



## **Aktualisierung der Allgemeinen Methoden Version 4.0**

Gemeinsame Stellungnahme von GMDS und IBS-DR vom 15.05.2013

Autoren: Carsten Schwenke, Oliver Kuß, Uwe Siebert und Dieter Hauschke

Am 18.04.2013 wurde der Entwurf zur Aktualisierung einiger Abschnitte der Allgemeinen Methoden Version 4.0 sowie neue Abschnitte zur Erstellung der Allgemeinen Methoden Version 4.1 des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) zur Stellungnahme publiziert (IQWiG, 2013).

Die vorliegenden Entwürfe für Aktualisierungen und Ergänzungen des Methodenpapiers betreffen folgende Abschnitte:

- 2.1.1 Bericht
- 2.2.3 Review der Produkte des Instituts
- Neuer Abschnitt 3.1.4 Endpunktbezogene Bewertung
- 3.1.5 Zusammenfassende Bewertung (vorher Abschnitt 3.1.4)
- 3.3.3 Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V
- 7.3.8 Meta-Analysen
- Neuer Anhang: Rationale der Methodik zur Feststellung des Ausmaßes des Zusatznutzens

### **2.1.1 Bericht**

Kein Kommentar.

### **2.2.3 Review der Produkte des Instituts**

Kein Kommentar.

### **Neuer Abschnitt 3.1.4 Endpunktbezogene Bewertung**

Seite 8 letzter Absatz [1]:

"Ein wichtiges Kriterium zur Ableitung von Aussagen zur Beleglage ist die Ergebnissicherheit. [...] Falls die zugrunde liegenden Daten dies zulassen, lässt sich die statistische Unsicherheit als Standardfehler bzw. Konfidenzintervall von Parameterschätzungen quantifizieren und beurteilen (Präzision der Schätzung)."

Hier bietet es sich an, eingangs die beiden besprochenen Unsicherheitsarten (quantitative und qualitative Unsicherheit) auch mit den gängigen epidemiologischen Begriffen zu definieren.

Ergebnisunsicherheiten können durch zwei verschiedene Fehler auftreten:

1. Zufälliger Fehler verursacht durch eine Stichprobenziehung (statistische Unsicherheit)
2. Systematischer Fehler (Bias, systematische Verzerrung) im Wesentlichen beeinflusst durch Studiendesign bzw. korrekte Analyse. Hierzu gehören Informationsfehler, Selektionsfehler und Confounding.

### **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](mailto:info@gmds.de)  
Internet: [www.gmds.de](http://www.gmds.de)

### **Geschäftsführung**

Beatrix Behrendt

### **Präsidentin und Vize-Präsidenten**

Prof. Dr. Heike Bickeböller  
(Göttingen), Präsidentin

Prof. Dr. Paul Schmücker  
(Mannheim), 1. Vizepräsident

Prof. Dr. Johannes Haertling  
(Halle/Saale), 2. Vizepräsident

### **IBS-DR Geschäftsstelle**

Heike Krubert  
c/o Institut für Biometrie, Epidemiologie  
und Informationsverarbeitung  
Tierärztliche Hochschule  
Hannover  
Bünteweg 2  
D-30559 Hannover

Telefon: +49 (0) 511 953 79 51  
Telefax: +49 (0) 511 953 79 74

E-Mail: [biometrische-gesellschaft@tiho-hannover.de](mailto:biometrische-gesellschaft@tiho-hannover.de)

### **IBS-DR –Präsident und Vizepräsidentin**

Dr. Jürgen Kübler  
(Marburg), Präsident

Prof. Dr. Katja Ickstadt  
(Dortmund), Vizepräsidentin

"Das Institut verwendet die folgenden drei Kategorien zur Graduierung des Ausmaßes der qualitativen Ergebnissicherheit auf Einzelstudien- und Endpunktebene:

- **hohe qualitative Ergebnissicherheit:** Ergebnis einer randomisierten Studie mit niedrigem Verzerrungspotenzial.
- **mäßige qualitative Ergebnissicherheit:** Ergebnis einer randomisierten Studie mit hohem Verzerrungspotenzial.
- **geringe qualitative Ergebnissicherheit:** Ergebnis einer nicht randomisiert vergleichenden Studie. " [1]

Daraus leitet sich ab, dass Ergebnisse aus nicht-interventionellen vergleichenden Studien wie prospektive Kohortenstudien oder auch Register, in denen neben dem zu bewertenden Arzneimittel auch Vergleichspräparate einbezogen werden, zur Nutzenbewertung herangezogen werden können. Es sind also nicht zwingend ausschließlich nur randomisierte kontrollierte Studien (RCT) zur Bewertung heranziehbar, sondern auch zusätzliche Evidenz aus anderen Studien, solange sie vergleichend ausgeführt sind.

