

Methoden

Version 2.0 vom 19.12.2006

Erstellt durch das Steuergremium^a des Instituts

Kontakt:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG)

Dillenburger Straße 27

D-51105 Köln

Fon: +49 (0)221 / 35685-0

Fax: +49 (0)221 / 35685-1

Email: methoden@iqwig.de

^a Bastian H, Bender R, Ernst AS, Kaiser T, Kirchner H, Kolominsky-Rabas P, Lange S, Sawicki PT, Weber M

Allgemeine Anmerkungen:

Der erste Entwurf der Methoden des Instituts entstand im Herbst 2004 und wurde mit der Veröffentlichung am 1. November 2004 zur Diskussion gestellt. Nach dem Eingang von Stellungnahmen und Gutachten schloss sich im Februar 2005 eine Gesprächsrunde mit den Stellungnehmenden an, einschließlich der Beteiligung des wissenschaftlichen Beirats des Institutes. Daraus folgend entstand die erste Version 1.0) vom 1. März 2005. Im Jahr 2006 wurde das Dokument überarbeitet und zwei aufeinander folgende Entwürfe zur Diskussion gestellt; einmal ein interner Entwurf vom 27.04.2006, dann ein im Internet veröffentlichter zweiter Entwurf vom 28.09.2006. Unter Berücksichtigung der eingegangenen Stellungnahmen zu beiden Entwürfen entstand die vorliegende zweite Version (Version 2.0) vom 19. Dezember 2006.

Das Dokument wird zukünftig jährlich auf eine notwendige Überarbeitung hin überprüft, es sei denn, Fehler im Dokument oder wesentliche Entwicklungen legen eine vorzeitige Aktualisierung nahe.

Für jedes durch das Institut erstellte Dokument gilt die jeweils zum Zeitpunkt der Veröffentlichung gültige Methodenversion.

In diesem Dokument werden bei im Singular verwendeten Substantiven immer die weibliche und die männliche Form benutzt. Aus Gründen der besseren Lesbarkeit wird im Plural auf eine Geschlechtsdifferenzierung verzichtet. So umfasst zum Beispiel der Begriff „Patienten“ sowohl Patientinnen als auch Patienten.

Präambel

Der Gesetzgeber hat mit der Gesundheitsreform 2003 (Gesundheits-Modernisierungsgesetz, GMG) die Etablierung eines neuen staatsunabhängigen Instituts im deutschen Gesundheitswesen vorgesehen. Der Gemeinsame Bundesausschuss (G-BA) hat hierzu im Juni 2004 eine rechtsfähige Stiftung des privaten Rechts errichtet. Der ausschließliche Zweck der Stiftung ist die Errichtung und Unterhaltung des „Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen“ (IQWiG). Organe der Stiftung sind der Stiftungsrat sowie ein fünfköpfiger Vorstand. Das Institut ist eine Einrichtung der Stiftung unter wissenschaftlich unabhängiger Leitung. Beratende Gremien des Institutes sind ein Kuratorium sowie ein wissenschaftlicher Beirat. Das Kuratorium besteht aus 30 Mitgliedern. Der wissenschaftliche Beirat wird durch den Vorstand bestellt und besteht aus sechs bis zwölf Wissenschaftlern. Der Sitz des Institutes ist bis 2008 Köln.

Das Institut hat die Aufgabe, den Nutzen, die Qualität und die Wirtschaftlichkeit von medizinischen Leistungen wissenschaftlich zu bewerten. Dazu gehören unter anderem die Bewertung von Behandlungsleitlinien, die Abgabe von Empfehlungen zu strukturierten Behandlungsprogrammen (DMP), die Nutzenbewertung von Arzneimitteln sowie die Herausgabe von Gesundheitsinformationen.

Das Institut widmet sich Fragen, die für die Qualität und die Wirtschaftlichkeit der von der gesetzlichen Krankenversicherung (GKV) erbrachten Leistungen von grundsätzlicher Bedeutung sind. Es hat dabei alters-, geschlechts- und lebenslagenspezifische Besonderheiten zu berücksichtigen. Das Institut wird als fachlich unabhängige wissenschaftliche Einrichtung insbesondere auf folgenden Gebieten tätig:

- Recherche, Darstellung und Bewertung des aktuellen medizinischen Wissensstandes zu diagnostischen und therapeutischen Verfahren bei ausgewählten Krankheiten,
- Erstellung von wissenschaftlichen Ausarbeitungen, Gutachten und Stellungnahmen zu Fragen der Qualität und Wirtschaftlichkeit der im Rahmen der gesetzlichen Krankenversicherung erbrachten Leistungen,
- Bewertung evidenzbasierter Leitlinien für die epidemiologisch wichtigsten Krankheiten,
- Abgabe von Empfehlungen zu Disease-Management-Programmen,
- Bewertung des Nutzens von Arzneimitteln,
- Bereitstellung von für alle Bürger verständlichen allgemeinen Informationen zur Qualität und Effizienz in der Gesundheitsversorgung.

Aufgabe des Instituts ist es, den G-BA auf diesen Gebieten durch Abgabe von Empfehlungen bei der Wahrnehmung seiner gesetzlichen Aufgaben zu unterstützen und durch seine Arbeit zu einer fortwährenden Verbesserung der Qualität der medizinischen Versorgung der Bevölkerung beizutragen. Das Institut soll eine eigenständige wissenschaftliche Kompetenz entwickeln, um die an das Institut herangetragen Forschungsfragen zu beantworten, um versorgungsrelevante medizinische Inhalte und Konzepte zu bewerten und um für Patienten relevanten Forschungsbedarf zu ermitteln. Die vom Institut erarbeiteten Informationen werden an das Bundesministerium für Gesundheit (BMG) und an den G-BA weitergegeben und der breiten Öffentlichkeit vermittelt.

Das Institut erfüllt seine Aufgaben durch die Erstellung von themenbezogenen Berichten auf Anforderung des G-BA und des BMG. Darüber hinaus hat es dort wissenschaftliche Untersuchungen zu initiieren, zu koordinieren und deren Ergebnisse zu publizieren, wo das versorgungsrelevante medizinische Wissen zu vervollständigen ist. Dazu führt das Institut ein regelmäßiges Screening der Literatur nach versorgungsrelevanten medizinischen Innovationen durch, bewertet diese und gibt die Ergebnisse in allgemeinverständlicher Form weiter. Aufgrund der internationalen Literatur und eigener Recherchen kann das Institut Vorschläge für innovative versorgungsrelevante Forschung erarbeiten, diese anregen, in der Durchführung begleiten und deren Ergebnisse publizieren.

Das Steuergremium des Institutes, das sich aus der Institutsleitung und den Ressortleitungen zusammensetzt, erstellt und modifiziert die Methoden und Abläufe der Institutsarbeit. Diese Methoden werden publiziert, um die Institutsarbeit transparent zu machen. Eine Diskussion der Methoden mit dem Ziel einer fortlaufenden Verbesserung ist ausdrücklich gewollt.

Nicht in jedem Fall werden sich alle Schritte innerhalb eines Bewertungsprozesses vorab detailliert und prozedural darstellen lassen. Einzelne Vorgehensweisen sind unter anderem von der jeweiligen Fragestellung, von der vorliegenden wissenschaftlichen Evidenz und von eingehenden Stellungnahmen abhängig. Demzufolge ist das vorliegende Dokument eher als Leitfaden bei der Bewertung einer medizinischen Maßnahme zu verstehen. Im Berichtsplan und im Vorbericht wird der auf den jeweiligen Einzelfall bezogene Bewertungsprozess ausgestaltet und dargelegt.

Um die bereits vorhandenen Ressourcen sinnvoll und effizient zu nutzen, wird das Institut fachliche Vorarbeiten anderer nationaler und internationaler Institutionen im Gesundheitswesen berücksichtigen und gegebenenfalls auf diesen aufbauen.

Inhaltsverzeichnis

Präambel	i
Inhaltsverzeichnis	1
1. Wissenschaftliche Methoden und Biometrie	3
1.1 Darstellung von Effekten und Risiken	3
1.2 Beurteilung statistischer Signifikanz	4
1.3 Beurteilung klinischer Relevanz	5
1.4 Subgruppenanalysen	7
1.5 Aspekte der Beurteilung der Studienqualität	9
1.6 Ermittlung des Schadenpotenzials medizinischer Interventionen	11
1.7 Beurteilung von Studien mit „überholten“ Methoden	14
1.8 Beurteilung verschiedener Studienarten	15
1.9 Rangordnung verschiedener Studienarten/Evidenzgrade	16
1.10 Bezug zwischen Studientyp/-art und Fragestellung	17
1.11 Beurteilung nicht oder nicht vollständig publizierter Daten	17
1.12 Beurteilung der Konsistenz publizierter Daten	18
1.13 Umgang mit Rohdaten	20
1.14 Beurteilung von Ergebnisunsicherheiten	20
1.15 Beurteilung nicht verblindbarer Verfahren	21
1.16 Berücksichtigung datenschutzrechtlicher Aspekte/Vertraulichkeit	22
1.17 Berücksichtigung ethischer Aspekte	22
1.18 Darstellung von Biasarten	24
1.19 Nachweis der Verschiedenheit	26
1.20 Nachweis der Gleichheit	27
1.21 Meta-Analysen	29
1.22 Adjustierung und multifaktorielle Verfahren	31
1.23 Bewertung qualitativer Studien	33
1.24 Einsatz von Konsultationstechniken	34
1.25 Begutachtung (externes Review)	36
2. Spezielle Beurteilung medizinischer und versorgungsrelevanter Inhalte	38
2.1 Nutzenbewertung in der Medizin	38
2.2 Medikamentöse und nichtmedikamentöse Interventionen	45
2.3 Diagnostische Verfahren	48

2.4	Screening	53
2.5	Gesundheitsökonomie	56
2.6	Leitlinien und Disease-Management-Programme	65
2.7	Systematische Übersichten und HTA-Berichte	72
2.8	Prognose	74
2.9	Individuelle Risikoabschätzung	76
2.10	Beurteilung bevölkerungsweiter Interventionen/Prävention	77
2.11	Beschreibung der Art und Größe des Placeboeffektes	79
3.	Evidenzbasierte Gesundheitsinformationen für Bürger und Patienten	82
3.1	Ziel	82
3.2	Informationssystem	82
3.3	Entwicklung der Informationsprodukte	86
3.4	Publikation	92
3.5	Evaluation und Aktualisierung	92
4.	Berichtserstellung	96
4.1	Produkte	96
4.2	Auswahl externer Sachverständiger	97
4.3	Gewährleistung der fachlichen Unabhängigkeit	99
4.4	Erstellung von Berichten	101
4.5	Erstellung von „Rapid Reports“ und Arbeitspapieren	104
4.6	Veröffentlichung von Berichten	106
4.7	Informationsbeschaffung	106
4.8	Fragestellungsbezogene Beweisführung	111
4.9	Prioritätensetzung	112
4.10	Zeitpunkte der Berichtserstellung	113

Eine Hauptursache der Armut in den Wissenschaften ist meist eingebildeter Reichtum. Es ist nicht ihr Ziel, der unendlichen Weisheit eine Tür zu öffnen, sondern eine Grenze zu setzen dem unendlichen Irrtum.

Bertolt Brecht. Leben des Galilei. Frankfurt: Suhrkamp. Uraufführung, erste Version 1943, Schauspielhaus Zürich.

1. Wissenschaftliche Methoden und Biometrie

1.1 Darstellung von Effekten und Risiken

Die Darstellung von Interventions- oder Expositionseffekten ist zunächst eindeutig an eine klare Zielvariable zu knüpfen. Die Betrachtung einer alternativen Zielvariablen verändert auch die Darstellung und die Stärke eines möglichen Effekts. Die Wahl eines geeigneten Effektmaßes hängt grundsätzlich vom Messniveau der betrachteten Zielvariablen ab. Bei stetigen Variablen können Effekte in der Regel mit Hilfe von Mittelwerten sowie Differenzen von Mittelwerten – unter Umständen nach geeigneter Gewichtung – dargestellt werden. Bei kategoriellen Zielvariablen kommen die üblichen Effekt- und Risikomaße aus Vierfeldertafeln in Frage [1]. Nach Festlegung eines primären Effektmaßes zur Datenanalyse wird das Institut wenn möglich zur deskriptiven Darstellung einen Effekt sowohl durch ein absolutes (zum Beispiel absolute Risikoreduktion oder Number Needed to Treat) als auch durch ein relatives Maß (zum Beispiel relatives Risiko oder Odds Ratio) beschreiben. Eine übersichtliche Zusammenfassung der Vor- und Nachteile der üblichen Effektmaße gibt Kapitel 8 des Cochrane Reviewer Handbooks [2]. Besondere Aspekte, die beim Vorliegen ordinaler Daten berücksichtigt werden müssen, beschreibt Agresti [3,4].

Zwingend erforderlich ist es, bei jeder Effektschätzung den Grad der statistischen Unsicherheit zu beschreiben. Häufige Methoden hierfür sind die Berechnung des Standardfehlers sowie eines Konfidenzintervalls. Wann immer möglich wird das Institut adäquate Konfidenzintervalle für Effektschätzungen mit der Information, ob es sich um ein- oder zweiseitige Konfidenzgrenzen handelt sowie dem gewählten Konfidenzniveau, angeben. In der medizinischen Forschung wird üblicherweise das zweiseitige Konfidenzniveau 95% verwendet, in manchen Situationen aber auch 90% oder 99%. Eine Übersicht über die häufigsten Methoden zur Berechnung von Konfidenzintervallen geben Altman et al. [5].

Je nach Datensituation (zum Beispiel sehr kleine Stichproben) und Fragestellung ist zur Einhaltung des Konfidenzniveaus die Anwendung exakter Methoden bei der Intervallschätzung von Effekten und Risiken in Erwägung zu ziehen. Eine aktuelle Diskussion exakter Methoden gibt Agresti [6].

Literatur

- [1] Bender R. Interpretation von Effizienzmaßen der Vierfeldertafel für Diagnostik und Behandlung. Med Klin 2001; 96: 116-121. Erratum: Med Klin 2001 96: 181
- [2] Deeks JJ, Higgins JPT, Altman DG. Analysing and presenting results. In: Higgins JPT, Green, S, editors. Cochrane Handbook for Systematic Reviews of Interventions 4.2.5 [updated May 2005]; Section 8. In: The Cochrane Library, Issue 3, 2005. Chichester: Wiley; 2005.
- [3] Agresti A. Categorical Data Analysis, 2nd Ed. New York: Wiley; 2002.
- [4] Agresti A. Modelling ordered categorical data: Recent advances and future challenges. Stat Med 1999; 18: 2191-2207.
- [5] Altman DG, Machin D, Bryant TM, Gardner MJ, editors. Statistics with Confidence: Confidence Intervals and Statistical Guidelines, 2nd Ed. London: BMJ Books; 2000.
- [6] Agresti A. Dealing with discreteness: Making 'exact' confidence intervals for proportions, differences of proportions, and odds ratios more exact. Stat Meth Med Res 2003; 12: 3-21.

1.2 Beurteilung statistischer Signifikanz

Mit Hilfe statistischer Signifikanztests ist es möglich, a priori formulierte Hypothesen unter Kontrolle der Irrtumswahrscheinlichkeit 1. Art zu testen. Die Konvention, von einem statistisch signifikanten Resultat zu sprechen, falls der p -Wert das Signifikanzniveau 0,05 ($p < 0.05$) unterschreitet, ist in vielen Fällen durchaus sinnvoll. Je nach Fragestellung und Aussage kann es erforderlich sein, ein möglicherweise kleineres Signifikanzniveau zu fordern. Umgekehrt gibt es Situationen, in denen auch ein höheres Signifikanzniveau akzeptabel sein kann. Solche Ausnahmen wird das Institut immer explizit begründen.

Bei der Interpretation von p -Werten ist eine Reihe von Aspekten zu betrachten. Es muss im Detail klar sein, auf welche Fragestellung und welche Datensituation sich die Signifikanzangabe bezieht und wie die statistische Hypothesenformulierung lautet. Insbesondere muss klar sein, ob es sich um eine ein- oder zweiseitige Hypothese handelt [1] und ob die betrachtete Hypothese als Teil eines Mehrhypothesenproblems zu sehen ist [2]. Beide Aspekte – ob eine Hypothese ein- oder zweiseitig zu formulieren ist und ob für multiples Testen adjustiert werden muss – werden in der wissenschaftlichen Literatur immer wieder kontrovers diskutiert.

Bezüglich der Hypothesenformulierung ist traditionell von einem zweiseitigen Testproblem auszugehen. Ausnahmen hiervon bilden zum Beispiel Nichtunterlegenheitsstudien (siehe auch Abschnitt 1.20). Die Formulierung eines einseitigen Hypothesenproblems ist grundsätzlich immer möglich, bedarf jedoch einer genauen Begründung. Im Fall einer einseitigen Hypothesenformulierung sind die Anwendung einseitiger Signifikanztests sowie die Berechnung einseitiger Konfidenzgrenzen angebracht. Zur besseren Vergleichbarkeit mit zweiseitigen statistischen Verfahren wird in einigen

Guidelines für klinische Studien eine Halbierung des üblichen Signifikanzniveaus von 5% auf 2,5% gefordert [3]. Das Institut folgt dem zentralen Grundprinzip, dass die Hypothesenformulierung (ein- oder zweiseitig) und das Signifikanzniveau a priori klar festzulegen sind und wird Abweichungen von den üblichen Festlegungen (einseitige anstelle zweiseitiger Hypothesenformulierung, Signifikanzniveau ungleich 5%) begründen bzw. auf entsprechende Begründungen in der Primärliteratur achten.

Ist eine betrachtete Hypothese eindeutig Teil eines Mehrhypothesenproblems, so ist eine adäquate Adjustierung für multiples Testen erforderlich. Eine Übersicht, in welchen Situationen dies der Fall ist und welche Methoden hierfür zur Verfügung stehen, geben Bender & Lange [4]. Wenn sinnvoll und möglich wird das Institut Methoden zur Adjustierung für multiples Testen anwenden.

Das Institut wertet ein statistisch nicht signifikantes Ergebnis nicht als Evidenz für das Nichtvorhandensein eines Effekts (Abwesenheit bzw. Äquivalenz) [5]. Zum Nachweis einer Gleichheit wendet das Institut adäquate Methoden für Äquivalenzhypothesen an (siehe auch Abschnitt 1.20).

Als prinzipielle Alternative zu statistischen Signifikanztests können Bayes'sche Methoden angesehen werden [6,7]. Das Institut wird in Abhängigkeit der Fragestellung nach Bedarf auch Bayes'sche Methoden verwenden (siehe auch Abschnitt 1.14).

Literatur

- [1] Bland JM, Altman DG. One and two sided tests of significance. *BMJ* 1994; 309: 248.
- [2] Tukey JW. Some thoughts on clinical trials, especially problems of multiplicity. *Sc* 1977; 198: 679-684.
- [3] ICH E9 Expert Working Group. ICH Harmonised Tripartite Guideline: Statistical Principles for Clinical Trials. *Stat Med* 1999; 18: 1905-1942.
- [4] Bender R, Lange S. Adjusting for multiple testing – when and how? *J Clin Epidemiol* 2001; 54: 343-349.
- [5] Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ* 1995; 311: 485.
- [6] Spiegelhalter DJ, Freedman LS. Bayesian approaches to randomised trials. *J. R. Stat. Soc. A* 1994; 157: 357-416.
- [7] Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. An introduction to Bayesian methods in health technology assessment. *BMJ* 1999; 319: 508-512.

