

## Leserbriefe

# Verfahrensweisen und Methoden zur Nutzenbewertung von Arzneimitteln in Deutschland – Ein Auftragsgutachten

Zum Supplement Nr. 7 in der DMW 49 am 5.12.2008<sup>1</sup>

## Zuschrift Nr. 1



Trudy Bekkering und Jos Kleijnen haben im Auftrag des deutschen Verbands forschender Arzneimittelhersteller (VFA) ein Gutachten zum Verfahren der Nutzenbewertung von Arzneimitteln in Deutschland erstellt. Bislang wurden solche Nutzenbewertungen alleine durch das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) erarbeitet. Das Gutachten wurde jetzt im European Journal of Health Economics [3] in englischer Sprache und in der Deutschen Medizinischen Wochenschrift auf deutsch [2] praktisch zeitgleich veröffentlicht.

Einer der beiden Autoren (Jos Kleijnen) war als externer Sachverständiger an mehreren Nutzenbewertungen des IQWiG aktiv beteiligt. Trotzdem sind in dem Gutachten die Prozesse des IQWiG aus unserer Sicht nicht angemessen beschrieben und seine Rahmenbedingungen nicht adäquat berücksichtigt. Im Folgenden möchten wir diese Aspekte erläutern.

Das IQWiG wurde 2004 gegründet. Seine aktuellen Methoden [8] sind eine stabile Grundlage für Entscheidungen im deutschen Gesundheitswesen und genügen voll den wissenschaftlichen und rechtlichen Anforderungen [4]. Weitere Anpassungen und Erweiterungen werden aber mit Sicherheit sinnvoll werden. Deshalb nehmen wir Anregungen gerne auf, soweit sie auf die deutsche Situation anwendbar sind. Hier liegt jedoch die entscheidende Schwäche des Gutachtens: Bekkering und Kleijnen erheben das NICE („National Institute for Health and Clinical Excellence“, Großbritannien) zum internationalen Standard und unterschätzen dabei die Bedeutung der strukturellen, gesetzlichen und kulturellen Vorgaben des deutschen Gesundheitssys-

tems. HTA-Institutionen (HTA= health technology assessment) mit wichtigen nationalen Aufgaben sind immer an unterschiedliche Kontexte angepasst [9, 13]. Es ist sozusagen internationaler Standard, dass sich die Verfahren der Nutzenbewertung von Land zu Land unterscheiden. Eisenberg hat das als „Globalize the evidence, localize the decision“ umschrieben [6].

Für internationale Pharmaunternehmen beinhalten solche lokalen Unterschiede allerdings eine zusätzliche Unsicherheit und Erschwernis. Die Kritik von Bekkering und Kleijnen ist deshalb für uns nicht neu. Der VFA und seine Mitgliedsfirmen haben die Vorgehensweise des Instituts schon in der Vergangenheit wiederholt kritisiert und insbesondere das NICE als Vorbild präsentiert [14, 15]. Bereits im Dezember 2004 hatte ein Gutachter im Auftrag des VFA, die IQWiG-Methoden 1.0 insbesondere mit denen des NICE verglichen [11].

Das neuerliche Gutachten von Bekkering und Kleijnen behandelt im Wesentlichen folgende Themen: Scoping, Umgang mit externen Sachverständigen, Stellungnahmen und Reviews, internationale EbM-Standards (EbM=evidenzbasierte Medizin) und das Verhältnis von wissenschaftlicher Evaluation zu Interpretation und Entscheidung (in der NICE-Terminologie als „assessment“ und „appraisal“ bezeichnet).

Grundsätzlich stimmen wir Bekkering und Kleijnen zu: Transparenz ist essenziell, ein Scoping-Prozess ist notwendig, EbM-Standards sollten gelten und das Verhältnis von „Assessment“ und „Appraisal“ sollte klar definiert sein. Wir sehen aber gute Gründe, warum das IQWiG keine Kopie des NICE sein kann.

