

Kommentierende Synopse der Fachpositionen zur Kosten- Nutzenbewertung für Arzneimittel

– Erstellt im Auftrag des Bundesministeriums für Gesundheit –

Verfasser:

Dr. Gerd Antes
(Deutsches Cochrane Zentrum)

Prof. Dr. Karl-Heinz Jöckel
(Universität Duisburg-Essen)

Prof. Dr. Thomas Kohlmann
(Universität Greifswald)

Prof. Dr. Heiner Raspe
(Universität Lübeck)

Prof. Dr. Jürgen Wasem
(Universität Duisburg-Essen)

Freiburg, Essen, Greifswald und Kiel

Oktober 2007

Executive summary

1. Das Bundesministerium für Gesundheit (BMG) hat eine Arbeitsgruppe von Wissenschaftlern mit einer kommentierenden Synopse der Fachpositionen zur Kosten-Nutzenbewertung für Arzneimittel beauftragt. Der Arbeitsgruppe gehörten an:

- Dr. Gerd Antes (Chochrane Zentrum Deutschland)
- Prof. Dr. K.-H. Jöckel (Universität Duisburg-Essen)
- Prof. Dr. Thomas Kohlmann (Universität Greifswald)
- Prof. Dr. Heiner Raspe (Universität Lübeck)
- Prof. Dr. Jürgen Wasem (Universität Duisburg-Essen)

Der Adressatenkreis dieser Synopse ist das Fachpublikum, so dass im folgenden Text die fachlichen Termini nicht gesondert erläutert, sondern als bekannt vorausgesetzt werden.

2. Der Begriff der evidenzbasierten-Medizin (EbM) hat seit seiner Einführung 1991 eine enorme Verbreitung erfahren. EbM ist ein umfassendes – ursprünglich im klinischen Kontext formuliertes – Konzept, in dem patientenorientierte Entscheidungen ausdrücklich auf wissenschaftlicher Basis getroffen werden sollen. Neben der wissenschaftlichen, externen Evidenz sind die ärztliche Erfahrung sowie Präferenzen und Werte von Patienten und Angehörigen Grundlage der Entscheidungsfindung. Ein zumindest 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 Entscheidungsprozessen gibt. Weitgehend etabliert und akzeptiert ist ein Fünf-Schritt-Schema mit 1. Formulierung der Fragestellung, 2. Systematische Suche, 3. Bewertung der identifizierten Evidenz, 4. Synthese der Evidenz und 5. Interpretation der Ergebnisse, das heute in unterschiedlicher Intensität für die Aufbereitung von externer Evidenz international üblich ist.

Andererseits verbirgt sich unter dem Kürzel EbM eine solche Vielfalt an unterschiedlichen Interpretationen, dass der erreichbare kleinste Nenner „evidenzbasiertes Vorgehen“ zu sein scheint. Nur in dieser Interpretation lässt sich die ge-

setzliche Forderung damit vereinbaren, dass im deutschen Kontext die Systemebene einbezogen ist, also eigentlich die Evidence-based Healthcare (EBHC) statt EbM genannt werden müsste. Die damit verbundene Vielfalt an Strukturen und Vorgehensweisen aus dem internationalen Raum lassen keine eindeutige Umsetzung der Forderung „nach anerkannten internationalen Standards der evidenzbasierten Medizin (EbM)“ zu, sondern damit wird nur ein Rahmen abgesteckt, in dem die Operationalisierung, Prozeduren und die Wahl geeigneter Methoden festgelegt werden müssen. Die Konzepte sind eng mit der klinischen Epidemiologie verbunden, stützen sich darüber hinaus jedoch auf Methoden aus diversen Fachgebieten. Evidenzbasiertes Vorgehen beinhaltet eine Reihe von nicht abschließend beantworteten, kontrovers diskutierten Fragen, die weitere wissenschaftliche Diskussion erfordern, damit jedoch keine eindeutigen Antworten erwarten lassen. Transparenz in allen Schritten ist deswegen eine zentrale Forderung, um Methoden und Ergebnisse beurteilen zu können.

Insbesondere ist wichtig, dass die Vorgaben aus dem EbM/EBHC-Feld keine unmittelbaren Angaben zur Integration zur Kostenbewertung machen. Allerdings hat sich in den letzten Jahren ein methodisches Fundament für die empirische Arbeit weiterentwickelt, dessen Grundprinzipien – vor allem der Schutz vor Bias und die Kontrolle vor Zufallsfehlern – auch bei der Methodenwahl für die Kosten-Nutzen-Bewertung berücksichtigt werden sollte.

3. Insgesamt stellt sich das Grundproblem der Kosten-Nutzen-Bewertung aus biometrischer und epidemiologischer Sicht wie folgt dar:
 - zum interessierenden Medikament gibt es verschiedene Studien, die nicht unbedingt alle – z. T. aufgrund regulatorischer Auflagen – den selben Endpunkt im Visier haben.
 - Die Studien weisen unterschiedliche Ein/Ausschlusskriterien auf, womit die Frage der externen Validität gestellt werden muss
 - manche Studien sind international, manche auch nur national, manche gar nicht publiziert, was bei der zusammenschauenden Bewertung die Frage nach dem publication bias aufwirft.

- Die Studienlandschaft bildet im Allgemeinen nur unzureichend die medizinische Versorgungspraxis ab, womit sich die Forderung nach weiteren (epidemiologischen) Studien ergibt.
- 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: die Methodik dieser Fächer ist geeignet, Licht in das Dunkel zu bringen. Das Zauberwort hierzu lautet: Transparenz. Diese erzeugen - qua definitionem – der Begriff des Studienprotokolls/der SOPs, das/die prospektiv und nachvollziehbar festlegen, wie mit unterschiedlichen Situationen umzugehen ist. In einem Studienprotokoll einer klinisch oder epidemiologischen Studie wird prospektiv festgelegt, wie 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 (nach Möglichkeit 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. 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.

4. Das zentrale Instrument der vergleichenden Nutzenbewertung von Arzneimitteln und anderen Behandlungsformen ist international das sog. Health Technology Assessment (HTA). Zu seiner Bearbeitung sind – auftragsbezogen, d.h. variabel und flexibel – diverse „Methoden und Kriterien“ (§ 35 Abs. 1 Satz 6) zu nutzen. Ihre Auswahl erfolgt auf der Grundlage von fachlich anerkannten internationalen

Standards der evidenzbasierten Medizin. Diese ist wenigen Grundprinzipien verpflichtet.

Zu diesen zeitlich stabilen Prinzipien gehören nach der prioritären Berücksichtigung der Prinzipien des Wohltuns und der Schadensvermeidung in der Patientenbehandlung („beneficence, nonmaleficence“ nach Beauchamps und Childress 4.1994) vor allem

- eine möglichst vollständige Vergegenwärtigung der thematisch einschlägigen Originalveröffentlichungen einschließlich ihrer Lücken und Grenzen,
- 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.

Dagegen sind die in § 35b Abs. 1 SGB V genannten Standards ein bewegliches Ziel; sie liegen nirgendwo in kodifizierter Form vor und verändern sich vergleichsweise rasch mit den Fortschritten der klinischen Wissenschaften, der klinischen Epidemiologie und Statistik 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 Nutznießer, Nutzenqualitäten, Nutzenausmaß, Eintrittswahrscheinlichkeit sowie zeitliche Aspekte. Ähnlich differenziert sind Merkmale von Schadensrisiken zu vergegenwärtigen, um Netto- und Zusatznutzen analysieren zu können. Die Evidenzanforderungen an die Schadensvermutung sind im Sinne des ethisch und rechtlich gebotenen Vorsorgeprinzips geringer als die an den Nutzennachweis.

Wirksamkeit und positiver Nutzen müssen in aller Regel durch sog. konfirmatorische Studien im Design der randomisierten klinischen Prüfung belegt werden. Die Untersuchung von Schadenspotentialen kann auf Studien anderen Zu-

schnitts kaum verzichten (Phase 4-Studien, Registerstudien, Anwendungsbeobachtungen).

Die vom IQWiG und G-BA genutzte Unterscheidung von (rein wissenschaftlichem und wertfreien) Assessment und umfassenden Appraisal im Versorgungskontext überzeugt nicht vollständig. Auch im Assessment sind zahlreiche Wertentscheidungen enthalten. Diese sind in jedem HTA offen zu legen und zu diskutieren. Dabei sind auch die Perspektiven, Interessen und Präferenzen verschiedener Gruppen sog. Stakeholder zu berücksichtigen. Ein umfassende und auch ethische Aspekte berücksichtigende Nutzenbewertung erfordert ein multi- und interdisziplinäres Vorgehen.

5. 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. Bei Durchführung einer gesundheitsökonomischen Bewertung ist die Definition der einzunehmenden Perspektive von wesentlicher 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 Ausgabenbereiche jenseits der GKV zu berücksichtigen, sofern nicht unbestritten solche Auswirkungen irrelevant sind. Der geeignete Ort darüber zu entscheiden, ob solche Auswirkungen bereits apriori als irrelevant eingeschätzt werden können, ist ein beim G-BA im Kontext der Auftragsvergabe anzusiedelnder Scoping-Workshop.

Im Gegensatz zur Beurteilung des isolierten medizinischen Nutzens, ist es das Ziel gesundheitsökonomischer Studien, deren Ergebnisse in Erstattungsentscheidungen einfließen, die Routine-Versorgung der Patienten abzubilden. Auch entspricht es dem internationalen gesundheitsökonomischen Standard, Nutzen (und Kosten) grundsätzlich über einen so langen Zeitraum untersucht werden, dass sämtliche für den Alternativenvergleich relevanten – auch langfristigen – Effekte einbezogen werden. Dies bedingt oftmals die Verwendung weiterer Daten als in einer isolierten Nutzenbewertung und ihrer Verknüpfung mittels entscheidungsanalytischer Modellierungen. An Modelle sind strenge Anforderungen zu stellen.

Kostendaten können im allgemeinen aus multinational angelegten RCTs nicht übernommen werden. Daher ist es in solchen Fällen unvermeidlich, die Kostendaten auch aus anderen Datenquellen zu generieren.

Es entspricht dem internationalen Standard in der Gesundheitsökonomie, dass künftige Kosten und Nutzen diskontiert werden. Hingegen bestehen in der internationalen Diskussion keine einheitlichen Vorstellungen darüber, welchen Betrag die Diskontrate annehmen soll.

Die Feststellung nach Gegenüberstellung von Zusatznutzen und Zusatzkosten, ob die Kostenübernahme durch die Versichertengemeinschaft bei der ermittelten Kosten-Nutzen-Relation „angemessen und zumutbar“ (§ 35b 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.

7. 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. Aufgrund 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 aufgegeben werden.

Inhaltsverzeichnis

Executive summary	2
Inhaltsverzeichnis	9
0. Vorbemerkung	11
1. Einführung: Neuregelungen durch das GKV-WSG.....	12
1.1. Verwendungszusammenhänge für die Kosten-Nutzen-Bewertung	13
1.2. Voraussetzungen für eine Nutzen- bzw. Kosten-Nutzen-Bewertung	14
1.3. Vergleichsinterventionen für Nutzen- und Kosten-Nutzen-Bewertung.....	15
1.4. Methoden des IQWiG bei der Nutzen- und Kosten-Nutzen-Bewertung.....	15
1.5. Beteiligungsrechte bei der Bewertung durch das IQWiG.....	16
2. Evidence Based Medicine	16
2.1 Die Entstehung des Begriffs Evidenzbasierte Medizin	17
2.2 EbM als integrierendes Konzept diverser Fachdisziplinen	18
2.3 Die Grundelemente der EbM-Konzepte	19
2.4 Die Rolle von EbM im Gesundheitssystem: Evidenzbasiertes Vorgehen.....	22
2.5 Anerkannte internationale Standards der evidenzbasierten Medizin – Gemeinsamkeiten und Grenzen	23
2.6 Fazit	28
3. Medizinische Biometrie und Epidemiologie	28
3.1. Datengrundlagen	30
3.2. Studientypen	30
3.3. Endpunkt	33
3.4. Bias	34
3.5. Interne und externe Validität.....	35
3.6. Zusammenfassung	36

4. Inhaltliche und methodische Aspekte der Nutzenbewertung nach § 35b Abs. 1 SGB V.....	37
4.1 Vorbemerkungen	37
4.2 „Nutzen“ – ein komplexer Begriff	39
4.3 Nutzenabschätzung – unter welchen Erhebungsbedingungen?.....	43
4.4 Indirekte Evidenz	44
4.5 Fehlende Evidenz.....	45
4.6 Offene und versteckte Werturteile	46
4.7 Die Ausformulierung der Forschungsfrage und die Beteiligung von Interessengruppen	47
4.8 Einzelfallentscheidungen oder Priorisierung.....	48
5. Gesundheitsbezogene Lebensqualität als Kriterium zur Nutzenbewertung und als Parameter in der gesundheitsökonomischen Kosten-Nutzwert-Analyse	48
5.1. Grundlagen.....	48
5.2. Gesundheitsbezogene Lebensqualität in Kosten-Nutzwert-Analysen	51
5.2.1. Besonderheiten und offene Fragen der Lebensqualitätsmessung für Kosten-Nutzwert-Analysen	53
6. Inhaltliche und Methodische Aspekte der Kosten-Nutzen-Bewertung	59
6.1. Die Perspektive einer gesundheitsökonomischen Evaluation	59
6.2. Nutzen in einer gesundheitsökonomischen Evaluation	63
6.3. Kosten in einer gesundheitsökonomischen Evaluation.....	68
6.4. Diskontierung von Kosten und Nutzen	71
6.5. Bildung einer Kosten-Effektivitäts-Relation.....	71
7. Anforderungen an den Bewertungsprozess	74
8. Literatur	76

0. Vorbemerkung

Das Bundesministerium für Gesundheit (BMG) hat eine Arbeitsgruppe von Wissenschaftlern mit einer kommentierenden Synopse der Fachpositionen zur Kosten-Nutzenbewertung für Arzneimittel beauftragt. Die kommentierende Synopse wird hiermit vorgelegt. Die Federführung für das Gesamtprojekt lag bei Jürgen Wasem.

Mit einer in kurzer Zeitperspektive erstellten kommentierten Synopse kann keine umfassende Behandlung sämtlicher Aspekte der Nutzen- und Kosten-Nutzen-Bewertung geleistet werden. Auch zielt das Papier nicht in erster Linie darauf, konkrete Umsetzungsfragen zu erörtern. Vielmehr fokussiert es auf einige grundsätzliche Aspekte.

Das Papier ist wie folgt aufgebaut: Ausgangspunkt für die Synopse sind die neuen gesetzlichen Regelungen des GKV-WSG (Abschn. 1; Federführung: Wasem). Es schließen sich Überlegungen zur evidenzbasierten Medizin an (Abschn. 2; Federführung: Antes). Es folgt ein Abschnitt über medizinische Biometrie und Epidemiologie (Abschn. 3; Federführung: Jöckel). Anschließend befassen wir uns mit inhaltlichen und methodischen Aspekte der Nutzenbewertung nach § 35b Abs. 1 SGB V (Abschn. 4; Federführung: Raspe). Anschließend stellen wir unsere Überlegungen zur gesundheitsbezogenen Lebensqualität als Kriterium zur Nutzenbewertung und als Parameter in der gesundheitsökonomischen Kosten-Nutzwert-Analyse vor (Abschn. 5; Federführung: Kohlmann). Des Weiteren befassen wir uns mit inhaltlichen und methodischen Aspekten der Kosten-Nutzen-Bewertung nach § 35b SGB V (Abschn. 6; Federführung: Wasem). Schließlich geht es um Anforderungen an den Prozess der Bewertung durch den Gemeinsamen Bundesausschuss und das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (Abschn. 7; Federführung: Wasem).

Das vorliegende Papier ist von den Autoren konsentiert worden, soweit nicht an einzelnen Stellen auf unterschiedliche Positionen hingewiesen wird.

1. Einführung: Neuregelungen durch das GKV-WSG

Ausgangspunkt für die vorliegende kommentierende Synopse sind die durch das Gesetz zur Stärkung des Wettbewerbs in der gesetzlichen Krankenversicherung (GKV-WSG)¹ vorgenommenen Änderungen im Recht der gesetzlichen Krankenversicherung. Insbesondere zwei Veränderungen sind für den vorliegenden Zusammenhang von Bedeutung:

- § 35b SGB V ist durch das GKV-WSG neu gefasst worden. Die Vorschrift ermächtigt den Gemeinsamen Bundesausschuss (G-BA) nunmehr, das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) nach § 139b Abs. 1 SGB V auch zu beauftragen, das Kosten-Nutzen-Verhältnis von Arzneimitteln zu bewerten. Solche Bewertungen können für jedes erstmals verordnungsfähige Arzneimittel mit patentgeschützten Wirkstoffen sowie für andere Arzneimittel, die von Bedeutung sind, erstellt werden. Die Vorschrift macht weitere Vorgaben über Gegenstand, Methoden und Beteiligungsrechte bei der Durchführung von solchen Bewertungen. Nähere Bestimmungen dazu werden auch durch Änderung von § 139a Abs. 4 u. 5 SGB V in die Regelungen über das IQWiG aufgenommen.
- In § 31 SGB V wird Absatz 2a eingefügt. Die Vorschrift regelt, dass die Spitzenverbände der Krankenkassen (bzw. ab 2008 der Spitzenverband Bund), für Arzneimittel, die nicht in eine Festbetragsgruppe nach § 35 SGB V einzu beziehen sind, einen Höchstbetrag festsetzen sollen, bis zu dem die Krankenkassen die Kosten tragen. Dieser Höchstbetrag ist auf Grund einer Bewertung nach § 35b Abs. 1 Satz 3 SGB V festzusetzen, sofern der Höchstbetrag nicht im Einvernehmen mit dem pharmazeutischen Unternehmer festgelegt wird. Die Vorschrift regelt, unter welchen Voraussetzungen die Festlegung eines Höchstbetrages nicht stattfinden darf und nennt die Voraussetzungen, die vorliegen müssen, damit eine Kosten-Nutzen-Bewertung durchgeführt werden kann.

¹ Gesetz vom 26.03.2007, Bundesgesetzblatt I S. 378.

Bereits seit dem GKV-Modernisierungsgesetz von 2003² konnte der G-BA das IQWiG mit Analysen zum Nutzen und zur Wirtschaftlichkeit der im Rahmen der gesetzlichen Krankenversicherung erbrachten Leistungen beauftragen. Für Arzneimittel war allerdings aufgrund einer Spezialvorschrift abweichend von dieser generellen Regelung nur eine Nutzen-Bewertung zulässig³. Seit 1.4.07 ist nunmehr auch für Arzneimittel eine Kosten-Nutzen-Bewertung vorgesehen.

Der Gesetzgeber geht hierbei davon aus, dass die bisher bereits mögliche isolierte Nutzenbewertung bestehen bleiben soll und zusätzlich die Kosten-Nutzen-Bewertung eingeführt wird. So spricht er in § 35b Abs. 1 Satz 1 SGB V davon, dass das IQWiG beauftragt werden könne, „den Nutzen oder das Kosten-Nutzen-Verhältnis“ von Arzneimitteln zu bewerten. In der Gesetzesbegründung des GKV-WSG ist von einer „Erweiterung“ des Auftrages des IQWiG auch zu einer Kosten-Nutzen-Bewertung die Rede.⁴ Auch spricht er davon, dass bestimmte Verfahrensregelungen auch „für bereits begonnene Nutzenbewertungen“ gelten sollen⁵, was ebenfalls darauf hindeutet, dass diese Form der Bewertung neben der Kosten-Nutzen-Bewertung weiterhin möglich bleibt. Hieran ändert auch der – missverständliche – Wortlaut der Vorgabe in Absatz 1 Satz 3 nichts, welcher lautet „Die Bewertung erfolgt durch Vergleich mit anderen Arzneimitteln und Behandlungsformen unter Berücksichtigung des therapeutischen Zusatznutzens für die Patienten im Verhältnis zu den Kosten.“ Die Vorgabe einer vergleichenden Bewertung mit Therapiealternativen gilt für sämtliche in § 35b Abs. 1 Satz 1 u. 2 angesprochenen Bewertungen, also auch für isolierte Nutzenbewertungen. Jedoch ist eine Bewertung des therapeutischen Zusatznutzens im Verhältnis zu den Kosten nur bei einer Kosten-Nutzenbewertung relevant.

1.1. Verwendungszusammenhänge für die Kosten-Nutzen-Bewertung

Die vom Gesetz nunmehr vorgesehene Kosten-Nutzen-Bewertung von Arzneimitteln findet in zwei möglichen Zusammenhängen Verwendung:

- Kosten-Nutzen-Bewertungen sind Grundlage für die Festsetzung von Höchstbeträgen nach § 31 Abs. 2a SGB V für Arzneimittel, die nicht in eine

² Gesetz vom 14.11.2003, BGBl. I S. 2190.

³ § 35b sowie § 139a Abs. 3 Nr. 5 SGB V in der bis 31.03.2007 geltenden Fassung.

⁴ Vgl. BT-Drucksache 16/3100, Besondere Begründung zu Nr. 20 (§ 35b).

⁵ § 35b Abs. 1 Satz 8 SGB V.

Festbetragsgruppe nach § 35 SGB V einzubeziehen sind, sofern die Arzneimittel in das Höchstbetragssystem einzubeziehen sind und sofern die Festsetzung des Höchstbetrages nicht im Einvernehmen mit dem pharmazeutischen Unternehmer erfolgt. Hier ist Adressat einer Bewertung der Spitzenverband Bund der Krankenkassen.

- Kosten-Nutzen-Bewertungen sind darüber hinaus eine Entscheidungsgrundlage für den G-BA bei den Arzneimittelrichtlinien. § 35b Abs. 2 SGB V sieht vor, dass die Bewertungen dem G-BA „als Empfehlung zur Beschlussfassung“ für die Arzneimittelrichtlinien nach § 92 Abs. 1 Satz 2 Nr. 6 SGB V dienen sollen. Insbesondere können sie eine Grundlage für einen Beschluss über Leistungseinschränkungen oder Leistungsausschlüsse wegen fehlender Wirtschaftlichkeit eines Arzneimittels nach § 92 Abs. 1 Halbsatz 3 SGB V darstellen.⁶

1.2. Voraussetzungen für eine Nutzen- bzw. Kosten-Nutzen-Bewertung

Voraussetzung für eine Nutzen-Bewertung und für eine Kosten-Nutzen-Bewertung ist jeweils ein entsprechender Auftrag an das IQWiG durch den gemeinsamen Bundesausschuss. Unter den Rahmenbedingungen des § 139b Abs. 2 SGB V kommt auch das BMG als Auftraggeber in Betracht.

Voraussetzung für einen Auftrag an das IQWiG ist, dass es sich bei den Arzneimitteln, die der Bewertung unterzogen werden sollen, entweder um patentgeschützte Arzneimittel oder um „andere Arzneimittel, die von Bedeutung sind“, handelt.⁷ Hierbei wird insbesondere an Arzneimittel mit einem hohen Ausgabenvolumen für die gesetzliche Krankenversicherung zu denken sein.

Sofern Auftragszweck die Festsetzung eines Höchstbetrages ist, ist geregelt, dass eine Kosten-Nutzen-Bewertung erst erstellt werden kann, wenn hinreichende Erkenntnisse über die Wirksamkeit des Arzneimittels nach den Grundsätzen der evidenzbasierten Medizin vorliegen können.⁸ Dies bedeutet nicht, dass diese Erkenntnisse real vorliegen müssen, vielmehr reicht die Möglichkeit hierzu aus („können“). Auch legt der hier verwendete Begriff der Wirksamkeit mit seiner Konnotation mit dem Arzneimittelgesetzes nahe, dass klinische Studien aus dem Zulassungsverfahren, sofern

⁶ Vgl. auch BT-Drucksache 16/3100, Besondere Begründung zu Nr. 20 (§ 35b).

⁷ § 35b Abs. 1 Satz 2 SGB V.

⁸ § 31 Abs. 2a Satz 6 SGB V.

sie geeignete Daten enthalten, für eine Kosten-Nutzenbewertung herangezogen werden. Für den Verwendungszweck der Beschränkung oder des Ausschlusses von Leistungen durch die Arzneimittelrichtlinien des G-BA ist diese Vorgabe nicht vorgesehen.

1.3. Vergleichsinterventionen für Nutzen- und Kosten-Nutzen-Bewertung

Wie schon angesprochen, sind nach § 35b Abs. 1 Satz 3 Vergleichsintervention für eine Nutzenbewertung wie auch für eine Kosten-Nutzen-Bewertung sowohl andere Arzneimittel als auch andere Behandlungsformen, also nicht-medikamentöse Therapien. Die (medikamentöse oder nicht-medikamentöse) Behandlung, mit der das zu untersuchende Arzneimittel verglichen wird, muss für die Versorgungspraxis in der GKV relevant sein, da ansonsten kein „therapeutischer Zusatznutzen“ für die Patienten, der durch die Nutzenbewertung ermittelt (und bei der Kosten-Nutzen-Bewertung zudem in das Verhältnis zu den Kosten für die Versichertengemeinschaft gesetzt) werden soll, resultiert.

