

Stellungnahme zum Entwurf des Methodenpapiers 4.2  
(Version vom 18.06.2014)  
des Instituts für  
Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG)

---

07. August 2014

Prof. Dr. Axel Mühlbacher  
Hochschule Neubrandenburg

Hochschule Neubrandenburg  
Brodaer Straße 2  
17033 Neubrandenburg

Tel.: +49 3 95 5693-3108/ 3304  
Fax: +49 3 95 5693-73108  
Email: [muehlbacher@hs-nb.de](mailto:muehlbacher@hs-nb.de)

## Gliederung

<b>Einleitung und Gegenstand</b>	<b>4</b>
<b>Begründung der Fokussierung der Stellungnahme</b>	<b>6</b>
<b>Stellungnahme Kapitel 4: Nutzen</b>	<b>10</b>
Maß des Gesamtnutzens als relevante Entscheidungsgröße	10
Alternative Vorgehensweise für die Bestimmung des Gesamtnutzens	10
Skalenniveau des transformierten Nutzens	12
<b>Stellungnahme Kapitel 3: Nutzenbewertung medizinischer Interventionen</b>	<b>14</b>
Evidenzgrad über die Werturteile in der frühen Nutzenbewertung	14
Patientenzufriedenheit und Nutzendimensionen	14
Begriff und Definition des Nutzen	15
Fakultative Würdigung von Schaden und Nutzen	17
Würdigung der Nutzen- und Schadensaspekte	17
Beliebigkeit der Abwägung von Nutzen und Schaden	19
<b>Stellungnahme Kapitel 2: Produkte des Instituts</b>	<b>20</b>
Informationsbeschaffung und wissenschaftliche Bewertung	20
Potenzialbewertung: Zielkriterien und Werturteile	21
Potenzialbewertung: Indikationsübergreifende Vergleichsgrößen	22
<b>Stellungnahme Kapitel 1: Das Institut</b>	<b>23</b>
Darstellung der wissenschaftlichen Grundlagen	23
Die Bereitstellung von Informationen über Prioritäten und Werturteile	24
Anforderungen an die Evidenz der Werturteile und Präferenzen	25
<b>Lösungsvorschlag: Messen, Bewerten &amp; Entscheiden</b>	<b>26</b>
Phasen der Nutzenbewertung	27
Empfehlungen von Arbeitsschritten bei der Nutzenbewertung	31
Anforderungen an das Skalenniveau	32
<b>Zusammenfassung</b>	<b>33</b>
<b>Literatur / Quellen</b>	<b>36</b>

## *Abkürzungsverzeichnis*

AHP	Analytic Hierarchie Process
CA	Conjoint Analyse
EbM	Evidenzbasierte Medizin
EG	Effizienzgrenze
FNB	Frühe Nutzenbewertung
GBA	Gemeinsamer Bundesausschuss
GKV	Gesetzliche Krankenversicherung
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
KNB	Kosten-Nutzen Bewertung
MdG	Maß des Gesamtnutzens
SGB	Sozialgesetzbuch
ZN	Zusatznutzen

## Einleitung und Gegenstand

Seit der Vorstellung des ersten Entwurfs des Methodenpapiers im Jahr 2009 wurden zahlreiche Stellungnahmen geschrieben und die methodischen Grundlagen der Nutzenbewertung in Deutschland diskutiert. Führt bereits der Vorschlag zur Effizienzgrenze (EG) mit dem Maß des Gesamtnutzens zu kontroversen, wissenschaftlichen Diskussionen, dann werfen die Anpassungen an die Regelungen des AMNOG im Kontext der frühen Nutzenbewertung bzw. die Bestimmung des Ausmaßes des Zusatznutzens weitere methodische Fragen auf. Die am 28. November 2013 publizierte Fassung "Allgemeine Methoden 4.1" zeigte Änderungen im Kapitel 3 und dort die Operationalisierung des Ausmaßes eines Zusatznutzens (Kapitel 3.3.3), unter anderem durch die Methoden der Präferenzmessung und Ansätze der multikriteriellen Entscheidungsfindung. Im Anhang wurde eine umfassende Begründung zur Rationale der zugrundeliegenden Methodik der Operationalisierung angefügt. In dem jetzt vorliegenden Entwurf wurden die Methoden zur Kosten-Nutzen Bewertung integriert, wesentliche Inhalte des Kapitels 3 (frühe Nutzenbewertung) gestrichen und in einem neuen Kontext für das Kapitel 4 (Kosten-Nutzen Bewertung) aufbereitet.

**Evidenzbasierte Medizin als Grundlage:** Heute – fünf Jahre nach dem ersten Entwurf – sind die Frage- und Problemstellungen des Messens von Zielkriterien und Nutzendimensionen im Zusammenhang mit der evidenzbasierten Medizin ausführlich dargestellt. Die praktische Implementierung der evidenzbasierten Medizin, die höchsten Anforderungen an das Studiendesign und die tiefgründige Analyse der Ergebnissicherheit garantieren die bestmögliche Information für die Beurteilung der kausal begründeten Effekte einer Intervention im Vergleich mit dem Status Quo oder Placebo. Die Messung der klinischen und nicht-klinischen Effekte ist eine notwendige Bedingung für die Entscheidung über den Einsatz, die Erstattungsfähigkeit und die Festsetzung der Preise von Gesundheitstechnologien. Es ist eine notwendige, jedoch keine hinreichende Bedingung.

**Werturteile und Präferenzen:** Eine rationale Entscheidung ist mit diesen Informationen nicht möglich. Über das gesamte Methodenpapier bleiben Fragen zur Bewertung der gemessenen klinischen oder nicht-klinischen Effekte offen. Informationen über Werturteile und Präferenzen der Betroffenen werden bei der Gewichtung der Zielkriterien nicht systematisch berücksichtigt. Es entsteht der Eindruck, dass der Nutzen einer Intervention bei der frühen Nutzenbewertung (FNB) anders bewertet werden soll, als bei der Kosten-Nutzen Bewertung (KNB). Wenn ja, dann stellen sich folgende Fragen: Auf Basis welcher methodischen und theoretischen Grundlagen unterscheiden sich diese Bewertungsverfahren? Welche wissenschaftlichen Annahmen und Werturteile sind Basis der unterschiedlichen Bewertung? Welche spezifischen Anforderungen werden an das Skalenniveau bei der frühen Nutzenbewertung bzw.

der Kosten-Nutzen Bewertung gestellt? Unklar bleibt, mit welchen Annahmen und Werturteilen das Maß des Gesamtnutzens (MdG) bzw. das Ausmaß des Zusatznutzens (ZN) in einem eindimensionalen Nutzenwert quantifiziert werden soll. Da eine konsistente wissenschaftliche Nutzenbewertung für die FNB und die KNB des Institutes für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) eine wesentliche Voraussetzung für konsistente und valide Entscheidungen über die Erstattungsfähigkeit oder Preisfestsetzung sind, kann von einem erheblichen Handlungsbedarf ausgegangen werden.

**Gegenstand der Stellungnahme:** Das Institut weist zu Beginn darauf hin, dass „Nur die als Änderungen gekennzeichneten Textteile (...) Gegenstand des Stellungnahmeverfahrens“ sind (S. ii). Inhaltliche Änderungen in einigen Abschnitten mögen begründeten Änderungs- oder Klärungsbedarf in anderen Abschnitten nach sich ziehen (auch wenn diese nicht im neuen Entwurf geändert wurden). Im neuen Entwurf werden konkrete Verweise auf andere Abschnitte in den Änderungen gemacht. So wird im ersten Satz des Abschnitts 4.3 direkt auf den Abschnitt 3.3 verwiesen und festgestellt, dass in diesem Abschnitt die Methoden „mit denen der Nutzen im Rahmen der Nutzenbewertung bestimmt wird“ beschrieben werden (S. 86). Trotz der zahlreichen Änderungen in den Kapiteln 3 und 4 wurden in diesem Abschnitt keine Änderungen oder Anpassungen vorgenommen. In Abschnitt 3.1.5 wird wiederum für die themenspezifische Konkretisierung und gleichzeitige Würdigung von Nutzen und Schaden auf Abschnitt 4.3.3 verwiesen (S. 49). In Abschnitt 4.3.3 sollen die Methoden für die dem Kapitel 3 zugrundeliegenden Entscheidungen erläutert werden. Die Einschränkung der Stellungnahme auf die gelb markierten Bereiche ist nicht zielführend, da dem Stellungnehmenden allgemein auch die Möglichkeit eingeräumt werden sollte, auf Kritikpunkte der vergangenen Jahre einzugehen, wenn diese im Methodenpapier keine Berücksichtigung gefunden haben. Im Sinne der Verfasser sollte es auch sein, dass die Stellungnehmenden ebenso auf Verbesserungspotential oder Korrekturbedarf hinweisen, auch wenn die Textpassagen nicht Gegenstand der Änderungen waren. Neue wissenschaftliche Erkenntnisse können die Beurteilung des Methodenpapiers über die Zeit beeinflussen, gerade dann, wenn diese Abschnitte in der Fortschreibung der Methoden nicht berücksichtigt wurden. Eine Einschränkung der Stellungnahmen auf die tatsächlich geänderten Textstellen entspricht nicht der Idee die (Fach-)Öffentlichkeit an der Weiterentwicklung der Methoden zu beteiligen. Zur Konkretisierung der adressierten Frage- und Problemstellungen des aktuellen Entwurfes des Methodenpapiers in der Version 4.2 ist es notwendig zunächst einzelne Aspekte aus dem neuen Kapitel 4 herauszuarbeiten. Fortfolgend werden dann erst die resultierenden Fragen in den vorangehenden Kapiteln 3, 2, 1 diskutiert.

## Begründung der Fokussierung der Stellungnahme

Für die Begründung weiterführender entscheidungsunterstützender Informationen als Grundlage für die Nutzenbewertung des IQWiG im Kontext der Entscheidungen der Selbstverwaltung, soll hier die Analogie „von zwei Läufern in einem fairen Rennen“ (S.7) im Methodenbericht aufgegriffen werden.

**Messen der Leistung:** Anhand des Laufwettbewerbs kann eindrücklich die Problematik des Messens und der Ergebnissicherheit verdeutlicht werden. Es besteht die Möglichkeit, die Zeit oder aber auch die zurückgelegte Wegstrecke zu messen. Um einen fairen Wettkampf zu garantieren, müssen die Ausgangsvoraussetzungen für beide Läufer gleich sein. Laufen die Wettkämpfer gleichzeitig gegeneinander, braucht es noch nicht einmal eine Stoppuhr, um einen Sieger zu bestimmen (Ein Messinstrument wird erst benötigt, wenn der Wettkampf zeitversetzt stattfindet). Je näher die zwei Läufer beieinander liegen, desto genauer muss die Messung erfolgen. Je knapper das Ergebnis desto schwerwiegender ist die Ergebnisunsicherheit. Messverfahren und Ergebnisunsicherheit müssen daher transparent sein.

**Entscheidung über die Platzierung:** Es können unterschiedliche Entscheidungsregeln zur Bestimmung der Rangordnung genutzt werden. Die Entscheidungsregeln haben Auswirkungen auf die Durchführung der Messung. Wird das Minimalprinzip für die Festlegung eines Siegers herangezogen, dann wird ein fixer Streckenabschnitt bestimmt, welchen die Läufer so schnell wie möglich zurücklegen müssen (um zu gewinnen, bedarf es der minimalen Zeit). Denkbar wäre aber auch das Maximalprinzip, wobei der Sieger des Wettkampfes die längste Strecke innerhalb eines bestimmten Zeitbudgets zurücklegen muss. Überträgt man dieses Beispiel auf die Bewertung von Gesundheitstechnologien, so stellt sich die Frage, ob die Entscheidungslogik von einem festen Budget ausgeht (Maximalprinzip) oder ob ein(e) zuvor bestimmte Leistung/Qualität/Nutzen mit dem minimalen Mitteleinsatz bewerkstelligt werden soll (Minimalprinzip). Unabhängig von der Entscheidungsregel, ist die Rangreihung – also die Feststellung des Siegers – in diesem Beispiel des Wettrennens sehr einfach, da es sich hier nur um eine Disziplin handelt.