### Scoping

Das Gutachten beschreibt ausführlich den Scoping-Prozess des NICE mit der Empfehlung, ihn in Deutschland zu übernehmen. In der Forderung, das IQWiG solle ein Scoping á la NICE einsetzen, liegt ein häufiger

Verständnisfehler: Während beim NICE die Wahl der Bewertungsthemen, die methodische Festlegung des Assessments und das Appraisal unter dem Dach einer Institution stattfinden, ist in Deutschland zumeist der G-BA (= Gemeinsamer Bundesausschuss) als Auftraggeber für die Wahl der Bewertungsthemen verantwortlich und trifft später die Entscheidung (Appraisal), die dann unter anderem auf den Ergebnissen der Nutzenbewertung beruht. Der GBA beauftragt das IQWiG mit dem Assessment. Kennzeichnend ist aber, dass der Auftrag bereits wesentliche Vorgaben für die Fragestellung enthält, z.B. sind Patientengruppen und die zu prüfende Technologie definiert.

Das Institut ist dann jedoch unabhängig in der Frage, wie es einen Auftrag fachlich-methodisch bearbeitet. Diese enger begrenzte Rolle des IQWiG im Vergleich zu der des NICE ist ein Grund dafür, warum in Deutschland die präzise Formulierung der Fragestellung als schrittweiser Prozess stattfindet. So lädt das IQWiG in der Regel Patientenvertreter zu einem eigenen Treffen ein, um sie in die Findung der patientenrelevanten Endpunkte einzubinden.

Wir denken nicht, dass der Prozess weniger transparent ist als der des NICE. Das IQWiG stellt seinen Vorschlag zur methodischen Ausgestaltung (vorläufigen Berichtsplan) in einer öffentlichen Anhörung zur breiten Diskussion. In diese Diskussion sind alle Gruppen eingebunden, die auch Bekkering und Kleijnen fordern. Unser Prozess beinhaltet sowohl die Möglichkeit zur schriftlichen Stellungnahme als auch zur Teilnahme an mündlichen Erörterungen. Die dabei vorgebrachten Argumente werden dokumentiert und vom IQWiG schriftlich gewürdigt (siehe unten). Dieser Diskussions- und Würdigungsprozess wird nach Veröffentlichung des Vorberichts noch einmal wiederholt.

Aus dem Auftraggeber-Auftragnehmer-Verhältnis zwischen G-BA und IQWiG ergibt sich aber, dass sich der Präzisionsprozess des IQWiG nicht auf Aspekte erstrecken kann, die durch den Auftrag vorgegeben sind.

### Würdigung der Stellungnahmen

Bekkering und Kleijnen behaupten, dass das IQWiG Stellungnahmen nicht offen evaluiere und dass es daher unklar sei, ob sie zu Änderungen im Bericht geführt hätten oder nicht. Diese Wahrnehmung

<sup>1</sup> Beitrag frei verfügbar unter <http://www.thieme-connect.de/ejournals/toc/dmw/68327>

erstaunt uns, weil die Dokumentation und Würdigung der Stellungnahmen ein Standardelement jeder IQWiG-Nutzenbewertung ist; dazu werden eigene Dokumente erstellt und auf der Internetseite [www.iqwig.de](http://www.iqwig.de) veröffentlicht. Diese Würdigungen schließen eine Beschreibung ein, welche Konsequenzen die in den Stellungnahmen vorgebrachten Argumente für das Projekt haben.

Es ist allerdings die Regel, dass wir viele und ausführliche Stellungnahmen erhalten, die in vielen vorgebrachten Argumenten identisch sind. In dieser Situation haben wir uns aus Effizienzgründen entschieden, identifizierte Argumente nur einmal zu würdigen und zu kommentieren. Wir nehmen den Hinweis, dass Bekkering und Kleijnen unsere Würdigungen nicht als solche wahrnehmen, jedoch als Anregung, die redaktionelle Darstellung zu überprüfen.

### Umgang mit Reviewern und Reviews

Bekkering und Kleijnen fordern eine Veröffentlichung externer Reviews. Es muss betont werden, dass der zentrale Review der IQWiG-Nutzenbewertungen durch das öffentliche Stellungnahmeverfahren stattfindet, an dem jeder teilnehmen kann, nicht nur „Stakeholder“ [8]. Alle Stellungnahmen werden dann veröffentlicht (siehe oben). Gesonderte „Reviews“ beauftragt das IQWiG nur als Element der internen Qualitätssicherung. Im Übrigen räumen Bekkering und Kleijnen ein, dass auch das NICE externe Reviews nicht veröffentlicht.