1.3 Beurteilung klinischer Relevanz

Die klinische Relevanz eines Effekts oder Risikos ist grundsätzlich nicht am p -Wert ablesbar. Die statistische Signifikanz ist eine Wahrscheinlichkeitsaussage, in die neben der Stärke eines möglichen Effekts auch die Variabilität der Daten und der Stichprobenumfang eingehen. Bei der Inter-

pretation der Bedeutsamkeit von p -Werten muss insbesondere der Stichprobenumfang der zu Grunde liegenden Studie berücksichtigt werden [1]. In einer kleinen Studie ist ein sehr kleiner p -Wert nur bei einem deutlichen Effekt zu erwarten, während in einer großen Studie auch bei verschwindend kleinem Effekt eine deutliche Signifikanz nicht ungewöhnlich ist [2,3]. Folglich lässt sich die klinische Relevanz eines Studienergebnisses auf keinen Fall am p -Wert ablesen.

Für die Beurteilung der klinischen Relevanz von Studienergebnissen gibt es noch kein breit akzeptiertes methodisches Vorgehen. Nur in wenigen Leitlinien finden sich Hinweise auf die Definition von relevanten beziehungsweise irrelevanten Unterschieden zwischen Gruppen. Ein erster Ansatz zur Beurteilung der klinischen Relevanz von Studienergebnissen ist die Bewertung des Effektschätzers und des dazugehörigen Konfidenzintervalls (siehe auch Abschnitt 1.1) mit Hilfe medizinischer Sachkenntnis. Ein formales Relevanzkriterium kann die Beurteilung der (im Falle von günstigen Effekten) unteren Konfidenzgrenze für den Effektschätzer beziehungsweise die Anwendung eines statistischen Tests mit Verschiebung der Nullhypothese zum statistischen Nachweis relevanter klinischer Effekte sein [4]. Eine weitere Möglichkeit liegt darin, ein Relevanzkriterium individuell zu formulieren, zum Beispiel im Sinne einer Responderdefinition [5]. Darüber hinaus spielt die individuelle Einschätzung der Betroffenen eine wesentliche Rolle. Anhaltspunkte hierzu wird im Einzelfall die Darlegung patientenrelevanter Endpunkte liefern (siehe auch Abschnitt 2.1). Das Institut wird die Bewertung der klinischen Relevanz problemorientiert unter Berücksichtigung dieser Aspekte vornehmen.

Literatur

- [1] Royall RM. The effect of sample size on the meaning of significance tests. *Am Stat* 1986; 40: 313-315.
- [2] Feinstein AR. Invidious comparisons and unmet clinical challenges. *Am J Med* 1992; 92: 117-120.
- [3] Hung HMJ, O'Neill RT, Bauer P, Köhne K. The behavior of the P -value when the alternative hypothesis is true. *Biometrics* 1997; 53: 11-22.
- [4] Windeler J, Conradt C. Wie können „Signifikanz“ und „Relevanz“ verbunden werden? *Med Klin* 1999; 94: 652-655.
- [5] Kieser M, Röhmel J, Friede. Power and sample size determination when assessing the clinical relevance of trial results by 'responder analyses'. *Stat Med* 2004; 23: 3287-3305.

1.4 Subgruppenanalysen

Subgruppenanalysen werden in der methodischen Literatur sehr kritisch diskutiert [1,2]. Die Interpretation ihrer Ergebnisse wird im Wesentlichen durch drei Aspekte erschwert:

- **Kein Beweischarakter:** Subgruppenanalysen werden selten a priori geplant und sind selten Bestandteil des Studienprotokolls (gegebenenfalls auch von Amendments). Wenn Subgruppenanalysen „post hoc“ stattfinden, können ihre Ergebnisse nicht als methodisch korrekte Prüfung einer Hypothese betrachtet werden.
- **Multiples Testen:** Bei der Analyse mehrerer Subgruppen besteht eine vom Signifikanzniveau abhängige, manchmal recht hohe Wahrscheinlichkeit, dass Ergebnisse einer Subgruppe statistische Signifikanz erreichen, obwohl es sich in Wahrheit um ein zufälliges Ergebnis handelt. Deshalb muss wie in anderen Situationen, in denen ein multiples Testproblem besteht, das Signifikanzniveau angemessen angepasst werden (siehe auch Abschnitt 1.2).
- **Geringe Power:** Auch wenn eine Subgruppenanalyse kein signifikantes Ergebnis zeigt, ist dieses Ergebnis nicht verlässlich. Oft erreichen Subgruppen nicht die Stichprobengröße, die für das (inferenzstatistische) Aufdecken moderater Unterschiede nötig ist. Anders ist die Situation, wenn schon bei der Fallzahlkalkulation eine für die Subgruppenanalyse angemessene Power berücksichtigt und ein entsprechend erhöhter Stichprobenumfang geplant wurde [3].

Die Ergebnisse von Subgruppenanalysen sollten bei Vorliegen eines der drei oben angesprochenen Aspekte unter starkem Vorbehalt in Bewertungen einfließen und nicht das Ergebnis der primären Analyse dominieren; dies gilt umso mehr, falls das primäre Studienziel nicht erreicht wurde.

Darüber hinaus sind Subgruppenanalysen nicht interpretierbar, wenn das subgruppenbildende Merkmal nach Therapiebeginn (nach Randomisierung) erhoben wurde, z.B. so genannte „Responder-Analysen“.

Der statistische Nachweis unterschiedlicher Effekte zwischen verschiedenen Subgruppen muss anhand eines adäquaten Homogenitäts- bzw. Interaktionstests vorgenommen werden. Der Befund, dass in einer Subgruppe ein statistisch signifikanter Effekt beobachtet wurde, in einer anderen aber nicht, kann (inferenzstatistisch) nicht als das Vorliegen eines Subgruppeneffekts interpretiert werden.

Ungeachtet der oben getroffenen Einschränkungen können für manche Fragestellungen Subgruppenanalysen die auf absehbare Zeit beste wissenschaftliche Evidenz für die Bewertung von Effekten in eben diesen Subgruppen darstellen [4], da z.B. ethische Überlegungen gegen eine Reproduk-

tion der entsprechenden Ergebnisse in einer Validierungsstudie sprechen können. Einen Überblick über Indikationen zur Anwendung von Subgruppenanalysen gibt Rothwell [5]. Eine wichtige Indikation für Subgruppenanalysen stellt eine mögliche Heterogenität des Effekts in verschiedenen klar voneinander abgrenzbaren Patientenpopulationen dar [5,6]. Gibt es a priori Kenntnisse über einen möglichen Effektmodifikator (z.B. Alter, Pathologie), so ist es sogar erforderlich, eine mögliche Heterogenität bezüglich des Effekts in den verschiedenen Patientengruppen vorab zu untersuchen. Ist eine solche vorhanden, so ist der geschätzte Gesamteffekt über alle Patienten nicht sinnvoll interpretierbar [6]. Daher ist es wichtig, dass Kenntnisse über eine mögliche Heterogenität der Patientengruppen beim Studiendesign adäquat berücksichtigt werden. Möglicherweise ist sogar die Durchführung mehrerer Studien notwendig [7].

Den Goldstandard für Subgruppenanalysen stellen Auswertungen bezüglich a priori gebildeter Subgruppen dar, wobei eine anhand der Subgruppen stratifizierte Randomisierung und eine für die Datenanalyse adäquate statistische Methode (Homogenitätstest, Interaktionstest) vorgenommen bzw. angewandt wurden [8].

Das Institut beurteilt unter Berücksichtigung obiger Aspekte Ergebnisse von Subgruppenanalysen sehr vorsichtig, schließt diese aber auch nicht grundsätzlich aus der Bewertung aus.

Literatur

- [1] Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med* 1992; 116: 78-84.
- [2] Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet* 2000; 355: 1064-1069.
- [3] Brookes ST, Whitely E, Egger M, Davey Smith G, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: Risks of subgroup-specific analyses; power and sample size for the interaction test. *J Clin Epidemiol* 2004; 57: 229-236.
- [4] Freemantle N. Interpreting the results of secondary end points and subgroup analyses in clinical trials: should we lock the crazy aunt in the attic? *BMJ* 2001; 322: 989-991.
- [5] Rothwell PM. Treating individuals 2: Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet* 2005; 365: 176-186.
- [6] Kraemer HC, Frank E, Kupfer DJ. (2006): Moderators of treatment outcomes: Clinical, research, and policy importance. *JAMA* 2006; 296: 1286-1289.
- [7] Grouin JM, Coste M, Lewis J. Subgroup analyses in randomized clinical trials: Statistical and regulatory issues. *J Biopharm Stat* 2005; 15: 869-882.
- [8] Cui L, Hung HM, Wang SJ, Tsong Y. Issues related to subgroup analysis in clinical trials. *J Biopharm Stat* 2002; 12: 347-358.

1.5 Aspekte der Beurteilung der Studienqualität

Bei der Beurteilung der allgemeinen Qualität von Studien spielt eine Reihe von Aspekten eine Rolle. Grundsätzlich soll in einer Studie von der Planung über die Durchführung bis zur Auswertung und Präsentation nach einem anerkannten standardisierten Konzept vorgegangen werden. Hierzu gehört ein Studienprotokoll, in dem alle wichtigen Methoden und Vorgehensweisen beschrieben sind. Die üblichen Standards hierbei bilden bei (randomisierten) klinischen Versuchen die Grundprinzipien der „*Good Clinical Practice*“ (GCP) [1,2] und bei epidemiologischen Studien die Leitlinien und Empfehlungen zur Sicherung von „*Guter Epidemiologischer Praxis*“ (GEP) [3]. Ein zentrales Qualitätskriterium von Studien ist dabei, ob die Studie auch so ausgewertet wurde wie in der Planung vorgesehen. Dies ist Publikationen in der Regel nur sehr unzuverlässig zu entnehmen. Ein Abschnitt über die Fallzahlplanung kann hierbei jedoch zumindest Hinweise liefern. Darüber hinaus ist ein Vergleich mit dem eventuell zuvor veröffentlichten Studienprotokoll beziehungsweise einer Publikation zum Studiendesign sinnvoll.

Wichtige Stellungnahmen, die zur Verbesserung der Qualität von Publikationen erarbeitet wurden, sind das CONSORT (Consolidated Standards of Reporting Trials) -Statement für randomisierte klinische Versuche [4] und das dazugehörige Erläuterungsdokument [5], das CONSORT-Statement für Cluster-randomisierte Studien [6], das QUOROM (Quality of Reporting of Meta-analyses) -Statement für Meta-Analysen randomisierter Studien [7], das TREND (Transparent Reporting of Evaluations with Non-randomized Designs) -Statement für nicht-randomisierte Interventionsstudien [8], das STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) -Statement für Beobachtungsstudien in der Epidemiologie [9], die MOOSE (Meta-analysis of Observational Studies in Epidemiology) -Checkliste für Meta-Analysen von Beobachtungsstudien in der Epidemiologie [10], das STARD (Standards for Reporting of Diagnostic Accuracy) -Statement für Diagnosestudien [11] und das dazugehörige Erläuterungsdokument [12]. Wird sich in einer Publikation nicht an diese Standards gehalten, so kann dies ein Indikator für Mängel in der entsprechenden Studie sein. Weitere wichtige Quellen, die grundlegende Aspekte zur Beurteilung der Studienqualität beschreiben, sind die Standardwerke [13-15].

Als Hilfestellung zur Beurteilung der Qualität von Studien wurden verschiedene Systeme entwickelt, wie beispielsweise der Qualitätsindex von Chalmers et al. [16]. Eine Übersicht über Systeme zur Einschätzung der Qualität randomisierter Studien geben Moher et al. [17], eine allgemeine Übersicht bezüglich solcher Systeme für Studien mit unterschiedlichen Designs geben West et al. [18]. Bei der Anwendung solcher Systeme in der Praxis ist allerdings Vorsicht geboten [17], da die Anwendung unterschiedlicher Systeme an ein- und demselben Studienpool zu sehr unterschiedlichen Ergebnissen bezüglich der Qualitätseinordnung und daraus abgeleiteter Schlussfolgerungen

führen kann [19]. Ein einheitliches, allgemein gültiges formales System zur Beurteilung der Studienqualität gibt es nicht [17,18]. Das Institut wird daher die Beurteilung der Studienqualität in Anlehnung an oben zitierte Quellen [1-15] problemorientiert vornehmen.

Zentrale Aspekte zur Beurteilung randomisierter kontrollierter Studien (Randomized Controlled Trials, RCTs) durch das Institut stellen ein adäquates Concealment, das heißt die Unvorhersehbarkeit und Verheimlichung der Gruppenzuteilung (zum Beispiel durch externe Randomisierung bei nicht verblindbaren Studien), die verblindete Zielgrößenerhebung bei nicht verblindbaren Studien (siehe auch Abschnitt 1.15), die adäquate Anwendung des „*Intention-to-treat*“-Prinzips sowie die Festlegung auf ein eindeutiges Hauptzielkriterium beziehungsweise die angemessene Berücksichtigung eines möglichen Multiplizitätsproblems (siehe auch Abschnitt 1.2) dar.

Die Beurteilung formaler Kriterien liefert wesentliche Anhaltspunkte für die Qualität von Studien. Das Institut wird jedoch immer eine über rein formale Aspekte hinausgehende Beurteilung vornehmen, um beispielsweise Fehler, Widersprüche und Ungereimtheiten in Publikationen darzustellen und deren Relevanz für die Ergebnisinterpretation prüfen.

Literatur

- [1] Kolman J, Meng P, Scott G. Good Clinical Practice. Standard Operating Procedures for Clinical Researchers. Chichester: Wiley; 1998.
- [2] ICH Steering Committee. Official web site for the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). <http://www.ich.org/> [Zugriff am 20.10.2004].
- [3] Arbeitsgruppe Epidemiologische Methoden der Deutschen Arbeitsgemeinschaft für Epidemiologie (DAE). Leitlinien und Empfehlungen zur Sicherung von Guter Epidemiologischer Praxis (GEP). Mit Änderungen nach Evaluation, April 2004. <http://www.dgepi.de/doc/Empfehlungen.doc> [Zugriff am 17.11.2006].
- [4] Moher D, Schulz KF, Altman DG for the CONSORT Group. The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann Intern Med* 2001; 134: 657-662.
- [5] Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne DR et al. The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Ann Intern Med* 2001; 134: 663-694.
- [6] Campbell MJ, Elbourne DR, Altman DG for the CONSORT Group. CONSORT statement: Extension to cluster randomised trials. *BMJ* 2004; 328: 702-708.
- [7] Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF et al. Improving the quality of reports of meta-analyses of randomised controlled trials: The QUOROM statement. *Lancet* 1999; 354: 1896-1900.
- [8] Des Jarlais DC, Lyles C, Crepaz N for the TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: The TREND statement. *Am J Public Health* 2004; 94: 361-366.

- [9] STROBE statement: Checklist of essential items Version 3 (Sept 2005). <http://www.strobe-statement.org/PDF/STROBE-Checklist-Version3.pdf> [Zugriff am 31.10.2005].
- [10] Stroup DF, Berlin IA, Morton SC. Meta-analysis of observational studies in epidemiology: A proposal for reporting. *JAMA* 2000; 283: 2008-2012.
- [11] Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Ann Intern Med* 2003; 138: 40-44.
- [12] Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM et al. The STARD statement for reporting studies of diagnostic accuracy: Explanation and elaboration. *Ann Intern Med* 2003; 138: W1-12.
- [13] Egger M, Davey Smith G, Altman DG, editors. *Systematic Reviews in Health Care: Meta-Analysis in Context*. London: BMJ Books; 2001.
- [14] Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions 4.2.5 (updated May 2005)*. In: *The Cochrane Library, Issue 3, 2005*. Chichester: Wiley; 2005.
- [15] Guyatt G, Rennie D, editors. *Users' Guide to the Medical Literature*. Chicago, IL: AMA Press; 2002.
- [16] Chalmers TC, Smith H, Blackburn B, Silverman B, Schroder B, Reitman D, Ambroz A. A method for assessing the quality of a randomized control trial. *Control Clin Trials* 1981; 2: 31-49.
- [17] Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists. *Control Clin Trials* 1995; 16: 62-73.
- [18] West S, King V, Carey TS, Lohr, K.N., McKoy N., Sutton, S.F., Lux, L. *Systems to Rate the Strength of Scientific Evidence. Evidence Report/Technology Assessment No. 47 (Prepared by the Research Triangle Institute-University of North Carolina Evidence-based Practice Center under Contract No. 290-97-0011). AHRQ Publication No. 02-E016*. Rockville, MD: Agency for Healthcare Research and Quality. April 2002. <http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=hstat1.chapter.70996> [Zugriff am 13.03.2006].
- [19] Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999; 262: 1054-1060.

1.6 Ermittlung des Schadenpotenzials medizinischer Interventionen

Die Anwendung einer medizinischen Intervention gleich welcher Art (medikamentös, nichtmedikamentös, chirurgisch, diagnostisch, präventiv etc.) birgt das Risiko unerwünschter Wirkungen in sich. „Unerwünschte Wirkungen“ bezeichnet dabei alle diejenigen Ereignisse und Wirkungen, die einen individuell wahrgenommenen oder objektiv feststellbaren physischen oder psychischen Schaden darstellen, der zu einer mehr oder weniger starken, kurz- oder langfristigen Verkürzung der Lebenserwartung, Erhöhung der Morbidität oder Beeinträchtigung der Lebensqualität führt oder führen kann. Für eine Begriffsklärung ist anzumerken, dass bei Verwendung des Begriffs

„Unerwünschte Wirkungen“ ein kausaler Zusammenhang zur Intervention angenommen wird, während dies bei Verwendung des Begriffs „Unerwünschte Ereignisse“ noch offen bleibt [1].

Das Schadenpotenzial beschreibt das Risiko für das Auftreten unerwünschter Wirkungen bei der Anwendung einer medizinischen Intervention. Die Darstellung des Schadenpotenzials ist wesentlicher und gleichberechtigter Bestandteil der Nutzenbewertung einer Intervention (siehe Abschnitt 2.1). Sie gewährleistet eine informierte populationsbezogene, aber auch individuelle Nutzen-Schaden-Abwägung [2]. Voraussetzung dafür ist, dass die Effektstärken einer medizinischen Intervention sowohl für die erwünschten als auch für die unerwünschten Wirkungen aus den vorliegenden Daten im Vergleich z.B. zu therapeutischen Alternativen beschrieben werden können.

