

**Zentrale Schlussfolgerungen  
der Kommentierenden Synopse der Fachpositionen zur Kosten-  
Nutzen-Bewertung für Arzneimittel  
unter Berücksichtigung der Ergebnisse  
der Fachtagung am 18. Juni 2007 und  
der wissenschaftlichen Klausurtagung am 7./8. November 2007**

---

- erstellt im Auftrag des Bundesministeriums für Gesundheit -

Autoren:

Prof. Dr. K.-H. Jöckel

Prof. Dr. T. Kohlmann

Prof. Dr. H. Raspe

Prof. Dr. J. Wasem

Essen, Lübeck und Greifswald, den 20.12.2007

0. *Kurz gefasst für den eiligen Leser*

- Das ursprünglich für den klinischen Kontext entwickelte Fünf-Schritt-Schema der Methodik der Evidenzbasierten Medizin hat sich auch für die Systemebene etabliert und bewährt: 1. Formulierung der Fragestellung, 2. Systematische Suche, 3. Bewertung der identifizierten Evidenz, 4. Synthese der Evidenz und 5. Interpretation der Ergebnisse.
- Fachlich international anerkannte Grundprinzipien für HTA sind vor allem
  - möglichst vollständige Vergegenwärtigung der thematisch einschlägigen Originalveröffentlichungen,
  - umfassende Prüfung und Kontrolle der Validität der Daten- und Informationsbasis,
  - Prüfung der Praxisrelevanz und Anwendbarkeit der Evidenz im jeweils gegebenen Praxiskontext und
  - Transparenz, Fairness und Rationalität aller Arbeitsschritte.

Diese zeitlich stabilen Grundprinzipien müssen jeweils in methodische Standards „übersetzt“ werden.

- Nutzen setzt den Nachweis der Wirksamkeit voraus. Randomisierte klinische Studien (RCTs) stellen den „Gold-Standard“ für den Nachweis von Wirksamkeit und Nutzen dar. Allerdings kann es notwendig werden, für weitere Nutzenbelege auch andere Studientypen heranzuziehen. Bei der Auswertung und Interpretation von Studien niedrigerer Evidenzgrade ist eine besonders sorgfältige Prüfung auf systematische Verzerrungen notwendig.
- In der Gesundheitsökonomie hat sich die Berechnung qualitätskorrigierter Lebensjahre (quality-adjusted life years, QALYs) zu einem de facto-Standard entwickelt. Offene Fragen bei der Anwendung von QALYs bestehen u.a. hinsichtlich der Referenzkollektive für die Bewertung von Gesundheitszuständen, der Übertragbarkeit von Studienergebnissen aus anderen Ländern und der Vergleichbarkeit von mit unterschiedlichen Messinstrumenten erhobenen Befunden. In vielen Behandlungsfeldern fehlt es noch an Nutzwertdaten.

- Bei der Kosten-Nutzen-Bewertung kann der Zusatznutzen krankheitsübergreifend, aber auch krankheitsspezifisch ermittelt werden. Die Nutzenbewertung ist optimal so auszugestalten, dass zugleich auch die notwendigen Informationen für eine Kosten-Nutzen-Bewertung erhoben werden können. Kosten- und Nutzen-Inkremente sind in Form der inkrementellen Kosten-Effektivitäts-Relation gegenüber zu stellen.
- In einer Kosten-Nutzen-Bewertung kann Modellierung aus einer Reihe von Gründen für die Ermittlung der inkrementellen Kosten-Effektivitäts-Relation eine zentrale Rolle spielen. Es bestehen international Standards der Modellierung und ihrer Qualitätssicherung.
- Die Festsetzung eines Erstattungshöchstbetrages auf Basis einer Kosten-Nutzen-Bewertung spiegelt die Zahlungsbereitschaft für den Zusatznutzen wider. Bei indikationsspezifischer Messung des Zusatznutzens ist über die verschiedenen Indikationen hinweg Konsistenz bei der Festsetzung des Erstattungshöchstbetrages anzustreben.
- Regelmäßig sollten bei der Kosten-Nutzen-Bewertung Auswirkungen auf weitere Kostenarten und Ausgabenbereiche jenseits der GKV berücksichtigt werden. Innerhalb der jeweils gewählten Perspektive sind sämtliche Ressourcenverbräuche einzubeziehen.
- Beim G-BA sollte zur Operationalisierung und vor Vergabe des Auftrags ein Scoping-Workshop eingerichtet werden.
- Durch ein zu Beginn einer Kosten-Nutzen-Bewertung formuliertes (günstigen Falls von allen Beteiligten konsentiertes) Protokoll ist die notwendige Transparenz über den Bewertungsprozess herzustellen. Bei der Abfassung des Protokolls sollten der biometrischen Methodik und der epidemiologischen Übertragbarkeit ein besonderes Gewicht beigemessen werden.
- Zur Sicherung der gesetzlich geforderten und auch internationalen Standards entsprechenden Beteiligungsmöglichkeiten sollte die mündliche wissenschaftliche Erörterung zu vorgelegten Dokumenten des IQWiG obligatorisch sein. Hierbei muss sichergestellt sein, dass die erforderliche methodische Kompetenz und