1.4. Methoden des IQWiG bei der Nutzen- und Kosten-Nutzen-Bewertung

Das IQWiG gibt sich gem. § 35b Abs. 1 Satz 5 SGB V seine Methoden „auftragsbezogen“ selber. Der Gesetzgeber des GKV-WSG führt hierbei nunmehr allerdings insoweit eine Bindung des Institutes ein, indem er vorschreibt, dass die Methoden „auf der Grundlage der in den jeweiligen Fachkreisen anerkannten internationalen Standards der evidenzbasierten Medizin und der Gesundheitsökonomie“ bestimmt werden müssen. In § 139a Abs. 4 SGB V wird darüber hinaus präzisiert: „Das Institut hat zu gewährleisten, dass die Bewertung des medizinischen Nutzens nach den international anerkannten Standards der evidenzbasierten Medizin und die ökonomische Bewertung nach den hierfür maßgeblichen international anerkannten Standards, insbesondere der Gesundheitsökonomie erfolgt.“ Für die ökonomische Bewertung sind neben den Standards der Gesundheitsökonomie offenbar auch andere international anerkannte Standards relevant.

1.5. Beteiligungsrechte bei der Bewertung durch das IQWiG

Das GKV-WSG verstärkt die Beteiligungsrechte der betroffenen Kreise sowie von Sachverständigen bei der Festlegung der Methoden und der Durchführung der Bewertungen durch das IQWiG. § 35b Abs. 1 Satz 6 SGB V verpflichtet das Institut zu einer „hohen Verfahrenstransparenz“ und einer „angemessenen Beteiligung der in § 35 Abs. 2 und § 139a Abs. 5 Genannten“ – dies sind Sachverständige der medizinischen, pharmazeutischen und gesundheitsökonomischen Wissenschaft und Praxis, die Arzneimittelhersteller und die Berufsvertretungen der Apotheker, bei der Beurteilung von Arzneimitteln der besonderen Therapierichtungen auch Sachverständige dieser Therapierichtungen, sowie die für die Wahrnehmung der Interessen der Patientinnen und Patienten und der Selbsthilfe chronisch Kranker und behinderter Menschen maßgeblichen Organisationen sowie die oder der Beauftragte der Bundesregierung für die Belange der Patientinnen und Patienten. Der Ausschussbericht weist zudem darauf hin, dass die Nennung dieser Gruppen Beteiligter die Hinzuziehung zusätzlicher Sachverständiger betroffener anderer Bereiche und Professionen nicht ausschließt.⁹ Der Begründung des Gesetzentwurfes ist darüber hinaus zu entnehmen, dass „insbesondere auch... international anerkannte Sachverständige“ in das Beteiligungsverfahren eingebunden werden sollen.¹⁰

Den zu Beteiligten ist in allen wichtigen Abschnitten des Bewertungsverfahrens Gelegenheit zur Stellungnahme zu geben. Die Stellungnahmen sind in die Entscheidungen des Instituts einzubeziehen, es hat sich also mit ihnen bei der Abfassung des Berichtsplanes und der Zwischen- und Endberichte auseinander zu setzen und dies angemessen zu dokumentieren.

2. Evidence Based Medicine

Das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) soll für die Evaluation des Nutzens wie auch des Kosten-Nutzen-Verhältnisses von Arzneimitteln auftragsbezogen Methoden und Kriterien für die Erarbeitung von Bewertungen auf der Grundlage der in den jeweiligen Fachkreisen anerkannten internationa-

⁹ Vgl. BT-Drucksache 16/4247, Besonderer Teil, zu Artikel 1, zu Nr. 117 (§ 139a).

¹⁰ Vgl. BT-Drucksache 16/3100, Besondere Begründung zu Nr. 20 (§ 35b).

len Standards der evidenzbasierten Medizin (EbM) und der Gesundheitsökonomie (§35 SGB V) bestimmen.

Ob es einen solchen Standard gibt und ob – was für das weitere Vorgehen von grundsätzlicher Bedeutung ist – dieser Standard zu einer eindeutigen Operationalisierung führt, wird im folgenden durch einen Blick auf die Entstehung des Begriffs EbM und auf die gegenwärtige damit verbundene heutige Begriffswelt untersucht. Diese Betrachtung ist auch angesichts der häufigen Kontroversen der letzten 10 Jahre notwendig und sinnvoll, da viele der Auseinandersetzungen auf Missverständnissen aufgrund inkonsistenter Benutzung von Schlagworten dieses Themenbereichs beruhen.

Die vorliegende Beschreibung und Diskussion bezieht sich auftragsgemäß mit der Arzneimitteltherapie nur auf einen Ausschnitt der EbM, die sich seit Entstehung der Konzepte auf Diagnostik, Therapie und Prognose erstreckt..

2.1 Die Entstehung des Begriffs Evidenzbasierte Medizin

Während die wissenschaftstheoretischen Grundlagen der EbM in manchen Darstellungen mit dem Frankreich des 19. Jahrhunderts in Verbindung gebracht werden, wird der Begriff in seiner heutigen Form und Bedeutung zum ersten Mal 1991 in einer nordamerikanischen Zeitschrift [ACP Journal Club 1991] eingeführt.

Die moderne Entwicklung bekam die entscheidenden Impulse durch eine Arbeitsgruppe an der McMaster Universität in Hamilton in Kanada, die Ende der 70er Jahre unter der Leitung von David Sackett begann, Konzepte zu entwickeln, um Entscheidungen im klinischen Alltag in expliziter Form auf die Aussagen von klinischen Studien (im Sinne von patientenorientierten Studien, also Studien an Patienten) zu gründen. Öffentlich wurde diese Arbeit durch eine 1981 beginnende Artikelserie im Canadian Medical Association Journal, in der unter dem Titel *Critical Appraisal* die Regeln für die Nutzung von Evidenz dargestellt wurden. Aus dem Vorschlag, die ersten Ansätze unter *Bringing Cricritical Appraisal to the Bedside* zu integrieren, entwickelte sich ein Gesamtkonzept für die medizinische Praxis, das Wissen und die medizinische Literatur als wesentlichen Bestandteil von klinischen Entscheidungen forderte. Wegen der unter den Fachkollegen verbreiteten Ablehnung dieses Ansatzes wurde die damalige Idee, dafür als Oberbegriff *Scientific Medicine* einzuführen, ver-

worfen. Statt dessen wurde der Begriff *evidence-based medicine* vorgeschlagen, der sich durchsetzen ließ und seit dem die Gesundheitsforschung und –versorgung sowie benachbarte Gebiete in beispielloser Form durchdrungen hat. Bemerkenswert ist, wie ein im universitären Umfeld geprägter Begriff seinen Weg in Institutionen des Gesundheitssystems, in die Medien und zu den Patienten gefunden hat. Darin liegt jedoch vermutlich auch der Grund für die Vielfalt der Interpretationen von EbM und die damit verbundenen anhaltenden Kontroversen.

2.2 EbM als integrierendes Konzept diverser Fachdisziplinen

EbM ist ein integrierendes Gesamtkonzept mit Methoden und Techniken, die einzeln in unterschiedlichen Fachdisziplinen entwickelt und angewendet wurden. Das übergeordnete Ziel ist, ärztliches Entscheiden und Handeln mit expliziten Regeln enger an die Ergebnisse patientenrelevanter Studien binden.

Die Entwicklung hatte seit Beginn eine didaktische Komponente, was sich u. a. darin ausdrückt, dass das von Sackett et al. erstmalig und nun bereits in 3. Auflage erschienene Standardwerk *Evidence-based Medicine* [Straus et al. 2005] jeweils den Untertitel *How to Practice and Teach EbM* trägt. Hintergrund ist die enge Verbindung von EbM-gemäßigem Vorgehen mit dem problemorientierten Lernen.

Methodisch hat die EbM sich seit Beginn der Entwicklung bei den Bibliotheks- und Kommunikationswissenschaften, der medizinischen Statistik, der Biometrie, der Epidemiologie und der Informatik und weiteren einerseits bedient und andererseits Impulse für deren Weiterentwicklung gegeben. Als Vorläufer und parallel zur EbM hat sich die Klinische Epidemiologie als methodische „Grundlagenwissenschaft“ (*Clinical epidemiology: a basic science for clinical medicine* [Sackett 1985]) zur EbM entwickelt. Sie definiert sich als Wissenschaft zur Untersuchung der Variabilität von Krankheitsverläufen und ihrer Abhängigkeit von diagnostischen und therapeutischen Maßnahmen [Weiss (2006). *Clinical Epidemiology* (3rd ed.). The Study of the Outcome of Illness]. Entscheidend ist hier die besondere Perspektive auf das klinische Problem, wobei vielfach auf Techniken aus der Epidemiologie, der Medizinstatistik und der Biometrie zurückgegriffen wird.

Die nahe Verwandtschaft von EbM und Klinischer Epidemiologie zeigt sich auch darin, dass neben dem wohl bekanntesten Lehrbuch von Fletcher et al. (2005) ein weite-

res relevantes Lehrbuch von den Autoren stammt, die die EbM in Kanada vorbereitet und eingeführt haben [Haynes et al. (2005)]. Fletcher & Fletcher definieren Klinische Epidemiologie als die Analyse der Bestimmungsgrößen und Auswirkungen klinischer Entscheidungen, während Haynes et al. im Titel die Erforschung der klinischen Praxis betonen. Die unmittelbare Nähe zur EbM zeigt sich in diesen Zielen und in den Daten des erstmaligen Erscheinens dieser Bücher von 1982 bis 1986 (letzte Auflagen 2005-2006). Diese die moderne Entwicklung beschreibenden Lehrbücher der klinischen Epidemiologie systematisieren die Entwicklung, die bereits in den 60er Jahren mit der Einrichtung entsprechender Lehrstühle in Buffalo und McMaster begann und untrennbar mit der Rolle von Alvan Feinstein [Feinstein 1984] verbunden ist. Ohne Berücksichtigung der Entwicklung der Klinischen Epidemiologie [Sackett (2002)] ist ein Verständnis der EbM kaum möglich.

Bezüglich des Auftrags an G-BA und IQWiG, Kosten-Nutzen-Bewertungen durchzuführen, ist bemerkenswert und festzuhalten, dass in der oben beschriebenen Entwicklung gesundheitsökonomische Aspekte nicht auftauchen. Die Aktivitäten konzentrieren sich ausdrücklich auf das Arzt-Patienten-Verhältnis und die Optimierung der ärztlichen bzw. gemeinsamen Entscheidungen zum Wohle des Patienten. Diese bewusst enge klinische Fokussierung, die in Deutschland in Bezug zur Entstehungsgeschichte von manchen als Sackett-EbM bezeichnet wird, ist bereits Mitte der neunziger Jahre scharf wegen des mangelnden Systembezugs kritisiert worden, bezeichnenderweise auch von Autoren aus dem gesundheitsökonomischen Umfeld (z. B. Maynard 1997).

2.3 Die Grundelemente der EbM-Konzepte

Die EbM, wie sie inzwischen in unzähligen Artikeln und Büchern beschrieben und diskutiert worden ist, hat im Kern eine einfache Struktur. Entscheidungen für ein therapeutisches Vorgehen sollen drei Komponenten vereinigen:

- die Ergebnisse patientenorientierter klinischer Studien, die als externe Evidenz bezeichnet wird,
- die Erfahrung und Kompetenz des Arztes, die als interne Evidenz bezeichnet wird,
- die Werte und Präferenzen von Patienten und Angehörigen.

Als Definition liest man oft den in diesem Zusammenhang wohl meistzitierten Satz aus einem Editorial von Sackett et al. (1997), nach dem EbM der gewissenhaft, ausdrückliche und vernünftige Gebrauch der gegenwärtig besten externen, wissenschaftlichen Evidenz für Entscheidungen in der medizinischen Versorgung individueller Patienten ist. Diese Aussage ist jedoch eher eine nicht sehr präzise formulierte Wunschvorstellung denn eine Definition. Die für die Umsetzung notwendige Spezifizierung bzw. Operationalisierung für die Berücksichtigung der externen Evidenz hat zu folgendem 5-Schritt-Schema geführt, das heute in diesem Feld als Arbeitsschema akzeptiert ist:

1. Formulierung des klinischen Problems in eine operationalisierte, suchtaugliche Frage,
2. Systematische Identifikation von Evidenz in der relevanten Literatur,
3. Kritische Beurteilung und Qualitätsbewertung der aufgefundenen Quellen nach Prinzipien der klinischen Epidemiologie,
4. Anwendung der Evidenz in Einklang mit der klinischen Erfahrung und der durch den Patienten gegebenen Situation,
5. Evaluation des Erfolgs der durchgeführten Maßnahme.

Diese Strukturierung bildet weitgehend einheitlich den Ausgangspunkt für evidenzbasiertes Vorgehen, wird jedoch immer wieder verkürzt dargestellt. Zwei Aspekte sind dabei besonders zu betonen:

- EbM stützt sich nicht nur auf die äußere Evidenz, sondern ausdrücklich auch auf die ärztliche Erfahrung und Kompetenz.
- EbM leitet die äußere Evidenz nicht nur aus randomisierten kontrollierten Studien (RCT) ab, sondern stützt sich im ersten Schritt auf die gesamte relevante Literatur. Die Einschränkung auf bestimmte Studientypen bzw. Studiendesigns orientiert sich an der Robustheit gegenüber möglichen Bias-Einflüssen und muss vorab im Studienprotokoll festgelegt werden.

Die einzelnen Schritte beinhalten einerseits weit verbreitete und benutzte Kernfunktionen, andererseits aber auch stark divergierende Details, wie in hunderten von Artikeln und etlichen Lehrbüchern beschrieben wird. So wird für den ersten Schritt das sogenannte PICO-Schema (Patients, Intervention, Comparison, Outcome) weitgehend einheitlich zugrunde gelegt, während im zweiten Schritt die Literatursuche große Unterschiede aufweist. Die Beschränkung auf den englischen Sprachraum sowie

die Auswahl der einbezogenen Datenquellen können erhebliche Auswirkungen auf die Ergebnisse haben. Der methodisch schwierigste und am wenigsten konsensfähige Schritt ist der dritte, da es dort weder eine international gültige Vereinbarung bzgl. der Qualitätskriterien innerhalb eines Studiendesigns gibt noch eine allgemein konsentierende Regel, welche Studientypen (Designs) als Evidenz herangezogen werden sollen.

Die zu recht häufig kritisierte statische Orientierung an Evidenzstufen (z. B. Centre for Evidence-Based Medicine 2001), deren einzige Grundlage das Studiendesign ist, wird zunehmend ergänzt und auch verdrängt durch das GRADE-System [Guyatt 2007, Kunz 2007], das eine Auf- oder Abwertung von Studien anhand von Qualitätsmerkmalen zulässt und damit auch die Qualität von Durchführung und Auswertung von Studien berücksichtigt.

Die Gewichtung von äußerer und innerer Evidenz und von Patientenpräferenzen ist ein von vielen Einflüssen abhängiger Prozess, der ausgestaltet werden muss. Wie dies aussehen muss, folgt nicht eindeutig aus den konzeptionellen Vorgaben der EbM. Das größte Konfliktpotential liegt hier in der oft unüberwindlichen Trennung zwischen den unterschiedlichen Welten der von außen oktroyierten externen Evidenz und der die internen Evidenz bestimmenden ärztlichen Praxis.

Die grundsätzliche Einbeziehung der gesamten relevanten Literatur (also nicht nur von RCTs) wird relativiert durch die unterschiedliche Aussagekraft (Validität) von einzelnen Studientypen. Die Validität wird bestimmt durch die Anfälligkeit für systematische Fehler (Bias) in den Ergebnissen. Die Methodik der klinischen Epidemiologie macht zwar Vorgaben, nach denen die Validität einzelner Studientypen hierarchisch angeordnet werden (*levels of evidence*), diese stehen jedoch in regelmäßiger Kritik. Teils werden sie aus prinzipiellen Gründen nicht anerkannt, häufiger ist jedoch das Fehlen von hochwertigen Studien und die damit verbundenen Frage, wieweit in der Hierarchie herunterzugehen ist. Ebenso wird kontrovers diskutiert, ob und wieweit bei Vorhandensein von Studien der obersten Stufe auch Studien niedriger Stufen eingeschlossen werden sollen. Für die Interpretation von fehlender oder unzureichender Evidenz gibt es von wissenschaftlicher Seite wenig Unterstützung. Die wissenschaftlich begründete Feststellung, in solchen Fällen nur geringe Aussagesicherheit zu erreichen, führt zwangsläufig zur stärkeren Berücksichtigung anderer Kriterien. Erwartungen und Einschätzungen sowie Finanzierungsquellen können eine beträchtliche

Biasquelle sein, wie in den letzten Jahren in einer Reihe empirischer Arbeiten gezeigt wurde [z. B. Bekelman et al. 2003]. Erschwerend kommt hinzu, dass die Synthese von Ergebnissen aus Studien unterschiedlichen Designs methodisch ungelöst ist. Damit bleibt nur die qualitative, nebeneinanderstehende Darstellung der Ergebnisse, was erheblich mehr Interpretationsspielraum lässt als eine methodisch konsentrierte Synthese.

Aus diesen Gründen ist es unausweichlich, dass die Frage der einzuschließenden Studien mit das größte Konfliktpotential birgt. Die EbM liefert dafür keine Ausführungsvorschriften, sondern nur Prinzipien. Diese sollten bei der Methodenwahl des IQWiG berücksichtigt werden, von ihnen kann jedoch keinesfalls ein Automatismus für evidenzbasierte Methoden erwartet werden.

2.4 Die Rolle von EbM im Gesundheitssystem: Evidenzbasiertes Vorgehen

Nach der oben beschriebenen Entwicklung der auf das Arzt-Patienten-Verhältnis fokussierten EbM wurden deren Elemente sehr rasch von anderen Gruppen und Institutionen der Gesundheitssysteme aufgegriffen, oft wegen der engen Fokussierung [Maynard 1997] kritisiert und entsprechend erweitert. Als Oberbegriff der weiteren Sichtweise hat sich *Evidence-based Health Care* etabliert [Gray, M. (2001)], unter dem sich eine vielfältige Begriffs- und Konzeptwelt subsumiert, die die Vielfalt unter dem Begriff EbM noch bei weitem übertrifft.

Eine kleine Auswahl zeigt die Erweiterung der Begriffswelt (evidenzbasiert = eb): eb practice, eb decision making, eb management, eb policy making, eb patient choice etc. Speziell in den angelsächsischen Ländern ist die Forderung nach evidenzbasiertem Vorgehen inzwischen längst über die Medizin hinausgewachsen und hat viele andere Fachgebiete erfasst. Dass es unterschiedliche Realisierungen dieser Konzepte gibt, zeigt z. B. Olsson (2007). Auch für die Politik gibt es Bestrebungen, Entscheidungen durch die Bereitstellung der relevanten Evidenz vorzubereiten und diese Abläufe zum Gegenstand wissenschaftlicher Analyse zu machen (z. B. Zeitschrift *Evidence and Policy*).

Bei manchen der in vielen Ländern inzwischen aufgebauten Agenturen, die Evidenz systematisch für die Entscheidungsfindung aufbereiten, findet sich auch der Begriff *evidence-based practice* oder auch *best practice*. Die Agency for Healthcare Research and Quality (AHRQ) der USA fördert in Nordamerika 13 Evidence-Based

Practice Centres und bietet dazu ein breites Spektrum an methodischen Aussagen zum evidenzbasierten Vorgehen, das als ein Modell für die Spezifizierung und die Umsetzung der Prinzipien der EbM genutzt werden kann.

2.5 Anerkannte internationale Standards der evidenzbasierten Medizin – Gemeinsamkeiten und Grenzen

Die Forderung des Gesetzgebers an die Arbeit des IQWiG, Methoden auf der Basis der anerkannten internationalen Standards der evidenzbasierten Medizin zu bestimmen und anzuwenden, stößt auf eine wissenschaftliche Welt, die diese Methoden durch theoretische und empirische Arbeiten und im wissenschaftlichen Diskurs stetig weiterentwickelt. Ohne hier auf die Bedeutung des Begriffs „Standard“ weiter einzugehen, ist festzustellen, dass es einen solchen Standard in kodifizierter Form nicht gibt und auch in Zukunft nicht geben wird. Die Gründe liegen einerseits im Wesen der Wissenschaft selbst, andererseits aber in der EbM auch gerade in den integralen Bestandteilen Interne Evidenz und Werte und Präferenzen. Als akzeptierter Standard könnten die in Abschn. 2.3 beschriebenen Grundlagen der EbM angesehen werden, die allerdings selbst für einen kurzen Zeitraum keinesfalls zwingend zu einem eindeutigen Methodenspektrum für eine Institution wie das IQWiG führen, sondern eine Fülle von Interpretationen und Realisierungen dieser Prinzipien zulassen. Dies sollte nicht als Defizit angesehen werden, sondern ist gerade die Produktivkraft von Wissenschaft und muss so akzeptiert werden, auch wenn es mit dem Wunsch nach festen Vorgaben und Regeln zu kollidieren scheint.

Ein Weg, Standards festzustellen, besteht im Blick auf die etablierten Agenturen in Ländern mit längerjähriger EbM-Orientierung in der Gesundheitsversorgung, wie z. B. NICE in UK, AHRQ in USA oder NICS in Australien. Weitere Quellen bieten die Vorgaben der Organisationen, die evidenzbasierte Leitlinien erstellen, sowie Organisationen wie die Cochrane Collaboration, die eine Vielzahl von methodischen Arbeiten insbesondere zur Bias-Kontrolle liefern.

Daraus und orientiert an den EbM-Prinzipien von Abschn. 1.3 (und unter Nutzenaspekt diskutiert in Abschn. 3.) lassen sich – bei allen Widersprüchen – folgende Eckpfeiler eines evidenzbasierten Vorgehens festhalten:

1. Das 5-Schritt-Schemas aus Abschn. 2.3 liegt in entsprechend modifizierter Form allem Eb-gerechtem Vorgehen zugrunde. Für Institutionen wie das IQWiG werden die letzten beiden Schritte, die die klinische Anwendung betreffen, ersetzt durch 4. Synthese der relevanten Evidenz und 5. Interpretation der Ergebnisse.

2. Schritt 1 (Formulierung der Fragestellung) baut auf die vier Elemente des PICO-Prinzips (siehe Abschn. 2.3). Unverzichtbar erscheint hier die multidisziplinäre Zusammenarbeit von Klinikern der entsprechenden Fachrichtung und Methodikern. Zur Frage, wieweit hier Interessengruppen einbezogen werden, geben die EbM-Konzepte keine Auskunft. Etabliert und unverzichtbar ist jedoch die Forderung der Offenlegung von potentiellen Interessenkonflikten seitens aller Beteiligter. Die praktische Gestaltung (z. B. Scoping Workshops) unterliegt vielen Einflüssen, die aus Sicht der EbM-Prinzipien nicht spezifiziert werden.

Unverzichtbar und damit Standard ist aus EbM-Sicht der Vergleich mit einer geeigneten Kontrollgruppe und die Forderung nach patientenbezogenen Zielvariablen als Outcome.

3. Schritt 2 (Literatursuche) kann als allgemeiner Bestandteil eines EbM-gerechten Vorgehens heute als Standard angesehen werden. Allerdings ist auch hier die Realisierung von erheblichen Unterschieden geprägt. Unterschiedlich gesehen wird hier z. B., ob und wieweit über die englischsprachige Literatur hinausgegangen werden muss, welche Datenbanken einbezogen werden sollen und welcher beliebig erweiterbare Aufwand notwendig ist.

4. Schritt 3 (kritische Bewertung der Artikel) kann als Bestandteil des EbM-Vorgehens heute ebenfalls als Standard angesehen werden, allerdings gibt es hier die größten Divergenzen sowohl bzgl. des praktischen Vorgehens wie auch der prinzipiellen Sichtweisen. Zum einen gibt es keine konsentierte Qualitätsskala für die Bewertung selbst für Studien einheitlichen Designs, sondern nur eine Reihe von Qualitätskriterien, die für die Qualitätsbewertung in unterschiedlicher Auswahl und Gewichtung angewendet werden (adäquate Kontrollgruppe, Verblindung, Randomisierung, Geheimhaltung der Behandlungsreihenfolge etc.). Einheitliche Messlatte ist (bzw. sollte sein) das Potential der einzelnen Maßnahmen, systematische Verzerrungen zu minimieren, oder in moderner Diktion: die Fähigkeit zum Schutz vor Bias (Bias-Protektion).

Noch weniger konsensfähig ist die Frage der Einbeziehung unterschiedlicher Studientypen. Die randomisierte kontrollierte Studie wird als Basis des Wirksamkeitsnachweises weithin gefordert und anerkannt, während gleichzeitig ihre Begrenztheit bzgl. Nebenwirkungserfassung außer Frage steht. Bezüglich der Frage des Wirksamkeitsnachweises durch andere Studientypen (insbesondere Nachweis von Kausalzusammenhängen) gibt es unübersehbare Gegensätze, die an dieser Stelle sehr deutlich festgehalten werden sollen. Für die Diskussion RCT vs. Beobachtungsstudien sind die allgemeinen Aussagen in vielen wissenschaftlichen Artikeln ausführlich diskutiert (siehe z. B. Pocock 2000). Für konkrete Bewertungen sind an RCTs und andere Studien wissenschaftlich begründete Qualitätskriterien anzulegen, unter denen vor allem kleine Beobachtungsstudien oder Fallserien unter unkontrollierten Bedingungen nur selten Berücksichtigung finden dürften. Qualitätsansprüche an Studien nehmen in der EbM und in den methodischen Wissenschaften eine zentrale Rolle ein und sind unverzichtbar.