Darstellung, Analyse und Bewertung des potenziellen Schadens einer medizinischen Intervention im Rahmen einer systematischen Übersicht sind jedoch in vielen Fällen ungleich schwerer als die des Nutzens (siehe auch Abschnitt 2.1). Dies betrifft insbesondere unerwartet auftretende unerwünschte Ereignisse [1]. Typischerweise sind Studien so konzipiert, dass sie den Einfluss einer medizinischen Intervention auf wenige, vorab definierte Wirksamkeitsendpunkte messen sollen. Die Ergebnisse bezüglich unerwünschter Ereignisse in solchen Studien hängen stark von der zu Grunde gelegten Methodik der Erfassung ab [3,4]. Des Weiteren ist die zielgerichtete Forschung zur Erhebung insbesondere der seltenen schwerwiegenden unerwünschten Wirkungen einschließlich der Darstellung eines kausalen Zusammenhangs mit der medizinischen Intervention erheblich unterrepräsentiert [5-7]. Darüber hinaus ist die Qualität der Berichterstattung bezüglich unerwünschter Ereignisse in Einzelstudien mangelhaft, was kürzlich zu einer Erweiterung des CONSORT-Statements für randomisierte klinische Studien geführt hat [8]. Schließlich wird die systematische Bewertung unerwünschter Wirkungen einer Intervention auch dadurch erschwert, dass die Kodierung in Literaturdatenbanken diesbezüglich unzureichend ist und die zielgerichtete Suche nach relevanter wissenschaftlicher Literatur deshalb häufig ein unvollständiges Bild liefert [9].

Folge der oben genannten Hindernisse ist, dass in vielen Fällen trotz enormer Anstrengungen die Aussagenunsicherheit bei der Darstellung des Schadenpotenzials höher sein wird als bei der Darstellung positiver Effekte [10]. Notwendig ist hier, eine sinnvolle Balance zwischen Vollständigkeit der Aufarbeitung und Ressourcenaufwand zu finden. Daraus folgend wird es notwendig sein, die Aufarbeitung und Darstellung auf relevante unerwünschte Wirkungen zu beschränken. Als relevant können dabei insbesondere diejenigen unerwünschten Wirkungen bezeichnet werden, die

- den Nutzen der Intervention teilweise oder ganz aufwiegen könnten,
- sich zwischen zwei oder mehr ansonsten gleichwertigen Therapieoptionen erheblich unterscheiden könnten,

- insbesondere bei denjenigen von mehreren Therapieoptionen auftreten, die besonders effektiv sein könnten,
- eine Dosis-Wirkungs-Beziehung haben könnten,
- von Patienten als besonders wichtig angesehen werden,
- von schwerwiegender Morbidität oder gar erhöhter Mortalität begleitet sind oder in Folge mit einer erheblichen Beeinträchtigung der Lebensqualität einhergehen.

Im Interesse der Patientensicherheit und des ärztlichen Axioms des „primum nil nocere“ wird das Institut bei der Aufarbeitung und Darstellung unerwünschter Wirkungen folgende Grundsätze beachten:

- Grundlage der Auswahl relevanter unerwünschter Wirkungen nach den oben genannten Kriterien bildet die Zusammenstellung derjenigen unerwünschten Wirkungen und Ereignisse, die bei der Entscheidungsfindung für oder gegen den Einsatz der zu evaluierenden Intervention einen wesentlichen Stellenwert haben. Die Zusammenstellung erfolgt im Rahmen der Vorrecherche zur jeweiligen Fragestellung insbesondere auf Grundlage der Daten kontrollierter Interventionsstudien, in denen zielgerichtet der Nutzen der Intervention untersucht wurde, sowie auf Basis vorliegender epidemiologischer Studien (zum Beispiel Kohorten- oder Fall-Kontroll-Studien), Pharmakovigilanzdaten, Informationen von Zulassungsbehörden etc. Im Einzelfall können hier auch Ergebnisse aus Tierexperimenten sowie aus Experimenten zur Überprüfung eines pathophysiologischen Konstrukts hilfreich sein.
- Liegen aus den oben genannten Datenquellen Hinweise für das Vorliegen einer unerwünschten Wirkung vor, so wird dieses so lange als möglich angesehen, bis diese unerwünschte Wirkung durch die Ergebnisse zielgerichteter Forschung mit hinreichender Sicherheit ausgeschlossen werden kann. Dies gilt insbesondere für schwerwiegende unerwünschte Wirkungen. Die Evidenzhierarchie entspricht dabei derjenigen für Therapiestudien (siehe auch die Abschnitte 1.9 und 1.10). Unter „hinreichender Sicherheit“ kann dabei z.B. verstanden werden, dass die entsprechende(n) Untersuchung(en) in ihrem Design und ihrer Planung primär auf den Nachweis der Nichtunterlegenheit der zu evaluierenden Intervention gegenüber anderen Therapieoptionen (oder Placebo bzw. keiner Intervention, je nach Fragestellung) gerichtet war(en) und eine adäquate Definition der Nichtunterlegenheit beinhaltet/n.

Literatur

- [1] Council of Europe: Committee of Experts on Management of Safety and Quality in Health Care (SP-SQS). Expert Group on Safe Medication Practices. Glossary of terms related to patient and medication safety.
http://www.who.int/entity/patientsafety/highlights/COE_patient_and_medication_safety_gl.pdf [Zugriff am 17.9.2006].
- [2] Ziegler DK, Mosier MC, Buenaver M, Okuyemi K. How much information about adverse effects of medication do patients want from physicians? *Arch Intern Med* 2001; 161: 706-713.
- [3] Bent S, Padula A, Avins AL. Brief communication: Better ways to question patients about adverse medical events. A randomized, controlled trial. *Ann Intern Med* 2006; 144: 257-261.
- [4] Ioannidis JPA, Mulrow CD, Goodman SN. Adverse events: The more you search, the more you find. *Ann Intern Med* 2006; 144: 298-300.
- [5] Edwards JE, McQuay HJ, Moore RA, Collins SL. Reporting of adverse effects in clinical trials should be improved: Lessons from acute postoperative pain. *J Pain Symptom Manage* 1999; 18: 427-437.
- [6] Ioannidis JPA, Lau J. Completeness of safety reporting in randomized trials – an evaluation of seven medical areas. *JAMA* 2001; 285: 437-443.
- [7] Bonhoeffer J, Zumbrunn B, Heining U. Reporting of vaccine safety data in publications: Systematic review. *Pharmacoepidemiol Drug Saf* 2005; 14: 101-106.
- [8] Ioannidis JPA, Evans SJW, Gøtzsche PC, O’Neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: An extension of the CONSORT statement. *Ann Intern Med* 2004; 141: 781-788.
- [9] Derry S, Loke YK, Aronson JK. Incomplete evidence: The inadequacy of databases in tracing published adverse drug reactions in clinical trials. *BMC Med Res Methodol* 2001; 1: 7.
- [10] Loke YK, Price D, Herxheimer A on behalf of the Cochrane Adverse Effects Subgroup. Including adverse effects. In: Higgins JPT, Green S, editors: *Cochrane Handbook for Systematic Reviews of Interventions* 4.2.5 [updated May 2005]; Appendix 6b. In: *The Cochrane Library*, Issue 3, 2005. Chichester: Wiley; 2005.

1.7 Beurteilung von Studien mit „überholten“ Methoden

Tritt der Fall ein, dass bei einer Projektarbeit des Instituts ältere Studien beurteilt werden müssen, die den heute gültigen Qualitätsansprüchen (siehe auch Abschnitt 1.8) nicht genügen, weil sie in einer Zeit geplant und durchgeführt wurden, in denen diese Ansprüche noch nicht existierten, so wird das Institut die Nachteile und Mängel dieser Studien aufzeigen und mögliche Konsequenzen diskutieren. Ein Unterschied zum Umgang mit neueren Studien, die ähnliche Qualitätsmängel haben, wird jedoch nur dann vorgenommen, wenn dies aus der Fragestellung oder anderen Umständen der Bewertung heraus klar begründbar ist.

1.8 Beurteilung verschiedener Studienarten

Es werden hier nur die wichtigsten Designs zusammengefasst, die bei der Nutzenbewertung (siehe auch Abschnitt 2.1) in der medizinischen Forschung eine Rolle spielen. Zunächst kann man unterscheiden zwischen Beobachtungsstudien und Interventionsstudien. Innerhalb der Beobachtungsstudien gibt es häufig den ersten Informationsgewinn aus Fallberichten oder Fallserien. Diese sind anfällig für Verzerrungen aller Art, so dass hier, je nach Fragestellung, nur sehr beschränkt eine verlässliche Evidenz ableitbar ist (siehe auch Abschnitt 1.9). Bevölkerungsbezogene Querschnittstudien sind geeignet zur Schätzung von Krankheitsprävalenzen. Weitere wichtige Studientypen der Epidemiologie sind Fall-Kontroll-Studien [1], in denen retrospektiv nach Expositionen von Fällen und Kontrollen gefragt wird, und Kohortenstudien [2], in denen bestimmte Gruppen (Kohorten) über einen Zeitverlauf hinweg beobachtet werden. Kohortenstudien sind in diesem Sinne prospektiv angelegt, allerdings gibt es auch retrospektive Kohortenstudien, in denen die Exposition aus der Vergangenheit erfasst wird (häufig in der Arbeits- oder auch Pharmakoepidemiologie anzutreffen). Grundsätzlich sind prospektive Studien retrospektiven Designs vorzuziehen. Allerdings sind zum Beispiel Fall-Kontroll-Studien häufig die einzige praktikable Möglichkeit, Informationen über Zusammenhänge zwischen Expositionen und seltenen Erkrankungen zu gewinnen.

Im Bereich der Interventionsstudien ist an erster Stelle eine Kontrollgruppe zu fordern. Aus einem reinen Vorher-Nachher-Vergleich in einem Design mit abhängigen Stichproben ohne Kontrollgruppe lässt sich in der Regel nicht der Effekt einer Intervention ableiten. Ausnahmen bilden Krankheitsbilder mit einem deterministischen (oder nahezu deterministischen) Verlauf (z.B. diabetisches ketoazidotisches Koma). Gütekriterien, welche die Aussagekraft kontrollierter Studien erhöhen, sind Randomisierung und Verblindung (siehe auch die Abschnitte 1.5 und 1.18). Häufige Designs bei klinischen Studien bilden parallele Gruppen [3], Cross-over-Studien [4] und Cluster-randomisierte Studien [5]. Sind in einer Studie Zwischenauswertungen geplant, so muss dies mit Hilfe eines adäquaten sequentiellen Designs berücksichtigt werden [6].

Diagnose- und Screeningstudien können sehr unterschiedliche Ziele haben, so dass hiervon die Wahl eines adäquaten Designs abhängt (siehe auch die Abschnitte 2.3 und 2.4).

Zur Untersuchung von genetischen Faktoren, die die Entstehung und Verteilung von Krankheiten verursachen können, ist in den letzten Jahren die relativ neue Disziplin der genetischen Epidemiologie entstanden [7]. In diesem Bereich gibt es im Rahmen von Kopplungs- und Assoziationsstudien eine Reihe neuer spezieller Studiendesigns.

Literatur

- [1] Breslow NE, Day NE. Statistical Methods in Cancer Research Vol. I: The Analysis of Case-Control Studies. Lyon: Int. Agency for Res. on Cancer; 1980.
- [2] Breslow NE, Day NE. Statistical Methods in Cancer Research Vol. II: The Design and Analysis of Cohort Studies. Lyon: Int. Agency for Res. on Cancer; 1987.
- [3] Pocock SJ. Clinical Trials: A Practical Approach. Chichester: Wiley; 1983.
- [4] Jones B, Kenward MG. Design and Analysis of Cross-Over Trials. London: Chapman & Hall; 1989.
- [5] Donner A, Klar J. Design and Analysis of Cluster Randomization Trials in Health Research. London: Arnold; 2000.
- [6] Whitehead J. The Design and Analysis of Sequential Clinical Trials. Chichester: Ellis Horwood; 1983.
- [7] Khoury MJ, Beaty TH, Cohen BH. Fundamentals of Genetic Epidemiology. New York: Oxford University Press; 1993.

1.9 Rangordnung verschiedener Studienarten/Evidenzgrade

Es gibt verschiedene Ansätze, bestimmten Studientypen bestimmte Evidenzgrade zuzuordnen, um so eine Rangfolge bezüglich der Validität der Evidenz aus verschiedenen Studienarten im Rahmen einer systematischen Übersicht oder für die Entwicklung von Leitlinien zu erstellen [1,2]. Es gibt jedoch kein System, das allgemein akzeptiert und universell in allen systematischen Übersichten einsetzbar ist [3]. Aufgrund der Komplexität der Studienbewertung (siehe Abschnitt 1.5) lässt sich aus einer Evidenzhierarchie auch kein abschließendes Qualitätsurteil bilden. Das Institut orientiert sich im Allgemeinen an folgender grober Hierarchie von Studientypen, die weitgehend akzeptiert wird und auch weitgehend mit der Evidenzklassifizierung des G-BA übereinstimmt [4]. Zumindest im Rahmen von Therapiestudien wird der höchste Evidenzgrad systematischen Übersichten von RCTs zugeordnet. Auf dem nächsten Platz erscheinen einzelne RCTs. Diese werden in einigen Einteilungen noch in solche hoher oder weniger hoher Qualität eingestuft, wobei die Vermengung von Konzept- und Ergebnisqualität von anderen Autoren wiederum kritisiert wird [5]. Danach folgen nicht randomisierte Interventionsstudien, prospektive Beobachtungsstudien, retrospektive Beobachtungsstudien, nicht experimentelle Studien (Fallserien und Fallberichte) und – mit niedrigstem Evidenzgrad – Expertenmeinungen ohne wissenschaftliche Begründung. Dieses grobe System wird das Institut der jeweiligen Situation und Fragestellung anpassen und bei Bedarf detaillierter darstellen [2].

Literatur

- [1] Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ et al. Users' guides to the medical literature. IX. A method for grading health care recommendations. JAMA 1995; 274: 1800-1804.
- [2] Harbour R, Miller J, for the Scottish Intercollegiate Guidelines Network Grading Review Group. A new system for grading recommendations in evidence based guidelines. BMJ 2001; 323: 334-336.
- [3] Katrak P, Bialocerkowski AE, Massy-Westropp N, Kumar VSS, Grimme KA. A systematic review of the content of critical appraisal tools. BMC Med Res Methodol 2004; 4: 22.
- [4] Gemeinsamer Bundesausschuss. Verfahrensordnung des Gemeinsamen Bundesausschusses vom 20. September 2005, veröffentlicht im Bundesanzeiger 2005, S. 16998, in Kraft getreten am 1. Oktober 2005, zuletzt geändert am 18. April 2006, veröffentlicht im Bundesanzeiger 2006, S. 4876, in Kraft getreten am 7. Juli 2006.
<http://www.g-ba.de/cms/upload/pdf/richtlinien/2006-07-07-VerfO.pdf> [Zugriff am 30.11.2006].
- [5] Windeler J, Ziegler S. EBM-Splitter: Evidenzklassifizierungen. Z ärztl Fortbild Qual sich 2003; 97: 513-514.

1.10 Bezug zwischen Studientyp/-art und Fragestellung

Der randomisierte kontrollierte Versuch gilt als Studientyp höchster Qualität. Allerdings muss dies in Bezug zur Fragestellung gesehen werden. Nur beim Wirksamkeitsnachweis einer Intervention ist in der Regel eine randomisierte kontrollierte Studie das zu fordernde Design. Bei anderen Fragestellungen kommen alternative Studiendesigns in Frage. Die wichtigsten Situationen hierbei sind die folgenden: Zur Untersuchung der Diskriminationsfähigkeit diagnostischer Methoden genügt in vielen Fällen eine Querschnittsstudie (siehe auch Abschnitt 2.3). Das optimale Design zur Erforschung prognostischer Faktoren ist eine prospektive Kohortenstudie. Zur Untersuchung des Zusammenhangs zwischen Expositionen und sehr seltenen Erkrankungen kommen Fall-Kontroll-Studien in Frage. Sollen allerdings diagnostische Verfahren oder Prognosefaktoren zusammen mit den aus dem Informationsgewinn resultierenden Konsequenzen (zum Beispiel Einleitung einer Therapie) als Strategie geprüft werden, ist wiederum die randomisierte Studie das Design der Wahl (siehe auch Abschnitte 2.3, 2.4 und 2.10).

1.11 Beurteilung nicht oder nicht vollständig publizierter Daten

In der Praxis ergibt sich häufig das Problem, dass zur Beurteilung von Publikationen notwendige Daten oder Informationen ganz oder teilweise fehlen. Dies betrifft in erster Linie so genannte „graue“ Literatur und Abstracts, aber auch vollständige Publikationen. Darüber hinaus ist es mög-

lich, dass Studien zum Zeitpunkt der Bewertung einer Technologie durch das Institut (noch) nicht publiziert sind.

Das Ziel des Instituts ist es, die Bewertung auf einer möglichst vollständigen Datenbasis vorzunehmen (siehe auch Abschnitt 4.7.3). Im Fall fehlender relevanter Information versucht das Institut deshalb, diese fehlenden Informationen zu vervollständigen. Dies geschieht unter anderem durch den Kontakt mit den Autoren von Publikationen oder den Sponsoren von Studien. Je nach Art des zu erarbeitenden Produkts (siehe auch Kapitel 4) kann es allerdings aufgrund zeitlicher Vorgaben nur eingeschränkt möglich sein, nicht publizierte Informationen zu erfragen.

Ein häufig auftretendes Problem besteht darin, dass wichtige Angaben zur Durchführung einer Meta-Analyse (z.B. Varianzen der Effektschätzer) fehlen. In vielen Fällen lassen sich jedoch die fehlenden Daten aus den vorhandenen Angaben berechnen oder zumindest schätzen [1-3]. Das Institut wird nach Möglichkeit solche Verfahren anwenden.

Bei nur teilweise vorhandenen Informationen bzw. bei Verwendung geschätzter Werte wird die Robustheit der Ergebnisse ggf. mit Hilfe von Sensitivitätsanalysen untersucht und diskutiert. Dies kann beispielsweise im Sinne von Best-Case- und Worst-Case-Szenarien durchgeführt werden. Ein Worst-Case-Szenarium kann jedoch hier nur als Beleg der Robustheit eines gefundenen Effekts verwendet werden. Aus einem Worst-Case-Szenarium, bei dem ein vorher gefundener Effekt nicht bestätigt wird, kann nicht ohne weiteres abgeleitet werden, dass ein solcher Effekt nicht nachgewiesen ist. Beim weitgehenden oder vollständigen Fehlen relevanter Informationen kann es vorkommen, dass eine Publikation nicht beurteilt werden kann. In diesen Fällen wird lediglich dargestellt, dass es zu einem bestimmten Thema weitere Daten gibt, die aber zur Beurteilung nicht zur Verfügung stehen.

Literatur

- [1] Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Stat Med* 1998; 17, 2815-2834.
- [2] Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol* 2005; 5: 13.
- [3] Di Pietrantonj C. Four-fold table cell frequencies imputation in meta analysis. *Stat Med* 2006; 25: 2299-2322.

1.12 Beurteilung der Konsistenz publizierter Daten

Für die Einschätzung der Aussagekraft publizierter Ergebnisse überprüft das Institut die Daten auf Konsistenz (Plausibilität, Vollständigkeit). Unplausible Daten entstehen zum einen durch eine fehlerhafte Darstellung der Ergebnisse (Tipp-, Formatierungs- oder Rechenfehler) aber auch durch

eine mangel- beziehungsweise fehlerhafte Beschreibung der Methodik oder sogar durch gefälschte oder erfundene Daten [1]. Inkonsistenzen können innerhalb einer Publikation bestehen, aber auch zwischen verschiedenen Publikationen zu ein- und derselben Studie.