Anforderung an eine Studie zum Nachweis eines Belegs:

"Soll aus lediglich einer Studie im Ausnahmefall ein Nutzenbeleg abgeleitet werden, so sind an eine solche Studie und ihre Ergebnisse besondere Anforderungen zu stellen" mit Referenz auf die Richtlinie "European Medicines Agency. Points to consider on application with: 1. meta-analyses; 2. one pivotal study [online]. 31.05.2001". Diese Richtlinie wurde erarbeitet für nichtonkologische Indikationen, in denen regelhaft mindestens zwei RCT gefordert werden, um eine Marktzulassung für ein neues Arzneimittel zu erhalten. In onkologischen Indikationen und bei Indikationserweiterungen ist dagegen häufig nur eine pivotale Studie gefordert.

Es stellt sich nun insbesondere für die frühe Nutzenbewertung gemäß § 35a SGB V die Frage, ob in den Indikationen, in denen die Zulassung auf einer Studie basiert und sogenannte Megatrials nicht möglich sind, höchstens ein "Hinweis" erreicht werden kann. Somit wäre die Ergebnissicherheit in diesen Indikationen beschränkt auf drei Kategorien (Hinweis, Anhaltspunkt, keines von beiden). Hier wären Überlegungen wünschenswert, wie für diese Indikationen ebenfalls vier Kategorien ermöglicht werden können, insbesondere in der frühen Nutzenbewertung, zu welcher nur die Zulassungsstudien vorliegen. Zudem wäre eine Klärung der Regeln wünschenswert, wann eine einzige Studie einen Beleg ermöglicht, da sich in den bisherigen Bewertungen nur ein fünfstelliger Stichprobenumfang als hinreichend dargestellt hat. In der Bewertung von Ticagrelor schreibt das IQWiG dazu "Wegen der besonderen Güte und der ausreichenden Größe der PLATO-Studie konnten aus den Daten Belege, z. B. für einen Zusatznutzen, abgeleitet werden", allerdings ohne Beschreibung, was unter "besonderer Güte" zu verstehen ist [2]. Hier wäre ein Hinweis auf die Richtlinie zu Meta-Analysen sinnvoll, in der die entsprechenden Qualitätsmerkmale benannt sind.

In der Tabelle 2 zu den Anforderungen an die Beleglage für die unterschiedlichen Aussagesicherheiten beim Vorliegen von Studien derselben qualitativen Ergebnissicherheit wird unter Hinweis und Anhaltspunkt aufgeführt:

Hinweis	≥2	hoch	heterogen mäßig gleichgerichtet
Anhaltspunkt	≥2	mäßig	heterogen mäßig gleichgerichtet

Da zwei Studien mit gleichgerichteten Effekten immer deutlich gleichgerichtet sind (siehe Entwurf, Seite 10), sollte es lauten:

Hinweis	≥3	hoch	heterogen mäßig gleichgerichtet
Anhaltspunkt	≥3	mäßig	heterogen mäßig gleichgerichtet

### 3.1.5 Zusammenfassende Bewertung (vorher Abschnitt 3.1.4)

"Eine Möglichkeit der gleichzeitigen Würdigung von Nutzen und Schaden ist die Gegenüberstellung der endpunktbezogenen Nutzen- und Schadenaspekte. Dabei werden die Effekte auf alle Endpunkte (qualitativ oder semiquantitativ) gegeneinander abgewogen mit dem Ziel, zu einer endpunktübergreifenden Aussage zum Nutzen bzw. Zusatznutzen einer Intervention zu kommen."

Es wäre wünschenswert, eine genauere Spezifizierung zu bekommen, nach welchen Kriterien diese Nutzen- zu Schadenabwägung durchgeführt wird. Sinnvoll wäre eine medizinische Abwägung auf Basis der Intensität/Schwere des Schadens gegenüber der Größe des Nutzens. Diese Abwägung sollte allerdings nur unter Miteinbeziehung der jeweiligen Indikation und deren Parameter wie z.B. Lebensbedrohung und Schwere der Krankheitssymptome erfolgen unter Einbeziehung indikationsspezifischer Experten erfolgen. Hier wäre eine genauere Beschreibung der Methodik und des Vorgehens wünschenswert.

