankungen grundsätzlich möglich, aus der verfügbaren Stichprobe Nutzensaussagen für eine gewisse Population abzuleiten, auch wenn für Patienten mit z.B. kleinen Tumoren die Informationen zum interessierenden Merkmal aufgrund aufgebrauchter Proben weitgehend fehlen.

Wir bedanken uns für die Möglichkeit, unsere Überlegungen einbringen zu können.

Freundliche Grüße

Dr. Dirk Eyding
Bereich Evidenzbasierte Medizin

Dr. Sandra Janatzek
Bereich Evidenzbasierte Medizin

Literatur

1. European Agency for the Evaluation of Medicinal Products. Points to consider on application with 1. Meta-analyses; 2. One pivotal study: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC50003657.pdf, letzter Zugriff: 26.01.2017, 2001.
2. n.n. Neues zur oralen Antikoagulation bei Vorhofflimmern. *Arzneimittelbrief* 2017;51.
3. n.n. ROCKET-AF- Studie mit Rivaroxavan (XARELTO) - Validität der Daten steht weiterhin infrage. *arznei-Telegramm* 2016;47:97.
4. n.n. Rivaroxaban (XARELTO) rehabilitiert? *arznei-Telegramm* 2016;47:22.
5. n.n. Neue Antikoagulanzen vs. Cumarine: ein unerträglicher Grabenkampf. *arznei-Telegramm* 2015;46:89.
6. n.n. Neue orale Antikoagulanzen bei Vorhofflimmern - Ist Nichtverordnung ein Behandlungsfehler? *arznei-Telegramm* 2014;45:13-14.
7. Windeler J. Prostatakrebs im Frühstadium. Warum wir die beste Therapie nicht kennen: <http://www.spiegel.de/gesundheit/diagnose/prostatakrebs-im-fruehstadium-juergen-windeler-verteidigt-prefere-studie-a-1127596.html>, letzter Zugriff: 26.01.2017, 2016.
8. Kim C, Prasad V. Cancer Drugs Approved on the Basis of a Surrogate End Point and Subsequent Overall Survival: An Analysis of 5 Years of US Food and Drug Administration Approvals. *JAMA internal medicine* 2015;175:1992-4.
9. Windeler J. Bedeutung randomisierter klinischer Studien mit relevanten Endpunkten für die Nutzenbewertung. In: Gesundheitsforschungsrat des Bundesministeriums für Bildung und Forschung, ed. Diskussionsforum zur Nutzenbewertung im Gesundheitswesen Begriffsdefinitionen und Einführung Dokumentation des ersten gemeinsamen Workshops von GFR und IQWiG am 4 September 2007 in Berlin. Berlin, 2007:26-31.

A.1.25 – MSD SHARP & DOHME GmbH

Autoren:

- Rämisch, Christiane
- Sahakyan, Narine
- Scheuringer, Monika

MSD SHARP & DOHME GMBH
Lindenplatz 1
85540 Haar
Tel. + 49 89 45 61 - [REDACTED]
Fax + 49 89 45 61 - [REDACTED]
www.msdd.de



MSD SHARP & DOHME GMBH · Postfach 12 02 · 85530 Haar
Institut für Qualität und Wirtschaftlichkeit
im Gesundheitswesen (IQWiG)
Im Mediapark 8
50670 Köln

30. Januar 2017

**Stellungnahme der MSD SHARP & DOHME GMBH zu den Allgemeinen Methoden,
Entwurf der Version 5.0 vom 07.12.2016**

Sehr geehrte Damen und Herren,

vielen Dank für die Möglichkeit zur Kommentierung der Allgemeinen Methoden, Entwurf der
Version 5.0.

Anbei finden Sie eine Übersicht zu Anmerkungen von Seiten MSD zu den Allgemeinen
Methoden, Entwurf der Version 5.0.

Mit freundlichen Grüßen

[REDACTED]

Dr. Monika Scheuringer MPH
Head HTA

[REDACTED]

Narine Sahakyan MD, MPH
Senior Manager HTA

[REDACTED]

Christiane Rämisch
Senior Manager HTA

Abschnitt Seitenzahl	Zitat Allgemeine Methoden	Anmerkungen MSD															
2.2.1 Einbindung von Betroffenen in die Erstellung von IQWiG-Produkten <i>Seite 35/36</i>	„Betroffene können insbesondere sein: Patientinnen und Patienten, ihre Eltern oder sonstige Angehörige sowie potenzielle Teilnehmer an Präventionsmaßnahmen.“	Nach dem Verständnis von MSD ergibt sich keine <u>regelmäßige</u> Einbindung von Patientinnen und Patienten in die Erstellung von IQWiG Produkten, vielmehr wird die Formulierung als „Kann“-Regelung verstanden. Eine Einbindung von Patientinnen und Patienten ist aber regelhaft umzusetzen. Im Zusammenhang damit ist eine transparente Darstellung des Einbindungsprozesses, also eine nachvollziehbare Offenlegung der Kommunikation im Rahmen der Einbindung, zu fordern.															
	„„Ferner ist eine (erneute) Einbindung von Betroffenen verzichtbar, wenn ein Berichtsthema bereits im Rahmen einer früheren Konsultation bearbeitet wurde“.	Wird auf eine erneute Einbindung verzichtet, ist nach Ansicht von MSD ein entsprechender Hinweis im Bericht wünschenswert.															
	<p>„Tabelle 2 zeigt für alle Produkte des Instituts die Form der Betroffenen einbindung.“</p> <p>Tabelle 2: Institutprodukte und Formen der Einbindung Betroffener</p> <table border="1" data-bbox="703 1018 1319 1345"> <thead> <tr> <th>Institutprodukt</th> <th>Form der Einbindung Betroffener</th> </tr> </thead> <tbody> <tr> <td>Bericht Rapid Report</td> <td>mündliche Konsultation</td> </tr> <tr> <td>Dossierbewertung Kosten-Nutzen-Bewertung nach § 35b SGB V</td> <td>schriftliche Konsultation</td> </tr> <tr> <td>Potenzialbewertung Bewertung gemäß § 137h SGB V Addendum</td> <td>keine Einbindung</td> </tr> <tr> <td>Gesundheitsinformation</td> <td>mündliche Konsultation, Nutzertestungen, Erfahrungsberichte</td> </tr> <tr> <td>Arbeitspapier</td> <td>mündliche oder schriftliche Konsultation nach Bedarf</td> </tr> <tr> <td>HTA-Bericht</td> <td>keine Einbindung</td> </tr> <tr> <td colspan="2">HTA: Health Technology Assessment; SGB: Sozialgesetzbuch</td> </tr> </tbody> </table>	Institutprodukt	Form der Einbindung Betroffener	Bericht Rapid Report	mündliche Konsultation	Dossierbewertung Kosten-Nutzen-Bewertung nach § 35b SGB V	schriftliche Konsultation	Potenzialbewertung Bewertung gemäß § 137h SGB V Addendum	keine Einbindung	Gesundheitsinformation	mündliche Konsultation, Nutzertestungen, Erfahrungsberichte	Arbeitspapier	mündliche oder schriftliche Konsultation nach Bedarf	HTA-Bericht	keine Einbindung	HTA: Health Technology Assessment; SGB: Sozialgesetzbuch	
Institutprodukt	Form der Einbindung Betroffener																
Bericht Rapid Report	mündliche Konsultation																
Dossierbewertung Kosten-Nutzen-Bewertung nach § 35b SGB V	schriftliche Konsultation																
Potenzialbewertung Bewertung gemäß § 137h SGB V Addendum	keine Einbindung																
Gesundheitsinformation	mündliche Konsultation, Nutzertestungen, Erfahrungsberichte																
Arbeitspapier	mündliche oder schriftliche Konsultation nach Bedarf																
HTA-Bericht	keine Einbindung																
HTA: Health Technology Assessment; SGB: Sozialgesetzbuch																	

Abschnitt Seitenzahl	Zitat Allgemeine Methoden	Anmerkungen MSD
		unumgänglich. Dies müsste sich also auch in Tabelle 2 widerspiegeln.
3.1.4 Endpunktbezogene Bewertung <i>Seite 51</i>	<p>„Aus lediglich 1 Studie kann im Ausnahmefall für einen Endpunkt ein Beleg für eine spezifische (Teil-)Population abgeleitet werden, wenn ein Studienbericht gemäß ICH-Richtlinien vorliegt, die auch sonst für einen Beleg geforderten Bedingungen erfüllt sind und darüber hinaus die Studie folgende besondere Anforderungen erfüllt:</p> <ul style="list-style-type: none"> • Die Studie ist multizentrisch, in jeden Studienarm wurden mindestens 1000 Patienten eingeschlossen und es gibt mindestens 10 Zentren.“ 	<p>Die Voraussetzung, dass jeder Studienarm mindestens 1000 Patienten einschließt, zu erfüllen, ist für viele Indikationen (z. B. Onkologie) nicht umsetzbar.</p> <p>Das Ziel einer Studie ist es, spezifische Studienfragen mit statistischen Verfahren zu untersuchen. Mit Hilfe einer studienfragebezogenen Fallzahlberechnung wird jene Stichprobengröße ermittelt, die erforderlich ist, um einen vorher bestimmten medizinisch relevanten Unterschied auf einem festgelegten Signifikanzniveau α mit einer definierten Teststärke nachweisen zu können [1].</p> <p>Der Studienplaner ist auch aus ethischen Gründen angehalten, die Fallzahl für jede Studie spezifisch zu ermitteln. Bei Vorliegen von starken Therapieeffekten widerspricht das Fortführen der Studie dem ethischen Grundsatz, nicht zu schaden.</p> <p>MSD sieht die Forderung nach mindestens 1000 Patienten in jedem Studienarm zur Ableitung eines Belegs für einen Zusatznutzen kritisch. Diese Forderung widerspricht nicht nur den geltenden internationalen Standards in der klinischen Forschung sondern auch der ethischen Unbedenklichkeit von Forschungsvorhaben.</p>

Abschnitt Seitenzahl	Zitat Allgemeine Methoden	Anmerkungen MSD
7 Evidenzbasierte Gesundheitsinformation für Bürgerinnen und Bürger <i>Seite 137</i>		<p>Auf www.Gesundheitsinformation.de werden zum Zwecke der „verständliche[n] Vermittlung von entscheidungsrelevantem Wissen zu gesundheitlichen Fragen“ auch die Ergebnisse der Gutachten, die das IQWiG im Auftrag des G-BA im Rahmen der Frühen Nutzenbewertung von Arzneimitteln erstellt hat, zusammenfassend dargestellt. Eine Aktualisierung der Informationen basierend auf der finalen Entscheidung des G-BA fehlt, es wird lediglich darauf verwiesen, dass der G-BA basierend auf dem Gutachten und der eingegangenen Stellungnahmen über den Zusatznutzen entscheidet.</p> <p>Im Sinne der Transparenz und Vollständigkeit ist anzustreben, die finale Entscheidung des G-BA nachträglich in die dargestellten Informationen einzuarbeiten.</p>
8.1.2 Suche in Studienregistern A) Aufbau einer Suchstrategie <i>Seite 159</i>	<p>„Suchen in Studienregistern sollten eine hohe Sensitivität ausweisen, einfach gehalten werden und nach Möglichkeit nur ein Konzept (in der Regel Indikation oder Intervention) berücksichtigen [257,352]. Es wird dabei zuerst mit den Begriffen zu dem Konzept gesucht, das sich am besten abbilden lässt und trotz hoher Sensitivität die wenigsten Treffer generiert. Nur bei zu vielen Treffern wird die Suche mit dem zweiten Konzept weiter eingeschränkt. Eine weitere Einschränkung (nach Status oder Phase) ist aufgrund der unterschiedlichen Qualität der einzelnen Einträge nur im Ausnahmefall vorzunehmen.“</p>	<p>Eine Suche in öffentlich zugänglichen Studienregistern (clinicaltrials.gov, EU Clinical Trials Register, Klinische Prüfungen PharmNet.Bund, International Clinical Trials Registry Platform Search Portal) für die Nutzenbewertung gemäß § 35a SGB V ist sinnvoll mit wirkstoff- und indikationsbezogenen Suchbegriffen durchzuführen.</p> <p>Nach Ansicht von MSD soll die Suche in Studienregistern nach beiden Konzepten analog zu der bibliografischen Literaturrecherche durchgeführt werden.</p>

Abschnitt Seitenzahl	Zitat Allgemeine Methoden	Anmerkungen MSD
9.3.8 Meta-Analysen C) Geringe Zahl von Ereignissen Seite 190	„Bei sehr seltenen Ereignissen kann die Peto-Odds-Ratio-Methode verwendet werden, die keinen Korrekturterm beim Vorliegen von Nullzellen erfordert [63,146]. Die Anwendung dieser Methode ist allerdings nur adäquat, wenn die zu schätzenden Effekte nicht zu groß sind und kein unbalanciertes Design vorliegt [71].“	<p>Gemäß dem <i>Cochrane handbook for systematic reviews of interventions</i> [2] ist das Peto-Odds-Ratio im Fall von Ereignisraten unter 1 % das stärkste und am wenigsten verzerrte Effektmaß, welches mittlerweile ein in dieser Situation etabliertes Maß darstellt. Zwar wird auch eingeräumt, dass ungleiche Gruppengrößen und außerordentlich große Behandlungseffekte zu einer verzerrten Schätzung führen können, allerdings wurde ungeachtet dessen die Peto-Methode als die Methoden mit der besten Performance identifiziert.</p> <p>Nach Ansicht von MSD ist das Peto-Odds-Ratio weiterhin dann angezeigt, wenn sich Ereignisraten unter 1 % ergeben, so dass die Notwendigkeit einer Nullzellenkorrektur vermieden wird. Da es sich bei dem in Brockhaus et al. (2014) [3] beschriebenen Verfahren um ein derzeit nicht etabliertes, nicht extern validiertes und auch in anderen Quellen nicht explizit empfohlenes Verfahren handelt, wird eine zusätzliche Berücksichtigung der dargelegten Grenzwerte, innerhalb derer das Peto-Odds-Ratio als valider Schätzer angesehen wird, kritisch gesehen.</p>
9.3.9 Indirekte Vergleiche Seite 193	„Allerdings müssen adäquate Methoden für indirekte Vergleiche verwendet werden. Die Anwendung nicht adjustierter indirekter Vergleiche (d. h. die Verwendung einzelner Arme aus verschiedenen Studien) wird abgelehnt [45].“	Betrachtet man die letzten Nutzenbewertungen insbesondere im Bereich der chronischen Hepatitis C, dann lässt sich feststellen, dass nicht adjustierte indirekte Vergleiche nicht nur Verwendung, sondern auch für die Beschreibung eines Zusatznutzens Akzeptanz finden. Dies begründet sich im Erreichen eines dramatischen Effektes, wie er auch in den Allgemeinen Methoden in Abschnitt 3.2.2

Abschnitt Seitenzahl	Zitat Allgemeine Methoden	Anmerkungen MSD
		<p>beschrieben ist. MSD geht davon aus, dass dieses Vorgehen weiterhin möglich ist und entsprechend gewürdigt wird, wenn keine RCTs entweder für einen direkten Vergleich oder adjustierte indirekte Vergleiche zur Verfügung stehen. Eine Ergänzung im entsprechenden Abschnitt ist wünschenswert.</p>
<p>9.3.10 Subgruppenanalysen <i>Seite 195</i></p>	<p>„Damit überhaupt eine gewisse Wahrscheinlichkeit für einen solchen Nachweis vorliegt, werden Subgruppenanalysen bei binären Daten regelhaft nur dann durchgeführt, falls mindestens 10 Ereignisse in einer der Subgruppen aufgetreten sind. Mit einem Ereignis ist hier gemeint, dass dieses im Verlauf einer Beobachtung auftritt und zu Baseline noch nicht eingetreten ist (z. B. Erlangen einer Virusfreiheit bei infizierten Personen).“</p>	<p>Grundsätzlich wird es als sinnvoll angesehen, Subgruppenanalysen nur dann durchzuführen, wenn diese auch valide zu interpretieren sind. Es ist allerdings unklar, wie sich die Einschränkung auf „mindestens 10 Ereignisse in einer der Subgruppen“ begründet. Dabei ist nicht nur die Einschränkung auf die Zahl 10 kritisch zu sehen, sondern auch die Einschränkung basierend auf der Zahl der Ereignisse statt der Fallzahl.</p> <p>Nimmt man beispielsweise an, es liegt eine Studie mit 2 Behandlungsarmen à 50 Patienten vor. Betrachtet man nun eine Subgruppenanalyse nach Merkmal X mit gleichmäßiger Verteilung, d.h. 25 Patienten pro Subgruppe, dann wäre in mindestens einer Subgruppe eine Ereignisrate von 40 % erforderlich, um eine Subgruppenanalyse durchzuführen. Es stellt sich hier die Frage, ob bei Fallzahlen in dem hier beschriebenen Umfang eine Einschränkung auf mindestens 10 Ereignisse in einer Subgruppe zielführend ist oder ob nicht sinnvollerweise auf Subgruppenanalysen verzichtet werden sollte, wenn nicht die Fallzahl in mindestens einer Subgruppe über 10 liegt.</p>

Abschnitt Seitenzahl	Zitat Allgemeine Methoden	Anmerkungen MSD
		<p>Im Rahmen der Nutzenbewertung muss sich zusätzlich die Frage gestellt werden, inwieweit das beschriebene Vorgehen zur Subgruppenanalyse die Ableitung des Zusatznutzens beeinflusst. Betrachtet man beispielsweise den Endpunkt <i>Schwerwiegende unerwünschte Ereignisse</i> und identifiziert insgesamt einen Nachteil zuungunsten der Intervention. Bei näherer Betrachtung in Form von Subgruppenanalysen zeigt sich jedoch, dass sich dieser Nachteil beispielsweise nur bei Patienten mit schwerer Erkrankung ergibt. Schwerwiegende unerwünschte Ereignisse treten selten auf, so dass eine Ereigniszahl über 10 in mindestens einer Subgruppe unter Umständen nicht erreicht wird. In diesem Fall würde auf eine Subgruppenanalyse verzichtet werden, auch wenn diese bestätigen würde, dass sich das Schadenspotential womöglich nur in einer Subgruppe zeigt, die ggf. auch zu der kleineren Population im Versorgungsalltag gehört. Auch hier stellt sich die Frage, ob eine regelhafte Einschränkung auf mindestens 10 Ereignisse in einer Subgruppe zielführend ist.</p> <p>Unabhängig von den zuvor angemerkten Punkten geht MSD davon aus, dass im Rahmen der Nutzenbewertung auch auf Seiten der pharmazeutischen Unternehmen Subgruppenanalysen nur dann verpflichtend darzustellen sind, wenn diese auch valide zu interpretieren sind, d.h. wenn die in den Allgemeinen Methoden formulierten Vorgaben zutreffend sind.</p>

Abschnitt Seitenzahl	Zitat Allgemeine Methoden	Anmerkungen MSD
9.3.10 Subgruppenanalysen Seite 195	„Effektmodifikation durch mehr als ein Subgruppenmerkmal (Wechselwirkung höherer Ordnung): Liegt für einen Endpunkt beispielsweise ein Unterschied sowohl zwischen 2 Altersgruppen als auch zwischen Männern und Frauen vor, so ist zur Interpretation eine Auswertung separat für jede Altersgruppe getrennt für Männer und Frauen nötig (also eine Analyse zu 4 Subgruppen). Eine solche Auswertung liegt jedoch selten vor.“	Nach Ansicht von MSD ist eine Auswertung zu 4 Subgruppen als eine post-hoc durchgeführte Subgruppenanalyse zu betrachten. Diese ist aus verschiedenen methodischen Gründen als problematisch anzusehen und kann nicht als korrekte Prüfung einer Hypothese betrachtet und interpretiert werden. Aufgrund kleinerer Fallzahlen in den Subgruppen und kleineren Ereigniszahlen in den Behandlungsarmen wäre diese Art der Analyse nicht durchführbar bzw. nicht sinnvoll interpretierbar. Post-hoc-Analysen zu 4 Subgruppen sind daher nicht dazu geeignet, definitive Schlüsse über das Ausmaß des Behandlungseffekts in einer bestimmten Subgruppe zu ziehen.

Literaturverzeichnis

- [1] Statistical Principles for clinical trials E9. ICH Harmonised Tripartite Guideline. 1998. URL:
http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf.
- [2] Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.handbook.cochrane.org.
- [3] Brockhaus AC, Bender R, Skipka G. The Peto odds ratio viewed as a new effect measure. *Stat Med* 2014; 33(28): 4861-4874.

A.1.26 – Novartis Pharma GmbH

Autoren:

- Gartner-Freyer, Daniela
- Hentschke, Christian
- Jost, Marco
- Klebs, Sven

**Stellungnahme der Novartis Pharma GmbH zum Entwurf des
IQWiG „Allgemeine Methoden Version 5.0“ vom 07.12.2016**

**Novartis Pharma GmbH
Roonstraße 25, 90429 Nürnberg**

Stellungnahme der Novartis Pharma zum Entwurf des IQWiG:

„Allgemeine Methoden Version 5.0“ vom 07.12.2016

Die Stellungnahme der Novartis Pharma GmbH (im folgenden NVS) bezieht sich auf die im Vergleich zur Version 4.2 „Allgemeine Methoden“ des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (im folgenden IQWiG) vom 07.12.2016 vorgenommenen Änderungen sowie auf einzelne grundsätzliche Aspekte der „Allgemeine Methoden“ des IQWiG.

A) Abschnitt 1.2.4 Die Bedeutung der Ergebnissicherheit

Auf Seite 8 führt das IQWiG aus, dass es in der Regel nur auf Studien mit ausreichender Ergebnissicherheit zurückgreife. Begründet wird dies mit zu zeit- und kostenintensiven Recherchen. Nach z.B. §5 der Arzneimittel-Nutzenbewertungsverordnung sind in diesem Verfahren grundsätzlich alle Evidenzstufen zu berücksichtigen. Weder die Zeit- noch die Kostenintensität werden dort als Gründe angeführt, davon abzuweichen und eine regelhafte Einschränkung vorzunehmen. Entsprechend ist auch im Methodenpapier auf die bestverfügbare Evidenz abzustellen.

B) Abschnitte 2.1 Produktspezifische Verfahrensabläufe und 2.2.7 Wissenschaftliche Beratung

Auf Seite 15 ff. beschreibt das IQWiG die unterschiedlichen Verfahrensabläufe bei seinen Produkten. Dabei erfolgt eine Differenzierung, bei welchen Produkten eine Anhörung durch das IQWiG erfolgt und bei welchen nicht.

So ist bei Rapid Reports kein öffentliches Stellungnahmeverfahren vorgesehen (S. 15). Zwar sollen Rapid Reports zeitnah finalisiert werden, es ist aber nicht nachvollziehbar, warum nicht zumindest eine öffentliche Anhörung/schriftliche Stellungnahme durchführbar ist, da es keine gesetzlich vorgegebenen Fristen für die Erstellung dieses Produkts gibt.