die medizinischen Wissenschaften angemessen vertreten sind. Hierzu sollte eine Verfahrensordnung entwickelt werden.

### *1. Evidenzbasierte Medizin auf der klinischen Ebene und auf der Systemebene*

Evidenzbasierte Medizin (EbM) ist ein umfassendes Konzept, das ursprünglich im klinischen Kontext formuliert wurde. Neben der wissenschaftlichen, externen Evidenz sind die ärztliche Erfahrung sowie Präferenzen und Werte von Patienten und Angehörigen Grundlage der Entscheidungsfindung in der klinischen Situation. Ein strukturell konsentiertes Vorgehen lässt sich jedoch nur für das Vorgehen für die Gewinnung der externen Evidenz feststellen, während die Einbeziehung von Erfahrung und von Werten aus den methodischen Betrachtungen weitgehend herausgelassen wird, wenngleich es auch dort Arbeiten z. B. zu formalisierten gemeinsamen Entscheidungsprozessen gibt.

Das für den klinischen Kontext entwickelte Fünf-Schritt-Schema der Methodik der EbM hat sich auch für die Allokationsentscheidungen auf der Systemebene – wo teilweise der Begriff „Evidence based Health Care“ eingeführt ist – etabliert und bewährt: 1. Formulierung der Fragestellung, 2. Systematische Suche, 3. Bewertung der identifizierten Evidenz, 4. Synthese der Evidenz und 5. Interpretation der Ergebnisse.

### *2. Biometrie und Epidemiologie als Basiswissenschaft*

Das Grundproblem der Kosten-Nutzen-Bewertung stellt sich aus biometrischer und epidemiologischer Sicht wie folgt dar:

- Zum interessierenden Medikament gibt es verschiedene Studien, die nicht unbedingt alle – z. T. aufgrund regulatorischer Auflagen – denselben Endpunkt im Visier haben.
- Die Studien weisen unterschiedliche Ein/Ausschlusskriterien auf, womit die Frage der externen Validität gestellt werden muss.
- Manche Studien sind nur international, manche auch nur national, manche gar nicht publiziert, was bei der zusammenschauenden Bewertung die Frage nach dem publication bias aufwirft.

- Vielfach fehlen Studien, die auf die Versorgungspraxis ausgerichtet sind; ggfs. sind weiteren Studien (z.B. pragmatic trials, epidemiologische Studien) erforderlich.
- Es liegen quantitativ und qualitativ unterschiedliche Informationen zur Versorgungsnotwendigkeit und Versorgungspraxis vor.

In dieser Situation erhalten die medizinische Biometrie und Epidemiologie als Basiswissenschaft des hier betrachteten Vorgehens eine besondere Relevanz: Studienprotokolle und SOPs erzeugen die notwendige Transparenz. Beide legen prospektiv und nachvollziehbar fest, wie mit unterschiedlichen Situationen und mit zum Zeitpunkt des Studienbeginns noch unklaren Bedingungen umzugehen ist.