Ein Problem vieler Publikationen sind unvollständige Angaben im Methodik- und Ergebnisteil. Insbesondere die Darstellung von Lost-to-follow-up-Patienten, Therapieabbruchern etc. sowie deren Art der Berücksichtigung in den Auswertungen sind oft nicht transparent.

Es ist daher angezeigt, mögliche Inkonsistenzen in den Daten aufzudecken. Dazu wird das Institut beispielsweise Rechenschritte überprüfen und die Angaben in Text, Tabellen und Grafiken vergleichen. Ein in der Praxis häufiges Problem bei Überlebenszeitanalysen sind Inkonsistenzen zwischen den Angaben zu Lost-to-follow-up-Patienten und den Patienten unter Risiko bei der grafischen Darstellung von Überlebenskurven. Bei bestimmten Endpunkten, zum Beispiel Gesamtmortalität, kann die Anzahl der Lost-to-follow-up-Patienten berechnet werden, wenn die Kaplan-Meier-Schätzer mit den Patienten unter Risiko zu einem Zeitpunkt vor der minimalen Follow-up-Zeit abgeglichen werden. Zur Aufdeckung erfundener und gefälschter Daten können statistische Techniken hilfreich sein [1].

Finden sich relevante Inkonsistenzen in der Ergebnisdarstellung, besteht in der Regel das Ziel des Institutes in der Klärung dieser Inkonsistenzen und/oder Beschaffung fehlender Informationen, beispielsweise durch den Kontakt mit den Autoren oder die Anforderung von kompletten Studienberichten und weiteren Studienunterlagen. Dabei ist allerdings zu bedenken, dass erstens Autorenanfragen, besonders wenn die Publikation länger zurückliegt, häufig nicht beantwortet werden und dass sie zweitens zu weiteren Inkonsistenzen führen können. Insofern ist im Einzelfall eine Abwägung von Aufwand und Nutzen solcher Anfragen sinnvoll und notwendig. Lassen sich Inkonsistenzen nicht klären, wird der mögliche Einfluss der Inkonsistenzen auf die Effektstärken (Stärke des Bias), die Ergebnisunsicherheit (Erhöhung der Fehlerwahrscheinlichkeiten) und die Präzision (Breite der Konfidenzintervalle) vom Institut abgeschätzt. Dazu können beispielsweise Sensitivitätsanalysen durchgeführt werden. Besteht die Möglichkeit, dass Inkonsistenzen die Ergebnisse relevant beeinflussen, wird dieser Sachverhalt dargestellt und die Ergebnisse werden sehr zurückhaltend interpretiert.

Literatur

- [1] Al-Marzouki S, Evans S, Marshall T Roberts I. (2005): Are these data real? Statistical methods for the detection of data fabrication in clinical trials. *BMJ* 2005; 331: 267-270.

1.13 Umgang mit Rohdaten

Im Rahmen der wissenschaftlichen Bewertung medizinischer Leistungen ist eine der Hauptaufgaben des Instituts das Sammeln und die Analyse bereits publizierter Daten aus systematischen Literaturrecherchen. Es ist jedoch möglich, dass das Institut für bestimmte Fragestellungen auch bisher noch nicht analysierte Rohdaten auswertet, die von externen Quellen zur Verfügung gestellt werden (zum Beispiel Krankenkassen). Eine sinnvolle Analyse solcher Daten setzt voraus, dass die Rahmenbedingungen klar sind, unter denen diese Daten erhoben wurden und die Plausibilität und Qualität der Daten überprüfbar sind. Insbesondere ist darauf zu achten, dass wesentliche Qualitätskriterien eingehalten werden. Zum Beispiel sollten die Daten bei Therapiestudien unter Einhaltung des GCP-Standards entstanden sein (siehe auch Abschnitt 1.5). Darüber hinaus wird in den meisten Fällen die Vorlage eines Studienprotokolls zur adäquaten Beurteilung notwendig sein. Im Umgang mit Rohdaten werden datenschutzrechtliche Aspekte berücksichtigt (siehe auch Abschnitt 1.16).

1.14 Beurteilung von Ergebnisunsicherheiten

Grundsätzlich ist jedes Ergebnis einer empirischen Studie oder einer systematischen Übersicht von empirischen Studien unsicher. Hierbei ist zu unterscheiden zwischen qualitativer und quantitativer Ergebnisunsicherheit. Die qualitative Unsicherheit wird bestimmt durch das Studiendesign, aus dem sich Evidenzgrade ableiten lassen (siehe Abschnitt 1.9), sowie durch die Studienqualität, die in Abhängigkeit des Studiendesigns beurteilt werden muss (siehe Abschnitt 1.5). Bei systematischen Übersichten spielen außerdem die Qualität der Recherchestrategie sowie möglicherweise die Wahl meta-analytischer Verfahren zur Datenzusammenfassung eine Rolle (siehe Abschnitt 1.21).

Neben der qualitativen Ergebnisunsicherheit gibt es quantitativ messbare Unsicherheiten aufgrund statistischer Gesetzmäßigkeiten. Die statistische Unsicherheit einer Parameterschätzung, die bedingt ist durch die begrenzte Stichprobe, lässt sich in Form von Standardfehler und Konfidenzintervall quantifizieren und beurteilen. Wann immer möglich soll ein adäquates Konfidenzintervall angegeben werden mit der Information, ob es sich um ein- oder zweiseitige Konfidenzgrenzen handelt, sowie dem gewählten Konfidenzniveau (siehe auch Abschnitt 1.1).

Es darf jedoch nicht übersehen werden, dass diese Berechnungen unter der Annahme durchgeführt werden, dass das gewählte statistische Modell richtig ist und dass keine anderweitigen systematischen Fehler und Verzerrungen vorliegen. Die Unsicherheiten, die dadurch entstehen, dass die tatsächlichen Gegebenheiten mehr oder weniger stark vom gewählten statistischen Modell abweichen, bleiben hier unberücksichtigt [1,2]. Es gibt formale Ansätze, auch diese allgemeinen Modellunsicherheiten zu berücksichtigen, zum Beispiel durch Bayes'sche Methoden [3] oder mit Hilfe von

Simulationstechniken [4]. Diese Verfahren sind aber noch nicht ausreichend entwickelt und untersucht, um routinemäßig in der Praxis verwendet werden zu können [5-7]. Bei Bedarf wird das Institut den Einsatz dieser Methoden jedoch in Erwägung ziehen. Eine qualitative Beurteilung der allgemeinen Ergebnisunsicherheit wird nach Bedarf auf der Basis der aktuellen Literatur zum jeweiligen Thema durchgeführt. Nach wie vor sind die klassischen Kausalitätskriterien von Hill [8] hier eine Hilfe.

Literatur

- [1] Chatfield C. Model uncertainty, data mining and statistical inference (with discussion). J R Stat Soc A 1995; 158: 419-466.
- [2] Buckland ST, Burnham KP, Augustin NH. Model selection: An integral part of inference. Biometrics 1997; 53: 603-618.
- [3] Draper D. Assessment and propagation of model uncertainty (with discussion). J R Stat Soc B 1995; 57: 45-97.
- [4] Phillips CV. Quantifying and reporting uncertainty from systematic errors. Epidemiology 2003; 14: 459-466.
- [5] Hjort NL, Claeskens G. Frequentist model average estimators. J Am Stat Assoc 2003; 98: 879-899.
- [6] Wang D, Zhang W, Bakhai A. Comparison of Bayesian model averaging and stepwise methods for model selection in logistic regression. Stat Med 2004; 23: 3451-3467.
- [7] Augustin N, Sauerbrei W, Schumacher M. (2005): The practical utility of incorporating model selection uncertainty into prognostic models for survival data. Stat Modelling 5, 95-118.
- [8] Hill AB. The environment and disease: Association or causation? Proc R Soc Med 1965; 58: 295-300.

1.15 Beurteilung nicht verblindbarer Verfahren

Zur Vermeidung systematischer Abweichungen (Verzerrung, Bias) einer Effektschätzung vom wahren Effekt (siehe auch Abschnitt 1.18) werden kontrollierte Studien, wenn möglich, randomisiert und doppelblind durchgeführt. In einigen Situationen ist jedoch eine Verblindung der Intervention nicht möglich. Auch nicht verblindete Studien können interpretierbare Ergebnisse liefern. Wichtige Instrumente zur Vermeidung von Bias sind hier wiederum Randomisierung und die Wahl geeigneter Zielvariablen. In nicht verblindbaren Studien ist es zentral wichtig, dass eine adäquat verdeckte Zuteilung (Allocation Concealment) der Patienten zu den zu vergleichenden Gruppen gewährleistet ist (siehe auch Abschnitt 1.5). Weiterhin ist es erforderlich, dass die Zielvariable unabhängig vom (unverblindeten) Behandler ist bzw. unabhängig vom Behandler verblindet erhoben wird (verblindete Zielgrößenerhebung). Falls eine verblindete Zielgrößenerhebung nicht möglich ist, sollte ein „harter“ objektiver Endpunkt gewählt werden (beispielsweise Mortalität), der in

seiner Ausprägung und in der Stringenz der Erfassung praktisch nicht durch denjenigen, der den Endpunkt (unverblindet) erhebt, beeinflusst werden kann.

1.16 Berücksichtigung datenschutzrechtlicher Aspekte/Vertraulichkeit

Die Verarbeitung personenbezogener oder personenbeziehbarer Daten findet innerhalb des Institutes nach Maßgaben der jeweiligen Bundesdatenschutzgesetzgebung statt. Für deren Einhaltung ist die oder der vom Institut bestellte Datenschutzbeauftragte zuständig.

Darüber hinaus ist es zukünftig denkbar, dass das Institut auch im Rahmen von Forschungsarbeiten personenbeziehbare sowie in Ausnahmefällen personenbezogene Daten verarbeitet. Soweit diese ursächlich an dritter Stelle erfasst wurden oder werden, sind entsprechende Erklärungen über die Einhaltung der gesetzlichen Bestimmungen abzugeben, andererseits ist vorab jeweils im Einzelfall sorgfältig zu prüfen, inwieweit die notwendigen gesetzlichen Voraussetzungen vorliegen (Einverständniserklärungen, Aufklärung etc.).

Weiterhin sollte angestrebt werden, primär personenbezogene Daten in anonymisierter oder pseudonymisierter Form durch Dritte anzunehmen und weiterzuverarbeiten, was in den meisten Fällen für die Forschungszwecke und einzelnen wissenschaftlichen Fragestellungen ausreichend sein dürfte. Insbesondere können damit eventuell Vorbehalte zur Weitergabe von Daten an das Institut ausgeräumt werden.

Werden dem Institut Daten übermittelt, die nicht publiziert werden dürfen, so können diese nicht inhaltlich in Nutzenbewertungen des Institutes einfließen, da dies dem Transparenzgebot widerspricht (siehe Abschnitt 4.7.3).

Hinsichtlich der vertraulichen Behandlung von Unternehmensdaten sind gegebenenfalls geeignete Erklärungen gegenüber Dritten abzugeben, die eine Wahrung der Vertraulichkeit nach außen gewährleisten. Neben der notwendigen technischen Infrastruktur zur Wahrung der Datensicherheit finden sich in allen Arbeitsverträgen, die das Institut mit seinen Mitarbeitern abschließt, entsprechende Passagen, die die Mitarbeiter zur Vertraulichkeit verpflichten. Entsprechende Verpflichtungen müssen im Einzelfall auch extern beauftragte Personen oder Einrichtungen gegenüber dem Institut abgeben.

1.17 Berücksichtigung ethischer Aspekte

Das Institut wird getragen von dem primären Ziel, mit qualitativ hochwertiger Arbeit die Gesundheitsversorgung der Menschen in Deutschland zu verbessern. Dabei stehen für das Institut die Ma-

ximierung des allgemeinen, aber auch des individuellen Patientennutzens sowie die Stärkung der Patientenautonomie durch Gesundheitsbildung und Information im Vordergrund. Die Methoden der evidenzbasierten Medizin werden hierbei als notwendige und hilfreiche Werkzeuge verstanden. Das Institut wird diese Werkzeuge gewissenhaft anwenden und ihre Grenzen berücksichtigen.

Das Institut ist sich darüber hinaus seiner Position im deutschen Gesundheitswesen und insbesondere seiner Verantwortung gegenüber den Personen und Institutionen, die Gesundheitsleistungen in Anspruch nehmen, durchführen, finanzieren oder entwickeln, bewusst. Auch wenn durch den Gesetzgeber eine strikte Trennung zwischen wissenschaftlicher Bewertung einerseits und Entscheidung für oder gegen die Aufnahme einer medizinischen Intervention in die Regelversorgung andererseits vorgesehen wurde, ist sich das Institut darüber im Klaren, dass die Ergebnisse der Institutsarbeit mittel- oder unmittelbar Einfluss auf die medizinische Versorgung nehmen können. Daraus folgend ist die Berücksichtigung der möglichen oder wahrscheinlichen Folgen, die die Berichterstattung des Institutes für einzelne Personen, Bevölkerungs- oder Berufsgruppen sowie Institutionen oder Unternehmen hat oder haben wird, elementarer Bestandteil der Institutsarbeit. Die Einbeziehung einzelner Vertreter der von der jeweiligen Institutsarbeit tangierten Gruppen und Institutionen wird hierbei unterstützend sein. Das Institut sieht sich dabei primär verantwortlich für die Belange aller Bürger und achtet deshalb auf Transparenz und Unabhängigkeit. Besonders wichtig ist für das Institut, dass Aussagen seiner Berichte und anderen Arbeitsergebnissen nicht von Interessen einzelner Gruppen beeinflusst werden können. Innerhalb des Instituts soll Transparenz sicherstellen, dass keine verdeckten Einflussnahmen möglich sind.

Das Institut wird sich Fragen der Verteilungsgerechtigkeit nicht verschließen. Begrenzte Ressourcen führen dazu, dass vermehrte Investitionen in einem Bereich des Gesundheitswesens zwangsläufig eine Beschränkung in anderen Bereichen mit sich bringen. Das Institut wird vermitteln, dass die Entscheidung für oder gegen eine medizinische Methode einem gewissenhaften Abwägungsprozess von allgemeingültigen Prioritäten entspringen muss. In diesem Zusammenhang sieht sich das Institut insbesondere auch in der Verantwortung gegenüber Minderheiten und bereits benachteiligten Bevölkerungsgruppen.

Für das Institut stehen ethische Belange auch bei eigenen Forschungen im Vordergrund. Auch im Rahmen einer Berichterstellung sind bei Forschungsaktivitäten die Abwägung des Für und Wider solcher Vorhaben für die Betroffenen sowie gegebenenfalls die für Studien vorgesehene Inanspruchnahme einer Beratung in ethischen Fragen bei der Planung und Durchführung erforderlich.

1.18 Darstellung von Biasarten

Unter Bias versteht man eine systematische Abweichung der Schätzung eines Effekts aus Studiendaten vom wahren Effekt. Es gibt vielfältige mögliche Ursachen, die einen Bias hervorrufen können [1]. Nur die wichtigsten Biasarten werden im Folgenden beschrieben; eine ausführliche Übersicht über verschiedene Biasarten in unterschiedlichsten Situationen gibt Feinstein [2].

Ein *Selection Bias* entsteht durch eine Verletzung der Zufallsprinzipien bei Stichprobenziehungen. Speziell beim Vergleich von zwei Gruppen kann ein Selection Bias zu systematischen Unterschieden zwischen den Gruppen führen. Sind dadurch wichtige Confounder in den Gruppen ungleich verteilt, so sind die Ergebnisse eines Vergleichs in aller Regel nicht mehr interpretierbar. Beim Vergleich von Gruppen ist die Randomisierung die beste Methode zur Vermeidung eines Selection Bias, da die entstehenden Gruppen sich sowohl bezüglich bekannter als auch unbekannter Confounder nicht systematisch unterscheiden. Die Gewährleistung einer Strukturgleichheit ist allerdings nur bei ausreichend großen Stichproben gegeben. In kleinen Studien kann es trotz Randomisierung zufällig zu bedeutsamen Gruppenunterschieden kommen. Beim Vergleich von Gruppen mit bestehender Strukturungleichheit kann der Effekt bekannter Confounder mit Hilfe multifaktorieller Methoden berücksichtigt werden (siehe auch Abschnitt 1.22), es bleibt aber das Problem eines systematischen Unterschieds zwischen den Gruppen durch nicht bekannte oder nicht ausreichend erhobene Confounder.

Eine Verzerrung durch unterschiedliche Behandlungen (mit Ausnahme der zu untersuchenden Intervention) wird als *Performance Bias* bezeichnet. Neben der Vergleichbarkeit der Gruppen bezüglich potenzieller prognostischer Faktoren spielen die Behandlungsgleichheit und die Beobachtungsgleichheit aller Probanden eine entscheidende Rolle. Eine Verletzung der Beobachtungsgleichheit kann zu einem *Detection Bias* führen. Die Verblindung ist ein wirksamer Schutz vor beiden Biasarten, die in der Epidemiologie als *Information Bias* zusammengefasst werden.

Protokollverletzungen und Studienaustritte können das Studienergebnis systematisch verzerren, was als *Attrition Bias* bezeichnet wird. Zur Vermeidung von Attrition Bias kann das Intention-to-treat-Prinzip eingesetzt werden, das besagt, dass alle randomisierten Probanden in der Analyse berücksichtigt werden und zwar in der durch die Randomisierung zugeordneten Gruppe, unabhängig von Protokollverletzungen.

In Diagnosestudien muss die Untersuchung des diagnostischen Tests in einem angemessenen Spektrum von Patienten erfolgen. Unterscheidet sich das Untersuchungskollektiv systematisch von der Patientenpopulation, in der der Test zur Anwendung kommen soll, so kann dies zum *Spectrum*

Bias führen (siehe auch Abschnitt 2.3). Um diesen Bias zu vermeiden, muss der diagnostische Test in einem repräsentativen Patientenkollektiv untersucht werden.

Bei der Untersuchung von Screeningprogrammen besteht die Gefahr, dass durch Vorverlagerung der Diagnose eine nur scheinbare Verlängerung der Überlebenszeit beobachtet wird, die jedoch in Wahrheit auf nicht vergleichbare Anfangszeitpunkte zurückzuführen ist (*Lead Time Bias*). Eine verlängerte Überlebenszeit kann ebenso vorgetäuscht werden, wenn eine Screeningmaßnahme bevorzugt milder oder langsamer verlaufende Frühformen einer Krankheit erkennen kann (*Length Bias*). Einen Schutz vor diesen Verzerrungsmechanismen bietet die Durchführung einer randomisierten Studie zur Beurteilung der Wirksamkeit einer Screeningmaßnahme (siehe auch Abschnitt 2.4).

Ein allgemeines Problem bei der Schätzung von Effekten ist eine Verzerrung durch *Messfehler* und *Fehlklassifikation* in den erhobenen Studiendaten [3,4]. In der Praxis sind Messfehler kaum zu vermeiden und es ist bekannt, dass auch nicht differentielle Messfehler zu einem Bias in der Effektschätzung führen können. Im Fall eines einfachen linearen Regressionsmodells mit zufälligem klassischen Messfehler in der erklärenden Variablen kommt es zum *Dilution Bias*, das heißt zu einer Verzerrung des Schätzers in Richtung des Nulleffekts. In anderen Modellen und komplexeren Situationen sind jedoch Verzerrungen in alle Richtungen möglich. Je nach Fragestellung sollte die Stärke möglicher Messfehler diskutiert werden, und bei Bedarf sollten Methoden zur Adjustierung von Messfehlerverzerrungen angewendet werden.