**Komplexes Entscheidungsproblem:** Das Messen der Zeit oder der Wegstrecke ist nur eine verkürzte Darstellung des Bewertungsproblems bei der Bewertung von Gesundheitstechnologien. Realistischer ist die Suche nach dem besten Athleten: hierzu bedarf es mehrerer Wettkämpfe, wie zum Beispiel bei einem Zehnkampf. Es zählen nicht nur Schnelligkeit, sondern auch andere Disziplinen wie Kraft, Ausdauer oder Koordination. Analog der klinischen Realität müssen unterschiedliche Bewertungskriterien (Disziplinen) für die Auswahl des Siegers herangezogen werden, damit die

Überlegenheit eines Athleten A gegenüber B bestimmt werden kann. Um einen fairen Wettkampf zu garantieren, müssen die Disziplinen (Zielkriterien) den Athleten vorab bekannt sein.

**Vergleichbarkeit der Leistungen:** Man kann nicht davon ausgehen, dass einer der Wettkämpfer alle Disziplinen gewinnt. Dem Vergleich der Leistungen in allen zehn Disziplinen müssen Werturteile zugrunde gelegt werden. Aus diesem Grund ist die Basis der fairen Entscheidung, dass vor dem Vergleich die Annahmen über die Vergleichbarkeit bzw. die Methoden des Vergleichs der gemessenen Ergebnisse nachvollziehbar gemacht werden (vgl. das Punktesystem im Zehnkampf).

**Gewichtung der Leistungen:** In einem weiteren Schritt ist zu klären, welcher Wert den einzelnen Disziplinen beigemessen wird. Im Falle des Zehnkampfes gehen wir davon aus, dass jede Disziplin mit dem gleichen Anteil in die Bewertung der Gesamtleistung eingeht (Was auch eine bewusste Entscheidung sein kann, dann aber einer entsprechenden Rationale bedarf). Denkbar wäre es aber auch, dass der Sprint höher gewichtet werden soll als die Langstrecke, die Langstrecke höher wie das Kugelstoßen und das Kugelstoßen höher wie der Speerwurf. Demnach müssen die gemessenen Leistungen entsprechend ihrer Relevanz bei der Ermittlung des Gesamtsiegers gewichtet werden. Diese Gewichtung könnte abhängig sein von den Wertmaßstäben der Athleten oder von Experten. Die Gewichtung und die Annahmen über die Vergleichbarkeit haben erheblichen Einfluss auf die Gesamtwertung, deshalb sollten die zugrundeliegenden Werturteile transparent gemacht werden. Vorab sollten ausreichend Informationen vorliegen, um über die Gewichtung der Disziplinen zu entscheiden.

**Verzerrungen der Wettkampfergebnisse:** Werden die einzelnen Wettkampfdisziplinen festgelegt, nachdem der Sportler bereits trainiert hat, kann dies zu Verzerrungen und strategischen Entscheidungen führen. Sind weder dem Publikum noch dem Sportler die Annahmen über die Vergleichbarkeit oder die Gewichtung der einzelnen Disziplinen bekannt, werden die Rangplätze intransparent und ohne rationale Begründung vergeben. Vor diesem Hintergrund ist ein Training für den Sportler nicht sinnvoll, da die Rangordnung einer gewissen Beliebigkeit entspricht und durch den Athleten nicht in ausreichendem Maße beeinflusst werden kann. Weiterhin kann man davon ausgehen, dass die Beliebigkeit des Gesamtergebnisses nicht vom Publikum akzeptiert wird.

**Bewertung im Gesundheitswesen:** Bei der Bewertung von Gesundheitstechnologien werden häufig mehrere Zielkriterien formuliert (z. B. Mortalität, Morbidität, Lebensqualität), durch Nutzendimensionen abgebildet (z.B. Verlängerung der Lebensdauer, Verbesserung des Gesundheitszustandes, Verkürzung der Krankheitsdauer, Verringerung der Nebenwirkung, Verbesserung der Lebensqualität) und durch

patientenrelevante Endpunkte gemessen. Weder bei der Identifikation der Zielkriterien, noch bei der Gewichtung der gemessenen Zielerreichung patientenrelevanter Endpunkte berücksichtigt das IQWiG Evidenz. Die „Schiedsrichter“ verfügen weder über wissenschaftliche Studien hinsichtlich der Prioritäten der Patientenpopulation(en), noch machen Sie Ihre eigene Einschätzung der Relevanz der Bewertungskriterien transparent. Fraglich ist, wie ein „fairen Wettkampf“ in diesem Kontext zustande kommen kann.

**Skalenniveau:** Das Ergebnis der Bewertung kann in unterschiedlichen Skalen präsentiert werden. Die Anforderung an das Skalenniveau ist abhängig von der Art der Entscheidung. Bei einem sportlichen Wettkampf muss eine Aussage getroffen werden, ob ein Unterschied zwischen Athlet A und Athlet B vorliegt (nicht in welcher Höhe). Das heißt eine Verortung des Bewertungsergebnisses auf einer Ordinalskala ist ausreichend. (Dem Wettkämpfer wird auf einer Ordinalskala ein Rangplatz zugewiesen: Wir kennen den Gewinner, wissen aber nicht, um wie viel besser der Gewinner ist). Soll aber ein Preisgeld entsprechend der sportlichen Leistung verliehen werden, d.h. entsprechend der Höhe des Erfolges von A über B, dann muss das gemessene und bewertete Ergebnis auf einer Kardinalskala (Intervallskala) abgetragen werden können (d.h. der Bewertungsprozess muss im Ergebnis den Abstand zwischen A und B abbilden können). Der unterschiedliche Informationsbedarf bei der Entscheidung stellt unterschiedliche Anforderungen an die Methoden und Instrumente der (Nutzen-)Bewertung, ändert aber nichts an der Forderung einer sachgerechten Auswahl der Disziplinen, der Vergleichbarkeit der Disziplinen und der Gewichtung der Wettkämpfe hinsichtlich Ihrer Bedeutung.

**Fazit:** Die Bewertung der Zielkriterien und der Vergleich der gemessenen Effekte ist ein wesentlicher Bestandteil der Nutzenbewertung, egal ob bei der Operationalisierung der gemessenen Effekte, in der frühen Nutzenbewertung oder der Kosten-Nutzen Bewertung. Eine faire Entscheidung bei multiplen Zielkriterien ohne eine explizite Gewichtung ist unmöglich. Die Aggregation eines Maßes für den Gesamtnutzen oder die Bestimmung des Ausmaßes des Zusatznutzens beim Vergleich von Therapie A gegen B ist ohne ein Werturteil im Hinblick auf die Vergleichbarkeit und Gewichtung der Zielkriterien ebenso nicht möglich. Dieses Werturteil kann subjektiv oder objektiv getroffen werden. Werturteile können transparent oder intransparent im Entscheidungsprozess verwendet werden. Jeder Sportler trainiert im Vertrauen auf einen fairen Wettkampf. Ohne eine transparente und objektive Transformation der gemessenen Ergebnisse in eine Entscheidung über Erfolg und Misserfolg sind die Entbehrungen im Training irrational. Es könnte argumentiert werden, dass die Beurteilung der Messung Bestandteil der Aufgaben des IQWiG ist, wobei das Treffen von Wertentscheidungen in den Aufgabenbereich der Selbstverwaltung fällt bzw. beim GBA liegt. Konkret: Das IQWiG bewertet die Messung, also die

Begründung kausaler Effekte, und der GBA trifft auf Basis des Berichtes bei der Feststellung des Ausmaßes des Zusatznutzens die notwendigen Wertentscheidungen. Diese Argumentation greift zu kurz, da eine rationale Entscheidung durch den GBA die Informationen über die Wertmaßstäbe der Bevölkerung und Präferenzen der Patienten benötigt. Diese Informationen sollten demnach Gegenstand des Dossiers und Gegenstand des Berichtes des IQWiG an den GBA sein.

## Stellungnahme Kapitel 4: Nutzen

### Maß des Gesamtnutzens als relevante Entscheidungsgröße

In Abschnitt 4.3.2 (Endpunkte) definiert das Methodenpapier das „Maß des Gesamtnutzens“, als „eine Aggregation der Bewertung von Nutzen und Schaden in einer Größe, wobei unterschiedliche patientenrelevante Endpunkte zu einem einzigen Maß zusammengefasst werden.“ (S. 87).

*Position und Fragen: Unklar bleibt, ob die Aggregation eine notwendige Bedingung für die Entscheidungsfindung darstellt? Oder, ob eine Kosten-Nutzenbewertung auch sinnvoll ist, wenn mehrere Endpunkte „nebeneinander dargestellt werden“ (S. 87) und für jeden „patientenrelevanten Endpunkt eine eigene Effizienzgrenze erstellt“ (S. 87) wird?*

### Alternative Vorgehensweise für die Bestimmung des Gesamtnutzens

Abschnitt 4.3.3 (Maß des Gesamtnutzens) nennt das qualitätsadjustierte Lebensjahr (QALY), das behinderungsbereinigte Lebensjahr (DALY) oder das Healthy Years Equivalent (HYE) als unterschiedliche Maße, um den Gesamtnutzen einer Intervention auszudrücken.

*Position und Fragen: Keinen Sinn macht die Feststellung, dass „je nach methodischem Ansatz oder ökonomischer Theorie in der wissenschaftlichen Literatur die Begriffe Präferenzen (Preferences), Nutzwerte bzw. Teilnutzenwerte (Utilities) oder Werte (Values) verwendet“ (S. 87) werden (Auch fraglich ist ein Verweis an dieser Stelle auf das Buch Drummond et al. 2005 mit annähernd 400 Seiten). Diese Begrifflichkeiten werden weder als Synonyme verwendet, noch haben sie in unterschiedlichen methodischen Ansätzen eine ähnliche Bedeutung. (Dies wird auf den annähernd 400 Seiten von Drummond et al. 2005 auch ausgeführt). Auf eine Diskussion der methodischen Grundlagen und Begrifflichkeiten wird im Methodenpapier gänzlich verzichtet. Diese Diskussion wäre hilfreich, da diese Begrifflichkeiten unterschiedliche Theorieansätze betreffen (z. B. bei der Definition des Nutzenbegriffes) und in unterschiedlichen Kontexten verwendet werden (so bedarf es der Gewichtung von klinischen oder nicht-klinischen Endpunkten durch Präferenzgewichte, damit ein Nutzwert aggregiert werden kann). Insofern sind in diesem Absatz die Ausführungen irreführend, da nicht alle der oben genannten Begrifflichkeiten unter dem Begriff „Gewicht“ abgehandelt werden können (siehe Aussage auf S. 87).*

*Position und Fragen: Es wäre sinnvoll zu erfahren, wie diese Gewichte "mit deren Hilfe einzelne patientenrelevante Endpunkte in ein Maß des Gesamtnutzens überführt werden können" (S. 87), wissenschaftlich hergeleitet werden? Es ist sicherlich falsch zu behaupten, dass der QALY ein derartiges Gewicht darstellen kann. Vielmehr werden wahlbasierte Verfahren (z. B. Standard Gamble, Time Trade off, Discrete Choice Experimente) eingesetzt, um Gewichte zu ermitteln, damit erhobene Gesundheitszustände in Nutzenwerte überführt werden können. Methoden der Präferenzmessung können sowohl bei der Gewichtung von nicht-klinischen Endpunkten (z. B. bei Gesundheitszuständen bzw. Lebensqualität) als auch klinischen Endpunkten (z. B. Zielkriterien oder Nutzendimensionen) eingesetzt werden.*