Vermutlich kann ein Teil der Divergenzen durch präzisere Formulierungen und Argumentationen aufgelöst werden, da viele der gegenwärtigen Diskussion sich durch große Oberflächlichkeit auszeichnen – entweder aufgrund mangelnder Kenntnisse oder aber auch zur Verdeckung darunterliegender anderweitiger Interessen. Für einen beträchtlichen Teil gilt das jedoch nicht, da viele wissenschaftliche Fragen nicht ausreichend gelöst sind.

Als zugrunde liegenden Standard könnte man hier die rigorose Orientierung an Forderungen nach der Kontrolle von systematischen und zufälligen Fehlern ansehen. Diese Sichtweise hat in den letzten 10 Jahren einen enormen Aufschwung erfahren, ohne dass man diesbezüglich heute jedoch von weitgehender Akzeptanz oder entsprechender Praxis sprechen kann.

5. Schritt 4 (Synthese der eingeschlossenen Evidenz) ist als Teil des Gesamtkonzepts grundsätzlich akzeptiert und weitverbreitete Praxis. In moderner Terminologie kann dieser Schritt als Teil einer systematischen Übersichtsarbeit (Systematic Review, SR) [Antes et al. 2006, Egger et al. 2001, Khan et al. 2003] qualitativ oder quantitativ erfolgen. Die quantitative Zusammenfassung, die einen Parameter für alle Studien gemeinsam schätzt, wird als Metaanalyse (MA) bezeichnet. Dieser Begriff ist älter und wird auch heute noch völlig uneinheitlich sowohl für die quantitative Zusammenfassung wie auch für das Gesamtkonzept SR (Bezeichnung dafür oft auch

Literature Review) benutzt. Die oft heftigen Anfeindungen von MAs sind einerseits prinzipieller Natur und können – ohne hier in Spekulationen zu ergehen – als grundsätzliche Abneigung gegen die zunehmende Berücksichtigung von patientenorientierten Daten gedeutet werden. Andererseits gibt es viele Beispiele, wo eine MA durchgeführt wurde, obwohl man es besser bei einer qualitativen Zusammenfassung belassen hätte. Insofern ist die immer wieder aufkommende „Äpfel-und-Birnen“ – Kritik sehr differenziert zu betrachten.

Die Bewertung der Synthesemethoden innerhalb der MA fällt in die Methodenwissenschaften (Biometrie und Statistik) und findet meistens außerhalb der Diskussionen um EbM statt. In Methodikerkreisen gibt es dazu erhebliche Meinungsverschiedenheiten, so dass auch hier keinesfalls von standardisierter Methodik gesprochen werden kann. Die MA ist eine gewichtete Mittelwertsbildung über Studien. Die verschiedenen Methoden sind äquivalent zu unterschiedlicher Gewichtung des Einflusses einzelner Studien, womit die Methodenwahl unmittelbar Einfluss auf die Ergebnisse hat. Aus diesem Grund darf die Methodenwahl nicht datengetrieben sein, sondern sollte a priori (z. B. in einem Protokoll oder Analyseplan) festgelegt sein.

Eine zusätzliche Komplikation ist, dass die Synthesemethoden für RCTs wesentlich weiter entwickelt und robuster sind als für andere Studiendesigns. Damit hängt auch dieser scheinbar eher technische Aspekt mit dem umstrittenen Punkt zusammen, welche Studiendesigns in ein SR eingeschlossen werden sollten. Während für RCTs die Zusammenfassung der aggregierten Daten üblich ist und im allgemeinen als ausreichend angesehen wird, gilt für Beobachtungsstudien, dort für eine MA nach Möglichkeit auf die vollständigen Daten der einzelnen Studien zurückzugreifen, also auf die individuellen Daten der Studienteilnehmer. Dies wird zunehmend gemacht (auch für RCTs), ist allerdings zeitlich und logistisch so aufwendig, dass es in großem Umfang keine übliche Praxis werden kann. Damit ist auch hier eine markante Lücke festzustellen, die für jeden Einzelfall gefüllt werden muss. Auch hier kann also von einer Standardisierung nur im Sinne einer weiten Vorgabe gesprochen werden, die jeweils problem- und aufgabenorientiert gefüllt werden muss.

6. Schritt 5 (Interpretation der Ergebnisse) bedarf keiner besonderen Diskussion, da er Teil einer jeden wissenschaftlichen Arbeit - und damit Standard – ist. Auch hier ist allerdings zu betonen, dass es erhebliche Heterogenität bei der Interpretation von Ergebnissen gibt. Beispielhaft sei hier der Umgang mit unzureichenden Daten oder

mit nichtsignifikanten Ergebnissen genannt, die sich für die EbM in der Warnung „lack of evidence of effect is not evidence of lack of effect“ zusammenfassen lässt.

7. Für den Versuch, einen Standard zu definieren, gibt es von methodischer Seite als übergeordnetes Prinzip die Ausrichtung an maximaler Biaskontrolle und Kontrolle von Zufallsfehlern. Diese Forderungen sind zwar seit einigen Jahren in vieler Munde, werden jedoch in der Umsetzung eher als esoterische und vielleicht praxisferne Forderungen von Methodikern angesehen, wie etliche Artikel und Diskussionsveranstaltungen zeigen, wo in Methodendiskussionen der Biasbegriff gar nicht oder nicht adäquat auftaucht. Schutz vor Bias kann nicht nur als durchgängiges Prinzip (und damit als standardisierendes Fundament) für die Nutzenbewertung dienen, sondern sollte auch bei der Methodenwahl für die Kosten-Nutzen-Bewertung Berücksichtigung finden.

8. Zwei Prinzipien kann man heute unzweifelhaft als Standards bezeichnen: Die Forderung nach weitestgehender Transparenz und als eine Realisierung davon die Offenlegung potentieller Interessenkonflikte. Die Forderung nach Publikation von Studienprotokollen vor Studienbeginn, vorab festgelegte Analysepläne und deren Publikation dient einerseits zur Verhinderung der weitverbreiteten datengetriebenen (und vom Protokoll abweichenden) Analyse, Interpretation und Publikation von Studien, um daraus unvermeidbar folgende Biasquellen zu kontrollieren. Andererseits ist die transparente Darstellung des Vorgehens angesichts der großen Komplexität der Bewertungsarbeit unverzichtbar, da auch in den rein methodischen Teilen notwendigerweise eine Vielzahl von Einschätzungen und subjektiven Entscheidungen der Wissenschaftler einfließen, die nachvollziehbar dargestellt werden müssen (z. B. durch Auflistung der nicht berücksichtigten Studien mit entsprechender Begründung). Für die Einschätzung dieser subjektiven Anteile ist die Deklaration potentieller Interessenkonflikte für beteiligte Wissenschaftler von zentraler Bedeutung und heute Standard, um die Einschränkung der auf wissenschaftlicher Seite angenommenen Unabhängigkeit einschätzen zu können. Hier sind sehr hohe Anforderungen an die Autoren von Evidenzberichten/HTA-Reports/klinische Leitlinien/Patienteninformation und damit an deren Erklärung zu Interessenkonflikten zu stellen. Interesse an einem bestimmten Ausfall der Ergebnisse (vested interest in outcomes) sollte hier ausgeschlossen sein. Für Anhörungen bedeutet das, dass die Teilnehmer offen einer Interessengruppe zugehörig sind oder (als Wissenschaftler) durch eine entsprechen-

de und unverzichtbare Erklärung Unabhängigkeit beanspruchen oder auf Verbindungen hinweisen.

2.6 Fazit

Der Begriff der evidenzbasierten Medizin hat seit seiner Einführung 1991 eine enorme Verbreitung erfahren. Ein zumindest 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 Entscheidungsprozessen gibt. Unter dem Kürzel EbM verbirgt sich eine solche Vielfalt an unterschiedlichen Interpretationen, dass der erreichbare kleinste Nenner „evidenzbasiertes Vorgehen“ zu sein scheint. Nur in dieser Interpretation lässt sich die gesetzliche Forderung damit vereinbaren, dass im deutschen Kontext die Systemebene einbezogen ist, also eigentlich die Evidence-based Healthcare (EBHC) statt EbM genannt werden müsste. Die damit verbundene Vielfalt an Strukturen und Vorgehensweisen aus dem internationalen Raum lassen keine eindeutige Umsetzung der Forderung „nach anerkannten internationalen Standards der evidenzbasierten Medizin (EbM)“ zu, sondern damit wird nur ein Rahmen abgesteckt, in dem Methodenwahl und Vorgehen festgelegt werden müssen.

Insbesondere ist wichtig, dass die Vorgaben aus dem EbM/EBHC-Feld keine unmittelbaren Angaben zur Integration zur Kostenbewertung machen. Allerdings hat sich in den letzten Jahren ein methodisches Fundament für die empirische Arbeit weiterentwickelt, dessen Grundprinzipien – vor allem der Schutz vor Bias und die Kontrolle vor Zufallsfehlern – auch bei der Methodenwahl für die Kosten-Nutzen-Bewertung berücksichtigt werden sollten.

3. Medizinische Biometrie und Epidemiologie

Die Forderung einer Kosten-Nutzen-Bewertung von Arzneimitteln ist im Kern eine politische: Aufgrund der Knappheit der Ressourcen soll (idealerweise) die medikamentöse Behandlung einer Krankheit so erfolgen, dass der Nutzen (der Behandlung) in einem angemessenen Verhältnis zu den Kosten steht. Beide Begriffe, sowohl der des Nutzens als auch der der Kosten

- bedürfen einer Operationalisierung, so dass sie gemessen, bzw. berechnet werden können
- verlangen nach einer Definition des Mess- oder Berechnungsverfahrens
- bedürfen der Durchführung der Messung, respektive der Attestierung einer Größenordnung.

Da es keine absoluten Vorstellungen über das angemessene Verhältnis von Kosten zu Nutzen gibt, ist außerdem das Vorhandensein einer vergleichbaren Medikation oder konkurrierenden Therapieform erforderlich, für die vergleichbare Daten vorliegen.

Aus der Heranziehung der EBM als Standard ergibt sich, dass die für die Bewertung erforderlichen Basisdaten nicht auf einer rein wissenschaftlichen Extrapolation, z. B. vom Tierexperiment auf den Menschen beruhen dürfen, sondern dass hierfür wissenschaftliche Erkenntnisse erforderlich sind, die direkt am Menschen gewonnen wurden. Eine weitere Voraussetzung besteht darin, dass das Medikament bereits zugelassen worden ist, so dass seine Wirksamkeit im Rahmen des Zulassungsverfahrens nach AMG bereits nachgewiesen ist und die Verabreichung des Medikaments mit einem vertretbaren Risiko verbunden ist.

Zur Generierung und Bewertung der für die Anwendung am Menschen verfügbaren Erkenntnisse, bedarf es des Wissens um statistische und krankheitsbezogene Methoden zur Messung biologischer Effekte, sowie der Kenntnis über Messverfahren, Therapie- und Diagnoseverfahren. Hierdurch wird die Kernkompetenz der **Medizinischen Biometrie** angesprochen. Diese entfaltet sich aber nur, wenn Sie in Kooperation mit den betroffenen medizinischen Teildisziplinen entwickelt wird, die ihrerseits über spezifische Kenntnisse der Pathophysiologie, der Pharmakologie und der klinisch-medizinischen Wissenschaft für die spezielle Erkrankung verfügen. Um Kosten und Nutzen unter realistischen Bedingungen einschätzen zu können, bedarf es darüber hinaus des Wissens über das Vorhandensein von Krankheiten in der Bevölkerung, sowie über den Zusammenhang zwischen verursachenden, aber auch protektiven Faktoren und Erkrankungen und deren Verteilung in der Bevölkerung. Hinzu kommen muss die Fähigkeit, Aussagen über den realisierbaren Nutzen und reale Kosten in der Bevölkerung herleiten zu können. Dies berührt die Kernkompetenz der Epidemiologie. Die medizinische Biometrie und Epidemiologie sind mithin Basiswissenschaften für die vom Gesetzgeber formulierte Aufgabe der Kosten-Nutzen-

Bewertung, ihre kompetente Anwendung zieht sich wie ein roter Faden durch **alle** Schritte der Nutzen/Kosten-Nutzen-Bewertung. Hierzu zählt auch die klinische Epidemiologie, die das Bindeglied zwischen Biometrie und Epidemiologie darstellt, sowie die methodische Kernkompetenz des EbM abbildet (s. Abschn. 2).

3.1. Datengrundlagen

Alle am Menschen gewonnen Erkenntnisse zur Wirkung, zum Nutzen und zum Risiko medikamentöser Therapien stammen in dem hier diskutierten Kontext aus sogenannten Studien, für deren Planung, Auswertung und Bewertung die Kompetenz aus medizinischer Biometrie und Epidemiologie unverzichtbar ist. In der Zulassungsphase findet dies beispielsweise seinen Niederschlag darin, dass der "verantwortliche Biometriker" nicht mehr aus einer nach Good Clinical Practice (GCP) geplanten Studie wegzudenken ist. Aber auch nach der Zulassung verlangt die Fragestellung nach Arzneimittelnebenwirkungen die Expertise der Epidemiologie für das Design und die Durchführung entsprechender Studien. Im Zusammenhang der Nutzen-, aber auch der Kosten-Nutzen-Bewertung bietet es sich an, den Begriff der "Studie" breit zu fassen, so dass hierzu neben den Studienformen der Arzneimittelprüfung (s. u.) auch Ergebnisse aus epidemiologischen Studien, aber auch wissenschaftliche Auswertungen aus Registerdaten und Daten des Versorgungssystems zählen. Nicht zur Datenbasis – wenngleich als Hintergrund- und Umfeldinformation relevant – zählen tierexperimentelle Studien und in vitro, bzw. in vivo Experimente.

3.2. Studientypen

Hinsichtlich der verschiedenen am Menschen durchgeführten Studien hat sich in der Wissenschaft eine Klassifikation und Abgrenzung verschiedener Typen herausgebildet. Zunächst einmal unterscheidet man Beobachtungs- und Interventionsstudien. Bei einer reinen Beobachtungsstudie werden die Wirkungen unterschiedlicher Expositionen, hier der Medikamenteneinnahme, beobachtet, ohne dass seitens der Studienleitung ein Eingriff erfolgt. Im Gegensatz dazu steht die Interventionsstudie, die zu einem definierten Zeitpunkt eine Intervention, z. B. die Verordnung eines Arzneimittels, vornimmt und deren Wirkung beobachtet. Da es bei der Arzneimitteltherapie grundsätzlich um eine Intervention geht, erscheint auf den ersten Blick die Interventionsstudie der präferierte Studientyp zu sein. Allerdings lassen sich auch aus Beobachtungsstudien, z. B. bei der Beurteilung von Arzneimittelnebenwirkungen, we-

sentliche Erkenntnisse gewinnen, indem man beispielsweise die Nebenwirkungsrate unter einem speziellen Medikament mit dem anderer Medikamente rein beobachtend vergleicht. Zu den in diesem Umfeld, aber insbesondere bei der Arzneimittelzulassung, wichtigsten Studien gehört das sogenannte Randomized Clinical Trial (RCT), bei dem die Zuteilung einer Prüfsubstanz zum Vergleich mit einer Vergleichssubstanz rein zufällig geschieht. Damit ist die Intervention zufallsgesteuert und damit sind – zumindest in der Theorie – beobachtbare, differentielle Effekte zwischen Interventions- und Vergleichsgruppe ausschließlich der Intervention geschuldet (vgl. aber die Ausführungen zu externen Validität weiter unten). Daneben stehen die Beobachtungsstudien, die sich wie folgt klassifizieren lassen:

- Die Querschnittsstudie: bei diesem Studientyp werden zu einem festen Zeitpunkt aus einer Grundgesamtheit nach einem definierten Prinzip Individuen gezogen und der statistische Zusammenhang zwischen einer Exposition (Medikamenteneinnahme) und einem Gesundheitszustand ermittelt.
- Die Fall-Kontroll Studie: hier wird eine Zahl Erkrankter (Fälle) einer Zahl von Gesunden (Kontrollen) gegenübergestellt, um beispielsweise festzustellen, ob bei den Erkrankten ein höherer Anteil ein definiertes Medikament genommen hat. Ein höherer Anteil der Medikamenteneinnahme bei den Fällen im Vergleich zu den Kontrollen deutet dabei auf ein problematisches Nebenwirkungsspektrum des betrachteten Medikaments hin.
- Die Kohortenstudie : hier werden Personen mit einer Medikation A im Vergleich zu Personen mit einer Medikation B beobachtet, um auf Grund unterschiedlicher Nebenwirkungs- oder Erfolgsraten auf differente Nebenwirkungsrisiken, bzw. Heilungschancen zu schließen.

Wenngleich Querschnitts-, Fall-Kontroll- und Kohortendesign eher mit einer Beobachtungsstudie in Verbindung gebracht werden, so sind diese Studientypen auch grundsätzlich im Zusammenhang mit Interventionsaktivitäten denkbar. Man spricht hierbei von quasi experimentellen Designs.

Im Bereich der Arzneimittelprüfung hat sich in den vergangenen Jahrzehnten eine Systematik herauskristallisiert, die sich an den Phasen der Arzneimittelprüfung orientiert und gleichzeitig Implikationen für die präferierten Studiendesigns aufweist, vgl. hierzu Tabelle 1.

Tabelle 1: Phasen der Arzneimittelprüfung und angewendete Studiendesigns¹⁾ (vereinfachte Darstellung)

	Phase I	Phase II	Phase III	Phase IV
Ziel	Ermittlung pharmakologischer Eigenschaften	Bestimmung von Ansprechraten/Verträglichkeit	Wirksamkeitsnachweis Nutzen / Risikobestimmung	Ermittlung von Sicherheit und Effektivität
dominanter Studientyp	Beobachtungsstudie	Beobachtungsstudie RCT	RCT	Kohorten/ Fall-Kontroll-Studie
Zielgruppe	Gesunde/"aus-therapierte" Patienten	Patienten	Patienten	Patientengruppen Population
Biometrie ²⁾	+	+	++	+
Epidemiologie ²⁾	-	-	+	++

¹⁾ modifiziert nach Jöckel et al. (2007)²⁾ Rolle der Biometrie / Epidemiologie: - weniger relevant, + relevant, ++zwingend erforderlich

Typischerweise liegen zu einem Medikament bei einer gegebenen Indikationsstellung Studien der Phase III am Menschen vor, die ihre Existenz hauptsächlich den regulatorischen Auflagen der Arzneimittelprüfung verdanken. Diese werden in der Regel mit Mitteln der pharmazeutischen Industrie finanziert. Daneben gibt es die Ergebnisse sog. Investigator Initiated Trials (IIT)-Studien, bei denen typischerweise Indikationsausdehnungen durch Studien aus der Wissenschaft heraus angegangen werden. Diese Studien folgen einem klaren Studienprotokoll mit einem definierten gesundheitlichen Endpunkt, einer klaren Studienhypothese, einer definierten Interventions- und Vergleichstherapie. Als Studientyp dominiert das RCT.

In der Kosten-Nutzen-Bewertung geht es aber nun um die Zusammenschau aller verfügbaren Evidenz. Dabei geht es um die Frage, welcher Nutzen und welche Kosten entstehen würden, wenn statt eines Standardpräparats (dem sog. Komparator) in der Versorgung das Medikament eingesetzt würde, für das der Prüfauftrag erteilt wurde. Hierzu liegen in aller Regel niemals direkte Daten vor, sondern es handelt sich im Kern immer um eine Extrapolation wissenschaftlicher Studiendaten am Menschen auf eine real life Situation. Kern des Problems ist es dabei, dass nicht nur solche Studien in die Analyse einbezogen werden müssen, die genau dieser Fragestellung

entsprechen, sondern auch solche, bei denen beispielsweise der gesundheitliche Endpunkt "mituntersucht" wurde. Deswegen ist es erforderlich in diesem Kontext auf drei zentrale Begriffe einzugehen:

- den Endpunkt
- den Begriff des Bias und
- die externe und interne Validität.

3.3. Endpunkt

Üblicherweise wird ein Medikament für eine bestimmte Indikation verordnet, z.B. bei der Diagnose Bluthochdruck. Da aber der Bluthochdruck per se im Regelfall dem Patienten keine Probleme bereitet, ist die Wahl des Blutdrucks als Endpunkt problematisch. Was bringt es, ein Antihypertensivum zu verordnen, wenn die Wahrscheinlichkeit für die Entstehung eines Myokardinfarkts oder eines Schlaganfalls nicht verändert wird? Zwar ist der Bluthochdruck ein Risikofaktor für jede dieser Erkrankungen, aber es ist nicht nachgewiesen, dass jede medikamentöse Senkung des Blutdrucks automatisch die Erkrankungsrisiken für die vorher genannten Krankheiten in vergleichbarer Weise vermindert. Wir bezeichnen daher den Blutdruck auch als Surrogatendpunkt. Grundsätzlich sind Studien mit direkten Endpunkten, z.B. dem Myokardinfarkt, solchen mit Surrogatendpunkten vorzuziehen, denn sie machen einen großen Teil des oben angesprochenen Extrapolationsprozesses vermeidbar. Andererseits sind Situationen vorstellbar, in denen keine hinreichende Evidenz für den gesundheitlich relevanten Endpunkt vorliegt, bzw. bei ausschließlicher Berücksichtigung nur solcher Studien zum geforderten gesundheitlichen Endpunkt die Datenlage zu dürrig wird. Ob ein Surrogatparameter als Endpunkt in Frage kommt, muss für jede medizinische Situation neu entschieden werden. Dabei sind Aspekte wie die Qualität der verfügbaren Studien, pharmakologisches Wissen, aber auch der gewählte Komparator zu berücksichtigen. Eine ausschließlich an statistischen Gesichtspunkten orientierte Einschätzung der Adäquanz von Surrogatendpunkten, wie etwa bei den sogenannten Prentice-Kriterien, vgl. Prentice (1989) greift zu kurz, insbesondere dann, wenn sie sich auf isolierte Studien bezieht. Hierzu sei exemplarisch auf die Diskussion und zitierte Literatur im Editorial von Baker (2006) hingewiesen. Beispiele für Surrogatendpunkte sind CD4 positive Lymphozytenzahl bei HIV,

positiver Antititer bei Impfungen, gesenkter Cholesterinspiegel bei Herz-Kreislaufkrankheiten, Tumorgröße bei Krebserkrankungen.

Formale Kriterien zur Zulässigkeit von Surrogatendpunkten können zu streng, aber auch zu weich sein, zumindest was die **Bewertung** des Nutzens angeht. Aus wissenschaftlicher Sicht ist es nicht vertretbar, Evidenz auf der Basis von Surrogatendpunkten unberücksichtigt zu lassen. Vertretbar erscheint es hingegen, bei starker Evidenz für einen Surrogatendpunkt und gleichzeitig fehlender oder schwacher Evidenz für den „wahren“ Endpunkt zu einer vorläufigen positiven Kosten-Nutzeneinschätzung zu kommen. Gleichzeitig ist dann aber zu fordern, dass nach einer gewissen Karenz, belastbare Studien und Daten für den „wahren“ Endpunkt nachgeliefert werden.

3.4. Bias

Mit dem englischen Begriff bias (auf Deutsch Verzerrung) bezeichnet man die systematische Abweichung eines Studienergebnisses von der Wahrheit. Ein typisches Problem in diesem Zusammenhang betrifft die Frage des sog. publication bias, die daraus resultiert, dass positive Ergebnisse faktisch eine höhere Wahrscheinlichkeit zur Publikation haben als negative. Was dabei „positiv“ und was „negativ“ heißt, hängt vom jeweiligen Setting ab. So ist beispielsweise eine Studie, die eine bessere Heilungschance durch ein Präparat für eine bestimmte Krankheit gegenüber einem Vergleichspräparat zeigt, leichter zu veröffentlichen, als eine, die eine solche Überlegenheit nicht zeigt. Aber auch umgekehrt lässt sich eine Studie, die eine relevante Arzneimittelnebenwirkung zeigt, leichter publizieren, als eine solche, die keinen Unterschied für dieses Präparat hinsichtlich des allgemeinen Nebenwirkungsspektrums aufweist. Zwar haben in der Vergangenheit die medizinisch-wissenschaftlichen Zeitschriften verschiedene Anstrengungen unternommen, einen solchen publication bias, wenn nicht zu vermeiden, so zumindest zu minimieren. Allerdings ist trotz aller Bemühungen in diese Richtung (z.B. Studienregister) nicht von einer völligen Eliminierung dieses publication bias auszugehen. Diese Form des bias berührt die Frage der Auswahl der Studien, die für die Evidenzgewinnung zur Verfügung steht. Daneben gibt es aber in den Studien selbst auch weitere Verzerrungsquellen. Diese sind insbesondere für die weiter oben genannten Beobachtungsstudien zu berücksichtigen: wenn beispielsweise Personen, die eine spezielle Zielmedikation präferieren, eine besonders gute Prognose einer Krankheit aufweisen, darf es nicht verwundern, dass

die mit der Zielmedikation behandelten Patienten in einer Kohortenstudie bessere Krankheitsverläufe aufweisen als solche Personen, die diese Zielmedikation nicht gewählt haben. Eine Sonderstellung nimmt in diesem Zusammenhang sicherlich das RCT ein, weil es durch die zufällige Zuteilung von Prüf- und Vergleichsmedikation (idealerweise unter Verblindung des Patienten und Verblindung des Therapeuten) diese Verzerrungsquellen ausschließt.