Bei Dossierbewertungen und Potentialbewertungen verweist das IQWiG auf die Stellungnahmeverfahren des G-BA zu den jeweiligen Vorgängen. Dies unterscheidet sich von der Kosten-Nutzen-Bewertung, die letztendlich auch vom G-BA initiiert wird. Hier sieht das Institut abweichend eine eigene Anhörung vor. Dies ist zu begrüßen und auch im Kontext mit dem Abschnitt 2.2.7 auf Seite 41 zu betrachten, in dem das IQWiG z.B. im Rahmen der Mitarbeit im EUnetHTA-Netzwerk eine wissenschaftliche Beratung anbietet. Die Erfahrung zeigt, dass im Rahmen der frühen Nutzenbewertung häufig Daten/Analysen im Rahmen des Stellungnahmeverfahrens nachgereicht werden müssen, weil die im Dossier eingereichten Daten nach Meinung des IQWiG unzureichend/unpassend seien. Die Schaffung einer Möglichkeit einer grundsätzlichen wissenschaftlichen Beratung durch das IQWiG auch außerhalb des EUnetHTA wäre daher zu begrüßen. Die entsprechende Beratung durch den G-BA bietet derzeit zwar einen groben Orientierungsrahmen, geht aber nicht immer auf Details der operativen Umsetzung ein. Zielgerichtete Analysen bzw. die Vermeidung von nicht-akzeptierten Analysen könnten auch dazu beitragen, den unnötigen Aufwand und die z.T. hohen Kosten für die jeweilige Erstellung zu vermeiden.

Bei der Kosten-Nutzen-Bewertung sieht das IQWiG auf Seite 25 eine Stellungnahmefrist von 3 Wochen vor. Betrachtet man den Umfang der Pilotprojekte (die Kosten-Nutzen-Bewertung G09-01 umfasst fast 600 Seiten und ist damit ca. zehnmal so umfangreich wie ein IQWiG-Bericht zur frühen Nutzenbewertung), ist diese Frist deutlich zu kurz angesetzt und sollte eher mindestens 6 Wochen betragen.

Auch bei den Gesundheitsinformationen ist lediglich eine nicht-öffentliche Stellungnahme vorgesehen (S. 30). Wenn ein Stellungnahmeprozess grundsätzlich vorgesehen ist, ist es nicht nachvollziehbar, warum dieser nicht auch öffentlich durchführbar ist.

Grundsätzlich ist daher nach unserer Meinung bei den unterschiedlichen IQWiG-Produkten ein öffentliches Stellungnahmeverfahren vorzusehen. Ferner ist die grundsätzliche Möglichkeit der wissenschaftlichen Beratung durch das IQWiG zu detaillierten operativen Fragestellungen wünschenswert.

C) Abschnitt 2.2.1 Einbindung von Betroffenen in die Erstellung von IQWiG-Produkten

Im Abschnitt 2.2.1 (S. 35ff) wird die Einbindung von Patienten (Betroffenen) in die Erstellung von IQWiG-Produkten thematisiert. Methodisch wird dabei, abhängig vom Produkt, eine mündliche oder schriftliche Konsultation oder keine Einbindung der Betroffenen formuliert. Diese Einbindung dient dabei „.....primär zu Beginn der Berichtsbearbeitung der Festlegung patientenrelevanter Endpunkte und relevanter Subgruppen“. Der Stellenwert und die Berücksichtigung von Patientenpräferenzstudien, wie sie das IQWiG z.B. für Patienten mit Hepatitis C (Conjoint Analyse) und Depression (Analytic Hierarchy Process) selbst durchgeführt und publiziert hat, wird dabei nicht konkretisiert. Aus unserer Sicht stellen Patientenpräferenzstudien eine valide Methode dar, um z.B. im Rahmen der Verfahren der Nutzenbewertung nach § 35a SGB V a) relevante Endpunkte zu bestimmen und b) eine Gewichtung der Endpunkte a priori festzulegen.

Entsprechend fehlt aus unserer Sicht in den vorliegenden Methoden eine Beschreibung zur Anwendung/Berücksichtigung von Patientenpräferenzstudien durch das IQWiG. Durch die Berücksichtigung von Patientenpräferenzstudien und den daraus gewonnen Ergebnissen würde die Beteiligung von Betroffenen/Patienten einen höheren Stellenwert als im bisherigen Verfahren erreichen. Die vorliegende Methodik ist entsprechend zu ergänzen, um beim Vorliegen methodisch adäquat durchgeführter Patientenpräferenzstudien die Bewertung aus Patientensicht angemessen zu berücksichtigen. Zusätzlich sollten Mindestanforderungen an Patientenpräferenzstudien definiert werden, damit diese bei der Planung und Durchführung berücksichtigt werden können. Zusammenfassend würden die im Rahmen von Patientenpräferenzstudien identifizierten patientenrelevanten Endpunkte und deren Gewichtung – im Gegensatz zum bisherigen Verfahren – auf der Basis eines systematischen und transparenten Ansatzes in die Bewertung einfließen.

Der Abschnitt beschreibt ferner die unterschiedlichen Produkte, bei denen Betroffene eingebunden werden können. Im Rahmen der Transparenz wäre für die Einbindung Betroffener im Methodenpapier klarzustellen, dass die Ergebnisse der Einbeziehung öffentlich in anonymisierter Weise zur Verfügung gestellt werden. Gleiches gilt analog für die Beiträge der externen Sachverständigen zu den unterschiedlichen IQWiG-Produkten. Es ist nicht nachvollziehbar, warum diese Dokumente nicht in die öffentliche Bewertung eingehen.

D) Abschnitt 3.1.4 Endpunktbezogene Bewertung

Zur Ableitung eines Belegs aus einer einzelnen Studie werden in Abschnitt 3.1.4 besondere Anforderungen definiert. Während die Mehrzahl dieser Anforderungen plausibel zur Einschätzung der Ergebnissicherheit von Studienergebnissen beitragen, erscheint die Anforderung von mindestens 1000 Patienten pro Studienarm zu undifferenziert und nicht transparent.

Zum einen werden wesentliche Kriterien für eine hohe Belastbarkeit von Studienergebnissen, die z.B. die interne und externe Validität einer Studie betreffen, in anderen Abschnitten als qualitative Ergebnissicherheit ausgeführt, finden jedoch im Abschnitt zur Ableitung eines Belegs aus einer einzelnen Studie keine explizite Beachtung. Zum anderen steht die pauschale Forderung von 1000 Patienten pro Studienarm im Kontrast zur ausgewogenen Berücksichtigung von medizinischen, ethischen und ökonomischen Anforderungen an eine Fallzahlplanung. In einer geeigneten Fallzahlplanung ist die Größe einer Studie durch die Größe des zu erwartenden Effektes der Power und des geforderten Signifikanzniveaus reguliert. Eine statistische Sicherheit des Ergebnisses (an anderer Stelle im Methodenpapier als quantitative Ergebnissicherheit bezeichnet) ist somit ausreichend durch das festgelegte nominelle Signifikanzniveau geregelt und bedarf aus wissenschaftlicher Sicht keiner zusätzlichen Regulierung. Für die Ableitung eines Belegs aus einer einzelnen Studie sollte eine Regulierung der Studiengröße also ausschließlich relativ durch ein strengeres nominelles Signifikanzniveau erfolgen. Zusätzlich ist hier ist das geforderte Niveau von $p < 0,001$ nicht allgemein etabliert und scheint zu konservativ. Frühere Versionen des Methodenpapiers referenzierten hier das EMA Dokument (CPMP/EWP/2330/99, Points to consider on application with 1. Meta-Analyses; 2. one pivotal study').

Nach Ansicht von Novartis ist eine Orientierung an dem o.g. EMA Dokument wesentlich sinnvoller, um einzuschätzen, ob eine einzelne Studie zur Ableitung eines Belegs geeignet ist.

E) Abschnitt 4.3.4 Datengrundlage

In Abschnitt 4.3.4 wird beschrieben, wie die Datengrundlage für die Kosten-Nutzen-Bewertung ermittelt werden soll. Hierbei wird bei der Literaturrecherche bei der Gesamtbewertung des Nutzens auf eine fokussierte Recherche abgestellt. Es ist unklar, warum nicht eine systematische Recherche nach Abschnitt 8.1. und 8.2 durchgeführt wird, wie auch bei der Beschaffung der einzelnen Nutzenparameter.

F) Abschnitte 5.2 Identifikation von Versorgungsstandards und 5.3 Empfehlungen für Disease Management Programme

Im Abschnitt 5.3 (S. 119) beschreibt das IQWiG die Auswahl von Empfehlungen basierend auf Grade of Recommendation (GoR) und Level of Evidence (LoE). Es beschreibt ferner im Abschnitt 5.2 (Seite 115), dass z.B. die Sichtung evidenzbasierter Leitlinien auf Basis formaler Kriterien erfolgt, und eine inhaltliche Prüfung nicht vorgenommen wird. Dies ist am Ende vor allem relevant für die Ableitung von Empfehlungen für Disease-Management-Programme (Abschnitt 5.3). Durch die strikte formale Umsetzung einer Bewertung nur aufgrund der Kriterien GoR und LoE kann es zu Schlussfolgerungen

kommen, die unter Betrachtung der konkreten Versorgungssituation nicht plausibel sind. Gerade bei Fragestellungen, bei denen es nur eine eingeschränkte Evidenz gibt, führt die gewählte hohe Hürde dazu, dass die bestverfügbare Evidenz gar nicht weiter bewertet wird. Liegt zu speziellen Fragestellungen nur eine eingeschränkte Evidenz vor, führt die Zensierung der bestverfügbaren Evidenz bei diesen Sachverhalten dazu, dass am Ende gar keine DMP-relevanten Empfehlungen übrigbleiben. Es ist davon auszugehen, dass betroffene Patienten eher davon profitieren würden, wenn bei komplexen Krankheitsbildern im DMP nach der bestverfügbaren Evidenz behandelt würde, als gar keine Informationen zur Behandlung im DMP zur Verfügung zu stellen. Insgesamt ergibt sich aus dem beschriebenen Vorgehen, dass die isolierte Betrachtung von GoR und LoE als alleinige Faktoren nicht zielführend ist. Eine Bewertung, ob eine Empfehlung DMP-relevant ist oder nicht, kann nur unter Berücksichtigung des Gesamtversorgungskontextes erfolgen. Eine inhaltliche Beurteilung bei der Auswahl ist daher nach unserer Meinung zwingend erforderlich.

G) Abschnitt 5.2.2 Übertragbarkeit auf das deutsche Gesundheitssystem

In Abschnitt 5.2.2 wird beschrieben, dass bei der Recherche zu internationalen Leitlinien grundsätzlich davon ausgegangen wird, dass Leitlinien aus Mitgliedsstaaten der *Organisation for Economic Co-operation and Development (OECD)* von besonderer Relevanz sind. NVS geht davon aus, dass in begründeten Fällen außerdem weitere Länder in der Recherche ergänzt werden können bzw. Evidenz/Leitlinien aus weiteren Ländern im Rahmen des Stellungnahmeverfahrens eingereicht werden können.

H) Abschnitt 8.1.1 Suche in bibliographischen Datenbanken

Bei den zur Suche herangezogenen Datenbanken benennt das IQWiG im Unterabschnitt D) PubMed als akzeptierte Oberfläche. Seit Mai 2016 enthält Medline via Ovid auch folgende Komponenten: „Publisher, In-Data-Review, In-Process and PubMed-not-MEDLINE records from NLM.“

Somit ist der Inhalt einer Recherche über Ovid gleichwertig zu einer solchen mit Pubmed (via NLM) und kann unserer Meinung nach alternativ dazu verwendet werden. Zur Vermeidung einer unnötigen Doppelrecherche sollte nach unserer Meinung klargestellt werden, dass eine explizite Recherche über Pubmed in diesen Fällen obsolet ist und auch andere Oberflächen wie Ovid, die die entsprechende Aktualität aufweisen, alternativ herangezogen werden können.

I) Abschnitt 8.2. Fokussierte Informationsbeschaffung

Im Abschnitt 8.2 werden bei der Suchstrategie mögliche Einschränkungen bzw. Anpassungen aufgeführt. Aus unserer Sicht sollte es darüber hinaus grundsätzlich akzeptiert werden, bestimmte Publikationstypen (z.B. Kongress-Abstracts) durch NOT-Suche zu eliminieren, wenn diese in der Folge bei der Sichtung/Bewertung des Rechercheergebnisses ohnehin aussortiert werden. So stellt die vom IQWiG selbst herangezogenen Referenz 186 dar, dass die umfassende Suche von Studien unter Abstracts bei Kongressen und Präsentationen zeitaufwändig und von fraglichem Nutzen sei.

J) **Abschnitt 9.3.9 Indirekte Vergleiche**

Die Anforderungen des IQWiG an indirekte Vergleiche sind im Abschnitt 9.3.9 zusammengefasst. Hier wird u.a. ausgeführt: „Es werden ausschließlich adjustierte indirekte Vergleiche über adäquate Brückenkomparatoren akzeptiert. Hierzu zählen insbesondere das Verfahren von Bucher et al. [88] sowie die oben erwähnten Methoden der Netzwerk-Meta-Analysen.“ Aus unserer Sicht sollte die vorliegende Methodik nicht explizit bestimmte Formen von indirekten Vergleichen ausschließen, sondern es ermöglichen, neue Ansätze und Verfahren im Rahmen der Bewertung einzureichen. Eine Nichtberücksichtigung in der Bewertung durch das IQWiG sollte begründet erfolgen. Dieses Verfahren würde es ermöglichen, auch wissenschaftliche neue Ansätze in einem Verfahren prüfen zu lassen. Eine Methode, die möglicherweise sonst alleine aufgrund dieser Einschränkungen nicht berücksichtigt werden würde, sind die Matching Adjusted Indirect Comparisons (MAIC). MAIC benötigt nicht zwangsläufig einen Brückenkomparator, da es Ziel der Methode ist eine Adjustierung von Unterschieden in den Patientenpopulationen über Studien hinweg vorzunehmen. Dies erlaubt es z.B. individuelle Daten aus einer Studie gegenüber publizierten aggregierten Daten hinsichtlich ausgewählter Patientencharakteristika anzupassen. Somit wird eine homogene Patientenpopulation über beide Studien hinweg erreicht. Die Ergebnisse der Studie, aus der die individuellen Daten stammen, können dann gewichtet neu berechnet werden und die Endpunkte der Studien verglichen werden. Es ist uns bekannt, dass auch dieses Verfahren eines indirekten Vergleichs Limitationen aber auch Stärken besitzt. Eine generelle Ablehnung dieser Methode ist uns wissenschaftlich nicht plausibel – wie bei allen indirekten Vergleichen sollte untersucht werden, ob es mögliche Schwächen gibt, die zur Ablehnung des Testes führen oder aber bei mangelnder Ergebnissicherheit zu einer Abwertung führen, wie es die IQWiG Methodik vorsieht.

K) **Abschnitt 9.3.10 Subgruppenanalysen**

In Abschnitt 9.3.10 wird das Vorgehen für die Durchführung von Subgruppenanalysen beschrieben. Es wird dabei auch auf methodische Schwächen von Subgruppenanalysen bei der Ableitung relevanter Entscheidungen für den zielgerichteten Einsatz medizinischer Interventionen eingegangen. Konkret werden hier die häufig zu geringe Power, die oft nicht gegebene Vergleichbarkeit der Behandlungsgruppen innerhalb der Subgruppen und eine mögliche Nichtberücksichtigung von Effektmodifikation durch mehr als ein Subgruppenmerkmal genannt. Weiter unten im Abschnitt wird darauf hingewiesen, dass post hoc auf Studienebene durchgeführte Subgruppenanalysen kritisch zu interpretieren sind. Es fehlt hier jedoch der berechnete Hinweis auf die multiple Testthematik, der in der Version 4.2 des Methodenpapiers noch enthalten ist („Bei der Analyse zahlreicher Subgruppen besteht eine manchmal recht hohe Wahrscheinlichkeit, dass die Ergebnisse irgendeiner Subgruppe statistische Signifikanz erreichen, obwohl es sich in Wahrheit um ein zufälliges Ergebnis handelt“). Da sich an dieser Problematik nichts verändert hat und diese auch nicht mit der Problematik einer post hoc Durchführung von Subgruppenanalysen gleichzusetzen ist, wäre es sinnvoll, einen entsprechenden Hinweis wieder aufzunehmen. Ein diesbezüglicher Hinweis wäre zudem konsistent mit der Vorgehensweise, nur noch Belege für Subgruppeneffekte zum Niveau $\alpha = 0,05$ zu berücksichtigen.

Zusätzlich ist zu bedenken, dass die Forderung der Darstellung von Effektmodifikationen höherer Ordnung bei mehr als zwei simultanen Effektmodifikatoren in der Praxis sehr schnell ziemlich umfangreich werden kann und an die Grenze der sinnvollen Interpretierbarkeit stößt. Vor dem

Hintergrund kleiner werdender Fallzahlen hat das IQWiG an anderer Stelle im Dokument eine Ereigniszahl von mindestens 10 Ereignissen pro Merkmalskombination für eine sinnvolle Interpretierbarkeit einer Effektmodifikation durch Subgruppen festgelegt. Davon abgesehen, existiert hier jedoch kein klares Kriterium, bis zu welcher Ordnung Effektmodifikation geprüft und dargestellt werden sollte.

Bei der Gesamtbetrachtung der ausgeführten methodischen Schwächen von Subgruppenanalysen, insbesondere im Fall von nur einer Studie, scheint diese Methodik im Allgemeinen keine hohe Ergebnissicherheit bei der Einschätzung von differenziellem Nutzen bzw. Schaden von medizinischen Interventionen erzielen zu können. Aufgrund der dargestellten Problematiken können hier häufig Fehleinschätzungen in beide Richtungen auftreten. Andererseits ist natürlich nachvollziehbar, dass die durch Subgruppenanalysen generierte Evidenz, oft die beste verfügbare Evidenz darstellt. Dies wird in entsprechendem Abschnitt weiter unten vom IQWiG aufgegriffen und als Begründung angeführt, um Bewertungen trotzdem auf Ergebnisse aus Subgruppenanalysen stützen zu können. Wie im Methodenpapier dargestellt, wird jedoch zur Bewertung nicht einheitlich und regelhaft die beste verfügbare Evidenz berücksichtigt. In Kapitel 7.4. wird z.B. beschrieben, dass nur Metaanalysen, die auf RCTs basieren, mit in die Entscheidungsfindung mit einbezogen werden sollten. Ähnliches gilt für den oben ebenfalls genannten Abschnitt 1.2.4. Im Hinblick auf die Tragweite von abgeleiteten Bewertungen, sollte im Methodenpapier mit der Berücksichtigung bester verfügbarer Evidenz jedoch einheitlich umgegangen werden. Wir möchten das IQWiG bitten, in Betracht zu ziehen, sich in allen Bereichen auf die beste verfügbare Evidenz zu stützen, natürlich mit der Einschränkung einer zurückhaltenden Interpretierbarkeit der Ergebnisse bei reduzierter Ergebnissicherheit.

Literatur:

CPMP. Points to consider with 1. meta-analyses; 2. one pivotal study. 31 May 2001.

http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003657.pdf

A.1.27 – Pfizer Deutschland GmbH

Autoren:

- Kühnast, Sarah
- Leverkus, Friedhelm

Stellungnahme der Pfizer Deutschland GmbH zum Entwurf des IQWiG

„Allgemeine Methoden Version 5.0 (ENTWURF)“ vom 07.12.2016

Die Pfizer Deutschland GmbH möchte positiv hervorheben, dass das IQWiG sich dazu entschieden hat, seinen Entwurf für Version 5.0 des Methodenpapier und die darin enthaltenen Änderungen gegenüber der Version 4.2 zur wissenschaftlichen Diskussion in Form eines Stellungnahmeprozesses zu öffnen. Daher möchten wir uns gerne an der Diskussion beteiligen und reichen folgende Stellungnahme ein.

Im Rahmen der Stellungnahme werden dabei folgende Punkte kommentiert :

- Verwendung adäquater Methoden für Indirekte Vergleiche
- Meta-Analysen
- Subgruppenanalysen

1 Verwendung adäquater Methoden für Indirekte Vergleiche

Das IQWiG beschreibt in Abschnitt 9.3.9 Methoden für indirekte Vergleiche und deren Anwendung im Rahmen von Bewertungen des Nutzens von Arzneimitteln mit neuen Wirkstoffen sowie bei Kosten-Nutzen-Bewertungen. Die Anwendung nicht-adjustierter indirekter Vergleiche (d. h. die Verwendung einzelner Arme aus verschiedenen Studien) lehnt das IQWiG grundsätzlich ab.

- Wir möchten darauf hinweisen, dass die in Abschnitt 9.3.9 beschriebene grundsätzliche Ablehnung nicht-adjustierter indirekter Vergleiche etwas unklar ist. Im Falle des Vorliegens eines dramatischen Effekts, welcher in Abschnitt 3.2.2 erläutert wird, können auch nicht-adjustierte indirekte Vergleich akzeptiert werden. Auch wenn es sich dabei um Spezialfälle handelt, wäre die Ergänzung eines entsprechenden Hinweises bzw. Verweises in Abschnitt 9.3.9 wünschenswert.

Im Vergleich zur Version 4.2 führt das IQWiG im Zuge der grundsätzlichen Ablehnung nicht-adjustierter indirekter Vergleiche insbesondere Matching-Verfahren auf. Als Beispiel wird die Methode des *Matching Adjusted Indirect Comparison* (MAIC; Signorovitch et al., 2010) referenziert.

- Generell stimmen wir damit überein, dass Ergebnisse aus einem MAIC bei fehlendem Brückenkompator eine verringerte Ergebnissicherheit gegenüber adjustierten indirekten Vergleichen über einen gemeinsamen Brückenkompator aufweisen und vergleichbar mit dem Niveau einer Beobachtungsstudie sind.
- Unserer Ansicht nach erscheint der Ansatz des MAIC jedoch im Vergleich zu anderen bekannten Verfahren in der Situation eines fehlenden Brückenkompators (z.B. naive indirekte Vergleiche, Modellierungsverfahren) am geeignetsten zur Schätzung des Effekts und sollte nicht kategorisch abgelehnt werden. Es sollte vielmehr geprüft werden, ob ein indirekter Vergleich unter Verwendung der MAIC-Methode als adäquat zur Ableitung des Zusatznutzens angesehen werden kann. Beispielsweise wäre eine Anwendung des MAIC in der Situation eines dramatischen Effekts, d.h. unter Berücksichtigung der Effektgröße, denkbar und sinnvoll. Im Vergleich zum naiven indirekten Vergleich können mit Hilfe des MAIC Verzerrungsaspekte zumindest teilweise berücksichtigt werden.
- Eine weitere Situation, in denen die Anwendung des MAIC als hilfreich angesehen werden kann, ist das Vorliegen einarmiger Studien wie es bei Orphan Drugs und Conditional Approvals der Fall sein kann. Anstatt diese Studien gar nicht anzusehen, könnte mit Hilfe des MAIC unter Berücksichtigung der Vergleichbarkeit prognostischer und klinisch relevanter

Merkmale (Effektmodifikatoren) und potenzieller Verzerrungsaspekte hilfreiche Information gewonnen werden. Je nach Größe des Effekts sollte auch hier geprüft werden, ob der MAIC als adäquates Verfahren zur Ableitung des Zusatznutzens dargestellt werden kann.