In analoger Weise sollte für eine Kosten-Nutzen-Bewertung ein Protokoll formuliert werden, das festlegt (günstigen Falls von allen Beteiligten konsentiert), wie man zu einer wissenschaftlich fundierten Einschätzung der Kosten und des Nutzens eines konkreten Medikaments kommen möchte. Analog zum Studienprotokoll einer klinischen Studie sind hier Amendments vorzusehen, die es gestatten, auf neu aufgetretene und unvorhergesehene Situationen adäquat zu reagieren. Im Kern handelt es sich dabei um ein abgestimmtes Verfahren, bei dem Konsens darüber erzielt wird, welche Studien in die Auswertung und in welcher Weise einbezogen werden und wie Datenlücken modellierend überbrückt werden können. Die Ergebnisermittlung erfolgt damit eindeutig definierten Standards, die a priori festgelegt werden, bei Bedarf entsprechend modifiziert werden und – und das ist das zentrale Ergebnis – für jeden nachvollziehbar überprüft werden können.

Diese Vorgehensweise wird sowohl der Forderung nach international anerkannten Standards als auch nach Transparenz im Sinne des GKV-WSG gerecht.

### 3. *Health Technology Assessment*

Das zentrale Instrument der umfassenden Bewertung von Arzneimitteln und anderen Behandlungsformen ist international das sog. Health Technology Assessment (HTA). Zu einem vollständigen HTA gehört über die Nutzenbewertung hinaus auch die Bearbeitung etwa ethischer, sozialer, ökonomischer und juristischer Aspekte. Das Europäische Netzwerks EuNetHTA ([www.eunethta.net](http://www.eunethta.net)) nennt insgesamt neun Unter-

kapitel eines HTA-Berichts, die drei zentralen Abschnitte beziehen sich auf die Sicherheit, die Effektivität und die ökonomischen Aspekte der untersuchten Gesundheitstechnologie.

In Deutschland wird das IQWiG in der Regel nicht mit einem umfassenden HTA beauftragt, sondern sein Auftrag ist vom Gesetzgeber (bei Arzneimitteln) auf eine Nutzen- oder Kosten-Nutzen-Bewertung eingeschränkt. Fachlich international anerkannte Grundprinzipien der eine Nutzenbewertung leitenden evidenzbasierten Medizin sind (hier als umfassende Methodologie verstanden) – nach der prioritären Berücksichtigung der Prinzipien des Wohltuns und der Schadensvermeidung in der Patientenbehandlung („beneficence, nonmaleficence“ nach Beauchamps und Childress 5.2001) – vor allem

- eine möglichst vollständige Vergegenwärtigung der thematisch einschlägigen Originalveröffentlichungen einschließlich ihrer Lücken und Grenzen nach dem Prinzip der „besten verfügbaren Evidenz“,
- eine umfassende Prüfung und Kontrolle ihrer Daten- und Informationsbasis in Hinblick auf ein breites Spektrum möglicher Verzerrungen (Biases, systematische Fehler), auf Zufallseinflüsse und weitere Unsicherheiten,
- eine Prüfung der Praxisrelevanz und Anwendbarkeit der Evidenz im jeweils gegebenen Praxiskontext und
- allgemein die Transparenz, Fairness und Rationalität („reasonableness“) aller Arbeitsschritte unter Berücksichtigung verschiedener Perspektiven und Gruppen.

Diese zeitlich stabilen Grundprinzipien müssen jeweils in methodische Standards „übersetzt“ werden. Diese verändern sich notwendigerweise mit den Fortschritten der klinischen Wissenschaften, der klinischen Epidemiologie und Biometrie und von Ethik und Recht.

Nutzen ist eine komplexe Größe; ihre Diskussion setzt den Nachweis der Wirksamkeit einer bestimmten Untersuchungs- und Behandlungsmethode voraus. Im Einzelnen zu bestimmen sind dann Nutznießer, Nutzenqualitäten, Nutzensausmaß, Eintrittswahrscheinlichkeit sowie Eintrittsbeginn und Nachhaltigkeit. Ähnlich differenziert sind Merkmale von Schadensrisiken zu vergegenwärtigen, um Netto- und Zusatznutzen analysieren zu können.