3.5. Interne und externe Validität

Damit weist das RCT anscheinend eine optimale Validität auf. Der Vergleich der beiden Gruppen kann – sofern sich ein Kontrast zwischen Interventions- und Vergleichstherapie findet – nicht durch Mängel der internen Validität erklärt werden. Patienten – gleich welcher Art ihre besonderen Charakteristika sind – haben die gleiche Chance, die Prüf- wie die Vergleichsmedikation zu erhalten. Etwas anders sieht dies aus, wenn man sich mit der Frage beschäftigt, ob dies auch für die externe Validität gilt. Wenn Personen, die sich willig einer RCT unterziehen, grundsätzlich andere Charakteristika aufweisen, als solche die nicht zur Teilnahme einer solchen Studie bereit sind, dürfen Therapieunterschiede nicht ohne weiteres auf die allgemeine Bevölkerung extrapoliert werden. Noch deutlicher wird dies, wenn wir die für eine Studie zugelassenen Patienten im Rahmen des Studiengeschehens mit den späteren Anwendern des Medikaments vergleichen. In einer Studie werden klare Ein- und Ausschlusskriterien definiert, die die verfügbaren Patientenkollektive sehr homogen machen. In der allgemeinen Versorgung hingegen werden auch solche Patienten therapiert, die nicht unbedingt diesen strengen Ein- und Ausschlusskriterien genügen. Daher kann der Nutzen, aber auch ein potentieller Schaden in der allgemeinen Versorgung deutlich anders ausfallen als in der Studie. Wenn zusätzlich Patientenpräferenzen (Bereitschaft zur Teilnahme an einem Randomised Clinical Trial) eine Beziehung zur Prognose besitzen, wird die Sache ganz unübersichtlich. In diesem Falle müssen wir von einer eingeschränkten externen Validität der zur Verfügung stehenden Studien ausgehen.

Zentral ist bei der Diskussion um externe und interne Validität der Begriff der Übertragbarkeit von Studienergebnisse in die Praxis. Dahinter steht letztlich die Frage, wie aus den Studiendaten auf die Performance eines Medikaments in einer (fiktiven) zukünftigen Versorgungspraxis extrapoliert wird. Hier stellt sich dann auch die Fra-

ge, ob nicht für die Zukunft die Qualität dieses Extrapolationsprozesses prospektiv überprüft werden sollte.

3.6. Zusammenfassung

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

1. zum interessierenden Medikament gibt es verschiedene Studien, die nicht unbedingt alle – z. T. aufgrund regulatorischer Auflagen – den selben Endpunkt im Visier haben.
2. Die Studien weisen unterschiedliche Ein/Ausschlusskriterien auf, womit die Frage der externen Validität gestellt werden muss
3. manche Studien sind international, manche auch nur national, manche gar nicht publiziert, was bei der zusammenschauenden Bewertung die Frage nach dem publication bias aufwirft.
4. Die Studienlandschaft bildet im Allgemeinen nur unzureichend die medizinische Versorgungspraxis ab, womit sich die Forderung nach weiteren (epidemiologischen) Studien ergibt.
5. Es liegen quantitativ und qualitativ unterschiedliche Informationen zur Versorgungsnotwendigkeit und Versorgungspraxis vor.

In dieser Situation erhalten die medizinische Biometrie und Epidemiologie eine besondere Relevanz: die Methodik dieser Fächer ist geeignet, Licht in das Dunkel zu bringen. Das Zauberwort hierzu lautet: Transparenz. Diese erzeugen - qua definitionem – der Begriff des Studienprotokolls/der SOPs, das/die prospektiv und nachvollziehbar festlegen, wie mit unterschiedlichen Situationen umzugehen ist. In einem Studienprotokoll einer klinischen oder epidemiologischen Studie wird prospektiv festgelegt, wie 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 (nach Möglichkeit 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 aufgetre-

tene 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. Die Ergebnisermittlung folgt 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.

4. Inhaltliche und methodische Aspekte der Nutzenbewertung nach § 35b Abs. 1 SGB V

4.1 Vorbemerkungen

a) Nach § 35b Abs. 1 Satz 1 kann das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) vom Gemeinsamen Bundesausschuss (G-BA) oder vom Bundesministerium für Gesundheit (BMG) beauftragt werden, „den Nutzen oder das Kosten-Nutzen-Verhältnis von Arzneimitteln zu bewerten“.

In diesem Abschnitt geht es allein um die eigenständige oder die einer Kosten-Nutzen-Bewertung vorausgehende Nutzenbewertung von Arzneimitteln und anderen Behandlungsformen.

b) Eine Nutzenbewertung ist nach unserer Auffassung ein anspruchsvolles Forschungsvorhaben. Insofern erscheint der Begriff Nutzen-„Bewertung“ problematisch. Er wird jedoch akzeptabel durch den unten auszuführenden Hinweis, dass in einen solchen Forschungsauftrag und seine Durchführung vielfältige Wertentscheidungen eingehen. Dieses Faktum erschwert es, der von IQWiG und G-BA in Anlehnung an das britische National Institute for Health and Clinical Excellence (NICE) favorisierten Unterscheidung von „assessment“ und „appraisal“ umstandslos zu folgen. „Assessment“ soll sich dabei auf die neutrale „wertfreie“ Darstellung der Datenbasis durch das IQWiG beziehen, „appraisal“ auf die explizite Bewertung der gesamten Evidenzlage in einem „umfassenden Abwägungsprozess“ durch den G-BA (§ 20 Abs. 1 Verfahrensordnung). Zusätzlich zu normativen Überlegungen und Entscheidungen sind

im „assessment“ vielfältige inhaltliche Fragen zu klären; hierzu ist eine intime Kenntnis des jeweiligen klinischen Krankheitsbildes und seiner Behandlungsformen unabdingbar. Eine umfassende Nutzenbewertung erfordert die Zusammenstellung eines multidisziplinären Teams.

c) „Die Bewertung erfolgt durch Vergleich mit anderen Arzneimitteln und Behandlungsformen unter Berücksichtigung des Zusatznutzens für die Patienten im Verhältnis zu den Kosten.“ (Satz 3) Wir gehen davon aus, dass eine vergleichende Bewertung auch im Falle einer eigenständigen Nutzenbewertung erforderlich ist.

d) „Das Institut bestimmt auftragsbezogen über die Methoden und Kriterien für die Erarbeitung der Bewertung nach Satz 1 auf der Grundlage der in den jeweiligen Fachkreisen anerkannten internationalen Standards der evidenzbasierten Medizin und der Gesundheitsökonomie.“ (Satz 5)

e) Damit ergibt sich folgende Staffelung von Determinanten einer vergleichenden Nutzenbewertung: Diese bedient sich einer Reihe von auftragsbezogenen Methoden und Kriterien auf der Grundlage von fachlich anerkannten internationalen Standards der evidenzbasierten Medizin im Dienste weniger Grundprinzipien und im Kontext ethischer und (sozial)rechtlicher Normen wie sie z.B. in § 70 SGB V formuliert sind.

Die EbM verpflichtet nach der prioritären Berücksichtigung der Prinzipien des Wohltuns und der Schadensvermeidung in der Patientenbehandlung („beneficence, non-maleficence“ nach Beauchamp und Childress 4.1994) vor allem zu

- einer möglichst vollständigen Vergegenwärtigung der thematisch einschlägigen Originalveröffentlichungen einschließlich ihrer Lücken und Grenzen,
- einer umfassenden Prüfung und Kontrolle ihrer Daten- und Informationsbasis in Hinblick auf ein breites Spektrum möglicher Verzerrungen (Biases, systematische Fehler), auf Zufallseinflüsse und weitere Unsicherheiten,
- einer Prüfung der Praxisrelevanz und Anwendbarkeit der Evidenz im jeweils gegebenen Praxiskontext und
- allgemein zu Transparenz, Fairness und Rationalität („reasonableness“) aller Arbeitsschritte unter Berücksichtigung verschiedener Perspektiven.

So arbeitet der Standard „systematische Literatursuche“ in definierten Datenbanken, durch Handsuche und Verfolgung von Literaturzitaten dem ersten Prinzip, die Kontrolle von Behandlungsabbrüchen und Noncompliance dem zweiten Prinzip, die kriti-

sche Prüfung von Surrogatparametern dem dritten Prinzip und die Beratungsbeteiligung u.a. von Patienten, Industrievertretern und Klinikern dem vierten Prinzip zu.

f) Schon jetzt ist hervorzuheben, dass die Standards und Grundprinzipien der evidenzbasierten Medizin (EbM) nirgendwo in kodifizierter Form vorliegen. Während die Grundprinzipien als vergleichsweise stabil gelten können, ändern sich die Standards und Methoden mehr oder weniger rasch. Sie zu einem bestimmten Zeitpunkt festzustellen, erscheint möglich, erforderte aber eine weit reichende Literatur- und Dokumentenanalyse; dies ist an dieser Stelle allein schon aus Zeitgründen nicht möglich.

Der folgende Text konzentriert sich auf einige Brennpunkte der aktuellen Diskussion im Kontext der sog. evidenzbasierten gesundheitlichen Versorgung („evidence-based health care“, EbHC; in Abgrenzung zur evidenzbasierten klinischen Medizin) und eines ihrer zentralen Instrumente, der systematischen Bewertung medizinischer Verfahren (Health Technology Assessment, HTA).

g) § 135 Abs. 1 SGB V normiert die Überprüfung des Nutzens, der Notwendigkeit und der Wirtschaftlichkeit von Untersuchungs- und Behandlungsmethoden, die zu Lasten der Gesetzlichen Krankenversicherung (GKV) erbracht werden sollen. Die Reihenfolge der unbestimmten Rechtsbegriffe ist wichtig: eine Methode, die keinen Nutzen in Aussicht stellt, kann nicht notwendig und wirtschaftlich sein. Insofern sind wir der Auffassung, dass Wirtschaftlichkeitsprüfungen – in welcher Form auch immer (cf. Abschnitt 4) – einer Nutzenbewertung sachlich und zeitlich nachzuordnen sind.

4.2 „Nutzen“ – ein komplexer Begriff

Es ist aus unserer Sicht bemerkenswert, dass das Arzneimittelgesetz den Begriff des Nutzens in seinen ersten fünf Abschnitten, also auch in den Bereichen „Anforderungen an die Arzneimittel“ und „Zulassung der Arzneimittel“ nicht kennt. Er taucht erst im Abschnitt 6 im Zusammenhang mit dem „Schutz des Menschen bei der klinischen Prüfung“ auf (u.a. § 40 Abs. 1 Satz 3 Ziffer 2), ohne jedoch eine Definition erfahren zu haben. Vorher ist im AMG von „Qualität, Wirksamkeit und Unbedenklichkeit der Arzneimittel“ die Rede (§ 1), wobei die Unbedenklichkeit in § 5 AMG als erste „Anforderung an die Arzneimittel“ genannt wird. Auch die Rechtsprechungen des Bundessozialgerichts und des Bundesverfassungsgerichts kennen die Begriffe Wirksamkeit, Behandlungserfolg, Aussicht auf Heilung etc., aber nicht den Begriff des Nutzens. Dagegen trennt die Verfahrensordnung des G-BA ausdrücklich „Wirksamkeit“ und

„Nutzen“ (§ 17). Diese unübersichtliche Terminologie erfordert eine Klärung. Der Gebrauch des Begriffs Nutzen in § 135 Abs. 1 SGB V und seine inhaltliche patientenorientierte Bestimmung in § 35b Absatz 1 Satz 4 lässt ein umfassendes Verständnis vermuten und legt eine Binnendifferenzierung nahe. Wir plädieren für folgende Unterscheidungen:

Wirksamkeit: Wirksamkeit beschreibt ein generelles Merkmal eines Arzneimittels (oder einer anderen Behandlungsmethode); es erfordert den Nachweis, dass beobachtbare Wirkungen (eintretende oder ausbleibende Ereignisse oder Veränderungen) kausal auf seine Anwendung zurückgeführt werden können. Dieser Nachweis erfordert bestimmte „Versuchsanordnungen“ (P. Martini), in aller Regel die der kontrollierten randomisierten klinischen Studie („randomised controlled trial“, RCT). Von dieser Regel gibt es wenige Ausnahmen, z.B. im Bereich der Anästhesiologie (sog. „all or none“-Situationen; cf. Glasziou et al 2007). Ein weiteres beliebtes Beispiel bezieht sich auf die Unnötigkeit, die präventive Wirksamkeit von Fallschirmen bei Sprüngen aus Flugzeugen mit Hilfe eines RCTs nachzuweisen. Solche Grenzsituationen sind in der Medizin aber (zunehmend) selten, gerade im Bereich präventiver Interventionen. In aller Regel sind heute kleinschrittige Innovationen zu evaluieren. Fast immer werden daher zum Wirksamkeitsnachweis kontrollierte und randomisierte Studien benötigt; dabei ergibt sich die Wirksamkeit aus der Differenz oder dem Verhältnis der in Raten oder Mittelwerten aggregierten Wirkungen zweier oder mehrerer Behandlungsbedingungen bzw. -gruppen (Interventionsbedingung/gruppe, Kontrollbedingung/gruppe). Der Nachweis ist hierbei „nur“ statistischer Natur und muss gegen Zufall (aber auch allfällige systematische Verzerrungen, Biases) gesichert werden. Die Verzerrungen können auf allen Stufen des Forschungsprozesses zwischen Studienplanung und Publikation wirksam werden.

An Einzelfällen und Fallserien können zwar positive wie negative Wirkungen beobachtet, aus ihnen können Wirksamkeitshypothesen abgeleitet werden, der Nachweis von Wirksamkeit erfordert aber konfirmatorische – kontrollierte und im besten Fall auch randomisierende - Studiendesigns.

Wirksamkeit wird in Hinblick auf positiv bewertete Wirkungen untersucht. Allerdings stellt sich die Kausalitätsfrage auch für sog. unerwünschte Wirkungen. Hier sind jedoch (nach § 5 AMG) die Evidenzanforderungen geringer. Es reicht im Sinne des Prinzips der Risikovorsorge der „begründete Verdacht“ – ein Nachweisniveau, das für

die Behauptung von (positiver) Wirksamkeit inakzeptabel wäre. Im Feld der Arzneimittelzulassung und –bewertung scheint eine ethische Intuition zu gelten, die uns eher risikoavers und nutzenskeptisch als rein nutzenaffin sein lässt. Diese wird von der EbM grundsätzlich (z.B. Sackett 2006, S. 192/3) geteilt. Auch das für die ambulante Versorgung geltende Verbot neuer Methoden unter Erlaubnisvorbehalt spricht in diese Richtung (§ 135 SGB V). Eine andere Relation gilt in der klinischen Situation, vor allem vor lebensbedrohenden Krankheiten. Dort treten Risikoabwägungen in den Hintergrund. Dennoch kann auch hier der Nutzen nicht einfach behauptet werden. Nach dem Urteil des Bundesverfassungsgerichts vom 6.12.2005 (1 BvR 347/98) muss eine – auf Indizien gestützte - „nicht ganz entfernt liegende Aussicht auf Heilung oder auf eine spürbare positive Einwirkung auf den Krankheitsverlauf“ bestehen; gefordert sind Daten wenigstens zweier sich wechselseitig kontrollierender Anwendungsbeobachtungen.

Das wesentliche Ergebnis einer konfirmatorischen Studie ist, dass die Wirksamkeitshypothese aufrechterhalten oder abgelehnt werden kann. Denkbar ist auch ein inkonklusives Ergebnis (Wirksamkeit unsicher).

Grundsätzlich zu berücksichtigen ist der „optimism bias“: die erste Studie zu einem Arzneimittel bringt in aller Regel besonders positive Ergebnisse (cf. Chalmers and Mathews 2006). Deshalb sind grundsätzlich weitere Studien zu fordern, bevor der Wirksamkeitsnachweis als ausreichend geführt gelten kann. Anders gesagt: eine Studie ist nicht viel besser als keine Studie.

Nutzen: Nutzen ist mehr und anderes als Wirksamkeit. An ihm lassen sich qualitative und quantitative Merkmale unterscheiden.

Die erste Frage richtet sich auf die Nutznießler/Benefiziäre. Auch wenn § 35b SGB V ganz auf die behandelten Patienten abstellt, so ist es doch geboten, weitere mögliche Nutznießer ins Auge zu fassen: zuerst ist an Angehörige (z.B. von Demenzkranke) zu denken, dann an weitere Gruppen: Impfungen schützen auch Nichtgeimpfte. Allerdings wird man die Marktzulassung oder GKV-Finanzierung einer therapeutischen Methode nicht mit einem Nutzen für ferner stehende Benefiziäre, z.B. die Anwender der Methoden, die Heilkunde oder medizinische Wissenschaft, den Hersteller oder die Volkswirtschaft begründen können.

Die zweite Frage richtet sich auf die Nutzenqualitäten. Immer geht es nach § 35b Abs. 1 Satz 4 allein um klinischen, d.h. gesundheitlichen Nutzen, der direkt durch den

Einsatz der Behandlungsmethode bewirkt werden kann. Ein solcher Nutzen wird unterschiedlich relevant sein. Der Vermeidung von Todesfällen wird man eine grundsätzlich größere Bedeutung zuschreiben müssen (auch aus verfassungsrechtlicher Sicht, cf. Zentrale Ethikkommission 2007) als der Verhinderung oder Linderung von schwerem Leiden und erheblichen Störungen der sozialen Teilhabe, der Besserung der Lebensqualität, dem Abbau von Risikofaktoren (u.a. auch der sog. Noncompliance) oder der Steigerung des Wohlbefindens.

In diesem Zusammenhang ist die Unterscheidung klinisch relevanter Endpunkte von sog. Surrogatparametern wichtig. Surrogatparameter können einmal Teil einer Kausalkette zwischen Intervention und Endpunkt („intermediate outcome“) sein, zum anderen mit diesem nur korreliert sein („proxy“). Im günstigen Fall sind sie mit dem eigentlich interessierenden „Outcome“ statistisch eng und verlässlich assoziiert. Diese Assoziation ist in jedem Einzelfall und für jede Untersuchungsgruppe zu belegen. So war z.B. die Erhöhung der Knochendichte (durch hoch dosiertes Natriumfluorid; cf. Riggs et al. 1990) nicht mit einer Erniedrigung der Frakturrate in der Interventionsgruppe, sondern im Gegenteil mit deren Erhöhung verbunden.

Weiter sind das Ausmaß des Nutzens (Risikodifferenz, Mittelwertdifferenz, relatives Risiko, Odds und Hazard Ratio), seine Eintrittswahrscheinlichkeit (NNT), sein Eintrittsbeginn und seine Nachhaltigkeit abzuschätzen und zu bewerten. Schließlich fallen die Unsicherheit der Effektschätzung und die Konsistenz der gesamten Evidenzlage ins Gewicht.

Schaden: Ähnlich differenziert wird man die unvermeidlichen Belastungen und Schadensrisiken der Behandlungsmethoden vergegenwärtigen müssen. Auch hier sind die möglicherweise von Schädigungen Betroffenen ebenso wie Schadensqualitäten und –relevanz, Schadensausmaß, Eintrittswahrscheinlichkeit (NNH), Eintrittszeitpunkt und Schadensdauer sowie die Evidenzlage abzuschätzen.

Es ist schon darauf hingewiesen worden, dass das Vorsorgeprinzip zu einer negativen Bewertung schon dann führt, wenn sich „nach dem jeweiligen Stand der wissenschaftlichen Erkenntnisse der begründete Verdacht“ ergibt, dass ein Arzneimittel „bei bestimmungsgemäßem Gebrauch schädliche Wirkungen (hat), die über ein nach den Erkenntnissen der medizinischen Wissenschaft vertretbares Maß hinausgehen“ (§ 5 Abs. 2 AMG). Mit anderen Worten: Die Evidenzanforderungen sind im Bereich Schaden/Risiken geringer als im Bereich Nutzen/Chancen. Zusätzlich ist zu berücksichti-

gen, dass kontrollierte randomisierte Studien Grenzen haben, wenn es um die umfassende Vergegenwärtigung von Schadenpotentialen (besonders im Bereich seltener und schwerer Nebenwirkungen) geht.

Netto-Nutzen: Als Nettonutzen wird im Allgemeinen das bezeichnet, was vom Gesamtnutzen nach Abzug aller Schäden übrig bleibt. Seine Abschätzung steht vor ungelösten Problemen: einerseits gibt es Outcome-Parameter (z.B. Lebensqualitätsinventare), deren Messung Nutzen und Schaden integriert. Ein niedriger Messwert kann durch fehlenden Nutzen, durch merklichen Schaden oder durch eine Kombination beider bedingt sein. Andererseits ist es bis heute unklar, wie heterogene Schaden- und Nutzenindikatoren systematisch gegeneinander abgewogen werden können. Wie soll man z.B. vorgehen, wenn der mögliche Nutzen in der Reduktion oder Vermeidung neurologischer Defizite besteht, der mögliche Schaden sich aber in einer Zunahme von bakteriellen Infektionen (Pneumonie, Sepsis) bemerkbar macht. Es fehlt an einer allseits akzeptierten einheitlichen Währung, in die Relevanz, Ausmaß, Häufigkeit und Zeitcharakteristika ganz unterschiedlicher Nutzen- und Schadenspotentiale „gewechselt“ werden können.¹¹

Zusatznutzen: Noch schwieriger ist die Bestimmung und Beurteilung des Zusatznutzens. Hierzu müssen die Nettonutzen mehrerer unterschiedlicher Behandlungsbedingungen mit einander verglichen und gegeneinander abgewogen werden.

4.3 Nutzenabschätzung – unter welchen Erhebungsbedingungen?

Zahlreiche Einwände gegen den RCT beziehen sich auf dessen artifiziellen Charakter, der Nutzen nur unter angeblich optimalen aber unrealistischen Bedingungen prüfe. Dies würde generell zu einer Überschätzung von Wirksamkeit und Nutzen führen. Entsprechend wird in der englischen Literatur zwischen „efficacy“ (Wirksamkeit/Nutzen unter Studienbedingungen) und „effectiveness“ (Nutzen unter Alltagsbedingungen) differenziert. Diese Unterscheidung ist relevant und wird auch von der Verfahrensordnung des G-BA wiederholt (§ 17 Abs. 1). Allerdings ist es keineswegs ausgemacht, dass „effectiveness“-Studien regelmäßig zu (für die geprüfte Methode) günstigeren Ergebnissen kommen. Auch wenn sie gelegentlich den Nutzen überschätzen mögen, so finden sich doch einerseits auch gleiche Wirksamkeits/Nutzen-Ergebnisse von „efficacy“- und „effectiveness“-Studien; und andererseits sind es oft

¹¹ In der gesundheitsökonomischen Analyse wird über die Nutzwerte sowie die Zahlungsbereitschaft (vgl. Abschn. 5 u. 6) eine Integration der unterschiedlichen Effekte angestrebt.

die „realistischeren“ Studienansätze, die eine umfassendere und genauere Abschätzung des Schadenspotentiale erlauben. Wir halten es für nicht ausgeschlossen, dass z.B. chirurgische oder psychotherapeutische Methoden in den Händen erfahrener Kliniker einen größeren Nutzen stiften, als es nach den durchschnittlichen Ergebnissen eines RCT zu vermuten ist. Und natürlich kann auch das Gegenteil der Fall sein.

Besonders für die Beurteilung von Schadenspotentialen werden nicht nur Daten aus RCTs, sondern auch au/s schwächeren Studiendesigns benötigt, z.B. aus Phase 4-Studien, Registerstudien und qualitativ hochwertigen nicht-interventionellen Studien (Anwendungsbeobachtungen nach den Empfehlungen des BfArM und PEI vom 9.5.2007). RCTs sind aufgrund ihres begrenzten Umfangs und ihrer begrenzten Nachbeobachtungszeit generell wenig geeignet, schwere aber seltene Nebenwirkungen zu erfassen. Sie sind ein Instrument des Wirksamkeitsnachweises. Eine umfassende Nutzenbewertung kann deshalb weder auf realistische oder auch „pragmatische“ RCTs (Tunis et al 2003) noch auf ergänzende Befunde aus nicht randomisierten und auch nicht kontrollierten Studien verzichten (Black 1996). Für den Wirksamkeitsnachweis („can it work?“) ist der möglichst pragmatisch anzulegende RCT mit hoher interner Validität unverzichtbar; für die realitätsgerechte Nutzenbewertung wird man Ergebnisse aus anderen Studientypen fordern und fördern müssen.

Dies bedeutet für die Literatursuche, dass man sie nicht a priori auf die sog. beste Evidenz (aus RCTs) begrenzen kann, und es bedeutet für die Literaturanalyse, dass man jedenfalls dann Evidenz auch geringerer Qualität prüfen muss, wenn sie zu „positiven“ Wirksamkeits- und Nutzenbelegen in Widerspruch steht. Dort drohen Evidenzfallen durch blendende oder auch falsch-positive Studien (cf. Knipschild 2002; Joannidis 2005).