- Zudem sollte das MAIC-Verfahren auch als mögliche Alternative zum Verfahren nach Bucher et al. in der Situation eines gemeinsamen Brückenkomparators in Betracht gezogen werden, falls die zu vergleichenden Studien als nicht hinreichend ähnlich bzgl. relevanter Effektmodifikatoren angesehen werden (Kühnast et al.). Mit Hilfe des MAIC ist die Berechnung eines adjustierten Effektschätzers möglich ohne die Randomisierung innerhalb der im indirekten Vergleich berücksichtigten Studien zu brechen.

2 Meta-Analysen

Das IQWiG beschreibt in Abschnitt 9.3.8 A) „Allgemeines“ Modelle zur Schätzung des Gesamteffekts und der Konfidenzintervalle im Rahmen einer Meta-Analyse. Im Unterschied zur vorigen Version 4.2 der Allgemeinen Methoden empfiehlt das IQWiG bei der Durchführung von Meta-Analysen mit zufälligen Effekten nicht mehr die Methode nach DerSimonian & Laird, sondern die Knapp-Hartung-Methode.

Zudem adressiert das IQWiG Limitationen der Anwendung von Meta-Analysen mit zufälligen Effekten für den Fall sehr weniger Studien (weniger als 5). Bisher sollten vorrangig Modelle mit zufälligen Effekten angewendet werden und nur in begründeten Ausnahmefällen auf Modelle mit festen Effekten zurückgegriffen werden, wobei unklar blieb, was das IQWiG als begründeten Ausnahmefall ansah. Für den Fall sehr weniger Studien soll nun die Anwendung des Modells mit festen Effekten oder eine qualitative Zusammenfassung in Betracht gezogen werden.

- Wir begrüßen die neuen methodischen Entwicklungen und deren Berücksichtigung im Entwurf der Version 5.0 der Allgemeinen Methoden sehr.
- Im Fall sehr weniger Studien (weniger als 5) weisen Modelle mit zufälligen Effekten Limitationen auf, die zu sehr breiten (zu konservativen) Konfidenzintervallen führen können. In diesem Zuge möchten wir zudem darauf hinweisen, dass die Verwendung von Modellen mit zufälligen Effekten für Meta-Analysen im Kontext der Nutzenbewertung in Anbetracht des zugrunde liegenden Modellgedankens eher fraglich scheint. Üblicherweise werden in der Nutzenbewertung Meta-Analysen auf Basis der Studiensituation der Zulassung durchgeführt. Liegt dabei mehr als eine Studie vor, so handelt es sich häufig um Replikationen eines Studienprotokolls in einem sehr ähnlichen Setting. In dieser Situation kann unsere Ansicht entsprechend von einem zugrunde liegenden Modell mit festen Effekten ausgegangen werden. In der Nutzenbewertung von Arzneimitteln liegen in der Regel nur sehr wenige Studien vor. In allen bisherigen Bewertungsverfahren, in denen Meta-Analysen durchgeführt wurden, basierten diese auf zwei oder drei Studien. Demzufolge ist die Anwendung von Modellen mit festen Effekten im Rahmen von Nutzenbewertungen eher als Standard als als Ausnahme anzusehen. Die Formulierung „sollte in Erwägung gezogen werden“ sollte unter diesem Hintergrund präziser formuliert werden.

3 Subgruppenanalysen

Das IQWiG beschreibt in Abschnitt 9.3.10 die Änderung der Schwellenwerte für den Test auf Effektmodifikationen. Die bisher verwendete Hinweis-Schwelle ($\alpha = 0,2$) wird gestrichen. Es verbleibt der Schwellenwert von $\alpha = 0,05$ für einen Beleg auf Effektmodifikation.

- Wir erachten die Streichung des Schwellenwerts für einen Hinweis auf Interaktion als sinnvolle und positive Entwicklung.
- Zudem möchten wir darauf hinweisen, dass die Bezeichnung „post hoc“ für ein einzelnes klinisches Protokoll im Sinne konfirmatorischer Analysen von Bedeutung ist, im Rahmen der Nutzenbewertung spielt die Differenzierung zwischen „präspezifiziert“ und „post hoc“ aus unserer Sicht jedoch keine Rolle. Die im AMNOG aufgeführten Subgruppen können zwar für eine klinische Prüfung post hoc sein aufgrund fehlender Präspezifizierung im Protokoll, sind aber im Rahmen der Nutzenbewertung präspezifiziert.
- Wir begrüßen, dass das IQWiG die Problematik des multiplen Testens nicht in den Vordergrund stellt. Die Nutzenbewertung sieht sich in diesem Kontext mit anderen Problemen konfrontiert als eine klinische Prüfung. Dennoch sehen wir, dass mit der Vielzahl an Interaktionstest, die im Rahmen der Nutzenbewertung regelmäßig durchgeführt werden, die Rate der falsch-positiven Ergebnisse steigt. Es ist fraglich, ob der klassische Interaktionstest, der einen festen Platz bei klinischen Prüfungen hat, hier sinnvoll ist. Es sollte daher geprüft werden, ob alternative Verfahren (z.B. Random Forrests) in dieser speziellen Situation der Fragestellung nicht näher kommen. Gerade im Hinblick auf mögliche zukünftige Implikationen von Interpretationen von Subgruppenergebnissen durch das anstehende Arzneimittelversorgungsstärkungsgesetz, sollte die Glaubwürdigkeit dieser geprüft und weitere Forschung zu alternativen Verfahren betrieben werden.

Referenzen:

Bucher HC, Guyatt GH, Griffith LE, Walter SD (1997). The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol*; 50(6): 683-691.

IQWiG Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2015). Allgemeine Methoden Version 4.2 vom 22.04.2015. Abrufbar unter https://www.iqwig.de/download/IQWiG_Methoden_Version_4-2.pdf

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2016). Allgemeine Methoden - Entwurf für Version 5.0 vom 07.12.2016. Abrufbar unter https://www.iqwig.de/download/IQWiG_Methoden_Entwurf-fuer-Version-5_final.pdf

Kühnast S, Schiffner-Rohe J, Rahnenführer J, Leverkus F. Evaluation of adjusted and unadjusted indirect comparison methods in benefit assessment - a simulation study for time-to-event endpoints. *Methods of Information in Medicine*; accepted for publication.

Signorovitch JE, Wu EQ, Yu AP, Gerrits CM, Kantor E, Bao Y, Gupta SR and Mulani PM (2010). Comparative Effectiveness Without Head-to-Head Trials: A Method for Matching-Adjusted Indirect Comparisons Applied to Psoriasis Treatment with Adalimumab or Etanercept. *Pharmacoeconomics*; 28(10): 935-945.

A.1.28 – Roche Diabetes Care GmbH

Autoren:

- Berndt, Karsten
- Casuso, Antonio
- Kraeling, Birgit



Allgemeine Methoden des Instituts für Wirtschaftlichkeit und Qualität im Gesundheitswesen (IQWiG)

Entwurf für Version 5.0 vom 07.12.2016:

**Stellungnahme der Roche Diabetes Care GmbH
vom 31.1.2017**

**Dr. Christiane Arras-Friederich
Head of Professional Relations**

Roche Diabetes Care Deutschland GmbH
DCEGR...6919
Sandhofer Strasse 116
68305 Mannheim/Germany

Phone: +49 621 759 [REDACTED]
[REDACTED]

**Karsten Berndt
Global Market Access Health Technology Assessment (HTA)**

Roche Diabetes Care GmbH
DCMRM...6164
Sandhofer Strasse 116
68305 Mannheim/Germany

Phone: +49 621 759 [REDACTED]

Fax: +49 621 759 [REDACTED]

Mobile: [REDACTED]
[REDACTED]

www.roche.com

ACCU-CHEK®

Stellungnahme der Roche Diabetes Care GmbH

Die Roche Diabetes Care GmbH bedankt sich für die Möglichkeit, in diesem Stellungnahmeverfahren zu der Weiterentwicklung der Allgemeinen Methoden des Instituts beitragen zu dürfen.

Das Institut übt im Rahmen seiner Aufgaben einen maßgeblichen Einfluss auf die Erstellung von wissenschaftlicher Evidenz aus.

Für die Ersteller solcher wissenschaftlicher Evidenz ist die bestmögliche Transparenz der Anforderungen zum Zeitpunkt der Studienplanung entscheidend. Demgegenüber legt das Institut im Laufe der Projekte wesentliche Anforderungen an die Evidenz fest.

Neben den bestehenden Aktivitäten und Vorgehensweisen im Rahmen der wissenschaftlichen Beratung (Kapitel 2.2.7) möchten wir vorschlagen, zu prüfen, ob weitere Möglichkeiten bestehen, die Transparenz der Studienanforderungen zu steigern, zumal die Kapazitäten für wissenschaftliche Beratung begrenzt sind.

So könnte das Institut Surrogatendpunkte und kombinierte Endpunkte, die das IQWiG generell akzeptiert, auf der Website des IQWiG in aktueller Form veröffentlichen.

Auch Relevanzkriterien spielen für die Evidenz-Generierung eine maßgebliche Rolle: So kann die in Kapitel 9.4 erwähnte Frage „Wie erleben Menschen diese Erkrankung / diese Intervention?“ auch in Bezug auf die klinische Relevanz bedeutende Anhaltspunkte geben. Insofern können die qualitativen Methoden neben der technischen Bestimmung von klinischen Relevanzgrenzen (Kapitel 3.1.1 und 9.3.3) wichtige inhaltliche Informationen für die Interpretation klinischer Relevanz beitragen.

A.1.29 – Roche Pharma AG

Autoren:

- Knoerzer, Dietrich

Stellungnahme zum Entwurf des IQWiG Methodenpapiers 5.0 durch die Roche Pharma AG

Themenkomplex: Belege aus einer Studie Abschnitt 3.1.4

Aus lediglich 1 Studie kann im Ausnahmefall für einen Endpunkt ein Beleg für eine spezifische (Teil-) Population abgeleitet werden, wenn ein Studienbericht gemäß 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 mindestens 1000 Patienten eingeschlossen* und es gibt mindestens 10 Zentren.

Kommentar:

- Zur Ableitung eines Belegs aus einer einzelnen Studie müssen besondere Anforderungen erfüllt sein. Frühere Versionen des Methodenpapiers referenzierten hier das EMA Dokument (CPMP/EWP/2330/99 ‚Points to consider on application with 1. Meta-Analyses; 2. one pivotal study‘). Hier werden (zu Recht) eine Vielzahl von Kriterien genannt, die für eine solche Studie erfüllt sein müssen, um diese als belastbar anzusehen.
- Das Methodenpapier 5.0 reduziert diese Kriterienzahl und definiert für 2 dieser Kriterien fixe Größen: Zentrenzahl ≥ 10 und Fallzahl ≥ 1000 .
 - Während das erste Kriterium mit einer Zentrenzahl ≥ 10 eine Konkretisierung des EMA-Dokuments in sinnvollem Größenrahmen darstellt, um die Variabilität zwischen Zentren zu beschreiben, ist das zweite Kriterium mit einer absoluten Fallzahl ≥ 1000 nicht nachvollziehbar. In einigen Indikationen wie z.B. der Onkologie sind Studien einer Fallzahl ≥ 1000 eher die absolute Ausnahme.
 - Die Fallzahlplanung soll in Abhängigkeit der gewünschten Power und für die entsprechende Fragestellung vorgenommen werden, um eine geeignete Fallzahl festzulegen. Eine zu kleine oder zu große Fallzahl ist weder medizinisch, ökonomisch noch ethisch vertretbar.
- Aus Sicht von Roche ist eine Orientierung an dem o.g. EMA Dokument wesentlich sinnvoller, um einzuschätzen, ob eine einzelne Studie zur Ableitung eines Belegs geeignet ist. Gleichzeitig wäre für diese Kriterien eine Konkretisierung (in sinnvollem und realisierbarem Rahmen) hilfreich.

Themenkomplex indirekte Vergleiche

Abschnitts 9.3.9

Folgende Passagen wurden in der Version 5.0 ergänzt bzw. geändert:

„Dies gilt auch für Methoden für indirekte Vergleiche, in denen über Modellierungen mit starken Annahmen über die unbekanntes Effekte [107] oder mithilfe von Verfahren aus dem Bereich der Kausalmodelle für Beobachtungsstudien mit untestbaren Annahmen [596] versucht wird, Effektschätzungen trotz fehlender Brückenkomparatoren zu ermöglichen. Es werden ausschließlich adjustierte indirekte Vergleiche über adäquate Brückenkomparatoren akzeptiert. Hierzu zählen insbesondere das Verfahren von Bucher et al. [88] sowie die oben erwähnten Methoden der Netzwerk-Meta-Analysen.“

Indirekte Vergleiche über einen Brückenkomparator sind solchen ohne Brückenkomparator vorzuziehen. In Fällen, in denen es nicht möglich ist, einen indirekten Vergleich über einen Brückenkomparator durchzuführen, sind diese indirekten Verfahren jedoch die einzige Methode um weitere Evidenz zu generieren. Für diese Verfahren sollte die Möglichkeit bestehen, sie bei einer dokumentierten Eignung auch für die Nutzenbewertung heranziehen zu können.

Matching-Verfahren via IPD

Dieser Themenkomplex wird im Methodenpapier 5.0 nicht behandelt. Hier seien beispielsweise Propensity Score Verfahren erwähnt. Es wäre wünschenswert, dass dieses Themenfeld ebenfalls bewertend dargestellt wird.

Matching-Verfahren nicht via IPD

Es sollte im Methodenpapier 5.0 dargestellt werden, welche Methoden für einen indirekten Vergleich ohne Brückenkomparator und ohne IPD zur Ableitung des Zusatznutzens akzeptabel sind.

Themenkomplex: Subgruppenanalysen (Zusammenführung und Aktualisierung der vorherigen Abschnitte zu Subgruppenanalysen)

Abschnitt 9.3.10

Zwei Aspekte sollen innerhalb dieses Kapitels betrachtet werden:

Multiples Testen

Folgende Passage aus der Version 4.2 findet sich in der Version 5.0 nicht wieder:

„Bei der Analyse zahlreicher Subgruppen besteht eine manchmal recht hohe Wahrscheinlichkeit, dass die Ergebnisse irgendeiner Subgruppe statistische Signifikanz erreichen, obwohl es sich in Wahrheit um ein zufälliges Ergebnis handelt“.

Da dieser Hinweis auf die multiple Testthematik vollkommen berechtigt ist, bitten wir das IQWiG, diese Passage in die Version 5.0 ebenfalls aufzunehmen.

Verlässlichkeit von Subgruppenanalysen

Folgende Passage wurde in der Version 5.0 ergänzt:

„Einerseits erfordern die oben aufgeführten Aspekte die Bewertung der Verlässlichkeit von Subgruppenanalysen; Sun et al. [627] identifizieren hierzu Kriterien. So sind beispielsweise Subgruppenanalysen zu Merkmalen, die nicht vor der Randomisierung, sondern im Studienverlauf erhoben wurden (beispielsweise Patienten mit vs. solche ohne Herzinfarkt unter der untersuchten Therapie), grundsätzlich nicht verlässlich. Andererseits können trotz dieser Einschränkungen Subgruppenanalysen für manche Fragestellungen die auf absehbare Zeit beste wissenschaftliche Evidenz für die Bewertung von Effekten in ebendiesen Subgruppen darstellen [237], da beispielsweise ethische Überlegungen gegen die Verifizierung beobachteter Effekte in weiteren Studien sprechen können“

Roche begrüßt die Ergänzung in Hinblick auf die Verlässlichkeit von Subgruppenanalysen, einschließlich des Verweises auf die Publikation von Sun (BMJ 2010). In einer späteren Arbeit des selben Autors (BMJ, 2012) werden weitere wichtige Kriterien zur Beurteilung der Glaubwürdigkeit von Subgruppenanalysen genannt, insbesondere:

- Das Vorhandensein indirekter Evidenz, z.B. eine biologische Rationale, Laboruntersuchungen oder Tierversuche, welche den scheinbaren Subgruppeneffekt stützt
- Die Anzahl von getesteten Hypothesen in Hinblick auf Subgruppenanalysen
- Die Konsistenz des beobachteten Subgruppeneffektes zu Ergebnissen vorangegangener Studien
- Die Konsistenz des beobachteten Subgruppeneffektes über ähnliche Endpunkte hinweg

Wir bitten das IQWiG, diese wichtigen Aspekte ebenfalls mit aufzunehmen.

A.1.30 – Sanofi-Aventis Deutschland GmbH

Autoren:

- Kienitz, Carsten
- Rosenfeld, Stephanie
- Schinzel, Stefan
- Theobald, Karlheinz

Stellungnahme der Sanofi-Aventis Deutschland GmbH zum IQWiG Methodenpapier Entwurf für Version 5.0

31.01.2017

Zum Stellungnahmeverfahren allgemein:

Am 07.12.2016 wurde das schriftliche Stellungnahmeverfahren zur Entwurfsversion 5.0 „Allgemeine Methoden“ eingeleitet¹. Auf seiner Homepage beschreibt das IQWiG das weitere Procedere wie folgt: *„Das Institut hat den Entwurf für den nächsten Überarbeitungsschritt seiner Allgemeinen Methoden vorgelegt: Interessierte Personen und Institutionen können eine Stellungnahme zu den inhaltlichen Aktualisierungen und Ergänzungen im Entwurf der Version 5.0 abgeben. Stellungnahmen zu der hier vorgestellten Methodik können Sie bis zum 31. Januar 2017 (12:00 Uhr) an die folgende E-Mail-Adresse schicken: methoden@iqwig.de.“*² Wie bereits im Verfahren zur Überarbeitung des Dokuments „Allgemeine Methoden Version 4.1“ im Jahr 2013 und „Allgemeine Methoden Version 4.2“ im Jahr 2014 geschehen, ist auch hier eine mündliche Anhörung im Nachgang zum schriftlichen Stellungnahmeverfahren unerlässlich. Neben der Verpflichtung des IQWiG gemäß § 139a SGB V, in regelmäßigen Abständen über Arbeitsprozesse und -ergebnisse einschließlich der Grundlagen für die Entscheidungsfindung öffentlich zu berichten, fördert solch ein Anhörungsverfahren den kontinuierlichen wissenschaftlichen Austausch.

Die Stellungnahme bezieht sich auf folgende Punkte:

- 2. Produkte des Institutes
- 3. Nutzenbewertung medizinischer Interventionen
- 6. HTA-Berichte
- 7. Evidenzbasierte Gesundheitsinformationen für Bürgerinnen und Bürger
- 9. Informationsbewertung
- Kritikpunkte aus der Stellungnahme vom 31. Juli 2014

Im weiteren Verlauf beziehen sich alle Seitenangaben auf „Allgemeine Methoden V 5.0“.

¹ IQWiG 2016: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden. Entwurf für Version 5.0 vom 07.12.2016, Köln 2016.

² IQWiG 2016a: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, <https://www.iqwig.de/de/methoden/methodenpapier.3020.html>, entnommen am 02.01.2017

Zu 2. „Produkte des Institutes“

Zu 2.1.1 „Bericht“

Nach Auftragserteilung durch den G-BA oder das BMG und der anschließenden Formulierung der wissenschaftlichen Fragestellung wird durch das IQWiG ein Berichtsplan erstellt, der die Grundlage für die Erstellung eines Vorberichtes bildet. Der Berichtsplan wird zur Anhörung gestellt, allerdings kann laut Methodenpapier *„in bestimmten Fällen [...] auf die Anhörung zum Berichtsplan verzichtet werden.“* (s. S. 20) Es ist unklar, in welchen Fällen das IQWiG auf eine Anhörung verzichten wird. Grundsätzlich sollte es für das Institut verpflichtend sein, Anhörungen zu allen Berichtsplänen und Vorberichten, insbesondere für Aufträge zur Erstellung von Gesundheitsinformationen, durchzuführen, da auch bei Informationsmaterialien zu bereits bestehenden Richtlinieninhalten größere Änderungen enthalten sein können. Falls es tatsächlich Fälle geben sollte, in denen eine Anhörung nicht sinnvoll oder notwendig erscheint, sollten aus Transparenzgründen an dieser Stelle eindeutige Kriterien definiert werden, die es erlauben nachzuvollziehen, weswegen auf eine Anhörung verzichtet wird. Darüber hinaus sollte die Einbindung von externe Gutachterinnen und Gutachter bei umfassenden oder grundlegenden inhaltlichen Änderungen nicht optional sein, sondern regelhaft stattfinden (s. Kapitel 2.1.8 „Gesundheitsinformationen“).

Anhörungen erfolgen nach Einreichung schriftlicher Stellungnahmen, wobei diese innerhalb einer Frist von mindestens 4 Wochen nach Veröffentlichung des Berichtsplans bzw. Vorberichts eingereicht werden können. Eine Fristdauer von 4 Wochen ist insbesondere für Fachgesellschaften mit ausgedehnten Abstimmungs- und Freigabe-Prozessen sehr kurz. Daher würden wir es begrüßen, wenn die Fristdauer, je nach Art und Umfang des Dokumentes, das zur Stellungnahme ausgeschrieben ist, ausgedehnt werden würde.

Nach Einholung der schriftlichen Stellungnahmen *„ist eine mündliche wissenschaftliche Erörterung mit Stellungnehmenden vorgesehen.“* (s. S. 20) Die Einbeziehung der Stellungnehmenden sollte nicht nur optional sein, sondern regelhaft stattfinden.

Zusätzlich soll an dieser Stelle darauf hingewiesen werden, dass innerhalb des Methodenpapiers V 5.0 keine Stringenz der verwendeten Bezeichnungen erkennbar ist. So ist von „Berichten“, „Berichtsplänen“, „Vorberichten“, „Projektskizzen“ und „Produktentwurf“ etc. die Rede, obwohl teilweise dasselbe gemeint zu sein scheint. Entsprechend sollte innerhalb des Methodenpapiers V 5.0 eine einheitliche Formulierung und klare Abgrenzung zwischen den Begrifflichkeiten vorgenommen werden, um Unklarheiten zu vermeiden.

Zu 2.1.10 „HTA-Bericht“

Im Rahmen der Erstellung eines HTA-Berichts wird der vorläufige Basisbericht zur Anhörung gestellt (s. S. 34).

