

**Verfahrensweisen
und
Methoden
zur Nutzenbewertung
von Arzneimitteln
in Deutschland**

Ein Gutachten von Trudy Bekkering und Jos Kleijnen

Leuven, Belgien und York, United Kingdom

08. April 2008

Inhaltsverzeichnis

Zusammenfassung.....	4
1. Hintergrund.....	9
2. Gesetzlicher Rahmen	12
3. Anforderungen an den Bewertungsprozess	14
3.1 Verfahrenstransparenz	15
3.1.1 Themenermittlung und Priorisierung.....	15
3.1.2 Ausschreibung und Auftragsvergabe.....	16
3.1.3 Definition der Forschungsfrage	16
3.1.4 Protokollerstellung.....	17
3.1.5 Evidenzbewertung.....	18
3.1.6 Veröffentlichung des Vorberichts.....	19
3.1.7 Review des Vorberichts	19
3.1.8 Empfehlungen und Abschlussbericht	20
3.1.9 Appeal und Planung der Berichtsaktualisierung.....	21
3.1.10 Fazit.....	21
3.2 Aktive Beteiligung der Betroffenen.....	22
3.2.1 Themenermittlung und Priorisierung.....	22
3.2.2 Definition der Forschungsfrage(n).....	23
3.2.3 Beteiligung in anderen Phasen der Bewertung	27
3.3 Fazit.....	28
4. Anforderungen an die Bewertungsmethoden.....	29
4.1 Evidenzbasierte Medizin.....	30
4.1.1 Beste verfügbare Evidenz	30
4.1.2 Interne vs. externe Validität	31
4.1.3 Die Rolle systematischer Reviews in der EbM.....	33
4.1.4 Bewertung der Evidenz für jedes Outcome	34
4.1.5 Evidenzbasierte Methoden zur Vermeidung von Bias in der Bewertung.....	34
4.2 Wahl der Vergleichsgrößen	36
4.3 Nutzen	37
4.4 Fazit.....	37
5. Internationale Standards / Von NICE und IQWiG angewandte Methoden.....	38
5.1 Die Methoden des NICE und des IQWiG.....	39
5.2 Der von NICE und IQWiG angewandte Bewertungsprozess	40
5.3 Fazit.....	44
6. Wie sollte eine Nutzenbewertung durchgeführt werden?.....	45
6.1 Themenermittlung und Priorisierung	48
6.2 Ausschreibung und Auftragsvergabe	49
6.3 Definition der Forschungsfrage	50
6.4 Protokollerstellung.....	52
6.5 Bewertung der Evidenz.....	53
6.5.1 Suchstrategie	53
6.5.2 Vorlage durch die Hersteller	54
6.5.3 Studienauswahl	54
6.5.4 Datenextraktionsverfahren	55
6.5.5 Datensynthese	56

6.5.6. Analyse von Untergruppen	56
6.6 Veröffentlichung des Vorberichts	57
6.7 Revision des Vorberichts	58
6.8 Empfehlungen und Endbericht	58
6.9 Appeal / Planung der Berichtsaktualisierung.....	59
6.10 Fazit.....	60
Literaturverzeichnis	61

Zusammenfassung

Der Gemeinsame Bundesausschuss (G-BA) definiert im deutschen Gesundheitswesen, was von den Krankenkassen zu erstatten ist. Zur Definition einer Vorschrift kann der G-BA u.a. Nutzenbewertungen in Auftrag geben, die einen Überblick über die wissenschaftliche Evidenz hinsichtlich Wirksamkeit und Nutzen einer Intervention bieten. Diese Abhandlung beschreibt die operationale Umsetzung der gesetzlichen Anforderungen im Hinblick auf die Nutzenbewertung von Arzneimitteln. Manchmal werden diese Nutzenbewertungen als "isolierte Nutzenbewertungen" bezeichnet, um sie von Nutzenbewertungen im Rahmen einer vollen ökonomischen Bewertung (Kosten-Nutzen-Bewertung) zu unterscheiden.

Der G-BA hat freie Wahl, an welches Institut er Bewertungsaufträge vergibt. Bis heute wurde jedoch die Mehrheit der Aufträge an das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) vergeben. Dennoch bezieht sich der Inhalt dieses Papiers und der Begriff „Institut“ auf alle Institute, die mit solchen Bewertungen beauftragt werden können.

Die gesetzlichen Rahmenbedingungen für Nutzenbewertungen sind im Sozialgesetzbuch V, § 35b Abs. 1, § 139a Abs. 4 bis 6 sowie § 139b Abs. 3 festgelegt. Hierin wird ausgeführt, dass das Institut

- einen hohen Transparenzgrad gewährleisten muss,
- für eine angemessene Beteiligung der maßgeblich zu Beteiligten bei der auftragsbezogenen Erstellung von Bewertungen sowie für eine Gelegenheit zur Stellungnahme in allen wichtigen Abschnitten des Bewertungsverfahrens zu sorgen hat,
- regelmäßig über seine Arbeitsfortschritte und -ergebnisse Bericht erstatten muss, und
- den Auftrag an externe Sachverständige vergeben muss.

Die gesetzlich geforderte hohe Verfahrenstransparenz soll das gesamte Verfahren durch eindeutige Berichterstattung zu allen Verfahren und Festlegungen während aller Phasen

der Nutzenbewertung erreicht werden. Die wichtigsten Mittel zur Verbesserung der Transparenz sind wie folgt:

1. Realisierung eines unterstützenden Scoping-Prozesses bei der Entwicklung der Forschungsfrage;
2. Trennung der Arbeit der externen Sachverständigen, die die Evidenzbewertung vornehmen, von der des Instituts, welches die Empfehlungen formuliert. Daher muss der von den externen Sachverständigen erstellte Vorbericht veröffentlicht werden, und zwar separat von allen nachfolgenden Änderungen oder Berichten (bzw. Berichtsentwürfen) des Instituts, welche die Empfehlungen des Instituts enthalten, und
3. Durchführung eines offenen Peer-Review-Verfahrens durch Veröffentlichung der Stellungnahmen mit namentlicher Nennung der Reviewer.

Auf der Grundlage der gesetzlichen Rahmenbedingungen muss das Institut eine angemessene Beteiligung der maßgeblichen Parteien ermöglichen. Hierzu gehören u.a. die maßgeblichen Organisationen, die die Interessen von Patienten vertreten, Sachverständige der medizinischen und gesundheitsökonomischen Wissenschaft und Praxis, die Berufsvertretungen von Apothekern und Arzneimittelherstellern sowie Experten für alternative Behandlungsmethoden. Patienten und Angehörige der Gesundheitsberufe ermöglichen neue Einblicke hinsichtlich der Forschungsprioritäten, Behandlungen und Outcomes.

Die maßgeblichen Parteien sollten ermittelt und kontaktiert werden, sobald die wesentlichen Elemente des Bewertungsauftrages umrissen worden sind. Danach sollten die maßgeblichen Parteien an der Definition der Forschungsfrage, der Protokollentwicklung und der Stellungnahme(n) zum Vorbericht beteiligt werden. Zur Beteiligung der maßgeblichen Parteien an der Definition der Forschungsfrage wird ein Scoping-Verfahren vorgeschlagen. Für die anderen Phasen sollten schriftliche Stellungnahmen gefolgt von mündlichen Diskussionen erfolgen. Zum Schluss sollte den maßgeblichen Parteien das Recht eingeräumt werden, aus rechtlichen Gründen Berufung

(„Appeal“) gegen die endgültige Entscheidung einzulegen. All diese Schritte bedeuten keinerlei Einbußen der wissenschaftlichen Unabhängigkeit des Instituts.

Aus den maßgeblichen Abschnitten der gesetzlichen Rahmenbedingungen im Hinblick auf die *Bewertungsmethoden* kann gefolgert werden, dass

- 1) das Institut die Durchführung der Bewertung nach international anerkannten Standards der evidenzbasierten Medizin (EbM) gewährleisten muss,
- 2) die Bewertung im Vergleich mit anderen Arzneimitteln und Behandlungsformen unter Berücksichtigung des zusätzlichen therapeutischen Nutzens für die Patienten erfolgt,
- 3) der Kriterienkatalog zur Bewertung des Patientennutzens mindestens die Verbesserung des Gesundheitszustands, die Verkürzung der Krankheitsdauer, die Verlängerung der Lebensdauer, die Verminderung von Nebenwirkungen und die Verbesserung der Lebensqualität beinhaltet.

Evidenzbasierte Medizin bezieht sich auf die Anwendung der besten verfügbaren Evidenz zur Beantwortung einer Forschungsfrage, die Informationen zu Fragen der Patientenversorgung liefern kann. Selbst bei Fragen der Wirksamkeit ist das optimale Studiendesign nicht grundsätzlich die randomisierte kontrollierte Studie (RCT), sondern hängt von der Forschungsfrage und den Outcomes ab. Um die Transparenz für alle Fragestellungen zu erhöhen, sollten die Evidenzlevel explizit angegeben werden. Es gibt keinerlei empirische Evidenz, die eine Mindestzahl der zu berücksichtigenden Studien vor Abgabe von Empfehlungen unterstützt. Um die bestverfügbare Evidenz für Forschungsfragen zu erhalten, sollte alle möglicherweise relevante Evidenz berücksichtigt werden; es sollte also für jede Fragestellung die bestverfügbare Evidenz herangezogen werden, um die Frage zu beantworten. Gegebenenfalls müssen also verschiedene Evidenzlevel für jedes Outcome berücksichtigt werden.

Systematische Reviews können auf vielfältige Weise einem Bias unterworfen sein. Manche Arten des Bias können vermieden werden, während andere nur berichtet werden können. In einigen Fällen kann der Einfluss des Bias untersucht werden. Bei den

Betrachtungen muss aufgezeigt werden, dass potenzielle Biasquellen angemessen berücksichtigt wurden.

Die angewandten Methoden anderer Institutionen, die Nutzenbewertungen durchführen, sind nützlich für die Interpretation des Begriffs "internationale Standards", denen das Institut folgen muss. Das NICE ist ein gutes Beispiel in diesem Zusammenhang. Das NICE zeigt, dass transparente Verfahren bei der Nutzenbewertung möglich sind, dies jedoch eine ausführliche Dokumentation erfordert. Das NICE hat ein offenes Verfahren für die Stellungnahme von Gutachtern eingerichtet, das für Verfahrenstransparenz sorgt. Obwohl das IQWiG in Deutschland um Stellungnahmen zu Protokoll und Vorbericht bittet, diese auf seiner Website veröffentlicht und seine Kommentare zu diesen Stellungnahmen veröffentlicht, werden die Stellungnahmen letztlich dennoch nicht offen ausgewertet und es bleibt daher ungewiss, ob diese Stellungnahmen zu Änderungen in den Berichten führen oder nicht. Die Beteiligung der maßgeblichen Parteien im NICE-Bewertungsprozess hingegen gewährleistet ein Verfahren, das für alle relevanten Beteiligten transparent ist.

Transparenz während des gesamten Verfahrens wird durch klare Berichterstattung über Verfahren und Festlegungen in allen Phasen der Nutzenbewertung gewährleistet. Im Rahmen eines Scoping-Prozesses wird zu einem angedachten Projektumfang zunächst schriftlich und dann in Form eines Scoping-Workshops Stellung genommen. Auf diese Weise können alle relevanten Aspekte gehört und im endgültigen Projektumfang berücksichtigt werden. Danach wird der Berichtsplan erstellt („Protokoll“), gefolgt von der Evidenzbewertung. Über die angewandten Methoden sollte vollständig berichtet werden, um den Lesern zu zeigen, dass die Bewertung wissenschaftlich korrekt durchgeführt wurde und dass Bias, wo möglich, vermieden wurde. Die maßgeblichen Parteien sollten die Möglichkeit haben, zum Protokollentwurf und zum Vorbericht Stellung zu nehmen. Jede Stellungnahme sollte dahingehend ausgewertet werden, ob sie zu Änderungen führen wird, und sowohl die Stellungnahme als auch die Auswertung sollten veröffentlicht werden, um die Transparenz dieses Prozesses zu gewährleisten. Dasselbe Verfahren sollte für die Peer-Review-Phase angewandt werden. Auf der Basis

des Endberichts der Evidenzbewertung erstellt das Institut Empfehlungen und der G-BA stuft abschließend die Relevanz der Evidenz ein („Appraisal“).

Beim Verfassen des Endberichts sollte eine Trennung zwischen der methodischen Evidenzbewertung („evidence assessment“) und der Relevanzeinstufung der Evidenz („evidence appraisal“) realisiert werden. Idealerweise sollte diese Trennung gesetzlich verfügt werden, um jegliche Verwirrung im Hinblick auf Zuständigkeitskonflikte zu vermeiden.

Ein solches Verfahren garantiert eine durchführbare Kombination von gesetzlichen Anforderungen hinsichtlich der Transparenz und Beteiligung von maßgeblichen Parteien mit internationalen Standards evidenzbasierter Medizin, um zu gewährleisten, dass die Nutzenbewertung von Arzneimitteln in Deutschland nach den höchsten Standards erfolgt.

1. Hintergrund

In Deutschland werden Gesundheitsleistungen für die breite Öffentlichkeit von etwa 90% der Bundesbürger durch gesetzliche Krankenkassen geregelt. Die durch die gesetzlichen Krankenkassen erstattungsfähigen Leistungen des Gesundheitswesens werden durch den Gemeinsamen Bundesausschuss (G-BA) definiert. Der G-BA wurde durch das GKV-Modernisierungsgesetz von 2004 eingerichtet und ersetzt die bisherigen vier parallel existierenden Ausschüsse: den Bundesausschuss Ärzte und Krankenkassen, den Bundesausschuss Zahnärzte und Krankenkassen, den Koordinierungsausschuss und den Ausschuss Krankenhaus. Das Bundesministerium für Gesundheit (BMG) hat die Aufsicht über den G-BA.

Die Aufgabenstellung des G-BA wird durch das Sozialgesetzbuch V definiert und in der G-BA-Verfahrensordnung (1) festgelegt. Die Hauptaufgabe des G-BA besteht darin, zu formulieren, welche Leistungen des Gesundheitswesens – darunter auch Arzneimittel – von den Krankenkassen zu erstatten sind. Zur Definierung einer Vorschrift kann der G-BA ein geeignetes Institut mit Nutzenbewertungen oder Kosten-Nutzen-Bewertungen beauftragen. Manchmal werden die Nutzenbewertungen als "isolierte Nutzenbewertungen" bezeichnet, um sie von Nutzenbewertungen im Rahmen einer vollen ökonomischen Bewertung (Kosten-Nutzen-Bewertung) zu unterscheiden. Diese Bewertungen bieten einen Überblick über die wissenschaftliche Evidenz zu Nutzenaspekten¹ eines Arzneimittels. Aufgrund der vom beauftragten Institut abgegebenen Empfehlungen werden vom G-BA Vorschriften formuliert, ob die jeweilige Medizin oder Technologie zu erstatten ist oder nicht und im Verlauf ggf. Höchstbeträge festgelegt. Bewertungen können für alle neu zugelassenen Arzneimittel (mit patentgeschützten Arzneistoffen) oder für jedes "relevante Medikament" in Auftrag gegeben werden. Die Entscheidung der Auftragsvergabe wird vom G-BA getroffen und stützt sich auf die Arbeit von G-BA-internen Arbeitsgruppen. Die zu berücksichtigenden

¹ Outcomes, bei denen eine Vorgehensweise gewählt wird, mit der kein Ressourcenverbrauch einhergeht. Diese können sich auf Veränderungen der Gesundheit oder des Wohlbefindens des Patienten beziehen bzw. auf psychologische und physische Nutzen, die anderen Personen (nicht dem Patienten selbst) entstehen, die von Drogenmissbrauch betroffen sind (Familien/Freunde des Patienten, Opfer von Verbrechen usw.).(2)

Arzneimittel werden von den Mitgliedern des G-BA bestimmt, aber die Kriterien sind unklar. Zurzeit wird die Auftragsentscheidung auf der Basis von Kostenimplikationen oder Aspekten epidemiologischer Relevanz getroffen. Wir konnten jedoch keine konkreteren Kriterien ermitteln.

Eine Nutzenbewertung bewertet die klinischen Nutzen- und Schadenspotenziale von Arzneimitteln und folgt dem Design eines systematischen Reviews. Die Schritte einer Nutzenbewertung sind in Abbildung 1 dargestellt. Seit dem 1. April 2007 hat der Gesetzgeber dem G-BA außerdem die Option eingeräumt, eine Kosten-Nutzen-Bewertung zu beauftragen². Eine Kosten-Nutzen-Bewertung enthält immanent stets eine Nutzenbewertung. Bei ökonomischen Auswertungen müssen internationale Standards angewandt werden (siehe unten).

Es gibt jedoch einen wichtigen grundsätzlichen Unterschied zwischen Nutzenbewertungen³ gefolgt von einer Kosten-Nutzen-Bewertung im Vergleich zu Nutzenbewertungen ohne nachfolgende Kosten-Nutzen-Bewertung („Isolierte Nutzenbewertung“):

Nutzenbewertungen mit der Absicht einer Kosten-Nutzen-Bewertung erfordern einen erweiterten Umfang im Hinblick der Berücksichtigung der Nutzen.

Abbildung 1: Übliche Schritte bei der Nutzenbewertung

Die Nutzenbewertung ist in die folgenden 9 Schritte unterteilt:

1. Definition der vorläufigen Forschungsfrage
2. Ausschreibung und Auftragsvergabe
3. Definition des Umfangs
4. Protokollerstellung (im Sinne des IQWiG: Berichtsplan)
5. Evidenzbewertung: Diese Phase beinhaltet die Literaturrecherche sowie die kritische Auswertung und Analyse der Literatur.
6. Veröffentlichung des Vorberichts

² Studien, bei denen ein Vergleich von zwei oder mehr Therapien oder Versorgungsalternativen vorgenommen wird und bei denen Kosten und Outcomes der Alternativen untersucht werden. (3)

³ Wir verwenden den Begriff Nutzenbewertungen, wenn sowohl eine solche Bewertung mit der Absicht vorgenommen wird, eine Kosten-Nutzen-Bewertung anzuschließen, als auch für eine isolierte Nutzenbewertung ohne die Absicht einer nachfolgenden Kosten-Nutzen-Bewertung.