4.4 Indirekte Evidenz

Evidenz wird als „indirekt“ bezeichnet, wenn sie sich nicht unmittelbar auf die interessierenden Methoden, Outcomes, Zeitpunkte und Populationen bezieht, wenn zwei Methoden nicht „head to head“ mit einander verglichen wurden, wenn überhaupt Glieder in Nachweisketten fehlen und wenn Nachweisketten mehr als zwei Glieder umfassen (GRADE Working Group 2004; Harris et al 2001). Es ergeben sich dann Fragen wie:

- Gelten die Daten aus einer oder mehreren Stichprobenuntersuchung auch für meine Patienten? Kann man von der untersuchten Wirksamkeit zu einem bestimmten Zeitpunkt auf Wirksamkeit zu früheren oder späteren Zeitpunkten schließen (Problem der externen Validität)?
- Beinhaltet eine Veränderung eines biologischen „paraklinischen“ Parameters eine Veränderung des klinischen Ergebnisses (Problem der Surrogatparameter)?
- Verbürgt die Messung eines Nutzens mit der Methode A (z.B. Fragebogen) auch einen Nutzen im Lichte der Methode B (z.B. Interview) (Problem des indirekten Outcomesvergleichs)?
- Wenn Behandlung A besser ist als Behandlung B und B gleich oder besser als C, ist dann A besser als C (Problem des indirekten Methodenvergleichs)?
- Bedingt die nie optimale Identifikation von Hochrisikopatienten eine intensiviertere Therapie mit der Folge einer Vermeidung von Todesfällen (Problem der mehrgliedrigen Kette)?

Nach den Vorschlägen der GRADE-Gruppe führt indirekte Evidenz zu einer Abwertung der Gesamtqualität der Beweislage, nicht aber zu ihrer vollständigen Verwerfung. Es bleibt daher erforderlich, auch indirekte Evidenz zu suchen und zu beurteilen - jedenfalls dort, wo eine „direkte“ Antwort auf eine versorgungsrelevante Wirksamkeits/Nutzenfrage nicht gegeben werden kann (cf. „one must rely on 2 or more bodies of evidence“ (Harris et al 2001, M-24).

4.5 Fehlende Evidenz

Eine offene Frage ist es, wie fehlende Evidenz zu bewerten ist. Grundsätzlich gilt der Satz, dass das Fehlen von Evidenz (für die Effektivität einer Methode) nicht identisch ist mit Evidenz für fehlende Effektivität. Allerdings ist daran festzuhalten, dass die Beweislast grundsätzlich bei dem liegt, der Wirksamkeit und Nutzen behauptet. Bei neuen Untersuchungs- und Behandlungsmethoden ist daher uneingeschränkt zu fordern, dass stichhaltige Unterlagen vorgelegt werden. Ein „in dubio pro“ wäre nicht vertretbar.

Es ist aber zu überlegen, ob es nicht begrenzte und bedingte Zulassungen zur GKV-Finanzierung geben könnte in Fällen, in denen Methoden bei an sich noch ungenügender Evidenz als potentiell wirksam, sehr nützlich oder für die Heilkunde bedeut-

sam bewertet werden. Hier könnte man sich an einem Modell orientieren, das in der Schweiz im Jahr 2001 noch möglich war. Dort konnten befristete Zulassungen unter bestimmten Bedingungen (u.a. Durchführung einer Studie, Beteiligung an einem RCT oder Register, Phase 4-Studie) erteilt werden. Für ein solches Vorgehen gibt es auch bei uns wenige und unsystematische Beispiele (Akupunktur, Hochdosischemotherapie bei Brustkrebs).

4.6 Offene und versteckte Werturteile

Das European Network for Health Technology Assessment hat im Juni 2007 ein sog. Core Model für einen HTA-Bericht vorgeschlagen (www.eunethta.net). Eine der Domänen soll sich mit „ethischen Aspekten“ beschäftigen. Hierzu werden sehr konkrete Vorschläge gemacht, u.a. im Rückgriff auf die oben schon angedeuteten ethischen Prinzipien von Beauchamp und Childress (2001; Gebot des Wohltuns, der Schadensvermeidung, der Respektierung der Patientenautonomie, der Gerechtigkeit).

Die typische Nutzenbewertung hat mit interessensgebundener Gewichtung die beiden ersten Gebote im Blick; wichtig ist es aber auch zu fragen, ob, wie und in welchem Maße eine Untersuchungs- und Behandlungsmethode die Patientenautonomie und die soziale Gerechtigkeit berührt. Wie wird z.B. Bedarfsgerechtigkeit gegen die Forderung der Gleichmäßigkeit der Versorgung abgewogen (cf. § 70 Abs. 1 SGB V)? Werden bestehende Versorgungsungleichheiten und damit eine „sozial bedingte Ungleichheit von Gesundheitsschancen“ (§ 20 Abs. 1 SGB V) gemildert oder verschärft? Welches „moral hazard“ beinhaltet eine neue Technologie?

Noch leichter zu übersehen sind Wertbezüge und –entscheidungen, die die Auswahl, Auswertung und Bewertung sowie die Synthese der vorfindlichen empirischen Evidenz betreffen, also schon vor und auf der Stufe des sog. Assessments wirksam werden. Welche Studien werden aus welchen Gründen aus den Analysen ausgeschlossen, z.B. weil die Patientenzahl als zu gering oder die Ausfallquoten als zu erachtet werden, weil Surrogatparameter gewählt wurden, zu früh oder zu spät nachbeobachtet wurde, die gewählte kritische Differenz zwischen zwei Methoden („minimally important difference“) oder die Hypothesenbildung (Differenz- oder Äquivalenzstudie, einseitige oder zweiseitige Testung) nicht angemessen schien, der Vergleich unfair organisiert wurde etc. pp.

Hier ist zu fordern, dass solche Wertentscheidungen vor und innerhalb einer Nutzenbewertung wenigstens transparent gemacht, besser noch ethisch begründet werden. Dazu wird man sich ausdrücklich medizinethischer Expertise versichern müssen. Es wäre ein Irrtum zu glauben, dass Wertbezüge erst auf der Ebene des umfassenden „appraisal“ im G-BA (§ 20 VO) ins Spiel kommen und dass der gesamte Vorlauf „rein wissenschaftlich“ und „wertneutral“ sei.

Wir weisen darauf hin, dass die angedeuteten impliziten Wertentscheidungen oft auch explizit gemacht werden können, z.B. durch qualitative Studien, die bisher in der EbM – ungerechtfertigter Weise – ein Schattendasein führen (cf. Meyer 2007).

4.7 Die Ausformulierung der Forschungsfrage und die Beteiligung von Interessengruppen

Grundsätzlich ist von pluralen, ja konfligierenden Interessen und Werthaltungen der verschiedenen „Stakeholder“ auszugehen. Patienten, Leistungserbringer, Kostenträger, die Industrie, die Politik und Wissenschaftler diverser Disziplinen verfolgen unterschiedliche Ziele, sie haben unterschiedliche Beurteilungsschemata und Interessen, ohne dass eine Perspektive anderen unter allen Umständen übergeordnet wäre. § 35b SGB V schreibt in Abs. 1 Satz 7 eine angemessene Beteiligung verschiedener Interessen(gruppen) vor. § 139a Abs. 5 verweist ausdrücklich auf Sachverständige der Wissenschaft und Praxis, Arzneimittelhersteller, Patienten- und Selbsthilfeorganisationen und den Beauftragten der Bundesregierung für die Belange der Patientinnen und Patienten. Wir regen an, die Beteiligung je nach Auftrag flexibel und großzügig zu gestalten und einen sog. Scoping Workshop vorzusehen, der Gelegenheit gibt, den G-BA-Auftrag in eine möglichst scharf umrissene Forschungsfrage zu überführen. Dabei können die Überlegungen der verschiedenen Interessentengruppen im Sinne wenigstens einer Verfahrens-, eventuell auch einer Beratungsbeteiligung zur Kenntnis zu nehmen (Francke und Hart 2001). Das gestufte Vorgehen von NICE könnte als Vorbild dienen.

Ein solcher Workshop hat das Potential, den Bereich der möglicherweise relevanten Gesichtspunkte und Argumente zu erweitern und zu vertiefen. Die Letztentscheidung bleibt nach § 35b Abs. 1 Satz 5 jedoch beim Institut.

4.8 Einzelfallentscheidungen oder Priorisierung

Ähnlich wie NICE und andere europäische Einrichtungen gleicher Zielsetzung entscheidet der G-BA auf der Basis von Evidenzberichten, systematischen Übersichten, HTAs u.ä. sozusagen von Fall zu Fall. Ansätze für eine Priorisierung (ZEKO 2007) sind bisher kaum zu erkennen. Dies kann zu Konflikten mit dem Prinzip der sozialen Gerechtigkeit Anlass geben: Ein solches Vorgehen führt zu einer zunehmenden Überlastung der gedeckelten finanziellen Ressourcen. So wird das bevölkerungsbezogene Mammographie-Screening in der angestrebten Routine etwa 400 Millionen Euro jährlich kosten, ohne dass klar ist, wodurch diese Mittel kompensiert werden (sollen). Um welche Leistungen und Aufwendungen entsprechenden Umfangs kann die GKV entlastet werden? Hierzu wären Prioritäten und Posterioritäten zu formulieren.

Allerdings ist das IQWiG gehalten, vergleichend zu bewerten. Um die Diskussion von Prioritäten zu befördern, sollte der Kreis der in den Vergleich einzubeziehenden „anderen Behandlungsformen“ möglichst weit gezogen werden und auch verhaltensmedizinische, psychotherapeutische, chirurgische etc. Methoden berücksichtigen. Dies würde eine vertikale Priorisierung in einem begrenzten Indikationsbereich beinhalten.

5. Gesundheitsbezogene Lebensqualität als Kriterium zur Nutzenbewertung und als Parameter in der gesundheitsökonomischen Kosten-Nutzwert-Analyse

5.1. Grundlagen

Die Analyse des Behandlungserfolgs anhand der im englischen Sprachgebrauch als „patient-reported outcomes“ (PRO) bezeichneten Zielgrößen hat in der klinisch-evaluativen Forschung in den vergangenen zwei Jahrzehnten eine ständig wachsende Bedeutung gefunden. Diese Entwicklung lässt sich besonders eindrücklich an der Anzahl der wissenschaftlichen Veröffentlichungen über Studien ablesen, in denen Zielgrößen aus dem Bereich der gesundheitsbezogenen Lebensqualität verwendet wurden (Garratt et al. 2002). Dass die Bedeutung dieser Zielgrößen auch aus regulatorischer Sicht anerkannt wurde, lässt sich z. B. durch die Veröffentlichung des Diskussionspapiers der European Medicines Agency über die Verwendung von Verfah-

ren zur Messung der gesundheitsbezogenen Lebensqualität (EMEA 2005) oder der von der US-amerikanischen Food and Drug Administration herausgegebenen Anleitung zur Verwendung von PRO-Instrumenten in der Entwicklung medizinischer Produkte (FDA 2006) belegen.

In der internationalen Literatur herrscht ein weitgehender Konsens darüber, dass mit dem Konstrukt der gesundheitsbezogenen Lebensqualität eine mehrdimensionale subjektive Beurteilung des Gesundheitszustands und der Auswirkungen der Krankheit und ihrer Behandlung auf das körperliche Befinden, die psychische Verfassung und die soziale Situation gemeint ist. In diesem Sinne wird die (gesundheitsbezogene) Lebensqualität in §35b des SGB V explizit als Kriterium zur Beurteilung des „Patientennutzens“ aufgeführt. Im Methodenpapier (Vers. 2.0) des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen wird dementsprechend die gesundheitsbezogene Lebensqualität als Zielgröße zur Bewertung des „allgemeinen patientenrelevanten medizinischen Nutzens“ genannt (IQWiG 2006).

Die Verfahren zur Messung der gesundheitsbezogenen Lebensqualität können nach ihrer Art und dem Verwendungszweck der Messergebnisse unterschiedlichen Gruppen zugeordnet werden: den diagnoseübergreifenden („generischen“), den krankheitsspezifischen und den präferenzbasierten Instrumenten. Bei den generischen und den krankheitsspezifischen Instrumenten handelt es sich in der Regel um Verfahren, die mit psychometrischen Techniken entwickelt wurden und mit denen die verschiedenen Dimensionen der gesundheitsbezogenen Lebensqualität in der Form eines Profils beschrieben werden.

- Die generischen Instrumente können in einem weiten Spektrum von Patientengruppen mit unterschiedlichen Erkrankungsarten eingesetzt werden. Sie sind darauf ausgerichtet, Problembereiche, die bei zahlreichen Krankheiten relevant sind (z.B. Beeinträchtigungen der körperlichen Funktionsfähigkeit, emotionale Beeinträchtigungen, Beeinträchtigungen des sozialen Lebens), in einer einheitlichen Form zu erfassen und ermöglichen dadurch direkte Vergleiche zwischen verschiedenen Patientengruppen und Behandlungsformen.
- Krankheitsspezifische Instrumente sind darauf ausgerichtet, besondere Formen der krankheitsbedingten Beeinträchtigungen, die nur bei einzelnen Erkrankungen auftreten, angemessen zu berücksichtigen und diese in den Messergebnissen abzubilden.

- Die präferenzbasierten Verfahren sind im Allgemeinen diagnoseübergreifend, sie folgen jedoch einem anderen Konstruktionsprinzip als die psychometrisch entwickelten Instrumente. Die Messung der gesundheitsbezogenen Lebensqualität mit präferenzbasierten Methoden resultiert in einem einzigen Messwert, einem Indexwert, der die Gesamtheit der relevanten Beeinträchtigungen abbilden soll. Solche Indexwerte der gesundheitsbezogenen Lebensqualität werden speziell im Zusammenhang mit gesundheitsökonomischen Kosten-Nutzwert-Analysen (s. u.) benötigt.

Die wesentlichen Eigenschaften der unterschiedlichen Instrumente und Verfahren zur Messung der gesundheitsbezogenen Lebensqualität sind in Tabelle 2 dargestellt.

Tabelle 2: Eigenschaften der diagnoseübergreifenden, krankheitsspezifischen und präferenzbasierten Verfahren zur Messung der gesundheitsbezogenen Lebensqualität

Art	Messergebnis	Eigenschaften/Anwendungsbereich
Diagnoseübergreifend („generisch“)	Profil	▪ Umfassende Messung allgemein bedeutsamer Dimensionen ▪ Vergleiche zwischen Interventionen und Indikationen ▪ Gute psychometrische Validierung, Verfügbarkeit von Normwerten
Krankheitsspezifisch	Profil	▪ "Maßgeschneidert" für bestimmte Diagnosegruppen ▪ Bessere Akzeptanz (Patienten, Kliniker) ▪ Evtl. höhere Änderungssensitivität
Präferenzbasiert	Index	▪ Eignung für gesundheitsökonomische Kosten-Nutzwert-Analysen

Alle drei genannten Verfahren zur Messung der gesundheitsbezogenen Lebensqualität können im Rahmen von Analysen des Nutzens diagnostischer und therapeutischer Maßnahmen verwendet werden. Die Auswahl einer angemessenen Methode und eines geeigneten Messinstruments muss sich an den konkreten Fragestellungen und Kontextbedingungen der Analysen orientieren. Für alle drei Verfahren sind international und national gut validierte Erhebungsinstrumente verfügbar und in entsprechenden Sammlungen, Kompendien und Datenbanken dokumentiert (z. B. McDowell 2006, Salek 1999, ProQolid www.qolid.org). Für deren spezielle Anwendung und die Ergebnisauswertung in klinischen Studien kann sowohl auf die weiter oben bereits genannten Empfehlungen der EMEA und der FDA als auch auf wissenschaftliche Publikationen (Fayers et al. 2005, Fairclough 2002, Staquet et al. 1998) zurückgegriffen werden.

5.2. Gesundheitsbezogene Lebensqualität in Kosten-Nutzwert-Analysen

Im Rahmen gesundheitsökonomischer Kosten-Nutzwert-Analysen werden den Kosten einer diagnostischen oder therapeutischen Intervention die in Termini der gesundheitsbezogenen Lebensqualität gemessenen Nutzwerte gegenüber gestellt. Die Messung der Lebensqualität wird mit präferenzbasierten Erhebungsmethoden durchgeführt und soll den Nutzwert eines Gesundheitszustands als Index auf einer Kardinal- bzw. Intervallskala abbilden. Die Skala ist dabei im Allgemeinen so konstruiert, dass der Wert 1 einem perfekten Gesundheitszustand, der Wert 0 dem Tod entspricht. Werte kleiner als 0 sind möglich und beschreiben Gesundheitszustände, die „schlechter als der Tod“ sind.¹²

Die Bestimmung von Nutzwerten eines Gesundheitszustands kann mit direkten Methoden, wie z. B. der Methode des „standard gamble“ (SG), des „time trade-off“ (TTO) oder der „rating scale“ (RS) erfolgen. Das SG-Verfahren gilt als die theoretisch stärkste, das RS-Verfahren als die theoretisch schwächste Methode der direkten Nutzwertbestimmung. Es ist diesen Verfahren gemeinsam, dass die Nutzwertbestimmung durch systematische Erfassung von Wahlentscheidungen zwischen Gesundheitszuständen („präferenzbasiert“) erfolgt. Eine nähere deutschsprachige Be-

¹² Eine abweichende Meinung unter den Verfassern führt hierzu aus: Hier wird die ganze Problematik des Konzepts deutlich. Was maßen wir uns an? Gibt es universell definierbare Gesundheitszustände, die schlechter als der Tod sind?

schreibung der praktischen Durchführung des SG-, TTO- und RS-Verfahrens ist bei Schöffski (2002) zu finden.

Wegen der geringen Praktikabilität und des hohen Erhebungsaufwands der direkten Verfahren werden jedoch indirekte Methoden der Nutzwertbestimmung weitaus häufiger angewandt. Hierbei werden die Nutzwerte für eine große Anzahl vordefinierter Gesundheitszustände unter Verwendung eines direkten Verfahrens im Rahmen einer Vorstudie („Valuierungsstudie“) bestimmt. In späteren Studien muss dann nur noch (z.B. mit Hilfe eines standardisierten Erhebungsinstruments, in der Regel eines schriftlichen Fragebogens) festgestellt werden, welcher der vordefinierten Gesundheitszustände bei einem Patienten vorliegt. Aus den Ergebnissen der Valuierungsstudie kann diesem Gesundheitszustand dann – unter Anwendung eines „Tarifs“ – der entsprechende Nutzwert zugeordnet werden.

Fast alle der international gebräuchlichen Fragebögen zur indirekten Nutzwertmessung stehen in deutscher Übersetzung zur Verfügung (vgl. Tabelle 3). Allerdings wurde nur für den EQ-5D-Fragebogen (EuroQol Group 1990) eine Umrechnung der Fragebogenangaben in Nutzwerte anhand von in Deutschland durchgeführten Valuierungsstudien entwickelt. Bei den übrigen Instrumenten muss die Nutzwertberechnung gegenwärtig noch auf der Grundlage von „Tarifen“ aus anderen Ländern erfolgen.

Die mit direkten oder indirekten Methoden erhobenen Nutzwerte können mit Angaben zur Lebenszeit verknüpft werden. Durch Berücksichtigung des in den Perioden einer definierten Lebenszeit vorliegenden Nutzwertes ergeben sich „qualitätskorrigierte Lebensjahre“ (quality-adjusted life years, QALYs), in die sowohl die Zeitkomponente (Quantität des Lebens) als auch die Nutzwert-Komponente (Qualität des Lebens) eingehen. Erst die Berechnung von QALYs bzw. der Differenz von QALYs unter verschiedenen Behandlungsszenarien ermöglicht in der gesundheitsökonomischen Kosten-Nutzwert-Analyse eine Betrachtung des Verhältnisses von Kosten und Nutzen einer medizinischen Intervention.¹³

¹³ Eine abweichende Meinung unter den Verfassern führt dazu aus: Während unbestritten ist, dass das Konzept der Lebensqualität in der Situation konkreter klinischer Studien eine Bereicherung ist – ermöglicht es doch patientenrelevanten Nutzen neben rein medizinischen Endpunkten zu berücksichtigen -, so ergeben sich in unserem setting einige Probleme. Zunächst einmal ist Lebensqualität über unterschiedliche Indikationsgebiete nicht oder nur sehr eingeschränkt vergleichbar. Man denke etwa an chronischen Kopfschmerz versus Krebserkrankung im finalen Stadium. Darüber hinaus handelt es sich immer um eine Orientierung am statistisch ermittelbaren Nutzen. Stellen wir uns beispielsweise für den Vergleich zweier Krebstherapien vor, dass das eine Medi-

Tabelle 3: Übersicht über standardisierte, in deutscher Sprache verfügbare Fragebögen zur indirekten Nutzwertbestimmung

Instrument	Acronym	Methode	Anzahl Fragen	Quelle
Short-Form Health State Classification	SF-6D	SG	10	Brazier et al. 2002, 2004
Quality of Well-Being Scale	QWB-SA	?	72	Kaplan et al. 1997
EuroQol	EQ-5D	TTO, RS	5	EuroQol Group 1990
15D	15D	RS	15	Sintonen 1995
Health Utilities Index	HUI	SG, RS	15	Torrance et al. 1996

5.2.1. Besonderheiten und offene Fragen der Lebensqualitätsmessung für Kosten-Nutzwert-Analysen

1. Die zur Durchführung der indirekten Methode der Nutzwertbestimmung erforderliche Valuierungsstudie kann in unterschiedlichen Kollektiven, insbesondere in Kollektiven aus der Allgemeinbevölkerung oder in Patientenkollektiven durchge-

kamant eine gute 2-Jahresüberlebensrate mit wenig Nebenwirkungen in dieser Zeit und anschliessend hoher Rezidivwahrscheinlichkeit, das andere eine hohe 5-Jahresüberlebensrate bei intermittierender Chemotherapie, verbunden mit kurzzeitig sehr eingeschränkter Lebensqualität verspricht. Die Beurteilung des Nutzens für eine junge Mutter mit Kindern im Alter von 11, 12 und 15 wird sicher anders ausfallen als für eine 78 jährige Frau oder einen 78 jährigen Mann mit vorbestehender koronarer Herzerkrankung. Aber nicht nur der Nutzen hängt von der individuellen Situation ab, sondern auch die Kosten, wenn man über die unmittelbaren Behandlungskosten hinausgeht. Im Fall der älteren Patienten sind Kosten aufgrund von Multimorbidität, im Fall der Mutter Kosten, die sich z.B. aus der Betreuung der Kinder ergeben, von Relevanz. Diese Überlegungen zeigen, dass bei der Verwendung von Lebensqualitätsindikatoren in der Kosten-Nutzenbewertung immer auch geprüft werden muss, ob für die spezifische Indikation von einer hinreichenden Homogenität der Kosten- und Nutzenindikatoren über die relevante Patienten- gruppe ausgegangen werden kann

führt werden. Vergleichsstudien zeigen, dass sich die Ergebnisse der Valuationsstudie je nach dem gewählten Referenzkollektiv unterscheiden. Nach eigenen Befunden (Moock et al. 2005) und publizierten Ergebnissen (Polsky et al. 2001) scheinen Personen aus der Allgemeinbevölkerung insbesondere die schlechten Gesundheitszustände ungünstiger einzuschätzen als Patienten.

Nach der heute von vielen Experten geteilten Auffassung sollte die Valuationsstudie in der Allgemeinbevölkerung durchgeführt werden, da diese Personengruppe das medizinische Versorgungssystem finanziert und in wirtschaftlicher Hinsicht von Allokationsentscheidungen auf der Basis von Nutzwertanalysen unmittelbar betroffen ist. Demgegenüber kann argumentiert werden, dass eine angemessene Beurteilung des Nutzwertes spezifischer Gesundheitszustände in besonderen Fällen nur von Patienten erfolgen kann, die mit den relevanten gesundheitlichen Beeinträchtigungen aus eigener Erfahrung vertraut sind. Werden Nutzwerte durch Methoden der direkten Messung bestimmt, sind es in der Regel ohnehin Patienten, die in diesem Zusammenhang befragt werden.

Eine Lösung dieser Kontroverse ist noch nicht in Sicht. Während die Vorgaben des britischen NICE klar den Bezug auf die Allgemeinbevölkerung favorisieren (NICE 2004), konstatieren die kanadischen Guidelines (CADTH 2006), dass die Allgemeinbevölkerung aus den oben genannten Gründen zwar die beste Referenz bietet, es wird jedoch nicht ausgeschlossen, in begründeten Fällen auch Patienten als Referenzkollektiv heranzuziehen. Im Idealfall, so die kanadischen Guidelines, sind Referenzwerte sowohl aus der Allgemeinbevölkerung als auch aus entsprechenden Patientenkollektiven verfügbar. Die Auswirkungen der Verwendung unterschiedlicher Referenzkollektive auf die Ergebnisse können dann im Rahmen von Sensitivitätsanalysen untersucht und beschrieben werden. Aus wissenschaftlicher und gesundheitspolitischer Sicht wäre es allerdings wichtig, genauere Informationen darüber zu erhalten, welche Gründe für die Unterschiede zwischen den in den Referenzkollektiven erhaltenen Ergebnissen verantwortlich sind. Die Entscheidung für oder gegen ein bestimmtes Referenzkollektiv könnte dadurch besser empirisch untermauert und die Erhebung von Präferenzurteilen in dem gewählten Kollektiv methodisch verbessert werden (Brazier et al. 2005).