An dieser Stelle fehlt eine Konkretisierung, wer unter welchen Voraussetzungen berechtigt ist, an der Anhörung teilzunehmen. Aus Sicht des DIMDI sollten *„Zur Unterstützung nachhaltiger Entscheidungen [sollten] vorhandene Informationen aus möglichst allen betroffenen Bereichen umfassend analysiert werden.“*³ Somit sollte auch der pharmazeutische Unternehmer, der ebenfalls einen „betroffenen Bereich“ darstellt, die Möglichkeit haben, seine Expertise aus klinischen Studien und der Versorgungsrealität einzubringen.

In der Abbildung 9 „Ablauf der Erstellung eines HTA-Berichts“ (s. S. 33) ist dargestellt, dass der von den externen Sachverständigen erstellte vorläufige Basisbericht zur Anhörung gestellt wird, aber die Herausgeberkommentare des IQWiG, die zusammen mit dem Basisbericht den finalen HTA-Bericht bilden, werden nicht zur Anhörung gestellt. Aus Transparenzgründen sollten sowohl der Basisbericht als auch die Herausgeberkommentare des IQWiG zur Anhörung gestellt werden.

³ DIMDI. HTA unterstützt Entscheidungen im Gesundheitswesen, [online] Link: <http://www.dimdi.de/static/de/hta/> entnommen am 03.01.2017

Des Weiteren wird in Tabelle 2 (s. S. 36) eine Einbindung Betroffener explizit ausgeschlossen, obwohl zu Beginn des Abschnitts 2.2.1 das Institut selbst mit entsprechenden Quellen belegt, dass *„Die Einbindung von Betroffenen im Rahmen der Erstellung von systematischen Übersichten **und** HTA-Berichten [...] mittlerweile etablierter internationaler Standard der Nutzenbewertung [ist].“*

Zu 2.2.1 „Einbindung von Betroffenen in die Erstellung von IQWiG-Produkten“

Nach Angaben des IQWiG können Betroffene *„Patientinnen und Patienten, ihre Eltern oder sonstige Angehörige sowie potenzielle Teilnehmer an Präventionsmaßnahmen.“* sein (s. S. 35) Diese Definition sollte erweitert werden für *„Betroffene einer spezifischen Erkrankung und Vertreter übergreifender Verbände“*.

Das IQWiG führt auf S. 36 näher aus, dass die Art der Einbindung Betroffener davon abhänge, wie viel Zeit für die Berichterstellung zur Verfügung stehe und ob das Berichtsthema vertraulich sei. In Tabelle 2 wird die Art der Einbindung (mündlich, schriftlich, keine Einbindung) schematisch dargestellt. Was fehlt, ist hier eine transparente Darstellung des Zeitrahmens der einzelnen Produkte des IQWiG für die Einbindung Betroffener und eine entsprechende Rationale dahinter. Außerdem ergeben sich unter Umständen in einem mündlichen Diskurs Aspekte, die bei einer rein schriftlichen Abfrage nicht thematisiert werden. Es ist auch nicht nachzuvollziehen, warum bei Potenzialbewertungen und HTA-Berichten überhaupt keine Einbeziehung von Patienten erfolgt.

Schließlich stellt das IQWiG auf S. 36 fest, dass eine erneute Einbindung von Betroffenen verzichtbar sei, wenn ein Berichtsthema bereits in Rahmen einer früheren Konsultation bearbeitet wurde. Auch dies ist nicht nachvollziehbar, da sich mit einer gewissen zeitlichen Latenz Änderungen der Therapie und damit Einschätzungen der Patienten zu ihrer Krankheitslast, zu Endpunkten usw. ergeben können.

Zu 3. „Nutzenbewertung medizinischer Interventionen“

Zu 3.1.4 „Endpunktbezogene Bewertung“

Laut Angaben der Allgemeinen Methoden V 5.0 bewertet das Institut jeden einzelnen patientenrelevanten Endpunkt anhand von Analysen vorhandener wissenschaftlicher Daten und trifft eine Aussage bzgl. des (Zusatz-)Nutzens basierend auf der Beleglage. Aus Sicht des IQWiG kann aus lediglich 1 Studie *„im Ausnahmefall für einen Endpunkt ein Beleg für eine spezifische (Teil-)Population abgeleitet werden, wenn ein Studienbericht gemäß 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 mindestens 1000 Patienten eingeschlossen und es gibt mindestens 10 Zentren.“* (s. S. 51)

Es wird mit keiner Referenz die Rationale belegt, warum 1000 Patienten aus mindestens 10 Zentren vorliegen müssen, um einen Beleg für einen Zusatznutzen erreichen zu können. Diese Festlegung erscheint daher willkürlich. Aus Transparenzgründen und aus Gründen der Nachvollziehbarkeit sollte an dieser Stelle eine entsprechende Quelle und eine Rationale genannt werden.

Weiterhin sei darauf hingewiesen, dass nicht allein die Patientenzahl, sondern auch die Effektgröße und die Kategorie des patientenrelevanten Endpunktes berücksichtigt werden sollte. Für einen Mortalitätsvorteil erscheint es fragwürdig, dass 1000 Patienten aus mindestens 10 Zentren vorhanden sein müssen, damit ein Beleg für einen Zusatznutzen anerkannt wird. Insbesondere bei Erkrankungen mit kleinen betroffenen Patientenpopulationen ist es nicht möglich, eine so hohe Patientenzahl für eine klinische Studie zu rekrutieren, was die Erfüllung dieser Anforderung faktisch unmöglich macht. Darüber hinaus wäre die Durchführung von mehreren, ähnlich konstruierten Studien, z. B. bei onkologischen Erkrankungen oder pädiatrischen Indikationen, ethisch nicht vertretbar. Die EMA betont, dass Im Rahmen der zentralen Zulassung eines Arzneimittels keine formale Anforderung für mehr als 1 pivotale Phase III-Studie im Studienprogramm besteht. Falls jedoch nur eine Phase III Studie als statistischer Beweis für Wirksamkeit vorhanden ist, wird die Einhaltung strenger Kriterien für diese Studie gefordert wie interne und externe Validität, klinische Relevanz, statische Signifikanz, Datenqualität und Datenkonsistenz.⁴ Mindest-Patientenzahlen pro Studienarm oder eine Mindestanzahl an Studienzentren, wie vom IQWiG gefordert, werden in dieser Guideline jedoch nicht thematisiert.

⁴ Committee for Proprietary Medicinal Products (2001): Points to consider on application with 1. Meta-analysis; 2. One pivotal study (CPMP/EWP/2330/99)
http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003657.pdf
 entnommen am 26.01.2017

Zu 6. „HTA-Berichte“

Zu 6.3 „Auswahl der Themen für die HTA-Berichte“

Bei der „Auswahl der Themen für die HTA-Berichte“ liegt die Prüfung und Gewichtung der vorgeschlagenen Themen allein beim IQWiG. Erst im zweiten Schritt, „Auswahlstufe 1“, kommt der sogenannte Auswahlbeirat hinzu. Hier findet also schon eine unter Umständen intransparente und willkürliche Vorselektion statt, die möglicherweise durch bestimmte Interessen gesteuert sein kann und keinerlei Kontrolle unterworfen ist.

Zu 6.3.1 „Auswahlkriterien“

Die Auswahlkriterien sind sehr allgemein formuliert, so dass ein breiter Interpretations- und Ermessenspielraum für das IQWiG eröffnet wird. So bleibt z. B. unklar, ab welcher Anzahl von Betroffenen oder ab welcher Kostengrenze das IQWiG positiv urteilen würde.

Zu 6.3.2 „Prüfung und Gewichtung der Themen“

Auf Seite 133, Absatz 3 heißt es: „Die eingereichten Vorschläge werden zeitnah durch das Institut geprüft und gegebenenfalls redaktionell und qualitativ bearbeitet.“

Auch an dieser Stelle räumt sich das IQWiG durch die ggf. redaktionelle und qualitative Bearbeitung einen hohen Ermessensspielraum ein. Zur Kontrolle der Bearbeitung durch das IQWiG, sollte der originäre Antragstext veröffentlicht werden.

Zu 6.3.3 „Erste Stufe des Auswahlverfahrens: Nominierung von Themen“ und 6.3.4 „Zweite Stufe des Auswahlverfahrens: Auswahl von Themen, zu denen HTA-Berichte erstellt werden“

In Abschnitt 6.3.3 ist die Zusammensetzung des Auswahlbeirates unseres Erachtens unzureichend, da dort keinerlei medizinisch-fachliche Expertise vorgesehen ist. Zusätzlich ist es nicht sinnvoll, dass der Auswahlbeirat nicht bereits in der vorgelagerten Phase, die derzeit ausschließlich beim IQWiG liegt, beteiligt wird. Dieses Problem setzt sich in Abschnitt 6.3.4 fort und es erfolgt auch hier keine Beteiligung des Auswahlbeirates auf dieser wichtigen Stufe.

Zu 6.4 „Sicherstellung der Qualität der HTA-Berichte“

Das Berichtsprotokoll wird nur durch das Institut begutachtet und es erfolgt keine öffentliche Diskussion interessierter/betroffener/fachlich kompetenter Kreise, wie es bei anderen IQWiG-Produkten üblich ist. Auch hier wäre eine Beteiligung Externer im Rahmen von Stellungnahme und Anhörung wünschenswert.

Zu 7. „Evidenzbasierte Gesundheitsinformationen für Bürgerinnen und Bürger“

Zu 7.2.1 „Themenkatalog gemäß des Generalauftrags“

Generell ist die Auswahl der Themen durch das IQWiG nicht transparent. Offenkundig liegen den unter www.gesundheitsinformationen.de dargestellten Themen andere Auswahlkriterien zugrunde als die Prävalenz der Erkrankungen (Beispiel „Frühbewertung von Arzneimitteln“).

Auf Seite 139, Absatz 3 heißt es: *„Zur Erstellung des Themenkatalogs wird primär der zweijährlich aktualisierte Versorgungsreport des Wissenschaftlichen Instituts der AOK (WIdO) [252] herangezogen.“*

Der Versorgungsreport des WIdO berücksichtigt nur einen Ausschnitt der GKV-Population (nach Angaben des IQWiG ca. 24 Millionen AOK-Versicherte, s. S. 139). Es fehlt eine Begründung, warum die genannte Datenbasis für die Fragestellung geeignet ist bzw. eine geeignete Stichprobe der GKV-Gesamtpopulation darstellen soll. Es stellt sich die Frage, warum an dieser Stelle nicht Daten verwendet werden, die die Gesamtheit der GKV-Population umfassen, z.B. die Daten, die nach § 267 SGB V für den Risikostrukturausgleich erhoben werden.

Zu 7.4 „Auswahl der Evidenz“, 7.5 „Auswahl der dargestellten Ergebnisse“, 7.6 „Wahl und Darstellung von Vergleichen“, 7.7 „Umgang mit Zahlen und Risikoangaben“ und 7.8 Berücksichtigung von Alters- und Geschlechtsunterschieden

In den genannten Abschnitten findet eine statistisch-methodisch getriebenen Bewertung unter nicht ausreichender Berücksichtigung alltäglicher und klinischer Aspekte statt. Allgemein wäre im gesamten Feld „Gesundheitsinformationen“ auch eine Einbeziehung klinischer Expertise erforderlich.

Auf Seite 144, Absatz 5 und 6 heißt es z.B.: *„Informationen über Behandlungsergebnisse beziehen sich auf Endpunkte, die für Patientinnen und Patienten relevant sind, also insbesondere auf die Sterblichkeit (Mortalität), die Beschwerden und Komplikationen (Morbidität) und die gesundheitsbezogene Lebensqualität. Dazu gelten grundsätzlich die in Abschnitten 3.1 und 3.2 der Methoden beschriebenen Grundlagen.“*

Zudem sind die Begleitumstände der Behandlung (z. B. Zeitaufwand, körperliche, seelische, soziale und auch finanzielle Belastungen) für Patienten oft wichtige Informationen.“

Es wird kurz kommentiert, dass Begleitumstände der Behandlung für die Patienten eine Rolle spielen können, zu dieser Thematik wird jedoch keine detaillierte Angabe zum methodischen Vorgehen gemacht. Stattdessen wird auf die Methodik in Kapitel 3 „Nutzenbewertung medizinischer Interventionen“ abgestellt, ohne dass eine Begründung erfolgt, warum für unterschiedliche Fragestellungen die gleiche Methodik herangezogen wird. Bei genauerer Betrachtung legt das IQWiG außerdem für die Gesundheitsinformationen nicht dieselben Maßstäbe an, z.B.

- Die Bewertung wird auf systematische Reviews hoher Qualität beschränkt (unabhängig von der Qualität der Einzelstudien).
- Evidenz, die älter ist als 3 Jahre, wird unabhängig vom Anwendungsgebiet ausgeschlossen.

Des Weiteren wäre es sinnvoll, die Aussagen aktueller deutscher Leitlinien explizit als relevante Informationsquelle einzubeziehen.

Dass IQWiG formuliert die Wichtigkeit von Begleitumständen der Behandlung für Patienten. Daher sollten auch weiterführende Endpunkte wie z.B. Patientenzufriedenheit oder klinisch relevante Surrogatendpunkte (LDLC) unbedingt berichtet werden.

Zu 7.9 „Anpassung an die Zielgruppe, 7.9.1 „Beteiligung von Betroffenen“ und 7.9.2 „Nicht öffentliches Stellungnahmeverfahren“

In Abschnitt 7.9 bzw. Abschnitt 7.9.1 sieht das Institut z.B. eine Einbeziehung von Betroffenen, von Surveys und Erfahrungsberichten vor. Die Einbindung klinischer Expertise scheint auch an dieser Stelle nicht vorgesehen zu sein. Dabei ist das Ziel explizit nicht nur die Verständlichkeit, sondern auch die wissenschaftliche Präzision. Darüber hinaus ist davon auszugehen, dass Mediziner, die im Umgang mit betroffenen Patienten erfahren sind, auch über Expertise zur Einschätzung der Verständlichkeit von Gesundheitsinformationen verfügen.

In Abschnitt 7.9.2 wird das nicht öffentliche Stellungnahmeverfahren beschrieben. Vor dem Hintergrund der Zielsetzung *„Verbesserung des Verständnisses medizinischer und wissenschaftlicher Informationen, darunter auch des Konzepts der evidenzbasierten Medizin“* (s. S. 138) bleibt unklar, warum das Stellungnahmeverfahren nicht öffentlich ist und die Stellungnahmen und Würdigungen nicht veröffentlicht werden.

Zu 7.13 „Transparenz über Verfasser und Herausgeber“

Auf Seite 150, Absatz 2 heißt es: *„Auf der Website gesundheitsinformation.de und iqwig.de stellt das Institut seine Grundlagen und Finanzierung als gemeinnütziger und wissenschaftlich unabhängiger Herausgeber von Gesundheitsinformationen dar.“*

Im Sinne der Transparenz, sollten nicht nur das IQWiG sondern auch andere Institutionen und Verbände, die an den erstellten und veröffentlichten Informationen mitgewirkt haben, genannt werden.

Zu 9. „Informationsbewertung“

Wir begrüßen die folgenden Änderungen bzw. Anpassungen:

- In Abschnitt 9.3.8 „Meta-Analysen, A) Allgemeines“ begrüßen wir die Möglichkeit, im Fall weniger Studien (weniger als 5) auch Meta-Analysen mit Studien als festem Effekt einbringen zu können. Diese Änderung ist gerade vor dem Hintergrund der frühen Nutzenbewertung sinnvoll, da in der Regel sehr wenige Studien zur Verfügung stehen.
- In Abschnitt 9.3.10 „Subgruppenanalysen“ begrüßen wir die Beschränkung auf Subgruppenanalysen bei binären Endpunkten mit mindestens 10 Ereignissen in einer der Subgruppen. Außerdem finden wir es richtig, dass der gemeinsame Gesamteffekt in Subgruppen nicht mehr in Frage gestellt wird, wenn paarweise statistische Tests auf Heterogenität oder Interaktion einen p-Wert zwischen 0,05 und 0,2 ergeben.

Zu 9.3.8 Meta-Analysen, C) Geringe Zahl von Ereignissen

Auf Seite 190, 4. Absatz heißt es: *„Bei sehr seltenen Ereignissen kann die Peto-Odds-Ratio-Methode verwendet werden, die keinen Korrekturterm beim Vorliegen von Nullzellen erfordert [63,146]. Die Anwendung dieser Methode ist allerdings nur adäquat, wenn die zu schätzenden Effekte nicht zu groß sind und kein unbalanciertes Design vorliegt [71].“*

Das Institut führt für die genannten Fälle „große Effekte“ und „unbalanciertes Design“ nicht aus, was in diesem Fall ein geeignetes alternatives relatives Effektmaß wäre.

Zu 9.3.9 Indirekte Vergleiche

Auf Seite 192, 3. Absatz heißt es: *„Darüber hinaus ist eine wesentliche Voraussetzung für die Berücksichtigung eines indirekten Vergleichs, dass dieser auf die interessierende Fragestellung insgesamt ausgerichtet ist und nicht nur selektiv auf z. B. einzelne Endpunkte.“*

Sanofi ist der Meinung, dass es auch im Rahmen indir. Vergleiche durchaus Fragestellungen gibt, die sich auf Teilaspekte der interessierenden Fragestellung bzw. auf einzelne Endpunkte konzentrieren. Das IQWiG selbst reduziert seine Bewertungen (auch innerhalb indir. Vergleiche) auf Endpunkte, die es zuvor als „patienten-relevant“ identifiziert hat.

Auf Seite 193, 3. Absatz heißt es weiter: *„Neben den Annahmen einer ausreichenden Ähnlichkeit und Homogenität der paarweisen Meta-Analysen, die auch hier erfüllt sein müssen, muss bei Netzwerk-Meta-Analysen zusätzlich eine ausreichende Konsistenz der geschätzten Effekte aus direkter und indirekter Evidenz gegeben sein.“*

An dieser Stelle ist unklar auf was sich die „Ähnlichkeit“ bezieht. Sinnvollerweise sollte hier die Ähnlichkeit von Studien gemeint sein. Da „unzureichende Ähnlichkeit von Studien“ im Rahmen der frühen Nutzenbewertung bislang der häufigste Ablehnungsgrund von indirekten Vergleichen ist, sollte es Aufgabe des IQWiG sein, Kriterien oder eine geeignete Methodik für die Ähnlichkeit von Studien zu entwickeln.

Zu 9.3.10 Subgruppenanalysen

Auf Seite 194, 1. Absatz heißt es: *„Auch wenn Subgruppenanalysen für die Optimierung der Behandlung sinnvoll sind, werden sie in der methodischen Literatur teilweise kritisch diskutiert [26,498]:“*

Zusätzlich zu den im Anschluss aufgezählten drei Problemfeldern (geringe Power; Vergleichbarkeit zwischen Behandlungsgruppen innerhalb von Subgruppen, die bei der Randomisierung nicht stratifizierend berücksichtigt wurden; Wechselwirkungen höherer Ordnung) fehlt eine kritische Auseinandersetzung mit dem Multiplizitätsproblem, also der Tatsache, dass mit wachsender Zahl von Subgruppenanalysen das Risiko ansteigt, falsch positive 'Belege' für Effektmodifikationen zu finden. Mit dem gewählten Signifikanzniveau für Heterogenitäts- oder Interaktionstests ($\alpha = 0,05$) erwartet man pro 20 durchgeführten Subgruppenanalysen, in denen de facto keine Effektmodifikation vorliegt, ein falsch positives Ergebnis.

Zu 9.3.13 Auswertung abhängiger Daten

Auf Seite 199, 4. Absatz heißt es: *„Die einfachen Standardverfahren der medizinischen Statistik gehen von unabhängigen Beobachtungseinheiten aus. In bestimmten Situationen (z. B. clusterrandomisierte Studien oder Untersuchungen an Augen oder Zähnen), bei denen mehrere Beobachtungseinheiten eines Clusters oder eines Patienten in die Analyse eingehen, ist diese Unabhängigkeit nicht gegeben. Die Anwendung der einfachen Standardverfahren für unabhängige Beobachtungseinheiten führt dann zu einer Unterschätzung der Varianz und möglicherweise zu falschen Signifikanzaussagen, da die errechneten p-Werte zu klein sind [14].“*

Der letzte zitierte Satz ist nicht allgemein gültig. Z.B. beobachtet man bei der Erfassung eines quantitativen Merkmals in einer Studie mit einem bilateral-paired Design (wie es u.a. in der Dermatologie anzutreffen ist) regelmäßig, dass die auf verschiedenen Seiten der Patienten (z.B. Unterarme) unter unterschiedlichen topischen Therapien gemessenen Werte positiv korreliert sind. In solchen Fällen führt die Anwendung der einfachen Standardverfahren für unabhängige Beobachtungseinheiten dann zu einer *Überschätzung* der Varianz und in Folge zu *zu hohen* p-Werte.

Kritikpunkte aus der Stellungnahme vom 31. Juli 2014

Im Folgenden möchten wir noch Kritikpunkte aus der letzten Stellungnahme vom 31. Juli 2014 aufgreifen, welche nach wie vor relevant und uns wichtig sind.

Zu 2. „Produkte des Institutes“

Zu 2.1. „Produktspezifische Verfahrensabläufe“

Hinsichtlich Potenzialbewertungen wird hier festgelegt, dass Potenzialbewertungen im Auftrag des G-BA durchgeführt werden und entsprechende Anhörungsverfahren nicht beim IQWiG erfolgen. Erst im Falle einer Erprobung würde der G-BA ein Stellungnahmeverfahren zur Erprobungsrichtlinie durchführen. (s. S. 16). Im Licht des Transparenzgrundsatzes ist es nicht nachvollziehbar, warum das IQWiG ein Anhörungsverfahren im Rahmen der Potenzialbewertung ausschließt, denn auch unter der gebotenen Vertraulichkeit sollte den betroffenen Parteien die Möglichkeit gegeben werden, zu offenen Fragen Stellung zu nehmen.

Zu 2.1.5. „Potenzialbewertung“

In der Abbildung 5 (s. S. 26) wird der Ablauf der Potenzialbewertung durch das IQWiG schematisch dargestellt, ohne dass jedoch der Prozess selbst inhaltlich transparent dargestellt wird. Auch eine Einbeziehung von externen Sachverständigen ist nur fakultativ vorgesehen. Da bei der Potenzialbewertung von Medizinprodukten regelhaft zu erwarten ist, dass entsprechende Evidenz nur begrenzt vorhanden ist, sollte es für das Institut verpflichtend sein, externe Sachverständige heranzuziehen. Insbesondere der interne Qualitätssicherungsprozess für die Potenzialbewertung sollte transparent im Methodenpapier dargelegt werden.

Zu 2.1.8. „Gesundheitsinformation“

Es ist nicht nachvollziehbar, warum im Zuge der Gesundheitsinformationserstellung eine externe Stellungnahme nur auf einen kleinen Kreis beschränkt werden soll, nämlich Auftraggeber, Kuratorium und weitere Gremien des Instituts/ IQWiG. Im Kuratorium sind die Organisationen der Leistungserbringer und die Selbstverwaltungsorgane der Trägerorganisationen des Gemeinsamen Bundesausschusses, inklusive Patientenvertretern, repräsentiert. Was fehlt, ist die Möglichkeit, dass auch Personen und Institutionen in der Wissenschaft, Hersteller von pharmazeutischen Produkten und Medizinprodukten und sonstige Interessierte offiziell Stellung nehmen können. Es wird daher gefordert, die Möglichkeit der Stellungnahme auf einen größeren Kreis Interessierter und Betroffener zu erweitern (s. S. 29/30).