7. Revision des Vorberichts
8. Veröffentlichung des Abschlussberichts
9. Vorlage beim Auftraggeber
10. Appeal / Planung der Berichtsaktualisierung

Sowohl für die Nutzenbewertung als auch für die Kosten-Nutzen-Bewertung hat die Gesetzgebung konkrete Vorgaben hinsichtlich der anzuwendenden wesentlichen Kriterien und Verfahrensorderungen gemacht. Ökonomische Auswertungen nach internationalen Standards der Gesundheitsökonomie wurde von von der Schulenburg et al. ausführlich dargestellt. (4) Diese Abhandlung beschreibt die operationale Umsetzung der gesetzlichen Anforderungen bei der Nutzenbewertung von Arzneimitteln.

Gemäß seiner Verfahrensordnung ist es dem G-BA freigestellt, ein Institut mit der Durchführung einer Nutzenbewertung zu beauftragen (§ 38, Teil F Satz 1 und 2 der VO des G-BA). Bis jetzt wurde jedoch der größte Teil der Aufträge an das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) vergeben. Das IQWiG wurde mit dem GKV-Modernisierungsgesetz von 2004 gegründet. Zu den Aufgaben des Instituts gehört die Durchführung einer Nutzenbewertung von Arzneimitteln (vgl. §§ 35b Abs. 1 und 139b Abs. 1 und 2 SGB V). Der Inhalt dieser Abhandlung und die Verwendung des Begriffs „Institut“ gilt einheitlich für alle Institute, die mit der Durchführung solcher Bewertungen beauftragt werden können.

Diese Abhandlung besteht aus drei Teilen. Im ersten Teil werden die methodischen Anforderungen für die Verfahrensweisen des Instituts auf der Grundlage des gesetzlichen Rahmens erörtert (Abschnitte 2, 3 und 4). Der zweite Teil beschreibt Methoden, die im Ausland bei der Durchführung von Nutzenbewertungen angewandt werden (Abschnitt 5). Diese Methoden bilden die internationalen Standards, denen das Institut aufgrund der gesetzlichen Rahmenbedingungen Folge leisten muss. Der letzte Teil (Abschnitt 6) beschreibt auf der Basis der vorangegangenen Abschnitte ausführlich die empfohlenen Verfahrensweisen und Methoden.

2. Gesetzlicher Rahmen

In diesem Abschnitt wird der gesetzliche Auftrag für die Methoden und Verfahrensweisen des Instituts dargestellt und kommentiert. Die Konsequenzen des gesetzlichen Auftrags werden separat für den Bewertungsprozess (Abschnitt 3) und die Bewertungsmethoden (Abschnitt 4) beschrieben.

Der gesetzliche Rahmen für Nutzenbewertungen ist im Sozialgesetzbuch V festgelegt. Die relevanten Abschnitte des gesetzlichen Rahmens sind wie folgt:

§ 35b Abs. 1: "(1) 1 Das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen kann nach § 139b Abs. 1 und 2 beauftragt werden, den Nutzen oder das Kosten-Nutzen-Verhältnis von Arzneimitteln zu bewerten. 2 Bewertungen nach Satz 1 können für jedes erstmals verordnungsfähige Arzneimittel mit patentgeschützten Wirkstoffen sowie für andere Arzneimittel, die von Bedeutung sind, erstellt werden. 3 Die Bewertung erfolgt durch Vergleich mit anderen Arzneimitteln und Behandlungsformen unter Berücksichtigung des therapeutischen Zusatznutzens für die Patienten im Verhältnis zu den Kosten.⁴ 4 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, bei der wirtschaftlichen Bewertung auch die Angemessenheit und Zumutbarkeit einer Kostenübernahme durch die Versicherungsgemeinschaft⁵, angemessen berücksichtigt werden. 5 Das Institut bestimmt auftragsbezogen über die Methoden und Kriterien für die Erarbeitung von Bewertungen nach Satz 1 auf der Grundlage der in den jeweiligen Fachkreisen anerkannten internationalen Standards der evidenzbasierten Medizin und der Gesundheitsökonomie. 6 Das Institut gewährleistet bei der auftragsbezogenen Erstellung von Methoden und Kriterien und der Erarbeitung von Bewertungen hohe Verfahrenstransparenz und eine angemessene Beteiligung der in

⁴ Aus der ökonomischen Perspektive ist diese Formulierung inkorrekt; es sollte "Zusatzkosten" anstatt "Kosten" heißen.

⁵ Dies bezieht sich auf eine Auswertung aus der GKV-Perspektive. Es sollte jedoch angemerkt werden, dass die meisten internationalen Richtlinien eine gesellschaftliche oder volkswirtschaftliche Perspektive für solche Auswertungen verlangen. (4)

§ 35 Abs. 2 und § 139a Abs. 5 Genannten. 7 Das Institut veröffentlicht die jeweiligen Methoden und Kriterien im Internet. 8 Die Sätze 3 bis 7 gelten auch für bereits begonnene Nutzenbewertungen."

§ 35b Abs. 2: "(2) 1 Die Bewertungen nach Absatz 1 werden dem Gemeinsamen Bundesausschuss als Empfehlung zur Beschlussfassung nach § 92 Abs. 1 Satz 2 Nr. 6 zugeleitet. 2 Sie sind in geeigneten Abständen zu überprüfen und erforderlichenfalls anzupassen. 3 Bei Vorliegen neuer wissenschaftlicher Erkenntnisse ist die Bewertung auf Antrag der Hersteller zu überprüfen."

§ 139a Abs. 4: "(4) 1 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. 2 Es hat in regelmäßigen Abständen über die Arbeitsprozesse und -ergebnisse einschließlich der Grundlagen für die Entscheidungsfindung öffentlich zu berichten.

(5) 1 Das Institut hat in allen wichtigen Abschnitten des Bewertungsverfahrens Sachverständigen der medizinischen, pharmazeutischen und gesundheitsökonomischen Wissenschaft und Praxis, den Arzneimittelherstellern sowie den für die Wahrnehmung der Interessen der Patientinnen und Patienten und der Selbsthilfe chronisch Kranker und behinderter Menschen maßgeblichen Organisationen sowie der oder dem Beauftragten der Bundesregierung für die Belange der Patientinnen und Patienten Gelegenheit zur Stellungnahme zu geben. 2 Die Stellungnahmen sind in die Entscheidung einzubeziehen.

(6) Zur Sicherstellung der fachlichen Unabhängigkeit des Instituts haben die Beschäftigten vor ihrer Einstellung alle Beziehungen zu Interessenverbänden, Auftragsinstituten, insbesondere der pharmazeutischen Industrie und der Medizinprodukteindustrie, einschließlich Art und Höhe von Zuwendungen offen zu legen."

§ 139b Abs. 3: "1 Zur Erledigung der Aufgaben nach § 139a Abs. 3 Nr. 1 bis 5 hat das Institut wissenschaftliche Forschungsaufträge an externe Sachverständige zu vergeben. 2 Diese haben alle Beziehungen zu Interessenverbänden, Auftragsinstituten, insbesondere der pharmazeutischen Industrie und der Medizinprodukteindustrie, einschließlich Art und Höhe von Zuwendungen offen zu legen."

3. Anforderungen an den Bewertungsprozess

Aus dem Inhalt von Abschnitt 2 kann man zu dem Schluss kommen, dass - unter anderen Instituten - das IQWiG mit der Nutzenbewertung von Arzneimitteln beauftragt werden kann. Der gesetzliche Auftrag legt eine Anzahl von spezifischen Anforderungen an das Verfahren solcher Bewertungen dar:

1) Das Institut muss hohe Transparenz gewährleisten. Verfahren, Methoden und Kriterien sollten im Internet veröffentlicht werden. In Abschnitt 3.1 argumentieren wir, dass – beginnend mit der Themenermittlung und Priorisierung – hohe Transparenz für alle Phasen der Bewertung gelten sollte.

2) Das Institut muss eine angemessene Beteiligung der maßgeblichen Parteien bei der auftragsbezogenen Erstellung von Bewertungen sowie eine Gelegenheit zur Stellungnahme in allen wichtigen Abschnitten des Bewertungsverfahrens einräumen. Angemessene Beteiligung bedeutet die Möglichkeit, zu allen wichtigen Abschnitten des Verfahrens einen Beitrag leisten zu können. Die Stellungnahmen müssen in die Dokumentation einbezogen werden. In Abschnitt 3.2 wird erklärt, wann die maßgeblichen Parteien teilnehmen sollten (Verfahren), während Abschnitt 6 beschreibt, wie die Teilnahme dieser Parteien am besten ablaufen sollte (Methoden). Zu den maßgeblichen Parteien gehören zumindest die folgenden:

- Relevante Organisationen, die Patienteninteressen vertreten,
- Sachverständige der medizinischen, pharmazeutischen und gesundheitsökonomischen Wissenschaft und Praxis, sowie
- Berufsvertretungen der Apotheker und Arzneimittelhersteller sowie Experten für alternative Behandlungsmethoden.

Diese Parteien werden im restlichen Dokument als "maßgebliche Parteien" bezeichnet. Durch die Teilnahme solcher maßgeblicher Parteien würden aber auch Verpflichtungen entstehen, z.B. alle relevanten Informationen für die Nutzenbewertung zur Verfügung zu stellen.

3) Das Institut muss in regelmäßigen Abständen über seine Arbeitsfortschritte und – ergebnisse berichten. Diese Bedingung ist eng verbunden mit dem Gebot der Transparenz. Nachfolgend werden wir argumentieren, dass die Arbeit der externen Experten aus Gründen der Transparenz getrennt von allen späteren Änderungen oder Berichten (bzw. Berichtsentwürfen) des Instituts veröffentlicht werden muss.

4) Das Institut muss den Auftrag an externe Sachverständige vergeben. Externe Experten sowie Angestellte des Instituts müssen potenzielle Interessenkonflikte offen darlegen.

5) Die Bewertung sollte in regelmäßigen Abständen aktualisiert werden. Falls neue Evidenz verfügbar wird, wird die Bewertung auf Wunsch der Hersteller revidiert.

3.1 Verfahrenstransparenz

Gemäß den Anforderungen in § 35b Abs. 1 Satz 6 SGB V muss das Institut eine hohe Verfahrenstransparenz und Beteiligung gewährleisten. Transparenz ist die Grundvoraussetzung für jegliche Forschung, da nur so gezeigt werden kann, dass der Prozess wissenschaftlich korrekt durchgeführt wurde und ein Bias so weit wie möglich vermieden wird. Transparenz ist für die beauftragten Institute ein wesentliches Instrument, um zu zeigen, dass man auf die Validität der Ergebnisse vertrauen kann. Die Transparenz des Gesamtprozesses sollte durch klare Berichterstattung zu Verfahren und Kriterien während aller Phasen der Nutzenbewertung erreicht werden. Der nachfolgende Abschnitt beschreibt die wesentlichen Momente bei einer Nutzenbewertung, in denen Transparenz erforderlich ist, um diese Anforderung zu erfüllen.

3.1.1 Themenermittlung und Priorisierung

Wir argumentieren, dass – als Teil eines transparenten Bewertungsverfahrens – der Prozess, der zur Bewertung führt, d.h. die Themenermittlung und Priorisierung, ebenfalls

klar sein sollte. Obwohl das endgültige Thema durch politische Gründe beeinflusst sein mag, sollte der eigentliche Vorgang so transparent wie möglich sein. Die Themenermittlung sollte für die Öffentlichkeit und damit auch für alle maßgeblichen Parteien offen sein. Zu diesem Zweck sollten das BMG und der G-BA ein Verfahren zur Themenermittlung und Priorisierung einrichten, das die Öffentlichkeit mit einbindet. Dies könnte durch Veröffentlichung der Kriterien erreicht werden, nach denen potenzielle Themen zur Beauftragung ausgewählt werden, sowie der Kriterien für die Priorisierung, falls es mehrere passende Themen gibt. Ein vergleichbares Verfahren wird bereits beim DIMDI realisiert, und eine Auswertung dieser Erfahrungen könnte nützliche Informationen liefern. (5)

3.1.2 Ausschreibung und Auftragsvergabe

Um legislative Anforderungen zu erfüllen, muss das Institut wissenschaftliche Forschungsaufträge an externe Sachverständige vergeben. Dennoch verlangt die Gesetzgebung nicht, dass externe Experten für alle Bewertungen beauftragt werden. In einigen Fällen mag das Institut entscheiden, den Bewertungsauftrag an eine interne Expertengruppe, d.h. Mitarbeiter des IQWiG, zu vergeben. Dies ändert jedoch nicht die operationale Umsetzung der gesetzlichen Anforderungen im Hinblick auf die Nutzenbewertung. Um Transparenz zu gewährleisten, sollte klargestellt werden, unter welchen Umständen externe Experten erforderlich bzw. nicht erforderlich sind. Diese Kriterien sollten veröffentlicht werden.

Zur Rekrutierung von externen Sachverständigen werden öffentliche und internationale Ausschreibungen bevorzugt, da diese sicherstellen, dass eine unbegrenzte Anzahl von Experten ihre Dienstleistungen anbieten können. Die Auswahl von Experten sollte anhand nachvollziehbarer und objektiver Kriterien erfolgen. Diese Phase könnte transparent gestaltet werden, in dem das Ausschreibungsverfahren sowie die Kriterien zur Expertenauswahl veröffentlicht werden, z.B. auf der Webseite des Instituts.

3.1.3 Definition der Forschungsfrage

Der erste Schritt einer Bewertung ist die Definition der Forschungsfrage. Dies ist eine der wichtigsten Phasen einer Bewertung, da schlecht fokussierte Fragen zu unklaren

Entscheidungen führen, welche Forschung einbezogen werden sollte und wie diese zusammenzufassen ist. (6) Alle maßgeblichen Parteien sollten in diese Phase eingebunden werden, wie in § 35b Abs. 1 SGB V gefordert. Ein solches Verfahren gewährleistet, dass alle wichtigen Aspekte gehört werden und somit alle Perspektiven bei der Formulierung der Forschungsfrage berücksichtigt werden. Hierfür wird ein Scoping-Prozess vorgeschlagen, der ausführlich in Abschnitt 6.3 beschrieben wird. Ein Scoping-Prozess zielt darauf ab, die Grenzen der Bewertung im Hinblick auf die vier Elemente des PICO-Systems abzustecken: Patientenpopulation, Intervention, Vergleichsintervention und Outcomes. Der Scoping-Prozess kann die gesamte Bewertung beschleunigen, vorausgesetzt, dass genügend Input von den maßgeblichen Parteien miteinbezogen wurde.

Um diese Phase transparent zu gestalten, sollte klargestellt werden, welche Parteien zu der Forschungsfrage beigetragen haben, sowie welche Vorschläge berücksichtigt wurden und welche nicht. Deshalb sollten alle Vorschläge der beteiligten Parteien, die zum Entwurf der Forschungsfrage Stellung nehmen, veröffentlicht werden, zusammen mit einer Dokumentation, ob ein Vorschlag als relevant erachtet wurde oder nicht, und in die Forschungsfrage einbezogen wurde. Auch Protokolle von Sitzungen, bei denen Vorschläge und Stellungnahmen erörtert wurden, sollten veröffentlicht werden. Zuvor sollte für eine Möglichkeit der Geheimhaltung von vertraulichen Patienteninformationen gesorgt werden.

3.1.4 Protokollerstellung

Auf der Basis der Forschungsfrage wird nun ein Protokoll verfasst (in der Terminologie des IQWiG: "Berichtsplanentwurf"). Zur Verbesserung der Transparenz werden PICO-Kriterien angewandt, um die Forschungsfrage(n) zu fokussieren:

- Patientenpopulation: Auf welche Patienten bezieht sich die Bewertung?
- Intervention: Welches Arzneimittel wird bewertet (einschließlich Berücksichtigung der Dosierung und Anwendungsmethoden)?

- Vergleichsintervention: Was ist/sind die gegenwärtigen Standardbehandlung(en) (einschließlich der Begründung für die Wahl dieser Behandlung(en) als Standard)?
- Outcomes: Welche Outcomes sind wichtig und patientenspezifisch?

Ferner enthält das Protokoll die folgenden Punkte: Hintergrundinformationen, die zu(r) Forschungsfrage(n) führen, Suchstrategie, Studienauswahlkriterien und –verfahren, Bewertung der Studienqualität, Datenextraktionsstrategie und Synthesemethode für die extrahierten Daten.

Der Protokollentwurf sollte für Stellungnahmen zur Verfügung stehen. Er sollte von allen maßgeblichen Parteien kommentiert werden können, und eine mündliche Anhörung sollte hiernach ebenfalls stattfinden. Zur Verbesserung der Transparenz sollten der Protokollentwurf, alle Stellungnahmen der maßgeblichen Parteien sowie die Auswertung dieser Stellungnahmen im Internet veröffentlicht werden. Auch Protokolle von Sitzungen zur Erörterung aller relevanten Stellungnahmen sollten publiziert werden.

3.1.5 Evidenzbewertung

Die Bewertung der Evidenz umfasst alle Schritte eines systematischen Reviews der Evidenz, von der Literaturrecherche bis zur Evidenzsynthese. Obwohl alle Schritte dieses Verfahrens ausführlich im Protokoll beschrieben werden, ist es wahrscheinlich, dass während der Bewertung neue Entscheidungen getroffen werden müssen oder vorherige Entscheidungen unausführbar erscheinen. Ferner sind subjektive Urteile während dieses Verfahrens unvermeidlich. Eine transparente Vorgehensweise versichert dem Nutzer, dass ein strenger Ansatz verfolgt, subjektive Urteile aufgedeckt und Bias soweit wie möglich vermieden wurde. Transparenz während dieser Phase ist gewährleistet, wenn der Leser alle während dieses Prozesses unternommenen Schritte nachvollziehen kann. Jegliche am Protokoll vorgenommenen Änderungen sollten begründet, mitgeteilt und als Ergänzung veröffentlicht werden sowie während einer mündlichen Anhörung zur Diskussion gestellt werden. Weitere Einzelheiten dazu, wie das Verfahren der Evidenzbewertung ablaufen sollte, werden in Abschnitt 6.5 ausgeführt.

3.1.6 Veröffentlichung des Vorberichts

Der Vorbericht beschreibt das Verfahren und die vorläufigen Ergebnisse der Bewertung.

Zur Verbesserung der Transparenz sollte der Bericht so weit wie möglich in

Übereinstimmung mit den Empfehlungen der Berichtsrichtlinien für systematische

Reviews verfasst werden (QUOROM-Richtlinien, www.equator-network.org). (7)

Um Evidenz auf klare Art und Weise zu präsentieren, sollten Evidenztabelle verwendet

werden. Eine Standardmethode für die Erstellung von Evidenztabelle wurde bislang

nicht ermittelt, hauptsächlich aus dem Grund, dass dies von der Forschungsfrage

abhängt.(8) Dennoch sollten alle Resultate und Charakteristika der eingeschlossenen

Studien, die die Ergebnisse evtl. beeinflusst haben oder für die Verallgemeinerbarkeit von

Ergebnissen relevant sind, so präsentiert werden, dass die eingeschlossenen Studien leicht

miteinander zu vergleichen sind.