### Veröffentlichung der Namen der externen Sachverständigen

Das IQWiG ist gesetzlich verpflichtet, bei Nutzenbewertungen regelhaft mit externen Sachverständigen zusammenzuarbeiten. Es trifft zu, dass das IQWiG die Namen dieser externen Sachverständigen erst relativ spät im Laufe des Verfahrens veröffentlicht. Hintergrund ist, dass einige unserer externen Sachverständigen selbst um diesen Schutz bitten, weil sie fürchten, unter Druck gesetzt zu werden. Wir bedauern diese Situation, müssen sie aber vorerst akzeptieren. Wir müssen gewährleisten, dass Transparenz nicht denen schadet, die mit dem IQWiG kooperieren.

### Veröffentlichung der Zwischendokumente der externen Sachverständigen

▼  
Bekkering und Kleijnen [2] fordern, dass das IQWiG „Berichte“ der externen Sach-

verständigen publiziert. Eine solche Veröffentlichung ist im Prozess des NICE sinnvoll, da dort die externen Sachverständigen alleine das Assessment übernehmen.

In Deutschland ist das Assessment aber gerade eine **gemeinsame** Aufgabe von IQWiG und externen Sachverständigen. Eine Nutzenbewertung entsteht erst durch den gemeinsamen Input, so dass eine Veröffentlichung von Zwischendokumenten das Assessment zersplittern würde. Dieses gemeinsame und konsentrierte Assessment wird durch den Vorbericht dokumentiert. Darüber hinaus erfordern die mit den Arzneimittelherstellern getroffenen Vertraulichkeitsvereinbarungen zur Überlassung bislang unveröffentlichter Daten einen besonderen Umgang mit diesen Daten durch das IQWiG. Diese Daten müssen auch aus rechtlichen Gründen institutsintern gehandhabt werden. Im Unterschied zum NICE veröffentlicht das IQWiG dann aber alle Daten, die für die Nutzenbewertung relevant sind. Auch das halten wir für ein wichtiges Element der Transparenz.

Im Übrigen trägt das IQWiG allein die Verantwortung für die Qualität der Berichte, so dass auch das letzte Wort beim IQWiG liegen muss. Falls externe Sachverständige eine Sachlage nicht angemessen durch das IQWiG interpretiert sehen, steht es ihnen frei, das auch öffentlich zu kritisieren.

### Internationale EbM-Standards

Bekkering und Kleijnen [2] behaupten, dass die Bewertung der „best available evidence“ internationaler EbM-Standard sei. Das mag für die Leitlinien gelten, die das NICE erstellt – das gilt aber sicher nicht für die Erstellung von HTA-Berichten [12].

Hinzu kommt, dass es beim IQWiG zumeist um die Erarbeitung von Nutzenbewertungen im Kontext der deutschen Sozialgesetzgebung geht. Der Auftrag der Nutzenbewertung beinhaltet primär einen Nachweis von Kausalität. Schließlich sind es ja auch immer Kausalitätsbehauptungen, mit denen zum Beispiel Hersteller die Überlegenheit einzelner Arzneimittel begründen. Da ist es selbstverständlich, dass auch bei der Nutzenbewertung Kausalitätsnachweise geprüft werden. Es ist internationaler Standard, dass ein Kausalitätsnachweis hohe Anforderungen an die Ergebnissicherheit stellt, also an die interne Validität.

Das IQWiG erhebt an keiner Stelle randomisierte kontrollierte Studien zum „Dogma“, wie das Gutachten suggeriert, son-

dern legt die Studientypen nach Fragestellung fest, die ausreichend ergebnissicher sind. Das gilt auch für die Bewertung von Arzneimitteln.