Die von IQWiG und G-BA genutzte Unterscheidung von (rein wissenschaftlichem und wertfreiem) Assessment und umfassendem Appraisal im Versorgungskontext überzeugt nicht vollständig. Auch im Assessment sind zahlreiche Wertfragen und –entscheidungen enthalten. Dies beginnt bei der Auswahl der einzubeziehenden Literatur und endet noch nicht bei der Beurteilung der inferenzstatistisch geprüften kritischen Differenz zwischen Interventions- und Kontrollgruppe. Solche Wertentscheidungen sind in jedem HTA offen zu legen und zu diskutieren. Dabei sind auch die Perspektiven, Interessen und Präferenzen verschiedener Gruppen sog. Stakeholder zu berücksichtigen. Eine umfassende und auch ethische Aspekte berücksichtigende Nutzenbewertung erfordert ein multi- und interdisziplinäres Vorgehen.

#### *4. Zur Rolle von randomisierten kontrollierten klinischen Studien (RCTs) und anderen Studienformen*

Es besteht international wie national eine breite Übereinstimmung, dass „konfirmatorische“ RCTs den „Gold-Standard“ in Bezug auf die Frage des statistischen Nachweises der Wirksamkeit eines Arzneimittels und auch eines Zusatznutzens gegenüber einer Vergleichsbehandlung darstellen. International haben sich „Evidenzhierarchien“ durchgesetzt, die etwa auch der G-BA seiner Verfahrensordnung (§ 18) zugrunde legt.

Auch für die Kosten-Nutzen-Bewertung bilden RCTs eine wichtige Basis der Analyse. Der Grundsatz der evidenzbasierten Medizin, die jeweils beste verfügbare Evidenz heranzuziehen, kann es allerdings notwendig werden lassen, auch andere Studientypen (z.B. auch der populationsorientierten Versorgungsforschung) heranzuziehen:

- RCTs decken oft nicht alle notwendigen Vergleichsinterventionen ab, wie sie für die Nutzen- und die Kosten-Nutzen-Bewertung benötigt werden.
- Insbesondere für die Beurteilung der (langfristigen) Nutzen-Risiko-Relation und der Kosten-Effektivität ist die Laufzeit von RCTs häufig zu kurz; dies gilt vor allem bei der Beurteilung im Kontext chronischer Erkrankungen. Die Untersuchung von Schadenspotentialen kann auf Studien anderen Zuschnitts kaum verzichten (Phase 4-Studien, Registerstudien, Anwendungsbeobachtungen, Outcomes Research).

- RCTs können aufgrund des „künstlichen“ Settings der Studien (besondere Patientenauswahl, Monitoring) den „wahren“ Effekt über- aber auch unterschätzen. Sie spiegeln nicht immer die Effektivität unter Alltagsbedingungen wider. „Falsch positive“ RCTs können durch Ergebnisse von Beobachtungsstudien relativiert werden; das Gegenteil kann in bestimmten Konstellationen auch relevant sein.
- RCTs können zwar bei der Identifikation von Kostenstellen hilfreich sein, im allgemeinen werden RCTs aber für die Beantwortung der Frage nach den Kostenunterschieden zwischen den Behandlungsalternativen nicht ausreichen; hier sind nahezu regelhaft ergänzende andere Daten erforderlich. Dies gilt aufgrund der nationalen Besonderheiten der Gesundheitssysteme und ihrer spezifischen relativen und absoluten Ressourcenverbräuche bei multinationalen RCT um so mehr.

„Pragmatische“ kontrollierte Studien, die z.B. hinsichtlich Ein- und Ausschlusskriterien oder Vergleichsintervention realistischer sind, werden zunehmend als attraktive Alternativen eingeschätzt.

Bei der Auswertung von Studien niedrigerer Evidenzgrade ist eine besondere sorgfältige Prüfung notwendig, ob in den Studien eine angemessene Kontrolle hinsichtlich möglicher systematischer Verzerrungen insbesondere durch Störvariablen durchgeführt wurde.

Besondere methodische Aufmerksamkeit ist bei „indirekter Evidenz“ und „fehlender Evidenz“ angebracht. Evidenz wird dann als indirekt bezeichnet, wenn sie sich nicht unmittelbar auf die interessierende Methoden, Komparatoren, Outcomes, Zeitpunkte und Populationen bezieht. Man ist dann auf Modellierungen und Sensitivitätsanalysen angewiesen. Dies wird auch dann notwendig, wenn Evidenz ganz fehlt. Dies spricht nicht automatisch gegen die zu beurteilende Methode („absence of evidence is not evidence of absence“). Man wird sich immer fragen müssen, ob und wieweit noch Schätzungen möglich sind oder ob es nicht besser wäre, auf neue Evidenz zu setzen und weitere Studien auf den Weg zu bringen.