Ein ähnliches Problem stellen fehlende Werte dar. Fehlende Werte, die nicht durch einen Zufallsmechanismus zustande kommen, können ebenso einen Bias im Ergebnis verursachen [5]. Die möglichen Ursachen und Auswirkungen fehlender Werte sind daher im Einzelfall zu diskutieren und bei Bedarf sind statistische Verfahren erforderlich, die einen Bias berücksichtigen beziehungsweise ausgleichen (siehe Abschnitt 1.11).

Im Rahmen von systematischen Übersichten spielt noch der *Publication Bias* eine wichtige Rolle [6]. Da häufig signifikante Ergebnisse eher publiziert werden als nicht signifikante, kommt es bei der Zusammenfassung der publizierten Ergebnisse zu einer systematischen Verzerrung des gemeinsamen Effektschätzers. Zur Entdeckung und Berücksichtigung von Publication Bias können grafische Methoden wie der Funnel Plot [7] und statistische Methoden wie die Meta-Regression eingesetzt werden [8,9,10] (siehe Abschnitt 1.21).

Literatur

- [1] Sackett DL. Bias in analytic research. J Chron Dis 1979; 32: 51-63.

- [2] Feinstein AR. Clinical Epidemiology. The Architecture of Clinical Research. Philadelphia: WB Saunders Co.; 1985.
- [3] Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu, CM. Measurement Error in Nonlinear Models: A Modern Perspective, 2nd Ed. London: Chapman & Hall; 2006.
- [4] Cheng C-L, van Ness JW. Statistical Regression with Measurement Error. London: Arnold; 1999.
- [5] Little RJA, Rubin DB. Statistical Analysis with Missing Data, 2nd Ed. New York: Wiley; 2002.
- [6] Begg CB, Berlin JA. Publication bias: A problem in interpreting medical data (with discussion). J R Stat Soc A 1988; 151: 419-463.
- [7] Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple graphical test. BMJ 1997; 315: 629-634.
- [8] Sterne JAC, Egger M, Davey Smith G. Investigating and dealing with publication and other biases in meta-analysis. BMJ 2001; 323: 101-105.
- [9] Macaskill P, Walter SD, Irwig L. A comparison of methods to detect publication bias in meta-analysis. Stat Med 2001; 20: 641-654.
- [10] Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. JAMA 2006; 295: 676-680.

1.19 Nachweis der Verschiedenheit

Beim empirischen Nachweis, dass sich gewisse Gruppen bezüglich eines bestimmten Merkmals unterscheiden, sind verschiedene Aspekte zu beachten. Zunächst ist festzuhalten, dass der Ausdruck „Nachweis“ nicht als „Beweis“ im mathematischen Sinne zu verstehen ist. Mit Hilfe empirischer Studiendaten lassen sich Aussagen nur unter Zulassung gewisser Irrtumswahrscheinlichkeiten treffen. Durch Anwendung statistischer Methoden lassen sich diese Irrtumswahrscheinlichkeiten jedoch gezielt kontrollieren und minimieren, um auf diese Weise eine Aussage „statistisch nachzuweisen“. Die übliche Methode eines solchen statistischen Nachweises in der medizinischen Forschung ist die Anwendung von Signifikanztests (siehe auch Abschnitt 1.2). Diese Argumentationsebene ist zu trennen von der Beurteilung der klinischen Relevanz eines Unterschieds (siehe auch Abschnitt 1.3). Die Kombination beider Argumente liefert in der Praxis eine adäquate Beschreibung eines Unterschieds aufgrund empirischer Daten.

Bei der Anwendung eines Signifikanztests zum Nachweis eines Unterschieds sind a priori die wissenschaftliche Fragestellung und darauf basierend die Zielvariable, das Effektmaß und die statistische Hypothesenformulierung festzulegen. Damit eine Studie groß genug ist, um einen Unterschied entdecken zu können, ist es notwendig, den benötigten Stichprobenumfang vor Beginn der Studie zu berechnen. Dazu sind in einfachen Situationen neben obigen Informationen zusätzlich die Angabe des klinisch relevanten Unterschieds sowie ein Schätzwert der Variabilität der Zielgröße er-

forderlich. Bei komplexeren Designs oder Fragestellungen sind zusätzlich weitere Angaben wie Korrelationsstruktur, Rekrutierungsschema, Drop-out-Schema usw. erforderlich [1,2].

Zur abschließenden Darstellung der Ergebnisse gehören neben der Signifikanzangabe für eine Aussage auch ein mit adäquaten Methoden berechnetes Konfidenzintervall für das gewählte Effektmaß (siehe auch Abschnitt 1.1), die deskriptive Angabe weiterer Effektmaße um unterschiedliche Aspekte des Ergebnisses darzulegen und eine Diskussion der klinischen Relevanz der Ergebnisse, die auf der Ermittlung patientenrelevanter Zielkriterien aufbauen sollte (siehe auch die Abschnitte 1.3 und 2.1).

Literatur

- [1] Desu MM, Raghavarao D. Sample Size Methodology. Boston: Academic Press; 1990.
- [2] Bock J, Toutenburg H. Sample size determination in clinical research. In: Rao CR, Chakraborty R, editors. Handbook of Statistics Vol. 8. Amsterdam: Elsevier; 1991: 515-538.

1.20 Nachweis der Gleichheit

Einer der häufigen schwerwiegenden Fehler bei der Interpretation medizinischer Daten ist es, ein nicht signifikantes Ergebnis eines traditionellen Signifikanztests als Evidenz für die Richtigkeit der Nullhypothese zu werten [1]. Zum Nachweis einer „Gleichheit“ ist die Anwendung von Methoden zur Untersuchung von Äquivalenzhypothesen erforderlich [2]. Wichtig ist hierbei zunächst das Verständnis, dass eine exakte „Gleichheit“ – also zum Beispiel, dass die Differenz der Mittelwerte zweier Gruppen exakt null beträgt – mit Hilfe statistischer Methoden nicht möglich ist. In der Praxis ist aber auch nicht der Nachweis einer exakten Gleichheit gefordert, sondern der Nachweis eines höchstens irrelevanten Unterschieds zwischen zwei Gruppen. Um dieses Ziel zu erreichen muss logischerweise zunächst definiert werden, was ein irrelevanter Unterschied ist, das heißt, es ist die Festlegung eines Äquivalenzbereichs erforderlich.

Um sinnvoll auf Äquivalenz schließen zu können, müssen – ähnlich wie beim Nachweis eines Unterschieds (siehe auch Abschnitt 1.19) – a priori die wissenschaftliche Fragestellung und darauf basierend die Zielvariable, das Effektmaß und die statistische Hypothesenformulierung festgelegt werden. Zusätzlich ist bei Äquivalenzstudien der Äquivalenzbereich klar zu definieren. Dies kann entweder zweiseitig, was zu einem Äquivalenzintervall führt, oder einseitig im Sinne einer „höchstens irrelevanten Unterschiedlichkeit“ beziehungsweise „höchstens irrelevanten Unterlegenheit“ erfolgen. Im letzteren Fall spricht man von einer „Nichtunterlegenheitshypothese“ [3-5].

Wie bei Überlegenheitsstudien ist es auch in Äquivalenzstudien erforderlich, den benötigten Stichprobenumfang vor Beginn der Studie zu berechnen. Welche Methode hierfür in Frage kommt, hängt von der genauen Hypothese und von der gewählten Analyseverfahren ab [6].

Zur Datenanalyse in Äquivalenzstudien müssen speziell für diesen Zweck entwickelte Methoden verwendet werden. Ein häufig angewendetes Verfahren ist die Konfidenzintervall-Einschluss-Methode. Liegt das berechnete Konfidenzintervall völlig im vorher definierten Äquivalenzbereich, so gilt dies als Nachweis der Äquivalenz. Um das Niveau $\alpha=0,05$ einzuhalten, genügt hierbei die Berechnung eines 90%-Konfidenzintervalls [2].

Im Vergleich zu Überlegenheitsstudien besitzen Äquivalenzstudien spezielle methodische Probleme. Zum einen ist es häufig schwierig, Äquivalenzbereiche sinnvoll zu definieren [7]. Zum anderen schützen zum Beispiel die üblichen Designkriterien Randomisierung und Doppelblindheit nicht mehr hinreichend sicher vor Verzerrungen [8]. Auch ohne Kenntnis der Therapiegruppe ist es zum Beispiel möglich, die Differenz der Behandlungsunterschiede zur Null und damit zur gewünschten Alternativhypothese hin zu verschieben. Des Weiteren ist mit dem Intention-to-Treat-Prinzip sehr vorsichtig umzugehen, da dessen inadäquate Anwendung eine falsche Äquivalenz vortäuschen kann [2]. Somit ist bei der Beurteilung von Äquivalenzstudien besondere Vorsicht geboten.

Literatur

- [1] Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ* 1995; 311: 485.
- [2] Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: The importance of rigorous methods. *BMJ* 1996; 313: 36-39.
- [3] D'Agostino RB, Massaro JM, Sullivan KM. Non-inferiority trials: Design concepts and issues - the encounters of academic consultants in statistics. *Stat Med* 2003; 22: 169-186.
- [4] Röhm J, Hauschke, Koch A, Pigeot. Biometrische Verfahren zum Wirksamkeitsnachweis im Zulassungsverfahren. Nicht-Unterlegenheit in klinischen Studien. *Bundesgesundheitsbl - Gesundheitsforsch - Gesundheitsschutz* 2005;48: 562-571.
- [5] ICH E9 Expert Working Group. ICH Harmonised Tripartite Guideline: Statistical Principles for Clinical Trials. *Stat Med* 1999; 18: 1905-1942.
- [6] Roebuck P, Elze M, Hauschke D, Leverkus F, Kieser M. Literaturübersicht zur Fallzahlplanung für Äquivalenzprobleme. *Inform Biom Epidemiol Med Biol* 1997; 28: 51-63.
- [7] Lange S, Freitag G. Choice of delta: Requirements and reality - results of a systematic review. *Biom J* 2005; 47: 12-27.
- [8] Senn S. Inherent difficulties with active control equivalence studies. *Stat Med* 1993; 12: 2367-2375.

1.21 Meta-Analysen

In der Literatur verwendete Begriffe wie *Literaturübersicht*, *systematische Übersicht*, *Meta-Analyse*, *gepoolte Analyse* oder *Forschungssynthese* sind häufig unterschiedlich definiert und nicht klar voneinander abgegrenzt [1]. Das Institut verwendet folgende Begriffe und Definitionen: Bei einer *nicht systematischen Übersicht* handelt es sich um eine Beschreibung und Bewertung von Studienergebnissen zu einer definierten Thematik ohne eine ausreichend systematische und reproduzierbare Identifikationsmethode der relevanten Forschungsergebnisse zu dieser Thematik. Eine quantitative Zusammenfassung von Daten mehrerer Studien wird als *gepoolte Analyse* bezeichnet. Wegen der fehlenden Systematik und der inhärenten subjektiven Komponente sind Übersichten und Auswertungen, die nicht auf einer systematischen Literaturrecherche basieren, sehr anfällig für Verzerrungen. Eine *systematische Übersicht* beruht auf einer umfassenden, systematischen Vorgehensweise und Studienbewertung, um mögliche Biasquellen zu minimieren. Eine systematische Übersicht kann – muss aber nicht – eine quantitative Zusammenfassung der Studienergebnisse beinhalten. Eine *Meta-Analyse* wird verstanden als eine statistische Zusammenfassung von Ergebnissen mehrerer Studien im Rahmen einer systematischen Übersicht. Eine Meta-Analyse basiert in den meisten Fällen auf aggregierten Studiendaten aus Publikationen. Dabei wird aus den in einzelnen Studien gemessenen Effektstärken, unter Berücksichtigung der Fallzahlen und der Varianzen, ein Gesamteffekt berechnet. Effizientere Auswerteverfahren sind möglich, wenn individuelle Patientendaten aus den Studien verfügbar sind. Unter einer *Meta-Analyse mit individuellen Patientendaten (IPD = individual patient data)* wird die Auswertung von Daten auf Patientenebene im Rahmen eines allgemeinen statistischen Modells mit festen oder zufälligen Effekten verstanden, bei dem die Studie als Effekt und nicht als Beobachtungseinheit eingeht. Unter einer *prospektiven Meta-Analyse* versteht das Institut die a priori geplante statistische Zusammenfassung der Ergebnisse mehrerer prospektiv gemeinsam geplanter Studien. Sollte es zur jeweiligen Fragestellung auch noch andere Studien geben, so müssen diese jedoch auch in der Auswertung berücksichtigt werden um den Charakter einer systematischen Übersicht zu bewahren.

Die übliche Darstellung der Ergebnisse einer Meta-Analyse erfolgt mittels Forest-Plots, in denen die Effektschätzer der einzelnen Studien und des Gesamteffekts inklusive Konfidenzintervallen grafisch aufgetragen werden [2]. Es kommen zum einen Modelle mit festen Effekten zum Einsatz, die (zum Beispiel durch die Inverse der Varianz) gewichtete Mittelwerte der Effektstärken liefern. Es werden zum anderen aber auch häufig Modelle mit zufälligen Effekten gewählt, bei denen eine Schätzung der Varianz zwischen den einzelnen Studien (Heterogenität) berücksichtigt wird. Die Frage, in welchen Situationen welches Modell eingesetzt werden soll, wird seit langem kontrovers diskutiert und muss für jede Analyse einzeln beantwortet werden [3,4]. Wenn es keine eindeutige

Begründung für die Wahl zwischen zufälligen und festen Effekten gibt, wird das Institut immer beide Methoden anwenden und im Fall von divergenten Ergebnissen diese als solche beschreiben.

Bevor eine Meta-Analyse durchgeführt wird, muss zunächst überlegt werden, ob die Zusammenfassung der betrachteten Studien überhaupt sinnvoll ist. Zum einen muss die Vergleichbarkeit der Studien bezüglich der Fragestellung gegeben sein, zum anderen ist die Heterogenität der Studien bezüglich der Ergebnisse zu untersuchen [5]. Für diesen Zweck gibt es spezielle neuere statistische Methoden, wie zum Beispiel das I^2 -Maß [6]. Für dieses Maß existieren Untersuchungen, die eine grobe Klassifizierung der Heterogenität zulassen (zum Beispiel niedrig: $I^2 \leq 25\%$, mittelmäßig: $25\% < I^2 \leq 50\%$, hoch: $I^2 > 50\%$) [7]. Ist die Heterogenität der Studien zu groß, so ist eine statistische Zusammenfassung der Studienergebnisse unter Umständen nicht sinnvoll [8]. In einem solchen Fall wird das Institut keine Meta-Analyse durchführen. In diese Entscheidung müssen jedoch neben statistischen Maßzahlen in jedem Fall auch inhaltliche Gründe einfließen, die nachvollziehbar darzustellen sind. In diesem Zusammenhang spielt auch die Wahl des Effektmaßes eine Rolle. Es kann sein, dass die Wahl eines bestimmten Maßes zu großer Studienheterogenität führt, ein anderes Maß jedoch nicht. Bei binären Daten sind relative Effektmaße häufig stabiler als absolute, da sie nicht so stark vom Basisrisiko abhängen [9]. In solchen Fällen sollte die Datenanalyse über ein relatives Effektmaß erfolgen, zur deskriptiven Darstellung sollten aber absolute Maße für spezifische Basisrisiken hieraus abgeleitet werden.

Bei großer Heterogenität der Studien ist es notwendig, mögliche Ursachen hierfür zu untersuchen. Unter Umständen lassen sich mittels Meta-Regressionen Faktoren finden, die die Heterogenität der Effektstärken erklären können [10,11]. In einer Meta-Regression wird die statistische Assoziation zwischen den Effektstärken der einzelnen Studien und den Studiencharakteristika untersucht, so dass möglicherweise Studiencharakteristika gefunden werden können, die einen Erklärungswert für die unterschiedlichen Effektstärken, also die Heterogenität, haben. Wichtig ist jedoch, dass man bei der Interpretation der Ergebnisse die Einschränkungen solcher Analysen berücksichtigt. Selbst wenn eine Meta-Regression auf randomisierten Studien basiert, kann durch eine Meta-Regression nur die Evidenz einer Beobachtungsassoziation abgeleitet werden, nicht jedoch ein kausaler Zusammenhang [10]. Besonders schwierig zu interpretieren sind Meta-Regressionen, die versuchen, eine Beziehung zwischen den unterschiedlichen Effektstärken und den durchschnittlichen Patientencharakteristika der einzelnen Studien aufzuzeigen. Solche Analysen unterliegen den gleichen Beschränkungen wie die Ergebnisse ökologischer Studien der Epidemiologie [12]. Aufgrund der hohen Anfälligkeit für Verzerrungen, die bei Analysen basierend auf aggregierten Daten auch nicht durch Adjustierung ausgeglichen werden können, sind hier gesicherte Schlussfolgerungen nur auf der Basis individueller Patientendaten möglich [10].

Literatur

- [1] Egger M, Smith GD, Altman DG, editors. Systematic Reviews in Health Care – Meta-analysis in Context. London: BMJ Books; 2001.
- [2] Lewis S, Clarke M. Forest plots: Trying to see the wood and the trees. *BMJ* 2001; 322: 1479-1480.
- [3] Engels EA, Schmid CH, Terrin N, Olkin I, Lau J. Heterogeneity and statistical significance in meta-analysis: An empirical study of 125 meta-analyses. *Stat Med* 2000; 19: 1707-1728.
- [4] Villar J, Mackey ME, Carroli G, Donner A. Meta-analysis in systematic reviews of randomized controlled trials in perinatal medicine: Comparison of fixed and random effects models. *Stat Med* 2001; 20: 3635-3647.
- [5] Glasziou PP, Sanders SL. Investigating causes of heterogeneity in systematic reviews. *Stat Med* 2002; 21: 1503-1511.
- [6] Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; 21: 1539-1558.
- [7] Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327: 557-560.
- [8] Deeks JJ, Higgins JPT, Altman DG. Analysing and presenting results. In: Higgins JPT, Green, S, editors. *Cochrane Handbook for Systematic Reviews of Interventions 4.2.5* [updated May 2005]; Section 8. In: *The Cochrane Library, Issue 3, 2005*. Chichester: Wiley; 2005.
- [9] Furukawa TA, Guyatt GH, Griffith LE. Can we individualize the 'number needed to treat'? An empirical study of summary effect measures in meta-analyses. *Int J Epidemiol* 2002; 31: 72-76.
- [10] Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002; 21: 1559-1573.
- [11] Van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med* 2002; 21: 589-624.
- [12] Greenland S, Morgenstern H. Ecological bias, confounding and effect modification. *Int J Epidemiol* 1989; 18: 269-274.