Zur Gewährleistung der Transparenz sollte der Bericht in der von den externen

Sachverständigen verfassten Form veröffentlicht werden. Die Forderung einer

Veröffentlichung gilt auch, wenn die Bewertung intern von Mitarbeitern des beauftragten

Instituts vorgenommen wird.

3.1.7 Review des Vorberichts

Der gesamte Bewertungsprozess bietet mehrere Möglichkeiten der Stellungnahme durch

die maßgeblichen Parteien. Alle Kommentare sowie die Auswertung jeder individuellen

Stellungnahme im Hinblick darauf, ob diese relevant ist und zu Änderungen führt, sollten

veröffentlicht werden. Ein solches Verfahren, wie es vom NICE bereits durchgeführt

wird, (9) wird die Transparenz des Reviewverfahrens verbessern.

Zusätzlich findet ein externes Review des Vorberichts statt. Ein ähnlicher Peer-Review-

Prozess hat sich bei wissenschaftlichen Veröffentlichungen etabliert. Dies ist eine

wichtige Arbeitsphase in der die Arbeit der Reviewgruppe überprüft werden kann. Der

Peer-Review-Prozess funktioniert am besten transparent. Um diese Transparenz zu

gewährleisten, sollte zunächst klar sein, wie das Institut Peer-Reviewer einlädt. Der

Einladungsprozess sowie die Kriterien für die Auswahl von Reviewern und für die

Durchführung des Reviews sollten veröffentlicht werden. Zweitens sollten die

Kommentare der Peer-Reviewer dokumentiert und ausgewertet werden, ähnlich dem

Verfahren für Stellungnahmen der maßgeblichen Parteien. Das Ergebnis der Auswertung sollte im Rahmen des Endberichts dokumentiert und veröffentlicht werden. Wann immer eine Stellungnahme als nicht relevant erachtet wird, sollte dies begründet werden. Nur so kann sichergestellt werden, dass die Kommentare der Reviewer berücksichtigt oder aus angemessenen Gründen ausgelassen wurden. Für weitere Einzelheiten siehe Abschnitt 6.7.

Um die Qualität des Peer-Review-Systems zu verbessern, haben zwei große medizinische Fachzeitschriften die Auswirkungen eines offenen Peer-Review-Systems untersucht, d.h. die Namen der Reviewer wurden den Autoren bekannt gegeben. (10;11) Beide Studien zeigten, dass sich ein offenes Reviewverfahren nicht negativ auf die Qualität dieser Reviews auswirkt. Daher gibt es keinen Grund, den Namen der Reviewer zu verheimlichen. Einige der Argumente für ein offenes Peer-Review-Verfahren sind größere Verantwortlichkeit, Fairness und Transparenz.

3.1.8 Empfehlungen und Abschlussbericht

Nach Veröffentlichung des Vorberichts der Evidenzbewertung gibt das Institut Empfehlungen ab. Da diese im Abschlussbericht enthalten sind, sind Berichte des Instituts mehr als eine Evidenzbewertung, und gehen in ein Appraisal über. Rechtlich gesehen ist die Evidenzeinstufung („Appraisal“) aufgrund dieses Berichts die Aufgabe des G-BA. Trotzdem beinhalten die Empfehlungen des Instituts ebenfalls eine Form der Relevanzeinstufung der Evidenz. Somit nimmt das Institut sowohl eine Bewertung als auch eine Relevanzeinstufung der Evidenz (ein „Appraisal“) vor, weshalb potenziell ein Interessenkonflikt entsteht. Dies sollte durch Trennung dieser zwei Schritte vermieden werden, ähnlich der Vorgehensweise des NICE. (9). In Tabelle 2 werden nachfolgend die Definitionen des NICE für Bewertung und Relevanzeinstufung der Evidenz vorgestellt.

Tabelle 2. Der Unterschied zwischen Evidenzbewertung und Relevanzeinstufung der Evidenz laut NICE.(12)

"Der <u>Bewertungs</u> prozess besteht aus einer objektiven Analyse der Qualität, Erkenntnisse
--

und Implikationen der verfügbaren Evidenz (hauptsächlich aus der Forschung), wie sie sich auf die Fragestellung und den Kontext der Relevanzeinstufung beziehen. Dahingegen handelt es sich bei dem Prozess der Relevanzeinstufung der Evidenz um eine Betrachtung der Ergebnisse des Bewertungsprozesses im Rahmen der zusätzlichen Informationen, die von den maßgeblichen Parteien wie z.B. Klinikern und Patientenexperten bereitgestellt werden. Die Einstufungsentscheidung ist eine Beurteilung der Wichtigkeit einer Reihe von Faktoren, die von Einstufung zu Einstufung unterschiedlich ausfallen."

Verwirrung kann ebenfalls vermieden durch ein offenes Verfahren, das die Ergebnisse der Evidenzbewertung (wie im Vorbericht der externen Experten dargelegt) getrennt von den Empfehlungen im Endbericht (durch das IQWiG) und den Ergebnissen der Relevanzeinstufung der Evidenz (durch den G-BA) präsentiert sowie eine klare Verbindung zwischen jeder Empfehlung und der Evidenz, auf denen diese Empfehlung basiert, aufzeigt. Idealerweise sollte jedoch – ähnlich dem vom NICE angewandten Verfahren – eine Trennung zwischen Evidenzbewertung und Relevanzeinstufung der Evidenz, d.h. zwischen Assessment und Appraisal und damit zwischen den Kompetenzen des IQWiG und des G-BA vollzogen werden.

3.1.9 Appeal und Planung der Berichtsaktualisierung

Laut § 35b Satz 2 und 3 SGB V müssen die Bewertungen in angemessenen Zeitabständen überprüft und ggf. angepasst werden. Ferner ist die Bewertung auf Antrag der Hersteller zu revidieren, wenn neue wissenschaftliche Evidenz verfügbar ist. Die Hersteller sollten alle Informationen zur Nutzenbewertung zur Verfügung stellen. In diesem Sinne wäre eine Registrierung der laufenden klinischen Studien nützlich. Transparenz ist während dieses Schritts gewährleistet, wenn die Verfahren für Appeal und Aktualisierung nach objektiven Kriterien veröffentlicht werden.

3.1.10 Fazit

Auf der Basis des gesetzlichen Rahmens muss das Institut hohe Verfahrenstransparenz gewährleisten. Transparenz sollte über den gesamten Prozess hinweg durch klare

Berichterstattung über Verfahren und Kriterien während aller Phasen der Nutzenbewertung erreicht werden. Die wichtigsten Mittel zur Transparenzverbesserung sind:

1. Realisierung eines unterstützenden Scoping-Prozesses bei der Entwicklung der Forschungsfrage.
2. Veröffentlichung der Stellungnahmen aller beteiligten Parteien zusammen mit einer Begründung, ob diese Stellungnahmen zu Änderungen der endgültigen Dokumente geführt haben oder nicht.
3. Trennung des Ergebnisses der Evidenzbewertung („Assessment“) von dem der Relevanzeinstufung der Evidenz („Appraisal“) durch Veröffentlichung der Arbeit der externen Sachverständigen.
4. Durchführung eines offenen Peer-Review-Verfahrens durch Veröffentlichung der Stellungnahmen mit namentlicher Nennung der Reviewer.

3.2 Aktive Beteiligung der Betroffenen

Auf der Basis der gesetzlichen Anforderungen muss das Institut eine angemessene Beteiligung der maßgeblichen Parteien bei der auftragsbezogenen Erstellung von Bewertungen ermöglichen. Angemessene Beteiligung bedeutet die Möglichkeit, zu allen wichtigen Phasen des Verfahrens einen Beitrag leisten zu können. Hierzu gehören die Arbeitsschritte von der Präzisierung der Fragestellung bis einschließlich zur Relevanzeinstufung der Evidenz („Appraisal“) beim G-BA. Dieser Abschnitt konzentriert sich auf das Verfahren (Wann sollten sich die maßgeblichen Parteien beteiligen?), während die Methoden (Wie sollten sich die maßgeblichen Parteien beteiligen?) ausführlich in Abschnitt 6 beschrieben werden.

3.2.1 Themenermittlung und Priorisierung

Die Themenermittlung sollte der Öffentlichkeit zugänglich sein und daher auch den maßgeblichen Parteien. Das Verfahren sollte öffentlich durchgeführt werden. Ein Beispiel der Zusammenarbeit zwischen Klinikern und Patienten wird von der Database of Uncertainties about the Effects of Treatments (DUETs) realisiert,(13) wo

Forschungsfragen zu den Auswirkungen von Behandlungen nur in die Datenbank miteinbezogen werden, wenn sie sowohl von Patienten als auch von Klinikern gestellt wurden.

3.2.2 Definition der Forschungsfrage(n)

Um bei der Definition der Forschungsfrage(n) eine Unterstützung durch aktive Beteiligung der maßgeblichen Parteien zu ermöglichen, wird ein Scoping-Prozess empfohlen. Der Zweck des Scoping-Prozesses ist es, einen Rahmen für die Bewertung zu definieren. Die Kernpunkte von Interesse, z. B. Population, Intervention und Outcomes, sind so klar wie möglich zu beschreiben. Ein solcher Scoping-Prozess besteht aus zwei Schritten. Zunächst wird ein erster Entwurf der Forschungsfrage mit Skizzierung aller relevanten Parameter und Hintergründe an alle maßgeblichen Parteien mit der Bitte um schriftliche Stellungnahmen geschickt. Danach wird ein Scoping-Workshop abgehalten, bei dem alle Kommentare und Ansichten diskutiert werden. Ein solches Verfahren gewährleistet, dass alle relevanten Aspekte gehört und in den endgültigen Arbeitsumfang miteinbezogen werden. Dies sollte zu Forschungsfragen führen, die für alle beteiligten Parteien relevant sind, wobei die wissenschaftliche Unabhängigkeit des Instituts dennoch gewährleistet bleibt.

Das Institut sollte den Scoping-Workshop organisieren, an dem der G-BA, das Institut, die externen Sachverständigen und die maßgeblichen Parteien teilnehmen sollten. Durch die Teilnahme des G-BA sollte sichergestellt werden, dass sich die Forschungsfrage(n) unmittelbar auf die klinische Fragestellung des G-BA bezieht (bzw. beziehen) und die Ergebnisse der Bewertung daher die Anforderungen des G-BA erfüllen. Alternativ könnte ein einphasiger Scoping-Prozess durchgeführt werden, der vom G-BA organisiert wird.

Das jeweilige Forschungsinstitut wird vom G-BA beauftragt. Dem potenziellen Auftrag liegt eine politische Fragestellung zugrunde, für die eine Entscheidung benötigt wird, z. B.: Was sind die besten Behandlungsmöglichkeiten für Gehirntumorpatienten? Danach wird die Forschungsfrage anhand der PICO-Kriterien weiterentwickelt: Meinen wir im Hinblick auf die Teilnehmer beispielsweise primäre oder sekundäre Tumoren,

beziehen wir uns auf Kinder oder Erwachsene? Wie sollte die Intervention definiert werden? Radiotherapie, Chemotherapie und/oder chirurgische Behandlung? Wird die Protonenbestrahlung miteinbezogen, wie viele Einrichtungen für Protonentherapie gibt es in Deutschland? Welche Vergleichsintervention wäre verfügbar, nützlich und ethisch vertretbar, d. h. Protonenbestrahlung, wachsaues Beobachten, Scheinbehandlung? Wenn es um die Outcomes geht, konzentrieren wir uns auf die Mortalität oder sind wir an Morbidität, Lebensqualität und unerwünschten Ereignissen interessiert? Die mindestens zu berücksichtigenden Outcome-Dimensionen sind im Sozialgesetzbuch V definiert. Ferner sollten die Studientypen, die zur Beantwortung jeder einzelnen Frage innerhalb der Fragestellung relevant sind, spezifiziert werden. Dabei sollten die ausgewählten Studiendesigns nicht als Mittel zum Ausschluss anderer Designs von der Bewertung benutzt werden. Für jede Fragestellung sollte die jeweils bestverfügbare Evidenz gesucht werden, die zum Zeitpunkt der Bewertung verfügbar ist. Dieser Aspekt wird in Abschnitt 4.1 erörtert. Die daraus resultierende Frage sollte für das deutsche Gesundheitswesen relevant sein und außerdem das Problem behandeln, für das eine Entscheidung benötigt wird. Dies ist eine wichtige Phase des Projekts und sollte angemessene Beachtung finden. Alle potenziell maßgeblichen Parteien sollten in dieser Phase ermittelt und kontaktiert werden.

Danach wird ein vorläufiger Themenentwurf formuliert, der mit den Beiträgen der Scoping-Gruppe besser präzisiert werden kann. Zu einer solchen Gruppe gehören alle maßgeblichen auftragsspezifischen Parteien. Die Beratung der maßgeblichen Parteien gewährleistet, dass alle relevanten Aspekte berücksichtigt wurden. Daher muss diese Gruppe sorgfältig zusammengesetzt werden und sollte auch aus Vertretern der Patientenorganisationen und Pflegern bestehen. Auch Sachverständige aus Wissenschaft und Praxis in Medizin, Pharmazie und Gesundheitsökonomie sollten beteiligt werden. Für eine Bewertung im Bereich Diabetes würden z. B. Kliniker mit starkem klinischen Hintergrund sowie Kliniker mit wissenschaftlicher Erfahrung gebraucht. Ferner gibt es viele Gesundheitsberufe, die sich mit Diabetespatienten beschäftigen, wie z. B. Krankenpfleger, Physiotherapeuten, Heimpflegepersonal, Diätassistenten sowie Experten in Wundpflege und Vorsorge.

Mindestens müssen die gesetzlich vorgeschriebenen Teilnehmer, die Berufsvertretungen von Apothekern und Arzneimittelherstellern sowie Patienten beteiligt werden. Darüber hinaus kann je nach Thema die Scoping-Gruppe um zusätzliche Kreise erweitert werden, wie z. B. Angehörige von Patienten oder Experten in Alternativtherapien. Dies führt zu besserer Akzeptanz der Bewertungsergebnisse und zu einer unverzichtbaren Sammlung von Fachwissen im Bewertungsprozess.

Die Beteiligung von Stakeholdern in der Forschung wird allgemein als wichtig erachtet. (14, 15) Einige empirische Evidenz dafür stammt aus Studien, die sich auf die Beteiligung von Patienten oder Verbrauchern konzentriert haben, zu denen frühere, gegenwärtige oder zukünftige Anwender, betreuende Personen und Personen, die diese Gruppen vertreten, gehörten. Es wird davon ausgegangen, dass eine breite Beteiligung zu einer Forschung führt, die für die Bedürfnisse und Belange der Menschen relevanter ist, die verlässlicher ist und die mit größerer Wahrscheinlichkeit praktisch umgesetzt wird. (16) Auf der Basis einer Forschungsarbeit schlussfolgern Hanley et al (17), dass die Beteiligung von Verbrauchern bei der Studienplanung und -durchführung von kontrollierten Studien zunimmt, was von den meisten Wissenschaftlern anscheinend gutgeheißen wird. Ein weiteres Argument für die Beteiligung der maßgeblichen Parteien bezieht sich auf subjektive Beurteilungen, die bei jeder Bewertung vorkommen. Ein Scoping-Workshop kann subjektive Beurteilungen (Werturteile) zwischen verschiedenen Gruppen ausgleichen.

Empirische Evidenz zeigt, dass der individuelle Bias von Personen in multidisziplinären Gruppen besser ausgeglichen werden kann. Beispielsweise kann bei Vorlage derselben Evidenz eine einzelne spezialisierte Gruppe zu anderen Schlussfolgerungen gelangen als eine multidisziplinäre Gruppe. Ein multidisziplinäres Gremium kann differenziertere Standpunkte liefern als Gremien, die zur Gänze aus Praktikern bestehen, die die Interventionen anwenden. (18) Coulter et al (19) haben gezeigt, dass die Zusammensetzung eines Gremiums die Bewertungen beeinflusst und dass diejenigen, die ein Verfahren anwenden, dieses eher als angemessen bewerten als jene, die das Verfahren nicht anwenden.

Das NICE sieht den Beitrag der Gesundheitsberufe als einzigartig an, durch den unter professioneller Sichtweise beschrieben wird, welchen Rang die zu bewertende Technik in der gegenwärtigen klinischen Praxis einnimmt. (12) Kliniker können daher Evidenz zu folgenden Fragen liefern:

- Variationen von Patientengruppen, insbesondere zum unterschiedlichen Baseline-Risiko der Erkrankung und unterschiedlicher Ausmaße des Nutzens für verschiedene Patientenuntergruppen,
- Besondere Umstände, unter denen eine Behandlung erbracht wird, einschließlich der Notwendigkeit von Begleitbehandlungen, des Settings, in dem die Behandlung angewandt wird, sowie der Anforderungen für zusätzliche professionelle Leistungen; und
- Behandlungen, die gegenwärtig als Standardpraxis angewandt werden, und falls diese sich von dem unterscheiden, was als beste Praxis gilt.

Es gibt empirische Evidenz, die aufzeigt, dass die Präferenzen von Patienten und Gesundheitsdienstleistern sich im Hinblick auf Forschungsprioritäten, Behandlung und Outcomes unterscheiden. Ein Review ließ z. B. eine Reihe von nicht zusammenpassenden Prioritäten für die Gesundheitsforschung zwischen Mitgliedern der Gesundheitsberufe und der Öffentlichkeit erkennen. (20) Devereaux et al. (21) zeigten eine beträchtliche Variabilität zwischen Ärzten und Patienten in ihrer Abwägung von potenziellen Outcomes im Zusammenhang mit Vorhofflimmern und seiner Behandlung. Ferner bestand erhebliche Variabilität innerhalb der Gruppe von Patienten und innerhalb der Gruppe von Ärzten. (21) Unterschiede bei den Präferenzen von Patienten und Gesundheitsdienstleistern sind schwierig vorherzusagen, unterscheiden sich in ihrer Tendenz und Größe und sind oft krankheitsspezifisch. (22) Chard et al. (23) zeigten, dass die Standpunkte von Medizinern und Patienten zur Behandlung von Osteoarthritis nicht übereinstimmten. Aufgrund des Inputs von Patienten wurde „Fatigue“ („Ermüdung“) als zentrales Outcome bei der Auswertung von Interventionen für rheumatoide Arthritis hinzugefügt. (24) Andere beispielhafte Studien zeigen mangelnde Übereinstimmung bei wichtigen Stakeholdern, darunter Patienten, Familienangehörige und

Gesundheitsdienstleister, in Bezug auf gewünschte Outcome-Prioritäten in der Jugendpsychiatrie (25), bei Schizophrenie (26) und in der Rheumatologie.(27) Solche Ergebnisse zeigen, dass der Input von Patienten und Mitgliedern der Gesundheitsberufe während der Formulierungsphase der Forschungsfrage besonders relevant ist.