Sofern es Hinweise auf einen dramatischen Effekt gibt, werden sogar Fallserien betrachtet, wie zum Beispiel bei der Nutzenbewertung bestimmter Formen der Stammzelltransplantation bei akuten Leukämien [7]. Allerdings sind dramatische Effekte bekanntermaßen eine seltene Ausnahme. Bei klein(er)en Unterschieden sind prospektiv geplante, vergleichende Studien mit Sicherstellung der Strukturgleichheit nötig, um mit einer ausreichenden Ergebnissicherheit zu einer Aussage zu kommen [16]. Das ist internationaler Standard. Nicht das Beharren auf diesem Prinzip, sondern das Abweichen davon bedarf einer besonderen Begründung [5,10]. Wenn es keinen Grund gibt, die „best available evidence“ als ausreichend ergebnissicher zu betrachten, dann teilt das IQWiG das seinem Auftraggeber so mit. Im Übrigen kann während des Anhörungsverfahrens Evidenz jeglichen Typs vorgebracht werden, insbesondere mit Hinweis auf dramatische Effekte, die dann vom IQWiG bewertet wird.

Wenn keine ausreichend sichere Aussage möglich ist, erlaubt das weitere Verfahren innerhalb der Entscheidungsfindung des G-BA dann aber sehr wohl die Abwägung anderer Argumente [4]. Außerdem hat der G-BA ausdrücklich die Möglichkeit, eine Entscheidung auszusetzen, um weitere Evidenz mit ausreichender Ergebnissicherheit abzuwarten.

### Verhältnis von Assessment zu Appraisal

Es trifft zu, dass Assessment und Appraisal in Großbritannien anders voneinander abgegrenzt sind als in Deutschland. Nutzenbewertungen des IQWiG sind ausdrücklich mit der im Sozialgesetzbuch V festgelegten Pflicht verbunden, eine Empfehlung abzugeben [1]. Die Formulierung des Gutachtens, dass es dem IQWiG lediglich „gestattet“ sei, Empfehlungen abzugeben, ist deshalb irreführend. Das IQWiG formuliert seine Empfehlung in Form eines kurzen Fazits, das eine Beschreibung der gefundenen Nutzenbelege enthält und macht eventuell Vorschläge für (weitere) Forschung.

### International herrscht Verfahrensvielfalt

Wir finden es im Übrigen überraschend, dass Bekkering und Kleijnen aus der Beschreibung einer einzigen Institution (dem

NICE) einen „internationalen Standard“ ableiten. Nach unserer Sichtung der publizierten Vergleiche internationaler HTA-Organisationen kommen wir vielmehr zu der Schlussfolgerung, dass jede Institution in ihren nationalen Strukturen und Diskussionsprozessen jeweils anders vorgeht [13, 17].

Eine Gleichsetzung von IQWiG und NICE ist auch deshalb unmöglich, weil beide Institute zum Beispiel bei der Bewertung von Arzneimitteln grundsätzlich andere Funktionen haben. Das NICE ist eine 4.-Hürde-Institution, die über den Zugang neuer Technologien in den Markt entscheidet. Das IQWiG ist ein Institut der nachträglichen Nutzenbewertung von Technologien, die längst im Markt sind. Dieser Unterschied schafft zum Beispiel ein fundamental unterschiedliches Verhältnis zu den „Stakeholdern“, wie etwa Pharmafirmen, für die Bekkering und Kleijnen ihr Gutachten erstellt haben: Im englischen Gesundheitswesen haben die Firmen eine hohe Motivation, zu einer Beschleunigung der Verfahren im NICE beizutragen, weil jeder Tag Verzögerung Umsatzverlust bedeutet. In Deutschland gibt es hingegen eine systembedingte Motivation, die Arbeit des IQWiG so lange wie möglich hinauszuzögern, weil sich solange am Status Quo nichts ändert, bis der G-BA auf der Basis des IQWiG-Berichts eine Entscheidung getroffen hat. Schon allein aus diesem Grund müssen sich die Verfahren des NICE und IQWiG unterscheiden.