1.22 Adjustierung und multifaktorielle Verfahren

Vor allem in nicht randomisierten Studien spielen multifaktorielle Verfahren, die es ermöglichen, den Effekt von Confoundern auszugleichen, eine zentrale Rolle (siehe auch Abschnitt 1.18) [1]. Ein weiteres wichtiges Anwendungsgebiet für multifaktorielle Methoden sind Studien mit mehreren Behandlungen [2]. Die Darstellung von Ergebnissen multifaktorieller Verfahren ist jedoch in der medizinischen Fachliteratur leider oft unzureichend [3,4]. Um die Qualität einer solchen Analyse bewerten zu können, sind eine Darstellung der wesentlichen Aspekte im Rahmen der statistischen Modellbildung notwendig [5,6] sowie Angaben zur Güte des gewählten Modells (Goodness-of-Fit) [7]. Die wichtigsten Informationen sind in der Regel hierbei

- eine eindeutige Beschreibung und a priori Festlegung der Zielvariablen und aller potenziell erklärenden Variablen,
- das Messniveau und die Kodierung aller Variablen,
- Angaben zur Selektion von Variablen und Wechselwirkungen,
- eine Beschreibung, wie die Modellannahmen überprüft wurden,
- Angaben zur Modellgüte,
- eine Tabelle mit den wesentlichen Ergebnissen (Parameterschätzung, Standardfehler, Konfidenzintervall) für alle erklärenden Variablen.

Je nach Fragestellung haben diese Informationen eine unterschiedliche Bedeutung. Geht es im Rahmen eines Prognosemodells (siehe Abschnitte 2.8 und 2.9) um eine gute Prädiktion der Zielvariablen, so ist eine hohe Modellgüte wichtiger als bei einem Gruppenvergleich, bei dem man nach wichtigen Confoundern adjustieren muss.

Besonders kritisch ist eine unzureichende Darstellung der Ergebnisse multifaktorieller Verfahren dann, wenn es durch die (nicht ausreichend transparent dargestellte) statistische Modellierung zu einer Verschiebung von Effekten in einen „gewünschten“ Bereich kommt, der bei univariabler Betrachtung so nicht erkennbar ist. Ausführliche Erläuterungen zu den Anforderungen im Umgang mit multifaktoriellen Verfahren sind in verschiedenen Übersichten und Leitlinien in der Literatur zu finden [1,8,9].

In eigenen regressionsanalytischen Berechnungen wird das Institut moderne Verfahren einsetzen [10]. Sehr vorsichtig zu interpretieren sind hierbei Ergebnisse multifaktorieller Modelle, die aus einem Variablenselektionsprozess hervorgegangen sind. Falls bei der Modellwahl solche Variablenselektionsverfahren unumgänglich sind, wird eine Form der Backward-Elimination eingesetzt, da diese dem Verfahren der Forward-Selection vorzuziehen ist [10]. Wichtig ist hierbei eine sorgfältige Vorauswahl der potenziellen erklärenden Variablen unter sachwissenschaftlicher Kenntnis [11]. Bei Bedarf werden auch moderne Verfahren wie die Lasso-Methode eingesetzt [12]. Bei der Modellierung stetiger Einflussgrößen wird das Institut im Bedarfsfall auf flexible Modellansätze wie zum Beispiel das Verfahren der Fractional Polynomials zurückgreifen [13,14], um eine adäquate Beschreibung nicht monotoner Zusammenhänge zu ermöglichen.

Literatur

- [1] Katz MH. Multivariable analysis: A primer for readers of medical research. *N Engl J Med* 2003; 138: 644-650.
- [2] McAlister A, Straus SE, Sackett DL, Altman DG. Analysis and reporting of factorial trials: A systematic review. *JAMA* 2003; 289: 2545-2553.

- [3] Bender R, Grouven U. Logistic regression models used in medical research are poorly presented (Letter). *BMJ* 1996; 313: 628.
- [4] Müllner M, Matthews H, Altman DG. Reporting on statistical methods to adjust for confounding: A cross-sectional survey. *Ann Intern Med* 2002; 136: 122-126.
- [5] Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; 15: 361-387.
- [6] Royston P. A strategy for modelling the effect of a continuous covariate in medicine and epidemiology. *Stat Med* 2000; 19: 1831-1847.
- [7] Hosmer DW, Lemeshow S. The importance of assessing the fit of logistic regression models: A case study. *Am J Public Health* 1991; 81: 1630-1635.
- [8] Bender R, Grouven U. Ordinal logistic regression in medical research. *J R Coll Physic London* 1997; 31: 546-551.
- [9] Bagley SC, White H, Golomb BA. Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain. *J Clin Epidemiol* 2001; 54: 979-985.
- [10] Harrell FE Jr. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer; 2001.
- [11] Derksen S, Keselman HJ. Backward, forward, and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *Br J Math Statist Psych* 1992; 45: 265-282.
- [12] Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc B* 1996; 58: 267-288.
- [13] Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling. *Appl Stat* 1994; 43: 429-467.
- [14] Sauerbrei W, Royston P. Building multivariable prognostic and diagnostic models: Transformation of the predictors by using fractional polynomials. *J R Stat Soc A* 1999; 162: 71-94.

1.23 Bewertung qualitativer Studien

Qualitative Forschungsmethoden werden eingesetzt, um subjektives Erleben und individuelles Handeln sowie die soziale Realität zu erforschen und zu verstehen [1-4]. Bei der Verwendung qualitativer Studien besteht für das Institut das Ziel, Hypothesen für die Arbeit des Instituts zu generieren und die Interpretation von Daten zu unterstützen sowie Einblick in Patientenerfahrungen und Patientensichtweisen zu erhalten.

Quantitative Forschung arbeitet primär mit Zahlen unterschiedlicher Größenordnung und ist von einer starken Standardisierung gekennzeichnet, wobei jedoch auch persönliche und soziale Erfahrungen berücksichtigt werden können. Im Gegensatz dazu stehen bei der qualitativen Forschung subjektive Daten im Vordergrund [1].

Die Hauptaufgabe des Institutes bei der Bewertung von sozialwissenschaftlichen Studien besteht darin, festzustellen, ob das Forschungsdesign, die Studienqualität und die Reliabilität für die untersuchte Fragestellung angemessen sind. Bezüglich der Validität der Kriterien für die Durchführung, die Beurteilung und die Synthese qualitativer Studien gibt es im Vergleich zu anderen Forschungsrichtungen einen schwächeren allgemeinen Konsens [1-5]. Das Institut wird definierte Kriterien zur Bewertung qualitativer Studien im Rahmen systematischer Übersichten und Gesundheitstechnologiebewertungen nutzen [1].

Literatur

- [1] Dixon-Woods M, Agarwal, S., Young B, Sutton A. Integrative approaches to qualitative and quantitative evidence. London: NHS Health Development Agency; 2004.
- [2] Murphy E, Dingwall R, Greatbatch D, Parker S, Watson P. Qualitative research methods in health technology assessment: A review of the literature. *Health Technol Assess* 1998; 2: iii-ix, 1-274.
- [3] McClelland S. Qualitative Research Methods: Their Role in Health Services Research. NHS Management Briefing, National Electronic Library for Health. 00&04, February 2001. <http://libraries.nelh.nhs.uk/healthmanagement> [Zugriff am 22.10.2004].
- [4] Harden A, Garcia J, Oliver S, Rees R, Shepherd J, Brunton G, Oakley A. Applying systematic review methods to studies of people's views: An example from public health research. *J Epidemiol Community Health* 2004; 58: 794-800.
- [5] Thomas J, Harden A, Oakley A, Oliver S, Sutcliffe K, Rees R, Brunton G, Kavanagh J. Integrating qualitative research with trials in systematic reviews. *BMJ* 2004; 328: 1010-1012.

1.24 Einsatz von Konsultationstechniken

Entsprechend den an das Institut herangetragenen Fragen und Aufgaben kann es notwendig sein, verschiedene Konsultationstechniken einzusetzen. Vor dem Hintergrund einer tiefgründigen, aber dennoch schnellen Bearbeitung werden dies üblicherweise Rapid-Appraisal-Methoden, wie zum Beispiel Fokusgruppen, sein. Rapid-Appraisal-Methoden können jedoch verschiedene Stärken und Schwächen aufweisen. Sie können sehr stark in ihrer Repräsentativität und Validität variieren. Deshalb wird sich das Institut beim Einsatz dieser Methoden an existierenden Qualitätskriterien für die Auswahl, Durchführung und Analyse solcher Techniken orientieren. Dabei werden ethische Aspekte insbesondere bei den Verfahren berücksichtigt, bei denen die Möglichkeit besteht, dass Teilnehmer einen Schaden erleiden könnten (zum Beispiel bei Fokusgruppen).

Dem Einsatz von Konsultationstechniken wird eine Suche nach relevanten qualitativen Daten vorausgehen. Die aus der Konsultation resultierenden Ergebnisse werden dann im Kontext mit vorhandenen Ergebnissen relevanter und detaillierterer qualitativer Studien interpretiert.

Folgende Konsultationstechniken können vom Institut eingesetzt werden:

- Interviews mit key informants (Schlüsselinformanten) [1],
- Fokusgruppen [2-4],
- Gruppeninterviews, Gruppenmeetings und Konsultationen [5-7],
- Surveys und Meinungsforschung (einschließlich Online-Umfragen und Feedback-Mechanismen),
- gelegentlich auch Konsensustechniken, wie zum Beispiel Delphi-Techniken [8] und partizipatorische Evaluationen [9].

Das Institut kann ebenfalls Health Impact Assessments erarbeiten [10]. Dabei werden sowohl qualitative als auch quantitative Methoden eingesetzt. Grundlage für den Einsatz von Health Impact Assessments ist eine klare und transparente Vorgehensweise, auch bezüglich möglicher Auswirkungen von Entscheidungen bezüglich der Gleichheit und sozialen Gerechtigkeit.

Die verschiedenen Methoden zur Erfassung von Meinungen und Einstellungen unterscheiden sich bedeutend in ihrer jeweiligen Reliabilität und Validität. Das Institut wird darauf achten, dass die Sichtweisen von benachteiligten Bevölkerungsgruppen angemessen in diesem Prozess berücksichtigt werden.

Das Ressort „Gesundheitsinformation“ verfolgt unter anderem das Ziel, die allgemeinen Gesundheits- und Wissenschaftskennnisse der Bevölkerung (Health Literacy) zu fördern. Dabei soll einerseits das Verständnis von wissenschaftlichen Begriffen in Bezug auf Gesundheit und auf Evidenzbasierung im Gesundheitssystem erhöht sowie andererseits gesellschaftliches Interesse für die Institutsarbeit geweckt werden. Die Öffentlichkeit soll aktiv in diesen Prozess einbezogen werden. Dazu kann das Ressort die Methoden der Konsultation, der Konsensentwicklung und der öffentlichen Entscheidungsfindung [11] einsetzen und weiterentwickeln.

Methoden, die sich auf die Bevölkerungsebene beziehen wie beispielsweise Online-Befragungen, Bürgerjurs [11] und öffentliche Untersuchungen, werden häufig bei Entscheidungen im Rahmen von Ressourcenallokationen im Gesundheitswesen verwendet [11,12]. Dabei haben sich Bürgerjurs als besonders effektiv für die Untersuchung komplexer Themen erwiesen. Einige dieser Methoden können vom Institut adaptiert werden, um die oben beschriebenen Ziele zu erreichen.

Literatur

- [1] USAID Center for Development Information and Evaluation. Conducting key informant interviews. Performance Monitoring and Evaluation Tips 1996; No 2. http://www.usaid.gov/pubs/usaid_eval/pdf_docs/pnabs541.pdf [Zugriff am 22. 10.2004].
- [2] USAID Center for Development Information and Evaluation. Conducting focus group interviews. Performance Monitoring and Evaluation Tips 1996; No 10. http://www.usaid.gov/pubs/usaid_eval/pdf_docs/pnaby233.pdf [Zugriff am 22.10.2004].

- [3] Aylward P. Conducting research with focus groups. Staff Development Session. Flinders University of South Australia.
<http://www.flinders.edu.au/staffdev/courses/research/resources/Focusgroups.pdf>
 [Zugriff am 22.10.2004].
- [4] Dixon-Woods M, Agarwal, S., Young B, Sutton A. Integrative Approaches to Qualitative and Quantitative Evidence. London: NHS Health Development Agency; 2004.
- [5] National Resource Centre for Consumer Participation in Health. Feedback, Participation and Diversity: A Literature Review. Canberra: Commonwealth of Australia, 2000.
<http://www.participateinhealth.org.au> [Zugriff am 22.10.2004].
- [6] National Health and Medical Research Council, Consumers Health Forum of Australia. Statement on Consumer and Community Participation in Research. Canberra: Commonwealth of Australia, 2002.
http://www.nhmrc.gov.au/publications/_files/r22.pdf [Zugriff am 17.11.2006].
- [7] National Resource Centre for Consumer Participation in Health. Methods and Models of Consumer Participation. Melbourne: National Resource Centre for Consumer Participation in Health, 2004.
<http://www.participateinhealth.org.au> [Zugriff am 22.10.2004].
- [8] Liberati A, Sheldon TA, Banta HD. EUR-ASSESS Project Subgroup report on Methodology. Methodological guidance for the conduct of health technology assessment. Int J Tech Assess Health Care, 1997; 13: 186-219.
- [9] USAID Center for Development Information and Evaluation. Conducting a participatory evaluation. Performance Monitoring and Evaluation Tips 1996; No 1.
http://www.usaid.gov/pubs/usaidd_eval/pdf_docs/pnabs539.pdf [Zugriff am 22.10.2004].
- [10] Barnes R, Scott-Samuel A. Health Impact Assessment: A Ten Minute Guide. Liverpool: International Health Impact Assessment Consortium, 2000.
<http://www.ihia.org.uk/hiaguide.html> [Zugriff am 25.10.2004].
- [11] Ryan M, Scott DA, Reeves C, Bate A, van Teijlingen ER, Russell EM, Napper M, Robb CM. Eliciting public preferences for healthcare: A systematic review of techniques. Health Technol Assess, 2001; 5: 1-186.
- [12] Hicks N, Harford J. Summary Report on Consumer Participation in Resource Allocation. Melbourne: National Resource Centre for Consumer Participation in Health; 2000.

1.25 Begutachtung (externes Review)

Zur Absicherung der Qualität von Institutsberichten und anderen Produkten des Instituts werden je nach Fragestellung und Produkt Stellungnahmen von Sachverständigen sowie von Bürgern und Patienten und deren Angehörigen eingeholt. Spezielle Aspekte des externen Reviewprozesses werden in den entsprechenden Kapiteln näher erläutert.

Die Frage, wie effektiv bestimmte Verfahrensweisen beim externen Review sind, wird erst in jüngerer Zeit durch gezielte Studien untersucht. Bislang gibt es allerdings nur wenig aussagekräftige Untersuchungen dazu [1,2]. Insbesondere gibt es nicht ausreichend valide Interventionsstudien. Der Stellenwert der üblichen Verfahren in medizinischen Zeitschriften [2,3], einschließlich der Bewer-

tung durch Konsumenten und Patienten [4], ist nach bislang vorliegenden Untersuchungen noch nicht ausreichend geklärt [3]. Daher wird das Institut sein Reviewsystem evaluieren und aktualisieren.

Literatur

- [1] Rennie D. Editorial peer review: Its development and rationale. In: Godlee F, Jefferson T, editors. Peer Review in Health Sciences, 2nd Ed. London: BMJ Books, 2003; 1-13.
- [2] Fletcher RH, Fletcher SW. The effectiveness of journal peer review. In: Godlee F, Jefferson T, editors. Peer Review in Health Sciences, 2nd Ed. London: BMJ Books, 2003; 62-75.
- [3] Jefferson TO, Rudin M, Brodney FS, Davidoff F. Editorial peer review for improving the quality of reports of biomedical studies. The Cochrane Database of Method Rev 2006, Issue 1. Chichester: Wiley, 2006.
- [4] Bastian H. Non-peer review: Consumer involvement in research review. In: Godlee F, Jefferson T, editors. Peer Review in Health Sciences, 2nd Ed. London: BMJ Books, 2003; 248-262.

2. Spezielle Beurteilung medizinischer und versorgungsrelevanter Inhalte

Die Extraktion von Daten aus den bei der Literaturrecherche als relevant angesehenen Publikationen (siehe auch Abschnitt 4.7) und deren Bewertung werden strukturiert erfolgen und dokumentiert werden. Bei der Dokumentation sollten die vom Institut entwickelten Extraktionsbögen Verwendung finden. Falls diese Extraktionsbögen nicht verwendet werden sollen, muss dies begründet und mit der Projektleiterin oder dem Projektleiter des Instituts abgestimmt sowie ein alternativer Extraktionsbogen entwickelt, abgestimmt und angewendet werden.

2.1 *Nutzenbewertung in der Medizin*

Allgemeiner patientenrelevanter medizinischer Nutzen

Zur Definition des patientenrelevanten medizinischen Nutzens ist zunächst eine Abgrenzung der Begriffe "Notwendigkeit" und "Nutzen" erforderlich. Nach § 27 SGB V ist eine medizinische Maßnahme notwendig, wenn sie eine Krankheit erkennen, heilen, ihre Verschlimmerung verhüten oder Krankheitsbeschwerden lindern kann. Der Begriff der Notwendigkeit geht über den Begriff des Nutzens hinaus. Der Nachweis des Nutzens ist demnach eine notwendige aber keine hinreichende Bedingung für den Nachweis der Notwendigkeit. Die Beurteilung der Evidenz kann darin münden, dass das Vorliegen eines Nutzens (bzw. Schadens) einer Maßnahme belegt ist, das Fehlen eines Nutzens belegt ist oder sein Vorliegen oder Fehlen nicht belegt ist. In Ergänzung zu der Beurteilung der Wirksamkeit und der Sicherheit haben die Berichte des Institutes primär die Beschreibung des Nutzens und des Schadens aller Arten medizinischer Interventionen zum Ziel. Da sich der Nutzen einer Maßnahme auf die Patientin oder den Patienten beziehen soll, beruht diese Bewertung auf Ergebnissen wissenschaftlicher Untersuchungen zur Beeinflussung patientenrelevanter Studienendpunkte. Dabei werden sowohl die beabsichtigten als auch die unbeabsichtigten Effekte der Interventionen berücksichtigt.

Diese Effekte können krankheits- und behandlungsbedingte Veränderungen bzw. die Beeinflussung insbesondere folgender Zielgrößen beinhalten:

1. Mortalität,
2. Morbidität (Beschwerden und Komplikationen),
3. gesundheitsbezogene Lebensqualität,
4. interventions- und erkrankungsbezogener Aufwand,
5. Patientenzufriedenheit.

Eine im Sinne der Patienten positive Veränderung dieser Aspekte wird als direkter patientenrelevanter medizinischer Nutzen definiert, eine negative Veränderung als direkter patientenrelevanter medizinischer Schaden. Wenn möglich, wird das Verhältnis dieser beiden Größen als Nutzen-Schaden-Relation ausgedrückt.

Diagnostische Maßnahmen können indirekt nutzen, indem sie eine notwendige Voraussetzung für therapeutische Interventionen sind, durch die das Erreichen einer Wirkung auf die der unter 1. bis 5. genannten Zielgrößen möglich wird. Darüber hinaus können diagnostische Tests in patientenrelevanter Weise das Treffen persönlicher Lebensentscheidungen ermöglichen.

Die Inanspruchnahme von Leistungen des Gesundheitssystems kann ebenfalls eine patientenrelevante Größe sein. Die direkte Relevanz sollte aber durch eine gleichzeitige Beeinflussung der Behandlungszufriedenheit belegt werden. Außerdem ist zu beschreiben, inwieweit ein Nutzen im Hinblick auf die Inanspruchnahme von Leistungen des Gesundheitssystems möglicherweise mit einem Schaden im Hinblick auf Lebenserwartung, Beschwerden und Komplikationen und/oder die gesundheitsbezogene Lebensqualität, den Behandlungsaufwand und die Patientenzufriedenheit verbunden ist.