3.2.3 Beteiligung in anderen Phasen der Bewertung

Beteiligung sollte in allen wichtigen Schritten des Bewertungsverfahrens stattfinden. Abgesehen vom oben genannten Scoping-Prozess zur Entwicklung der Forschungsfrage gehören hierzu insbesondere:

- Schriftliche Stellungnahmen und mündliche Diskussion zum Protokoll,
- Schriftliche Stellungnahmen und mündliche Diskussion zum Vorbericht,
- Appeal („Einspruch“) gegen die endgültige Entscheidung des G-BA.

Die Form der Beteiligung ist in Abschnitt 6 dieser Abhandlung beschrieben.

Es gibt zunehmend empirische Evidenz, die die Bedeutung einer solchen Beteiligung bei der Forschungsplanung und -gestaltung belegt. Ein systematisches Review hat gezeigt, dass die Hauptbegründung dafür, Krebspatienten in Forschung, Politik, Planung und Praxis einzubinden, in der einzigartigen Perspektive besteht, die Patienten zur Forschung beitragen können. Die Auswirkung der Beteiligung wurde jedoch wenig untersucht. (28) Im Rahmen einer Pilotstudie von kurzer Dauer wurden Verbraucher in alle Phasen eines HTA eingebunden. Die Verbraucher leisteten einzigartige Beiträge zum HTA-Programm. Bei der Suche nach Forschungsthemen war das persönliche Gespräch mit einer Verbrauchergruppe produktiver als die Durchsicht von Berichten der Verbraucherforschung oder Kontakt zu Gesundheitsinformationsdiensten. Die Verbraucher waren bereit, aktiv als Mitglieder des Gremiums bei der Weiterentwicklung und Priorisierung von Themen sowie bei der Stellungnahme zu Forschungsplänen und -berichten mitzuwirken. (29) Einige Fallreihen haben beschrieben, dass der Input von Patienten zu Veränderungen bei den Methoden, Verfahren und Maßen geführt hat, die bei der Studienplanung eines RCT angewandt wurden, z. B. zu Brustkrebs (30) und Schlaganfall. (31)

Die Erfahrungen des NICE weisen auf die Möglichkeit hin, Stakeholder an Entscheidungen im Gesundheitswesen zu beteiligen, obgleich dies ein Engagement der gesamten Organisation sowie besondere Vereinbarungen auf der Geschäftsführungsebene erfordert und unter Umständen kostspielig sein kann. (32)

Obwohl die Teilnahme von Stakeholdern (sowohl Patienten als auch Personen aus allen relevanten Gruppen der Gesundheitsberufe) einer der Bereiche des AGREE-Instruments ist, das international angewandt wird, um die Qualität und Berichterstattung von klinischen Leitlinien zu bewerten, (33) gibt es nur wenig empirische Evidenz über die Bedeutung der Teilnahme von Gesundheitspersonal an der klinischen Forschung. Die Consumers' Advisory Group for Clinical Trials (CAG-CT) ist jedoch ein gutes Beispiel für den Wert eines solchen Beitrags. Diese Gruppe besteht sowohl aus Brustkrebspatienten sowie aus Mitgliedern der Gesundheitsberufe im Bereich Brustkrebs. Marsden et al (34) haben beschrieben, wie diese Gruppe einen angemessenen Beitrag zur Planung einer landesweiten randomisierten Brustkrebsstudie geleistet hat.

3.3 Fazit

Auf der Basis des gesetzlichen Rahmens muss eine angemessene Beteiligung der maßgeblichen Parteien bei allen Schritten des Verfahrens ermöglicht werden. Empirische Evidenz zeigt, dass Patienten und Gesundheitsdienstleister ihre eigenen Präferenzen in Bezug auf Forschungsprioritäten, die Behandlung und Outcomes haben.

Die maßgeblichen Parteien sollten ermittelt und kontaktiert werden, sobald die Bewertung hinreichend skizziert vorliegt. Danach sollten die maßgeblichen Parteien an der Definition der Forschungsfrage, der Protokollentwicklung und der Stellungnahme zum Vorbericht beteiligt werden. Während aller Phasen sollten schriftliche Stellungnahmen gefolgt von einer mündlichen Diskussion ermöglicht werden. Für die Forschungsfrage sollte jedoch ein explorativer Scoping-Workshop durchgeführt werden, an dem der G-BA, das Institut, die maßgeblichen Parteien sowie die externen

Sachverständigen teilnehmen. Schlussendlich sollten die maßgeblichen Parteien das Recht haben, gegen die Entscheidung des G-BA ein Appeal („Berufung“) einzulegen.

4. Anforderungen an die Bewertungsmethoden

Aus den einschlägigen Abschnitten des gesetzlichen Rahmens in Bezug auf die *Bewertungsmethoden* kann man schlussfolgern, dass

- das Institut gewährleisten muss, dass die Bewertung des medizinischen Nutzens nach international anerkannten Standards der evidenzbasierten Medizin (EbM) vorgenommen wird,
- die Nutzenbewertung im Vergleich zu anderen Arzneimitteln und/oder Behandlungsformen unter Berücksichtigung des zusätzlichen therapeutischen Patientennutzens durchgeführt wird. Dies erfordert die Definition eines oder mehrerer gegenwärtiger Behandlungsstandards, mit denen eine (neue) Intervention verglichen wird. Vielfach angewandte Kointerventionen sollten in die Bewertung einbezogen werden können.
- Der gesetzlich vorgeschriebene Mindestkatalog zur Bewertung von Patientennutzen umfasst die folgenden Kriterien:
 - Verbesserung des Gesundheitszustandes,
 - Verkürzung der Krankheitsdauer,
 - Verlängerung der Lebensdauer,
 - Verminderung von Nebenwirkungen sowie
 - Verbesserung der Lebensqualität.

In den nachfolgenden Abschnitten argumentieren wir, dass im Prinzip jedes zu betrachtende Outcome zu einer separaten Forschungsfrage führt und dass eine separate Ermittlung und Betrachtung der angemessenen Studientypen für jede Fragestellung nötig ist, um die jeweils beste verfügbare Evidenz zu finden.

4.1 Evidenzbasierte Medizin

Das Institut muss gewährleisten, dass die Nutzenbewertung nach anerkannten Standards der evidenzbasierten Medizin vorgenommen wird. EbM (oder besser gesagt: evidenzbasierte Gesundheitsversorgung) stellt die Integration der besten Forschungsevidenz mit klinischer Expertise und Patienten-Werten bei Entscheidungen über die Patientenversorgung dar. (35) Der Inhalt von EbM ist zeitnah, da sie sich auf die beste gegenwärtig verfügbare Evidenz bezieht, was auch als "bestverfügbare Evidenz" bezeichnet wird. Daher gibt es einen Unterschied zwischen optimaler Evidenz (wenn z. B. ein RCT durchgeführt werden kann) und der besten verfügbaren Evidenz (wenn z. B. keine RCTs für ein bestimmtes Outcome durchgeführt wurden und die beste Evidenz sich daher z. B. auf Kohortenstudien bezieht). Die beste verfügbare Evidenz impliziert die Existenz einer Hierarchie von Evidenzleveln.

4.1.1 Beste verfügbare Evidenz

Aus wissenschaftlicher Sicht ist die randomisierte kontrollierte Studie (RCT) das stärkste Studiendesign bei der Auswertung der Wirksamkeit von therapeutischen Interventionen. (36) Das Grundprinzip bei RCTs besteht darin, den Patienten nach dem Zufallsprinzip einer der gewählten Anzahl oder Arten von Interventionen zuzuweisen, um deren Auswirkungen zu vergleichen. Die zufällige Zuordnung der Patienten gewährleistet, dass die Patientengruppen durch Vermeidung des Selektionsbias vergleichbar sind. Eine Einschränkung von RCTs besteht darin, dass sie oft strenge Einschlusskriterien anwenden und folglich große Teile der Zielpopulation ausschließen. Ferner können sich jene Patienten, die beschließen, nicht an der Studie teilzunehmen, von denjenigen unterscheiden, die teilnehmen. Daraus ergibt sich, dass die in Studien dokumentierten Auswirkungen vielleicht nicht für jene Auswirkungen repräsentativ sind, die beobachtet würden, wenn die Interventionen auf die gesamte Zielpopulation angewandt würde. Ferner erkennen RCTs nur das, wonach sie suchen, und sie haben oft einen begrenzten Untersuchungszeitraum. Es ist beispielsweise unwahrscheinlich, dass RCTs seltene Nebenwirkungen oder Langzeitwirkungen berichten. Daher stellt eine RCT nicht immer das einzige bzw. optimale Studiendesign für Fragestellungen zu den Auswirkungen der

Gesundheitsversorgung dar, das optimale Studiendesign hängt also von den zu bewertenden Outcomes ab.

4.1.2 Interne vs. externe Validität

Das Ausmaß der Minimierung von Bias in einer klinischen Studie wird als interne Validität bezeichnet. Interne Validität ist dabei definiert als das Ausmaß, bis zu dem die Ergebnisse einer Studie in Bezug auf die untersuchten Aspekte korrekt sind. (37) Externe Validität bezieht sich hingegen auf das Ausmaß, bis zu dem die Ergebnisse einer Studie verallgemeinerbar auf andere Aspekte sind, (37) z. B. andere Patientenpopulationen und Interventionen einschließlich unterschiedlicher Vergleichstechnologien. Das Studiendesign von RCTs ist normalerweise durch hohe interne Validität geprägt, manchmal auf Kosten der Aussagefähigkeit zur Anwendbarkeit, der externen Validität. Die hohe interne Validität wird auch als Wirksamkeit einer Intervention unter Idealbedingungen ("efficacy") bezeichnet. Observationsstudien können dagegen höhere externe Validität auf Kosten der internen Validität haben, und dabei die Bedingungen der Behandlungspraxis besser erfassen als Studien in einem klinischen Setting (RCTs). Die hohe externe Validität wird auch als Wirksamkeit unter Alltagsbedingungen ("effectiveness") bezeichnet. Beide Studiendesigns können einen Beitrag zur "bestverfügbaren Evidenz" leisten, und die Einschränkungen des jeweiligen Studiendesigns sollten bei der Formulierung von Empfehlungen berücksichtigt werden.

Basierend auf dem Prinzip der "bestverfügbaren Evidenz" ist keine Schlussfolgerung möglich, dass es zu wenig Evidenz gibt, um eine Nutzenbewertung durchzuführen. Die Beauftragung einer Nutzenbewertung wird veranlasst durch die Fragestellung, zu der der G-BA eine Entscheidung treffen muss. Solch eine Entscheidung kann im Prinzip auf dem niedrigsten Evidenzlevel getroffen werden (Level-V-Evidenz, Expertenmeinung). Falls keine RCTs durchgeführt wurden oder falls die RCTs keine (validen) Informationen für das jeweilige Outcome berichten, sollten andere Studien bewertet und die Ergebnisse einbezogen werden. Aus diesem Grunde sollte die Stärke der analysierten Evidenz stets zusammen mit der Empfehlung präsentiert werden.

Es gibt keinerlei empirische Evidenz, die eine Mindestzahl der zu berücksichtigenden Studien vor Abgabe von Empfehlungen unterstützt. Ganz im Gegenteil deutet die empirische Evidenz darauf hin, dass manchmal eine Studie genug ist und manchmal wiederum viele Studien noch zu wenig Evidenz liefern. Drei Studien haben die Ergebnisse von Meta-Analysen mit denen von klinischen Studien verglichen. (38-40) Ein systematischer Vergleich dieser drei empirischen Bewertungen kam zu dem Schluss, dass die Nichtübereinstimmung vielleicht weniger ausgeprägt ist für primäre Outcomes und dass die Gesamthäufigkeit von signifikanten, über den Zufall hinausgehenden Nichtübereinstimmungen 10 bis 25% beträgt. (41;42) Nichtübereinstimmung kann auch bei klinischen Studien (43) und bei Meta-Analysen vorliegen. (44). Diese Diskrepanzen legen nahe, dass ein dogmatischer Ansatz in Bezug auf eine benötigte Mindestanzahl von Studien schwer zu vertreten ist. Gemäß der Definition von evidenzbasierter Medizin sollte stattdessen die bestverfügbare Evidenz verwendet werden, um die klinische Fragestellung zu beantworten. Alle Evidenz sollte sorgfältig überprüft werden, um festzustellen, wie gut sie der klinischen Fragestellung entspricht im Hinblick auf die Merkmale von Patienten, Interventionen und Outcome-Ereignissen. Um zu bestimmen, welche Evidenz die beste ist, müssen solche Faktoren wie Studiengröße und Studienqualität untersucht werden, um die Validität der Studie oder Meta-Analyse herauszufinden.

Die Wichtigkeit, Evidenz aus anderen Studiendesigns als nur aus RCTs zu verwenden, wird durch die aktualisierten Methoden des NICE veranschaulicht, in denen es heißt: "Evidenz, die nicht aus RCTs stammt, ist notwendig, und zwar nicht nur für Situationen, in denen RCTs nicht verfügbar sind, sondern auch, um Informationen zu RCTs zu ergänzen, wenn diese verfügbar sind." (45) Diese Aussage ist in den aktuellen NICE-Methoden deutlicher als zuvor, wo es noch hieß: "In einigen Situationen kann Evidenz, die nicht aus RCTs stammt, erforderlich sein, um die aus RCTs verfügbare Evidenz zu ergänzen". (12)

Zur Berücksichtigung der Unsicherheit in der verfügbaren Evidenz haben Claxton et al (46) die Verwendung eines neuen Entscheidungsrahmens für das NICE vorgeschlagen.

Die Analyse kombiniert explizit alle verfügbaren Datenkalkulationen in Bezug auf die Unsicherheit und ermittelt, welche Intervention für eine bestimmte Patientengruppe die höchste erwartete Kosteneffektivität hat. Dieser Ansatz bedarf weiterer Entwicklung.

4.1.3 Die Rolle systematischer Reviews in der EbM

Die durch die Definition der EbM eingeführte Evidenzhierarchie wird normalerweise auf der Basis der internen Validität der jeweiligen Studien als "Evidenzlevel" bezeichnet. Die Evidenzlevel zielen darauf ab, die einer Schlussfolgerung zugrundeliegende Evidenz transparent zu machen. Es gibt mehrere Ansätze, um die Stärke der Evidenz zu bewerten (z. B. (47, 48)). Für Fragestellungen der Wirksamkeit unter Alltagsbedingungen verwendet das NICE z. B. ein Vierstufensystem, um die verschiedenen Studiendesigntypen nach ihrer relativen internen Validität zur Einschätzung des relativen Behandlungseffekts einzustufen:

- Level 1: Randomisierte klinische Studien (RCTs),
- Level 2: Kontrollierte Observationsstudien, z.B. Kohortenstudien und Fallkontrollstudien,
- Level 3: Observationsstudien ohne Kontrollgruppe, z. B. Fallserien, und
- Level 4: Sachverständigengutachten.(12)

Aus der Sicht der EbM ist das systematische Review von RCTs die beste und nützlichste verfügbare Evidenz. (35) Daher sollte dieses Studiendesign das höchste Evidenzlevel darstellen. Dies wird veranschaulicht durch das Einstufungssystem des Oxford Centre for Evidence-based Medicine (49), das eine fünfstufige Hierarchie erstellt hat, in der die systematischen Reviews von RCTs für Fragen der Wirksamkeit von Behandlungsinterventionen unter Alltagsbedingungen auf der höchsten Stufe stehen. Dieses Schema stellt auch Evidenzlevel für andere Studientypen wie prognostische und diagnostische Studien dar. Immer, wenn ein systematisches Review für die Forschungsfrage schon vorliegt, sollte genau untersucht werden, ob ein neues systematisches Review dann noch erforderlich ist. Die Entscheidung, ein bereits existierendes Review auszuschließen, z. B. wenn dessen Qualität schlecht ist, sollte im Vorbericht begründet werden.

4.1.4 Bewertung der Evidenz für jedes Outcome

Da die Nutzenbewertung auf der Grundlage von verschiedenen patientenrelevanten Outcomes vorgenommen wird, muss die Bewertung für jedes dieser Outcomes ebenfalls auf dem Prinzip der "bestverfügbaren Evidenz" durchgeführt werden. Der Auftrag zur Nutzenbewertung eines Arzneimittels führt üblicherweise zu einer Reihe von Forschungsfragen, daher kann für jede Forschungsfrage (oder jeden Outcomeaspekt) eine separate Suche erforderlich sein. In diesem Fall sollten für jede Frage (jeder Outcomeaspekt) alle in Frage kommenden Studientypen bewertet werden. Falls es für die relevante Fragestellung (jeden Outcomeaspekt) mehrere hochqualitative RCTs gibt, kann die Evidenz von einem niedrigeren Level, wie z. B. nichtrandomisierte Studien und Observationsstudien, ausgeschlossen werden. Wenn es jedoch nur wenige bzw. nur kleine RCTs oder solche von schlechter Qualität gibt, sollte die Evidenz auf niedrigeren Levels ebenfalls berücksichtigt werden.

Danach wird die Evidenz separat für jedes Outcome auf ähnliche Art und Weise ausgewertet. Dies erfordert einen eher flexiblen als dogmatischen Ansatz in Abhängigkeit davon, ob das höchste verfügbare Evidenzlevel Antworten auf die Fragen liefert. Oft erhält man darüber erst im Laufe der Bewertung Klarheit.

4.1.5 Evidenzbasierte Methoden zur Vermeidung von Bias in der Bewertung

Für den Bereich der systematischen Reviews gibt es zwei international führende Organisationen, die Handbücher entwickelt haben, wie systematische Reviews von randomisierten klinischen Studien durchzuführen sind: die Cochrane Collaboration (6) und das Centre for Reviews and Dissemination (CRD). (50) Beide Handbücher beschreiben explizite Methoden, wie man den Bias im Review begrenzen und verlässlichere Ergebnisse liefern kann. Beides ist erforderlich, um valide Schlussfolgerungen zur Entscheidungsfindung zu erlangen. Es folgt ein Überblick über wichtige Einschränkungen bei systematischen Reviews, die zu Bias führen können, und wie diese behandelt werden sollten. Für weitere Einzelheiten zu diesen Einschränkungen sowie zu anderen Methoden, Bias bei der Bewertung zu vermeiden, verweisen wir auf die Handbücher.