Das Methodenpapier unterscheidet zwei Ansätze zur Dokumentation eines Maßes des Gesamtnutzens: A) den „QALY als Maß des Gesamtnutzens“ und B) die „Erhebung von Präferenzen zur Erstellung eines Maßes des Gesamtnutzens“. Im Kern geht es bei dieser Unterscheidung darum, dass bei der ersten Alternative patientenberichtete subjektive Angaben zur Lebensqualität bzw. dem Gesundheitszustand (nicht-klinische Endpunkte) mit Hilfe der Präferenzmessung gewichtet werden, um dann einen Nutzenwert abzuleiten. Das zweite Maß basiert auf einem breiteren Ansatz (hier bezeichnet mit „Multikriterielle Entscheidungsfindung“) und gewichtet klinische und nicht-klinische Endpunkte auf Grundlage wissenschaftlich abgeleiteter Präferenzen. Beiden Verfahren ist also gemein, dass nur mit Hilfe der Gewichtung ein Nutzenwert als Maß des Gesamtnutzens abgeleitet werden kann. Unterschiede ergeben sich durch die subjektive Datenerhebung (patient reported outcomes in A) im Gegensatz zur objektiven Datenerhebung (clinician reported outcomes in B). Für die Gewichtung des QALYs werden grundsätzlich nur theoriebasierte Verfahren der Präferenzmessung eingesetzt. Traditionell die Lotterieverfahren wie der Standard Gamble und Time Trade Off. In jüngster Zeit werden zunehmend fortgeschrittene Stated Preference Methoden, wie das Discrete Choice Experiment bzw. die wahlbasierte Conjoint Analyse genutzt. Für die Variante B) diskutiert das IQWiG die Verfahren des Analytic Hierarchy Process sowie die (wahlbasierte) Conjoint Analyse. Das Methodenpapier verweist darauf, dass es „noch ungelöste methodische Probleme beim Einsatz dieser Verfahren“ gibt, „sodass gegenwärtig eine routinemäßige Anwendung dieser Methoden nicht vorgesehen ist“. Eine weitere Erläuterung dieser Aussage wird jedoch nicht vorgenommen. Eine wissenschaftliche Argumentation hinsichtlich dieser Aussagen bzw. weiterführende Zitate sind nicht zu finden. In diesem Kontext werden weitere Behauptungen aufgestellt, die weder durch Literatur untermauert sondern lediglich argumentativ gestützt werden.

*Position und Fragen: Geht das IQWiG davon aus, dass beide genannten Ansätze zur Operationalisierung des Gesamtnutzens systematisch und konsistent zu gleichen Ergebnissen kommen? Auf welchem Weg können die Ergebnisse beider Verfahren vergleichbar gemacht werden? Vor dem Hintergrund, dass hier zum einen subjektive Daten (patient reported outcomes) als auch objektive Daten (clinician reported outcomes) bei der Transformation der Effekte in einen Nutzenwert zugrunde gelegt werden, stellt sich die Frage nach der Ausschließlichkeit: Spielen bei der Ableitung des Maßes für den Gesamtnutzen über QALYs die klinisch gemessenen Endpunkte keine Rolle mehr?*

*Folgende Aussagen sind falsch: „Aus der Wahl der Szenarien wird dann über ein Regressionsmodell ein Gewichtungsfaktor für jedes Attribut errechnet“ (S. 89). Das Regressionsmodell schätzt lediglich die Koeffizienten für die Ausprägungen (Level) der Attribute. „Für die CA gilt, dass maximal 6 bis 7 Attribute einbezogen werden können.“ (S. 89) Die einschlägige Literatur zeigt, dass diese Aussage falsch ist, es können und werden mehr Attribute verwendet. „Beim AHP-Verfahren gibt es diese Grenze nicht“ (S. 89). Auch der AHP ist im selben Ausmaß begrenzt (was ausdrücklich durch den Begründer Saaty festgehalten wurde (Saaty 1990, Seite 20), denn es können nicht unendlich viele Paarvergleichsfragen durchgeführt werden. „Weiterhin scheint das AHP-Verfahren die Befragten vor geringere kognitive Anforderungen zu stellen“ (S. 89). Diese Aussage ist falsch, da die Paarvergleiche vom Teilnehmer einer Befragung verlangen, dass er den Nutzen einer Eigenschaft auf einer 9-stufigen Likert Skala hinsichtlich ihres Zusatznutzens bewertet. Ferner müssen beim AHP stets zwei Entscheidungen getroffen werden: Zum einen, welches Kriterium hat die größere Bedeutung und zum anderen, um wie viel größer die Bedeutung des jeweiligen Attributes ist. Dagegen verlangt die wahlbasierte Conjoint Analyse nur die Entscheidung zwischen zwei Alternativen. Es ist in der einschlägigen psychologischen und ökonomischen Literatur bekannt, dass eine numerische Bewertung des Nutzens auf einer Kardinalskala für den einzelnen Teilnehmer mit sehr großen Problemen verbunden ist.*

### Skalenniveau des transformierten Nutzens

In Abschnitt 4.3.1 wird die Skala definiert, auf der ein transformierter Nutzen abgetragen werden soll. Über den „Bereich, der für die Definition des patientenrelevanten Zusatznutzens von Bedeutung ist“ (S. 86) soll der Nutzenwert auf einer kardinalen Nutzenskala abgetragen werden. Im Methodenbericht wird diese Skala mit „approximativ kardinalskaliert“ bezeichnet.

*Position und Fragen: Diese Bezeichnung ist irreführend, da im relevanten Bereich eine Kardinalskala eingefordert wird. Nachdem die Kapitel 3 und 4 Eingangs in ihren methodischen Grundlagen verknüpft*

*wurden, gelten diese Anforderungen hinsichtlich des Skalenniveaus des Gesamtnutzens nur für die Kosten-Nutzen Bewertung oder auch (entsprechend der Verweise aus den Abschnitten in Kapitel 3) für die frühe Nutzenbewertung?*

**Fazit Kapitel 4:** Die einführende Argumentation dieser Stellungnahme zeigt die Notwendigkeit der Berücksichtigung von Werturteilen und Präferenzen. Folgt man diesen Ausführungen, gilt es in diesem Abschnitt und in den Kapiteln 3 und 2 die Trennung zwischen „Messen“ und „Bewerten“ herauszustellen und deutlich die Anforderungen an die Methoden und Instrumente der Gewichtung von klinischen und nicht-klinischen Zielkriterien zu formulieren. Gleich zu Beginn des Abschnitts 4.3 wird festgestellt, dass die „Methoden, mit denen der Nutzen im Rahmen der Nutzenbewertung bestimmt wird, (...) in Abschnitt 3.3 beschrieben“ (S. 86) wurden. Diese Aussage ist unzureichend, da im Abschnitt 3.3 keine Aussagen zur Bewertung des Nutzens, sondern nur Aussagen hinsichtlich der richtigen Messung der Effekte gemacht werden (siehe die Hinweise im folgenden Abschnitt). Auf Basis dieses Verweises wird das Kapitel 3 auch umfassend in die Stellungnahme einbezogen.

## Stellungnahme Kapitel 3: Nutzenbewertung medizinischer Interventionen

### Evidenzgrad über die Werturteile in der frühen Nutzenbewertung

In Abschnitt 3.1.1 deutet das Methodenpapier rudimentär an, dass realisierte positive und negative Effekte sich gegenseitig kompensieren (S. 38 und 39). Konkreter, dass eine Verlängerung der Lebensdauer durch eine stärkere Nebenwirkung kompensiert werden kann. Zudem können „Nutzen- und Schadensaspekte (...) eine unterschiedliche Wichtigkeit für die Betroffenen haben“ (S. 40). In diesem Zusammenhang weißt das Methodenpapier lapidar auf die Quelle der zugrundeliegenden Information hin, „die sich ggf. durch die qualitative Erhebung oder bereits bei der Beratung durch Betroffene, Patientenvertretungs- und / oder Verbraucherorganisationen (...) abzeichnet“ (S. 40).

*Position und Fragen: Welcher Evidenzgrad wird bei der Berücksichtigung dieser Werturteile toleriert? Reicht in diesem Kontext die anekdotische Evidenz der qualitativen Befragung von nicht weiter spezifizierten Personengruppen? In welcher Form wird diese Evidenz über die Werte und Prioritäten der Patienten dokumentiert? In welcher Form wird das durch qualitative Erhebungen gewonnene Wissen den Entscheidungsträgern zur Verfügung gestellt? In welcher Form werden diese Entscheidungsgrundlagen im Bericht an den GBA aufbereitet? Mit welcher wissenschaftlichen Begründung ist die qualitative Erhebung ausreichend – würden diese Anforderungen auch für den pharmazeutischen Unternehmer im Rahmen der Berichterstattung des Dossiers gelten?*

### Patientenzufriedenheit und Nutzendimensionen

Vorrangig zu berücksichtigende Zielgrößen sind nach § 35 Abs. 1b SGB V die Mortalität, die Morbidität und die Lebensqualität, da sie eine Feststellung krankheits- und behandlungsbedingter Veränderungen erlauben (S. 39). Die „Zufriedenheit der Patientinnen und Patienten“ wurde gestrichen.

*Position und Fragen: Welche wissenschaftlichen Erkenntnisse haben dazu geführt, die Patientenzufriedenheit nach fünf Jahren zu streichen und zukünftig nicht mehr als ergänzende Zielgröße anzuerkennen?*

Wie die Patientenzufriedenheit, werden ohne Kommentar, die in der Gesetzgebung konkretisierenden Nutzendimensionen des § 35b SGB V (Verbesserung des Gesundheitszustandes, Verkürzung der

Krankheitsdauer, Verlängerung der Lebensdauer, Verringerung der Nebenwirkung, Verbesserung der Lebensqualität) gestrichen.

*Position und Fragen: Welche wissenschaftlichen Erkenntnisse haben dazu geführt, diese konkretisierenden Nutzendimensionen im Kapitel 3 zu löschen? Werden diese Nutzendimensionen nicht mehr anerkannt?*

## Begriff und Definition des Nutzen

Grundsätzlich bleibt unklar, welche Zielkriterien aus den Zielgrößen oder Nutzendimensionen abgeleitet werden können. Die Ableitung von Zielkriterien bzw. patientenrelevanten Endpunkten ist abhängig von der dem Methodenpapier zugrundeliegenden Definition des Nutzens. Im Methodenbericht wird der Begriff „Nutzen“ als „kausal begründete positive Effekte“ einer medizinischen Intervention auf patientenrelevante Endpunkte definiert (S. 38). Hier scheint eine Definition gewählt, die sich ausschließlich an den klinischen Effekten orientiert, also ein „medizinischer Nutzenbegriff“ (soweit es diesen als eigenständige Definition geben kann). Vermutlich umfasst diese Definition nicht den ökonomischen Nutzen, der beim Patienten entsteht oder das Konzept des Erwartungsnutzens. Im Internet Glossar definiert das IQWiG einen „engen“ und einen „weiten“ Nutzenbegriff. „In seiner engen Bedeutung lehnt er sich an die EbM an und spiegelt den reinen medizinischen Nutzen zur Beurteilung einer Maßnahme wider (= Gesundheitseffekte / Outcomes)“ (IQWiG Glossar im Internet). Auf diesen Nutzenbegriff wird zu Beginn des Kapitels 3 zurückgegriffen. Erst in seiner weiten Bedeutung bezieht sich der Nutzenbegriff auf den beim Patienten entstandenen Nutzen oder den Erwartungsnutzen (unter Unsicherheit). Der Nutzen wird dann nicht nur als ein Gesundheitseffekt einer Intervention definiert, „sondern berücksichtigt auch den Wert, den der Patient diesem Effekt zuschreibt (sogenannte Nutzwerte / Utilities)“ (IQWiG Glossar im Internet). Dieser Begriff lehnt sich scheinbar an die neoklassische Theorie an, wonach der Nutzen nicht mehr über die Befriedigung von Bedürfnissen definiert ist, sondern sich durch die Wahlentscheidung zwischen zwei Handlungsalternativen definiert, d.h. über den Wert den der Konsument einer Intervention, Handlung oder Alternative beimisst. Wird eine Behandlung A einer Behandlung B vorgezogen, so hat A aus der Perspektive des Entscheiders (z.B. Arzt oder Patient) einen höheren Nutzen. Die Präferenz für A resultiert aus einem höheren Nutzen von A im Vergleich mit B (aus der Perspektive des Entscheiders). Der durch den Experten oder Konsumenten prognostizierte Nutzen muss nicht immer dem tatsächlich eintretenden Nutzen entsprechen. Man spricht vom Erwartungsnutzen, wenn zum Zeitpunkt der Entscheidung aufgrund von Unsicherheit der Nutzen nicht genau quantifiziert werden kann.