Die vielfach durch Vertreter von Arzneimittelherstellern in Deutschland erhobene und von Bekkering und Kleijnen geteilte Forderung, bei der Nutzenbewertung „alle“ Evidenzlevel zu recherchieren, zu sichten und zu bewerten, würde auch den Aufwand für die Projekte massiv erhöhen und die Bearbeitungszeit unangemessen ausdehnen. Das Fehlen von Studien mit ausreichender Ergebnissicherheit würde dadurch trotzdem nicht kompensiert, da sich klare Empfehlungen nur aus Studien mit ausreichender Ergebnissicherheit ableiten lassen.

Angesichts der wesentlichen Kontextunterschiede zwischen NICE und IQWiG halten wir das deutsche Verfahren der Nutzenbewertung für gleichwertig und nicht für weniger transparent. Jeder Interessierte hat Gelegenheit zu Mitarbeit und Kritik. Gleichzeitig garantiert das Verfahren die wissenschaftliche Unabhängigkeit der Nutzenbewertung.

Aus unserer Sicht liegt die größte Gefährdung von Transparenz ohnehin in der Tat-

sache, dass nach wie vor Pharmafirmen und andere die Ergebnisse von Studien nicht, nicht vollständig oder nur verzögert veröffentlichen. Wir stimmen deshalb völlig mit der Forderung von Bekkering und Kleijnen überein, dass „Hersteller alle Informationen zur Nutzenbewertung zur Verfügung stellen“ sollten. Wir sind jetzt gespannt, wie die Mitgliedsfirmen des VFA diesen Aspekt des Gutachtens umsetzen.

#### Literatur

- 1 Gesetz zur Stärkung des Wettbewerbs in der gesetzlichen Krankenversicherung (GKV-Wettbewerbsstärkungsgesetz – GKV-WSG). 26.03.2007. Bundesgesetzblatt 2007; Teil 1 (11): 378–473
- 2 Bekkering GE, Kleijnen J. Verfahrensweisen und Methoden zur Nutzenbewertung von Arzneimitteln in Deutschland. Dtsch Med Wochenschr 2008; 133: S225–S246
- 3 Bekkering G, Kleijnen J. Procedures and methods of benefit assessments for medicines in Germany. The European Journal of Health Economics 2008; 9: 5–29
- 4 Bertelsmann H, Roters D, Bronner D. [Benefits and benefits assessment: the principles of evidence-based medicine and health technology assessment as a decisional basis for the federal joint committee]. Z Arztl Fortbild Qualitatssich 2007; 101: 455–462
- 5 Committee on Reviewing Evidence to Identify Highly Effective Clinical Services. Knowing what works in health care: A roadmap for the nation. Washington, DC: Institute of Medicine (IOM), 2008 (The National Academies Press; Band URL: <http://www.nap.edu/catalog/12038.html>)
- 6 Eisenberg JM. Globalize the evidence, localize the decision: evidence-based medicine and international diversity. Health Aff (Millwood) 2002; 21: 166–168
- 7 Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Stammzelltransplantation bei den Indikationen Akute lymphatische Leukämie (ALL) und Akute myeloische Leukämie (AML) bei Erwachsenen. Abschlussbericht N05–03A. Köln: IQWiG, 2007 URL: [http://www.iqwig.de/download/N05-03A\\_Abschlussbericht\\_Stammzelltransplantation\\_be\\_ALL\\_und\\_AML.pdf](http://www.iqwig.de/download/N05-03A_Abschlussbericht_Stammzelltransplantation_be_ALL_und_AML.pdf)
- 8 Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden 3.0. Köln: IQWiG, 2008 URL: [http://www.iqwig.de/download/IQWiG\\_Methoden\\_Version\\_3\\_0.pdf](http://www.iqwig.de/download/IQWiG_Methoden_Version_3_0.pdf)
- 9 Liberati A, Sheldon TA, Banta HD. EUR-ASSESS project subgroup report on methodology: methodological guidance for the conduct of health technology assessment. Int J Technol Assess Health Care 1997; 13: 186–219
- 10 Norris SL, Atkins D. Challenges in Using Non-randomized Studies in Systematic Reviews of Treatment Interventions. Ann Intern Med 2005; 142 (12\_Part\_2): 1112–1119
- 11 OHE Consulting. Note for the VFA – The Institute for Quality and Economic Efficiency in the Health Care Sector (IQWiG) [Online]. 2004 [Zugriff am 25.11.2008] URL: <http://www.vfa.de/download/SAVE/de/politik/artikelpo/nutzenbewertung.html/dritte/OHE-Stellungnahme-Entwurf-IQWiG-Methoden-Dez04.pdf>
- 12 Royle P, Waugh N. Literature searching for clinical and cost-effectiveness studies used in health technology assessment reports carried out for the National Institute for Cli-