Publikationsbias

Für Studien, die vorteilhafte Ergebnisse präsentieren, ist die Wahrscheinlichkeit größer, veröffentlicht und in systematische Reviews einbezogen zu werden, wodurch ein Bias wirken kann. (51) Als Mittel, um bei der Recherche von Studien gegen Bias vorzugehen, wird eine umfangreiche Suchstrategie umgesetzt: Die Suche erfolgt in mehreren Datenbanken sowie unter Verwendung von mehreren Quellen (verschiedene elektronische Datenbanken, manuelle Suche, Studienregister, Referenzlisten, etc.). Wo möglich, sollten statistische Methoden angewandt werden, um die Evidenz auf Publikationsbias zu untersuchen. (6, 50) Publikationsbias bei den zugrundeliegenden Studien schwächt die Aussagekraft von Schlussfolgerungen in einem systematischen Reviews.

Reporting-Bias (Berichts-Bias)

Innerhalb von Studien ist die Wahrscheinlichkeit größer, dass vorteilhafte Ergebnisse berichtet werden. Daher ist die Wahrscheinlichkeit auch größer, dass diese in systematische Reviews einbezogen werden. (51) In einem systematischen Review sollten jedoch alle Ergebnisse zu allen relevanten Outcomes dargestellt werden. Dies hängt jedoch immer noch von den Outcomes ab, die in den Primärstudien berichtet werden. Die Registrierung von Studien ist das einzige Mittel, um das Problem des Berichtsbias zu vermindern.

Methodische Qualität von Primärstudien

Die Qualität der in das Review einzubeziehenden Primärstudien ist von größter Wichtigkeit. Es liegt empirische Evidenz vor, die zeigt, dass eine mangelnde Studienqualität die Ergebnisse systematischer Reviews verzerren kann. (52) Daher muss die methodische Qualität aller einzubeziehenden Studien bewertet werden. Auch der Einfluss der Qualität der einbezogenen Studien auf ihre Ergebnisse sollte untersucht werden. (52) Die Verwendung von Summenscores aus Qualitätsskalen ist problematisch. (6) Auf der Basis der empirischen Evidenz sollten die versteckte Zuordnung („Concealment of Allocation“), die Verblindung der Outcomebewertung und der

Umgang mit Studienabbrüchen in der Analyse allgemein bewertet werden. (52) Es ist nicht immer möglich, alle Anforderungen an die methodische Qualität in einer Studie abzubilden, so ist es z. B. nicht möglich, Patienten für psychologische Interventionen zu verblinden.

4.2 Wahl der Vergleichsgrößen

Die Vergleichsgröße ist entweder die bestmögliche Behandlung oder die gegenwärtige Routinebehandlung. Obwohl es sich bei der besten Behandlung um die bevorzugte Vergleichsgröße handelt, sollten Behandlungen, die die deutsche Routineversorgung repräsentieren, ebenfalls in die Auswertung einbezogen werden. Abhängig von regionalen Unterschieden kann es mehrere Vergleichsbehandlungen geben. Die Vergleichsgröße muss so präzise wie möglich definiert werden, besonders wenn sich die Anwendung im Vergleich zu der zu bewertenden Intervention unterscheidet. Die Wahl einer (oder mehrerer) Vergleichsgröße(n) muss während des Scoping-Prozesses erörtert und im Protokoll begründet werden.

Während des Zulassungsverfahrens werden Arzneimittel üblicherweise mit Placebo verglichen. Solche Studien beantworten die Frage, ob das Arzneimittel wirksamer ist als Placebo. Für Nutzenbewertungen aus der Perspektive des Gesundheitswesens werden Kopf-an-Kopf-Studien bevorzugt, bei denen ein Arzneimittel mit einem anderen verglichen wird, falls es sich bei der Vergleichstherapie um die gegenwärtige Standardtherapie handelt. Kopf-an-Kopf-Studien sollten auf die gleiche Art und Weise ausgewertet werden wie Studien mit Placebo-Kontrolle. Wenn die Zuordnung zu beiden Behandlungen nach dem Zufallsprinzip erfolgt ist, handelt es sich bei solchen Studien um Level-1-Evidenz. Wenn nur Studien mit Placebo-Kontrolle verfügbar sind, kann der Zusatznutzen von Arzneimitteln durch Verwendung von adjustierten indirekten Vergleichen eingeschätzt werden. (53, 54)

4.3 Nutzen

Nutzenbewertungen untersuchen Nutzen (und Schaden) eines Arzneimittels. Die gesetzlichen Anforderungen verlangen, dass folgende Nutzendimensionen in die Bewertung einfließen:

- Verbesserung des Gesundheitszustandes,
- Verkürzung der Krankheitsdauer,
- Verlängerung der Lebensdauer,
- Verringerung von Nebenwirkungen, und
- Verbesserung der Lebensqualität.

Es handelt sich hierbei um eine minimale Liste, andere Nutzenparameter sollten nach Bedarf und abhängig von der Forschungsfrage einbezogen werden. Der Scoping-Prozess sollte deshalb alle relevanten Outcomes ermitteln, die in die Bewertung einfließen sollten, z. B. auch Patientenzufriedenheit.

Nutzen ist ein subjektives Konzept, und Patienten, Kliniker und Forscher haben möglicherweise unterschiedliche Meinungen, wie Nutzen zu definieren ist. Selbst innerhalb einer Gruppe von Patienten oder innerhalb einer Gruppe von Berufsexperten kann "Nutzen" unterschiedlich interpretiert werden. Ferner hat das Bewertungsthema einen Einfluss darauf, was als Nutzen betrachtet wird. Für jede Bewertung sollte der Nutzen deshalb während des Scoping-Prozesses definiert werden. Hierfür ist Input von allen maßgeblichen Parteien entscheidend. Die unterschiedlichen Nutzentypen können unterschiedliche Studiendesigns beim Review erfordern. Nutzen, der als geringere Mortalität oder Morbidität definiert wird, kann z. B. unter Verwendung von RCTs bewertet werden. Allerdings sind Observationsstudien vielleicht besser geeignet, um die Frage zu beantworten, falls Nutzen als weniger unerwünschte Ereignisse definiert wird.

4.4 Fazit

Das Institut muss gewährleisten, dass die Bewertung nach international anerkannten Standards der EbM durchgeführt wird. Um die Transparenz zu erhöhen, sollte für jede

Fragestellung dargelegt werden, welche Evidenzlevel für die Bewertung herangezogen wurden. Um die bestverfügbare Evidenz für die Forschungsfrage zu erhalten, sollte alle Evidenz berücksichtigt werden, um zu ermitteln, welche Studien die bestverfügbare Evidenz zur Beantwortung der Fragestellung darstellen. Das optimale Studiendesign – selbst für Fragen der Wirksamkeit unter Idealbedingungen – ist nicht immer das RCT, sondern dies hängt von der Forschungsfrage und den zu betrachtenden Outcomes ab. In Abhängigkeit von den Fragestellungen und Outcomes müssen ggf. separate Suchstrategien getrennt für jedes Outcome verfolgt werden.

Es gibt viele Arten und Weisen, wie Bias in systematische Reviews einfließen kann. Einige Arten von Bias können vermieden werden, während andere nur berichtet werden können. Für andere Biasarten wiederum kann der Einfluss des Bias untersucht werden. Reviews müssen zeigen, dass potenzielle Biasquellen angemessen behandelt wurden. Die Vergleichsgröße ist entweder die beste Behandlung oder die gegenwärtige Routinebehandlung. Es kann mehrere Vergleichsbehandlungen geben. Dies sollte während des Scoping-Prozesses diskutiert werden.

Für jede Bewertung sollte der Nutzen während des Scoping-Prozesses definiert werden, da der Input von allen maßgeblichen Parteien für dieses Thema entscheidend ist.

5. Internationale Standards / Von NICE und IQWiG angewandte Methoden

Die deutsche Gesetzgebung verlangt in §35b SGB V die Anwendung von internationalen methodischen Standards. Das Nichtvorhandensein einer überregionalen Organisation, die verbindliche Standards definiert, bedeutet nicht, dass es keine international anerkannten methodischen Standards gibt. Es gibt eine Reihe von Organisationen und Kooperationen, die Richtlinien für ihre Bewertungen veröffentlicht haben. Diese Richtlinien reflektieren weltweit anerkannte Methoden, die unserer Ansicht nach als internationale Standards angesehen werden sollten.

Dieser Abschnitt bietet einen Überblick über eine Auswahl dieser Verfahren und Methoden. Zu diesem Zweck werden die Methoden der Institute, die Technologien auf nationaler Ebene für die jeweilige Regierung oder im Regierungsauftrag bewerten, mit den vom IQWiG angewandten Methoden verglichen. Wir konzentrieren uns auf das

National Institute for Health and Clinical Excellence (NICE), da es sich hier um ein führendes Institut handelt, dessen Methoden größtenteils für die Öffentlichkeit zugänglich sind. Die Methoden der EUR-ASSESS-Gruppe, die über die beste Praxis bei der Durchführung und Berichterstattung zu Health Technology Assessments berichtet (8), der Canadian Agency for Drugs and Technology in Health (CADTH, früher CCOHTA), (55) sowie mehrerer europäischer Institutionen wie der französischen Gesundheitsbehörde (Haute Autorité de santé, HAS), (56) dem niederländischen Gesundheitsversicherungsgremium (College voor Zorgverzekeringen, CvZ), (57) und dem Dänischen Institut für Health Technology Assessment (36) wurden bewertet, aber es war aufgrund der begrenzt verfügbaren Einzelheiten zu deren Methoden kein aussagekräftiger Vergleich möglich.

5.1 Die Methoden des NICE und des IQWiG

Da die Methoden von Instituten mit ihrer Organisation zusammenhängen, werden diese Aspekte nachfolgend skizziert. Das NICE ist eine unabhängige Organisation innerhalb des britischen Gesundheitsdienstes National Health Service (NHS) und wird zentral aus Regierungsgeldern gefördert. (58) Das Institut entscheidet für den NHS in England und Wales über die Anwendung von ausgewählten neuen und etablierten Technologien. Ferner nimmt das Institut auf Anfrage des britischen Gesundheitsministeriums und der walisischen Regierung Relevanzeinstufungen von Gesundheitstechnologien vor. (12)

Das IQWiG ist als regierungsunabhängige gemeinnützige Stiftung privaten Rechts gegründet worden. Das Institut ist rechtsfähig und in diesem Sinne eine unabhängige Organisation. Das IQWiG ist für die wissenschaftliche Auswertung von Nutzen, Schaden, Qualität und Wirtschaftlichkeit von Gesundheitsleistungen zuständig. Es ist dafür verantwortlich, den G-BA bei der Erfüllung seiner legislativen Pflichten zu unterstützen, indem es Empfehlungen vorlegt und zur ständigen Verbesserung der Versorgungsqualität für die Öffentlichkeit beiträgt. (59) Das IQWiG wird durch den G-BA über Zuschläge für stationäre und ambulante medizinische Leistungen mit einem definierten Prozentsatz

finanziert. Reguläre formale Rechtsklagen sind bei Nutzen- oder Kosten-Nutzen-Bewertungen explizit ausgeschlossen (§ 35b Abs. 4 SGB V).

Die Beziehung zwischen G-BA und IQWiG ist in der Satzung des Instituts in § 7 Abs. 5 Satz 2 wie folgt definiert: "Die auf der Grundlage des §91 Abs. 3 SGB V vom Gemeinsamen Bundesausschuss beschlossenen Verfahrensregelungen sind, soweit es die Einbeziehung des Instituts betrifft, zu beachten. Die in der Verfahrensordnung zu regelnden methodischen Anforderungen an die wissenschaftliche, sektorübergreifende Bewertung von Maßnahmen und die Anforderungen an die fachliche Unabhängigkeit von Sachverständigen sind in enger Abstimmung mit der Institutsleitung zu definieren." Dadurch ist das IQWiG verpflichtet zu gewährleisten, dass es seine Arbeit auf solche Art und Weise verrichtet, dass der G-BA gemäß seiner Verfahrensordnung mit dem IQWiG zusammenarbeiten kann.

5.2 Der von NICE und IQWiG angewandte Bewertungsprozess

Der Prozess des NICE ist ausführlich beschrieben worden. (9, 12) Die Beteiligung der maßgeblichen Parteien ist einer der wichtigsten Grundsätze im Prozess des NICE. Organisationen, die an der Relevanzeinstufung teilnehmen möchten, werden so früh wie möglich ermittelt, sobald die Minister eine vorläufige Entscheidung in Bezug auf die Liste der einzustufenden Technologien getroffen haben. (9) Diese Organisationen werden zur Teilnahme am Scoping- und Bewertungsprozess als Ratgeber oder Kommentatoren eingeladen. (9) Bei den Ratgebern handelt es sich um Organisationen, die an der Bewertung und Relevanzeinstufung teilnehmen, wie z. B. Hersteller bzw. Kostenträger der Technologie oder nationale Patientenorganisationen. Sie können zu verschiedenen Dokumenten und Verfahrensschritten Stellung nehmen, können einen schriftlichen Beitrag einreichen und gegen die endgültige Entscheidung Berufung einlegen. Zu den Stellungnahmeberechtigten gehören Organisationen wie z. B. die Hersteller von Vergleichstechnologien. Sie können die verschiedenen Dokumente nur kommentieren. Alle Stellungnahmen durch die Ratgeber oder Stellungnahmeberechtigte, einschließlich der darauf gegebenen Antworten, werden auf der Website veröffentlicht.

Der Scoping-Prozess besteht aus der Vorlage von Schriftstücken sowie einer Sitzung, die als Scoping-Workshop bezeichnet wird, bei der die aufgeworfenen Fragestellungen erörtert werden. Der Scoping-Workshop wurde anlässlich der Aktualisierung der Methoden im Jahre 2004 zur Phase der schriftlichen Stellungnahme hinzugefügt, (9, 60) was darauf hindeutet, dass der ausschließliche Einsatz von schriftlichen Stellungnahmen keine optimale Lösung für den Prozess ist. Die Evidenzbewertung wird von einer unabhängigen Gruppe durchgeführt, während das NICE die Relevanzeinstufung (das „Appraisal“) der Evidenz vornimmt und Empfehlungen formuliert.

Um die Arbeit des NICE auszuwerten und zu verbessern, wurde das Institut zunächst im Jahre 2001/02 einer Evaluation unterzogen, gefolgt von einer weiteren Auswertung im Jahre 2007/08. Ein darauf folgender Bericht beschrieb Arbeitsinhalte und Arbeitsweise des NICE, welche Änderungen seit seiner Gründung vorgenommen wurden, und mit welchen neuen Herausforderungen es konfrontiert wird. Die Autoren kamen zu dem Schluss, dass das Institut eine wichtige Aufgabe unter schwierigen Umständen erfüllt, und formulierten Empfehlungen zur Verbesserung seiner Funktionsweise.(61)

Der Hauptunterschied zwischen der Situation in Deutschland und England in Bezug auf den Scoping-Prozess besteht im formalen Aufbau der Institute.

Beim NICE ist der Scoping-Prozess der Beginn aller Aufträge und führt zu einem detaillierten Protokoll, das direkt von den beauftragten Sachverständigen verwendet wird.

Der Scoping-Prozess ist ein offenes und transparentes Verfahren mit Teilnahme der externen Sachverständigen, und ihre Namen und alle Produkte werden veröffentlicht.

Beim IQWiG ist die Situation weniger transparent. Das IQWiG kann den vom G-BA entwickelten Umfangsentwurf verändern, was zu einer Bewertung führen kann, die nicht ganz den Erfordernissen des G-BA entspricht. In seinen Methoden beschreibt das IQWiG die Notwendigkeit, die externen Sachverständigen geheim zu halten, was eine Teilnahme am Scoping-Prozess und ein transparentes Reviewverfahren direkt behindert. Unserer Ansicht nach sollten die Namen der externen Sachverständigen kurz nach der Beauftragung veröffentlicht werden. Ein weiteres Problem besteht in der Aufteilung der gesetzlichen Zuständigkeit zwischen G-BA und IQWiG. Der G-BA ist aufgrund der gesetzlicher Regelungen die allein als autorisiert zu betrachtende Institution, die eine

Relevanzeinstufung (ein „Appraisal“) vornehmen kann, während das IQWiG und seine externen Sachverständigen dieser gesetzlichen Regelung folgend nur die Evidenzbewertung durchführen sollen. Dennoch ist es dem IQWiG gestattet, Empfehlungen abzugeben, was zu einer Überschneidung zwischen Bewertung („Assessment“) und Relevanzeinstufung („Appraisal“) führen kann.

Das IQWiG bindet die maßgeblichen Parteien in Teile des Bewertungsprozesses ein. Die Prozesse sind jedoch unklar. In Bezug auf das Scoping eines Projekts entwirft das IQWiG die (grobe) Definition der Forschungsfragen. Falls nötig, wird diese Definition von der Projektgruppe unter Einbindung der externen Sachverständigen (falls erforderlich) präzisiert. Individuelle Betroffene, Patientenvertreter und/oder Verbraucherorganisationen werden "regelmäßig in die themenbezogene Definition patientenrelevanter Endpunkte einbezogen."(62)

Das IQWiG stellt seine Berichtspläne und Vorberichte ins Internet mit dem Ziel, schriftliche Stellungnahmen einzuholen. Der Umfang der Stellungnahmen, die abgegeben werden können, ist jedoch beschränkt, ebenso wie die Zeitspanne, innerhalb derer kommentiert werden darf. Der Zeitraum für Stellungnahmen zu den Produkten wird als zu kurz erachtet.(63) Im Ergebnis bleibt jedoch unklar, wie die Stellungnahmen behandelt werden und ob alle eingereichten Stellungnahmen veröffentlicht werden. Nur Änderungen des Berichtsplans werden in den Berichten beschrieben, was bedeutet, dass die Kommentare der Reviewer nur indirekt dadurch bekannt werden, dass das IQWiG diesen zustimmt. Ferner werden nicht auf Deutsch verfasste Stellungnahmen ausgeschlossen. Sachverständige müssen der deutschen Sprache mächtig sein, was den Pool von Experten einschränkt.

Tabelle 1 präsentiert einen Überblick über die Methoden des IQWiG (64) und des NICE. (45) Beide Institute haben im November 2007 einen Aktualisierungsentwurf ihrer Methoden veröffentlicht. Die Informationen in der nachfolgenden Tabelle basieren auf den jeweiligen Aktualisierungsentwürfen der Methoden, da diese alle Änderungen der jüngsten Vergangenheit beinhalten dürften.