Der folgende Abschnitt zu den Methoden der Präferenzenintegration ist neu:

S. 13 [1]: „Häufig werden sogenannte Nutzwerte für Gesundheitszustände erhoben, die die von den Befragten positiv wie negativ empfundenen Aspekte in einer Indexzahl ausdrücken sollen. Unter Integration der Dauer der entsprechenden Gesundheitszustände können diese Nutzwerte bspw. in sogenannte qualitätsadjustierte Lebensjahre (QALYs = Quality-Adjusted Life Years) überführt werden. Aufgrund der ethischen und methodischen Probleme gerade der häufig verwendeten QALYs [122,137,138,520] sollten alternative Verfahren der multikriteriellen Entscheidungsfindung oder der Präferenzenerhebung angewendet werden. Dazu zählen u. a. der Analytic Hierarchy Process (AHP) und die Conjoint-Analyse (CA).“

Die Zusammenführung verschiedener Outcome-Dimensionen wie Lebensdauer, Morbidität und gesundheitsbezogener Lebensqualität ist ein komplexes Thema und es gibt weder ein einziges allgemeingültiges krankheitsübergreifendes Konzept, noch ist es einfach so, dass eines der genannten Instrumente durch „ethische und methodischen Probleme“ zu disqualifizieren wäre. Auch ist es nicht so, dass sich diese Instrumente gegenseitig ausschließen, sondern sie können komplementär eingesetzt werden. Die Empfehlungen hierzu sollten wissenschaftlich nüchtern gegeben werden. Dabei ist insbesondere zu berücksichtigen, dass der Entscheider auf der einen Seite unterschiedliche Dimensionen wie Lebensdauer und Lebensqualität bzw. unterschiedliche Dimensionen der Lebensqualität getrennt betrachten möchte und auf der anderen Seite diese verschiedenen Dimensionen integriert bzw. miteinander verrechnet werden müssen, wenn die Abwägung systematisch und nachvollziehbar stattfinden soll. Gerade die Getrennthaltung verschiedener Dimensionen wird dann schwierig nach-

zuvollziehen, wenn das IQWiG diese Dimensionen dann in einem nicht weiter definierten gewichteten Score wieder zusammenführen muss.

Siehe dazu im IQWiG Dokument auf S. 13 [1]: *„Eine weitere Möglichkeit der gleichzeitigen Würdigung besteht darin, die verschiedenen patientenrelevanten Endpunkte zu einem einzigen Maß zu aggregieren. In diesem Fall würden die Aussagen des Instituts für jeden einzelnen patientenrelevanten Endpunkt gewichtet z. B. in einen Summenscore einfließen.“*

Bei einer solchen Zusammenführung müssen ähnliche Annahmen gemacht werden, wie sie dem QALY zugrunde liegen, z.B. dem Abwägen zwischen Lebensdauer und Lebensqualität (s. dazu auch [5-7]). Es wäre wünschenswert, wenn das IQWiG an dieser Stelle keine inhaltliche Vorentscheidungen für Methoden trifft, sondern explizit festhält, dass die Auswahl der Integrationsmethode von der Indikation und Fragestellung abhängt und hierfür grundsätzlich alle Verfahren (QALY, AHP und CA) in Frage kommen, und kontextspezifisch geprüft werden müssen.

Insgesamt erscheint die Argumentation zum QALY-Konzept gegenüber dem ansonsten sehr seriös und wissenschaftlich gehaltenen Aktualisierungsdokument vergleichsweise unwissenschaftlich, was dem Gesamtdokument schadet. So wird auf S. 13 a) eine wissenschaftlich fehlerhafte Argumentation vorgenommen und b) werden die Argumente mit Literaturstellen belegt, die die Argumente nicht durchgängig stützen. Dies ist besonders problematisch, da beim Leser der (falsche) Eindruck entsteht, dass die zitierten Literaturstellen sinngemäß den Text im Methodenhandbuch belegen.

So handelt es sich bei den vier zitierten Literaturstellen (122, 137, 138, 520) vorwiegend um den Einsatz von Präferenzmaßen im Kontext der Gesundheitsökonomie bzw. Ressourcenallokation, nicht um den Kontext der Nutzen-Schaden-Abwägung.

Ein Beispiel: Das Dokument des Deutschen Ethikrats. Nutzen und Kosten im Gesundheitswesen: zur normativen Funktion ihrer Bewertung; Stellungnahme. Berlin: Deutscher Ethikrat; 2011. URL: <http://www.ethikrat.org/dateien/pdf/stellungnahme-nutzen-und-kosten-im-gesundheitswesen.pdf>.