nical Excellence appraisal system. Health Technol Assess 2003; 7: iii, ix–51

- 13 Schwarzer R, Siebert U. Methods, procedures, and contextual characteristics of HTA and health policy decision-making: Comparison of leading HTA agencies in Germany, UK, France and Sweden [Online]. 2008 [Zugriff am 18.11.2008] URL: <http://www.egms.de/en/meetings/gmds2008/08gmds043.shtml>
- 14 VFA. Stellungnahme des VFA zum Entwurf der allgemeinen Methoden 3.0 des IQWiG vom 15.11.2007 [Online]. 2007 [Zugriff am 25.11.2008] URL: <http://www.vfa.de/download/SAVE/de/politik/artikelpo/nutzenbewertung.html/vfapapers/vfastellungnahme-iqwig-methoden-3p0.pdf>
- 15 VFA. Positionspapier „Nutzenbewertung nach Methodenpapier 2.0 des IQWiG“ [Online]. 2007 [Zugriff am 14.11.2008] URL: <http://www.vfa.de/download/SAVE/de/politik/positionen/pos-nutzenbewertung.html/pos-nutzenbewertung.pdf>
- 16 Windeler J. Bedeutung randomisierter klinischer Studien mit relevanten Endpunkten für die Nutzenbewertung. In: Gesundheitsforschungsrat des Bundesministeriums für Bildung und Forschung (Hrsg). Diskussionsforum zur Nutzenbewertung im Gesundheitswesen: Begriffsdefinitionen und Einführung; Dokumentation des 1. gemeinsamen Workshops von GFR und IQWiG am 4. September 2007 in Berlin. Bonn: Gesundheitsforschungsrat, 2007: 26–31.
- 17 Zentner A, Velasco-Garrido M, Busse R. Methods for the comparative evaluation of pharmaceuticals [Online]. 2005 [Zugriff am 14.11.2008] URL: [http://portal.dimdi.de/de/hta/hta\\_berichte/hta122\\_bericht\\_de.pdf](http://portal.dimdi.de/de/hta/hta_berichte/hta122_bericht_de.pdf)

Dr. rer. medic. Klaus Koch

PD Dr. med. Stefan Lange

Institut für Qualität und Wirtschaftlichkeit im

Gesundheitswesen (IQWiG)

Dillenburger Straße 27

51105 Köln

Tel. 0221/356 85-156

Fax 0221/356 85-845

eMail [koch@iqwig.de](mailto:koch@iqwig.de)

[www.iqwig.de](http://www.iqwig.de)

DOI 10.1055/s-0028-1124000

#### Zuschrift Nr. 2

▼ Mit Interesse haben wir den Artikel der Kollegen Bekkering und Kleijnen „Verfahrensweisen und Methoden zur Nutzenbewertung von Arzneimitteln in Deutschland“ gelesen [2]. Die Autoren schreiben „in den nachfolgenden Abschnitten argumentieren wir, dass im Prinzip jedes zu betrachtende Outcome zu einer separaten Forschungsfrage führt und dass eine separate Ermittlung und Betrachtung der angemessenen Studientypen für jede Fragestellung nötig ist, um die jeweils beste Evidenz zu finden“. Weiter heißt es, „muss die Bewertung für jedes outcome ebenfalls auf dem Prinzip der „bestverfügbaren Evidenz“ durchgeführt werden.“