2. Die Übertragbarkeit von Ergebnissen der gesundheitsökonomischen Evaluation von einem Land in ein anderes kann durch verschiedene Faktoren beeinträchtigt werden. Zu diesen Faktoren gehören z. B. die Krankheitshäufigkeit, die Preise und Verfügbarkeit medizinischer Leistungen und die Existenz landesspezifischer Behandlungsmuster (Boulenger et al. 2005). Im Falle von Kosten-Nutzwert-Analysen kann die Übertragbarkeit von Ergebnissen zusätzlich dadurch eingeschränkt sein, dass bei der indirekten Methode zur Nutzwertmessung ein Tarif verwendet wurde, der für das Ursprungsland der gesundheitsökonomischen Studie gilt und sich von dem für das eigene Land geltenden Tarif unterscheidet. Im Extremfall existiert für das eingesetzte indirekte Verfahren der Nutzwertmessung noch kein Tarif im eigenen Land, so dass unbekannt bleibt, ob und in welchem Umfang eine Vergleichbarkeit überhaupt besteht.

Für den EQ-5D-Fragebogen, eines der am häufigsten verwendeten Verfahren zur indirekten Nutzwertmessung, ergaben sich in einem Vergleich von deutschen mit britischen Ergebnissen aus Valuierungsstudien in der Allgemeinbevölkerung substantielle Unterschiede in der Beurteilung der Gesundheitszustände (Claes et al. 1999). Busschbach et al. (2003) zeigten in einem Vergleich der deutschen, britischen und spanischen Tarife des EQ-5D, dass die beobachteten Unterschiede nicht völlig durch Strukturunterschiede in der soziodemographischen Zusammensetzung der Referenzkollektive erklärt werden können. Ein Vergleich der Ergebnisse der britischen und der US-amerikanischen Valuierungsstudien ergab ähnliche Resultate (Johnson et al. 2005).

Als Konsequenz aus diesen Befunden ergibt sich die Notwendigkeit, die Übertragbarkeit von in anderen Ländern mit der Nutzwertmessung erzielten Ergebnisse sorgfältig zu prüfen. Bei einer Verwendung solcher Ergebnisse z. B. bei Allokations- oder Erstattungsentscheidungen, ist diese Prüfung besonders wichtig. Es existieren bereits Versuche, international harmonisierte Tarife zu entwickeln, die eine einheitliche Berechnung von Nutzwerten in multi-nationalen Studien ermöglichen (Greiner et al. 2003). Weitere Studien unter Anwendung dieser Tarife müssen aber belegen, dass die nationalen Besonderheiten in ihnen ausreichend gut repräsentiert werden.

3. Im Kontrast zur großen Anzahl der verfügbaren diagnoseübergreifenden und krankheitsspezifischen Verfahren zur psychometrischen Messung der gesund-

heitsbezogenen Lebensqualität existiert nur eine kleine Zahl von Instrumenten zur indirekten Nutzwertmessung (vgl. Tabelle 2). Unmittelbare Vergleiche („head-to-head“) der mit diesen präferenzbasierten Instrumenten erzielten Messergebnisse zeigen, dass trotz der einheitlichen Skalennormierung (perfekte Gesundheit: 1, Tod: 0) merkliche Unterschiede zwischen den mit verschiedenen Instrumenten erfassten Nutzwerten bestehen. Es findet sich in der Regel zwar eine ausreichend hohe lineare Korrelation zwischen den Messergebnissen, diese stimmen jedoch in relevanten Parametern (theoretischer Wertbereich, Verteilungscharakteristika, Änderungssensitivität) nicht so gut überein, dass die Instrumente als äquivalent gelten könnten. Entsprechende Ergebnisse aus Deutschland zu diesem Sachverhalt aus einem Vergleich aller fünf in Tabelle 2 aufgeführten Instrumente wurden von Moock et al. (2005) veröffentlicht. Ein Review von Kopec et al. (2003) illustriert anhand hypothetischer Fälle die methodischen Gründe für die geringe Übereinstimmung der mit verschiedenen Instrumenten bestimmten Nutzwerte.

4. Insgesamt ist aus diesen Befunden zu schließen, dass auf Nutzwerte bezogene Ergebnisse, die mit unterschiedlichen Instrumenten generiert wurden, nicht unmittelbar vergleichbar sind. Dies gilt mit hoher Wahrscheinlichkeit auch für die Vergleichbarkeit von Differenzen zwischen Nutzwerten, die in der gesundheitsökonomischen Kosten-Nutzwert-Analyse betrachtet werden. Hervorzuheben ist an dieser Stelle jedoch, dass die präferenzbasierten Instrumente trotz mangelnder Äquivalenz der Messwerte hinreichend und vergleichbar gut in der Lage sind, zwischen Personen und Patientengruppen mit unterschiedlich stark ausgeprägter Erkrankungsschwere zu diskriminieren (z.B. Sach et al. 2007, Barton et al. 2007a, 2007b, Marra et al. 2005). Dies deutet darauf hin, dass diesen Instrumenten ein gemeinsames Konstrukt zugrunde liegt, das jedoch nicht auf einer einheitlichen Skala abgebildet wird. Die im Vergleich zu anderen Instrumenten der Lebensqualitätsmessung große Anzahl von Ergebnissen aus „head-to-head“-Vergleichen - Janssen et al. (2007) konnten allein 23 Studien identifizieren, in denen der EQ-5D-Fragebogen mit dem HUI- und/oder dem SF-6D-Index verglichen wurde – bietet darüber hinaus die Möglichkeit, die Unterschiede zwischen den mit verschiedenen Instrumenten gemessenen Nutzwerten nach Richtung und Umfang zu beurteilen und im Rahmen von Nutzenbetrachtungen zu berücksichtigen.

5. Die vergleichende Bewertung von medizinischen Interventionen anhand der Betrachtung qualitätskorrigierter Lebensjahre (QALYs) geht vom Zutreffen verschiedener Voraussetzungen aus. Zu diesen Voraussetzungen gehören einige „technische Axiome“ (der additiven Separierbarkeit, des konstanten proportionalen Trade-Offs und der Risiko-Neutralität). Darüber hinaus muss für die vergleichende Bewertung anhand von QALYs in der gegenwärtig zumeist eingenommenen Perspektive die Unabhängigkeit der QALYs von Kontextbedingungen gelten. Diese Voraussetzung wird in der internationalen Literatur unter dem Motto „a QALY is a QALY is a QALY“ diskutiert. Sie besagt, dass ein berechnetes qualitätskorrigiertes Lebensjahr unabhängig von allen anderen Merkmalen des Kontext jeweils den selben „Wert“ hat, also z. B. unabhängig von Alter, Krankheits-schwere, bisheriger und zukünftiger Krankheitsdauer ist. Diese Prämisse gilt allerdings nicht nur für die hier diskutierte Betrachtung der qualitätskorrigierten Lebensjahre. Auch bei alleiniger Berücksichtigung der Lebensjahre ohne Einbezug von Nutzwerten würde die Unabhängigkeit des „Werts“ dieser Jahre von Kontextbedingungen erfüllt sein müssen.

In empirischen Studien und Übersichten konnte verschiedentlich gezeigt werden, dass die genannten Voraussetzungen, sowohl die technischen Axiome (Bleichrodt et al. 1997, Dolan et al. 2003) als auch die Unabhängigkeit von Kontextbedingungen, empirisch nicht in vollem Umfang zutreffen. Nach den Ergebnissen von Dolan et al. (2005) wird der „Wert“ von QALYs in der Gesellschaft insbesondere in Abhängigkeit von Krankheitsmerkmalen (u. a. Erkrankungsschwere und Krankheitsverlauf) beurteilt. Daneben kommen bei der gesellschaftlichen Beurteilung von QALYs Aspekte der Verteilungsgerechtigkeit zum Tragen, die in herkömmlichen Analysen nicht berücksichtigt werden. Zur Lösung dieser Probleme sind bereits verschiedene Ansätze vorgeschlagen worden. Diese reichen von Methoden der statistischen Modellierung (z. B. Krabbe et al. 1998) bis hin zur Einführung von empirisch gewonnenen „equity weights“ (Dolan et al. 2006).

6. Zur Planung und Durchführung von Studien, in denen Nutzwerte und QALYs bestimmt werden, sowie zur Ergebnispräsentation der entsprechenden Ergebnisse existieren keine international konsentierten Empfehlungen oder Leitlinien. Es ist insofern verständlich, dass in solchen Studien sowohl unterschiedliche methodische Vorgehensweisen als auch unterschiedliche Arten der Ergebnisdarstellung gewählt werden. Besonders die Darstellung der Methodik und der Ergebnisse ist

nach verschiedenen Reviews bei Kosten-Nutzwert-Analysen hinsichtlich Informationsgehalt und Transparenz verbesserungswürdig (Neumann et al. 2000, Richardson et al. 2004, Griebisch et al. 2005). In ihrer Übersicht über 18 in Deutschland durchgeführten Kosten-Nutzwert-Analysen konnten Schwappach und Boluarte (2007) die bereits in den früheren Reviews identifizierten Schwächen bestätigen. Während die identifizierten Defizite in der methodischen Anlage und der Berichterstattung dieser Studien die Anwendbarkeit und Aussagekraft der Kosten-Nutzwert-Analyse nicht grundsätzlich in Frage stellen, weisen sie doch deutlich auf einen erheblichen Bedarf an Harmonisierung und Standardisierung der Methodik der Nutzwertmessung und der Präsentation ihrer Ergebnisse in Veröffentlichungen hin.

Die dargestellten Besonderheiten der Lebensqualitätsmessung im Kontext von Kosten-Nutzwert-Analysen lassen erkennen, dass bei ihrer Anwendung eine Reihe offener Fragen und Probleme bestehen. Die präferenzbasierte Lebensqualitätsmessung teilt im Hinblick auf die Probleme der Vergleichbarkeit, Übertragbarkeit und der messtheoretischen Prämissen das Schicksal anderer Zielgrößen (z. B. der psychometrischen Lebensqualitätsmessung und der Erfassung von Kosten). Die Fragen und Probleme der Nutzwertmessung wurden vielfach und werden auch gegenwärtig in der wissenschaftlichen Forschung diskutiert und untersucht.

Bei der Nutzwertmessung und der Verwendung dieser Zielgröße in gesundheitsökonomischen Analysen handelt es sich um ein vergleichsweise junges, aber sehr dynamisches Forschungsfeld, in dem das Konstrukt der qualitätskorrigierten Lebensjahre trotz aller Einschränkungen eine führende Rolle spielt. Alternativen zur Nutzwertbestimmung (z. B. das Verfahren des „person trade-off“, Nord 1995) oder zur Berechnung qualitätskorrigierter Lebensjahre (z. B. „disability-adjusted life years“ DALY, Murray et al. 1997; „healthy-year equivalents“ HYE, Gafni et al. 1993; „saved young life equivalents“ SAVE, Nord 1992) konnten ihre Überlegenheit bisher nur begrenzt unter Beweis stellen. Empirische Ergebnisse zu diesen Alternativen liegen bisher nur sehr eingeschränkt vor, sie werden in der gesundheitsökonomischen Evaluation – im Unterschied zu qualitätskorrigierten Lebensjahren – faktisch kaum eingesetzt.

Die gegenwärtige Dominanz des QALY-Ansatzes wird in einer vergleichenden Übersicht über die Maßgaben der gesundheitsökonomischen Evaluation pharmazeutischer Produkte in 9 europäischen und außereuropäischen Ländern unterstrichen. In

7 der Länder wurden qualitätskorrigierte Lebensjahre als bevorzugte Ergebnisgröße in der Kosten-Nutzwert-Analyse identifiziert (Zentner et al. 2006, 2005).

Zunehmende Aufmerksamkeit besonders unter Gesundheitsökonominnen haben in der letzten Zeit Ansätze erfahren, mit denen die Nutzenkomponente direkt in monetären Einheiten erfasst werden können und die insofern einen Übergang von der Kosten-Nutzwert- zur Kosten-Nutzen-Analyse darstellen. Trotz einer wachsenden Anzahl internationaler Publikationen über Ansätze der „Zahlungsbereitschaft“ (Klose 1999) sind die Erfahrungen mit diesen Methoden in Deutschland noch begrenzt.

6. Inhaltliche und Methodische Aspekte der Kosten-Nutzen-Bewertung

In diesem Papier kann kein umfassender Überblick über die international anerkannten Standards der Gesundheitsökonomie gegeben werden; hierfür sei auf die einschlägige Literatur verwiesen (etwa: CADTH 2006; Drummond et al 2005; Gold et al 1996). Im Folgenden sollen vielmehr einige zentrale Aspekte, die in der jüngsten deutschen Diskussion zu Kontroversen geführt haben, erörtert werden.

6.1. Die Perspektive einer gesundheitsökonomischen Evaluation

Bei Durchführung einer gesundheitsökonomischen Bewertung ist die Definition der einzunehmenden Perspektive von wesentlicher Bedeutung. Darüber, welche Perspektive das IQWiG bei der Durchführung seiner gesundheitsökonomischen Bewertungen einnehmen sollte, sind in der letzten Zeit kontroverse Ansichten geäußert worden.

Bei der „Perspektive“ geht es um die Entscheidung, aus wessen Blickwinkel das IQWiG den Vergleich des zu untersuchenden Medikamentes mit anderen medizinischen Interventionen vornehmen soll. Grundsätzlich bezieht sich dies sowohl auf die Effekt-Seite („Nutzen“), als auch auf die Kosten-Seite, allerdings wird in der Debatte hierüber zumeist nur auf die Kostenseite abgestellt. In der internationalen wie nationalen gesundheitsökonomischen Diskussion haben sich hierbei insbesondere die folgenden Perspektiven mit Blick auf die Kostenseite herausgebildet:

- *Patienten-Perspektive:* Bei Einnahme der Patienten-Perspektive werden Kosten von Krankheit und Behandlung nur soweit in die Analyse einbezogen, wie sie die Patienten selber zu tragen haben. Sofern Behandlungen Gegenstand des Leistungskataloges des sozialstaatlichen Gesundheitssystems (in Deutschland: der GKV) sind, bedeutet dies oft, dass nur geringe Anteile der entstehenden Kosten aus dieser Perspektive zu berücksichtigen sind. Oft – insbesondere bei chronischen Erkrankungen – fallen aber auch in hohem Maße Kosten bei Patienten an – etwa Fahrtkosten zur Behandlung, krankheitsbedingte Kosten einer Haushaltshilfe, Selbstbeteiligungen, vorübergehende oder permanente Einkommensausfälle, nicht-erstattungsfähige Leistungen (z.B. ausgeschlossene Arzneimittel) oder nicht zugelassene Leistungserbringer (z.B. Heilpraktiker). Volkswirtschaftliche Kosten des Arbeitsausfalls werden nur insoweit berücksichtigt, wie die Patienten durch Minderung ihres verfügbaren Einkommens davon betroffen sind.
- *Krankenkassen-Perspektive:* Bei Einnahme dieser Perspektive werden Kosten soweit berücksichtigt, wie die GKV für sie aufzukommen hat. Aufgrund von Selbstbeteiligungen, Leistungsausgrenzungen, Budgetierungen etc. stimmt diese Perspektive häufig nicht mit den tatsächlichen Kosten, die im Gesundheitswesen entstehen, überein. Auch werden Kosten, die bei der Pflegeversicherung oder anderen (sozialstaatlichen wie privaten) Trägern von Gesundheitsausgaben anfallen, nicht berücksichtigt. Volkswirtschaftliche Kosten des Arbeitsausfalls werden nur insoweit berücksichtigt, wie die Krankenkassen mit Krankengeldzahlungen beteiligt sind.¹⁴
- *Gesundheitssystem-Perspektive:* In diesem, in jüngerer Zeit verstärkt angewendeten, Ansatz werden alle direkten medizinischen Kosten, die anfallen, berücksichtigt, unabhängig davon, wer sie zu tragen hat. In der Gesundheitssystem-Perspektive werden daher Behandlungskosten, die von den Krankenkassen gezahlt werden, ebenso berücksichtigt, wie Kosten, die etwa die Pflegekassen oder der Patient selber aufwenden. Dagegen werden volkswirtschaftliche Kosten des Arbeitsausfalls nicht berücksichtigt, da sie nicht als Teil des Gesundheitssystems angesehen werden. Auch nicht-medizinische Kosten (z.B. Fahrtkosten, aber z.B. auch Kosten im Bildungssystem oder anderen gesellschaftlichen Bereichen) werden nicht berücksichtigt.

¹⁴ Vgl. zur Operationalisierung der GKV-Perspektive in Deutschland: Krauth, et al. 2005.

- *Gesellschaftliche Perspektive:* Die gesellschaftliche Perspektive gilt als die umfassendste Sicht auf die Effekte von Krankheit und Behandlung. Sowohl direkte medizinische Kosten als auch nicht-medizinische Kosten (wie z.B. krankheitsbedingte Haushaltshilfe, Fahrtkosten, aber auch Kosten z.B. im Bildungssystem) wie schließlich auch volkswirtschaftliche Kosten des Arbeitsausfalles werden grundsätzlich in die Analyse einbezogen, und zwar unabhängig davon, wer diese Kosten jeweils zu tragen hat. Aus der gesellschaftlichen Perspektive wird eine Bewertung der Ressourcenverbräuche nach ihren Opportunitätskosten (Kosten entsprechen dem Nutzen in der nächst best möglichen Verwendungsmöglichkeit) angestrebt.
- Für die deutsche Situation wird schließlich auch diskutiert:
- *Sozialversicherungsperspektive:*¹⁵ Die Sozialversicherungsperspektive ist breiter als die Perspektive der gesetzlichen Krankenversicherung, weil etwa auch die Kosten oder Ersparnisse, die bei der Pflegeversicherung anfallen, oder die Kosten von Frühverrentung bei den gesetzlichen Rentenversicherungsträgern berücksichtigt werden. Demgegenüber werden von den Patienten zu leistende Zahlungen, nicht-medizinische Kosten oder Produktivitätsausfälle, die nicht bei den Sozialversicherungsträgern anfallen (sondern von den Patienten, den Arbeitgebern oder allgemein von der Gesellschaft zu tragen sind), nicht berücksichtigt.

Nationale (z.B. Hannoveraner Konsensgruppe 1999) wie internationale (z.B. Gold et al. 1996) wissenschaftliche gesundheitsökonomische Guidelines empfehlen regelmäßig, *in jedem Falle auch* die gesellschaftliche Perspektive zu berücksichtigen. Dies ist von der Überlegung getragen, dass die wirtschaftlichste Verwendung knapper Mittel dann zu erreichen ist, wenn ein möglichst breiter Blick auf die Wirkungen von Krankheit und Behandlung geworfen wird. Die Einnahme der gesellschaftlichen Perspektive muss als der in der internationalen *wissenschaftlichen gesundheitsökonomischen Community* präferierte Standard gelten.

Bei den Agenturen, die in anderen Ländern Entscheidungen über Ressourcenallokationen treffen oder hierfür vorbereitende Evaluationen durchführen müssen, ist das Bild hingegen uneinheitlich (Zentner u. Busse 2006): teilweise wird prioritär die Kostenträgerperspektive eingenommen, teilweise die gesellschaftliche Perspektive. Soweit die Kostenträgerperspektive eingenommen wird, wird verschiedentlich (etwa von

¹⁵ Vgl. etwa IQWiG, Methoden. Version 2.0 vom 19.12.2006, S. 58.

der kanadischen Agentur CADTH 2006) darauf hingewiesen, dass getrennt davon die Ergebnisse einer weiter gefassten Perspektive berichtet werden sollen, wenn zu erwarten ist, dass dies die Entscheidung beeinflussen könnte.

Vertreter der Krankenkassen haben in jüngerer Zeit vorgeschlagen, das IQWiG solle ausschließlich die GKV-Perspektive berücksichtigen und dies unter anderem mit dem hohen Stellenwert, den auch das Bundesverfassungsgericht der Beitragssatzstabilität einräumt, begründet.¹⁶ Die pharmazeutische Industrie vertritt demgegenüber die Position: „Kosten-Nutzen-Bewertungen sind grundsätzlich aus Sicht der gesellschaftlichen Perspektive durchzuführen.“¹⁷ Das IQWiG (2006: 58) vertritt eine Position ähnlich derjenigen (oben erwähnten) der CADTH (2006), wenn es in seinem Methodenpapier 2.0 ausführt: „Für die gesetzliche Aufgabenstellung des IQWiG nimmt die Kostenbetrachtung aus der Perspektive der gesetzlichen Krankenversicherung ...eine besondere Stellung ein. Über die GKV-Perspektive hinaus wird das Institut in Abhängigkeit von der Fragestellung die für eine Bewertung der Wirtschaftlichkeit geeignete Perspektive wählen. So kann es z.B. sinnvoll sein, bei der Bewertung von Interventionen, die sich auf Rehabilitation oder Arbeitsunfähigkeit auswirken, von der Perspektive der Sozialversicherungsträger oder von einer gesamtgesellschaftlichen Perspektive auszugehen.“

Nach Auffassung der Verfasser dieses Papiers muss sichergestellt sein, dass der G-BA und (soweit die Festsetzung eines Höchstbetrages für Arzneimittel betroffen ist) der (künftige) Spitzenverband Bund der Krankenkassen in die Lage versetzt werden, bei ihren Entscheidungen regelhaft Auswirkungen auf andere Ausgabenbereiche jenseits der GKV zu berücksichtigen, sofern nicht unbestritten solche Auswirkungen irrelevant sind. Der geeignete Ort darüber zu entscheiden, ob solche Auswirkungen bereits apriori als irrelevant eingeschätzt werden können, ist nach unserer Auffassung ein beim G-BA im Kontext der Auftragsvergabe anzusiedelnder Scoping-Workshop: Sofern in diesem Kontext nicht überzeugend dargelegt wird, dass Kosteneffekte jenseits der GKV nicht zu erwarten sind, sollte eine über die GKV hinausgehende Perspektive eingenommen werden. Damit ist selbstredend *keine* Bindung des G-BA oder des Spitzenverbandes Bund der Krankenkassen auf diese Perspekti-

¹⁶ So etwa Vertreter der GKV auf der Fachtagung im BMG am 18. Juni 2007.

¹⁷ Vgl. Meyer, H.-W.: Vorschlag des VFA zur Umsetzung der Regelungen des GKV-WSG zur isolierten Nutzenbewertung und zur Kosten-Nutzen-Bewertung. Download unter: http://www.bmg.bund.de/nn_1145310/DE/Themenschwerpunkte/Gesundheit/Arzneimittel/meier-vfa,templateId=raw,property=publicationFile.pdf/meier-vfa.pdf

ve bei der *Entscheidung* verbunden – wenn also etwa die Bewertung durch das IQWiG zeigt, dass höheren Ausgaben für die GKV deutlich höhere Entlastungen an anderer Stelle gegenüber stehen, so können G-BA oder Spitzenverbände sich gleichwohl entscheiden, nur die GKV-Ausgaben in ihr Entscheidungskalkül einzubeziehen – sie sollten dies jedoch zumindest *wissentlich* tun.

6.2. Nutzen in einer gesundheitsökonomischen Evaluation

Die Bewertung des „Kosten-Nutzen-Verhältnisses“ (§ 35b Abs. 1 Satz 1 SGB V) soll nach dem Willen des Gesetzgebers erfolgen „durch Vergleich mit anderen Arzneimitteln und Behandlungsformen unter Berücksichtigung des therapeutischen Zusatznutzens für die Patienten im Verhältnis zu den Kosten“ (§ 35b Abs. 1 Satz 3 SGB V).

Eine Bewertung, bei der ein „Kosten-Nutzen-Verhältnisses“ ermittelt wird, ist in der Ökonomie üblicherweise ein Verfahren, bei dem nicht nur die Kosten sondern auch der Nutzen in Geldeinheiten (Euro) ausgedrückt wird. Der „Charme“ eines solchen Ansatzes liegt u.a. darin, dass eine direkte Gegenüberstellung von zwei Geldbeträgen (für Kosten und für Nutzen) resultiert; damit wird ein direkter Vergleich möglich, ob der Nutzen die Kosten überwiegt (Sen 2000). Solche Verfahren sind auch in der Gesundheitsökonomie in den letzten Jahren entwickelt worden (Johannesson 1996). Insbesondere sind hierbei Verfahren zum Einsatz gekommen, in denen die (fiktive) „Zahlungsbereitschaft“ der Patienten/Versicherten für den Zusatznutzen erfasst und bewertet und den Kosten gegenübergestellt wird. Hierbei wird entweder nach der Bereitschaft für die Therapie selber zu zahlen oder nach der Bereitschaft, wegen der Aufnahme der Therapie in den Leistungskatalog der Krankenkassen höhere Beiträge zu zahlen, gefragt.¹⁸

Es ist allerdings davon auszugehen, dass der Gesetzgeber diesen Studientypus der „Kosten-Nutzen-Analyse“ (cost-benefit-analysis) bei der Neuformulierung von § 35b SGB V nicht vor Augen hatte. Zum einen stellt die Kosten-Nutzen-Analyse noch nicht den international üblichen Standard dar, sondern befindet sich – ungeachtet der großen methodischen Fortschritte, die hier in den letzten Jahren erzielt worden sind – im Gegensatz zu Kosten-Effektivitäts- oder Kosten-Nutzwert-Analysen noch eher in einem experimentellen Stadium. Zum anderen gibt der Gesetzgeber selber einen Hin-

¹⁸ Umgekehrt kann auch gefragt werden, um wieviel die Patienten *entschädigt* werden müssten, wenn die Leistung nicht zur Verfügung steht bzw. nicht in den Leistungskatalog der GKV aufgenommen wird.

weis auf sein Verständnis von Nutzen, indem er in § 35b Abs. 1 Satz 4 SGB V ausführt: „Beim Patienten-Nutzen sollen insbesondere die Verbesserung des Gesundheitszustandes, eine Verkürzung der Krankheitsdauer, eine Verlängerung der Lebensdauer, eine Verringerung der Nebenwirkungen sowie eine Verbesserung der Lebensqualität,... angemessen berücksichtigt werden.“ Zwar wäre es möglich, die vom Gesetzgeber genannten Dimensionen des Patienten-Nutzens im Rahmen einer Kosten-Nutzen-Analyse monetär zu bewerten, es erscheint jedoch zweifelhaft, ob der Gesetzgeber eine solche monetäre Bewertung beabsichtigt hat. Im Folgenden wird daher davon ausgegangen, dass *keine monetäre Bewertung* im Rahmen der gesundheitsökonomischen Bewertungen im Kontext der Aufträge für das IQWiG beabsichtigt ist.