Der Kontext und Anlass der Stellungnahme des Deutschen Ethikrats ist eher die Nutzen-Kosten-Bewertung des IQWiG. Dort wo es im Dokument um die reine Nutzenbewertung geht (Kapitel 3.1.3 Das Nutzenmaß QALY als Produkt aus Lebensqualität und Lebensdauer), wird das QALY-Konzept neutral mit seinen Vorzügen und methodischen Annahmen beschrieben. Auf S. 38 wird explizit auf die Plausibilität der QALY-Methode eingegangen, am Beispiel der „oftmals nebenwirkungsreichen Maßnahmen in der Tumorbehandlung wie Chemo- oder Strahlentherapie. Hier ist eine –manchmal fragliche – Verlängerung der Lebenserwartung mit einer nicht selten deutlich geminderten subjektiven ‚Erlebensqualität‘ der verbliebenen Lebenszeit verbunden.“ (S. 37)

Die Literatur zur Nutzen-Schaden-Abwägung, die sich durchaus vorsichtig positiv für das QALY-Konzept ausspricht, insbesondere, wenn das QALY-Konzept innerhalb einer Indikation angewendet wird, fehlt im IQWiG-Dokument. Siehe beispielsweise:

- Puhan MA, Singh S, Weiss CO, Varadhan R, Boyd CM. A framework for organizing and selecting quantitative approaches for benefit-harm assessment. BMC Med Res Methodol. 2012 Nov 19;12:173. doi: 10.1186/1471-2288-12-173.
- Garrison LP Jr, Towse A, Bresnahan BW: Assessing a structured, quantitative health outcomes approach to drug risk-benefit analysis. Health Aff (Millwood) 2007, 26(3):684-695.

Im IQWiG-Dokument (S. 13f) werden die Methoden Analytic Hierarchy Process und Conjoint-Analyse kurz beschrieben. Anwendungen der Methoden Analytic Hierarchy Process (AHP) und Conjoint-Analyse (CA) haben gezeigt, dass diese Konzepte nicht ohne Weiteres auf den Kontext der (Zusatz-) Nutzenbewertung anwendbar sind bzw. ebenfalls auf multiplen Annahmen beruhen, die in vielen

Entscheidungssituationen nicht erfüllt sind. Wünschenswert wäre deshalb eine kurze Beschreibung dazu, wie diese Instrumente im Kontext von Nutzenbewertungen durch das IQWiG eingesetzt werden sollen, und mit welchen Limitationen sie behaftet sind.

### **3.3.3 Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V**

In der Operationalisierung der Feststellung des Ausmaßes des Zusatznutzens werden im 1. Schritt die Kriterien zur Ableitung von Aussagen zur Beleglage nach Abschnitt 3.1.4 angewendet. Hier stellt sich die Frage, ob die Regeln der Nutzenbewertung auch für die frühe Nutzenbewertung gemäß § 35a SGB V gelten können oder ob hier andere Regeln gelten sollten, da die Voraussetzungen von Nutzenbewertung und früher Nutzenbewertung durch z.B. die unterschiedlich lange Präsenz der Produkte auf dem Markt und dementsprechend unterschiedliche Verfügbarkeit verwendbarer Evidenz als sehr unterschiedlich einzustufen sind.

Im zweiten Schritt wird das Ausmaß des Nutzens festgestellt. Hierbei ist zu berücksichtigen, dass, wie unter 3.1.4 und für Schritt 1 diskutiert, in der frühen Nutzenbewertung sehr häufig nur wenige oder sogar nur eine verwendbare Studie vorliegt und damit das zweiseitige 95% Konfidenzintervall des Behandlungseffekts auf wenigen Studien basiert, wohingegen in der Nutzenbewertung häufig viele Studien zur Verfügung stehen, so dass der Behandlungseffekt präziser geschätzt werden kann. Daraus ergibt sich wie schon für die Ergebnissicherheit die Frage, ob in der frühen Nutzenbewertung dieselben Kriterien herangezogen werden sollten, um den Ergebnissen die Kategorien des Ausmaßes des Zusatznutzens zuzuordnen.

Schritt 3 ist dagegen äquivalent für die Nutzen- wie auch frühe Nutzenbewertung verwendbar, da es sich hierbei um die Nutzen- zu Schadenabwägung handelt, die in beiden Bewertungen gleich verwendet werden können.

#### *C) Stetige oder quasi-stetige Zielgrößen mit jeweils vorliegenden Responderanalysen*

Für stetige und quasi-stetige Zielgrößen werden Responderanalysen auf Basis validierter bzw. etablierter Responsekriterien verlangt. Damit wird verlangt, dass Zielgrößen mit einem hohen Informationsgehalt dichotomisiert werden sollen, wodurch ein Großteil der Information der Daten verloren geht. In diversen Indikationen wird von der Europäischen Zulassungsbehörde ein metrischer Endpunkt als primärer Endpunkt verlangt, auf dessen Basis die Stichprobenplanung der Studie basiert (z.B. Diabetes mellitus [3] mit dem "change from baseline in HbA1c"). Zudem werden Morbiditäts- und Lebensqualitätsdaten häufig in Form eines "visual analogue scale" dargestellt, für die dann der "change from baseline" bestimmt wird, z.B. für die Bewertung von Änderungen des Symptoms Schmerz (siehe auch [4]).