*Position und Fragen: Es wäre sinnvoll zu prüfen, ob die dem Methodenpapier zugrundeliegende Definition des Nutzens neben der EbM auch den gesundheitsökonomischen Inhalten gerecht werden könnte. Der „reine medizinische Nutzen“ scheint hinsichtlich seines Wertes für den Patienten unbestimmt. Die Verbesserung des Patientennutzens ist jedoch eine relevante Handlungsmaxime. Der „enge“ Nutzenbegriff orientiert sich an den gemessenen klinischen Effekten und nicht an dem Wert, den ein Entscheider oder Konsument der Intervention zuweist (insofern ist dieser Nutzenbegriff nicht in der Ökonomie verankert). Man kann nicht davon ausgehen, dass dieser medizinische Nutzenbegriff mit dem Patientennutzen gleichzusetzen ist (gemessene patientenrelevante Endpunkte spiegeln genauso wenig den Patientennutzen wieder, wie Motorleistung und Benzinverbrauch den Nutzen eines Autos charakterisiert. Es kommt auf die Bewertung der Effekte aus der Perspektive des Konsumenten an).*

Zudem setzt das Glossar im Internet den „weiten“ Nutzenbegriff gleich mit dem Nutzwert. Der neoklassische und damit vorherrschende Nutzenbegriff leitet sich aus der Wahlhandlung bzw. über die Präferenzen ab. Der Nutzwert orientiert sich am klassischen Begriff, der durch das Ausmaß der Eignung eines Gutes zur Befriedigung der Bedürfnisse bestimmt wird.

*Position und Fragen: Offen bleibt dann, welcher ökonomische Nutzenbegriff handlungsleitend für die weiteren Ausführungen des Methodenpapiers ist. Weiterhin führt das Institut im Internet Glossar aus, dass diese „Nutzwerte (...) zur Gewichtung von Effekten genutzt werden und hierdurch die Bedeutung eines Effektes aus Sicht der Betroffenen erhöhen oder reduzieren“ können. Diese Feststellung ist falsch, da in die Berechnung der Nutzwerte bereits die gewichteten Effekte der Intervention eingegangen sind.*

*Haben die Ausführungen über den Nutzen oberflächlich gesehen wenig mit den geänderten Textzeilen im Entwurf des Methodenpapiers 4.2 zu tun, zeigt sich im folgenden, dass Missverständnisse aus der unterschiedlichen Definition des Nutzens gerade in den Kapiteln 3 und 4 nicht aufgelöst werden können. Zudem stellt sich für den Leser die Frage, ob beide Kapitel mit dem gleichen Nutzenbegriff arbeiten oder ob Kapitel 3 (zur FNB) und Kapitel 4 (zur KNB) einen unterschiedlichen Nutzenbegriff zu Grunde legen. Nach der Lektüre könnte man zu dem Ergebnis kommen, dass bei der frühen Nutzenbewertung der „enge“ Nutzenbegriff handlungsleitend ist, wobei bei der Kosten-Nutzen Bewertung der „weite“ Nutzenbegriff umgesetzt werden soll. Eine Begründung für die unterschiedliche Betrachtung ist in den Ausführungen nicht zu finden.*

## Fakultative Würdigung von Schaden und Nutzen

Der Begriff der „Nutzenbewertung“ wird definiert als der gesamte „Prozess der Evaluation medizinischer Interventionen hinsichtlich ihrer kausal begründeten positiven und negativen Effekte im Vergleich mit einer klar definierten anderen Therapie, einem Placebo (oder einer andersartigen Scheinbehandlung) oder keiner Behandlung. Dabei werden Nutzen- und Schadensaspekte zunächst endpunktbezogen evaluiert und dargestellt. Darüber hinaus ist eine gemeinsame Würdigung der endpunktbezogenen Nutzen- und Schadensaspekte möglich (...)“ (S. 40)

*Position und Fragen: Was heißt in diesem Kontext „gemeinsame Würdigung“? Warum ist eine gemeinsame Würdigung von Nutzen- und Schadensaspekten „möglich“ – ist eine Entscheidung ohne die Abwägung von Nutzen und Schaden für das IQWiG denkbar?*

## Würdigung der Nutzen- und Schadensaspekte

„Schaden“ wird definiert als kausal begründete, negative Effekte einer medizinischen Intervention auf patientenrelevante Endpunkte. Abschnitt 3.1.3 begründet die Relevanz der unerwünschten Wirkungen, unter anderem durch die Tatsache, dass der Nutzen teilweise oder ganz durch den Schaden aufgewogen werden könnte. Von einer Relevanz der negativen Effekte ist dann auszugehen, wenn unter anderem die Effekte „von den Patientinnen und Patienten als besonders wichtig angesehen werden“ (S. 43). Die Auswahl der potenziell relevanten unerwünschten Wirkungen erfolgt unter anderem auf Basis dieses Kriteriums (S. 44).

*Position und Fragen: Es bleibt offen, welche wissenschaftlichen Methoden und Instrumente zur Dokumentation der Wichtigkeit einer unerwünschten Wirkung herangezogen werden können? Bei einem unklaren Nutzenbegriff bleibt es offen, welche Nutzendimension und Zielkriterien bei der Abwägung von Nutzen und Schaden zur Bestimmung des Zusatznutzens gegenüber einer Alternative eingehen. Die gemeinsame Würdigung von Nutzen und Schaden verlangt einen transparenten Prozess, der die Wertentscheidungen, d.h. der die Gewichtung der positiven und negativen Effekte, offen legt. Wertentscheidungen der Abwägung sollten nicht Gegenstand „subjektiv geprägte(r) und damit häufig verzerrte(r) Bewertung(en)“ sein (dies ist auf S.4 in Abschnitt 1.2 als wesentliche Forderung der EbM vorgegeben). Die Gewichtung der Entscheidungskriterien als Grundlage des Abwägungsprozesses, ist die Voraussetzung für die gemeinsame Würdigung der endpunktbezogenen Nutzen- und Schadensaspekte.*

Bei der Definition des Begriffes der Nutzenbewertung wird für die Details der „gemeinsamen Würdigung der endpunktbezogenen Nutzen- und Schadensaspekte“ auf Abschnitt 3.1.4 verwiesen (S. 40). In Abschnitt 3.1.4 wird die Abwägung zwischen Nutzen und Schaden anhand der Stärke der Ergebnisunsicherheit vorgenommen. Dabei wird der (Zusatz-)Nutzen und Schaden zusammenfassend anhand der Beleglage in vier Abstufungen bezüglich der Aussagesicherheit getroffen (Beleg, ein Hinweis, ein Anhaltspunkt oder keine der drei Situationen). Anhand der Gewichtung der Ergebnissicherheit sollen die gemessenen Effekte zusammenfassend beschrieben werden. Diese Operationalisierung der Ergebnissicherheit, als Gewichtungsparameter der Studienergebnisse, dient der Beurteilung der Qualität des Messens.

*Position und Fragen: Kapitel 3.1.4 geht nicht darauf ein, wie Nutzen und Schaden aufgrund von Werturteilen und Prioritäten gegeneinander abgewogen werden können. Es gibt keinen Hinweis darauf, wie eine Würdigung der endpunktbezogenen Nutzen- und Schadensaspekte hinsichtlich der Wichtigkeit der gemessenen Effekte für den Konsumenten Eingang in die Bewertung findet. Die Relevanz der einzelnen Zielkriterien aus Sicht der Patienten oder Experten wird nicht systematisch berücksichtigt. Anstatt der systematischen Berücksichtigung von Expertenurteilen und Patientenpräferenzen, erfolgt eine implizite Gewichtung durch die Beurteilung der Ergebnissicherheit des Messens. Eine transparente Bewertung der Nutzen- und Schadensaspekte anhand expliziter Informationen, welche Endpunkte von den Patienten als besonders wichtig angesehen werden, ist nicht Gegenstand der Abwägung (siehe Forderung in 3.1.3). Eine Transparenz hinsichtlich der Gewichtung der patientenrelevanten Endpunkte durch die Experten ist in Abschnitt 3.1.4 auch nicht vorgesehen.*

*Diese Bewertung deckt sich auch mit der im Anhang dargestellten „Rationale der Methodik zur Feststellung des Ausmaßes des Zusatznutzens“. Hier wird die „Hierarchisierung der Zielgrößen“ (S. 206) und die Strukturierung entsprechend der „Bedeutung“ (S. 208) auf Basis der Messung der Effektstärken vorgenommen (keine systematische und strukturierte Gewichtung der Studienergebnisse). Die Ergebnisskala differenziert zwischen „erheblich“, „beträchtlich“ und „gering“. Eine Zuweisung der Effekte der Gesamtmorbidität, der schwerwiegenden Symptome und Nebenwirkungen, der gesundheitsbezogenen Lebensqualität und der nicht schwerwiegenden Symptome und Nebenwirkungen erfolgt über unbestimmte Begriffe (S. 209 und 210). Irrenführend ist auch der Verweis auf S. 208, dass das schrittweise Vorgehen einer Gesamtabwägung, d.h. die Zusammenfassung zu einer globalen Aussage zum Ausmaß des Zusatznutzens, in Abschnitt 3.3.3 beschrieben wird. Ein Hinweis zur systematischen Aggregation auf Basis wissenschaftlicher*

*Erkenntnisse zur Gewichtung der Endpunkte hinsichtlich der Relevanz oder Wichtigkeit ist in diesem Abschnitt nicht zu finden. In Abschnitt 3.3.3 wird der Prozess der Nutzenbewertung erläutert, ohne jedoch die Methoden oder Arbeitsschritte zur Herleitung des (Zusatz-)Nutzens zu erläutern. So hoch die Anforderungen an den Evidenzgrad und die Dokumentation der Unsicherheit sind, so gering sind die Anforderungen an die Evidenz hinsichtlich der Werte und Prioritäten der Bürger. Evidenz auf Basis der Einzelmeinungen von Patientenvertretern, Gutachten von Patientenorganisation und Surveys unter den zugelassenen Organisationen nähren den Verdacht, dass man den einzelnen Patienten und deren Vertretern mehr traut als den klinischen Experten. (Oder wie ist es zu erklären, dass unterschiedliche Anforderungen hinsichtlich der medizinischen Evidenz und der Evidenz über die Patientenpräferenzen gestellt wird?; vgl. hierzu Abschnitt 1.2.7).*

### **Beliebigkeit der Abwägung von Nutzen und Schaden**

In Abschnitt 3.1.5 wird wiederum die Nutzen-Schaden Abwägung als bewertendes Fazit zur zusammenfassenden Bewertung der Beleglage für jeden patientenrelevanten Endpunkt aufgegriffen. Basierend auf den Zielgrößen (siehe Abschnitt 3.1.1) stellt das Institut den Nutzen, den Schaden und ggf. die Nutzen-Schaden-Abwägung dar „soweit dies aufgrund der vorliegenden Daten möglich ist“ (S. 49). Ohne wissenschaftlich belastbare Aussagen zu Wertmaßstäben und Präferenzen für jeden patientenrelevanten Endpunkt ist eine gleichzeitige Bewertung von Nutzen und Schaden nicht möglich. Unklar bleibt wieder, wie die Gegenüberstellung der endpunktbezogenen Nutzen- und Schadensaspekte „qualitativ oder semiquantitativ wie in der frühen Nutzenbewertung nach § 35a SGB V“ erfolgt. Die wissenschaftlichen Methoden und Verfahren sind im Methodenbericht nicht konkretisiert.