Zu 2.2.2. „Auswahl externer Sachverständiger“, 2.2.3 „Gewährleistung der fachlichen Unabhängigkeit“ und 2.2.4 „Review der Produkte des Instituts“

Laut eigener Aussage vergibt das IQWiG Aufträge an externe Sachverständige in einem transparenten und nichtdiskriminierenden Wettbewerb. Dies erfolgt über eine Registrierung Interessierter in eine Sachverständigen-Datenbank beim IQWiG, aus der das IQWiG je nach Auftrag Sachverständige nach gesetzten Kriterien auswählt (s. S. 37). Um dem Grundsatz der Transparenz nachzukommen, wäre es jedoch wünschenswert und notwendig, das Auswahlverfahren, die in Frage kommenden Sachverständigen und auch die letztlich angewendeten Kriterien, die zur Wahl eines bestimmten Sachverständigen geführt haben, zu veröffentlichen.

Im Sinne der Transparenz wäre es weiterhin wünschenswert, dass auch die externen Reviews veröffentlicht werden, was bis dato nicht der Fall ist (s. S. 37).

Zu 3. „Nutzenbewertung medizinischer Interventionen“

Zu 3.8. „Potenzialbewertung“

Das IQWiG führt auf S. 75, 3. Absatz aus, dass „aufgrund der besonderen Zielsetzung [...] bei der Potenzialbewertung im Vergleich zur Nutzenbewertung deutlich niedrigere Anforderungen an die Evidenz zu stellen“ sind. Zugleich werden hier zusätzliche Grade der Ergebnissicherheit „gering“, „sehr gering“ und „minimal“ neu eingeführt. Für jeden dieser Grade werden zudem minimale Effektstärken definiert, für die wiederum in den Allgemeinen Methoden V 5.0, wie bereits schon in dem Entwurf zu 4.2 angemerkt, nur „ungefähre Grenzen“ festgelegt werden. (s. S. 75-77).

Diese neu eingeführten Effektstärken „gering“, „sehr gering“ und „minimal“ führen eher dazu, dass selbst die bestmöglich verfügbare Evidenz bei der Testung von Medizinprodukten sprachlich herabgewürdigt wird. Um der Tatsache Rechnung zu tragen, dass im Bereich von Medizinprodukten schon die bestmöglich generierbare Evidenz unterhalb dessen liegt, was bei Arzneimitteln im Rahmen von klinischen Studien möglich ist, sollte daher die Ergebnissicherheit bei Medizinprodukten deskriptiv dargestellt werden.

Das Design einer möglichen Erprobungsstudie, die sich aus einer Potenzialbewertung ergibt, sollte in jedem Fall im Dialog zwischen dem IQWiG, der Selbstverwaltung und dem Hersteller, ggf. unter Hinzunahme von externen Sachverständigen, erarbeitet werden. Wie bereits oben ausgeführt, könnte das z.B. im Rahmen eines Anhörungsverfahrens, die das Methodenpapier leider nicht vorsieht, diskutiert werden (s. oben Kommentar zu 2.1.5 „Potenzialbewertung“).

Zu 8. „Informationsbeschaffung“

Im einleitenden Teil von Kapitel 8 (vormals Kapitel 7) wird – ergänzend zur bisherigen Fassung des Methodenpapiers – festgelegt, dass außerhalb bestehender Regelungen unaufgefordert übermittelte Studiendaten seitens des Institutes bei der Abfassung seiner Berichte inhaltlich nicht berücksichtigt werden. Begründet wird dies mit dem potentiellen Risiko einer selektiven Auswahl der übermittelten Daten, was wiederum zu einer Verzerrung der Berichtsergebnisse führen könne. (s. S. 178 Allgemeine Methoden V 5.0) Auffällig ist, dass in dieser Begründung nun ganz explizit die Ergebnisse von Nutzenbewertungen genannt werden, während der eingefügte Abschnitt ansonsten bezüglich anderer möglicherweise betroffener Produkte des Instituts unspezifisch ist.

Eine derart pauschale Einschränkung hinsichtlich der Erhebung der Datengrundlage für Berichte aller Art, wie sie hier vorgesehen ist, könnte dazu führen, dass relevante Informationen bzw. neueste und aktuellste Evidenz in der Berichterstellung keine Berücksichtigung fänden, was dann umgekehrt ebenfalls eine Verzerrung des Berichtsergebnisses nach sich zöge. Diese Vorgehensweise erscheint auch nicht konform mit der im zur Kommentierung stehenden Methodenpapier selbst in Abschnitt 1.2.4, Absatz 4 aufgeführten zentralen Strategie der evidenzbasierten Medizin, der sich das Institut in seiner Arbeitsweise offenbar verpflichtet sieht („Eine weitere zentrale Strategie der EbM besteht darin, alle nach der Qualität ihres Designs und ihrer Durchführung angemessenen Studien zu einer Frage zu identifizieren und so den Stand des zuverlässigen Wissens zusammenzufassen...“, s. S. 7 Allgemeine Methoden V 5.0) und damit letztendlich auch nicht im Sinne der von einer zu bewertenden Therapie potentiell betroffenen Patienten.

Das in den Raum gestellte Risiko des Selektionsbias in der Begründung der geplanten Einschränkung hinsichtlich der Akzeptanz unaufgefordert eingereicherter Daten erscheint auch deshalb nicht stichhaltig, da es gemäß Abschnitt 8.5 des vorliegenden Entwurfs des Methodenpapiers (Überprüfung der Informationsbeschaffung) ohnehin zur Arbeitsweise des Instituts gehört, eingereichte Daten zu überprüfen – zumindest im Zusammenhang mit einer Dossierbewertung. Aber auch außerhalb von Dossierbewertungen bleibt es dem Institut jederzeit unbenommen zu prüfen, ob eine zu den unaufgefordert eingereichten Daten möglicherweise gegenteilige Evidenz verfügbar ist

und diese dann gegebenenfalls entsprechend in seine Bewertung mit einzubeziehen, bzw. die Daten wegen mangelnder Qualität unberücksichtigt zu lassen und insofern einen potentiellen Bias zu vermeiden.

Darüber hinaus erscheint dieses Risiko auch insofern eher artifiziell, als keinerlei Angaben dazu gemacht werden, inwieweit sich aus der bisherigen Arbeit des Instituts stichhaltige Belege dafür finden lassen, dass es in der Vergangenheit zu derartigen Verzerrungen, wie sie eingangs als Risiko beschrieben wurden, gekommen ist.

Zu 9. „Informationsbewertung“

Zu 9.1.3. „Rangordnung verschiedener Studienarten/ Evidenzgrade“

Das Institut führt zunächst aus, dass es verschiedene Ansätze gibt, Studientypen gemäß Ihrer Evidenzgrade in eine Rangfolge zu bringen, dass aber keines dieser Systeme allgemein akzeptiert und universell einsetzbar sei. Das Institut orientiert sich weitgehend an der Evidenzklassifizierung in der aktuellen Verfahrensordnung des Gemeinsamen Bundesausschusses⁵, die in der Arzneimittel-Nutzenbewertungsverordnung des Bundesministeriums für Gesundheit in § 5 (6) wie folgt festgelegt ist:

- | | | |
|----|------|---|
| 1. | I a | systematische Übersichtsarbeiten von Studien der Evidenzstufe Ib |
| 2. | I b | randomisierte klinische Studien (RCT) |
| 3. | II a | systematische Übersichtsarbeiten der Evidenzstufe IIb |
| 4. | II b | prospektiv vergleichende Kohortenstudien |
| 5. | III | retrospektiv vergleichende Studien |
| 6. | IV | Fallserien und andere nicht vergleichende Studien |
| 7. | V | Assoziationsbeobachtungen, pathophysiologische Überlegungen, deskriptive Darstellungen, Einzelfallberichte, nicht mit Studien belegte Meinungen anerkannter Experten, Konsensuskonferenzen und Berichte von Expertenkomitees. |

Dementsprechend kommt unter den verschiedenen Studientypen randomisiert kontrollierten Studien (RCT) der höchste Evidenzgrad zu. Das Institut betont, dass darüber hinaus RCT auch nach ihrer Qualität eingestuft werden.

Nicht-randomisierte Studien teilt das Institut grundsätzlich in vergleichende und nicht vergleichende Studien ein. Das Institut führt aus, dass bei nicht randomisierten vergleichenden Studien neben dem Studiendesign vor allem die Maßnahmen zur Kontrolle von Confoundern zu prüfen sind, dies jedoch die Bewertung des Verzerrungspotentials und nicht den Evidenzgrad berührt.

Wie bisherige Nutzenbewertungen gezeigt haben, wurden in internationalen Entwicklungsprogrammen neuer Arzneimittel häufig andere Komparatoren verwendet als die unter nationalen Gesichtspunkten vom Gemeinsamen Bundesausschuss gesetzten zweckmäßigen

⁵ Gemeinsamer Bundesausschuss. Verfahrensordnung des Gemeinsamen Bundesausschusses [online]. Link: https://www.g-ba.de/downloads/62-492-1282/VerfO_2016-07-21_iK-2016-10-29.pdf, entnommen am 02.01.2017

Vergleichstherapien. In der Arzneimittel-Nutzenbewertungsverordnung⁶ findet sich hierzu in § 5 (5) der Hinweis:

„Liegen keine direkten Vergleichsstudien für das neue Arzneimittel gegenüber der zweckmäßigen Vergleichstherapie vor oder lassen diese keine ausreichenden Aussagen über einen Zusatznutzen zu, können verfügbare klinische Studien für die zweckmäßige Vergleichstherapie herangezogen werden, die sich für einen indirekten Vergleich mit dem Arzneimittel mit neuen Wirkstoffen eignen.“

Der pharmazeutische Unternehmer hat also in diesen Situationen die Option, indirekte Vergleiche gegen die vom Gemeinsamen Bundesausschuss gesetzte zweckmäßige Vergleichstherapie durchzuführen und vorzulegen. Vor diesem Hintergrund erscheint es dringend geboten, dass das Institut in 9.1.3 ausführt, wie es indirekte Vergleiche bzw. Netzwerk-Metaanalysen in die Evidenzhierarchie eingruppiert.

Stand: 31.01.2017

Sanofi-Aventis Deutschland GmbH

Potsdamer Str. 8

10785 Berlin

⁶ M. Bullinger. Das Konzept der Lebensqualität in der Medizin – Entwicklung und heutiger Stellenwert. Z. Evid. Fortbild. Qual. Gesundh. wesen 2014; 108: 97-103

A.1.31 – SmartStep Consulting GmbH

Autoren:

- Ehlert, Andree
- Nitz, Inke
- Volmer, Timm

SmartStep Consulting GmbH, Alter Teichweg 25a, 22081 Hamburg

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

Hamburg, 30. Januar 2017

Stellungnahme zum Entwurf der IQWiG-Methoden – Version 5.0

Sehr geehrte Damen und Herren,

das IQWiG hat am 07.12.2016 den Entwurf für den nächsten Überarbeitungsschritt seiner Allgemeinen Methoden vorgelegt. Folgende Aspekte sollten aus unserer Sicht im Rahmen der Änderungen berücksichtigt werden.

Konkretisierungen und Ergänzungen in Bezug auf das Thema Medizinprodukte

- Das neue Methodenpapier wurde um die Bewertung von Medizinprodukten hoher Risikoklassen erweitert (§ 137h SGB V). Die Vorgehensweise entspricht im Wesentlichen dem der Potenzialbewertung nach § 137e SGB V – welche einige Unsicherheiten und Herausforderungen für die Antragsteller beinhaltet:
 - Kritisch zu sehen aus Sicht eines Dienstleisters, welcher Firmen aktiv bei der Erstellung der notwendigen Unterlagen unterstützt, ist der kurze Bearbeitungszeitraum für die Bewertungen nach § 137h. Dieser bedingt laut Darstellung eine „fehlende methodische Tiefe der Bearbeitung“ und erlaubt daraus resultierend nur eine geringere Aussagesicherheit zum Nutzen / Schaden. Aus unserer Sicht sollte ein Nutzenbewertungsverfahren nach § 137h mit seinen weitreichenden Konsequenzen für Hersteller und die Versorgung von Patienten eine zumindest dem Arzneimittelbereich vergleichbar verlässliche Aussage erlauben, welche auf einer ebenso vergleichbar sorgfältigen Bearbeitung des IQWiGs basiert. Gerade der Aspekt des Abwendens von Schaden, welcher öffentlich so betont wird, erfordert eine sorgfältige Abwägung des Nutzen-Schaden-Potenzials.

Biometrische Aspekte

- Bisher ist es im Rahmen der Nutzenbewertung gängige Praxis, Schwellenwerte von $\alpha = 0,20$ und $\alpha = 0,05$ für einen Hinweis bzw. Beleg des Vorliegens einer Effektmodifikation zu verwenden. Die Festlegung des Signifikanzniveaus auf $\alpha = 0,05$ wird begrüßt, da hierdurch eine Reduktion der Anzahl zu berücksichtigender Subgruppen in der Nutzenbewertung erfolgt. Das Ergebnis der Primäranalyse (Gesamtbewertung) rückt somit mehr in den Fokus, während die Wahrscheinlichkeit für nicht interpretierbare Zufallsergebnisse aus Subgruppen reduziert wird. Kritisch anzumerken erscheint aus Sicht von SmartStep die nicht einheitliche Darstellung des IQWiG zum Pro und Contra von Subgruppen und zu multiplen Testproblemen. Deutlich wird das Problem eventueller Doppelstandards in Abschnitt 9.3.10: *„Für den zielgerichteten Einsatz medizinischer Interventionen sind solche Subgruppenanalysen sinnvoll, da mit ihnen gegebenenfalls Teilpopulationen definiert werden können, für die eine Intervention einen Nutzen hat, und solche, für die dieselbe Intervention eher schadet als nutzt. [...] Solche Subgruppenanalysen sind dann im Sinne der systematischen Übersicht auch nicht als „post hoc“ zu bezeichnen, sondern entsprechen einer in dieser Übersicht zu überprüfenden Hypothese.“* Die Formulierung bietet eine scheinbar inhaltliche Begründung der Subgruppen- bzw. multiplen Testung, welche aus Sicht von SmartStep so nicht existiert. Zudem wird die Herausforderung, aus Subgruppenanalysen in systematischen Übersichten Hypothesen zu entwickeln, ohne das Problem des globalen Testniveaus zu thematisieren, vergrößert. Es bedarf aus Sicht von SmartStep einer erweiterten Auswahl der zitierten Literatur und eines konsequenten Verweises auf den explorativen Charakter bei „post hoc“ durchgeführten Subgruppenanalysen. Der in der Version 4.2 vorhandene Verweis auf das multiple Testproblem im Rahmen von Subgruppenanalysen wurde gestrichen, sollte aber zur Vermeidung von Missverständnissen wiederaufgenommen werden. Bei Subgruppenanalysen sollte daher ein korrigiertes Signifikanzniveau Anwendung finden und entsprechend im Methodenpapier diskutiert werden. Im Falle von multifaktoriellen Fragebögen sollte die Wahl des Signifikanzniveaus den Verfassern der jeweiligen Erhebungsinstrumente überlassen bleiben.
 - Die Formulierung *„Subgruppenanalysen können auch sozialrechtlich impliziert sein: Nach § 139a Absatz 2 SGB V ist es beispielsweise notwendig, alters-, geschlechts- und lebenslagenspezifische Besonderheiten zu berücksichtigen“* sollte aus dem Abschnitt 9.3 zu „speziellen biometrischen Aspekten“ gestrichen werden, da diese sozialrechtlichen Anforderungen nicht zur Begründung von Subgruppenanalysen im statistischen Sinne herangezogen werden können und sollten.
 - Die Aussage *„Damit überhaupt eine gewisse Wahrscheinlichkeit für einen solchen Nachweis vorliegt, werden Subgruppenanalysen bei binären Daten regelhaft nur dann durchgeführt, falls mindestens 10 Ereignisse in einer der Subgruppen aufgetreten sind.“* in Abschnitt 9.3.10 erfordert aus Sicht von SmartStep eine ausführlichere Begründung. Die Formulierung „damit überhaupt“ deutet auf ein gewisses Maß an Beliebigkeit im Umgang mit der

statistischen Power des Tests hin. Die anschließende Beschränkung auf „mindestens 10 Ereignisse“ stellt einen neu eingefügten Schwellenwert dar, der zu erklären ist. Entweder sollte der Satz gestrichen oder um eine ausführliche Diskussion inkl. Literaturbeleg ergänzt werden.

- In Abschnitt 9.3.10 sollte die Textstelle *„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.“* ebenfalls um den Hinweis auf ein multiples Testproblem sowie die Änderung des geforderten Signifikanzniveaus ergänzt werden.
- Das Signifikanzniveau für das Vorliegen einer Heterogenität im Rahmen von Meta-Analysen wurde ohne Begründung und ohne Verweis auf eine veränderte Literaturlage von $\alpha = 0,20$ auf $\alpha = 0,05$ abgesenkt. Unter gleichen Bedingungen erlaubt ein geringeres Niveau des Heterogenitätstests unter sonst gleichen Bedingungen häufiger eine metaanalytische Zusammenfassung von Einzelstudien. Die Auswirkungen im Rahmen der Nutzenbewertung sind möglicherweise erheblich. Die Änderung des Signifikanzniveaus sollte daher ausführlich wissenschaftlich diskutiert und begründet werden. Dabei geht es aus Sicht von SmartStep weniger um die Wahl des konkreten Signifikanzniveaus (welche in gewissem Sinne immer „willkürlich“ ist) als vielmehr um eine Begründung der veränderten Sichtweise, welche die bisherige Praxis der Nutzenbewertung darstellt. Im vorliegenden Text wird lediglich die geringe statistische Power des Heterogenitätstests diskutiert. Der Verweis auf D. Jackson [363] erscheint SmartStep ungeeignet, da dort eine flexible Handhabung bzw. Anpassung des Signifikanzniveaus favorisiert wird.
- Zusätzlich bleibt das Vorgehen im Falle von Meta-Analysen mit weniger als 5 Studien unklar. Eine Anzahl von weniger als 5 Studien stellt im Nutzenbewertungsverfahren für gewöhnlich den Regelfall dar. Daher wäre es umso wichtiger, dass dem Hinweis auf die Erwägung eines Modells mit festen Effekten eine konkrete Darstellung der notwendigen Anforderungen folgt. Dies bleibt in der vorliegenden Version offen.

Mit freundlichen Grüßen



Name

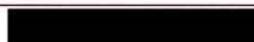


Name



Name

SmartStep Consulting GmbH



A.1.32 – SPECTARIS – Deutscher Industrieverband für optische, medizinische und mechatronische Technologien e. V. / ZVEI – Zentralverband Elektrotechnik- und Elektronikindustrie e. V. (gemeinsame Stellungnahme)

Autoren:

- Bätzel, Andreas
- Kuhlmann, Marcus

Stellungnahme von SPECTARIS und ZVEI

zum Entwurf des allgemeinen Methodenpapiers des IQWiG (Version 5.0)

Ansprechpartner aus den Verbänden

Andreas Bätzel

Referent Innovationen und Gesundheitspolitik

ZVEI – Zentralverband Elektrotechnik- und Elektronikindustrie e. V.

Fachverband Elektromedizinische Technik

Lyoner Straße 9, 60528 Frankfurt a.M.

Fon: +49 (0) 69 6302-

Fax: +49 (0) 69 6302-



www.zvei.org

Cordula Rapp

Projektmanagerin Medizintechnik

SPECTARIS. Deutscher Industrieverband für optische, medizinische und mechatronische Technologien e.V.

Werderscher Markt 15, D-10117 Berlin

Fon +49 (0)30 41 40 21-

Fax +49 (0)30 41 40 21-



www.spectaris.de

Berlin/Frankfurt, den 31. Januar 2017

Die Industrieverbände SPECTARIS und ZVEI vertreten die Interessen der Unternehmen aus den Bereichen Medizintechnologie und Diagnostik. Viele der Mitgliedsunternehmen gehören zu den kleinen und mittelständischen Unternehmen, die die Medizintechnologiebranche in Deutschland prägen. In Vertretung ihrer Mitglieder beteiligen sie sich am politisch-wissenschaftlichen Dialog und nehmen Stellungnahmerechte wahr.

Die Verbände beschränken sich nachfolgend darauf, nur zu ausgewählten Themen Stellung zu nehmen, die für ihre Mitglieder von besonderer Bedeutung sind. Sie behalten sich aber vor, im weiteren Verlauf des Verfahrens und der Diskussion zu einzelnen Punkten weitere Ausführungen zu machen.

Zu den einzelnen Themen:

Verfahren gemäß § 139b Absatz 5 SGB V

In §139b Absatz 5 wird auf die Möglichkeit hingewiesen, dass interessierte Einzelpersonen Vorschläge für Bewertungen machen können. In den Kapiteln 1.2.1.10 und 6 des Methodenpapiers erfolgen nähere Erläuterungen zu den Verfahren.

Es fehlen unserer Auffassung nach Kriterien, wie die Auswahl des IQWiG getroffen wird, welche der Anfragen für einen HTA-Bericht ausgewählt werden. Zwar werden in Kapitel 6.3.1 die entscheidenden Fragen erläutert, jedoch sind diese nur als „einige Kriterien“ bezeichnet. Hier sollten vertiefende Erläuterungen zu den Auswahlkriterien gemacht werden.

Ausgeschlossen wird im Methodenpapier die Bewertung von Arzneimitteln, jedoch ist nach dem Gesetz unseres Erachtens nicht vorgesehen, dass eine Produktbewertung stattfinden soll, sondern eine Bewertung von Methoden mit besonderer Patientenbedeutung, die von Einzelpersonen vorgeschlagen wurden.

Bewertung gemäß § 137h SGB V, insbesondere Kapitel 2.1.5 und 3.8 und 3.4

Es wird angeführt, dass das IQWiG optional zuzüglich zu den eingereichten Unterlagen eigene Recherchen durchführt. Wann das der Fall ist und welche Kriterien genau vorliegen müssen, ist allerdings aus dem Methodenpapier nicht ersichtlich. Unserer Ansicht nach muss hierzu detailliert dargestellt werden, welche Voraussetzungen vorliegen müssen, damit das IQWiG eine ergänzende Recherche als notwendig erachtet.

Anders als im Rahmen des § 137e Absatz 7 SGB V sind Medizinproduktehersteller nur unzureichend in das Verfahren zur Potenzialbewertung eingebunden. Ihnen obliegt nur eingeschränkt die Möglichkeit, die Informationen, die der Potenzialbewertung nach § 137h SGB V zu Grunde liegen, zu vervollständigen. Aus diesem Grund ist es von großer Bedeutung, inwieweit das IQWiG ergänzend Studien recherchiert.

