



gmds

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 (0)2236 33 19 958
Telefax: +49 (0)2236 33 19 959

E-Mail: info@gmds.de
Internet: www.gmds.de

Geschäftsführung

Beatrix Behrendt

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 Präsidiumscommission IQWiG angefertigt.

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

Mit freundlichen Grüßen

Prof. Dr. Ulrich Mansmann
- Präsident der GMDS -

Prof. Dr. Tim Friede
- Präsident der IBS-DR -

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

Anlagen: - Formblatt zur Stellungnahme
- Stellungnahme GMDS / IBS-DR

Dokument:	Allgemeine Methoden: Entwurf der Version 5.0
------------------	---

--

Formblatt zur schriftlichen Stellungnahme zum Methodenentwurf

Name, Vorname; Titel und Funktion des/der Stellungnehmenden
1. Mansmann, Ulrich ; Professor, Präsident der GMDS
2. Friede, Tim; Professor, Präsident der IBS-DR
3.
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
<input checked="" type="checkbox"/> im Namen folgender Institution/Organisation: GMDS und IBS-DR
<input type="checkbox"/> als Privatperson(en)

Kontaktdaten
Name: Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie GMDS e.V. / Deutsche Region der Internationalen Biometrischen Gesellschaft (IBS-DR) Anschrift: GMDS: GMDS-Geschäftsstelle, Industriestraße 154, D-50996 Köln IBS-DR: IBS-DR Geschäftsstelle, c/o Institut für Biometrie, Epidemiologie und Informationsverarbeitung Tierärztliche Hochschule Hannover, Bünteweg 2, D-30559 Hannover Tel./-Fax-Nr GMDS: (02236) 33 19 958, (02236) 33 19 959 IBS-DR: (0511) 953 79 51, (0511) 953 79 74 E-Mail-Adresse: GMDS: info@gmds.de, IBS-DR: biometrische-gesellschaft@tiho-hannover.de

Einwilligung

Ich willige ein, dass meine/unsere Stellungnahme im Wortlaut unter Nennung meines Namens sowie ggf. des Namens der Institution/Organisation im Internet veröffentlicht wird.

Rechtseinräumung

Sofern ich/wir *bisher unveröffentlichte* Dokumente der Stellungnahme beifüge(n) oder im Zusammenhang mit der Stellungnahme nachträglich überreichen, räume(n) ich/wir dem Institut für diese das inhaltlich, zeitlich und örtlich unbegrenzte, nicht ausschließliche Veröffentlichungsrecht, Vervielfältigungsrecht, Verbreitungsrecht und das Recht der

Dokument:	Allgemeine Methoden: Entwurf der Version 5.0
------------------	---

öffentlichen Wiedergabe, jeweils in allen gegenwärtig bekannten Formen, sowie das Archivierungs- und Datenbankenrecht ein.

Ich/wir garantiere(n), über diese Rechte verfügungsberechtigt zu sein, und stelle(n) das Institut von Ansprüchen Dritter frei, die im Zusammenhang mit einer Wahrnehmung der vorgenannten Rechte durch das Institut erhoben werden sollten. Zu den erstattungsfähigen Kosten zählen dabei auch die angemessenen Kosten einer Rechtsverfolgung und Rechtsverteidigung, die dem Institut zur Abwehr von Ansprüchen Dritter entstehen sollten. Das Institut wird Sie unverzüglich über vorzunehmende Maßnahmen der Rechtsverfolgung und Rechtsverteidigung informieren und Ihnen die Möglichkeit geben, Ihrerseits das Verfahren gegen den bzw. die Dritten zu führen.

Köln, Hannover, 31.01.2017

Ort/Datum

W. Asmann

Unterschrift 1

Tim Biele

Unterschrift 2

Unterschrift 3

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 gewichtet 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 widerspricht 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 Nutzensgewinn 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 Zensierung 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.