*Position und Fragen: Im Methodenbericht wird weiter ausgeführt, dass eine weitere Möglichkeit der gleichzeitigen Würdigung darin besteht, „die verschiedenen patientenrelevanten Endpunkte zu einem einzigen Maß zu aggregieren oder über ihre Gewichtung zu einer Gesamtaussage zu kommen.“ (S. 49) Diese Formulierung ist schwer vor dem Hintergrund der gängigen Methoden zu interpretieren, da die Aggregation von Effekten nur über die Gewichtung der Zielerreichungsgrade jedes Endpunktes möglich ist. Die Zusammenhänge der Gewichtung und Aggregation werden hier missverständlich dargestellt. Für die themenspezifische Konkretisierung und gleichzeitige Würdigung von Nutzen und Schaden wird weiter auf Abschnitt 4.3.3 verwiesen. Ist diesem Verweis zu entnehmen, dass die in Abschnitt 4.3.3 dargestellten Methoden auch der Kapitel 3 zugrundeliegenden Entscheidungen zur frühen Nutzenbewertung als Basis dienen? Demnach gelten für Kapitel 3 die gleichen Frage- und Problemstellungen wie für Kapitel 4.*

## Stellungnahme Kapitel 2: Produkte des Instituts

### Informationsbeschaffung und wissenschaftliche Bewertung

Nach der Übermittlung des Dossiers bezieht sich die inhaltliche Bewertung des Dossiers neben der medizinischen Expertise auch auf die Einbeziehung der Patientenperspektive. In Abschnitt 2.1.4 zeigt das Ablaufdiagramm (Abb. 4) zur Kosten-Nutzen Bewertung nach § 35b SGB V die Phase der „Informationsbeschaffung und wissenschaftlichen Bewertung“. Es kann sich in diesem Kontext nicht mehr um die Messung der klinischen Effektparameter handeln, da diese bereits in klinischen Studien gemessen und im Dossier des pharmazeutischen Unternehmers dokumentiert wurden. Die wissenschaftliche Bewertung kann sich zum einen auf die Bewertung der Qualität (Unsicherheit) der Messung beziehen, zum anderen könnte hier auch die Bewertung der Relevanz der Zielkriterien bzw. der Endpunkte (Gewichtung) für die Aggregation des Gesamtnutzens gemeint sein. Beide Aktivitäten sind anzunehmen, da das Ablaufdiagramm vorsieht, dass sowohl die medizinische Expertise über „externe Sachverständige“ als auch die „Patientenperspektive über Patienten / Patientenvertreter“ einbezogen werden sollen.

*Position und Fragen: Da beide o.g. Aktivitäten eine wesentliche methodische Voraussetzung für die Dokumentation des Patientennutzens darstellen, stellt sich hier die Frage, ob nicht beide Aktivitäten getrennt im Ablaufdiagramm gewürdigt werden sollten. Diese Maßnahme würde nicht nur die Interpretation des Ablaufdiagramms auf Seite 23 erleichtern, sondern auch die Ablaufdiagramme auf den Seiten 16 und 19. Klärungsbedarf besteht dann auch in Abgrenzung zur Abbildung 3 (Ablauf der Erstellung einer Dossierbewertung). Dort wird neben den externen Sachverständigen auch die Patientenperspektive durch Patienten und Patientenvertreter vertreten. Dabei nennt sich die Ablaufphase jetzt „Bewertung des Arzneimittels auf Basis der Dossierinhalte“. Im Text fehlt die Konkretisierung der Begrifflichkeit „wissenschaftliche Bewertung“. Das Weglassen der Begrifflichkeit „wissenschaftlich“ deutet darauf hin, dass eine Bewertung hinsichtlich der Relevanz der Zielkriterien (oder Endpunkte) nicht mehr auf wissenschaftlichen Kriterien beruhen sollte. Warum werden wissenschaftliche Informationen zur Patientenperspektive nicht explizit im Dossier ausgewiesen?*

*Entgegen der unterschiedlichen Darstellung in den Ablaufdiagrammen, beschreiben die Abschnitte 2.1.3 und 2.1.4 dann die „Einbeziehung der Patientenperspektive“ deckungsgleich. Auf Basis eines Fragebogens werden Patienten und Patientenorganisationen, welche über die in § 140f SGB V*

*benannten maßgeblichen Organisationen identifiziert wurden, zu relevanten Endpunkten und wichtigen Subgruppen befragt. Gelten hier andere Anforderungen an die Daten als im Abschnitt 1.2.7 (Nutzen im Einzelfall)? Auf Basis welcher wissenschaftlichen Methoden wird hier die Evidenz über Werturteile und Präferenzen auf Basis der Antworten von Patienten und Patientenorganisationen generiert? Grenzen sich die Änderungen in 2.1.4 gegenüber der Dossierbewertung bei der Berücksichtigung der Patientenperspektive im Ablaufschema ab? Gelten für die Abwägung des Nutzens und Schadens für die frühe Nutzenbewertung (Ausmaß des Zusatznutzens) die gleichen methodischen Voraussetzungen, wie bei der Abwägung von Nutzen und Schaden in der Kosten-Nutzen Bewertung (Maß des Gesamtnutzens)?*

### **Potenzialbewertung: Zielkriterien und Werturteile**

Neue Inhalte, wie die Methoden zur Potenzialbewertung in Abschnitt 2.1.5 (S. 25), offenbaren alte und neue Frage- und Problemstellungen. Damit Innovationen schneller in der Regelversorgung eingesetzt werden, können auf Antrag Erprobungsrichtlinien zu neuen Untersuchungs- und Behandlungsmethoden beschlossen werden (§ 137e SGB V). Neben dem Erlaubnisvorbehalt (§ 135 SGB V) in der vertragsärztlichen Versorgung und dem Verbotsvorbehalt (§ 137c SGB V) in der stationären Versorgung können so neue Untersuchungs- und Behandlungsmethoden in den Leistungskatalog der gesetzlichen Krankenversicherung aufgenommen und zeitlich befristet vergütet werden (insofern entscheidet das Potenzial über die Erstattungsfähigkeit). Der Vorteil der Potenzialbewertung besteht darin, dass ein Nutzen nicht hinreichend belegt sein muss. Ausreichend ist das Potenzial einer Behandlungsalternative, welche durch das IQWiG bewertet und durch den GBA bei der Entscheidung berücksichtigt wird.

*Position und Fragen: Das Methodenpapier macht deutlich, dass bei der Dokumentation des Potenzials nicht die gleichen Anforderungen an die Evidenz gestellt werden, wie bei der Nutzenbewertung von Arzneimitteln. Es wird darauf eingegangen, wie das Potenzial einer Methode gemessen werden kann, es werden keine Aussagen darüber getroffen, wie die gemessenen Effekte in die Bewertung des Potenzials letztendlich eingehen (wie die Werturteile über die Effektparameter in die Bewertung eingehen). Es bleibt unklar, welche Zielkriterien für die Bewertung des Potenzials herangezogen werden, ob diese Zielkriterien bei der Potenzialbewertung gewichtet werden und wie ein eindimensionales Maß für das Potenzial bestimmt werden kann.*

## Potenzialbewertung: Indikationsübergreifende Vergleichsgrößen

Mit den Beiträgen der Versicherten soll über einen zeitlich befristeten Zeitraum und mit einem abgestimmten Studiendesign Evidenz generiert werden. Die begründete Erwartung einer effektiven Behandlung wird anhand der Morbidität, Mortalität und Lebensqualität gemessen. Eine positive Beurteilung des Potenzials nach Antragstellung begründet keinen Anspruch auf Erprobung; alle eingereichten Anträge werden auf der jährlichen GBA Haushaltsberatung bewertet.

*Position und Fragen: Auf Basis eines Vergleichs der Potenziale der Methoden und der Wahrscheinlichkeit einer erfolgreichen Erprobung werden die Methoden zur Erprobung ausgewählt. Damit genügt nicht alleine eine Überschreitung eines zuvor definierten Schwellenwertes für die positive Bewertung. Die innerhalb eines Jahres eingereichten neuen Untersuchungs- und Behandlungsmethoden werden entsprechend des Potenzials verglichen. Bei der Potenzialbewertung werden indikationsübergreifende Vergleiche angestrebt. Demnach muss die Methode der Potenzialbewertung einen indikationsübergreifenden Vergleich von Untersuchungs- und Behandlungsmethoden ermöglichen. Die Methode muss im Ergebnis mindestens die Rangreihung anhand des gemessenen und bewerteten Potenzials über alle Entscheidungsalternativen garantieren. Diese Anforderung für die Methoden ist neu und in den Kapitel 3 und 4 in dieser Form nicht umgesetzt. Indikationsübergreifende Vergleiche wurden bis heute als nicht sinnvoll erachtet und im Sinne des SGB V bislang abgelehnt. Es bleibt bei der Potenzialbewertung unklar, wie diese multiplen Zielkriterien zu einer indikationsübergreifenden Kennzahl für das „Potenzial“ aggregiert werden, um dann auf Basis dieser Information eine Entscheidung über die Durchführung einer Erprobungsstudie zu treffen. Es bleibt unklar in welchem Umfang sich die Methode der Potenzialbewertung auf die methodischen Erläuterungen in Kapitel 3 und Kapitel 4 bezieht. Welche methodischen Grundlagen werden der Potenzialbewertung zugrunde gelegt?*

## Stellungnahme Kapitel 1: Das Institut

### Darstellung der wissenschaftlichen Grundlagen

Bei der Nutzenbewertung geht es darum, die notwendigen Informationen für eine evidenzbasierte Entscheidung zu generieren. In Abschnitt 1.2.1 werden aus der evidenzbasierten Medizin die Aufgaben des Instituts abgeleitet. Das Institut führt eine objektive Bewertung durch und dokumentiert, „mit welcher Sicherheit der Nutzen medizinischer Maßnahmen nachgewiesen“ (S.5) worden ist.