Tabelle 1. Übersicht über die wichtigsten Verfahrensfragen von IQWiG und NICE

	IQWiG	NICE
<i>Themenermittlung und Priorisierung</i> - Kriterien der Themenauswahl werden veröffentlicht* - Kriterien der Themenpriorisierung werden veröffentlicht	+/- ¹ - ²	+ ¹⁵ Na ¹⁶
<i>Ausschreibung und Auftragsvergabe</i> - Kriterien für Sachverständige werden veröffentlicht - Kriterien der Expertenauswahl werden veröffentlicht - Auftragsverfahren wird veröffentlicht	+ ³ - ⁴ - ⁵	+/- ¹⁷ - ¹⁸ -
<i>Forschungsfrage</i> - Maßgebl. Parteien nehmen an dieser Phase teil - Mündl. Anhörung wird in dieser Phase durchgeführt - Alle Stellungnahmen der maßgebl. Parteien werden veröffentlicht und ausgewertet	- ⁶ +/- ⁷ Na ⁸	+ ¹⁹ + + ²⁰
<i>Protokoll</i> - Das Protokoll wird veröffentlicht - Maßgebl. Parteien werden in dieser Phase einbezogen - Mündl. Anhörung wird in dieser Phase durchgeführt - Alle Stellungnahmen der maßgebl. Parteien werden veröffentlicht und ausgewertet	+ ⁹ +/- ¹⁰ +/- ⁷ +/- ¹¹	+ ²¹ - ²² - Na
<i>Vorbericht</i> - Bericht der Sachverständigen wird veröffentlicht - Maßgebl. Parteien werden in dieser Phase einbezogen - Mündl. Anhörung wird in dieser Phase durchgeführt - Alle Stellungnahmen der maßgebl. Parteien werden veröffentlicht und ausgewertet	- + +/- ^{7 / 12} +/- ¹³	+ ²³ + - +
<i>Reviewverfahren</i> - Einladungskriterien für Reviewer werden veröffentlicht - Alle Stellungnahmen der Reviewer werden veröffentlicht	- -	- ²⁴ -
<i>Endbericht</i> - Relevanzeinstufung der Ergebnisse wird separat von der Bewertung der Ergebnisse veröffentlicht - Die zugrundeliegende Evidenz für jede Empfehlung wird veröffentlicht	- +	+ ²⁵ +
<i>Berufung</i> - Berufung möglich?	Na ¹⁴	+ ²⁶

Kriterium wird: +: erfüllt, +/- teilweise erfüllt, oder - nicht erfüllt; Na=nicht anwendbar

* über das Internet zugänglich

- 1) Obwohl sich diese Methoden auf breitere Kriterien wie Krankheitsbelastung und Kostenbelastung beziehen, konnten keine spezifischeren Kriterien gefunden werden.
- 2) Das IQWiG erhält Aufträge zu mehreren Themen zur selben Zeit und entscheidet selbst, welche zuerst bearbeitet werden. Es ist unklar, wie diese Themen priorisiert werden.
- 3) Kriterien werden veröffentlicht, um die Erfahrung der Sachverständigen zu beschreiben (Methoden, Version 2, Seite 98).
- 4) Es ist unklar, wie die Sachverständigen aus dem Pool von Experten, die die Kriterien erfüllen, ausgewählt werden.
- 5) Das Ausschreibungsverfahren ist offen, aber die Entscheidungsfindung ist unklar.

- 6) In den neuen Methoden heißt es, dass die Forschungsfrage durch die Projektgruppe definiert wird (Methoden, Version 3, Seite 17, erster und dritter Absatz).
- 7) Mündliche Anhörungen sind jetzt eine Option in jedem Schritt des Verfahrens; allerdings entscheidet nur das IQWiG, ob eine Stellungnahme es wert ist, diskutiert zu werden. Hier fehlt es an Transparenz.
- 8) Das IQWiG beteiligt in dieser Phase keine maßgeblichen Parteien, daher gibt es keine Stellungnahmen.
- 9) Das Protokoll ist im Internet verfügbar, z. B. auf <http://www.iqwig.de/index.651.en.html>
- 10) Die maßgeblichen Parteien werden nur um schriftliche Stellungnahmen gebeten (Methoden, Version 2, Seite 102). Dennoch heißt es, dass sie einbezogen werden KÖNNEN, was bedeutet, dass die Beteiligung nicht immer durchgeführt wird. Es ist unklar, unter welchen Umständen sie nicht einbezogen werden.
- 11) Stellungnahmen werden im Internet veröffentlicht, aber es ist unklar, was mit Stellungnahmen gemacht wurde.
- 12) Eine mündliche Anhörung ist freigestellt. Es ist unklar, wann eine solche Anhörung durchgeführt wird (Methoden, Version 3, Seite 16).
- 13) Stellungnahmen werden veröffentlicht, aber es ist unklar, was mit den Stellungnahmen gemacht wird, z.B.: http://www.iqwig.de/download/N06-01A_Dokumentation_und_Wuerdigung_der_Stellungnahmen_zum_Vorbericht.pdf
- 14) Dies liegt außerhalb des Anwendungsbereichs der Methoden des IQWiG. Berufung ist auf der G-BA-Ebene möglich.
- 15) Die Kriterien werden im Verfahrensdokument veröffentlicht, auf Seite 3.
- 16) Es gibt in diesem Stadium keine echte Priorisierung: Angemessene Themen werden zum Scoping verwiesen. Auf der Basis des Scopings entscheiden die Minister, ob ein Thema beauftragt wird oder nicht.
- 17) Die akademischen Gruppen, die für das NICE Bewertungen vornehmen, werden durch die jeweiligen Ausschreibungen eingerichtet. Die Auswahl, welche Gruppe welche Bewertung vornimmt, hängt von der Kapazität, den Interessenkonflikten, der Expertise und den Präferenzen der Gruppen ab.
- 18) Das NICE arbeitet mit 7 akademischen Zentren zusammen (<http://www.nchta.org/publicationspdfs/infoleaflets/nice.pdf>). Keinerlei Nennung von Kriterien.
- 19) Ratgeber und Kommentatoren werden während eines Scoping-Prozesses konsultiert (Methoden, Seite 2 und 7).
- 20) Stellungnahmen werden veröffentlicht, z. B.: http://www.nice.org.uk/nicemedia/pdf/Asthma_children_comments_on_draftscope.pdf).
- 21) Das Protokoll wird veröffentlicht, z.B.: <http://www.nice.org.uk/guidance/index.jsp?action=byID&o=11711>.
- 22) Die Bewertungsgruppe entwickelt das Protokoll auf der Basis des Scoping-Prozesses (Methoden, Seite 7, Abschnitt 2.1.3).
- 23) Im Prozess heißt es, dass "Autoren für den Inhalt und die Qualität der Bewertung verantwortlich sind" (Verfahrensdokument, Abschnitt 4.4.1.5). Die Autoren werden im Bericht angegeben, z.B.: <http://www.nice.org.uk/nicemedia/pdf/AssessmentReportSenttoC&CAsthmaChildren.pdf>.
- 24) Zu dem Bericht der externen Sachverständigen kann von Stakeholdern zur Vorbereitung der Relevanzeinstufungssitzung Stellung genommen werden. Ferner wird der Endberichtsentswurf einem Peer-Review-Verfahren unterzogen sowie einem Review durch die Redakteure der Fachzeitschrift Health Technology Assessment.
- 25) Beschrieben in den Methoden, Seite 2.
- 26) Nur für Ratgeber (beschrieben im Leitfaden für Hersteller, Seite 15).

5.3 Fazit

Die angewandten Methoden anderer Institutionen, die Nutzenbewertungen durchführen, sind hilfreich bei der Interpretation des Begriffs "internationaler Standard", dem das Institut entsprechen muss. Das NICE zeigt, dass transparente Verfahren bei Nutzenbewertungen möglich sind, aber dass hierfür eine ausführliche Dokumentation notwendig ist. Seine Dokumente sollten anderen Instituten in dieser Hinsicht als Beispiel

dienen. Das NICE hat ein offenes und transparentes Verfahren umgesetzt in Bezug auf die Veröffentlichung der von externen Sachverständigen erstellten Bewertungen und die Stellungnahmen der Reviewer. Ferner vermeidet seine Trennung von Evidenzbewertung („Assessment“) und Relevanzeinstufung („Appraisal“) der Evidenz Interessenkonflikte während der letzten Phase der Bewertung.

Obwohl das IQWiG zu Stellungnahmen von Protokoll und Vorbericht aufruft und diese auf seiner Website veröffentlicht, werden die Stellungnahmen nicht klar einzeln ausgewertet. Daher ist es unklar, welche Stellungnahmen als relevant wahrgenommen werden oder nicht und ob diese miteinbezogen wurden. Die Teilnahme der maßgeblichen Parteien am Bewertungsprozess wird vom NICE ordnungsgemäß umgesetzt, was ein für alle maßgeblichen Parteien akzeptables Verfahren garantiert.

6. Wie sollte eine Nutzenbewertung durchgeführt werden?

Dieser Abschnitt beschreibt, wie eine Nutzenbewertung durchgeführt werden sollte. Er konzentriert sich auf in Deutschland vorzunehmende Bewertungen und basiert auf dem gesetzlichen Rahmen, welcher transparente Verfahren, Methoden und Kriterien sowie die Beteiligung von maßgeblichen Parteien und die Durchführung von Bewertungen nach anerkannten Standards der evidenzbasierten Medizin verlangt. Wo anwendbar, werden die Erfahrungen von ausländischen Institutionen, die solche Bewertungen durchführen, berücksichtigt. Für allgemeine methodische Leitlinien für Nutzenbewertungen verweisen wir auf die vom CRD (50) oder der Cochrane Collaboration (6) herausgegebenen Handbücher über systematische Reviews sowie auf das QUOROM-Statement (7) für allgemeine Leitlinien zur Berichterstattung für solche Bewertungen. Der Prozess der Nutzenbewertung wird in Abbildung 1 und 2 vorgestellt.

Abbildung 1

Prozess der Nutzenbewertung bis zur Publikation des Berichtsplans

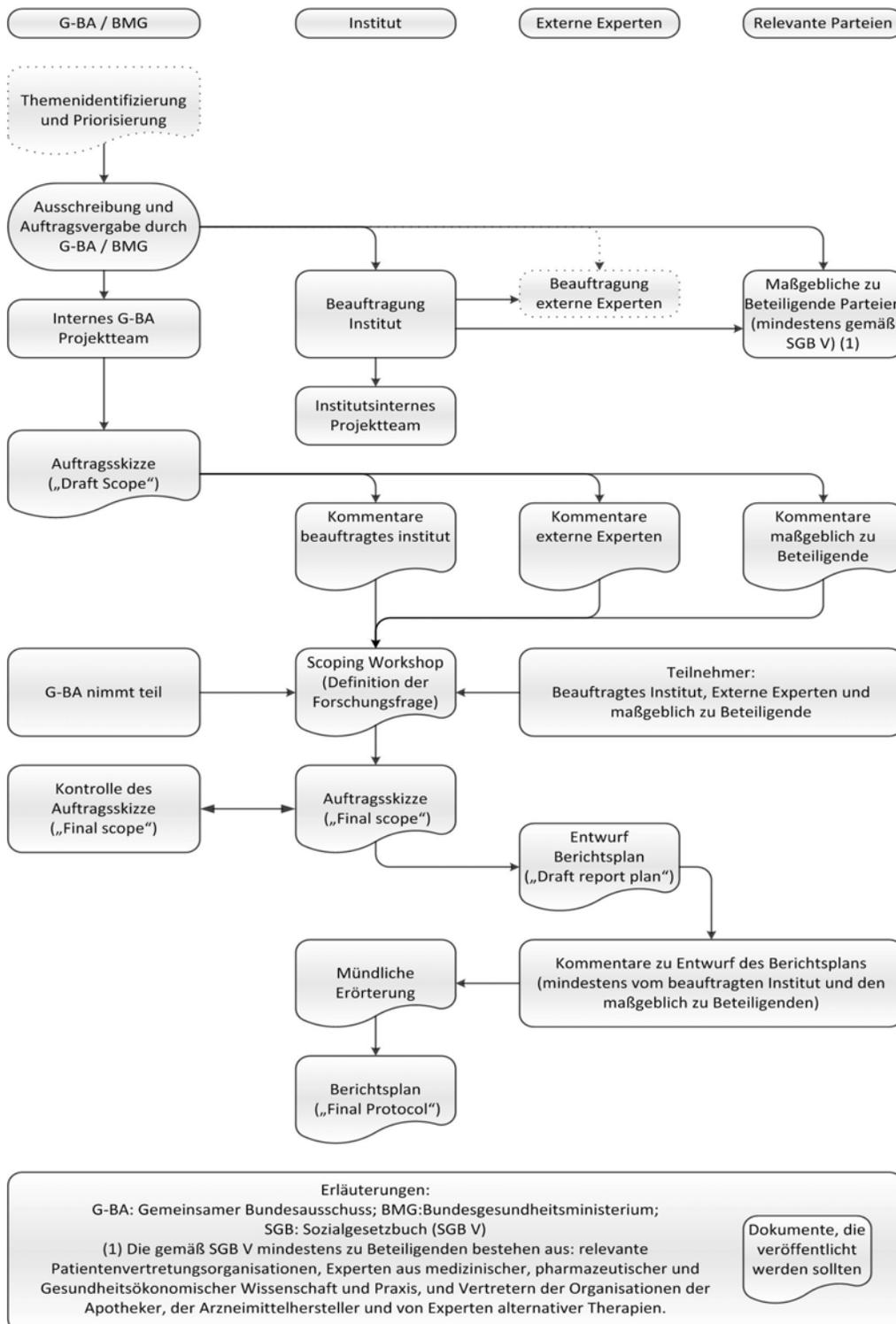


Abbildung 2

Prozess der Nutzenbewertung vom Berichtsplan bis zur Entscheidung



Der G-BA kann eine Nutzenbewertung oder eine volle ökonomische Bewertung (Kosten-Nutzen-Bewertung) beauftragen. Wenn eine volle ökonomische Bewertung beantragt wird, sollte zuerst eine Nutzenbewertung durchgeführt werden. Da die Methoden zur Bewertung des Nutzens für eine ökonomische Auswertung jedoch eine breitere Perspektive erfordern im Vergleich zu den Methoden zur Bewertung von Nutzen für eine Nutzenbewertung („isolierte Nutzenbewertung“), sollten die Methoden für die Nutzenbewertung erweitert werden, falls eine volle ökonomische Auswertung des Arzneimittels gewünscht wird. Außerdem sollte man wissen, dass eine Nutzenbewertung auf der Basis von RCTs wichtige Nachteile hat, wenn diese Bewertung die Grundlage für eine ökonomische Auswertung darstellen soll, z. B. aufgrund der selektiven Patientenpopulation, die sich von der allgemeinen Population unterscheidet. Für weitere Einzelheiten zu den Methoden von ökonomischen Auswertungen nach internationalen Standards verweisen wir auf von der Schulenburg et al (4) und Antes et al (63). Dieser Abschnitt konzentriert sich auf die Nutzenbewertung ohne nachfolgende ökonomische Auswertung.

6.1 Themenermittlung und Priorisierung

Zur Themenermittlung und Priorisierung für potenzielle künftige Nutzenbewertungen sollte ein öffentliches Verfahren entwickelt werden. Jeder sollte Themen vorschlagen können, z. B. durch Verwendung einer Website. Regulär sollte eine potenzielle Themenliste vom G-BA erstellt werden auf der Basis eines Kriteriensatzes, wie z. B.: "Die Intervention führt wahrscheinlich zu einem bedeutenden Gesundheitsnutzen in ganz Deutschland, wenn sie bei allen Patienten angewandt wird, für die sie indiziert ist." (9). Die Liste der Kriterien sollte auf der Website veröffentlicht werden. Die Themen sollten anschließend vom BMG überprüft werden, das entscheidet, welches Thema für eine Bewertung ausgewählt werden sollte. Die Gründe für die (Nicht-)Auswahl eines Themas aus der potenziellen Themenliste sollten zur Gewährleistung eines transparenten Verfahrens veröffentlicht werden.

Zur Vorbereitung der Nutzenbewertung wird ein Entwurf abgefasst, der die Bewertung hinreichend skizziert. Dieses Dokument enthält eine erste Beschreibung der

Patientenpopulation, die Intervention und ihr Setting, die Vergleichsintervention(en) und die vorgeschlagenen Outcome-Dimensionen und -maße. Dieses Verfahren sorgt dafür, dass die durchzuführende Bewertung die Bedürfnisse des G-BA erfüllt. Zur selben Zeit beauftragt der G-BA ein Institut mit der Bewertung. Das Thema wird auf der Website des G-BA veröffentlicht, um die maßgeblichen Parteien zu informieren, dass ein Bewertungsverfahren gestartet wurde, und um diese zur Teilnahme einzuladen.

6.2 Ausschreibung und Auftragsvergabe

Gemäß § 139b Abs. 3 SGB V muss das Institut wissenschaftliche Forschungsaufträge zur Durchführung von Nutzenbewertungen an externe Sachverständige vergeben. Obwohl öffentliche Ausschreibungen für alle Einzelbewertungen aus Gründen der Planung und zeitnahen Bearbeitung zu bevorzugen sind, könnte das Institut öffentliche Ausschreibungen mit Auftragsvergabe für eine festgelegte Anzahl von Bewertungen durchführen. Vergleichbar mit der Auftragsvergabepraxis des Deutschen Instituts für Medizinische Dokumentation und Information (DIMDI) könnten Aufträge zur Bearbeitung einer bestimmten Anzahl von Nutzenbewertungen jedes Jahr verbindlich ausgeschrieben werden. Auf internationaler Ebene ähnelt dieses Verfahren der Praxis des NICE, das mit sieben universitären Institutionen zusammenarbeitet⁽⁶⁵⁾ - das NICE führt keine eigenen Bewertungen durch.

Unabhängig vom Ausschreibungsverfahren müssen die Kriterien, die bei der Auswahl der externen Sachverständigen angewandt werden, sowie das Ausschreibungsverfahren selbst veröffentlicht und ins Internet gestellt werden. Der Pool von Sachverständigen sollte so breit wie möglich angelegt sein. Daher ist die Anwendung von sprachlichen Einschränkungen nicht empfehlenswert. Um ein hohes Maß an Verfahrenstransparenz zu gewährleisten, müssen die Namen der externen Experten innerhalb von 4 Wochen nach Auftragsvergabe im Internet sowie im Vor- und Endbericht veröffentlicht werden. Alle bei der Bewertung involvierten Experten sollten alle Interessenkonflikte im Hinblick auf das zu bewertende Produkt und die einbezogenen Alternativen offen legen sowie

Beziehungen zu den damit verbundenen Arzneimittelherstellern für die letzten drei Jahre. Angemessene Verfahrensprüfungen müssen etabliert werden.