### *5. Gesundheitsbezogene Lebensqualität als Bewertungskriterium für diagnostische und therapeutische Maßnahmen*

Parameter der subjektiven Gesundheit haben in den vergangenen Jahren in der Nutzen- und der Kosten-Nutzenbewertung eine zunehmende Bedeutung erlangt. Besonders berücksichtigt wurde dabei das Konzept der gesundheitsbezogenen Lebensqualität als wesentliches Bewertungskriterium für diagnostische und therapeutische Maßnahmen.

Das Instrumentarium zur Erfassung der gesundheitsbezogenen Lebensqualität umfasst Methoden zur diagnoseübergreifenden, krankheitsspezifischen und präferenzbasierten Lebensqualitätsmessung. Die präferenzbasierten Methoden ermöglichen eine Verdichtung der erhobenen Daten zu einem Index, der in gesundheitsökonomische Kosten-Nutzwert-Analysen durch Berechnung qualitätskorrigierter Lebensjahre (quality-adjusted life years, QALYs) einbezogen werden kann.

Bei der Berechnung qualitätskorrigierter Lebensjahre und bei der Einbindung dieser Parameter in die Kosten-Nutzen-Bewertung sind verschiedene Anwendungsvoraussetzungen zu klären. Klärungsbedarf besteht u. a. bei der Frage der Referenzkollektive für die Bewertung von Gesundheitszuständen, der Übertragbarkeit von Studienergebnissen aus anderen Ländern und der Vergleichbarkeit von mit unterschiedlichen Messinstrumenten erhobenen Befunden.

Kosten-Nutzwert-Analysen mit Berücksichtigung qualitätskorrigierter Lebensjahre stellen derzeit international einen de facto Standard dar. Sie werden in einer Vielzahl von europäischen und außereuropäischen Ländern bei Zulassungs- und Erstattungsentscheidungen einbezogen.

### *6. Das Verhältnis der Kosten-Nutzen-Bewertung zur Nutzenbewertung*

Eine Kosten-Nutzen-Bewertung vergleicht das jeweilige Arzneimittel mit einer oder mehreren zweckmäßigen Therapiealternativen. Hierbei werden Zusatzkosten und Zusatznutzen jeweils gegenüber einer Vergleichstherapie errechnet und in der inkrementellen Kosten-Effektivitäts-Relation (engl. abgekürzt: ICER) gegenüber gestellt. Für die Bestimmung des Zusatznutzens ist jeweils ein valider Parameter festzulegen; dies kann auch eine zusammengesetzte Größe (zum Beispiel Major Adver-

se Cardiac Events, MACE) sein. Es kann sich um krankheitsspezifische, aber auch krankheitsübergreifende Indikatoren (etwa QALYs) handeln.

In Deutschland wird es sowohl die Kosten-Nutzen-Bewertung als auch weiterhin eine isolierte Nutzenbewertung geben. Es ist sinnvoll, die Nutzenbewertung so auszugestalten, dass zugleich auch die notwendigen Informationen für eine Kosten-Nutzen-Bewertung, die den internationalen gesundheitsökonomischen Standards entspricht, erhoben werden können.

In einer Kosten-Nutzen-Bewertung kann Modellierung aus einer Reihe von Gründen für die Ermittlung der inkrementellen Kosten-Effektivitäts-Relation eine zentrale Rolle spielen. Insbesondere können durch Modellierung die Begrenzungen klinischer Studien (incl. RCTs) adressiert werden. Es bestehen internationale Standards der Modellierung; auch sind internationale Standards für die Bewertung der Qualität von Modellen verfügbar. Modelle müssen sich dabei immer auf das klinische Entscheidungsproblem der Kosten-Nutzen-Bewertung beziehen. Sie können nicht besser sein als ihr Daten-Input.