*Position und Fragen: Die Kommentierungen in den Kapitel 4 und 3 legen offen, dass es hierbei auch um die Beurteilung der Informationen über die folgenden Fragestellungen gehen müsste: Welche Endpunkte sind aus der Perspektive der Betroffenen und Experten relevant? Wie wurde die Relevanz der Entscheidungskriterien im Dossier belegt? Mit welcher Bedeutung gehen diese Endpunkte in einer Gesamtbetrachtung ein? Wie kann ein Gesamtnutzen oder Zusatznutzen auf Basis der Werturteile über die Zielkriterien abgeleitet werden? Darüber hinaus ist darzulegen, unter welchen Annahmen der Nutzen auf eine Dimension des Gesamtnutzens reduziert worden ist und wie die Effekte der Zielkriterien aggregiert wurden?*

*Es geht in diesem Kontext auch um gesundheitsökonomische und entscheidungstheoretische Fragestellungen, die neben der evidenzbasierten Medizin (auch fünf Jahre nach dem ersten Entwurf des Methodenpapiers) in den einführenden Darlegungen der wissenschaftlichen Grundlagen bislang keinen Platz finden. Es geht bei der Nutzenbewertung nicht nur um den kausalen Nachweis eines Effektes bzw. die Messung von Effektgrößen, sondern auch darum, unter welchen Prämissen Nutzen und Schaden abgewogen werden können, um mit dieser Information eine rationale Entscheidung zu ermöglichen.*

*Damit die Methode der Nutzenbewertung ausreichend zuverlässige Belege zur Verfügung stellt, dass „Therapie A für Patientinnen und Patienten mit einer bestimmten Krankheit besser ist als Alternative B“, müssen neben den Methoden der EbM auch die Methoden der Gesundheitsökonomie erläutert werden. Sowohl der Gesetzgeber, als auch das Methodenpapier bezeichnet neben der evidenzbasierten Medizin, die Gesundheitsökonomie als wissenschaftliche Basis der Nutzenbewertung. In Kapitel 1 des Methodenpapiers werden jedoch nur die Grundlagen der evidenzbasierten Medizin herangezogen um die methodischen Grundlagen des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) näher zu erläutern.*

*Würde man die wissenschaftlichen Grundlagen der Gesundheitsökonomie berücksichtigen, würde dies eine wesentliche Strategie der EbM unterstreichen. Ein grundlegendes Prinzip der EbM ist die Berücksichtigung der Präferenzen der betroffenen Patienten. Die Notwendigkeit der Abwägung von Nutzen und Schaden auf Basis der Patientenpräferenzen ist bereits in den Ausführungen zur evidenzbasierten Medizin explizit genannt (siehe hier die Erläuterungen in Abschnitt 1.2.1). Der Patient gilt im Sinne der evidenzbasierten Medizin als Betroffener, dessen Präferenzen strukturiert und systematisch zu berücksichtigen sind. Leider werden die Präferenzen als theoretische Schnittstelle der evidenzbasierten Medizin und der Gesundheitsökonomie in Kapitel 1 nicht diskutiert. In Abschnitt 1.2.4 (die Strategien der EbM) müssten dann weitere charakteristische Standardelemente aufgenommen werden. Die Punkte (1) bis (4) befassen sich nur mit dem Messen, jedoch nicht mit dem Bewerten der klinischen Evidenz. Wenn die EbM letztendlich darauf verweist, dass die Präferenzen der Betroffenen im Kontext der Behandlungsentscheidungen eine Rolle spielen müssen, dann könnte die Bewertung der Zielkriterien auch im Abschnitt 1.2.4 in einem gesonderten Punkt als zentrale Strategie aufgegriffen werden.*

*Die charakteristischen Standardelemente in Abschnitt 1.2.3 sind somit auch um die gesundheitsökonomischen Problem- und Fragestellungen zu erweitern. So müsste die Vergleichbarkeit von unterschiedlichen Messergebnissen, die Gewichtung von unterschiedlich gemessenen Zielerreichungsgraden und letztendlich die Definition einer Aggregationsregel zur Analyse eines Nutzenmaßes als eindimensionale Kennzahl auf einer ordinalen oder kardinalen Skala, diskutiert werden.*

## **Die Bereitstellung von Informationen über Prioritäten und Werturteile**

In Abschnitt 1.2.5 stellt das Methodenpapier die Bedeutung der Ergebnissicherheit für die Nutzenbewertung heraus. Im Kontext der Anforderungen an die Präzision und Verlässlichkeit kommt dem GBA die Aufgabe zu, die Unsicherheit bei der Entscheidungsfindung zu beachten. Hier konkretisiert das Methodenpapier: „Neben der wissenschaftlichen Sachlage bezieht er auch andere Aspekte in seine Entscheidung ein, wozu neben der Wirtschaftlichkeit auch die Bedürfnisse und Werte der Menschen gehören. In einer wissenschaftlich unsicheren Situation gewinnen diese Aspekte an Gewicht“ (S. 8).

*Position und Fragen: In diesem Kontext müsste klar herausgestellt werden, inwiefern diese Informationen über die Bedürfnisse und Werturteile durch den Antragsteller bereitgestellt werden bzw. wie diese Informationen für die Entscheidungsträger bewertet werden. Werturteile bzw. die Bedürfnisse der Menschen dienen der Bewertung der klinischen Evidenz. Sie sollten systematisch für die Bestimmung des Ausmaßes des Zusatznutzens und des Maßes für den Gesamtnutzen im Dossier bzw. der Kosten-Nutzen Bewertung bereitgestellt werden. Die Notwendigkeit derartiger Studien für die Dokumentation des Ausmaßes des Zusatznutzens und des Maßes des Gesamtnutzens, wird in den folgenden Kapiteln nur rudimentär berücksichtigt. Nach der Verschiebung der wesentlichen Textstellen, könnte aus den Ausführungen geschlossen werden, dass unterschiedliche Anforderungen an die frühe Nutzenbewertung gestellt werden sollen. Dieser verkürzten Berücksichtigung der Bedürfnisse der Menschen sollte bereits in Kapitel 1 klar entgegengetreten werden.*

### Anforderungen an die Evidenz der Werturteile und Präferenzen

Im Abschnitt 1.2.7 (Nutzen im Einzelfall) stellt das Methodenpapier fest, dass auf „Einzelfällen basierende Erfahrungen (...) für eine Nutzenbewertung ungeeignet“ sind, „da es nicht möglich ist, Ergebnisse im Einzelfall (d.h. ohne Vergleich) einer Intervention zuzuschreiben.“ Treffender Weise hält der Methodenbericht fest, dass Erfahrungen, welche auf Einzelfällen basieren nicht für die Nutzenbewertung geeignet sind.

*Position und Fragen: Leider versäumen die Autoren in diesem Kontext darauf hinzuweisen, dass im Status Quo die Prioritäten und Werte der Patienten auf Einzelmeinungen und Einzelfällen beruhen (Vgl. hiermit auch die Anforderungen an die Dokumentation der Wichtigkeit in 3.1.4, weiter unten im Text). Das Methodenpapier sollte zukünftig darauf verweisen, dass es wissenschaftlicher Studien bedarf, um die Präferenzen und Werturteile systematisch wissenschaftlich zu erfassen und zu dokumentieren. Inwieweit wissenschaftliche Methoden zur Dokumentation der Patientenpräferenzen als Informationen über die Prioritäten und Werte bei der frühen Nutzenbewertung herangezogen werden, bleibt offen.*

## Lösungsvorschlag: Messen, Bewerten & Entscheiden

In Abschnitt 3.3.3 wird das eigentliche Selbstverständnis des Methodenpapiers deutlich: „Die Nutzenbewertung erfolgt auf Basis der im vorliegenden Methodenpapier beschriebenen Standards der evidenzbasierten Medizin, die Bewertung der Kosten auf Basis der Standards der Gesundheitsökonomie.“ (S. 59) Hier geht das Institut von einer klaren Trennung zwischen den Disziplinen aus. Diese Trennung erscheint vor dem Hintergrund der wissenschaftlichen Erkenntnisse zur Nutzenbewertung falsch. Wenn die wissenschaftlichen Instrumente und Methoden zur Erstellung und Diskussion des Studiendesigns und der Messung von klinischen und nicht-klinischen Endpunkten in der evidenzbasierten Medizin große Fortschritte gemacht haben, so ist es doch die Ökonomie die in Theorien, Methoden und Instrumenten die Bewertung des Konsumentennutzens zum Gegenstand hat. Der Nutzenbegriff ist in den Medizinwissenschaften nicht einheitlich belegt, wogegen der Nutzenbegriff in den Wirtschaftswissenschaften Gegenstand eines langanhaltenden wissenschaftlichen Diskurses ist.

Diese strikte Trennung zwischen den beiden Disziplinen resultiert in Lücken und offenen Fragen zur Operationalisierung des Nutzens, unabhängig davon, ob er bei der frühen Nutzenbewertung oder der Kosten-Nutzen Bewertung eingesetzt wird. Diese Lücken hatten sich bereits bei der Stellungnahme zum Methodenpapier 4.1 offenbart: Bereits in den vorangegangenen Stellungnahmen wird gewünscht, dass das Vorgehen bei der Ableitung einer Gesamtaussage zum Ausmaß des Zusatznutzens genauer beschrieben wird. Das IQWiG gesteht ein, dass eine „Operationalisierung für diesen Schritt der Gesamtschau (...) zurzeit nicht (...) möglich“ ist (Würdigung der Stellungnahmen des Methodenpapiers 4.1, S. 8). Die Begründung des IQWiG: „Für die Gesamtabwägung existiert keine allgemein akzeptierte Methodik.“ (Würdigung der Stellungnahmen des Methodenpapiers 4.1, S. 8) Was genau unter einer allgemein akzeptierten Methodik verstanden wird bleibt unklar. Festzustellen ist, dass kaum eine der im Methodenpapier vorgestellten Methoden „allgemein“ akzeptiert wird. Gerade die Operationalisierung zur Feststellung des Ausmaßes des Zusatznutzens wird in den Stellungnahmen als zu starr kritisiert. Die vorgeschlagene indikationsübergreifende Gewichtung widerspricht jedem ethischen Empfinden, da hier deutlich wird, dass diese Gewichtung unabhängig von der Indikation und unabhängig von den zur Bewertung herangezogenen Zielkriterien ist. Jedes Entscheidungsproblem erfährt die gleiche „Gewichtung“. Eine schwere lebensbedrohliche Krankheit wird ebenso gewichtet wie eine chronische Krankheit, sogar im gleichen Indikationsgebiet werden austherapierte Patienten (Lebensqualität steht im Vordergrund) gleichgestellt mit Patienten deren Therapie auf die Heilung zielt. Dieses Vorgehen deckt sich weder mit dem allgemeinen Gerechtigkeitsempfinden noch mit den Wertmaßstäben der Bevölkerung, noch mit den Präferenzen der Patienten. (Es wird im Internet aber durch das IQWiG

darauf hingewiesen, dass die Präferenzen bei der Nutzenbewertung eine Rolle spielen und bei der Nutzenbewertung berücksichtigt werden sollten;

[https://www.iqwig.de/de/methoden/grundsatz/hintergrund\\_was\\_sind\\_patientenpraferenzen.3757.html](https://www.iqwig.de/de/methoden/grundsatz/hintergrund_was_sind_patientenpraferenzen.3757.html)

Im Hinblick auf diese Kritik antwortet das IQWiG in der Würdigung der Stellungnahmen: „Die im Entwurf dargestellte Operationalisierung ist in der Tat abstrakt und unabhängig von der Indikation.“ Das IQWiG muss eine „Feststellung zum Ausmaß treffen“, wobei dies mit „möglichst nachvollziehbaren Verfahren erfolgen sollte“. Ebenso wird gefordert das, dass „Verfahren möglichst wenige Werteentscheidungen (etwa zur unterschiedlichen „Wertigkeit“ von Indikationen) trifft, die dem G-BA aufgrund seiner Legitimation vorbehalten bleiben sollten.“ (Würdigung der Stellungnahmen des Methodenpapiers 4.1, S. 11) Die Trennung zwischen der Informationsaufbereitung und der tatsächlichen Entscheidung, also der Trennung der Aufgaben des IQWiG und des GBA macht Sinn. Jedoch muss der GBA in seiner Funktion auch in der Lage sein, auf den durch das IQWiG berichteten Daten zu entscheiden. Keinen Sinn macht die Tatsache, dass der GBA durch den Bericht des IQWiG, keinerlei belastbare Informationen über die Werturteile der Versicherten bzw. Präferenzen der Patienten bekommt. In mehreren Antworten verweist das IQWiG darauf, dass keine konkreten „Verbesserungs- oder Alternativvorschläge (...) dem Institut (...) bekannt“ sind, oder durch die Stellungnehmenden im Stellungnahmeverfahren gemacht wurden (Würdigung der Stellungnahmen des Methodenpapiers 4.1, S. 11). Bei der Nachfrage hinsichtlich der Pilotprojekte AHP und CA antwortete Herr Andreas Gerber: „Wir haben uns bewusst dazu entschieden, darauf diesmal nicht einzugehen, weil, wie Sie alle wissen, ein Papier zu AHP ist erschienen, zur Conjoint-Analyse wird es kommen. Und wir denken, dass es Sinn macht. (...) Wir werden im Grunde genommen das dann in einer neuen Version unterbringen. (...) Die inhaltliche Diskussion sollte man vor dem Hintergrund der beiden Arbeitspapiere führen und dann, wenn es in die Methoden Einzug findet.“ (Dokumentation und Würdigung der Stellungnahmen zu den IQWiG Methoden 4.1, S.69 und 70) Die Pilotprojekte wurden bereits jeweils publiziert, eine Anpassung des Methodenberichtes an die Methoden der Präferenzmessung stehen noch aus.