Interventionen können auch Auswirkungen auf indirekt Betroffene wie zum Beispiel Familienangehörige und Pflegepersonen haben. Diese Auswirkungen können im Rahmen der Institutsberichte gegebenenfalls auch berücksichtigt werden.

Es werden in erster Linie Endpunkte berücksichtigt, die zuverlässig und direkt konkrete Änderungen des Gesundheitszustandes abbilden. Dabei werden insbesondere einzelne Betroffene, Patientenvertretungs- und/oder Verbraucherorganisationen für die themenbezogene Definition patientenrelevanter Endpunkte einbezogen. Instrumente, die die Lebensqualität oder (andere) so genannte „Patient Reported Outcomes“ (PRO) erfassen, sollten als Voraussetzung für deren Einsatz in klinischen Studien aufgrund einer adäquaten Evaluation geeignet sein [1,2]. Darüber hinaus können für die Nutzenbewertung valide Surrogatendpunkte berücksichtigt werden.

Studiendauer

Ein wesentliches Kriterium bei der Auswahl der für die Nutzenbewertung relevanten Studien ist deren Dauer. Für die Evaluation einer Intervention zur Behandlung einer akuten Erkrankung, deren primäres Ziel z.B. die Verkürzung der Krankheitsdauer und die Reduktion der Akutsymptomatik im Vordergrund steht, ist die Forderung nach Langzeitstudien in der Regel nicht sinnvoll, es sei denn, es sind Spätkomplikationen zu erwarten. Auf der anderen Seite sind für die Evaluation von Interventionen zur Behandlung chronischer Erkrankungen Kurzzeitstudien in der Regel nicht für die vollständige Nutzenbewertung geeignet. Dies trifft insbesondere dann zu, wenn die Behandlung

über mehrere Jahre oder sogar lebenslang erforderlich ist. In solchen Fällen sind insbesondere Studien, die einen mehrjährigen Behandlungszeitraum abdecken, sinnvoll und wünschenswert. Da sowohl Nutzenaspekte als auch das Schadenpotenzial über die Zeit unterschiedlich verteilt sein können, ist bei Langzeittherapien eine aussagekräftige Nutzen-Schaden-Abwägung nur bei einer entsprechenden Studiendauer machbar. Einzelne Aspekte des Nutzens/Schadens können aber durchaus in Kurzzeitstudien betrachtet werden.

Allerdings ist davon auszugehen, dass ein patientenrelevanter Nutzen nie vor Eintritt der Wirksamkeit einer Intervention entstehen kann. Bezüglich des Auswahlkriteriums „Mindeststudiendauer“ orientiert sich das Institut daher primär an allgemein konsentierten Anforderungen bzgl. des Nachweises der Wirksamkeit. Für die Arzneimittelbewertung wird insbesondere auf Angaben in den entsprechenden indikationsspezifischen Leitlinien der Zulassungsbehörden zurückgegriffen (z.B. [3]). Da die Nutzenbewertung auch die Aspekte des Schadenpotenzials beinhaltet, sind bei der Festlegung der Mindeststudiendauer auch die diesbezüglich allgemein konsentierten Anforderungen relevant. Bei Langzeitinterventionen, wie oben beschrieben, wird darüber hinaus auch auf das Kriterium „Langzeitbehandlung“ aus entsprechenden Guidelines zurückgegriffen [4]. Im Einzelfall kann von diesem Vorgehen begründet abgewichen werden, z.B., falls es aus inhaltlichen Gesichtspunkten erforderlich ist, eine Nachbeobachtung über einen längeren Zeitraum zu fordern, spezielle (Unter-)Fragestellungen sich auf einen kürzeren Zeitraum beziehen oder wenn bei der Evaluation neu verfügbarer/zugelassener Interventionen und/oder der Evaluation von Technologien ohne adäquate Behandlungsalternative auch kurzfristige Wirkungen Gegenstand der Evaluation sind.

Surrogate des patientenrelevanten medizinischen Nutzens

Surrogatendpunkte werden in der medizinischen Forschung häufig als Ersatz für patientenrelevante Endpunkte verwendet, meist um Aussagen über den patientenrelevanten Nutzen früher und einfacher zu erhalten [5-7]. Die meisten Surrogatendpunkte sind jedoch in dieser Hinsicht nicht verlässlich und können bei der Nutzenbewertung irreführend sein [8,9]. Surrogatendpunkte werden daher im Rahmen der Nutzenbewertung des Institutes in der Regel nicht als Beleg für den Nutzen einer Intervention herangezogen, es sei denn, eindeutige Belege aus Interventionsstudien für einen plausiblen, starken, konsistenten und gleichgerichteten Zusammenhang zwischen der Änderung des Surrogats und der Änderung des patientenrelevanten Endpunkts sind vorhanden.

Ein gleichgerichteter Zusammenhang bedeutet, dass eine positive oder negative Veränderung des Surrogates mit einer Veränderung des patientenrelevanten Endpunktes konsistent in immer derselben Art und Weise einhergeht.

Die Validität eines Surrogates wird als nicht belegt betrachtet, wenn keine aussagekräftigen Studien zur Modifikation dieses Surrogates in Verbindung mit einer Beschreibung des entsprechenden patientenrelevanten Endpunktes vorliegen. Darüber hinaus wird ein Surrogat als nicht valide betrachtet, wenn in wissenschaftlichen Untersuchungen gezeigt wurde, dass eine Intervention

- einen Einfluss auf den Surrogatendpunkt hatte, nicht jedoch auf den patientenrelevanten Endpunkt, oder
- einen Einfluss auf den patientenrelevanten Endpunkt hatte, nicht jedoch auf den Surrogatendpunkt, oder
- inkonsistente Effekte beim Surrogat- und patientenrelevanten Endpunkt erzeugt.

Surrogatendpunkte von unklarer oder umstrittener Validität können zwar in den Berichten des Institutes dargestellt werden. Derartige Endpunkte sind aber nicht als Belege für den Nachweis des Nutzens einer Intervention geeignet.

Dramatischer Effekt

Ist der Verlauf einer Erkrankung sicher oder nahezu sicher vorhersagbar und bestehen keine Behandlungsoptionen zur Beeinflussung dieses Verlaufs, so kann der Nutzen einer medizinischen Intervention auch durch die Beobachtung einer Umkehr des (quasi-) deterministischen Verlaufs bei einer gut dokumentierten Serie von Patienten belegt werden. Wenn beispielsweise von einer Erkrankung bekannt ist, dass sie nach Diagnosestellung innerhalb kurzer Zeit in sehr vielen Fällen zum Tode führt, und in einer Fallserie beschrieben wird, dass nach Anwendung einer bestimmten Maßnahme die meisten Betroffenen eine längere Zeitspanne überlebt haben, so wird ein solcher „dramatischer Effekt“ ausreichend für einen Nutzenbeleg sein können. Wesentliche Voraussetzung für die Einordnung als „dramatischer Effekt“ ist die ausreichend sichere Dokumentation des schicksalhaften Verlaufs der Erkrankung in der Literatur und der Diagnosestellung bei den Patienten in der Fallserie. Dabei sollten auch mögliche Schäden der Maßnahme berücksichtigt werden.

Nutzen in kleinen Populationen

Es existiert kein überzeugendes Argument, bei kleinen Populationen (z.B. Patienten mit seltenen Erkrankungen oder Untergruppen häufiger Erkrankungen) von der Hierarchie der Evidenzgrade abzuweichen. Auch Patienten mit sehr seltenen Erkrankungen haben einen Anspruch auf eine möglichst sichere Information über Behandlungsoptionen [10]. Nicht randomisierte Studien benötigen aufgrund der Notwendigkeit der Adjustierung nach Störgrößen größere Patientenzahlen als randomisierte Studien. Allerdings kann es aufgrund der Seltenheit einer Erkrankung manchmal unmöglich sein, so viele Patienten in einer Studie einzuschließen, dass diese eine ausreichende statisti-

scher Macht besitzt. Eine meta-analytische Zusammenfassung kleinerer Studien kann in solchen Fällen besonders sinnvoll sein. Im Allgemeinen führen kleinere Stichproben zu einer geringeren Präzision einer Effektschätzung und damit einhergehend zu breiteren Konfidenzintervallen. Aufgrund der Bedeutung des vermuteten Effektes einer Intervention, seiner Größe, der Verfügbarkeit von Alternativen und der Häufigkeit und Schwere potenzieller therapiebedingter Schäden kann es durchaus notwendig sein, bei kleinen Stichproben größere p -Werte als 5% (z.B. 10%) für den Nachweis statistischer Signifikanz zu akzeptieren [11]. Dies muss aber *ex ante* geschehen und nachvollziehbar begründet werden. Ebenso kann bei kleineren Stichproben eher die Notwendigkeit bestehen, einen zu selten auftretenden patientenrelevanten Endpunkt durch Surrogate zu ersetzen. Solche Surrogatparameter müssen allerdings auch bei kleinen Stichproben valide sein [12].

Nutzen im Einzelfall

A) Subjektive Abwägung des potenziellen Nutzens im Vergleich zum potenziellen Schaden

Die Möglichkeit der individuellen Abwägung zwischen dem immer vorhandenen Risiko des Schadens einer medizinischen Intervention und der Chance des Nutzenereignisses setzt voraus, dass zuverlässige und auf den entsprechenden Einzelfall übertragbare allgemeine Daten vorliegen, die die Häufigkeit der Nutzen- und der Schadensereignisse beschreiben und den Grad der Unsicherheit in beiden Bereichen darstellen. Hierbei sollte die Häufigkeit dieser Ereignisse in Form von absoluten Risiken (zum Beispiel als Prozentangabe) für die entsprechende medizinische Intervention und die Alternativen angegeben werden. Die Alternativen sollten dabei gegebenenfalls vorhandene weitere mögliche medizinische Interventionen, aber auch die Möglichkeit des bewussten Nichthandelns beschreiben. Der Begriff des Nichthandelns bezieht sich dabei allein auf die zu betrachtende therapeutische Entscheidung und schließt nicht andersartige unterstützende Interventionen aus. Die Wertung der jeweiligen möglichen Schadens- und Nutzenereignisse in ihrer Größe und der Bedeutung für den Einzelfall ist subjektiv. Somit kann die diesbezügliche subjektive Abwägungsentscheidung nie allgemein richtig oder falsch sein.

B) Beleg des Nutzens im Einzelfall

Der Nachweis des tatsächlich eingetretenen Nutzens im Einzelfall kann immer nur erst *post hoc* geschehen. Da viele medizinische Interventionen sich auf eine Primär- oder Sekundärprävention eines unerwünschten Ereignisses richten, kann hier der Beleg des Nutzens für den Einzelfall nicht als Grundlage der medizinischen Entscheidung dienen. Daher werden die Ergebnisse kontrollierter Interventionsstudien als Grundlage für die Schätzung der Wahrscheinlichkeit des Eintretens derartiger Ereignisse zu Grunde gelegt. Es ist aber möglich, dass solche Studien nicht alle Ereignisse erfassen und/oder dass der Einzelfall von dem Gesamtergebnis abweicht, wenn zum Beispiel ein

Patient bei einer in der Regel lebensrettenden Operation doch verstirbt. Da sich solche Ereignisse für den Einzelfall nicht mit Sicherheit vorhersagen lassen und unumkehrbar sind, wird sich die medizinische Entscheidung in einem solchen Einzelfall auf den Regelfall stützen müssen.

In anderen Fällen, zum Beispiel bei reversiblen Ereignissen, kann im Einzelfall manchmal zuverlässig belegt werden, dass ein bestimmtes Ereignis bei einem bestimmten Menschen durch eine definierte medizinische Intervention hervorgerufen wird. Hierbei kann die medizinische Intervention „versucht“ werden, und die Fortsetzung dieser Maßnahme kann sich nach dem Ergebnis dieser Intervention richten (zum Beispiel bei der Schmerztherapie). Zu beachten sind dabei verschiedene Störgrößen, die die Interpretation des „Versuches“ behindern können, wie zum Beispiel der Effekt der Zeit und der Suggestion.

C) Versuche im Einzelfall

Versuche im Einzelfall sind in der praktischen Medizin häufig, werden aber meist unsystematisch durchgeführt und sind dadurch aufgrund der meist ungenügenden Kontrolle der Störgrößen nicht selten nur unzuverlässig interpretierbar [13].

An systematischen Versuchen im Einzelfall – den so genannten „n-of-1“-Studien – nimmt nur eine einzelne Patientin oder ein einzelner Patient teil, wobei die Zielgrößen vollständig und relativ schnell reversibel sein müssen [13,14]. Eine solche Studie erlaubt, wenn sie möglich und gut durchgeführt ist, eine Aussage darüber, ob eine einzelne Patientin oder ein einzelner Patient von einer bestimmten Behandlung profitiert. Derartige „n-of-1“-Studien bestehen aus mehreren Versuchs- und Kontroll-Studienperioden, die in zufälliger Reihenfolge bei einer Patientin oder einem Patienten angewendet werden [13]. Bei „n-of-1“-Studien werden also nicht die Patienten, sondern die Behandlungsperioden randomisiert. Optimalerweise sollte die Intervention verblindet und entweder aktiv- oder placebo-kontrolliert sein. Leider ist aber in Einzelversuchen die methodische Qualität häufig unzureichend. Die Schlussfolgerungen aus solchen suboptimalen Studiendesigns sind je nach Effektgröße, Stärke und Frequenz von Störgrößen mit erheblichen Fehlern behaftet [12]. Für den Beleg des Nutzens im Einzelfall bedarf die Verwendung eines nicht verblindeten Designs daher einer nachvollziehbaren Begründung.

Zusammenfassende Bewertung

Medizinische Maßnahmen werden vergleichend mit einer anderen klar definierten aktiven Maßnahme oder Scheinmaßnahme (z.B. Placebo) oder mit keiner Maßnahme bezüglich ihrer Auswirkungen auf definierte patientenrelevante Endpunkte in ihrem Nutzen und Schaden zusammenfassend beschrieben. Dafür wird zunächst für jeden vorher definierten patientenrelevanten Endpunkt

einzelnen aufgrund der Analyse vorhandener wissenschaftlicher Daten genau eine der fünf folgenden bewertenden Aussagen getroffen:

- 1) Ein Nutzen (Schaden) liegt vor
- 2) Hinweise liegen vor, dass ein Nutzen (Schaden) vorhanden ist
- 3) kein Nutzen (Schaden) liegt vor
- 4) Hinweise liegen vor, dass kein Nutzen (Schaden) vorhanden ist
- 5) kein Beleg für und kein Hinweis auf einen Nutzen (Schaden) liegen vor.

Eine Maßnahme wird als „1)“ bezeichnet, wenn ein Nutzen oder Schaden wissenschaftlich belegt wurde. Eine Maßnahme wird als „2)“ bezeichnet, wenn Hinweise vorliegen, diese aber nicht eindeutig und/oder einheitlich sind, dass ein Nutzen oder Schaden vorliegt. Eine Maßnahme wird als „3)“ bezeichnet, wenn wissenschaftlich belegt wurde, dass diese Maßnahme keinen Nutzen hat bzw. mit keinem Schaden verbunden ist. Eine Maßnahme wird als „4)“ bezeichnet, wenn Hinweise vorliegen, diese aber nicht eindeutig und/oder einheitlich sind, dass kein Nutzen oder Schaden vorliegt. Eine Maßnahme wird als „5)“ bezeichnet, wenn kein Beleg für und kein Hinweis auf einen Nutzen oder Schaden zum Beispiel wegen unzureichender oder widersprüchlicher Datenlage vorliegen.

Diese für jeden patientenrelevanten Endpunkt einzeln getroffenen Aussagen werden anschließend – soweit möglich – in einem bewertenden Fazit in Form einer Nutzen-Schaden-Abwägung zusammengefasst. Beim Vorhandensein von Belegen eines Nutzens und/oder eines Schadens bzgl. der Punkte 1-5 wird das Institut

- 1) ein Nutzenpotenzial
- 2) ein Schadenpotenzial und
- 3) eine Abwägung zwischen Nutzen- und Schadenpotenzial darstellen

soweit dies auf Grund der vorliegenden Daten möglich ist. Hierbei wird sich das Institut nach dem Prinzip der Risikovorsorge vorgehend, im Zweifel für die Annahme eines Schadenpotenzials entscheiden. Darüber hinaus werden alters-, geschlechts- und lebenslagenspezifische Besonderheiten berücksichtigt. Die genaue Beschreibung der Evaluation der Nutzen-Schaden-Abwägung ist themenspezifisch und sollte – wenn dies prospektiv möglich ist – im Berichtsplan und ansonsten im Vorbericht beschrieben werden.

Literatur

- [1] Food and Drug Administration. Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Online 2006/19/01.
<http://www.fda.gov/cder/guidance/5460dft.pdf> [Zugriff am 03.04.2006].
- [2] European Medicines Agency. Reflection Paper on the regulatory guidance for the use of Health related quality of life (HRQL) measures in the evaluation of medicinal products. Online 2005/07/27.
<http://www.emea.eu.int/pdfs/human/ewp/13939104en.pdf> [Zugriff am 03.04.2006].
- [3] CPMP/EWP/1080/00 Note for Guidance on Clinical Investigation of Medicinal Products in the Treatment of Diabetes Mellitus (CPMP adopted May 2002).
<http://www.emea.eu.int/pdfs/human/ewp/108000en.pdf> [Zugriff am 23.11.2006].
- [4] ICH E1: The Extent of Population Exposure to Assess Clinical Safety for Drugs Intended for Long-Term Treatment of Non-Life-Threatening Conditions.
<http://www.ich.org/LOB/media/MEDIA435.pdf> [Zugriff am 23.11.2006].
- [5] Prentice RL. Surrogate endpoints in clinical trials: Definition and operational criteria. *Stat Med* 1989; 8: 431-440.
- [6] Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med* 1996; 125:605-613.
- [7] ASCO Special Article. Outcomes of cancer treatment for technology assessment and cancer treatment guidelines. *J Clin Oncol* 1996; 14: 671-679.
- [8] Götzsche PC, Liberati A, Torri V, Rossetti L. Beware of surrogate outcome measures. *Int J Tech Assess Health Care* 1996; 12: 238-246.
- [9] Grimes DA, Schulz KF. Surrogate end points in clinical research: Hazardous to your health. *Obstet Gynecol* 2005; 105: 1114-1118.
- [10] Verordnung (EG) Nr. 141/2000 des Europäischen Parlaments und des Rates vom 16. Dezember 1999 über Arzneimittel für seltene Leiden. *Amtsblatt der Europäischen Gemeinschaften* vom 22.01.2000; L18: 1-5.
- [11] Committee for Medicinal Products for Human Use, EMEA. Guideline on the Choice of the Non-inferiority Margin. London, 27 July 2005.
<http://www.emea.eu.int/pdfs/human/ewp/215899en.pdf> [Zugriff am 27.10.2005].
- [12] Committee for Medicinal Products for Human Use, EMEA. Guideline on Clinical Trials in Small Populations. London, 17 March 2005.
<http://www.emea.eu.int/pdfs/human/ewp/8356105en.pdf> [Zugriff am 23.03.2006].
- [13] Schluter PJ, Ware RS. Single patient (n-of-1) trials with binary treatment preference. *Stat Med* 2005; 24: 2625-2636.
- [14] Jull A, Bennett D. Do n-of-1 trials tailor treatment? *Lancet* 2005; 365: 1992-1994.