Grundsätzlich können in einer gesundheitsökonomischen Evaluation, die international üblichen Standards entspricht, sehr unterschiedliche Parameter auf der Outcome-Seite erhoben (und den Kosten gegenübergestellt) werden, etwa im Rahmen von Kosten-Effektivitäts-Studien auch Veränderungen von rein klinischen Parametern (z.B. Veränderungen bestimmter Laborwerte), denen möglicherweise keine direkte Patientenrelevanz zukommt. Der gesetzliche Auftrag bezieht sich allerdings insbesondere auf solche Ergebnisparameter, die eine Patientenrelevanz (wozu auch valide Surrogatparameter zu rechnen sind; vgl. dazu oben Abschn. 3) aufweisen. Zutreffend weist daher das IQWiG (2006: 62) in seinem Methodenpapier 2.0 darauf hin, dass es in gesundheitsökonomischen Evaluationen grundsätzlich Endpunkte verwenden wird, die sich am Patientennutzen orientieren. Dabei verweist das Institut auf den entsprechenden Abschnitt des Methodenpapiers zur Nutzenbewertung in der Medizin.

Damit ist eine grundsätzliche Problematik angesprochen, zu der sich in den letzten Monaten eine Kontroverse in der deutschen Gesundheitspolitik entwickelt hat: nämlich nach dem Verhältnis einer isolierten Nutzenbewertung zur Bewertung des Nutzens im Rahmen einer gesundheitsökonomischen Bewertung. Hier sind sehr heterogene Vorstellungen in die Diskussion eingebracht worden: So hat der Vorsitzende des Gemeinsamen Bundesausschusses vorgeschlagen, bei neuartigen Arzneimitteln grundsätzlich ein „zweistufiges Verfahren“ vorzusehen, bei dem in der ersten Stufe eine isolierte Nutzenbewertung und daran anschließend in der zweiten Stufe eine

Kosten-Nutzen-Bewertung durchgeführt wird.¹⁹ Er plädiert dafür in die Nutzenbewertung bereits „ökonomische Komponenten“ zu integrieren, so dass dann die anschließende Kosten-Nutzen-Bewertung auf die Frage der Kosten und des angemessenen Verhältnisses von Nutzen und Kosten begrenzt werden kann. Der Leiter des IQWiG geht davon aus: „Die Basis für die Kosten-Nutzen-Bewertung ist zunächst die Nutzen-Bewertung“,²⁰ spricht hierbei jedoch nicht davon, dass bei dem ersten Schritt ökonomische Komponenten integriert werden sollten; nach seiner Auffassung bedarf es, wenn im ersten Schritt eine isolierte Nutzenbewertung durchgeführt wurde, im zweiten Schritt nur noch der Erhebung von Kosten. Demgegenüber vertritt der Verband Forschender Arzneimittelhersteller die These: „Eine Kosten-Nutzen-Bewertung kann nicht auf einer Nutzenbewertung aufbauen, da es sich hierbei um zwei verschiedene Konzepte handelt.“²¹ Aus den Erfahrungen entsprechender Institutionen im internationalen Vergleich können hier keine Schlussfolgerungen gezogen werden, da in den meisten Ländern vom Gesetzgeber entweder eine Nutzen-Bewertung oder eine Kosten-Nutzen-Bewertung vorgesehen ist, hingegen nicht beide Instrumente nebeneinander bestehen und eingesetzt werden.

Ob die isolierte Nutzenbewertung bereits zugleich die erste Stufe einer Kosten-Nutzen-Bewertung sein kann oder für die Kosten-Nutzen-Bewertung eine von einer isolierten Nutzenbewertung getrennte Analyse durchzuführen ist, hängt wesentlich davon an, ob eine isolierte Nutzenbewertung insoweit den vom Gesetzgeber eingeforderten internationalen Standards der Gesundheitsökonomie zur Nutzen-Komponente einer Kosten-Nutzen-Bewertung entsprechen würde.

Zunächst gilt auch für die gesundheitsökonomische Evaluation: „A sound clinical review of the interventions should form the basis of the evaluation. This should involve a systematic review of the available evidence on the efficacy and effectiveness of the intervention.“ (CADHTA 2006: 20) Der EbM-Grundsatz, hierbei die beste verfügbare Evidenz heranzuziehen, gilt auch für gesundheitsökonomische Evaluationen ohne Abstriche. Es gilt, auf dieser Basis den Zusatznutzen einer Intervention gegenüber

¹⁹ http://www.bmg.bund.de/nn_1145310/DE/Themenschwerpunkte/Gesundheit/Arzneimittel/hess-gba,templateId=raw,property=publicationFile.pdf/hess-gba.pdf

²⁰ Sawicki, P.: IQWiG Thesen zur Fachtagung: „Einführung der Kosten-Nutzen-Bewertung in Deutschland.“ Download unter:
http://www.bmg.bund.de/nn_1145310/DE/Themenschwerpunkte/Gesundheit/Arzneimittel/sawicki-iqwig,templateId=raw,property=publicationFile.pdf/sawicki-iqwig.pdf

²¹ Vgl. Meyer, H.-W.: Vorschlag des VFA zur Umsetzung der Regelungen des GKV-WSG zur isolierten Nutzenbewertung und zur Kosten-Nutzen-Bewertung. Download unter:
http://www.bmg.bund.de/nn_1145310/DE/Themenschwerpunkte/Gesundheit/Arzneimittel/meier-vfa,templateId=raw,property=publicationFile.pdf/meier-vfa.pdf

einer Vergleichsintervention zu bestimmen (und dann der Kostendifferenz gegenüber zu stellen). Gesundheitsökonomische Evaluationen zielen hierbei grundsätzlich auf den „net effect“ – unerwünschte Nebeneffekte (Risiken, Nebenwirkungen) sind daher zu berücksichtigen (Gold et al 1996: 135ff.).

Im Gegensatz zur Beurteilung des isolierten medizinischen Nutzens, der sich an dem konfirmatorischen Nachweise der medizinischen Wirksamkeit unter streng vorgegebenen Studienbedingungen orientiert, ist es das Ziel gesundheitsökonomischer Studien, deren Ergebnisse in Erstattungsentscheidungen einfließen, die Routine-Versorgung der Patienten abzubilden. Dieser Gewinn an externer Validität mit Blick auf die Relevanz für das Versorgungsgeschehen bedingt in vielen Fällen „Zugeständnisse“ bei der „internen“ Validität und damit eine weniger restriktive Auswahl bei den berücksichtigten Studientypen. Beispielhaft wird etwa die Compliance in RCTs oftmals nicht realitätsgerecht abgebildet, so dass Studien mit geringerer interner Validität (open label) angezeigt sein mögen. Auch Registerdaten u.ä. können relevant sein, z.B. wenn sie Informationen über die Alltagspopulationen, die bestimmte Interventionen bekommen, enthalten. In gesundheitsökonomischen Evaluationen wird die Evidenz zur Efficacy (aus RCTs) mit der Evidenz zu Effectiveness (aus anderen Datenquellen) verknüpft. Eine typische methodische Herangehensweise hierfür sind Modellierungen.

Ein weiteres Unterscheidungsmerkmal zwischen dem Nutzen in einer isolierten Nutzenbewertung und dem Outcome einer Kosten-Nutzen-Bewertung stellt der Zeitraum dar, der in der Studie berücksichtigt wird. Wesentliches Charakteristikum einer internationalen Standards entsprechenden gesundheitsökonomischen Kosten-Nutzen-Bewertung ist, dass Nutzen (und Kosten) grundsätzlich über einen so langen Zeitraum untersucht werden, dass sämtliche für den Alternativenvergleich relevanten – auch langfristigen – Effekte einbezogen werden. Denn es geht um einen wirtschaftlichen Einsatz knapper Mittel – dieser aber muss periodenübergreifend optimiert werden (Gold et al 1996: 139): „The analysis is carried out over a time horizon of as many years as necessary to capture the main effects...of the intervention.“

Wenn die gesundheitsökonomische Analyse Bestandteil einer klinischen Studie ist, wird häufig ein für den gesundheitsökonomischen Zweck zu kurzfristiger Zeithorizont festgelegt, um langfristige Aussagen tätigen zu können. Oftmals können keine hinreichend langen Zeiträume in einer Studie überblickt werden, um die aus gesundheits-

ökonomischer Sicht notwendige langfristige Betrachtung aus diesen Daten zu generieren. Wird etwa davon ausgegangen, dass die Lebenserwartung sich in den untersuchten Alternativen unterscheidet, wird dieser Effekt zumeist nicht vollständig in dem Zeitfenster der Untersuchung eintreten.

Immer dann, wenn der Zeitraum, in dem Kosten und Effekte anfallen, den Erhebungszeitraum übersteigt, ist es ebenfalls notwendig, die primäre Datenerhebung (die ihrerseits patientenrelevante intermediäre Endpunkte oder auch „nur“ – hinreichend valide – Surrogatparameter zum Inhalt gehabt haben kann) um Modellanalysen (z.B. Markov-Modelle, Discret Event Simulationen) zu ergänzen, in denen unterschiedlichste Datenquellen (z.B. epidemiologische Datenbestände, Registerdaten) miteinander verknüpft werden. Erhebliche Bedeutung kommt hierbei einer möglichst validen Schätzung der Eintrittswahrscheinlichkeiten der verschiedenen Outcomes im Zeitstrom zu: „For estimation of effectiveness in a CEA²², outcome probability values should be selected from the best designed (and least biased) sources that are relevant to the question and population under study“ (Gold et al 1996: 141).

Teilweise wird empfohlen, bei der Analyse zwei Zeithorizonte zu berücksichtigen: einen kurz- bis mittelfristigen, bei dem für den Alternativenvergleich nur Primärdaten genutzt werden, und einen langfristigen, bei dem auch Modelldaten einbezogen werden (AG Reha-Ökonomie 1999). Allerdings wird auch vorgeschlagen, bei chronischen Erkrankungen regelhaft primär einen lebenslangen Zeithorizont für die Analyse zu verwenden und den Effekt der Verwendung alternativer (kürzerer) Zeithorizonte in einer Sensitivitätsanalyse zu untersuchen (CADTH 2006: 23).

Modellierungen stellen damit einen zentralen Bestandteil gesundheitsökonomischer Evaluationen dar. Sie dienen zum einen dem Ziel, den Schritt von Efficacy zu Effectiveness zu vollziehen, wo dies nicht mit primären Studien möglich ist. Zum anderen sind sie erforderlich, um die notwendige Langfristigkeit der Betrachtung zu ermöglichen. Ein Verzicht auf Modellierungen in einer Kosten-Nutzen-Bewertung durch das IQWiG würde in den Fällen, in denen Modellierungen zur Erreichung der genannten Zwecke zielführend sind, nicht den internationalen Standards der Gesundheitsökonomie – und damit einer Vorgabe des Gesetzgebers – entsprechen.

Allerdings gilt zugleich auch: „Decision makers should be critical when reviewing the results of model-based evaluation“ (CADTH 2006: 23). Modellieren bedarf einer um-

²² CEA: Cost-effectiveness Analyses

fassenden sowohl methodischen als auch auf die jeweilige Intervention bezogenen fachlichen Expertise. Modelle müssen für den jeweiligen Zweck adäquat ausgestaltet sein, die Auswahl und Art der Verwendung der unterschiedlichen Daten muss angemessen sein, Modellstruktur und –ergebnisse müssen transparent und nachvollziehbar berichtet werden. Es liegen mehrere international akzeptierte Guidelines zur „Praxis guter Modellierung“ vor (Weinstein et al 2003, Philips et al 2004), an denen sich Modelle für den Zweck der Kosten-Nutzen-Bewertung in der GKV orientieren sollten.

Die vorstehenden Ausführungen haben das Verhältnis der isolierten Nutzenbewertung zur Betrachtung der Outcome-Seite im Rahmen der Kosten-Nutzen-Bewertung deutlich gemacht: Insoweit in der isolierten Nutzenbewertung nach dem Grundsatz der besten verfügbaren Evidenz bereits ebenfalls unter Nutzung von Studien mit einer höheren externen (und oft niedrigeren internen) Validität, ggfs. die unterschiedlichen Datenbasen über Modellierung miteinander verbindend, über die Untersuchung der Efficacy hinausgehend die Effektivität unter Alltagsbedingungen untersucht wurde, und sofern in der „reinen“ Nutzenbewertung bereits (ebenfalls wiederum Daten verschiedener interner Validität berücksichtigend, die ggfs. über Modellierung miteinander verknüpft werden) dem Langfristcharakter gesundheitsökonomischer Evaluationen Rechnung getragen wurde, ist die isolierte Nutzenbewertung als erster Baustein einer internationalen Standards der Gesundheitsökonomie entsprechenden Kosten-Nutzen-Bewertung verwendbar. Andernfalls kann die Kosten-Nutzen-Bewertung nicht auf einer isolierten Nutzenbewertung aufbauen, sondern ist der „Nutzenteil“ spezifisch für die gesundheitsökonomische Bewertung durchzuführen.

6.3. Kosten in einer gesundheitsökonomischen Evaluation

Dem (Zusatz-)Nutzen einer Intervention im Vergleich zu einer anderen Intervention sind in einer gesundheitsökonomischen Evaluation die (Zusatz-)Kosten im Vergleich zu den Kosten der Vergleichsintervention gegenüber zu stellen.

Es entspricht dem internationalen Standard in der Gesundheitsökonomie, dass hierbei sämtliche Kosten berücksichtigt werden müssen, die aus der jeweiligen Perspektive (siehe dazu oben) relevant sind. Dies bedeutet insbesondere, dass nicht nur die Kosten der zur Untersuchung anstehenden Intervention selber (also etwa: die Kosten

des zur Evaluation stehenden Arzneimittels und der Vergleichsmedikation) zu berücksichtigen sind. Vielmehr sind in jedem Falle auch bei der (sehr engen) Perspektive der gesetzlichen Krankenversicherung die weiteren direkten medizinischen Kosten ebenso wie direkte nicht-medizinische Kosten (z.B. Fahrtkosten, Haushaltshilfen) und indirekte Kosten (Krankengeld) soweit zu berücksichtigen, wie die Krankenkassen sie tragen und soweit sie in Folge der Erkrankung und Therapie eine Rolle spielen. Führt also – um dies zu illustrieren – ein neues Arzneimittel im Vergleich zur bisherigen Standardtherapie zu einer größeren Wahrscheinlichkeit der Vermeidung von schwerwiegenden Ereignissen, so sind die entsprechend daraus resultierenden verringerten Behandlungskosten der schwerwiegenden Ereignisse zu berücksichtigen. Auch Kosten, die bei der Behandlung von Nebenwirkungen auftreten, sind – unter Berücksichtigung der Eintrittswahrscheinlichkeit dieser Nebenwirkungen – zu berücksichtigen. Die Kostendifferenz zwischen Intervention und Vergleichstherapie bezieht sich auf die Aggregation der eigentlichen Interventionskosten mit den hier angesprochenen Kosten von Folge-, Fern- und Nebenwirkungen.

Wird eine weiter gefasste Perspektive als die der gesetzlichen Krankenversicherung eingenommen, sind entsprechend auch Kosten, die bei anderen Akteuren entstehen, zu berücksichtigen, bei der gesellschaftlichen Perspektive schließlich unabhängig davon, bei wem sie entstehen.

Auch für Kosteninformationen gilt grundsätzlich, dass die Qualität der Daten möglichst hoch sein sollte. Die Kosten stehen – wie gesehen – grundsätzlich in einem engen Zusammenhang zur Nutzenseite: Wenn sich zwischen zwei Alternativen die Outcomes unterscheiden, wird dies oftmals auch mit unterschiedlichen Ressourcenverbräuchen einhergehen – so führt beispielhaft ein durch ein neues Medikament zusätzlich verhinderter Schlaganfall zur Vermeidung der mit dem Schlaganfall assoziierten Kosten. Trotz des angesprochenen engen Zusammenhanges zwischen Nutzen- und Kosten-Seite wird es auch dort, wo in der Bewertung der Nutzenseite Daten aus randomisierten klinischen Studien eine zentrale Rolle spielen, oftmals nicht möglich sein, die Kostendaten direkt aus solchen Studien zu entnehmen. Dies liegt unter anderem darin begründet, dass multizentrische klinische Studien insbesondere für die Zulassung neuer Arzneimittel aber auch zum Nachweis des Zusatznutzens etwa von Medizinprodukten häufig multinational angelegt sind.

Schon die Frage, inwieweit die Ergebnisse multinationaler Studien auf der Nutzen-Seite ohne weiteres in gesundheitsökonomische Evaluationen für ein bestimmtes nationales Gesundheitssystem übernommen werden können, ist nicht trivial (Welte u. Leidl 1999); bei den Ressourcenverbräuchen kommt allerdings hinzu, dass sie aufgrund der Unterschiedlichkeit von Gesundheitssystemen oft nicht übertragbar sein dürften. So ist etwa die Schnittstelle zwischen der ambulanten und der stationären Behandlung in unterschiedlichen Gesundheitssystemen verschieden gezogen – inwieweit also eine bestimmte Intervention (z.B. im Zusammenhang mit – vermiedenen – Folgeerkrankungen) zu einem stationären Krankenhausaufenthalt oder zu ambulanten Arzt-Kontakten führt, wird nicht immer einheitlich sein. Dies gilt auch für die hierbei in Anspruch genommenen „Mengen“ an Gesundheitsleistungen, wie etwa die im internationalen Vergleich sehr unterschiedliche durchschnittliche Krankenhausverweildauer deutlich macht. Schließlich sind auch die nationalen Preise für identische medizinische Leistungen höchst unterschiedlich – und zwar sowohl unter einer Kostenträgerperspektive als auch der gesellschaftlichen Perspektive eigenen Sicht auf die Bewertung der Ressourcenverbräuche nach dem Prinzip der Opportunitätskosten.

Es ist aber schon aus Kostengründen für die Studiendurchführung kaum vertretbar, multinationale Studien in allen Ländern, in denen später Erstattungsentscheidungen Kostendaten verlangen, mit einer hinreichend großen Fallzahl durchzuführen, so dass aus diesen Studien die landesspezifischen Kostendaten generiert werden könnten. Daher ist es in solchen Fällen unvermeidlich, die Kostendaten auch aus anderen Datenquellen zu generieren. Auch entstehen in den RCTs oftmals Kostenkomponenten, die mit dem Studienprotokoll zusammenhängen (z.B. zusätzliche Angiographien bei bestimmten kardiologischen RCTs). Hierzu können für die Mengengerüste Daten aus Registern, von Krankenkassen, aus Leitlinien, ggfs. auch aus Expertenbefragungen genutzt werden; die Preisinformationen können – auch in Abhängigkeit von der Perspektive – insbesondere aus Abrechnungskatalogen ermittelt werden. Hierzu besteht – trotz aller Unterschiede im Detail, die den Verschiedenartigkeiten der Gesundheitssysteme geschuldet sind – zu den grundsätzlichen Vorgehensweisen ein hohes Maß an zu einem internationalen Standard geronnenen Konsens.²³

²³ Vgl. etwa Gold et al. (1996); Drummond et al. (2005). Vorschläge für unterschiedliche Detaillierungsgrade in Deutschland siehe in AG Rehaökonomie (1999) sowie Krauth, C. et al. (2005).

6.4. Diskontierung von Kosten und Nutzen

Es entspricht dem internationalen Standard in der Gesundheitsökonomie, dass künftige Kosten und Nutzen diskontiert, also abgezinst, werden. Die Praxis der Diskontierung künftiger Kosten und Nutzen ist darin begründet, dass Ökonomen davon ausgehen, dass Menschen bzw. Gesellschaften eine positive Zeitpräferenzrate haben, sprich: heutige Nutzen künftigen Nutzen vorziehen und künftige Kosten heutigen Kosten vorziehen. Der Umfang der Diskontierung hat Einfluss darauf, inwieweit künftige Kosten und Outcomes geringer gewichtet werden als heutige Kosten und Outcomes.

Es bestehen in der internationalen Diskussion keine einheitlichen Vorstellungen darüber, welchen Betrag die Diskontrate annehmen soll. Das britische NICE verwendet seit 2004 eine Diskontrate von 3,5 % für Kosten und Nutzen (NICE 2004). Das von der US-Administration eingesetzte Panel empfahl eine Diskontrate von 3 % (Gold et al 1996). Dies entspricht auch der Empfehlung der AG Reha-Ökonomie (1999) für Deutschland. Das US-Panel empfahl, aufgrund der großen Zahl von Studien, die mit einer Diskontrate von 5 % gearbeitet hatten, auch diese anzuwenden. Die international viel beachteten Empfehlungen von Drummond et al. (2005) sprechen sich wie das US-Panel für 3 % und 5 % aus. Eine Rate von 5 % entspricht auch der Empfehlung der kanadischen Guideline (CDHTA 2006) sowie des Hannoveraner Konsenses (1999).

Vereinzelt wird in der gesundheitsökonomischen Literatur die Position vertreten, Kosten sollten mit einer größeren Rate diskontiert werden sollten als Outcomes (Brower et al. 2005). Das britische NICE hat dies bis 2004 praktiziert. Weit überwiegend wird aber aus einer Reihe von systematischen Argumenten eine Gleichbehandlung von Kosten und Outcomes mit Bezug auf die Diskontrate gefordert.

Übereinstimmend empfehlen alle Guidelines, den Einfluss von unterschiedlichen Diskonraten durch Sensitivitätsanalysen (bei denen z.B. die Diskontrate von 0 % bis 7 % variiert wird) zu untersuchen.

6.5. Bildung einer Kosten-Effektivitäts-Relation

Über die Ermittlung des Zusatznutzens und der Zusatzkosten hinaus, ist wesentliche Aufgabe einer gesundheitsökonomischen Bewertung, beide Größen einander gege-

nüber zu stellen. Aus dieser Gegenüberstellung resultiert die „inkrementelle Kosten-Effektivitäts-Relation“ (IKER), die sich als Quotient von (diskontierten) Zusatzkosten und (diskontierten) Zusatznutzen darstellt. Sind also etwa die durch eine neue Intervention zusätzlich verhinderten Schlaganfälle im Vergleich zur gegenwärtigen Standardintervention Gegenstand der Untersuchung, dann sagt die IKER aus, welche zusätzlichen Kosten für die Verhinderung eines zusätzlichen Schlaganfalles gegenüber der Standardintervention bei Anwendung der neuen Intervention aufgewendet werden müssen. Handelt es sich um eine Untersuchung aus der Perspektive der gesetzlichen Krankenversicherung, so entsprechen die zusätzlichen Kosten zusätzlichen Ausgaben der Krankenkassen; bei der gesellschaftlichen Perspektive sind Zusatzkosten unterschiedlicher Akteure angesprochen, wobei die Budgetrelevanz für die gesetzliche Krankenversicherung zusätzlich ermittelt werden kann.

Der Gesetzgeber spricht in § 35b Abs. 1 Satz 4 SGB V davon, dass vom Institut im Rahmen der Gegenüberstellung von Kosten und Nutzen „auch die Angemessenheit und Zumutbarkeit einer Kostenübernahme durch die Versichertengemeinschaft angemessen berücksichtigt“ werden soll. Damit wird eine zentrale Aufgabe der gesundheitsökonomischen Evaluation deutlich: Im Ergebnis steht – sofern die Festsetzung eines Erstattungshöchstbetrages bei neuen Arzneimitteln durch den Spitzenverband Bund der GKV angesprochen ist – eine Entscheidung, bis zu welchem Betrag die GKV den vom pharmazeutischen Unternehmen vorgesehenen Preis übernehmen soll. Diese Entscheidung hängt wesentlich vom dem in der gesundheitsökonomischen Evaluation ermittelten Verhältnis zwischen Zusatznutzen und Zusatzkosten ab.

In Deutschland ist bislang – im Gegensatz zu anderen Ländern – keine Debatte darüber geführt worden, bis zu welchen Zusatzkosten ein Zusatznutzen als „angemessen und zumutbar“ angesehen wird. Eine solche Diskussion, die letztlich in einer auch ethisch fundierten Reflexion über den Wert des menschlichen Lebens mündet, muss aufgrund der gesetzlichen Vorgaben nunmehr geführt werden. Sie ist dann besonders komplex, wenn sich die ökonomische Bewertung durch das IQWiG nicht nur auf einen universellen Outcome-Parameter wie QALYs beschränkt, sondern bei unterschiedlichen Indikationen unterschiedliche Outcome-Parameter, hinsichtlich derer der Zusatznutzen gemessen und den Zusatzkosten gegenübergestellt werden soll, Verwendung finden.