Das beauftragte Institut sollte ein Internet-Logbuch über jede Bewertung führen, wie dies bereits vom IQWiG und vom NICE getan wird. In einem derartigen Logbuch sollten das Verfahren und alle Kriterien veröffentlicht werden, um die Anforderungen in Bezug auf hohe Verfahrenstransparenz zu erfüllen. Es sollte die Namen der externen Sachverständigen enthalten sowie den Zeitrahmen für das Projekt, der zuvor von allen Projektbeteiligten verabschiedet werden sollte. Ferner sollte es eine aktualisierte Erklärung zum Stand des Projekts und aller Produkte liefern, die um der Transparenz willen veröffentlicht werden sollten (z. B. erste Projektskizze zu Beginn des Scoping-Prozesses, endgültige Projektskizze, Protokollentwurf).

6.3 Definition der Forschungsfrage

Die Definition der Forschungsfrage ist die wichtigste Phase einer Bewertung, da hierbei die Grenzen der Bewertung festgelegt werden. Die maßgeblichen Parteien sollten bei allen wichtigen Teilen der Bewertung einbezogen werden und daher auch bei der Definition der Forschungsfrage. Patienten können in dieser Phase einen sinnvollen Beitrag leisten. Sie spielen eine wesentliche Rolle, da der Patientennutzen wie vom Gesetzgeber geplant zu berücksichtigen ist.

Um die Anforderungen in Bezug auf Transparenz zu erfüllen, sollte solch eine Beteiligung in zwei Schritten durchgeführt werden: zuerst schriftliche Stellungnahmen, dann ein Scoping-Workshop. Die erste Projektskizze wird an das Institut, die maßgeblichen Parteien und die externen Sachverständigen verschickt, die allesamt um eine schriftliche Stellungnahme innerhalb des vereinbarten Zeitrahmens gebeten werden. Alle eingereichten Stellungnahmen sollten von der Projektgruppe des Instituts ausgewertet und im Hinblick auf ihre Relevanz kommentiert werden. Aus Gründen der Transparenz sollten alle Stellungnahmen zusammen mit einer Begründung, warum einige Stellungnahmen ggf. als nicht relevant eingestuft wurden, veröffentlicht werden. Innerhalb des vereinbarten Zeitrahmens werden die Stellungnahme-Einreicher sowie

Vertreter des G-BA und die externen Sachverständigen, die die Bewertung vornehmen, zum Scoping-Workshop eingeladen. Dieser Workshop hat folgende Ziele:

- Diskussion und – falls notwendig – Vorschlag für die Überarbeitung der Fragestellung,
- Vorschlag von klinisch relevanten Vergleichstherapien,
- Vorschlag von patientenrelevanten Outcomes, einschließlich einer Definition und Operationalisierung von "Nutzen" für alle maßgeblichen Beteiligten,
- Vorschlag von relevanten Untergruppen, die mehr oder weniger Nutzen aus der Intervention ziehen könnten,
- Vorschlag einer auftragsbezogenen Methodik, inklusive Einschluss- und Ausschlusskriterien für die Auswahl der einzubeziehenden Literatur, und
- Hervorhebung von relevanten Fragen für die externen Sachverständigen zur Informationsbereitstellung für Protokollentwicklung und Relevanzeinstufung.

Der Scoping-Workshop sollte von einer unabhängigen Person geleitet werden, die als Moderator auftritt. Der Workshop hat das Ziel, eine Diskussion zu Umfang und Bewertung aus verschiedenen Perspektiven anzuregen, so dass eine abschliessende Projektskizze („Protokoll-Vorentwurf“) erstellt werden kann. Dies führt danach zur Entwicklung eines Protokollentwurfs durch das Auftragsinstitut. Der Moderator sollte für den breitest möglichen Konsens im Hinblick auf die oben genannten Ziele des Scoping-Workshops sorgen. Eine wortwörtliche Niederschrift der Diskussion des Workshops sollte angefertigt werden, die als Teil des Protokolls im Internet und im Endbericht eingesehen werden kann. Der Protokollentwurf (und das Protokoll) sollte möglichst genau die Fragestellung adressieren, die die Grundlage des Auftrags für diese Nutzenbewertung bildet. Um dies zu gewährleisten, sollte der G-BA in dieser Phase konsultiert werden.

6.4 Protokollerstellung

Auf der Basis der relevanten Vorschläge, die während des Scoping-Workshops gemacht wurden, wird der Protokollentwurf entwickelt. Das Protokoll enthält die folgenden Punkte:

- Hintergrundinformation mit Forschungsfrage(n),
- Suchstrategie,
- Studiauswahlkriterien und -verfahren,
- Bewertung der Studienqualität,
- Datenextraktionsstrategie und Synthese der extrahierten Evidenz.
- Das Protokoll enthält darüber hinaus einen Zeitplan für die Bewertung.

Die Forschungsfrage wird anhand der PICO-Kriterien operationalisiert, die die Patienten, die Intervention, die Vergleichsintervention sowie die Outcomes ausführlich definieren. Die Ergebnisse des Scoping-Workshops sollten die Grundlage für die Forschungsfrage bilden.

Basierend auf den Ergebnissen des Scoping-Workshops sollte das Protokoll außerdem zeigen, welche Studiendesigns idealerweise zur Beantwortung der Frage(n) herangezogen werden sollten. Dies sollte jedoch nicht als Kriterium zur Ausgrenzung von Studien niedrigerer Evidenzlevel dienen, da die Bewertung immer aufgrund der gegenwärtig bestverfügbaren Evidenz in Bezug auf die jeweilige Fragestellung vorgenommen werden sollte. Die Information des Reviews sollte vielmehr Hinweise in Bezug auf die Stärke der Evidenz für das Appraisal geben.

Das Protokoll sollte innerhalb des vereinbarten Zeitrahmens veröffentlicht und im Internet zugänglich gemacht werden. Danach können die maßgeblichen Parteien ihre Stellungnahmen vorlegen, woraufhin eine mündliche Anhörung zur Erörterung der Kommentare stattfindet. Die Stellungnahmen, zusammen mit der Dokumentation, ob ein Vorschlag als relevant erachtet und in die Forschungsfrage einbezogen wurde oder nicht, sollten im Internet sowie im Vor- und Endbericht veröffentlicht werden. Falls keine mündliche Anhörung stattfindet, sollte dies im Vorbericht begründet werden.

Um Berichtsbias so weit wie möglich zu vermeiden, sollten alle relevanten Outcomes, die in den Primärstudien berichtet wurden, bewertet werden. Wenn Outcomes nach Beginn der Bewertung hinzugefügt oder aus dem Protokoll herausgenommen werden, sollte dies in einem Zusatz zum Protokoll („Amendment“) begründet werden.

6.5 Bewertung der Evidenz

Um die gesetzliche Bedingung der Transparenz zu erfüllen und Werturteilen entgegenzuwirken, die ein unvermeidlicher Teil eines solchen Verfahrens sind, sollten alle Schritte der Evidenzbewertung klar dokumentiert werden. Alle Schritte in diesem Verfahren sollten mit der gleichen Transparenz durchgeführt werden, wenn die Nutzenbewertung von internen Mitarbeitern des Instituts vorgenommen wird.

6.5.1 Suchstrategie

Auf der Basis der im Protokoll genau dargelegten Suchstrategie führen die externen Sachverständigen eine Suche zur Ermittlung der relevanten Studien durch. Ziel der Literaturrecherche ist es, alle Studien zu ermitteln, die zur Zeit der Recherche zur Beantwortung der Fragestellungen in der Bewertung angemessen sein könnten. Die Suche sollte breit genug angelegt sein, um Literatur zu allen Forschungsfragen zu ermitteln. Gegebenenfalls sollten separate Recherchen für jede Fragestellung bzw. Aspekte durchgeführt werden, da die zu berücksichtigenden Studien und Studientypen unterschiedlich sein können.

Eine Reihe von Datenbanken sollte durchsucht werden, denn keine einzelne Datenbank ist umfangreich genug, um alle Veröffentlichungen aus allen medizinischen Fachzeitschriften zu erfassen. Allgemeine Datenbanken wie z. B. Medline, Embase und die Datenbanken der Cochrane Library sind ein guter Ausgangspunkt. Ansonsten sollten andere Datenbanken einbezogen werden, die zusätzliche Literaturhinweise zur Fragestellung liefern können. Da die Nutzenbewertung vorrangig angesichts der Situation im deutschen Gesundheitswesen durchgeführt wird, muss gewährleistet werden, dass insbesondere für Deutschland relevante Studien ermittelt werden. Dies kann durch eine

zusätzliche Recherche in den wichtigsten Datenbanken von deutschen Verlagen sowie durch eine ergänzende manuelle Suche der relevanten Fachzeitschriften erfolgen. Die Suche sollte umfassend sein: Weder Sprachbeschränkungen noch der Ausschluss von spezifischen Studiendesigns (wie z. B. Nicht-RCTs) oder Publikationstypen (wie z. B. Abstracts oder unveröffentlichte Studien) sollten erfolgen. Zusätzlich sollten die Literaturlisten von einschlägigen Veröffentlichungen genutzt werden, um ggf. in der eigenen Recherche übersehene Studien zu identifizieren. Relevante Veröffentlichungen ergeben sich z. T. während der Suche, aber sie können auch während des Scoping-Prozesses ermittelt werden. Die Liste der relevanten Veröffentlichungen, die überprüft wurden, sollte im Vor- und Endbericht aufgeführt werden.

Die Suchstrategie sollte klar dokumentiert werden. Das QUOROM-Statement empfiehlt die ausführliche Dokumentation von Informationsquellen (Datenbanken, Register, persönliche Akten, sachkundige Informanten, Behörden, manuelle Suche) und allen Einschränkungen (berücksichtigte Jahre, Veröffentlichungsstatus, Sprache der Veröffentlichung).(7)

6.5.2 Vorlage durch die Hersteller

Das Institut sollte die maßgeblichen Hersteller so früh wie möglich über die Durchführung der Nutzenbewertung informieren und diese bitten, Informationen zu relevanten veröffentlichten und unveröffentlichten Studien innerhalb des vereinbarten Zeitrahmens einzureichen. Für die Übermittlung von Studien und Daten kann der vom IQWiG und vom Verband Forschender Arzneimittelhersteller (VFA) vereinbarte Mustervertrag zum vertraulichen Umgang mit Herstellerdaten verwendet werden, um Betriebs- und Geschäftsgeheimnisse der Arzneimittelhersteller zu schützen. Alle in der Evidenzbewertung verwendeten Daten sollten zu einem Zeitpunkt veröffentlicht werden, der vom Institut mit den Herstellern vereinbart wird.

6.5.3 Studienauswahl

Alle Veröffentlichungen, die bei der Literaturrecherche ermittelt und von den Arzneimittelherstellern eingereicht wurden, werden unabhängig von mindestens zwei

Wissenschaftlern (externe Sachverständige) auf ihre Themenrelevanz unter Verwendung der Informationen aus Titel und Abstract untersucht. Die Auswahlkriterien sollten eindeutig dokumentiert werden. Alle Studien sollten in Bezug auf die zugelassenen Indikationen des bewerteten Arzneimittels einbezogen werden. Üblicherweise angewandte Kointerventionen sollten in die Bewertung mit einbezogen werden können. Daher sollten Kointerventionen für keine dieser Studien einen Grund zum Ausschluss darstellen.

Die ausgewählten Artikel werden in ihrer Volltextversion wiederum von mindestens zwei Wissenschaftlern ausgewertet. Für jede ausgeschlossene Veröffentlichung sollte der Grund für den Ausschluss dokumentiert werden. Die übrigen Studien werden in die Bewertung einbezogen und ihre Daten werden extrahiert.

Falls systematische Reviews zu einem dem Bewertungsthema ähnlichen Thema ermittelt wurden, sollten diese genau untersucht werden. Falls für eine Fragestellung bereits systematische Reviews durchgeführt wurden, müssen diese nach den Grundsätzen der EbM mit der höchsten Priorität überprüft, ausgewertet und dementsprechend berücksichtigt werden. Durch das Vorliegen eines methodisch guten, neueren systematischen Reviews zum selben Thema wäre die Bewertung (bzw. Teile davon) eventuell eine Redundanz und damit verzichtbar.

6.5.4 Datenextraktionsverfahren

Alle ausgewählten Studien bzw. Arbeiten sollten sorgfältig gelesen und nach relevanten Daten extrahiert werden. Bei fehlenden Daten sollten die Autoren zwecks zusätzlicher Informationen kontaktiert werden, um sicherzustellen, dass keine relevanten Informationen durch schlechte Berichterstattung verloren gehen. Für Anfragen an die Autoren sollte ein standardisiertes und transparentes Verfahren eingerichtet werden. Die Antworten auf alle Fragen einschließlich fehlender Antworten sollten veröffentlicht werden.

Ein wichtiges Element bei der Datenextraktion ist die Bewertung der methodischen Qualität. Eine solche Bewertung ist nur nützlich, wenn sie Folgen für den Rest des Reviews hat. Bewertungen der methodischen Qualität können zum Beispiel verwendet

werden, um die Qualitätsunterschiede für die Erklärung der Heterogenität der Studienergebnisse zu nutzen oder als Hilfe bei der Interpretation der Bewertungsergebnisse.

Es gibt mehrere Instrumente der Qualitätsbewertung von Studien, und die Wahl eines solchen Instrumentes hängt vom Zweck der Qualitätsbewertung ab. Daher sollten der Zweck der Qualitätsbewertung und die Wahl des Instrumentes im Protokoll begründet und die Ergebnisse der Qualitätsbewertung beschrieben werden. Da die Beurteilung von methodischer Qualität auf subjektiven Einschätzungen und Entscheidungen fußt, sollte bei solchen Bewertungen eine doppelte Datenextraktion angewandt werden.

6.5.5 Datensynthese

Das Ziel der Datensynthese ist die Zusammenfassung der Ergebnisse der als Informationsgrundlage herangezogenen Primärstudien. Sie kann unter Anwendung einer deskriptiven, qualitativen Synthese vorgenommen werden oder - falls möglich – unter Anwendung von quantitativen Methoden. Zur weiteren Information bezüglich dieser Methoden verweisen wir auf die vom CRD (50) oder der Cochrane Collaboration (6) herausgegebenen Handbücher zu systematischen Reviews.

6.5.6. Analyse von Untergruppen

Nutzenbewertungen haben den Nachteil, dass sich daraus resultierende Empfehlungen nicht auf jeden einzelnen Patienten übertragen lassen. Faktoren wie z. B. Alter, Geschlecht, Komorbidität oder Komedikation können die Wirksamkeit und Sicherheit einer Therapie beeinflussen. Um die Ergebnisse einer Nutzenbewertung auf bestimmte Zielgruppen zu fokussieren, wird je nach Möglichkeit die folgende Strategie vorgeschlagen:

1. Untergruppen, die für die zu bewertende Indikation relevant ist, sollten während des Scoping-Prozesses definiert werden.
2. Die verfügbaren Daten sollten auf die statistische Durchführbarkeit der jeweiligen Untergruppenanalysen hin geprüft werden.

3. Die definierten Untergruppen sowie die Gründe für die Durchführung oder Nichtdurchführung der jeweiligen Untergruppenanalysen müssen transparent dargelegt werden.
4. Ergebnisse in Bezug auf die relevanten Untergruppen sollten veröffentlicht werden. Wenn es aufgrund von statistischen Problemen nicht möglich ist, wichtige Untergruppen zu analysieren, sollten Trends mit einem angemessenen Hinweis auf die möglicherweise eingeschränkte Relevanz veröffentlicht werden.

Die Vorteile von Untergruppenanalysen sind verbesserte Aussagen in Bezug auf Nutzenoptimierung und Schadenverringerung. In weiterer Konsequenz kann dies einen positiven Einfluss auf die Kosten haben. Ein weiterer (technischer) Vorteil besteht darin, dass die Heterogenität von Studiendesign und Ergebnissen im Verlauf einer Nutzenbewertung weniger problematisch in der Handhabung ist.

6.6 Veröffentlichung des Vorberichts

Der von den externen Sachverständigen verfasste Vorbericht muss an den G-BA und die maßgeblichen Parteien gesandt werden. Er sollte ebenfalls im Internet veröffentlicht werden. Zusätzlich wird vom Institut ein internes Review sowie ein externes Peer-Review-Verfahren durchgeführt (siehe auch Abschnitt 6.7). Alle beteiligten Parteien werden dazu aufgefordert, innerhalb des vereinbarten Zeitrahmens schriftliche Stellungnahmen einzureichen.

Die Stellungnahmen sollten darauf fokussieren, ob die Festlegungen des Berichtsplans eingehalten wurden und die Schlussfolgerungen plausibel sind in Bezug auf die zugrunde liegenden Datenanalysen.

Die eingegangenen Stellungnahmen werden von den externen Sachverständigen ausgewertet und einzeln dokumentiert im Hinblick auf ihre Relevanz sowie darauf, ob sie einbezogen wurden oder nicht. Die vorgelegten Stellungnahmen sowie die Antwort der externen Sachverständigen sollten im vereinbarten Zeitrahmen veröffentlicht und stets in den nachfolgenden Vor- oder Endbericht aufgenommen werden.

Innerhalb des vereinbarten Zeitrahmens treffen sich die Vertreter der Stellungnahmen sowie die Vertreter des G-BA zu einer mündlichen Anhörung. Der Kern dieser Anhörung ist die wissenschaftliche Diskussion der umstrittenen Fakten mit dem Ziel, die Qualität und Akzeptanz der Bewertung zu verbessern. Die Anhörung wird durch einen von der Institutsleitung ernannten Vertreter geleitet. Wenn keine mündliche Anhörung stattfindet, sollte dies im Endbericht begründet werden.

6.7 Revision des Vorberichts

Zur Unterstützung der externen Sachverständigen sollte ein internes und externes Reviewverfahren stattfinden. Das Institut führt sein eigenes internes Review des Vorberichts durch, in dem insbesondere die Einhaltung der formalen Anforderungen des Protokolls überprüft wird. Parallel dazu werden die inhaltlichen Schlussfolgerungen sowie die Qualität der Studienbewertungen von externen Peer-Reviewern ausgewertet. Aus Gründen der Transparenz sollten die Kriterien, anhand derer die externen Sachverständigen für das Peer-Review-Verfahren für den Bericht ausgewählt werden, veröffentlicht werden, sowie auch das Verfahren, aufgrund dessen die Reviewer ausgewählt und eingeladen werden. Ferner (wie zuvor in Abschnitt 3.1.7 dargelegt) ist ein offenes Reviewverfahren der beste Weg, um Transparenz zu erreichen, was bedeutet, dass sowohl die Namen der Reviewer als auch ihre Stellungnahmen im Internet veröffentlicht werden. Die Reviewberichte werden auch an den Auftraggeber (BMG oder G-BA) weitergeleitet.