### *7. Ermittlung der Zahlungsbereitschaft für den Zusatznutzen*

Im deutschen Kontext wird die Kosten-Nutzen-Bewertung insbesondere bei der Festsetzung von Erstattungshöchstbeträgen für patentgeschützte Arzneimittel von Bedeutung sein. Die Festsetzung eines Erstattungshöchstbetrages spiegelt die Zahlungsbereitschaft der relevanten Entscheidungsträger für den Zusatznutzen wider. Einige Länder verfolgen den Ansatz, den Zusatznutzen in qualitätsadjustierten Lebensjahren (QALYs) zu messen und entsprechend eine finanzielle Obergrenze für ein zusätzlich gewonnenes QALY festzulegen. Demgegenüber besteht auch die Möglichkeit, die finanzielle Obergrenze jeweils für indikationsspezifisch operationalisierten Zusatznutzen zu ermitteln. Die Autoren weisen darauf hin, dass bei indikationsspezifischer Messung des Zusatznutzens über die verschiedenen Indikationen hinweg Konsistenz bei der Festsetzung des jeweiligen Erstattungshöchstbetrages anzustreben ist.

Bei der Ermittlung der Kosten für die Berechnung der inkrementellen Kosten-Effektivitäts-Relation ist die Definition der einzunehmenden Perspektive von wesent-

licher Bedeutung. Nach Auffassung der Autoren muss sichergestellt sein, dass der G-BA und der Spitzenverband Bund der Krankenkassen in die Lage versetzt werden, bei ihren Entscheidungen regelhaft Auswirkungen auf andere Kostenarten und Ausgabenbereiche jenseits der GKV zu berücksichtigen, sofern nicht unbestritten solche Auswirkungen irrelevant sind. Der geeignete Ort darüber zu entscheiden, ob solche Auswirkungen bereits a priori als irrelevant eingeschätzt werden können, ist ein beim G-BA im Kontext der Auftragsformulierung anzusiedelnder Scoping-Workshop. Innerhalb der jeweils gewählten Perspektive sind sämtliche Ressourcenverbräuche, nicht nur die Kosten der Arzneimittel, einzubeziehen. Stehen etwa höheren Preisen für Arzneimittel Einsparungen oder Mehrausgaben in anderen Bereichen (z.B. Krankenhaus) gegenüber, sind diese zu saldieren.

Die Feststellung nach Gegenüberstellung von Zusatznutzen und Zusatzkosten, ob die Kostenübernahme durch die Versicherungsgemeinschaft bei der ermittelten Kosten-Nutzen-Relation „angemessen und zumutbar“ (§ 35b Abs. 1 SGB V) ist, ist primär gesellschaftlicher Natur. Solche Entscheidungen gehören nach Auffassung der Autoren originär in die Hände der Politik oder zumindest des G-BA oder des Spitzenverbandes Bund, nicht aber des IQWiG, das als wissenschaftliche Einrichtung konzipiert ist.

#### *8. Anforderungen an den Bewertungsprozess*

Der Gesetzgeber des GKV-WSG hat festgelegt, dass das IQWiG bei der Erstellung der Bewertungen (und auch bereits zuvor bei der Festlegung seiner Methoden) „hohe Verfahrenstransparenz und eine angemessene Beteiligung“ der maßgeblichen Beteiligten zu realisieren hat (§ 35b Abs. 1 Satz 6 SGB V).

- Wir sehen einen Scoping-Workshop beim G-BA als sinnvollen Bestandteil eines solchen Beteiligungsprozesses an.
- Zur Sicherung angemessener Beteiligungsmöglichkeiten sollte die mündliche wissenschaftliche Erörterung zu vorgelegten Dokumenten des IQWiG obligatorisch sein. Hierbei muss sichergestellt sein, dass die erforderliche methodische Kompetenz und die medizinischen Wissenschaften (Theorie und Klinik) angemessen vertreten sind. Hierfür ist eine Verfahrensordnung zu entwickeln. Auf-

grund der vom Gesetzgeber gewünschten Einbindung insbesondere auch internationaler Experten sollte Englisch als Verhandlungssprache sowohl für das schriftliche Stellungnahmeverfahren als auch für die mündlichen Anhörungen zugelassen werden. Die vorgegebene Begrenzung von Stellungnahmen auf 10 Seiten sollte durch eine angemessenere Umfangsangabe ersetzt werden. Verhandlungssprache und Umfang der Unterlagen dürfen nicht zu einer Überlastung der Beteiligten führen.