Des Weiteren sollte die Veröffentlichung des Berichts des Institutes zeitgleich mit der Übermittlung an den GBA erfolgen und nicht erst 6 Wochen später, vor allem, damit alle Beteiligten im Stellungnahmeverfahren des GBA ausreichend Zeit haben, die Inhalte des Berichts angemessen zu berücksichtigen.

Unklar bleibt, welche Designanforderungen an die komplexe Beteiligung weiterer Krankenhäuser im Rahmen von Nutznachweisstudien gestellt werden, die an der begleitenden Studien zum zusätzlichen Evidenznachweis gestellt werden. Diesen speziellen Evidenznachweis spart das Methodenpapier des IQWiG aus. Es ist aber davon auszugehen, dass es im Aufgabenbereich des Instituts liegt, derartige Studiendesigns zu erstellen.

Wissenschaftliche Beratung, Kapitel 2.2.7

Erstmals wird in dem Methodenpapier auf die Beratungsfunktion des IQWiG hingewiesen, bei dem es in speziellen Fällen Studiensponsoren wissenschaftlich beraten kann.

Es fehlt an konkreten Erläuterungen, welche Voraussetzungen vorliegen müssen, damit das IQWiG Studiensponsoren zu Studiendesigns beraten darf. Zudem fehlt es an einer Erläuterung, in welchem Umfang das Befolgen der wissenschaftlichen Empfehlungen eine Anerkennung der wissenschaftlichen Evidenzgewinnung im In- und Ausland bewirken kann.

Nichtmedikamentöse therapeutische Interventionen, Kapitel 3.4 und 1

Grundsätzlich stellt das IQWiG die gleichen Anforderungen an die Evidenznachweise, unabhängig von den verwendeten Produkten. Nur in wenigen Ausnahmen werden keine verblindeten RCTs gefordert und Besonderheiten, wie z.B. Anwenderlernkurven, kurze Innovationszyklen, Patientenpräferenzen und fehlende Vergleichsinterventionen bei der Erstellung von Studiendesigns berücksichtigt.

Unserer Auffassung nach sind die Erläuterungen zu Ausnahmen und Anforderungen für Studien zum Evidenznachweis für Methoden unter Verwendung von Medizinprodukten genauer zu beschreiben und entsprechend angepasste Studiendesigns zu entwickeln. Die Möglichkeit, Zwischenbewertungen während des Studienverlaufs zu erheben, sollte eingeplant werden.

Generell weist das Methodenpapier in Kapitel 1 darauf hin, dass es keine einheitliche Definition für evidenzbasierte Medizin gibt, auch die Bewertung der Evidenz einzelner Studiendesigns in der Ergebnissicherheit scheint nicht einem allgemeingültigen Standard zu folgen.

Unabhängig von der wissenschaftlichen Auffassung des IQWiG, in welcher Form Evidenznachweise zu führen sind, gibt es offenbar noch keine allgemeingültigen, einheitlichen, wissenschaftlichen Standards für die Methodenbewertung. Es bleibt daher zu fordern, dass an einer Nachweismethode weiterhin erforscht und unterschiedliche Evidenznachweise auch weiterhin diskutiert werden müssen. Die Beschreibungen des Evidenznachweises im Rahmen des Methodenpapiers müssen daher einer ständigen wissenschaftlichen Überprüfung unterliegen. Wie dieser Pflicht nachzukommen ist, wird im Methodenpapier nicht beschrieben.

A.1.33 – Verband forschender Arzneimittelhersteller e. V. (vfa)

Autoren:

- Rasch, Andrej
- Werner, Sebastian

Stellungnahme

**zum Entwurf der IQWiG „Allgemeinen
Methoden“ Version 5.0**

Hausvogteiplatz 13
10117 Berlin
Telefon 030 206 04-0
Telefax 030 206 04-222
www.vfa.de

Einleitung

Seite 2/16

Das IQWiG hat inhaltliche Aktualisierungen und Ergänzungen in den „Allgemeinen Methoden“ des Instituts vorgenommen und den entsprechenden Entwurf der Version 5.0 zur Stellungnahme gegeben.

Der vfa nimmt Stellung zu Aktualisierungen in den folgenden Abschnitten:

- 2.2.1 Einbindung von Betroffenen
- 2.2.7 Wissenschaftliche Beratungen
- 3.1.4 Endpunktbezogene Bewertung
- 9.3.8 Meta-Analysen
- 9.3.9 Indirekte Vergleiche
- 9.3.10 Subgruppenanalysen
- 9.3.13 Auswertung abhängiger Daten

2.2.1 Einbindung von Betroffenen

Der Abschnitt wurde ergänzt:

- *Das IQWiG bindet bei der Berichterstellung (z. B. Dossierbewertung) Betroffene zu Beginn der Verfahren im Rahmen der Festlegung patientenrelevanter Endpunkte und relevanter Subgruppen ein. Bei der Auswahl von Teilnehmenden wird der Fokus auf tatsächlich Betroffene gerichtet, d.h. auf Patientinnen und Patienten, ihre Eltern oder sonstige Angehörige.*
- *Zur Einbindung von Betroffenen werden zwei Wege beschrieben: Zum einen kann ein persönliches Konsultationsgespräch geführt werden, bei dem Betroffene und Institutsmitarbeiter in kleinem Kreis über eine Erkrankung sprechen (z. B. Bericht, Rapid Report). Zum Zweiten können Sichtweise und Erfahrungen von Betroffenen schriftlich erfragt werden (z. B. Dossierbewertung).*
- *Die Findung wird regelhaft über die Patientenvertretung im G-BA organisiert, die selbst oder über ihre Mitgliedsorganisationen Betroffene zu einem Thema benennen soll. Ergänzend kann das IQWiG auch selbst Betroffene suchen oder anfragen über nationale oder lokale Selbsthilfeorganisationen oder -gruppen, über Krankenhäuser oder Arztpraxen, externe Sachverständige oder über sonstige Wege.*

Der vfa begrüßt das Vorgehen des IQWiG bei seiner Berichterstellung **Betroffene** einzubinden. Der vfa stimmt zu, dass Patientenvertreter oder Repräsentanten von Selbsthilfegruppen häufig selbst keine Patienten sind und so nicht in vergleichbarer Weise beurteilen können, wie Betroffene Symptome empfinden, ihre Funktionen und Aktivitäten wahrnehmen können, oder in ihrer Lebensqualität beeinträchtigt sind. Im Entwurf bleibt jedoch unklar, ob es sich bei der Einbindung um eine **Einzelperson** oder ggf. **Gruppen** von Betroffenen handelt. Grundsätzlich

ist eine Einbindung mehrerer Betroffener zu befürworten. Diese sollten den Behandlungsalltag in der Versorgung hinreichend widerspiegeln, in dem das zu bewertende Arzneimittel eingesetzt wird.

Seite 3/16

Das IQWiG bindet Betroffene bei der Festlegung **patientenrelevanter Endpunkte sowie relevanter Subgruppen** ein. Dieses Vorgehen ist sinnvoll und notwendig, da insbesondere die Fragestellung, welcher Endpunkt für den Patienten relevant ist, oftmals nur durch den Patienten selber sinnvoll beantwortet werden kann. Die Einbindung von Betroffenen greift nach Einschätzung des vfa für bestimmte Aspekte der Nutzenbewertung jedoch noch zu kurz: **Präferenzen** von Betroffenen hinsichtlich dieser Endpunkte sollten hier ebenso qualitativ erfasst werden. Im Rahmen von Dossierbewertungen, in denen stets eine Gesamtaussage zum Zusatznutzen getroffen wird, sind Patientenpräferenzen notwendig, um eine sinnvolle Abwägung von positiven und negativen Effekten zu gewährleisten. Eine solche Abwägungsentscheidung sollte nach Konsultation von Betroffenen stattfinden. Die Einbindung über ein persönliches Konsultationsgespräch ist auch im kurzen Zeitrahmen von Dossierbewertungen umsetzbar. Abwägungen ohne Berücksichtigung der Betroffenenperspektive können anderenfalls unangemessene Ergebnisse zur Gesamtaussage des Zusatznutzens liefern.

Die Initiative des IQWiG zur Einbindung von Betroffenen mit den oben formulierten Zielsetzungen ist grundsätzlich als positiv zu bewerten. Die Wirksamkeit einer solchen Maßnahme muss jedoch für Außenstehende überprüfbar sein. Deshalb ist eine umfängliche **Transparenz des Einbindungsprozesses** vom IQWiG zu fordern: Die Fragen des IQWiG, die Antworten der Betroffenen, mögliche Diskussionen sowie ein denkbar erreichter Konsens müssen nachvollziehbar und transparent offengelegt werden. Dabei sollten die Antworten von Betroffenen umfänglich gewürdigt werden und zudem abschließend dargelegt werden, wie die Einbindung der Betroffenen die Nutzenbewertung beeinflusst hat. Am Beispiel von Dossierbewertungen wäre demnach zu fordern, die im Rahmen der schriftlichen Einbindung gestellten Fragen des IQWiG und den entsprechenden Antworten der Betroffenen mit der Nutzenbewertung zu veröffentlichen. Zudem sollte die hier vom vfa vorgeschlagene qualitative Erfassung von Präferenzen von Betroffenen mittels persönlichem Konsultationsgespräch, nachvollziehbar dokumentiert mit der Nutzenbewertung veröffentlicht werden.

Die Findung bzw. **Auswahl der Betroffenen** soll gemäß Entwurf regelmäßig über die Patientenvertretung im G-BA organisiert werden. Ergänzend möchte das IQWiG auch eigenständig Betroffene suchen bzw. auswählen. Für beide Arten der Einbindung bleibt jedoch unklar nach welchen Kriterien eine Auswahl getroffen wird. Es ist notwendig ein standardisiertes Auswahlverfahren zur Beteiligung von Betroffenen zu verwenden. Dieses sollte **nachvollziehbar und transparent** sein. Das

IQWiG sollte die Organisationen, Krankenhäuser oder Arztpraxen benennen, wo Betroffene gesucht werden und die Schritte des Auswahlverfahrens transparent darlegen. Dies sollte ebenso für die Auswahl von Betroffenen über die Patientenvertretung im G-BA gelten. Auch hier sollte eine nachvollziehbare und transparente Auswahl von Betroffenen gewährleistet werden. Der vfa begrüßt ausdrücklich die Initiative des IQWiG eigenständig (d. h. ohne Rückgriff auf die Patientenvertretung im G-BA) zusätzliche Betroffene zu suchen bzw. auszuwählen. Der vfa befürwortet dabei das Vorgehen, Betroffene mit potenziellen finanziellen Interessenskonflikten von Konsultationen nicht kategorisch auszuschließen.

2.2.7 Wissenschaftliche Beratungen

Der Abschnitt wurde neu erstellt:

- *Das IQWiG beteiligt sich in speziellen Fällen an der wissenschaftlichen Beratung von Studiensponsoren (Scientific Advice), z. B. im Rahmen der Mitarbeit im europäischen Netzwerk EUnetHTA, mit dem Ziel Studiendesigns zu beraten, die aussagekräftige Daten für die Nutzenbewertung liefern.*
- *Um die Unabhängigkeit der Bewertung auch in den Fällen zu gewährleisten, in denen die Nutzenbewertung Studien enthält, zu denen das Institut eine Beratung durchgeführt hat, wird eine angemessene organisatorische Trennung von Beratung und Bewertung sichergestellt.*

Seit Einführung des AMNOG ist es vielfach zu **Diskrepanzen zwischen Bewertungen** von Zulassungsstudien durch EMA und G-BA bzw. IQWiG gekommen, aufgrund unterschiedlicher Anforderungen an Studiendesigns und zu erbringende Nachweise. Deshalb sind grundsätzlich Initiativen zu begrüßen, die eine wissenschaftliche Beratung von Studiensponsoren bezüglich dieser Anforderungen ermöglicht. Gemeinsame Beratungen eröffnen die Möglichkeit Diskrepanzen zu **reduzieren**.

Der vfa nimmt zur Kenntnis, dass sich das IQWiG an einer wissenschaftlichen Beratung von Studiensponsoren, z. B. im Rahmen der Mitarbeit im europäischen Netzwerk EUnetHTA beteiligt. Allerdings finden sich im Methodenentwurf **keine konkreten Erläuterungen** hinsichtlich der **Ausgestaltung** dieser Beteiligung, welches die Stellungnahme erschwert. Bspw. wäre es wünschenswert zu erfahren, in welchen „speziellen Fällen“ das IQWiG an Beratungen teilnimmt oder ob auch eine Mitarbeit auf europäischer Ebene in einem anderen Rahmen als EUnetHTA erfolgt. Insbesondere sollte klargestellt werden, ob IQWiG-Beratungen unter Beteiligung des G-BA vollzogen werden. Für den vfa ist es von großer Bedeutung, dass Institutionen, die Entscheidungsträger bei den Nutzenbewertungen sind, an diesen Beratungen teilnehmen.

Das IQWiG möchte laut Methodenentwurf eine angemessene **organisatorische Trennung von Beratung und Bewertung** erreichen, um die Unabhängigkeit der Bewertung auch in Fällen zu gewährleisten in denen das Institut eine Beratung durchgeführt hat. Eine solche organisatorische Trennung erscheint aus Sicht des vfa nebensächlich. Eher sollte sichergestellt werden, dass Beratungsinhalte auch verbindlich in der Bewertung berücksichtigt werden. Wird bspw. durch beteiligte Institutionen eine bestimmte Operationalisierung eines Endpunkts oder eine bestimmte Art der Auswertung empfohlen und vom Sponsor umgesetzt, sollte diese Umsetzung auch entsprechend in der Bewertung als „geeignetes Vorgehen“ gewürdigt werden. Um **verbindliche Beratungen sicherzustellen** wäre neben verbindlichen Beratungsprotokollen auch eine Zusammenarbeit von „Beratern“ und „Bewertern“ wünschenswert, die die gebotene Verbindlichkeit erhöhen könnte. Vor diesem Hintergrund erscheint es aus Sicht des vfa nicht zielführend eine organisatorische Trennung anzustreben.

3.1.4 Endpunktbezogene Bewertung

- *Im Abschnitt erfolgte eine Ergänzung der Anforderungen an Beleg aus einer Studie. Neu ist, dass in jeden Studienarm mind. 1000 Patienten eingeschlossen sein müssen.*

Es ist grundsätzlich nachvollziehbar, dass zur Ableitung eines Belegs aus einer Studie besondere Anforderungen erfüllt sein müssen. Die neue Anforderung einer fixen **Schwelle von mindestens 1000 Patienten** im Studienarm ist aber weder nachvollziehbar begründet noch inhaltlich sinnvoll. Nach Auffassung des vfa **sollte** sie deshalb **gestrichen werden**. In einigen Indikationen (z. B. Onkologie) sind Studien mit einer Fallzahl von über 1000 Patienten je Studienarm kaum realistisch. Die Einführung dieses Schwellenwerts würde somit für solche Indikationsgebiete das Ableiten eines Belegs aus einer Studie unmöglich machen. Die vorgeschlagene Ergänzung könnte in der Folge zur systematische Benachteiligung bestimmter Indikationen führen. Je nach Erkrankung, Fragestellung und Größe der zu erwartenden Effekte sind unterschiedliche Fallzahlen für Studien realisierbar, statistisch erforderlich und ethisch vertretbar. Die Anforderungen an die Ableitung eines Belegs aus einer Studie sollten daher die **indikationsspezifischen Besonderheiten** der Studiendurchführung **und** die Aspekte der zugrundeliegenden **Fallzahlplanung** adäquat **berücksichtigen**.

Vor diesem Hintergrund sollte nach Auffassung des vfa folgende Änderungen erfolgen: Die **Ableitung eines Belegs** aus einer Studie sollte in Anlehnung an Anforderungen der EMA vom **beobachteten Effekt und** dem zugehörigen **p-Wert** abhängen und nicht von einer fixen Schwelle der Fallzahl [1]. Dabei sollte für den beobachteten Effekt eine *etablierte*

Schwelle von $p < 0.05$ gelten, die „im hinreichendem Maße“ unterschritten wird (vgl. CPMP [1] „deutlich besser als das 5 %-Niveau“). Ein solches Kriterium wäre hinreichend variabel, um den indikationspezifischen Besonderheiten und der zugrundeliegenden Fallzahlplanung Rechnung zu tragen. Die Verwendung des vom IQWiG festgelegten Schwellenwerts von $p < 0,001$ ist nicht etabliert und zudem zu konservativ, um für alle Konstellationen gleichermaßen angemessen zu sein.

Eine weitere Anforderung an Beleg aus einer Studie ist, dass das Ergebnis innerhalb der Studie konsistent ist und hierfür Auswertungen für „**Teilmengen von Studienzentren**“ vorliegen. Für diese Anforderung erfolgt im Methodenentwurf keine Änderung. Die Erfahrungen aus abgeschlossenen Nutzenbewertungen zeigen jedoch, dass bezüglich der Umsetzung von Auswertungen zu „Teilmengen von Studienzentren“ Unklarheiten über ein adäquates Vorgehen bestehen. Hier wäre eine Klarstellung wünschenswert, wann eine Konsistenzprüfung über das (präspezifizierte) Merkmal „geografische Region“ eine adäquate Umsetzung der „Teilmengen von Studienzentren“ darstellen kann. Anderenfalls lassen sich widersprüchliche Bewertungen nicht erklären (z. B. akzeptierte Konsistenzprüfung über das Merkmal „geografische Region“ in der IQWiG Nutzenbewertung A14-28 und A16-15, keine Akzeptanz in den IQWiG Nutzenbewertung A15-52 oder A14-48).

9.3.8 Meta-Analysen

- *Das IQWiG empfiehlt bei der Durchführung von Meta-Analysen mit zufälligen Effekten nun die Knapp-Hartung-Methode. Bislang führte das IQWiG Meta-Analysen mit zufälligen Effekten nach DerSimonian und Laird-Methode durch.*
- *Studienergebnisse werden vom IQWiG regelhaft in Meta-Analysen statistisch zusammengefasst, wenn keine „zu große“ Heterogenität der Studienergebnisse vorliegt. Von einer statistischen Zusammenfassung wird nun abgesehen, falls der Heterogenitätstest einen p-Wert unter 0,05 liefert. Bislang wurde von einer Zusammenfassung abgesehen, wenn der Heterogenitätstest einen p-Wert unter 0,2 lieferte.*
- *Das IQWiG adressiert Limitationen der Anwendung von Meta-Analysen mit zufälligen Effekten für den Fall sehr weniger Studien (weniger als 5). Für diesen Fall sollte nun die Anwendung des Modells mit festem Effekt oder eine qualitative Zusammenfassung in Betracht gezogen werden. Kontextabhängig könnten aber auch alternative Verfahren in Betracht kommen, wie z. B. Bayes'sche Verfahren oder Methoden aus dem Bereich der generalisierten linearen Modelle. Bislang war es die Linie des IQWiG, dass vorrangig Modelle mit zufälligen Effekten angewendet werden sollten und nur in begründeten Ausnahmefällen auf Modelle mit festem Effekt ausgewichen wird.*

- *Das IQWiG konkretisiert die Bedingungen unter denen die Anwendung der Peto-Odds-Ratio-Methode adäquat ist. Dies ist der Fall, wenn die zu schätzenden Effekte nicht zu groß sind und kein unbalanciertes Design vorliegt.*

Seite 7/16

Die Anpassungen bezüglich der Durchführung von Meta-Analysen mit zufälligen Effekten mittels **Knapp-Hartung-Methode** werden vom vfa als sinnvoll erachtet. Der vfa begrüßt, dass neue methodische Entwicklungen aufgegriffen und im Entwurf des Methodenpapiers berücksichtigt und transparent gemacht werden. Problematisch ist jedoch, dass die im Rahmen der frühen Nutzenbewertung von Arzneimittel zu verwendenden Dossiertemplates noch die DerSimonian und Laird-Methode empfehlen. Eine Anpassung der Templates durch den G-BA wäre demnach hier von Nöten, um die Änderungen des methodischen Standards auch hier zu berücksichtigen.

Bislang wurde von einer **Zusammenfassung von Studienergebnissen** abgesehen, wenn der Heterogenitätstest einen p-Wert unter 0,2 lieferte. Dieser p-Wert wurde auf 0,05 abgesenkt. Der vfa hält diese Änderung für begrüßenswert und pragmatisch. Damit kann zukünftig besser vermieden werden, Heterogenität mit kleinem Umfang als bedeutsam zu identifizieren, obwohl diese möglicherweise nicht klinisch relevant ist. Eine Zusammenfassung von Studienergebnissen erscheint unter diesen Bedingungen sinnvoll.

Das IQWiG empfiehlt grundsätzlich die Anwendung von Meta-Analysen mit zufälligen Effekten. Das IQWiG würdigt im Entwurf aber auch ausdrücklich die Limitationen der Methode. Im Fall sehr weniger Studien (weniger als 5) kann demnach die Methode zu sehr breiten Konfidenzintervallen führen, sodass möglicherweise keine Aussagen zur Evidenzlage mehr möglich sind. Laut Entwurf sollte daher im **Fall sehr weniger Studien** die Anwendung des Modells mit festem Effekt bzw. alternative Verfahren in Erwägung gezogen werden.

Der vfa begrüßt die Änderungen der Empfehlungen zum methodischen Vorgehen bei wenigen Studien als Schritt in die richtige Richtung. Vor allem in der frühen Nutzenbewertung von Arzneimitteln liegen in der Regel nur sehr wenige Studien vor – zumeist zwei oder maximal drei. In diesen Fällen erscheint die Methode mit zufälligen Effekten als zu konservativ und damit ungeeignet. Der vfa hält deshalb die Empfehlung des IQWiG unter diesen Umständen ein **Modell mit festem Effekt** anzuwenden für **sinnvoll und richtig**. Dabei bedarf es nach Ansicht des vfa jedoch einer Präzisierung der verwendeten Formulierung „sollte in Erwägung gezogen werden“, da diese Interpretationsspielräume eröffnet. Hier sollte eine **deutlichere Formulierung** gewählt werden, wonach unter diesen Umständen *vorrangig* Modelle mit fixen Effekten angewendet werden sollten und nur in begründeten Ausnahmefällen auf Modelle

mit zufälligen Effekten ausgewichen werden soll. Aus Sicht des vfa ist es notwendig der besonderen Situation in der frühen Nutzenbewertung mit wenigen Studien ausreichend Rechnung zu tragen und die Anforderungen an die passenden Methoden hier so konkret wie möglich zu formulieren. Zudem sollte geprüft werden, inwieweit sich aus den Änderungen im Entwurf, Anpassungsbedarf im Dossiertemplate zur frühen Nutzenbewertung von Arzneimitteln ergibt.