## Phasen der Nutzenbewertung

Zielsetzung einer Intervention ist die positive Beeinflussung des Patientennutzens. Zielsetzung der Nutzenbewertung ist die Operationalisierung des Patientennutzens, um eine rationale Wahlentscheidung über Interventionen zu ermöglichen. Das Verfahren der Nutzenbewertung sollte eine nachvollziehbare Bewertung durch eine transparente Arbeitsweise garantieren, die dem Verfahren zugrundeliegenden wissenschaftlichen Grundlagen sollten dokumentiert werden und darüber hinaus sollte sich das Verfahren an den gesetzlichen Grundlagen orientieren.

---

Grundsätzlich kann die Nutzenbewertung in drei Phasen unterteilt werden:

- (1) Das **Messen** von kausalen Effekten einer Intervention, wobei klinische und nicht-klinische Zielgrößen beim Patienten, einer Patientenpopulation oder den Bürgern mit einem dafür geeigneten Studiendesign gemessen werden. Der Nutzenbegriff des IQWiG in der engen Fassung, bezieht sich auf die kausal begründete Wirksamkeit einer Intervention. Zielsetzung ist die Darlegung eines kausalen Zusammenhangs zwischen Intervention und positivem oder negativem Effekt. Dieser Wirksamkeitsnachweis ist Gegenstand der evidenzbasierten Medizin. Der Nachweis kausaler Effekte sollte die Prinzipien der Strukturgleichheit (Randomisierung), Behandlungsgleichheit (Verblindung) und der Beobachtungsgleichheit (Doppel-Verblindung) berücksichtigen. Gesundheit ist ein multidimensionales Konstrukt und wird in klinischen und nicht-klinischen Effektparametern gemessen. Zielgrößen sind die Mortalität, die Morbidität und die Lebensqualität. Unter Nutzendimensionen versteht man die Verbesserung des Gesundheitszustandes, Verkürzung der Krankheitsdauer, Verlängerung der Lebensdauer, Verringerung der Nebenwirkung und die Verbesserung der Lebensqualität. Endpunkte sollten zuverlässig und korrekt Änderungen des Gesundheitszustandes abbilden. Unter anderem sind die folgenden Fragestellungen relevant: Definition der Zielkriterien, Festlegung der Messgrößen, Zuordnung der Effekte im Experiment, Bewertung der Unsicherheit und Interpretation der Ergebnisse bzw. Ergebnissicherheit.
  
- (2) Die **Bewertung** der gemessenen Effekte erfolgt im nächsten Schritt. Kausal begründete Effekte können einen Nutzen oder Schaden für den Patienten darstellen. Nutzen im ökonomischen Sinne resultiert aus der Bewertung der klinischen Effekte und erfolgt in der Regel aus der Perspektive des Konsumenten der Leistung. Der Nutzen entsteht beim Konsument und wird zumeist durch die Entscheidung über Alternativen durch den Konsumenten bestimmt. Wenn aufgrund von Informationsasymmetrien ein Individuum (eine Gruppe von Individuen) keine Entscheidung zwischen Handlungsalternativen für sich selbst treffen kann, können die Entscheidungsrechte an Dritte (eine Person oder Gruppe) übertragen werden, welche eine Wahlhandlung zwischen zwei Alternativen für den Konsumenten vornehmen. Ein Nutzen wird also aus der (erwarteten) Wirksamkeit abgeleitet, entspricht aber nicht den Wirksamkeitsparametern. Nutzen ist eine bewertende Aussage des Wissens über kausale Effekte einer Intervention. Bei der Ableitung des Nutzens aus den Effektgrößen sind deshalb mehrere Wertentscheidungen notwendig. Unklar ist, in welchem funktionalen Zusammenhang

ein Effekt einen Nutzen beim Patienten begründet. Werden mehrere Endpunkte für die Nutzenbewertung herangezogen, müssen Nutzen und Schaden abgewogen werden. Präferenzgewichte sollten zuverlässig und korrekt die relative Wichtigkeit der Endpunkte abbilden. Die relative Wichtigkeit eines Zielkriteriums, bedeutet, dass sich die Relevanz eines Endpunktes immer in Relation zu der Bedeutung alternativer Zielkriterien bemisst. Das Bewerten ist Gegenstand der Ökonomie und Statistik, insbesondere der Entscheidungstheorie. Zielsetzung ist die systematische Aggregation unterschiedlicher Zielerreichungsgrade, d.h. die Zusammenfassung aller Informationen über die Effekte in einem (eindimensionalen) Nutzenmaß. Ergebnis der Bewertung ist die Bereitstellung von Informationen, d.h. indikationsspezifischen oder indikationsübergreifenden Kennzahlen, für eine Entscheidung. Der rationale Vergleich von Interventionen erfolgt anhand dieses eindimensionalen Nutzenmaßes, als notwendige Bedingung für die Ermittlung des Zusatznutzens einer Intervention über die Vergleichsintervention. Nettonutzen ist das positive Ergebnis der vergleichenden Abwägung von Nutzen und Schaden einer Intervention hinsichtlich der Zielkriterien bei der Behandlung von Patienten bzw. Patientengruppen. Zusatznutzen ist das positive Ergebnis der vergleichenden Abwägung von Interventionen hinsichtlich der bewerteten Nettonutzen für Patienten bzw. Patientengruppen. Folgende Aspekte können dabei im Mittelpunkt stehen: Relevanz der Zielkriterien, Werturteile im Hinblick auf die Gewichtung und Vergleichbarkeit der Effekte sowie Annahmen über die Aggregation multipler Zielkriterien.

- (3) Die **Entscheidung** über die Intervention(en) wird auf Basis der Informationen über den Nutzen getroffen. Eine rational begründete Entscheidung über Interventionen basiert auf vorab definierten Zielen und vorhandenen Werturteilen/Wertmaßstäben. Die Entscheidungskriterien können abhängig sein, von den Zielen der Effektivität und der Wirtschaftlichkeit einer Intervention. Gegenstand der Entscheidung können die alternativen Interventionen in einem Indikationsgebiet sein (indikationsspezifische Entscheidung). Denkbar ist auch eine Entscheidung über alle Indikationsgebiete (indikationsübergreifende Entscheidung). Der Gegenstand der Entscheidung (z. B. Erstattungsfähigkeit oder Preisfestsetzung) konkretisiert die Anforderungen an das Skalenniveau (z. B. ordinale oder kardinale Skalen). Soll die Wirtschaftlichkeit über alle Handlungsalternativen garantiert werden, dann muss notwendigerweise ein indikationsübergreifendes Maß des Gesamtnutzens den Kosten gegenübergestellt werden. Die Ökonomie und Statistik befasst sich in der Entscheidungstheorie mit der Frage nach der optimalen Entscheidung. Die rational begründete Entscheidung orientiert sich neben den kausal begründeten Effekten (Zielerreichungsgraden) auch an den

Präferenzen der Betroffenen, als Ausdruck der Werte (Wertvorstellungen, Werturteilen oder Wertmaßstäben). Zielsetzung der Entscheidung ist die rationale Festlegung auf die optimale Handlungsalternative (oft unter Unsicherheit). Folgende Aspekte können dabei interessant sein: Konkretisierung der Entscheidungslogik, Auswahl der Wirtschaftlichkeitsprinzipien (Minimumprinzip und Maximumprinzip), Unterstützung /Konfliktlösung der Verhandlungspartner und Berücksichtigung ethischer und moralischer Rahmenbedingungen.

Dieser umfassende Prozess über alle drei Phasen kann als Verfahren der Nutzenbewertung bezeichnet werden. Aufbauend auf dieser Klassifizierung kann unterschieden werden in die „Nutzenbewertung im engeren Sinne“, die sich mit der eigentlichen Übertragung der klinischen Effekte in ein Nutzenmaß befasst, und einer „Nutzenbewertung im weiteren Sinne“, welche die Nutzenbewertung durch die Messung der Effekte und die Entscheidung anhand der Nutzenmaße als notwendige Bestandteile des Verfahrens komplementiert.

In Abschnitt 1.2 wird die evidenzbasierte Medizin als die Grundlage der Methoden erörtert. Der Beitrag der EbM liegt darin, dass sich die medizinische Behandlung von Patienten nicht „auf Meinungen und Übereinkünfte stützt“ (S. 4), sondern auf Basis von „Belege(n), die mit möglichst objektiven wissenschaftlichen Methoden erhoben wurden“ (S. 4). Aus der Perspektive der Haushalte und Unternehmen ist die Umsetzung dieser Zielsetzung wünschenswert. Eine transparente Nutzenbewertung und Entscheidung über die Erstattungsfähigkeit und Preisfestsetzung, die sich konsistent am Patientennutzen und der Wirtschaftlichkeit orientiert, maximiert die Wohlfahrt für die Bevölkerung und schafft Planungssicherheit für das innovative Gesundheitsunternehmen. Es bedarf geeigneter Methoden und Instrumente, die über das gesamte Bewertungsverfahren, die Evidenz über die Effekte und die Werte berücksichtigt, um so einer „subjektiv geprägten und damit häufig verzerrten Bewertung vorbeugen“ zu können (S. 4). Dazu gehören die Messung klinischer und nicht-klinischer Effekte, die Bewertung dieser gemessenen Zielkriterien (klinische oder nicht-klinische Endpunkte) sowie die Anwendung einer spezifischen Entscheidungslogik. Die dafür notwendigen Methoden und Kriterien müssen ferner den anerkannten internationalen Standards der EbM, aber auch der Gesundheitsökonomie entsprechen.

Das vorliegende Methodenpapier konzentriert sich sehr stark auf die Frage- und Problemstellungen des Messens und damit auf die evidenzbasierte Medizin, ohne in ausreichendem Maß die gesundheitsökonomischen Fragestellungen aufzugreifen. Die Instrumente der Entscheidungstheorie werden bei der Problemlösung nicht ausreichend berücksichtigt. Diese Stellungnahme konzentriert sich

auf die Frage- und Problemstellungen des Bewertens (Phase 2) und damit im Wesentlichen auf den gesundheitsökonomischen Beitrag für die Nutzenbewertung. Die offenen Frage- und Problemstellungen des Entscheidens (Entscheidungslogik der Effizienzgrenze) werden hier nicht vertieft.

## Empfehlungen von Arbeitsschritten bei der Nutzenbewertung

In der evidenzbasierten Medizin gibt es die prominente Forderung nach der Berücksichtigung der Patientenpräferenzen, jedoch gibt es keine Anhaltspunkte wie dies geschehen soll. In der Ökonomie und Entscheidungstheorie gibt es ein reichhaltiges Spektrum an Methoden, Verfahren und Instrumenten. Folgende Fragestellungen und Arbeitsschritte sind notwendig, um ein Nutzenbewertungsverfahren an beiden Disziplinen auszurichten.

### Messen:

- (1) Welche Bewertungskriterien werden bei der Bewertung der Handlungsalternativen zugrunde gelegt? Bewerten der Relevanz von Zielkriterien und Nutzendimensionen, d.h. konkret Identifikation der wichtigsten Bewertungskriterien (patientenrelevante Endpunkte).
- (2) Wie wird das Studiendesign bewertet? Wie wird die Qualität der Daten bewertet? Wie wird die Qualität der Ergebnisse bewertet? Bewerten der Ergebnissicherheit, d.h. Einordnung des Evidenzgrad, der Repräsentativität und der Effektschätzer bzw. Konfidenzintervalle.