Das Gesetz kann so gelesen werden, als ob die wertende Entscheidung, ob „Angemessenheit und Zumutbarkeit einer Kostenübernahme durch die Versichertengemeinschaft“ vorliegt, durch das IQWiG getroffen werden soll. Dies erscheint allerdings wenig glücklich. Ein Beispiel soll dies illustrieren: Das IQWiG finde in seiner gesundheitsökonomischen Evaluation heraus, dass ein neues Arzneimittel den Eintritt von Erblindung bei einer bestimmten Erkrankung gegenüber dem bisherigen Standard im Durchschnitt der Patienten um 1 Jahr hinausschieben kann (Zusatznutzen gegenüber Standard); dies geht jedoch unter Berücksichtigung des vom pharmazeutischen Unternehmens angesetzten Arzneimittelpreises und der sonstigen Kosten mit zusätzlichen Kosten je Patienten gegenüber dem bisherigen Standard in Höhe von 22.000 Euro einher. Die Entscheidung, ob die Kostenübernahme durch die Versichertengemeinschaft bei dieser Kosten-Nutzen-Relation – die IKER beträgt 22.000 Euro je vermiedenem Jahr Blindheit – angemessen und zumutbar ist, ist primär gesellschaftlicher Natur. Solche Entscheidungen gehören nach unserer Auffassung 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. Das Institut sollte den Zusatznutzen und die Zusatzkosten in den verschiedenen Bereichen / Perspektiven darstellen, sich jedoch einer Empfehlung, ob diese Relation wirtschaftlich oder unwirtschaftlich ist, enthalten.

In jedem Falle gibt es für die Bewertung der Angemessenheit und Zumutbarkeit der Kostenübernahme keine internationalen Standards, wenn man von Kriterien wie Vernünftigkeit, Logik, interne Konsistenz etc. absieht. Auch die Tatsache, dass Größenordnungen wie etwa 30.000 £ je gewonnenem QALY für England (Devlin u Parker 2004; Culyer 2006) oder 50.000 \$ für die Privatversicherungen in den USA verschiedentlich als „Schwellenwerte“ empirisch identifiziert oder konzeptionell diskutiert werden, weist nicht auf internationale Standards hin sondern auf nationale Kontexte, in denen solche bewertenden Entscheidungen getroffen werden müssen. Standards gibt es „nur“ für die sachlichen Grundlagen der Bewertungen, d.h. die Evidenzbasis, ihre Quellen, Suchstrategien, Überprüfung auf Verzerrungen etc.. Auch in diesen sachlichen Grundlagen sind im Übrigen zahlreiche Werturteile enthalten, z.B. hinsichtlich der Vergleichsinterventionen, kritischen Differenzen, Fehlerniveaus, Messzeitpunkte.

7. Anforderungen an den Bewertungsprozess

In den vorangegangenen Abschnitten dieses Papiers haben wir uns mit inhaltlichen Aspekten der Nutzen- bzw. Kosten-Nutzen-Bewertung befasst. In diesem Abschnitt sollen einige Überlegungen zu den Anforderungen an den *Prozess* der Bewertungen angeschlossen werden. Der Gesetzgeber des GKV-WSG hat – auch in Reaktion auf Einwürfe Beteiligter zum bisherigen Verfahren – nunmehr 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). Zudem hat er (in § 139a Abs. 5 SGB V) festgelegt, dass diese Beteiligung „in allen wichtigen Abschnitten des Bewertungsverfahrens“ zu erfolgen hat. Der Begründung des entsprechenden Änderungsantrages ist zu entnehmen, dass das Institut damit insbesondere auch verpflichtet wird, „eine entsprechende Beteiligung vor Festlegung des Berichtsplans und den darauf folgenden Zwischen- und Endberichten durchzuführen.“²⁴

In den letzten Monaten ist die Frage intensiver diskutiert worden, auf welche Weise die Umsetzung eines Auftrages des Gemeinsamen Bundesausschusses in ein Studiendesign, also: einen Berichtsplan, des IQWiG erfolgen soll. Der Verband Forschender Arzneimittelhersteller (2007) hat einen Verfahrensvorschlag vorgelegt. Zentral ist in diesem Konzept, dass vor Beginn des eigentlichen Bewertungsverfahrens ein sogenannter „Scoping-Workshop“ durchgeführt werden soll, wie er etwa auch von NICE in England praktiziert wird. Auch in dem Gutachten von der Schulenburg (2007) für den VFA wird ein solcher Scoping-Workshop positiv beurteilt.

Die Fachtagung des BMG am 18.06. hat deutlich gemacht, dass sowohl die Vertreter des IQWiG als auch der Krankenkassen einem solchen Scoping-Workshop kritisch gegenüberstehen. Im Kern wurde argumentiert, man sei bisher auch ohne ein solches Instrument ausgekommen, das den Prozess nur komplizierter mache ohne wirklich benötigt zu werden. Die Verfasser dieses Papiers teilen diese kritische Einschätzung nicht. Nach unserer Auffassung ist ein solcher Workshop ein geeignetes Mittel, um die Beteiligungsrechte der Beteiligten auch bei der Operationalisierung und ersten Planung eines IQWiG-Auftrages zu wahren. Aus unserer Sicht sollte allerdings nicht das IQWiG sondern der Gemeinsame Bundesausschuss solche Sco-

²⁴ BT-Drucksache 16/4247, Zu Nummer 20, S. 32.

ping-Workshops regelhaft nach Bekanntgabe eines Auftrages an das IQWiG durchführen. Wir regen an, dass der Gesetzgeber eine entsprechende Verpflichtung für den Gemeinsamen Bundesausschuss normiert.

Mit den nunmehr für Berichtsplan und Berichte getrennten Stellungnahme- und Anhörungsmöglichkeiten hat der Gesetzgeber grundsätzlich wichtige Schritte für eine Mitwirkung der Beteiligten geschaffen. Das IQWiG hat in seinem Methodenpapier 2.0 sowie in mehreren „Leitfäden“, die im Internet herunterladbar sind, nähere Einzelheiten zu den Stellungnahmeverfahren festgelegt. Folgende Anmerkungen dazu erscheinen den Verfassern dieses Papieres sinnvoll:

- Das IQWiG legt in der Regel eine Frist von 4 Wochen für die Stellungnahme zu Berichtsplänen und Berichtsentwürfen fest. Insbesondere bei komplexen Aufträgen mit vielfältigen Facetten, die zu umfangreichen Dokumenten (500 S. u. mehr) führen, scheint diese Frist knapp.
- Für die Stellungnahmen gibt es Seitenbegrenzungen (5 Seiten bei Stellungnahme zu Berichtsplänen, 10 Seiten zu Berichtsentwürfen). Dies erscheint je nach Komplexität der zu kommentierenden Dokumente ebenfalls zumindest in Einzelfällen nicht sachgerecht und ist international (z.B. bei NICE) nicht vorgesehen.
- Mündliche wissenschaftliche Erörterungen sind nach dem Methodenpapier für das Institut „optional“. Sie sollen nur dann durchgeführt werden, wenn das Institut aufgrund der schriftlichen Stellungnahmen Besprechungsbedarf hat. Dies erscheint keine angemessene Beteiligungspraxis. Eine mündliche Erörterung wird international im allgemeinen durchgeführt. Bei ihr sollten (mindestens) alle Sachverständigen, die eine schriftliche Stellungnahme abgegeben haben, eingeladen werden.
- Da der Gesetzgeber wünscht, dass „insbesondere auch... international anerkannte Sachverständige“ in das Beteiligungsverfahren eingebunden werden sollen,²⁵ sollte sowohl für das schriftliche Stellungnahmeverfahren als auch für die mündlichen Anhörungen neben Deutsch Englisch als Verhandlungssprache zugelassen werden.

²⁵ Vgl. BT-Drucksache 16/3100, Besondere Begründung zu Nr. 20 (§ 35b).

8. Literatur

AG Reha-Ökonomie (Hessel et al) (1999): Gesundheitsökonomische Evaluation in der Rehabilitation. Teil I: Prinzipien und Empfehlungen für die Leistungserfassung. In: Verband Deutscher Rentenversicherungsträger (Hrsg.): Förderschwerpunkt „Rehabilitationswissenschaften“. Empfehlungen der Arbeitsgruppen „Generische Methoden“, „Routinedaten“ und „Reha-Ökonomie“. DRV-Schriften Bd. 16. Frankfurt, S. 106-193.

Antes, G., Diener, M. (2006). The Role of Systematic Reviews in Evidence-based Healthcare. Chin J Evid-based Med, Vol. 6(7), 468-9.

Baker SG(2006): Editorial: Surrogate Endpoints: Wishful Thinking or Reality. Journal of the National Cancer Institute, Vol 98, 502-503

Barton GR, Sach TH, Doherty M, Avery AJ, Jenkinson C, Muir KR. An assessment of the discriminative ability of the EQ-5D(index), SF-6D, and EQ VAS, using sociodemographic factors and clinical conditions. Eur J Health Econ 2007.

Barton GR, Sach TH, Avery AJ, Jenkinson C, Doherty M, Whynes DK et al. A comparison of the performance of the EQ-5D and SF-6D for individuals aged ≥ 45 years. Health Econ 2007.

Beauchamp TL, Childress JF: Principles of Biomedical Ethics. New York, Oxford, Oxford University Press 4. Aufl. 1994,

Bekelmann, J.E., Li, Y., Gross, C.P. (2003). Scope and impact of financial conflicts of interest in biomedical research: a systematic review. JAMA, 289(4), 454-65.

Black N: Why we need observational studies to evaluate the effectiveness of health care. BMJ 312 (1996) 1215-1218

Bleichrodt H, Johannesson M. The validity of QALYs: an experimental test of constant proportional tradeoff and utility independence. Med Decis Making 1997; 17(1):21-32.

Boulenger S, Nixon J, Drummond M, Ulmann P, Rice S, de Pouvourville G. Can economic evaluations be made more transferable? Eur J Health Econ 2005; 6(4):334-346.

Brazier J, Akehurst R, Brennan A, Dolan P, Claxton K, McCabe C et al. Should patients have a greater role in valuing health states? Appl Health Econ Health Policy 2005; 4(4):201-208.

Brouwer, W. B. F. et al. (2005). "Need for differential discounting of costs and health effects in cost effectiveness analysis." *British Medical Journal*. 331: S. 446-448.

Busschbach J, Weijnen T, Nieuwenhuizen M, Oppe S, Badia X, Dolan P et al. A comparison of EQ-5D time trade-off values obtained in Germany, the United Kingdom and Spain. In: Brooks R, Rabin R, de Charro F, editors. The measurement and valuation of health status using EQ-5D: A european perspective. Evidence from the EuroQol BIOMED research programme. Kluwer Academic Publishers, Dordrecht, Boston, London, 2003: 143-165.

Canadian Agency for Drugs and Technologies in Health (CADTH). HTA. Guidelines for the economic evaluation of health technologies: Canada. 3 ed. Canadian Agency for Drugs and Technologies in Health (CADTH), Ottawa, 2006.

Centre for Evidence-Based Medicine (2001). Levels of Evidence. Oxford. <http://www.cebm.net/index.aspx?o=1025>

Chalmers I, Matthews R: What are the implications of optimism bias in clinical research? Lancet 367 (2006)449-450

Claes Ch, Greiner W, Uber A, Schulenburg J-MGvd. An interview-based comparison of the TTO and VAS values given to EQ-5D states of health by the general german population. In: Greiner W, Schulenburg J-MGvd, Piercy J, editors. Plenary Meeting. Hannover 1998. 1st-2nd October. Discussion papers. Uni-Verlag Witte, Hannover, 1999: 13-74.

Culyer, A. J. (2006). "NICE's use of cost effectiveness as an exemplar of a deliberative process." *Health Economics, Policy and Law* 1: 299-318

Devlin, N. and D. Parkin (2004). "Does NICE have a cost-effectiveness Threshold and what other factors influence its decisions? A binary choice analysis." *Health Economics* 13(5): 437-452

Dolan P, Stalmeier P. The validity of time trade-off values in calculating QALYs: constant proportional time trade-off versus the proportional heuristic. *J Health Econ* 2003; 22(3):445-458.

Dolan P, Shaw R, Tsuchiya A, Williams A. QALY maximisation and people's preferences: a methodological review of the literature. *Health Econ* 2005; 14(2):197-208.

Dolan P, Tsuchiya A. The elicitation of distributional judgements in the context of economic evaluation. In: Jones AM, editor. *The elgar companion to health economics*. Edward Elgar, Cheltenham, 2006: 382-391.

Drummond, M., M. Sculpher, G. Torrance, B. O'Brien und G. Stoddart (2005). *Methods for the Economic Evaluation of Health Care Programmes*. Oxford, Oxford University Press.

Egger, M., Smith, G. D., Altman, D. G.(eds.) (2001). *Systematic Reviews in Health Care: Meta-Analysis in Context* (2nd ed.).BMJ Books.

European Medicines Agency (EMA). Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products. 2005. London.

EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy* 2007; 16:199-208.

Fairclough DL. Design and analysis of quality of life studies in clinical trials. *Interdisciplinary statistics*. Chapman & Hall/CRC, Boca Raton, London, New York, Washington D.C., 2002.

Fayers P, Hays R. *Assessing quality of life in clinical trials*. 2 ed. Oxford University Press, Great Clarendon Street, Oxford OX2 6DP, 2005.

Fletcher, R. H., Fletcher, S. W. (2005). *Clinical Epidemiology The Essentials* (4th ed.)

Food and Drug Administration (FDA). Guidance for industry. Patient-reported outcome measures: Use in medical product development to support labeling claims. Draft guidance. 2006.

Francke R, Hart D: *Bürgerbeteiligung im Gesundheitswesen*. Baden-Baden, Nomos 2001

Gafni A, Birch S, Mehrez A. Economics, health and health economics: HYE (healthy-years equivalent) versus QALYs (quality-adjusted live-year). *J Health Econ* 1993; 12(3):325-339.

Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. *BMJ* 2002; 324(7351):1417-1421.

Gemeinsamer Bundesausschuss: *Verfahrensordnung des Gemeinsamen Bundesausschusses*. Siegburg 2005, geänderte Fassung vom 07.07.2006

Glasziou P, Chalmers I, Rawlins M, McCulloch P: When are randomised trials unnecessary? Picking signal from noise. *BMJ* 334 (2007) 349-351

Gold, M. R., J. E. Siegel, L. B. Russel und M. C. Weinstein (1996). *Cost-Effectiveness in Health and Medicine*. New York & Oxford, Oxford University Press

GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004; 328(7454): 1490-1494

Gray, M. (2001). *Evidence-Based Healthcare*. Churchill Livingstone.

Green SB (1982). Patient heterogeneity and the need for randomized controlled clinical trials. *Controlled Clinical Trials*, 3, 189-198.

Green, SB, Byar, DP. (1984). Using observational data from registries to compare treatments: the fallacy of omnimetrics. *Statistics in Medicine*, 3, 361-370.

Greiner W, Weijnen T, Nieuwenhuizen M, Oppe S, Badia X, Busschbach J et al. A single European currency for EQ-5D health states. Results from a six-country study. *Eur J Health Econ* 2003; 4(3):222-231.

Griebsch I, Coast J, Brown J. Quality-adjusted life-years lack quality in pediatric care: a critical review of published cost-utility studies in child health. *Pediatrics* 2005; 115(5):e600-e614.

Grimes, D. A., Schulz, K. (2002). Bias and causal associations in observational research. *The Lancet*, Vol. 359, 248 – 252.

Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ. GRADE: An Emerging Consensus on Rating Quality of Evidence and Strength of Recommendations. (BMJ, accepted for publication, pending revision)

Hannoveraner Konsens Gruppe (1999). "Deutsche Empfehlungen zur gesundheitsökonomischen Evaluation - Revidierte Fassung des Hannoveraner Konsens." *Gesundheitsökonomie und Qualitätssicherung* 4: A62-A65

Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD et al: Current methods of the U. S: Preventive Services Task Force: A Review of the Process. *Am J Prev Med* 20, suppl. 1 (2003) 21-35

Haynes, B., Sackett, L., Guyatt, H. G., Tugwell, P. (2005). *Clinical Epidemiology. How to Do Clinical Practice Research* (3rd ed.)

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Methoden - Version 2.0 vom 19.12.2006. 19-12-2006.

Ioannides JPA: Why most published research findings are false. *PLoS Medicine* 2 (2005) 696-701

Janssen MF, Birnie E, Bonsel GJ. The discriminatory power of the EQ-5D, HUI2, HUI3 and SF-6D in 13 patient groups. Paper presented at 2007 EuroQol-Meeting, The Hague, Netherlands . 2007.

Jöckel KH, Hirche H, Neuhäuser M(2007): Hinweise zur Studienplanung, Biometrie und klinischen Epidemiologie. In: Rübben H (Hrsg.): *Uroonkologie*. 4. Auflage, Springer Medizin-Verlag, Heidelberg 2007, 13-28

Johannesson M, *Theory and Methods of Economic Evaluation of Health Care*. Kluwer, Dordrecht Boston London 1996.

Johnson JA, Luo N, Shaw JW, Kind P, Coons SJ. Valuations of EQ-5D health states: are the United States and United Kingdom different? *Med Care* 2005; 43(3):221-228.

Khan, S. K., Kunz, R., Kleijnen, J. Antes, G. (2003). *Systematic Reviews to Support Evidence-based Medicine*. The Royal Society of Medicine Press.

Klose T. The contingent valuation method in health care. *Health Policy* 1999; 47(2):97-123.

Knipschild P: The false positive therapeutic trial. *J Clin Epidemiol* 55 (2002) 1191-1195

Kopec JA, Willison KD. A comparative review of four preference-weighted measures of health-related quality of life. *J Clin Epidemiol* 2003; 56(4):317-325.

Krabbe PF, Bonsel GJ. Sequence effects, health profiles, and the QALY model: in search of realistic modeling. *Med Decis Making* 1998; 18(2):178-186.

Krauth, C., D. Charalobos-Markos, I. Brandes und J. Wasem (2005). "Die Perspektive der Gesetzlichen Krankenversicherung in der gesundheitsökonomischen Evaluation." *Zeitschrift für die gesamte Versicherungswissenschaft* 94: 215-256

Krauth, C. et al. (2005): Empirische Bewertungssätze in der gesundheitsökonomischen Evaluation - ein Vorschlag der AG Methoden der gesundheitsökonomischen Evaluation (AG MEG). *Gesundheitswesen* 67: S. 736-746.

Kunz R, Lelgemann M, Guyatt G, Antes G, Falck-Ytter Y, Schünemann HJ (2007). Von der Evidenz zur Empfehlung. In: Lehrbuch Evidenzbasierte Medizin in Klinik und Praxis. Deutscher Ärzteverlag.

Lyman, G. H., Kuderer, N. M. (2005). The strengths and limitations of meta-analyses based on aggregate data. *BMC Medical Research Methodology* 2005, 5:14 doi:10.1186/1471-2288-5-14.

Marra CA, Woolcott JC, Kopec JA, Shojania K, Offer R, Brazier JE et al. A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. *Soc Sci Med* 2005; 60(7):1571-1582.

Martini P: Experimentelle und therapeutisch-klinische Forschung, Kritik und Ethik. Festvortrag zur Eröffnung der Deutschen Therapiewoche 1950. *Die Therapiewoche* 1 (1950/51) 12-17

Maynard, A. (1997). Evidence-based medicine: an incomplete method for informing treatment choices. *The Lancet*, 349, 572-3..

McDowell I. Measuring health: A guide to rating scales and questionnaires. 3 ed. Oxford University Press, Oxford, 2006.

Meyer Th: Kritische Bewertung von qualitativen Studien. In: Kunz R, Ollenschläger G, Raspe H, Jonitz G, Donner-Banzhoff N (Hrsg.): Lehrbuch Evidenz-basierte Medizin in Klinik und Praxis. 2. überarbeitete und erweiterte Auflage. Köln, Deutscher Ärzteverlag 2007, S. 159-176

Moock J, Kohlmann T, Besch D, Druner K. Nutzentheoretische Lebensqualitätsmessinstrumente in der medizinischen Rehabilitation: Ein anwendungsbezogener Vergleich. *Z Med Psychol* 2005; 14:25-32.

Murray CJ, Acharya AK. Understanding DALYs (disability-adjusted life years). *J Health Econ* 1997; 16(6):703-730.

National Institute for Clinical Excellence (NICE). Guide to the Methods of technology appraisal. National Institute for Clinical Excellence, London, 2004.

Neumann PJ, Stone PW, Chapman RH, Sandberg EA, Bell CM. The quality of reporting in published cost-utility analyses, 1976-1997. *Ann Intern Med* 2000; 132(12):964-972.

Nord E. An alternative to QALYs: the saved young life equivalent (SAVE). *BMJ* 1992; 305(6858):875-877.

Nord E. The person-trade-off approach to valuing health care programs. *Med Decis Making* 1995; 15(3):201-208.

Olsson, T. M. (2007). Reconstructing evidence-based practice: an investigation of three conceptualisations of EBP. *Evidence & Policy*. Vol. 3, No. 2, 271-85.

Philips Z et al.: Review of guidelines for good practise in decision-analytic modelling in health technology assessment. *Health Technol Assess* 2004; 8 (36): 1-172.;

Polsky D, Willke RJ, Scott K, Schulman KA, Glick HA. A comparison of scoring weights for the EuroQol derived from patients and the general public. *Health Econ* 2001; 10(1):27-37.

Pocock, S., Elbourne, S. (2000). Randomized Trials or Observational Tribulations? *New England Journal of Medicine*. Vol. 342, No. 25,

Prentice RL(1989): Surrogate Endpoints in Clinical Trials: Definitions and Operational Criteria. *Stat. Med.* Vol 8, 431-440

Richardson G, Manca A. Calculation of quality adjusted life years in the published literature: a review of methodology and transparency. *Health Econ* 2004; 13(12):1203-1210.

Riggs LB, Hodgson SB, O'Fallon M et al: Effect of Fluoride treatment on the fracture rate in postmenopausal women with osteoporosis. *NEJM* 322 (1990) 802-809

Sach TH, Barton GR, Doherty M, Muir KR, Jenkinson C, Avery AJ. The relationship between body mass index and health-related quality of life: comparing the EQ-5D, EuroQol VAS and SF-6D. *Int J Obes (Lond)* 2007; 31(1):189-196.

Sackett, D. L. (2002). Clinical epidemiology: what, who, and whither. *Journal of Clinical Epidemiology*, 1161-1166.

Sackett D: The principles behind the tactics of performing therapeutic trials. In: Haynes, RB, Sackett D, Guyatt GH, Tugwell P: *Clinical Epidemiology*. Philadelphia, Baltimore, New York, Lippincott Williams & Wilkins 2006, 173-243

Salek S. The compendium of life instruments. John Wiley and Sons Ltd, 1999.

Schöffski O. Nutzentheoretische Lebensqualitätsmessung. In: Schoffski O, Graf v.d.Schulenburg M, editors. *Gesundheitsökonomische Evaluationen*. Studienausgabe. Springer, Berlin, Heidelberg, 2002: 277-284.

Schulenburg, J. M. von der (2007). Methoden zur Ermittlung von Kosten-Nutzen-Relationen für Arzneimittel in Deutschland. Hannover, Leibniz Universität, Wirtschaftswissenschaften

Schwappach DL, Boluarte TA. HEE-GER: a systematic review of German economic evaluations of health care published 1990-2004. *BMC Health Serv Res* 2007; 7:7.

Sen, Amartya: The Discipline of Cost-Benefit Analysis. In: *Journal of Legal Studies* 29 (2000) 2: S.931-952.

Staquet MJ, Hays RD, Fayers PM. Quality of life assessment in clinical trials. *Methods and practice*. Oxford University Press, Oxford, 1998.

Straus, Richardson, Glasziou, Haynes (2005). *Evidence-Based Medicine. How to Practice and Teach EBM*. Elsevier Churchill Livingstone.

Tunis SR, Stryer DB, Clancy CM: *Practical Clinical Trials*. JAMA 290 (2003) 1624-1632

Verband Forschender Arzneimittelhersteller: Pressekonferenz „Anforderungen des VFA an die Kosten-Nutzenbewertung in Deutschland.“ 28.03.2007 Berlin.

Weinstein M et al.: Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices – Modeling Studies. *Value in Health* 2003; 6 (1): 9-17

Weiss (2006). *Clinical Epidemiology* (3rd ed.) Oxford University Press

Welte, R. und R. Leidl (1999). Übertragung der Ergebnisse ökonomischer Evaluationsstudien aus dem Ausland auf Deutschland. In: Leidl, J. M. / von der Schulenburg / J. Wasem (Hrsg.): *Ansätze und Methoden der ökonomischen Evaluation- eine internationale Perspektive*. Nomos-Verlag. Baden-Baden. S. 171-202.

Zentner A, Velasco-Garrido M, Busse R. Methoden der vergleichenden Bewertung pharmazeutischer Produkte. DAHTA@DIMDI, Köln, 2005.

Zentner A, Busse R. Internationale Standards der Kosten-Nutzen-Bewertung [International standards of cost-effectiveness evaluation]. *Gesundh ökon Qual manag* 2006; 11:368-373.

Zentrale Ethikkommission bei der Bundesärztekammer: Priorisierung medizinischer Leistungen im System der Gesetzlichen Krankenversicherung (GKV). *Deutsches Ärzteblatt* 40 (104) 2007, A 2750-2754