In der Tat lassen sich absolute Differenzen auf Basis metrischer Endpunkte nicht über Indikationen standardisieren. Möglich wäre dies zum Beispiel mit relativen Effektgrößen wie Cohen's d. Es bleibt aber die Frage, inwieweit sich diese dann über Indikationen hinweg vergleichen lassen in Bezug auf statistischer Signifikanz verknüpft mit klinischer Relevanz und welche Grenzen dann für die Bewertung des Ausmaßes des Zusatznutzens verwendet werden sollen. Eine Dichotomisierung und der damit verbundene Verlust vorhandener Information kann aber nicht die Lösung für diese Herausforderung sein. Sicher macht es Sinn, für metrische Daten sowohl den absoluten wie auch relativen Ef-

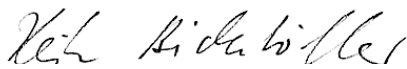
fekt zu betrachten, wenn denn letzterer möglich ist (d.h. eine adäquate Responderdefinition vorliegt, wie z.B. in der Indikation rheumatoide Arthritis. Zudem muss (indikationsspezifisch) definiert werden, wie dann das Ausmaß des Zusatznutzens bestimmt wird.

### **Abschnitt 7.3.8**

Die Änderungen im Kapitel 7.3.8 Meta-Analysen betreffen im Wesentlichen nur einen neu eingefügten Abschnitt zur Angabe von Prädiktionsintervallen im Unterabschnitt B) Heterogenität, dieser entspricht aus unserer Sicht dem gegenwärtigen Forschungsstand. Leider scheint sich der Inhalt der gemeinsamen Stellungnahme von GMDS und IBS-DR vom Februar 2012 zum Kapitel 7.3.8 noch nicht in der Aktualisierung niedergeschlagen zu haben. Wir ermutigen das IQWiG besonders, die routinemäßige Anwendung der Standardmodelle mit zufälligen und festen Effekten zu überdenken, wo doch die Nachteile dieser Methoden altbekannt sind und zahlreiche Methoden zu deren Verbesserung vorgeschlagen worden sind.

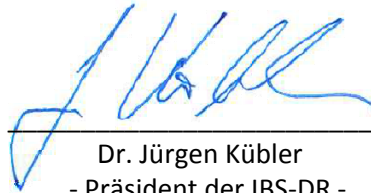
Köln/Hannover, 15. Mai 2013

Mit freundlichen Grüßen



---

Prof. Dr. Heike Bickeböller  
- Präsidentin der GMDS -



---

Dr. Jürgen Kübler  
- Präsident der IBS-DR -

Anhang: Literatur

## Literatur

- [1] IQWiG (2013): *Aktualisierung einiger Abschnitte der Allgemeinen Methoden Version 4.0 sowie neue Abschnitte zur Erstellung der Allgemeinen Methoden Version 4.1 vom 18.04.2013*. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), Köln.
- [2] IQWiG (2011): *Ticagrelor – Nutzenbewertung gemäß § 35a SGB V*. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), Köln, IQWiG-Berichte – Jahr 2011 Nr. 96.
- [3] CHMP (2012): *Guideline on clinical investigation of medicinal products in the treatment or prevention of diabetes mellitus*. Committee for medicinal products for human use (CHMP), CHMP/EWP/1080/00 Rev. 1
- [4] CHMP (2006): *Reflection paper on the regulatory guidance for the use of health related quality of life (HRQL) measures in the evaluation of medicinal products*. Committee for medicinal products for human use (CHMP), EMEA/CHMP/EWP/139391/2004
- [5] Marckmann G, Siebert U. Prioritäten in der Gesundheitsversorgung: Was können wir aus dem „Oregon Health Plan“ lernen? *Dtsch Med Wochenschr.* 2002;127:1601-4
- [6] Marckmann G, Siebert U. Kosteneffektivität als Allokationskriterium in der Gesundheitsversorgung. *Zeitschrift für medizinische Ethik.* 2002;48:171-190
- [7] Marckmann G. Priorisierung im Gesundheitswesen: Was können wir aus den internationalen Erfahrungen lernen?, in: *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 2009;103(2):85-91.