Der geübte Leser sieht, dass es sich hier und in anderen Stellen des Textes, insbesondere



im Abschnitt 4, um eine Widerspiegelung der konzeptionellen Ansätze des „Grading of Recommendations Assessment, Development and Evaluation“ (GRADE)-Systems handelt [1,3,4,5,6,7,8,9], dessen Nennung und Beschreibung den Lesern und der Diskussion wirklich helfen würden. Leider wird in dem Artikel das seit dem Jahre 2000 entwickelte und mit der Entwicklung in der methodologischen Forschung einhergehende GRADE-System weder erwähnt noch zitiert. Die Nennung des Ansatzes der „U.S. Preventive Services Taskforce“ beruht auf einem vorläufigen Schriftstück, das gegenwärtig unter Mitwirkung von Mitgliedern der GRADE-Arbeitsgruppe revidiert wird. Auch das NICE-System ist in dieser Form nicht mehr aktuell. GRADE ist eine systematische Weiterentwicklung der zitierten Literaturstelle von 1995 (Guyatt et al.), deren konzeptioneller Ansatz inzwischen teilweise überholt ist. Wie in anderen Gebieten der medizinischen Forschung, sind 13 Jahre ein langer Zeitraum, in dem Erkenntnisse verarbeitet werden und sollten. Viel schwerwiegender ist die Tatsache, dass die Autoren aber nicht konsequent eine (GRADE) Linie verfolgt haben und auf der Basis einer älteren Literaturstelle schreiben „Aus der Sicht der EBM ist das systematische Review von RCTs die beste und nützlichste verfügbare Evidenz. Daher sollte dieses Studiendesign das höchste Evidenzlevel darstellen.“ Leider ist der zweite Teil der Aussage in der modernen EBM isoliert so nicht mehr haltbar. Systematische Reviews stellen keine Form oder „level“ der Evidenz an sich dar; sie bilden vielmehr die Basis für eine Identifizierung und Bewertung der Evidenz. Die Evidenz in einem systematischen Review kann auf schlechten RCTs (randomisierte kontrollierte Studien) beruhen, eine anfängliche und gründliche Einzelbewertung der RCTs ist daher unbedingt nötig mit der Folge, dass der systematische Review RCTs niedriger Qualität zu Grunde haben kann und als systematischer Review deshalb nicht in die Evidenzhierarchie (als höchste Stufe) eingeordnet werden kann. Systematische Reviews (mit oder ohne Metaanalyse) müssen natürlich auch für beobachtende Studien durchgeführt werden, wenn sie relevant und informativ für die Fragestellung sind. Aus der Sicht der modernen EBM bleibt zu sagen, dass ein Sachverständigengutachten (oder expert opinion) auch keine Evidenz per se, sondern eine **Interpretation** der Evidenz für alle Studiendesigns unabhängig vom „level“ darstellt. Das Gutachten beruht somit auf Evidenz, die transparent dargestellt werden sollte,

ob als RCTs oder aufgrund der einzelnen Beobachtungen der Sachverständigen.

Da GRADE von internationalen Organisationen wie der WHO, NICE, CADTH, und in dem Bericht beschriebenen Institutionen benutzt wird und mit deren Einwirkung entwickelt wurde, hätte eine aktuelle, quellenbezogene Darstellung mit Referenz zum aktuellen Stand der Evidenzbewertung in „Institution“ deutlich mehr geholfen und die Diskussion auf ein anderes „level“ gehoben. Dennoch freuen wir uns, dass die Konzepte von GRADE, zusammengetragen von einer internationalen Arbeitsgruppe mit nunmehr über 120 Mitgliedern, auch außerhalb von GRADE weit vertreten werden und für Deutschland von Interesse sind. Die deutschsprachigen Mitglieder der GRADE-Arbeitsgruppe sind sehr daran interessiert und bemühen sich aktiv, eine informierte Diskussion in Deutschland über Evidenzbewertung und Abwägung von Schaden und Nutzen für Handlungsempfehlungen in der Gesundheitsversorgung zu unterstützen.