6.8 Empfehlungen und Endbericht

Auf der Basis der Stellungnahmen der maßgeblichen Parteien, der Beiträge während der mündlichen Anhörung sowie der Ergebnisse des internen und externen Reviews wird der Vorbericht vom Institut unter Beratung mit den externen Sachverständigen revidiert. Auf der Grundlage der Evidenz gibt das Institut daraufhin erste Empfehlungen ab. Der revidierte Vorbericht sowie die Empfehlungen stellen den Endbericht dar. Aus Gründen der Transparenz müssen die Empfehlungen klar gekennzeichnet werden. Nach dem

vereinbarten Zeitraum wird der Abschlussbericht auf der Website des Instituts veröffentlicht.

Danach beginnt die Relevanzeinstufungsphase („Appraisal“). Dies sollte vom G-BA durchgeführt werden, um die Phase der Evidenzbewertung („Assessment“) von der Phase der Relevanzeinstufung der Evidenz („Appraisal“) zu trennen. Während der Relevanzeinstufungsphase wird die Evidenz im Hinblick auf die besonderen Umstände in Deutschland beurteilt, und Empfehlungen für künftige Vollerstattung, Höchstbeträge oder andere mögliche Bestimmungen für Arzneimittel gemäß den gesetzlichen Regelungen werden formuliert. Die zu berücksichtigenden Faktoren werden sich zwar von Appraisal zu Appraisal unterscheiden, können dabei jedoch immer auch Aspekte der Wirtschaftlichkeit, Sicherheit, soziale und ethische Kriterien sowie Aspekte zu Besonderheiten des deutschen Gesundheitswesens beinhalten. Nach dem vereinbarten Zeitraum wird der Endbericht inklusive der Ergebnisse des G-BA-Appraisals auf der Website des Instituts veröffentlicht.

6.9 Appeal / Planung der Berichtsaktualisierung

Alle maßgeblichen Parteien sollten das Recht haben, gegen die Entscheidungen des G-BA Berufung einzulegen („Appeal“).

Die Nutzenbewertung stellt die bestverfügbare Evidenz zum Thema zum Zeitpunkt der Bewertung dar. Neue Evidenz könnte Konsequenzen für die Schlussfolgerung der Bewertung haben. Ein Review und ggf. eine nachfolgende Revision der Bewertung sollte alle 3 bis 5 Jahre geplant werden.

Gemäß § 35b Abs. 2 Satz 2 und 3 SGB V ist eine regelmäßige Überprüfung der Bewertungsergebnisse vorgeschrieben. Der Bericht muss in regelmäßigen Abständen innerhalb des G-BA beratschlagt werden. Unabhängig davon haben die Hersteller das Recht, eine Aktualisierung zu beantragen, wenn neue wissenschaftliche Evidenz verfügbar wird. Der G-BA oder das BMG haben über die Anträge zu entscheiden. Die Entscheidung und ihre Begründung müssen auf der Website des G-BA veröffentlicht werden.

6.10 Fazit

Diese Abhandlung beschreibt die operationale Umsetzung der gesetzlichen Anforderungen im Hinblick auf Nutzenbewertungen von Arzneimitteln in Deutschland. Aus rechtlicher Sicht erfordern solche Bewertungen die Teilnahme der maßgeblichen Parteien, Transparenz sowie den Einsatz von EbM nach international anerkannten Standards. Um die Voraussetzung der aktiven Beteiligung der maßgeblichen Parteien zu erfüllen, sollte ein Scoping-Workshop durchgeführt werden, um diese Parteien besser an der Definition der Forschungsfrage zu beteiligen, was eine wesentliche Phase der Bewertung darstellt. Transparenz während des gesamten Prozesses sollte durch Berichterstattung zu allen Arbeitsschritten und Kriterien in allen Phasen der Bewertung erreicht werden. Insbesondere der Input aller an der Definition der Forschungsfrage beteiligten Parteien sollte durch Veröffentlichung der Stellungnahmen klargestellt werden, zusammen mit einer Auswertung, ob dies zu Veränderungen der Forschungsfrage geführt hat oder nicht. Dieses Verfahren sollte genauso transparent angewandt werden, um den Input aller beteiligten Parteien an der Entwicklung des Protokolls und des Vorberichts zu zeigen. Die Stellungnahmen der Peer-Reviewer sollten ebenso veröffentlicht werden. Ferner verlangt das Transparenzgebot eine Trennung der Phasen der Evidenzbewertung („Assessment“) und der Relevanzeinstufung der Evidenz („Appraisal“). Die Anwendung der Grundsätze evidenzbasierter Medizin bedeutet, dass alle Evidenz bewertet werden sollte, um die bestverfügbare Evidenz zur Beantwortung jeder Forschungsfrage zu ermitteln. Ein solches Verfahren gewährleistet, dass die Nutzenbewertungen von Arzneimitteln in Deutschland nach den international üblichen hohen Standards durchgeführt werden.

Literaturverzeichnis

- (1) Gemeinsamer Bundesausschuss: Verfahrensordnung Fassung vom 20.09.2005, letzte Änderung am 18.04.2006, in Kraft getreten am: 07.07.2006. <http://www.g-ba.de/institution/aufgaben-und-struktur/Verfahrensordnung/> . 2006.
- (2) Cochrane Effective Practice and Organization of Care Group. Effective interventions unit. Evaluation guide 5: Designing an economic evaluation. <http://www.epoc.cochrane.org/Files/Website/Reviewer%20Resources/draftps.pdf>. 2008.
- (3) CRD. What is an economic evaluation? <http://www.york.ac.uk/inst/crd/nfaq1.htm>. 2008. CRD, York.
- (4) von der Schulenburg J, Vauth C, Mittendorf T, Greiner W: Methods for determining cost-benefit ratios for pharmaceuticals in Germany [Methoden zur Ermittlung von Kosten-Nutzen-Relationen für Arzneimittel in Deutschland]. Eur J Health Econ 2007; 8 Suppl 1:S5-31.
- (5) Göhlen B, Rüter A: HTA beim DIMDI. Z ärztl Fortbild Qual Gesundh wes 2007; 101:508-511.
- (6) Cochrane Handbook for Systematic Reviews of Interventions 4.2.6 [aktualisiert im September 2006]. Higgins JPT, Green S, editors. <http://www.cochrane.org/resources/handbook/hbook.htm> Zugriff am: 8. November 2007. 2007.
- (7) Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF: Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet 1999; 354(9193):1896-1900.
- (8) Velasco M, Perleth M, Drummond M, Gurtner F, Jorgensen T, Jovell A, Malone J, Ruther A, Wild C: Best practice in undertaking and reporting health technology assessments. Working group 4 report. Int J Technol Assess Health Care 2002; 18(2):361-422.
- (9) NICE: Guide to the technology appraisal process. London, National Institute for Clinical Excellence, 2004.
- (10) Van Rooyen S, Godlee F, Evans S, Black N, Smith R: Effect of open peer review on quality of reviews and on reviewers' recommendations: a randomised trial. BMJ 1999; 318(7175):23-27.

- (11) Godlee F, Gale CR, Martyn CN: Effect on the quality of peer review of blinding reviewers and asking them to sign their reports: a randomized controlled trial. *JAMA* 1998; 280(3):237-240.
- (12) NICE: Guide to the methods of technology appraisal. London, National Institute for Clinical Excellence, 2004.
- (13) DUET. Database of Uncertainties about the Effects of Treatments (DUETs): A resource to help prioritise new research. <http://www.duets.nhs.uk/Default.asp> Zugriff am: 12. Februar 2008. 2008.
- (14) Tallon D, Chard J, Dieppe P: Consumer involvement in research is essential. *BMJ* 2000; 320(7231):380-381.
- (15) Goodare H, Lockwood S: Involving patients in clinical research. Improves the quality of research. *BMJ* 1999; 319(7212):724-725.
- (16) INVOLVE Promoting public involvement in NHS, public health and social care research. <http://www.invo.org.uk/> Zugriff am: 10. Januar 2008. 2008.
- (17) Hanley B, Truesdale A, King A, Elbourne D, Chalmers I: Involving consumers in designing, conducting, and interpreting randomised controlled trials: questionnaire survey. *BMJ* 2001; 322(7285):519-523.
- (18) Kahan JP, Park RE, Leape LL, Bernstein SJ, Hilborne LH, Parker L, Kamberg CJ, Ballard DJ, Brook RH: Variations by specialty in physician ratings of the appropriateness and necessity of indications for procedures. *Medical Care* 1996; 34(6):512-523.
- (19) Coulter I, Adams A, Shekelle P: Impact of Varying Panel Membership on Ratings of Appropriateness in Consensus Panels - A Comparison of A Multidisciplinary and Single-Disciplinary Panel. *Health Services Research* 1995; 30(4):577-591.
- (20) Grant-Pearce C, Miles I, Hills P. Mismatches in priorities for health research between professionals and consumers. A report to the standing advisory group on consumer involvement in the NHS R&D Programme. 1998. Manchester UK, PREST: University of Manchester.
- (21) Devereaux PJ, Anderson DR, Gardner MJ, Putnam W, Flowerdew GJ, Brownell BF, Nagpal S, Cox JL: Differences between perspectives of physicians and patients on anticoagulation in patients with atrial fibrillation: observational study. *BMJ* 2001; 323(7323):1218-1222.
- (22) Montgomery AA, Fahey T: How do patients' treatment preferences compare with those of clinicians? *Qual Health Care* 2001; 10 Suppl 1:i39-i43.

- (23) Chard J, Dickson J, Tallon D, Dieppe P: A comparison of the views of rheumatologists, general practitioners and patients on the treatment of osteoarthritis. *Rheumatology (Oxford)* 2002; 41(10):1208-1210.
- (24) Kirwan JR, Minnock P, Adebajo A, Bresnihan B, Choy E, de Wit M, Hazes M, Richards P, Saag K, Suarez-Almazor M, Wells G, Hewlett S: Patient perspective: fatigue as a recommended patient centered outcome measure in rheumatoid arthritis. *J Rheumatol* 2007; 34(5):1174-1177.
- (25) Garland AF, Lewczyk-Boxmeyer CM, Gabayan EN, Hawley KM: Multiple stakeholder agreement on desired outcomes for adolescents' mental health services. *Psychiatric Services* 2004; 55(6):671-676.
- (26) Lee TT, Ziegler JK, Sommi R, Sugar C, Mahmoud R, Lenert LA: Comparison of preferences for health outcomes in schizophrenia among stakeholder groups. *Journal of Psychiatric Research* 2000; 34(3):201-210.
- (27) Kwok CK, Ibrahim SA: Rheumatology patient and physician concordance with respect to important health and symptom status outcomes. *Arthritis & Rheumatism-Arthritis Care & Research* 2001; 45(4):372-377.
- (28) Hubbard G, Kidd L, Donaghy E, McDonald C, Kearney N: A review of literature about involving people affected by cancer in research, policy and planning and practice. *Patient Educ Couns* 2007; 65(1):21-33.
- (29) Oliver S, Milne R, Bradburn J, Buchanan P, Kerridge L, Wally T, Gabbay J: Involving consumers in a needs-led research programme: a pilot project. *Health Expectations* 2001; 4:18-28.
- (30) Bradburn J, Maher J, Adewuyi-Dalton R, Grunfeld E, Lancaster T, Mant D: Developing clinical trial protocols: the use of patient focus groups. *Psycho-Oncology* 1995; 4:107-112.
- (31) Ali K, Roffe C, Crome P: What patients want: consumer involvement in the design of a randomized controlled trial of routine oxygen supplementation after acute stroke. *Stroke* 2006; 37(3):865-871.
- (32) Culyer AJ: Involving stakeholders in healthcare decisions--the experience of the National Institute for Health and Clinical Excellence (NICE) in England and Wales. *Healthc Q* 2005; 8(3):56-60.
- (33) The AGREE Collaboration: Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Qual Saf Health Care* 2003; 12(1):18-23.
- (34) Marsden J, Bradburn J: Patient and clinician collaboration in the design of a national randomized breast cancer trial. *Health Expectations* 2004; 7(1):6-17.

- (35) Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB: Evidence-based medicine. How to practice and teach EBM. ed 2nd, Edinburgh, Churchill Livingstone, 2000.
- (36) Danish Institute for Health Technology Assessment: Health Technology Assessment Handbook. ed 1st, Copenhagen, Danish Institute for Health Technology Assessment, 2001.
- (37) Fletcher RH, Fletcher SW, Wagner EH: Clinical epidemiology - the essentials. Baltimore, Williams & Wilkins, 1982.
- (38) Villar J, Carroli G, Belizan JM: Predictive ability of meta-analyses of randomised controlled trials. *Lancet* 1995; 345(8952):772-776.
- (39) LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F: Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med* 1997; 337(8):536-542.
- (40) Cappelleri JC, Ioannidis JP, Schmid CH, de Ferranti SD, Aubert M, Chalmers TC, Lau J: Large trials vs meta-analysis of smaller trials: how do their results compare? *JAMA* 1996; 276(16):1332-1338.
- (41) Ioannidis JP, Cappelleri JC, Lau J: Meta-analyses and large randomized, controlled trials. *N Engl J Med* 1998; 338(1):59-2.
- (42) Ioannidis JP, Cappelleri JC, Lau J: Issues in comparisons between meta-analyses and large trials. *JAMA* 1998; 279(14):1089-1093.
- (43) Furukawa TA, Streiner DL, Hori S: Discrepancies among megatrials. *J Clin Epidemiol* 2000; 53(12):1193-1199.
- (44) Jadad AR, Cook DJ, Browman GP: A guide to interpreting discordant systematic reviews. *CMAJ* 1997; 156(10):1411-1416.
- (45) NICE. Guide to the methods of technology appraisal. Draft for consultation (Nov 2007). 2007. NHS, National Institute for Health and Clinical Excellence.
- (46) Claxton K, Sculpher M, Drummond M: A rational framework for decision making by the National Institute For Clinical Excellence (NICE). *Lancet* 2002; 360(9334):711-715.
- (47) Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ: Users' guides to the medical literature. IX. A method for grading health care recommendations. Evidence-Based Medicine Working Group. *JAMA* 1995; 274(22):1800-1804.
- (48) Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, Atkins D, for the Methods Work Group Third U.S.Preventive Services Task Force. Current

- Methods of the U.S. Preventive Services Task Force: A Review of the Process. <http://www.ahrq.gov/clinic/ajpmsuppl/harris1.htm> Zugriff am: 12. Februar 2008. 2008.
- (49) Philips B, Ball C, Sackett D, Badenoch D, Straus S, Haynes B, Dawes M. Oxford Centre for Evidence-based Medicine Levels of Evidence (may 2001). <http://www.cebm.net/index.aspx?o=1025> Zugriff am: 8. November 2007. 2001.
- (50) CRD: Undertaking systematic reviews of research of effectiveness. CRD's guidance for those carrying out or commissioning reviews. ed 2nd, 2001.
- (51) Egger M, Dickersin K, Davey Smith G: Problems and limitations in conducting systematic reviews; in: Egger M, Davey Smith G, Altman DG, (eds): Systematic reviews in health care. Meta-analysis in context. London, BMJ Publishing Group, 2001, S. 43-68.
- (52) Jüni P, Altman DG, Egger M: Assessing the quality of randomised controlled trials; in: Egger M, Davey Smith G, Altman DG, (eds): Systematic reviews in health care. Meta-analysis in context. London, BMJ Publishing Group, 2001, S. 87-108.
- (53) Bucher HC, Guyatt GH, Griffith LE, Walter SD: The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol 1997; 50(6):683-691.
- (54) Song F, Altman DG, Glenny AM, Deeks JJ: Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. BMJ 2003; 326(7387):472.
- (55) CCOHTA. Guidelines for authors of CCOHTA Health Technology Assessment reports. Canadian Coordinating Office for Health Technology Assessment, editor. 2003.
- (56) Haute Autorité de santé. General method for assessing health technologies. source: http://www.has-sante.fr/portail/display.jsp?id=c_541195. 2007. Haute Autorité de santé.
- (57) Staal PC, Ligtenberg G. Beoordeling stand van de wetenschap en praktijk. 254. 2007. Diemen, College voor Zorgverzekeringen.
- (58) Martelli F, Torre GL, Ghionno ED, Staniscia T, Neroni M, Cicchetti A, Bremen KV, Ricciardi W: Health technology assessment agencies: An international overview of organizational aspects. Int J Technol Assess Health Care 2007; 23(4):414-424.
- (59) Bastian H, Bender R, Ernst AS, Kaiser T, Kirchner H, Kolominsky-Rabas P, Lange S, Sawicki PT, Weber M. Institute for Quality and Efficiency in Health Care. Methods (preamble). 2007. Cologne, Germany, IQWiG.

- (60) NICE: Guide to the technology appraisal process. London, National Institute for Clinical Excellence, 2001.
- (61) House of Commons HC. National Institute for Health and Clinical Excellence. First report of session 2007-08. Volume 1. Report, together with formal minutes. HC 27-I. 10-1-2008. London, The Stationery Office Limited.
- (62) Bastian H, Bender R, Ernst AS, Kaiser T, Kirchner H, Kolominsky-Rabas P, Lange S, Sawicki PT, Weber M. Institute for Quality and Efficiency in Health Care. Methods (version 2.0). 2007. Cologne, Germany, IQWiG.
- (63) Antes G, Jöckel KH, Kohlmann T, Raspe H, Wasem J. Kommentierende Synopse der Fachpositionen zur Kosten-Nutzenbewertung für Arzneimittel.
http://www.bmg.bund.de/cln_041/nn_1145310/DE/Themenschwerpunkte/Gesundheit/Arzneimittel/Fachtagung/kommentierte-synopse.templateId=raw,property=publicationFile.pdf/kommentierte-synopse.pdf. 2007. Bundesministerium für Gesundheit, Deutschland.
- (64) IQWiG. Allgemeine Methoden. *Entwurf für* Version 3.0 vom 15.11.2007. IQWiG, editor. 2007. Köln.
- (65) Collaborative centres NICE.
<http://www.nchta.org/publicationspdfs/infoleaflets/nice.pdf> Zugriff am: 16. Januar 2008.