Ergänzend, soll hier zudem erwähnt werden, dass die Öffnung der IQWiG-Methodik bezüglich der Anwendung weiterer „**alternativer Verfahren**“ ebenso als positiv zu bewerten ist. Vor dem Hintergrund der bestehenden Schwächen der Meta-Analysen mit zufälligen Effekten, sollte man neuen Verfahren die Möglichkeit geben, ihre Stärken zu zeigen.

Darüber hinaus wurden die Bedingungen unter denen die Anwendung der **Peto-Odds-Ratio-Methode** adäquat ist, auf Grundlage einer aktuellen Untersuchung von IQWiG-Mitarbeitern konkretisiert [2]. Der vfa begrüßt grundsätzlich die Anstrengungen des IQWiG neue methodischen Entwicklungen voranzutreiben. Jedoch werden in der Praxis der frühen Nutzenbewertung bereits konkrete Grenzwerte angewendet, die über die Darstellungen im Methodenentwurf hinausgehen. Diese Grenzwerte werden vom vfa kritisch gesehen. So wird bspw. in der erneuten Nutzenbewertung zu Sitagliptin (A16-44) explizit auf Tabelle 3 in Brockhaus et al. (2014) [2] verwiesen, wonach die Methode nur dann anwendbar wäre, wenn das beobachtete POR in Abhängigkeit vom jeweiligen Gruppengrößenverhältnis und einer 1,1-fachen tolerierten Abweichung zwischen den da angegebenen maximalen Effektstärken liegen müsse. Welches Ausmaß an Abweichung man hinzunehmen bereit ist, bestimmt dabei die Grenzwerte und damit die Frage, ob die Methode anwendbar ist. Die vom IQWiG in der Praxis tolerierte „1.1-fache“ Abweichung ist dabei willkürlich und ohne weitere Begründung festgelegt. Sie ergibt sich nicht zwangsläufig aus den Ergebnissen der Untersuchung, die ebenso die Möglichkeit einer größeren Abweichung eröffnen. Insgesamt sind deshalb die in der Praxis angewendeten Grenzwerte kritisch zu sehen. Zudem muss darauf hingewiesen werden, dass solche Grenzwerte nicht etabliert sind und von anderen Organisationen, wie z. B. Cochrane Collaboration nicht angewendet werden.

9.3.9 Indirekte Vergleiche

Im Abschnitt wurden u.a. die folgenden Aspekte aktualisiert:

- *Es wurde ergänzt, dass eine „adäquate Begründung“ erforderlich ist, um auf Methoden für indirekte Vergleiche zurückgreifen zu können.*
- *Aussagen zur Akzeptanz adjustierter indirekter Vergleiche sowie zur grundsätzlichen Ablehnung nicht adjustierter indirekter Vergleiche*

wurden konkretisiert. Demnach sind nach wie vor adjustierte indirekte Vergleiche die einzige akzeptable Methode, allerdings aktualisiert nur unter Verwendung „adäquater Brückenkomparatoren“. Die grundsätzliche Ablehnung nicht adjustierter indirekter Vergleiche wurde auf mögliche Methoden erweitert, in denen kein Brückenkomparator zur Verfügung steht und „über Modellierungen“ oder mithilfe von Adjustierungs- oder Matching-Verfahren „aus dem Bereich der Kausalmodelle für Beobachtungsstudien“ Effekte geschätzt werden.

- Die Ausführungen zu Methoden der Konsistenzprüfung wurden erweitert. Hier wurden mehrere Methoden benannt, jedoch keine Stellung zur Eignung der Methoden genommen. Das IQWiG verweist darauf, dass diese derzeit noch nicht ausreichend untersucht sind und sich kein Standard durchgesetzt hätte.
- Es wird ergänzt, dass für Situationen, in denen die Konsistenz nicht überprüfbar ist, die Ähnlichkeit mit besonderer Sorgfalt überprüft werden muss. Hier wird jedoch nicht weiter darauf eingegangen, wie konkret ein adäquates Vorgehen aussehen könnte.

Die **Beschreibungen** zum adäquaten Vorgehen bei indirekten Vergleichen bleiben trotz der vorgenommenen Aktualisierungen **weiterhin wenig konkret**. Aus Sicht des vfa wären Konkretisierungen dahingehend geboten, was genau „adäquate Brückenkomparatoren“ charakterisiert bzw. was eine „adäquate Begründung“ wäre, um auf Methoden für indirekte Vergleiche zurückzugreifen zu können. Des Weiteren wären weitere **Erläuterungen** in Hinblick auf die Konsistenzprüfung bei Netzwerk-Meta-Analysen **wünschenswert**. Da sich offenbar noch kein methodischer Standard etabliert hat, wäre es dennoch zu begrüßen, wenn das IQWiG eine Einschätzung zu möglichen Vor- bzw. Nachteilen der in der Literatur diskutierten Methoden abgeben könnte. Wenig konkrete Aussagen finden sich zudem für die Beschreibung einer geeigneten Methodik zur Bewertung der Ähnlichkeit von Studien im indirekten Vergleich. Bei der frühen Nutzenbewertung ist die unzureichende Ähnlichkeit von Studien aber eine der häufigsten Ablehnungsgründe für indirekte Vergleiche. Auch hier wäre es wünschenswert, wenn das IQWiG unterschiedliche **Ansätze diskutieren** würde.

In fast jedem dritten Verfahren der frühen Nutzenbewertung von Arzneimitteln mussten pharmazeutische Unternehmen in der Vergangenheit auf **indirekte Vergleiche** zurückgreifen. Daher kommt der Methodik für indirekte Vergleiche ein **hoher Stellenwert** zu. Vor dem Hintergrund der noch zahlreichen ungelösten methodischen Probleme, wäre es zu begrüßen, wenn verbindliche und ausreichende **Beratungen** zum adäquaten Vorgehen angeboten werden könnten.

Im Methodenentwurf werden **nicht adjustierte indirekte Vergleiche** (indirekte Vergleiche ohne Brückenkomparator) weiterhin **grundsätz-**

lich abgelehnt. Diese Ablehnung ist vor dem Hintergrund der zurückliegenden IQWiG-Bewertungspraxis **nicht nachvollziehbar**. Im Rahmen der frühen Nutzenbewertung von Arzneimitteln wurden bereits indirekte Vergleiche ohne Brückenkomparatoren unter bestimmten Bedingungen akzeptiert und für Aussagen zum Zusatznutzen herangezogen (z. B. Nutzenbewertung von Ledipasvir/Sofosbuvir bei chronischer Hepatitis C [A14-44]). In dieser Nutzenbewertung wurden die Bedingungen explizit beschrieben, unter denen nicht adjustierte indirekte Vergleiche akzeptiert werden. Beispielsweise könne der indirekte Vergleich akzeptiert werden, wenn der Effekt so groß ist, dass ausgeschlossen werden kann, dass er allein auf systematischer Verzerrung beruht (RR im Bereich von 10 oder höher). Die **Kriterien**, die in **der Bewertungspraxis** des IQWiG bereits Anwendung gefunden haben, sollten auch im Methodenentwurf **dargelegt** und diskutiert **werden**. Zudem sollte diskutiert werden, wie in solchen speziellen Fällen ein adäquates methodisches Vorgehen bei der Bewertung der Ähnlichkeit der Studien bzw. von möglichen Verzerrungsaspekten aussehen könnte.

Die **grundsätzliche Ablehnung** von indirekten Vergleichen ohne Brückenkomparator wurde im Methodenentwurf auf mögliche Methoden erweitert, in denen über Modellierungen oder mithilfe **von Adjustierungs- oder Matching-Verfahren** Effekte geschätzt werden. Diese kategorische Ablehnung von solchen Verfahren (z. B. MAIC-Methode) ist aus der Sicht des vfa **nicht sinnvoll**. Die Verfahren, die einen Vergleich ohne Brückenkomparator über eine Modellierung ermöglichen, werden stetig weiterentwickelt und systematisch auf ihre Eignung untersucht. Daher sollten sie nicht kategorisch abgelehnt werden. Vielmehr sollte die Möglichkeit bestehen, dass solche Verfahren bei einer dokumentierten Eignung auch für die Nutzenbewertung herangezogen werden können. So sollte z. B. geprüft werden, ob ein indirekter Vergleich unter Verwendung der MAIC-Methode ein adäquates Verfahren zur Ableitung des Zusatznutzens darstellt. Dabei sollte unter Angabe der Vergleichbarkeit aller prognostischen und klinisch relevanten Merkmale, die Richtung der potentiellen Verzerrung angegeben und bei der Bewertung berücksichtigt werden. Gerade bei konservativen Ergebnissen, d. h. die zuungunsten der zu überprüfenden Intervention verzerrt sind, sollte das Verfahren berücksichtigt werden. Zudem könnte, ähnlich wie bei der Akzeptanz von nicht adjustierten indirekten Vergleichen ohne Modellierungs-Verfahren (siehe oben), die Größe des Effekts berücksichtigt werden. Die Größenordnung des Effekts zur Akzeptanz von Vergleichen *mit* Modellierungs-Verfahren müsste jedoch deutlich unter der oben angewendeten Grenze von RR von 10 liegen, da mit den Modellierungs-Verfahren Verzerrungsaspekte (zumindest teilweise) kontrolliert werden. Der vfa ist deshalb der Auffassung, dass das IQWiG solche **Modellierungs-Verfahren nicht kategorisch ausschließen** sollte. Es sollte eher die Akzeptanz der Verfahren genau geprüft werden, unter Berücksichtigung

der Vergleichbarkeit der prognostischen und klinisch relevanten Merkmale, der Richtung einer potentiellen Verzerrung sowie der Größe des Effekts.

Seite 11/16

9.3.10 Subgruppenanalysen

Im neu verfassten Abschnitt wurden u.a. folgende Aspekte aktualisiert:

- *Änderung von Schwellenwerten: Streichung der Hinweis-Schwelle ($\alpha = 0,2$) für eine Effektmodifikation; Einführung eines neuen Schwellenwerts von mind. 10 Ereignissen in einer der Subgruppen bei binären Daten.*
- *Die in der methodischen Literatur diskutierten methodischen Limitationen von Subgruppenanalysen wurden überarbeitet. Entfallen ist die Problematik des multiplen Testens. Hinzugekommen sind Aspekte der Vergleichbarkeit von Behandlungsgruppen innerhalb von Subgruppen hinsichtlich prognostischer Faktoren (Imbalance) bei kleinen Fallzahlen sowie der Effektmodifikation durch mehr als ein Subgruppenmerkmal (Wechselwirkung höherer Ordnung).*
- *Es wird allgemein auf Kriterien zur Bewertung der Verlässlichkeit von Subgruppenanalysen hingewiesen. Anschließend erfolgen Ausführungen zur Notwendigkeit von Subgruppenanalysen trotz Limitationen: Begründung durch wissenschaftliche Notwendigkeit, beste wissenschaftliche Evidenz und sozialrechtliche Implikationen.*

Eine wesentliche Änderung bezieht sich auf das Signifikanzniveau für die Ergebnisse von Heterogenitäts- oder Interaktionstests zwischen Subgruppen. Ist ein Interaktionstest zum Niveau $\alpha = 0,05$ signifikant, wird von unterschiedlichen Effekten in Subgruppen ausgegangen bzw. werden die Subgruppenergebnisse mit jeweils getrennten Punktschätzern und Konfidenzintervallen berichtet. Im Vergleich zu Allgemeinen Methoden 4.2 entfällt damit Signifikanzniveau von $\alpha = 0,2$ für Hinweise auf unterschiedliche Effekte in den Subgruppen. Die **Streichung der Hinweis-Schwelle für eine Effektmodifikation** ist zu begrüßen, da damit die Wahrscheinlichkeit für falsch-positive Ergebnisse von Interaktionstests reduziert wird.

Für die Interpretation der Interaktionstests führt das IQWiG eine neue Regel bei binären Daten ein. Solche Analysen werden nur dann regelhaft durchgeführt, wenn **mindestens 10 Ereignisse** in einer der Subgruppen aufgetreten sind. Die Berücksichtigung eines Kriteriums für eine Mindestanzahl von Ereignissen ist grundsätzlich sinnvoll. Die Einführung des fixen Schwellenwerts „10“ ist im Methodenentwurf jedoch nicht hinreichend begründet und erscheint willkürlich. Eine Rationale für die Auswahl ist geboten.

Laut „Allgemeinen Methoden 4.2“, werden **Subgruppenanalysen in der methodischen Literatur** „sehr kritisch diskutiert“. Das IQWiG verweist hierbei auf zwei Literaturquellen [3,4]. Im vorliegenden Methodenentwurf verweist das IQWiG ebenfalls auf diese Literaturquellen, merkt jedoch an, dass die Subgruppenanalysen „teilweise kritisch diskutiert“ werden. Es ist nicht nachvollziehbar, aus welchen Gründen die Einschätzung relativiert wurde. Nach Auffassung des vfa werden Subgruppenanalysen **nach wie vor sehr kritisch diskutiert**. Die ursprüngliche Formulierung sollte daher beibehalten werden.

Einige der methodischen Limitationen von Subgruppenanalysen werden im Methodenentwurf nicht berücksichtigt. So entfällt im Vergleich zu „Allgemeinen Methoden 4.2“ das **Problem des multiplen Testens** bei Subgruppenanalysen. Bei einer großen Anzahl von Subgruppenanalysen können Ergebnisse irreführend sein [5]. Insbesondere im Rahmen der frühen Nutzenbewertung von Arzneimitteln kommt es regelhaft zu Analysen sehr zahlreicher Subgruppen. Dabei kann die Wahrscheinlichkeit bestehen, dass irgendein Subgruppenergebnis statistische Signifikanz erreicht, obwohl es sich in Wahrheit um ein zufälliges Ergebnis handelt. Diese Problematik sollte im Methodenpapier weiterhin benannt werden.

Ein im Methodenentwurf beschriebenes Problem ist die **Reduktion der Power** in Subgruppen. So stellt das IQWiG fest, dass Subgruppen oft nicht die Stichprobengröße erreichen, „die für das (inferenzstatistische) Aufdecken moderater Unterschiede nötig ist“. In der Folge muss „auch beim Vorliegen tatsächlicher Effekte das Ergebnis innerhalb einer Subgruppe nicht statistisch signifikant sein“. Dieses Problem wird zwar benannt, es findet sich jedoch **keine Lösung** zum Umgang mit dem **Power-Verlust**. Anzumerken ist, dass die beschriebene Problematik zudem dadurch verschärft wird, dass die Bestimmung des Zusatznutzenausmaßes auf Grundlage von Konfidenzintervallgrenzen mittels Schwellenwerten erfolgt (Anhang A). Dadurch kann der Nachweis eines Effektes innerhalb einer Subgruppe zusätzlich erschwert werden. Laut Methodenentwurf ist die Situation anders, „wenn schon bei der Fallzahlkalkulation eine für die Subgruppenanalyse angemessene Power berücksichtigt und ein entsprechend erhöhter Stichprobenumfang geplant wurde“. Dies ist zwar zutreffend, gilt jedoch nicht für die überwiegende Zahl der Subgruppenanalysen in der Praxis der frühen Nutzenbewertung.

Das IQWiG merkt zutreffend an, dass sich bei kleinen Fallzahlen die Behandlungsgruppen innerhalb der Subgruppen in prognostischen Faktoren unterscheiden können, falls keine stratifizierte Randomisierung nach Subgruppenmerkmalen erfolgt (**Problem der Imbalance prognostischer Faktoren bei kleinen Fallzahlen**). Die Ergänzung dieses Aspekts ist grundsätzlich sinnvoll. Aus Sicht des vfa ist jedoch anzumerken, dass eine stratifizierte Randomisierung nicht für alle Subgruppen

durchführbar ist. Gerade eine Verwendung von zu vielen Stratifizierungsmerkmalen erhöht das Risiko von Imbalancen innerhalb der einzelnen Merkmalskombinationen. Ferner können Subgruppen auch nachträglich regulatorisch gefordert werden, so dass eine Stratifizierung nicht möglich ist.

Ein neuer Aspekt ist die Beschreibung der Problematik der Effektmodifikation durch mehr als ein Subgruppenmerkmal (sog. **Wechselwirkung höherer Ordnung**). Der vfa stimmt zu, dass beschriebene Wechselwirkungen höherer Ordnung möglich sind. Gemäß IQWiG seien zur Interpretation der Subgruppenergebnisse separate überlappende Auswertungen (z. B. für jede Altersgruppe getrennt für Männer und Frauen) nötig. Anzumerken ist, dass einer vollständigen Auswertung aller denkbaren Wechselwirkungen höherer Ordnung Grenzen gesetzt sind. Hierfür sollten alternative methodische Ansätze geprüft werden, da der Interaktionstest an seine Grenze stößt. Zudem gilt das bereits oben beschriebene Problem der Reduktion der Power noch stärker für die immer kleiner werdenden Fallzahlen in den überlappenden Subgruppen. Ebenso verschärft sich das Problem des multiplen Testens. Auch die vom IQWiG festgelegte Mindestzahl von 10 Ereignissen für Subgruppenanalysen bei binären Daten wäre ggf. nicht mehr erfüllbar.

Nach den zuvor beschriebenen methodischen Limitationen stellt das IQWiG fest, dass eine **Bewertung der Verlässlichkeit von Subgruppenanalysen** erforderlich sei. Hinsichtlich der Kriterien einer solchen Bewertung wird auf die Arbeit von Sun et al. verwiesen [6]. Beispielhaft wird jedoch nur ein Kriterium aufgeführt (keine Verlässlichkeit von Subgruppenmerkmalen, die erst im Studienverlauf erhoben wurden). Die Diskussion zur Verlässlichkeit von Subgruppenanalysen auf der Basis der Arbeit von Sun et al. [6] ist damit unvollständig. In dieser Arbeit sowie in der folgenden Arbeit von Sun et al. 2012 [7] werden noch weitere Kriterien zur Beurteilung der Verlässlichkeit von Subgruppenanalysen genannt, die im vorliegenden Methodenentwurf jedoch nicht erwähnt werden. Dazu gehören, z. B. die biologische Rationale, die Konsistenz der Ergebnisse über mehrere Endpunkte oder die Konsistenz zu Ergebnissen ähnlicher Studien [6,7]. Nach Auffassung des vfa sollten diese **Kriterien** auch benannt und bei der Interpretation von Subgruppenanalysen **berücksichtigt werden**.

Im weiteren Verlauf des Methodenentwurfs finden sich nur unzureichende Ausführungen darüber, ob und inwieweit die benannten methodischen Limitationen bei der Interpretation von Subgruppenanalysen berücksichtigt werden. Vielmehr gibt das IQWiG eine Reihe von Gründen für eine **Notwendigkeit von Subgruppenanalysen** an, die die zuvor beschriebenen Limitationen relativieren. So merkt das IQWiG an, dass bei systematischen Übersichten eine Notwendigkeit zur Verwendung der

Ergebnisse von „*post-hoc*“ Subgruppenanalysen bestehe, sofern im Rahmen von systematischen Übersichten eben diese Subgruppen untersucht werden sollen. Trotz methodischer Limitationen könnten zudem Subgruppenanalysen „für manche Fragestellungen die auf absehbare Zeit beste wissenschaftliche Evidenz“ darstellen. Damit behält sich das IQWiG offenbar die Verwendung von Subgruppenanalysen zu Merkmalen vor, die im Studienverlauf erhoben wurden – ein Vorgehen, welches höchst fragwürdig erscheint. Darüber hinaus existiere eine sozialrechtliche Implikation nach § 139a Absatz 2 SGB V für die Berücksichtigung bestimmter Subgruppenanalysen.

Insgesamt entsteht damit im Methodenentwurf der Eindruck, dass unter bestimmten Umständen die methodischen **Limitationen** von Subgruppenanalysen **nicht berücksichtigt** werden müssen. Der vfa ist der Auffassung, dass bei der Interpretation von Subgruppenergebnissen die Limitationen stets berücksichtigt und Kriterien für die Bewertung der Verlässlichkeit der Auswertung angewendet werden sollten. Dazu gehört insbesondere die biologische Rationale oder die Konsistenz zu Ergebnissen ähnlicher Studien. Anzumerken ist, dass in der frühen Nutzenbewertung von Arzneimitteln solche o.ä. Kriterien regelhaft bei der Interpretation von Subgruppenergebnissen durch den G-BA berücksichtigt werden. Eine **Überprüfung der Verlässlichkeit** von Subgruppenanalysen erscheint daher sinnvoll, adäquat und geboten. Dies sollte ebenso für Fälle gelten, in denen das IQWiG die „Notwendigkeit“ von Subgruppenanalysen feststellt, bspw. aus „sozialrechtlichen Implikationen“.

Die Überprüfung der Verlässlichkeit von Subgruppenanalysen scheint insbesondere vor dem Hintergrund geboten aufgrund der sich abzeichnenden Entwicklung im Erstattungsbereich von Arzneimitteln durch das anstehende **Arzneimittelversorgungsstärkungsgesetz**. Hier ist vorgesehen, dass Medikamente für Subgruppen ohne belegten Zusatznutzen unter bestimmten Bedingungen von der Erstattung ausgeschlossen werden können. Zudem sollen Subgruppenunterteilungen im geplanten Arztinformationssystem hinterlegt werden. Dadurch könnten Subgruppenergebnisse in Zukunft eine noch bedeutendere Rolle für die Versorgung von Patienten spielen. Dabei besteht insbesondere die Gefahr, dass Patientengruppen, die eigentlich von neuen Arzneimitteln profitieren würden, diese dann nicht mehr bekommen. In einem sich so stark ändernden Rahmen, sind umso **höhere Anforderungen an die Glaubwürdigkeit von Subgruppenergebnissen** zu stellen.

9.3.13 Auswertung abhängiger Daten

Der Abschnitt wurde neu angelegt:

- *Wenn abhängige Daten vorliegen, so muss laut Methodenentwurf die damit verbundene Korrelationsstruktur durch Anwendung geeigneter Methoden für abhängige Daten berücksichtigt werden.*

- *Die Anwendung der einfachen Standardverfahren für unabhängige Beobachtungseinheiten führt laut Methodenentwurf zu einer Unterschätzung der Varianz und möglicherweise zu falschen Signifikanzaussagen.*
- *Die Evidenz aus Studien, die trotz korrelierter Daten einfache Standardverfahren verwendet haben, kann unter Umständen dennoch berücksichtigt werden, falls sich der damit verbundene Fehler gut einordnen lässt.*

Seite 15/16

Der vfa begrüßt die Ergänzungen bezüglich der Methodik zur Auswertung abhängiger Daten.

Das IQWiG gibt an, dass Studien, die trotz korrelierter Daten einfache Standardverfahren verwendet haben, unter Umständen dennoch berücksichtigt werden können, falls sich der damit verbundene Fehler gut einordnen ließe. Für diesen Aspekt, sollte ergänzt werden, unter welchen Bedingungen sich der Fehler gut einordnen lässt, d. h. mit welchen Methoden, in welchen Situationen, etc. dies möglich ist. Anzumerken ist zudem, dass dieser Maßstab einheitlich sowohl bei Nutzen als auch bei Schaden angewendet werden sollte.