#### Referenzen

- 1 *Atkins D, Best D, Briss PA et al.* Grading quality of evidence and strength of recommendations. *BMJ* 2004; 328 (7454): 1490–1494
- 2 *Bekkering GE, Kleijnen J.* Verfahrensweisen und Methoden zur Nutzenbewertung von Arzneimitteln in Deutschland. *Dtsch Med Wochenschr* 2008; 133: S221–248
- 3 *Guyatt GH, Oxman AD, Vist GE et al.* GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008; 336: 924–6
- 4 *Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schunemann HJ.* What is „quality of evidence“ and why is it important to clinicians? *BMJ* 2008; 336: 995–998
- 5 *Guyatt GH, Oxman AD, Kunz R et al.* Incorporating considerations of resources use into grading recommendations. *BMJ* 2008; 336: 1170–1173
- 6 *Guyatt GH, Oxman AD, Kunz R et al.* Going from evidence to recommendations. *BMJ* 2008; 336: 1049–1051
- 7 *Schunemann HJ, Oxman AD, Brozek J et al.* Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* 2008; 336: 1106–1110
- 8 *Schunemann HJ, Jaeschke R, Cook DJ et al.* An official ATS statement: grading the quality of evidence and strength of recommendations in ATS guidelines and recommendations. *Am J Respir Crit Care Med* 2006; 174: 605–614
- 9 *Schunemann HJ, Best D, Vist G, Oxman AD.* Letters, numbers, symbols and words: how to communicate grades of evidence and recommendations. *Cmaj* 2003; 169: 677–680

Prof. Dr. med. Holger J. Schünemann, M.Sc., Ph.D.  
Chair, Department of Clinical Epidemiology & Biostatistics  
Michael Gent Chair in Healthcare Research  
Professor of Clinical Epidemiology/Biostatistics/  
Medicine  
McMaster University, Hamilton, Kanada  
Prof. Dr. Regina Kunz, M.Sc.  
Institut für Klinische Epidemiologie

Universitätsspital Basel  
Hebelstrasse 10  
4031 Basel  
Schweiz  
Asst. Prof. Dr. med. Yngve Falck-Ytter  
Case Western Reserve University Cleveland  
VA Medical Center Cleveland  
10701 East Blvd.  
Cleveland, OH, 44106, USA  
DOI 10.1055/s-0028-1124001

#### Erwiderung

Wir begrüßen die Antwort von Koch und Lange auf unsere Vorschläge [1] außerordentlich, da es wichtig ist, die Meinung des IQWiG zu den von uns diskutierten Punkten zu erfahren. Wir sind allerdings nicht der Ansicht, dass einer unserer Vorschläge durch diese Antwort geändert werden müsste. Methoden und Verfahrensweisen entwickeln sich im Laufe der Zeit naturgemäß weiter. Wir hoffen, dass dieser Prozess der Weiterentwicklung dazu führen wird, dass das Vorgehen des IQWiG zukünftig einmal von allen Interessengruppen als hervorragend eingestuft wird.

Begrüßen möchten wir auch die Antwort von Schünemann und Kunz. Es ist in der Tat wichtig, das GRADE-System in diesem Zusammenhang zu nennen. Wir möchten alle Leser darin bestärken, GRADE beim Umgang mit Evidenzleveln zu berücksichtigen. Dem Großteil der Aussagen der Autoren können wir zustimmen. Jedoch sind wir der Meinung, dass auch systematische Übersichtsarbeiten für sich genommen neue Evidenz liefern können. Ein gutes Beispiel bietet die systematische Übersichtsarbeit zur Kortikoidgabe bei drohender Frühgeburtlichkeit, welche graphisch im Logo der Cochrane Collaboration dargestellt ist. Diese Arbeit lieferte neue Ergebnisse, die durch die Betrachtung einzelner Studien nicht gewonnen wurden könnten. Darüber hinaus kann die erhöhte Power von systematischen Übersichtsarbeiten neue Evidenz über Subgruppen liefern, die in einzelnen randomisierten Studien nicht gesehen werden kann.

#### Referenzen

- 1 *Bekkering GE, Kleijnen J.* Verfahrensweisen und Methoden zur Nutzenbewertung von Arzneimitteln in Deutschland. *Dtsch Med Wochenschr* 2008; 133: S225–S246

Jos Kleijnen, Trudy Bekkering  
Kleijnen Systematic Reviews Ltd  
Westminster Business Centre  
10 Great North Way  
Nether Poppleton  
York YO26 6RB  
United Kingdom  
DOI 10.1055/s-0028-1124002