### Bewerten:

- (3) Welche Ober- und Untergrenzen werden für die Berechnung der Zielerreichungsgrade festgelegt? Festlegung der Annahmen über die Vergleichbarkeit, d.h. Konkretisierung der Realisationspotentials und Bewertung der Interaktionen zwischen den Bewertungskriterien (patientenrelevante Endpunkte).
- (4) Welche Wertmaßstäbe werden der Bevölkerung, den Versicherten und den Patienten unterstellt? Wer trifft die Wertentscheidungen im Bewertungsverfahren? Sind die Wertentscheidungen transparent? Welche Evidenz ist Grundlage dieser Wertentscheidungen?
- (5) Bewerten der Präferenzstruktur über das einzelne Bewertungskriterium, d.h. Gewichtung der Ausprägungen von Bewertungskriterien (patientenrelevante Endpunkte). Bewerten der Präferenzstruktur zwischen den Bewertungskriterien, d.h. Gewichtung der Bewertungskriterien (patientenrelevante Endpunkte). Wie werden die Bewertungskriterien bei der Bewertung der Handlungsalternativen gewichtet?

**Entscheiden:**

- (6) Welche Annahmen liegen der Aggregation kausaler Effekte zugrunde? Konkretisierung der Annahmen zur Operationalisierung bzw. Aggregation des Nutzens, d.h. Festlegung der Nutzenfunktion.
- (7) Welche Annahmen gelten für die Entscheidungsregel? Festlegung auf eine Entscheidungsregel, d.h. Identifikation einer Allokationslogik.

**Anforderungen an das Skalenniveau**

Die Entscheidung, basierend auf der frühen Nutzenbewertung über die Erstattungsfähigkeit bedarf einer unterschiedlichen Informationsqualität gegenüber der Entscheidung über die Preise im Kontext der Kosten-Nutzen Bewertung.

Nach § 5 Abs. 4 Satz 1 der Arzneimittel-Nutzenbewertungsverordnung (AM-NutzenV) ist im Dossier und damit bei der Nutzenbewertung festzulegen, „in welchem Ausmaß ein Zusatznutzen vorliegt“. Die 6 Kategorien sind in § 5 Abs. 7 der AM-NutzenV vorgegeben: (1) erheblicher Zusatznutzen, (2) beträchtlicher Zusatznutzen, (3) geringer Zusatznutzen, (4) nicht quantifizierbarer Zusatznutzen, (5) kein Zusatznutzen belegt und (6) geringerer Zusatznutzen. Bei der Operationalisierung des Zusatznutzens geht das IQWiG davon aus, dass in der Arzneimittel-Nutzenbewertungsverordnung eine „Hierarchisierung dieser Zielgrößen intendiert“ ist (S. 206). Zudem kann davon ausgegangen werden, dass für die Bestimmung der 6 oben genannten Kategorien, der Zusatznutzen in einem eindimensionalen Nutzenmaß quantifiziert werden muss. (Ansonsten würde die Kategorie (4) keinen Sinn ergeben).

Geht man davon aus, dass sich die Teilnutzen unterschiedlicher patientenrelevanter Endpunkte auf ein Maß bzw. eine Dimension reduzieren lassen, dann muss festgelegt werden, welches Skalenniveau für die Entscheidung erforderlich ist. Man kann zwischen kardinalem und ordinalem Skalenniveau unterscheiden. Die kardinale Nutzenmessung misst der Nutzendifferenz von Intervention A und Intervention B eine Bedeutung zu. Die Nutzendifferenz zwischen zwei Interventionen hat in der ordinalen Nutzentheorie keine wesentliche Aussagekraft.

Bei der Entscheidung über die Erstattungsfähigkeit ist die Information über die Rangreihung der Interventionen ausreichend. (Eine Intervention A hat einen höheren Nutzen als eine Intervention B; der resultierende Zusatznutzen ist Grundlage der Entscheidung über den Festbetrag und die Erstattung.) Für diese Erstattungsentscheidung ist eine ordinale Nutzenfunktion für den Zusatznutzen ausreichend,

d.h. der Nutzen der Interventionen A und B muss lediglich in eine Rangreihung gebracht werden. Das ordinale Skalenniveau sagt über den Vergleich der Interventionen A und B nur aus, welche der Interventionen einen Zusatznutzen (höheren Nutzen) hat. Die ordinale Nutzenfunktion macht keine Aussagen darüber, „wie viel höher“ der Nutzen ist. Mit Hilfe der ordinalen Nutzenfunktion können die Präferenzen analytisch dargestellt werden, ohne dass die Nutzenvorstellungen der Patienten oder Experten quantifiziert werden müssen.

Bei der Entscheidung über die Festsetzung des Preises einer Gesundheitstechnologie ist ein kardinales Skalenniveau notwendig. Die kardinale Nutzenfunktion quantifiziert den Nutzen, d.h. der Nutzen einer Intervention kann als Zahl dargestellt werden. Die Nutzenmessung ist die Voraussetzung für eine rationale Entscheidung, welche die Wirtschaftlichkeit als Rationale in die Entscheidung integriert. Zielsetzung dieser Entscheidung ist die Nutzenmaximierung, d.h. die Optimierung der Gesundheitsversorgung unter Beachtung der Budgetrestriktion.

## Zusammenfassung

Diese Stellungnahme befasst sich nicht mit den Methoden und Kriterien der Messung klinischer und nicht-klinischer Erfolgsparameter, der Begründung unterschiedlicher Studiendesigns, biometrischen Anforderungen oder der sachgerechten Umsetzung der EbM. Auch die theoretischen Grundlagen der im Methodenpapier vorgeschlagenen Entscheidungslogik sollten hier nicht vertieft werden – es wird nicht darauf eingegangen, ob die Effizienzgrenze als Entscheidungsregel geeignet ist oder nicht. Diese Stellungnahme konzentriert sich auf die Bewertung der gemessenen klinischen und nicht-klinischen Effekte. Die Stellungnahme erfolgt vor dem Hintergrund der in der Gesundheitsökonomie, noch spezifischer der Entscheidungstheorie, diskutierten Methoden und Instrumente.

Damit konzentrieren sich die Ausführungen auf die Bewertung der gemessenen Effekte als Kern der Nutzenbewertung. Es geht darum, wie die Multidimensionalität des Patientennutzens (Mortalität, Morbidität, Lebensqualität) bei der Entscheidung berücksichtigt werden kann. Zudem geht es darum, wie die patientenrelevanten Endpunkte, bei der Abwägung des Nutzen und Schadens, der Aggregation eines Maßes des Gesamtnutzens und bei der Beurteilung des Ausmaßes des Zusatznutzens in die Entscheidung eingehen.

Das zur Diskussion stehende Methodenpapier 4.2 (Version vom 18.06.2014) unterstellt, dass eine klinische Studie prospektiv die Wirkung und den Wert einer Intervention am Patienten bzw. Patientengruppen untersucht. Diese Annahme ist falsch, da die Konkretisierung des Nutzens einer Intervention weitere Werturteile in die Untersuchung einbeziehen muss.

Eine klinische Studie kann nur die Wirkung einer Intervention untersuchen bzw. Wissen über die Wirksamkeit generieren. Die Bewertung des Studiendesigns resultiert in einer Gewichtung der Ergebnisse im Hinblick auf unterschiedliche Evidenzgrade. Die Bewertung der Effektstärken resultiert ferner in einer Gewichtung der Zielkriterien im Hinblick auf die Ergebnisunsicherheit. Diese Bewertung des Studiendesigns oder der Studienergebnisse endet nicht mit dem Wissen über den Nutzen, sondern erweitert nur das Wissen über die Wirksamkeit. Eine Dokumentation des Zusatznutzens oder des Gesamtnutzens ist auf Basis der kausal begründeten Effekte, ohne die Berücksichtigung von Wertmaßstäben oder Präferenzen, nicht möglich.

Angenommen man könnte davon ausgehen, dass die Messung der patientenrelevanten Endpunkte zu 100% die Wirksamkeit einer Intervention A im Vergleich zu einer Intervention B wiedergeben. Dann kann auf Basis dieses Wissens keine Aussagen über den Nutzen, den Nettonutzen oder den Zusatznutzen gemacht werden. Insofern ist eine rationale Entscheidung nur auf Basis dieser klinischen Evidenz nicht möglich. Solange ein Nutzenmaß über die Erstattungsfähigkeit (Zusatznutzen) oder Preisfestsetzung (Gesamtnutzen in Relation zu den Kosten) informieren soll, gilt es, die gemessenen Effekte der patientenrelevanten Endpunkte in ein wissenschaftlich begründetes Nutzenmaß zu überführen. Um den Nutzen oder Wert einer Intervention zu analysieren, bedarf es mehrerer Wertentscheidungen. Ohne diese Wertentscheidungen zu treffen, ist es nicht möglich, einen Nutzenwert zu generieren oder eine systematische und transparente Entscheidung zu treffen.

Die Berücksichtigung der Wertvorstellungen der Versicherten und die Präferenzen der Patienten ist eine notwendige Bedingung für eine systematische und transparente Entscheidungsfindung. Es ist nicht ausreichend geklärt, auf welcher Basis bzw. welcher Evidenz die Abwägung des Nutzens und Schadens erfolgen soll. Dies ist sowohl aus der Perspektive der evidenzbasierten Medizin als auch aus Sicht der Gesundheitsökonomie eine wesentliche Lücke des Methodenpapiers. Offen bleibt, wer diese Wertentscheidung trifft. Offen bleibt auch, welche Evidenz als Basis dieser Wertentscheidungen herangezogen wird. Gegenwärtig werden diese Wertentscheidungen nicht transparent gemacht, ohne wissenschaftliche Evidenz nur auf Basis impliziten Wissens.

Wertentscheidungen können explizit oder implizit getroffen werden. Wertentscheidungen können auf Basis subjektiver oder objektiver Daten getroffen werden. Zunehmend wird die Forderung laut, dass implizites Wissen der Entscheidungsgremien durch explizites Wissen über diese Wertentscheidungen zu ersetzen. Der Anspruch an die Wissenschaftlichkeit des Bewertungsverfahrens macht eine eindeutige und klare Kommunikation der Gewichtung der Zielkriterien notwendig. Explizites Wissen kann systematisch dokumentiert und über eine wissenschaftliche Veröffentlichung der Bürger-, Versicherten oder Patientenpräferenzen kommuniziert und korrigiert werden.

## Literatur / Quellen

Drummond MF, Sculpher MJ, Torrance GW, O'Brian BJ, Stoddart GL. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press; 2005.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (Hrsg.) (2014): Allgemeine Methoden; Entwurf für Version 4.2 vom 18.06.2014; Köln.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (Hrsg.) (2013): Dokumentation und Würdigung der Stellungnahmen zur „Aktualisierung einiger Abschnitte der Allgemeinen Methoden Version 4.0 sowie neue Abschnitte zur Erstellung der Allgemeinen Methoden Version 4.1“ Version 1.0 vom 28.11.2013

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (Hrsg.): Internet Glossar; URL: <https://www.iqwig.de/de/glossar.2727.html> (letzter Abruft: 06.08.2014)

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (Hrsg.): „Hintergrund: Was sind Patientenpräferenzen?“ URL: [https://www.iqwig.de/de/methoden/grundsatz/hintergrund\\_was\\_sind\\_patientenpraferenzen.3757.html](https://www.iqwig.de/de/methoden/grundsatz/hintergrund_was_sind_patientenpraferenzen.3757.html); (letzter Abruft: 06.08.2014)

Saaty, T. L. (1990), How to make a decision: The analytic hierarchy process, in: European Journal of Operational Research, 48(1):9-26.