Stand 31.01.2017

Referenzen

Seite 16/16

- [1] CPMP (2001) Points to consider on application with 1. Meta-Analysis; 2. One pivotal study (CPMP/EWP/2330/99)
- [2] Brockhaus AC, Bender R, Skipka G. The Peto odds ratio viewed as a new effect measure. *Stat Med* 2014; 33(28): 4861-4874.
- [3] Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet* 2000; 355(9209): 1064-1069.
- [4] Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med* 1992; 116(1): 78-84.
- [5] Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions, version 5.1.0 [online]. 2011. Kapitel 9.6.2 What are subgroup analyses? http://handbook.cochrane.org/chapter_9/9_6_2_what_are_subgroup_analyses.htm
- [6] Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ* 2010; 340: c117.
- [7] Sun et al. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. *BMJ* 2012;344: e1553

A.2 – Stellungnahmen von Privatpersonen

A.2.1 – Dintsios, Charalabos-Markos

Schriftliche Stellungnahme zum IQWiG Methodenentwurf Version 5.0

Dr. Ch.-Markos Dintsios

Düsseldorf, den 25.01.2017

[REDACTED]
[REDACTED]
[REDACTED]

Die Abgabe vorliegender Stellungnahme erfolgt als Privatperson.

Eine weitere institutionelle Stellungnahme mit Beteiligung des Stellungnehmenden bezieht sich auf andere Inhalte des Methodenentwurfs 5.0 und ist somit vollkommen unabhängig von dieser Stellungnahme zu betrachten.

6.1 Hintergrund und Ziel

In diesem Abschnitt wird auf das Sozialgesetzbuch (§ 139b (5) SGB V) rekurriert und u.a. die Aufgabe des IQWiG in Zusammenhang mit der Themenauswahl zur Anfertigung von Bewertungsberichten zu medizinischen Verfahren und Technologien (**HTA**) adressiert. Hier sollte in den Methodenentwurf noch angegeben werden, dass es eine Vorgängerinstitution, die **DAHTA** (Deutsche Agentur für HTA) beim DIMDI gab, die sich dieser Thematik schon seit mehr als eineinhalb Jahrzehnten zuvor angenommen hatte.

6.3 Auswahl der Themen für die HTA-Berichte

In diesem Abschnitt wird das Vorgehen zur Auswahl der Themen für die vom IQWiG laut Gesetz anzufertigender HTA-Berichte detailliert beschrieben. Nach Prüfung und Gewichtung der vorgeschlagenen Themen soll in einem **zweistufigen Verfahren** die Auswahl von bis zu 5 Themen, zu denen HTA-Berichte erstellt werden, erfolgen.

Im ersten Schritt ist ein eigens hierfür ins Leben gerufener Auswahlbeirat mit der Nominierung von 15 Themen für die Berichterstellung beauftragt, was insofern dieser Beirat repräsentativ für die verschiedenen Interessenkonstellationen im Gesundheitswesen und immer unter dem Primat der Patienteninteressen in Praxi sein sollte, zu begrüßen ist.

Im zweiten Schritt (Originalbezeichnung im Methodenentwurf 5.0 „Auswahlstufe 2“) werden die im Rahmen des ersten Schrittes nominierten 15 Themen durch die **Institutsleitung** des IQWiG bewertet, um im Nachgang zu einer Auswahl von bis zu 5 Themen, zu denen HTA-Berichte erstellt werden, im **Benehmen** mit Delegierten des **Stiftungsvorstands** des IQWiG und einer Vertreterin / eines Vertreters der KZBV zu gelangen.

In Abschnitt 6.3.4 (Zweite Stufe des Auswahlverfahrens: Auswahl von Themen, zu denen HTA-Berichte erstellt werden) wird festgehalten, dass zu den in der ersten Auswahlstufe nominierten Themenvorschlägen in der zweiten Stufe des Auswahlverfahrens eine **Vorabrecherche** nach Abschnitt 8.1.1 durchgeführt wird, auf deren Basis die Institutsleitung im Benehmen mit Vertretern der den Stiftungsrat bildenden Organisationen aus den in der ersten Stufe des Auswahlverfahrens bestimmten Themen bis zu 5 auswählt, für die HTA-Berichte erstellt werden.

Diese Ausführungen lassen spekulativ die Vermutung zu, dass die **Machbarkeit** des Unterfangens auf Basis einer wie auch immer gearteten Vorabrecherche als **Auswahlkriterium** von Bedeutung für Schritt 2 zu sein scheint. Da sicherlich auch Konstellationen eintreten können, in welchen mehr als 5 Themen machbar erscheinen, müssten weitere **Auswahlkriterien** herangezogen werden. Diese sollten ähnlich wie die unter Abschnitt 6.3.1 für den ersten Schritt

dokumentierten – auch wenn der Gewichtungsmodus noch nicht klar dargelegt wird – Kriterien auch explizit beim zweiten Schritt im Methodenentwurf zumindest genannt werden.

Hierzu könnte sich das IQWiG auch des Auswahlmodus der DAHTA bedienen, der vorbildlich und **demokratisch** für die **stimmberechtigten Institutionen** organisiert wurde, wie der Autor vorliegender Stellungnahme neben seiner ehemaligen IQWiG-Kuratoriumsmitgliedschaft aus seiner mehr als dreijährigen Mitgliedschaft im Kuratorium der DAHTA beim DIMDI zu berichten weiß. Ferner sollte hier festgehalten werden, dass eine finale Auswahl im Benehmen mit den Delegierten des Stiftungsvorstands des IQWiG viel Spielraum für die Institutsleitung des IQWiG bei der finalen Themenauswahl zulässt und ein Gremium, das nicht unbedingt primär Interessen von Patienten widerspiegelt, sondern durchaus Partikularinteressen verfolgen kann, kein Garant für die Berücksichtigung von Patienteninteressen sein muss. Auch stellt die Institutsleitung des IQWiG inklusive ihrer Stellvertretung wiederum keinen Garanten für die Berücksichtigung von Patienteninteressen bzw. für die Einordnung der Bedeutsamkeit von Themenvorschlägen für die **Patientenversorgung** dar, da diese trotz ihrer ärztlichen Ausbildung nicht klinisch aktiv ist und somit auch keinen Patientenkontakt im Rahmen des jeweiligen indikations- oder auch technologiebezogenen Versorgungskontexts pflegt. Darüber hinaus ist das Auswahlverfahren in seinem zweiten Schritt auch **manipulationsanfällig**, da theoretisch zumindest Themen, die für das Institut selber bzw. dessen Leitung von Interesse sind, auch über externe Personen im Sinne von „Erfüllungsgehilfen“ vorgeschlagen werden können und damit Eingang in die entsprechende **Themensammlung** finden können, um im Anschluss daran aufgrund des Auswahlmodus direkt institutsseitig priorisiert zu werden. Aus diesem Grund bietet sich eher ein **offenes Abstimmungsverfahren** für die finale Priorisierung in Anlehnung an das vom DAHTA in der Vergangenheit praktizierte Vorgehen an.

8. Informationsbeschaffung

Einleitend erfolgen einige Kommentare zu **definitivischen Abgrenzungen** von unterschiedlichen Begrifflichkeiten im Rahmen von Kapitel 8.

Informationsbeschaffung stellt kein Synonym von „**Information Retrieval**“ dar. Während in diesem Kapitel der gesamte Prozess des Information Retrieval dargestellt wird, also einschließlich Planung bzw. Definition von Recherchekomponenten, beschränkt sich die Informationsbeschaffung laut unterschiedlicher Publikationen der IQWiG Mitarbeiterin Hausner auf die Ausführung von Recherchen. Würde die Informationsbeschaffung des Instituts in den beschriebenen Arbeitsschritten ablaufen, gäbe es keinerlei Grund, eine eigene Methode zu beanspruchen. Auch müssten dann die entsprechenden Referenzen von Hausner et al. entfernt werden. Ferner wird keine Aussage getroffen, wer die Entscheidungen trifft, die **Fachexpertise** erfordern. Zudem wird nicht erläutert, warum im Text in extenso der Begriff Suche in all seinen Variationen verwendet wird, während man im Rahmen von Information Retrieval doch eher von **Recherche** spricht.

8.1 Umfassende Informationsbeschaffung

Hinsichtlich der Definition „**systematische Informationsbeschaffung**“ gelten die bereits im vorherigen Abschnitt getätigten Aussagen. Da es sich bei der von Hausner et al. beschriebenen Informationsbeschaffung lediglich um die Ausführung von Recherchen handelt, die auf die Zulieferung von Informationen angewiesen ist, allerdings auf eine Planung verzichtet, fehlen den Zuständigen Informationen zur Beurteilung, ob eine Informationsbeschaffung systematisch war

oder nicht. Hier fehlt somit der Hinweis im ersten Abschnitt des Kapitels auf die IQWiG-eigene Definition: Was ist IQWiG-systematisch und warum ist eine IQWiG-Informationsbeschaffung systematisch?

Des Weiteren wird die Darstellung der Berichterstattung der Informationsbeschaffung im jeweiligen Bericht **nicht spezifiziert** und diese Darstellung bleibt im Belieben des Instituts. Da unterschiedliche **Standards** existieren, ist eine Festlegung diesbezüglich leicht umsetzbar. Es wird zwar auf den **Publication Bias** eingegangen, das Phänomen des sogenannten **Retrieval Bias** wird allerdings im Kapitel überhaupt nicht abgehandelt.

Hinsichtlich der zu selektierenden **Datenbanken** bezieht sich die in der entsprechenden Textpassage zitierte Literatur ausschließlich auf Cochrane und NICE. Diese durchaus relevante Information, da dadurch das Vorgehen zweier von mehreren relevanten Institutionen in den Vordergrund gestellt wird, fehlt im Text. Außerdem werden keine hochwertigen Publikationen zu diesem Thema zitiert. Es soll angeblich eine **themenspezifische Ergänzung** erfolgen, ohne aber eine explizite **Verortung** im Rahmen des prozeduralen Ablaufs anzugeben.

Eine **externe Qualitätssicherung** scheint nicht vorgesehen zu sein und wird auch nicht im Methodenentwurf beschrieben, was durchaus befremdlich erscheint, ist der Inhalt dieses Kapitels als Ausgangspunkt für die Informationsgewinnung zu werten und somit durchaus einer externen Qualitätssicherung zuzuführen.

Darüber hinaus scheinen sich bei der Verwendung der **termini technici** einige gravierende Fehler eingeschlichen zu haben. So ist beispielsweise der Terminus **Suchquellen** nicht synonym mit **Informationsquellen**. Personen, z.B. Autoren, stellen keine Suchquellen dar, sondern Individuen für potenzielle Anfragen im Rahmen von Informationsbeschaffung.

Im letzten Absatz auf Seite 155 lautet es: „Publizierte Studien werden in Fachzeitschriften veröffentlicht und sind somit über bibliografische Datenbanken wie MEDLINE und Embase zu recherchieren. Hinweise auf unpublizierte Studien sowie Daten ergeben sich beispielsweise aus Einträgen aus Studienregistern sowie aus Studienberichten der Hersteller“. Der **Duktus** scheint falsch. Vielmehr müsste es heißen, dass folgende Quellen eingeschlossen werden und nicht, dass man folgende Quellen einschließen kann.

Zur Gegenüberstellung von **Snowballing** und **Similar articles-Recherchen** sollte festgehalten werden, dass similar articles-Recherchen in Medline Snowballing sind. In der nachfolgenden Beschreibung wird zudem, was Snowballing ist, wiederum sehr unzureichend das **comprehensive pearl growing** beschrieben.

Zur wissenschaftlichen Fragestellung und ihrer Aufteilung in einzelne **Konzepte** wird auf Seite 157 ff. Bezug genommen. Eigentlich lautet die korrekte Information Retrieval Terminologie „**Recherchekomponente**“. Die Fragestellung wird auch nicht „aufgeteilt“, sondern es werden die in der Fragestellung enthaltenen Recherchekomponenten **identifiziert** und **definiert**. Dabei können einzelne **PICO-Komponenten** durchaus auch mehrfach definiert werden und es existieren darüber hinaus auch **alternative Schemata**, die möglicherweise geeigneter sind als das PICO-Schema. Im Übrigen handelt es sich hier um einen Planungsschritt. Dies wird aber vom Institut strikt abgelehnt, bedarf es auch einer entsprechenden Fachexpertise. Nach den Publikationen von IQWiG-Mitarbeitern (Hausner) ist dies nicht Bestandteil des **so genannten „objektiven Ansatzes“**. Die Definition kann allerdings nicht automatisch erfolgen, sondern ist ausschließlich mit **Fachexpertise** möglich. Also entsteht hier ein klarer Widerspruch zwischen

dem vom IQWiG bzw. seinen Mitarbeitern in diversen Publikationen favorisierten Vorgehen und dem konventionellen Vorgehen nach dem PICO-Schema.

8.2 Fokussierte Informationsbeschaffung

Es existieren ferner **Systematiken** des Information Retrieval. Oberste Klassifikationsebene sind die „**systematischen**“ und „**nicht-systematischen**“ Recherchen. Eine Einteilung der Literaturrecherche in **umfassende** und **fokussierte** existiert in der Terminologie dieser Disziplin eigentlich nicht. Die „**umfassende systematische Recherche**“ ist der Fachterminus für eine bestimmte Recherchemethode.

Insbesondere stellt die „**fokussierte**“ **Informationsbeschaffung** ein **Eigenkonstrukt** dar, das so nicht im Information Retrieval anzutreffen ist. Im Rahmen des Information Retrievals gibt es **fokussierte Recherchefragestellungen**, wonach **punktueller Aspekte** recherchiert werden. Bei **Cluster-Recherchen** z.B. sind das aber auch systematische Recherchen – lediglich die Fragestellung ist sehr eng. Das Information Retrieval unterscheidet zwischen **umfassender, selektiver und zielgerichteter Recherche** als zugrundeliegende Strategie (englisch comprehensive, selective, purposive) (siehe bspw. Booth). Intendiert sind im Text des Methodenentwurfs höchstwahrscheinlich – allerdings schwer verständlich formuliert – Recherchen mit **reduzierter Sensitivität** und **erhöhter Präzision** (gegenüber der sonst offenbar unerheblichen Präzision), also womöglich nicht-systematische Recherchen. Die Referenz von Glenville passt dann nicht, da diese sich über **gesundheitsökonomische Filter** auslässt und auch nicht quantifiziert, was ausgewogene Sensitivität und Präzision meint.

Verfahren zu Erhöhung der Präzision gibt es seit Einführung von Medline 1966 bei Schlagworten (Major MeSH). Für **Freitextsuchen** hat Harter [1986] bereits eine Vielzahl an **Heuristiken** eingeführt. Sind stellen mittlerweile einen Standard bei wissensbasierten Recherchen. Aus dem Text geht das genaue Vorgehen der fokussierten Recherche allerdings nicht hervor. Limits als solche erhöhen nicht die Präzision. Außerdem existiert ein offensichtlicher Widerspruch zwischen „systematisch und transparent ...“ und Nichtdokumentation auf der folgenden Seite.

Zu den Systematischen Übersichtsarbeiten heißt es auf Seite 163 unten: „Außerdem können systematische Übersichten als Quelle für Primärstudien dienen, um auf Grundlage derer eine Nutzenbewertung durchzuführen [530]. Hierfür erfolgt vorab eine Prüfung der Informationsbeschaffung der identifizierten systematischen Übersicht(en) (siehe Abschnitt 8.5). Dann wird eine oder ggf. werden mehrere hochwertige und aktuelle systematische Übersicht(en) ausgesucht, deren Primärstudien extrahiert und dann selektiert werden. Bei diesem Vorgehen wird von der verwendeten systematischen Übersicht allein das Rechercheergebnis, nicht aber die Bewertung der eingeschlossenen Primärstudien oder die Datenextraktion übernommen. Zusätzlich erfolgt eine Aktualisierung der Informationsbeschaffung für den Zeitraum, der nicht durch die systematische Übersicht abgedeckt wird (siehe Abschnitt 8.1)“. Hier gilt es anzumerken, dass bei Einschluss des Rechercheergebnisses immer die ganze Recherche wiederholt werden müsste. Dies ist aber in der **Praxis** aufgrund der unvollständigen Dokumentation kaum möglich.

8.3 Orientierende Recherchen

Zur „**Orientierenden**“ **Informationsbeschaffung** wird folgende Definition präsentiert: „gezielt für bestimmte Fragestellungen nach geeigneten Daten zu suchen“. Dagegen definiert Information Retrieval nicht-systematische Recherchen mit eng definiertem Recherchethema als „selective

sampling“ oder „purposive sampling“. Im Text des Methodenentwurfs Version 5.0 ist aber nur die klassische **Vorrecherche** (pre-search) adressiert.

Der **sogenannten objektiven Ansatz** ist eigentlich nicht als solcher existent. Es handelt sich vielmehr um einen vom IQWiG so bezeichneten Ansatz. In der Realität geht es bei genauerer Betrachtung der eingesetzten Mittel um **Wortfrequenzstatistik**. EndNote wiederum ist ein **Literaturverwaltungsprogramm** und kein **Textanalysetool**. Die Verwendung „bereits bekannter Artikel“ hat nichts mit objektiv zu tun, sondern diese wurden subjektiv durch wen auch immer definiert und ist somit auch hochgradig **bias-** bzw. **fehlerträchtig**. Je heterogener die Intervention, Indikation oder Populationen, umso geringer werden Sensitivität und Präzision ausfallen.

Folgende **Widersprüche** in den getätigten Aussagen zur Informationsbeschaffung lassen sich identifizieren: Zur **Spracheinschränkung** erstens, dass a.) deutsch auch nicht-englisch ist und zweitens, dass die Aussage auf S. 155 impliziert, dass nur Publikationen, für die es einen bibliographischen Datensatz mit Abstract gibt, berücksichtigt werden. Somit würden Sekundärliteratur (z.B. HTA-Berichte) und graue Literatur oder Faktendaten usw. in der Regel ausgeschlossen. Zu **unpublizierten Daten** scheinen unterschiedliche Definitionen im IQWiG vorhanden zu sein [siehe hierzu Publikationen von Wieseler und Kreis]. Unpublizierte Daten sind **nicht qualitätsgeprüft** und sollten mit Bedacht und eher als Prüfdaten verwendet werden. Der ihnen durch das Institut beigemessene Stellenwert entspricht nicht immer ihrem Informationsgehalt.

Abschließend verweisen wir auf eine Reihe von Aussagen, die so nicht stimmen bzw. **fehlerhaft** sind oder **irreführend interpretiert werden könnten**:

Publizierte Studien und Datenbanken: Die Aussage stimmt so nicht. Studien werden z.B. in Fachzeitschriften publiziert und diese werden zum Teil in bibliographischen Datenbanken indexiert. „...sind somit zu recherchieren ...“ ist unzutreffend. Etwa ein Fünftel aller wissenschaftlichen medizinischen Zeitschriften sind in Medline indexiert (in unterschiedlicher Vollständigkeit). In Embase sind es noch weniger. Somit kann nur ein kleiner Teil der publizierten Studien über die Datenbanken Medline oder Embase identifiziert werden.

Bereitgestellte Informationen aus **anderen Suchquellen** (S. 156 oben): **Studienberichte** sind **Faktendaten**, keine Datenbanken. Die bereitgestellten Informationen aus Suchquellen werden auch nicht für (Daten-)Auswertungen verwendet. Was eine **zielgerichtete Auswertung** genau sein soll, bleiben die Autoren dieses Abschnitts dem Leser schuldig.

Studienfilter: Es gibt **keine validierten Studienfilter** als solche, sondern nur teilweise sehr unzureichend **evaluierte Filter**. Auch wenn manche Autoren das Wort „validiert“ inflationär verwenden, ihre Suchstrategien zu Filtern erklären, 1-3 Beispielthemen exemplarisch auswerten, hat das nichts mit „Validierung“ im eigentlichem Sinne zu tun. Selbst bei großen Prüfdatensatzmengen (UK und Canada) sind die Ergebnisse sehr unterschiedlich. Fast alle sogenannten **Standardfilter** müssen für eine adäquate Sensitivität und Präzision themenspezifisch angepasst werden. Dies wird derzeit grundlos vom IQWiG ignoriert.

Qualitätssicherung: „Aufgrund ihrer Komplexität sind Suchstrategien für bibliographische Datenbanken fehleranfällig“. Diese Aussage ist nicht Gegenstand des referenzierten Aufsatzes von Sampson. Auch sind Suchstrategien in bibliographischen Datenbanken nicht zwangsläufig komplex. Die Vollständigkeit vordefinierter Datensätze vor Durchführung der entwickelten Suchstrategien hat nichts mit der von Sampson diskutierten Qualitätssicherung und

Fehlervermeidung zu tun. Vielmehr wird durch einen solchen Schritt die Wahrscheinlichkeit von manuellen Manipulationen sehr hoch.

Weitere Suchtechniken: Bereits 1967 gab es eine Vielfalt an unterschiedlichen Recherchemethoden – ein systematisches Review zitierte ca. 50. Die Referenz [180] ist lediglich eine „Level of Evidence 5“-Quelle und nicht qualitätsgeprüft. Sie sollte deswegen auch in den Allgemeinen Methoden des IQWiG nicht Eingang finden. Diese Quelle wurde zudem von den Kapitelautoren selbst verfasst, gibt also ausschließlich deren Wissensstand wieder. Es muss gewiss nicht diskutiert werden, dass Boolesche Logik keine Recherchemethode ist. Das moderne Information Retrieval wurde bereits in den 50er und 60er Jahren entwickelt. Seither sind unzählige **Lehrbücher** erschienen. Allerdings wird entweder aus Unwissen oder auch gezielt keines zitiert. Die Bedeutung „**schwer recherchierbare**“ **Fragestellungen** erschließt sich dem Leser nicht, wird im Methodenentwurf kein Bezug genommen, für wen dies gilt. Etwa für die verantwortlich zeichnenden IQWiG-Mitarbeiter? Die explizit genannte Recherchemethode „**citation tracking**“ birgt ein hohes Potential für **unterschiedliche Biasarten**. Je nach Einsatz kann eine erhebliche Verzerrung des Rechercheergebnisses erfolgen. Diese wird im Methodenentwurf vollends ignoriert.

Literatur

Booth A. "Brimful of STARLITE": toward standards for reporting literature searches. J Med Libr Assoc. 2006 Oct;94(4):421-9, e205.

Harter S. Online Information Retrieval: Concepts, Principles, and Techniques. Orlando: Academic Press;1986.

Wieseler B, Wolfram N, McGauran N, Kerekes MF, Vervölgyi V, Kohlepp P, Kamphuis M, Grouven U. Completeness of reporting of patient-relevant clinical trial outcomes: comparison of unpublished clinical study reports with publicly available data. PLoS Med. 2013 Oct;10(10):e1001526.

Kreis J, Panteli D, Busse R. How health technology assessment agencies address the issue of unpublished data. Int J Technol Assess Health Care. 2014 Jan;30(1):34-43.