2.2 *Medikamentöse und nichtmedikamentöse Interventionen*

Die Beurteilung einer Studie zu einer medikamentösen oder nichtmedikamentösen Intervention hat zum Ziel, darzustellen, mit welcher Sicherheit sich ein Effekt oder das Fehlen eines Effekts aus dieser Studie ableiten lässt (Ergebnissicherheit). Darüber hinaus ist es notwendig zu beschreiben,

ob und in welchem Umfang sich das Studienergebnis auf lokale Gegebenheiten (zum Beispiel Population, Versorgungsbereich etc.) übertragen lässt, beziehungsweise welche lokalen Besonderheiten Auswirkung auf die Ergebnisse selbst beziehungsweise deren Interpretation haben oder haben könnten. Unter diesem Gesichtspunkt sind insbesondere Studien relevant, in denen die tatsächliche Versorgungsrealität so weit wie möglich abgebildet wird, ohne dass jedoch die im Folgenden beschriebenen Kriterien der Ergebnissicherheit außer Acht gelassen werden dürfen.

Die Ergebnissicherheit wird im Wesentlichen durch vier Komponenten beeinflusst

- das Studiendesign,
- die designspezifische interne Validität der Studien,
- die Konsistenz der Ergebnisse mehrerer Studien,
- die Größe eines zu erwartenden bzw. beobachteten Effekts.

Die Kriterien zur Beurteilung der internen Validität sind im ersten Kapitel der vorliegenden Methoden an verschiedenen Stellen ausführlich beschrieben. Sie kommen bei der Bewertung von Studien zu medikamentösen und nichtmedikamentösen Interventionen entsprechend zur Anwendung. Dabei können unterschiedliche Aspekte besondere Bedeutung in Abhängigkeit von der zu bearbeitenden Fragestellung haben.

Das Studiendesign hat insofern erheblichen Einfluss auf die Ergebnissicherheit, als dass mit epidemiologischen Studien, prospektiv oder retrospektiv, ein kausaler Zusammenhang zwischen Intervention und Effekt in der Regel nicht dargestellt werden kann, während die experimentelle Forschung grundsätzlich hierzu geeignet ist [1]. Dies gilt zumindest dann, wenn andere, das Ergebnis beeinflussende Faktoren ganz oder weitgehend ausgeschaltet werden, weshalb die randomisierte kontrollierte Studie den Goldstandard bei der Beurteilung medikamentöser und nichtmedikamentöser Interventionen darstellt [2].

Das Institut wird daher nur im Ausnahmefall auf nicht randomisierte Interventionsstudien oder epidemiologische Untersuchungen bei der Effektivitätsbeurteilung zurückgreifen. Diese Ausnahmen müssen begründet sein. Die Begründung kann einerseits in der Nichtdurchführbarkeit eines RCT liegen, z.B. bei starken Präferenzen für eine bestimmte Therapiealternative bei Therapeuten und/oder Patienten, andererseits darin, dass für die vorliegende Fragestellung auch andere Studienarten eine hinreichende Ergebnissicherheit liefern könnten. Bei Erkrankungen, die ohne Intervention nach kurzer Zeit sicher zum Tod führen, können beispielsweise mehrere konsistente Fallbeobachtungen darüber, dass eine bestimmte Intervention diesen zwangsläufigen Verlauf verhindert, hinreichende Ergebnissicherheit liefern [3] (siehe auch Abschnitt 2.1).

Die besondere Begründungspflicht für ein nicht randomisiertes Design bei der Prüfung von Arzneimitteln findet sich auch im Rahmen des Arzneimittelzulassungsrechts in den Arzneimittelprüfrichtlinien (Richtlinie 2001/83/EG, Abschnitt 5.2.5 [4]). Für Medizinprodukte wird abschwächend im Rahmen der Konformitätsbewertung in der EN ISO Norm 14155-2 (Abschnitt 4.7 [5]) zwar nicht die randomisierte kontrollierte Studie als Design der Wahl dargestellt, allerdings muss die Wahl des Designs überhaupt begründet werden. Auch die Verfahrensordnung des G-BA sieht die prioritäre Berücksichtigung von RCTs soweit möglich unabhängig vom Typ (medikamentös/nicht medikamentös) der zu bewertenden medizinischen Maßnahme vor (§20 [6]).

Falls durch Vorrecherchen der wissenschaftlichen Literatur erkennbar ist, dass für eine bestimmte medizinische Indikation keine oder nur wenige Studien höchster Evidenzgrade (RCTs) für die zu bewertende Intervention oder für in Frage kommende Alternativen vorliegen, können in die Berichte des Institutes unter Abwägung von Aufwand und Nutzen im Hinblick auf die resultierende mögliche Ergebnissicherheit auch Studien niedrigerer Evidenz eingeschlossen werden.

Das Institut wird im Rahmen der Erstellung des Berichtsplans (siehe auch Abschnitt 4.4) daher vorab festlegen, welche Studienarten aufgrund der Fragestellung als durchführbar und hinreichend Ergebnissicher (bei hoher interner Validität) angesehen werden. Studien, die diesem Qualitätsanspruch nicht mindestens entsprechen (siehe auch die Abschnitte 1.8 und 1.9), werden in den Bewertungsprozess nicht primär einfließen.

Die Übertragbarkeit des Studienergebnisses selbst muss schließlich in einem gesonderten Prozess überprüft werden, der von Design und Qualität der Studie zunächst unabhängig ist.

Literatur

- [1] Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. *Lancet* 2002; 359: 57-61.
- [2] Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. *The Cochrane Database of Methodology Reviews* 2002, Issue 4. Art. No.: MR000012. DOI: 10.1002/14651858.MR000012.
- [3] Liberati A, Sheldon TA, Banta HD. EUR-ASSESS Project Subgroup report on Methodology. *Methodological guidance for the conduct of health technology assessment. Int J Tech Assess Health Care*, 1997; 13: 186-219.
- [4] Anhang I der Richtlinie 2001/83/EG. Analytische, toxikologisch-pharmakologische und ärztliche oder klinische Vorschriften und Nachweise über Versuche mit Arzneimitteln. *Amtsblatt der Europäischen Union* vom 27.06.2003; L159: 49-94.
- [5] *Klinische Prüfung von Medizinprodukten an Menschen - Teil 2: Klinische Prüfpläne (ISO 14155-2:2003)*; Deutsche Fassung EN ISO 14155-2:2003. Berlin: Beuth; 2003.

- [6] Gemeinsamer Bundesausschuss. Verfahrensordnung des Gemeinsamen Bundesausschusses vom 20. September 2005, veröffentlicht im Bundesanzeiger 2005, S. 16998, in Kraft getreten am 1. Oktober 2005, zuletzt geändert am 18. April 2006, veröffentlicht im Bundesanzeiger 2006, S. 4876, in Kraft getreten am 7. Juli 2006.
<http://www.g-ba.de/cms/upload/pdf/richtlinien/2006-07-07-VerfO.pdf> [Zugriff am 30.11.2006].

2.3 *Diagnostische Verfahren*

Diagnostische Verfahren werden häufig auch heute noch mit Methoden evaluiert, die nicht den Grad der Zuverlässigkeit haben, wie diejenigen, die zur Prüfung von therapeutischen Verfahren eingesetzt werden [1,2]. Somit ist damit zu rechnen, dass nur in Ausnahmefällen die im Folgenden aufgelisteten Informationen zur Bewertung eines diagnostischen Verfahrens komplett vorliegen (siehe z.B. [3]). Wesentliche Basis für die Bewertung ist eine präzise Fragestellung, da Studien zu diagnostischen Verfahren mit unterschiedlichen Zielen durchgeführt werden und je nach Zielsetzung nicht alle Informationen für eine adäquate Bewertung relevant sind [4]. Im Rahmen der Nutzenbewertung eines diagnostischen Verfahrens wird es demnach erforderlich sein, im Berichtsplan (siehe Abschnitt 4.4) genau festzulegen, welche Informationen innerhalb welchen Studiendesigns für die zugrunde liegende Fragestellung erhoben werden sollen. In Einzelfällen kann es beispielsweise ausreichend sein, für die Bewertung der geringfügigen Modifikation eines bereits in seinem Nutzen gesicherten diagnostischen Verfahrens lediglich zu prüfen, ob die Intra-Test-Variabilität gleichwertig oder besser ist (siehe Abschnitt 2.3.2). Soll dagegen ein neues diagnostisches Prinzip in seiner Wertigkeit geprüft werden, sind ggf. vergleichende, randomisierte Studien erforderlich (siehe Abschnitt 2.3.3).

2.3.1 *Allgemeine Aspekte*

Zur Bewertung diagnostischer Verfahren sind insbesondere die folgenden Informationen relevant:

- klare Definition der zu diagnostizierenden/entdeckenden Erkrankung(en) bzw. allgemeiner, des zu entdeckenden Gesundheitsstatus (zum Beispiel Fitness)
- Angaben zur Prävalenz der zu diagnostizierenden/entdeckenden Erkrankung(en) in der zu untersuchenden Population und in Subpopulationen
- eindeutige Definition des „Goldstandard“, das heißt derjenigen Methode, mit der die zu entdeckende Erkrankung (der zu entdeckende Gesundheitsstatus) allgemein akzeptiert eindeutig erfasst werden kann
- genaue Beschreibung des diagnostischen Verfahrens mit Darstellung des materiellen und personellen Aufwands für das Erlernen, die Durchführung und die Auswertung

- Darstellung der Risiken (möglicher Schaden), die mit der Anwendung des diagnostischen Verfahrens verbunden sind, und der Akzeptanz sowie Zumutbarkeit sowohl für die Patienten als auch für das medizinische Personal und für die Allgemeinheit (zum Beispiel Umweltrisiken)
- Angaben zu den weiteren Konsequenzen, die sich aus den jeweiligen Befunden ergeben (zum Beispiel weiterführende Diagnostik, Therapie, andersartige nicht therapeutische Intervention, Monitoring, Lifestyle, Basis für informierte Entscheidungen), mit der Darstellung, inwieweit sich aus diesen Konsequenzen ein Nutzen oder ein Schaden für die Patientin oder den Patienten ergibt
- Angaben zu alternativen diagnostischen Verfahren und gegebenenfalls Darstellung der Vorteile des neuen gegenüber den alten Verfahren.

2.3.2 Testgütekriterien/Testeigenschaften

Nachfolgend sind spezielle Bewertungskriterien aufgelistet.

I. Technische Voraussetzungen

- Angaben zur analytischen Richtigkeit
- Angaben zur analytischen Sensitivität und Spezifität
- Angaben zur Reproduzierbarkeit (Wiederholbarkeit)
- Intratest-/Beurteilervariabilität
 - Intertest-/Beurteilervariabilität
 - Intra-Patient-Variabilität (Kurz- und Langzeit)
- Gegebenenfalls Angaben zur Übereinstimmung (von Messwerten) mit bereits etablierten Standardverfahren
- Angaben zu möglichen Störfaktoren, vor allem systematisch verzerrenden Einflüssen

Bei Studien zu den technischen Voraussetzungen ist darauf zu achten, dass die entsprechenden Kenngrößen (auch) unter klinischen Alltagsbedingungen und in der Anwendungssituation erhoben werden. Beispielsweise reichen in aller Regel Angaben zur Reproduzierbarkeit des diagnostischen Verfahrens allein bei Gesunden nicht aus.

II. Diskriminationsfähigkeit

- Angaben zur diagnostischen Sensitivität und Spezifität, alternativ Angaben zu Likelihood-Quotienten; bei quantitativen Verfahren ROC (Receiver Operating Characteristic) -Kurven mit Angabe eines geeigneten Trennpunkts (beziehungsweise von geeigneten Trennpunkten) und Begründung für dessen (deren) Wahl (Gewichtung von Sensitivität und Spezifität)

Bei Studien zur Diskriminationsfähigkeit lassen sich prinzipiell zwei Vorgehensweisen unterscheiden: zum einen Studien bei ausgewählten Personen mit bekanntem Krankheitsstatus, zum anderen die Prüfung in der Anwendungssituation bei unselektierten Personen und unbekanntem Krankheitsstatus [5-7]. Die erste Vorgehensweise ist in aller Regel (bei entsprechenden Ergebnissen) die Voraussetzung für die Durchführung einer (zumeist aufwändigeren) Studie gemäß der zweiten Vorgehensweise. Studien der ersten Vorgehensweise liefern typischerweise zu optimistische Schätzungen der diskriminatorischen Fähigkeit [8-10], insbesondere wenn schwer erkrankte Patienten und eindeutig gesunde Personen gezielt für den Vergleich ausgewählt werden [11], was bei einer Bewertung berücksichtigt werden sollte.

III. Vorhersage

- Angaben zu prädiktiven Werten (bei quantitativen Verfahren Angaben zu prädiktiven Werten bei gewähltem Trennpunkt [gewählten Trennpunkten])

Folgende methodische Grundprinzipien sollten darüber hinaus bei der Bewertung von Studien zur Evaluierung der Diskriminationsfähigkeit und Vorhersagefähigkeit diagnostischer Verfahren beachtet werden:

- Die Fragestellung und damit verbunden die Studienplanung müssen klar formuliert sein. Dies schließt eine Fallzahlplanung ein, die sich beispielsweise an der gewünschten Präzision der Schätzung (Breite des Konfidenzintervalls) und/oder dem Nachweis des Überschreitens einer Mindestschwelle für bestimmte Kenngrößen (beispielsweise Sensitivität, Spezifität, prädiktive Werte) orientieren kann.
- Die Studien sollten in einer Population von Patienten bzw. Personen, bei denen das Verfahren auch zur Anwendung kommen soll, durchgeführt werden (angemessenes Patientenspektrum beziehungsweise Vermeidung eines Spectrum Bias).
- Zu evaluierendes diagnostisches Verfahren und Goldstandard sollten voneinander unabhängig und verblindet beurteilt werden (wechselseitige Blindheit).
- Die Diagnosesicherung (Goldstandard) beziehungsweise die Art der Diagnosesicherung (unterschiedlicher Goldstandard) sollte nicht vom Ergebnis des zu evaluierenden diagnostischen Verfahrens abhängig gemacht werden (Gefahr des Verification Bias). Falls eine Diagnosesicherung nicht bei allen Patienten durchgeführt werden kann, sollte die Auswahl zufällig erfolgen. Dies löst das Problem allerdings auch nicht vollständig, insbesondere in Situationen mit niedrigen a priori Wahrscheinlichkeiten [12].

- Patienten mit nicht einzuordnenden, unklaren oder nicht interpretierbaren Testergebnissen müssen bei der Auswertung eine angemessene Berücksichtigung finden (kein Ausschluss ohne adäquate Begründung) [13].
- Falls das zu evaluierende diagnostische Verfahren in eine Diagnosestrategie eingebettet ist bzw. werden soll, ist eine isolierte Betrachtung dieses Verfahrens häufig nicht sinnvoll (Problem der Abhängigkeit der Testgütekriterien von der Testkombination).
- Falls das zu evaluierende diagnostische Verfahren Bestandteil des Goldstandards ist, treten ebenfalls besondere methodische Probleme auf, die einer genauen Erörterung und Berücksichtigung bedürfen.

Die Erfahrung zeigt, dass sich diese Prinzipien in publizierten Diagnosestudien häufig nicht wieder finden. Ein Grund dafür ist, dass die Berichterstattung über Diagnosestudien häufig unzureichend ist [10,14]. Es wird demnach notwendig sein, um einerseits diesem Umstand Rechnung zu tragen und andererseits überhaupt eine Aussage treffen zu können, die methodischen Defizite der einzelnen Studien im Zusammenhang mit den darin gefundenen Ergebnissen genau zu beschreiben und besondere Vorsicht bei der eventuellen (statistischen) Zusammenführung der Einzelergebnisse (im Sinne einer Meta-Analyse) walten zu lassen.

Im Hinblick auf eine weitestgehend einheitliche und erschöpfende Darstellung (Publikation) wurde in Analogie zum CONSORT-Statement für Therapiestudien eine ähnliche Empfehlung für Studien zu diagnostischen Verfahren publiziert [15]. Whiting et al. haben eine Checkliste zur Qualitätsbeurteilung von Diagnosestudien im Rahmen von systematischen Übersichten zusammengestellt [16].

2.3.3 Nachweis des Nutzens

Allein aus der Tatsache, dass eine Erkrankung durch ein bestimmtes diagnostisches Verfahren (besonders) gut nachgewiesen oder ausgeschlossen werden kann, lässt sich in der Regel nicht der Nutzen der Anwendung dieses diagnostischen Verfahrens im Sinne einer Verbesserung im Outcome für die Patienten ableiten. Vielmehr ergibt sich der Nutzen aus der resultierenden (zumeist therapeutischen) Konsequenz. Diese Wechselwirkung aus diagnostischer Information und Konsequenz kann, muss aber nicht selbst-evident sein [3]. Wenn Zweifel am Vorliegen einer solchen Wechselwirkung bestehen, sind vergleichende Studien erforderlich, beziehungsweise es werden solche Studien empfohlen [17]. Ähnliches gilt, wenn durch ein neues diagnostisches Verfahren im Vergleich zu einer etablierten Methode mehr beziehungsweise andere Fälle einer zu diagnostizierenden Erkrankung entdeckt werden [4,18].

Studien zum Nutznachweis diagnostischer Tests können z.B. als Vergleich von Patienten, bei denen das diagnostische Verfahren durchgeführt wird, mit solchen, bei denen das Verfahren nicht durchgeführt wird, konzipiert sein; für die Bewertung solcher Studien gelten im Wesentlichen die gleichen Anforderungen wie in Abschnitt 2.4.4 formuliert. Ein Nachteil derartiger Studien ist, dass der Wert der diagnostischen Information nicht von der sich daraus ableitenden Konsequenz getrennt werden kann, d.h. es kann zum Beispiel bei einem negativen Ausgang nicht differenziert werden, ob die diagnostische Information unzureichend ist oder (beispielsweise) die Therapie (bei denjenigen mit pathologischem Testergebnis) unwirksam.

Alternativ zur Durchführung des Tests kann auch die Offenlegung der Testergebnisse geprüft werden, das heißt es werden Personen miteinander verglichen, bei denen das Testergebnis bekannt ist, mit solchen, bei denen das Ergebnis verblindet bleibt [7]. Ein solches Vorgehen hat den Vorteil, den natürlichen Verlauf der Personen mit positivem Testergebnis beurteilen zu können.

In einem weiteren Designvorschlag wird in einer Therapieprüfung das zu evaluierende diagnostische Verfahren bei allen Patienten (unabhängig von der Therapiegruppe) angewendet, das Ergebnis aber für die Dauer der Prüfung für alle Patienten verblindet belassen. In einer solchen Studie lässt sich prüfen, ob die Patienten in Abhängigkeit vom Ergebnis des diagnostischen Tests einen unterschiedlichen therapeutischen Nutzen erfahren [6].

Literatur

- [1] Knottnerus JA, van Weel C, Muris JWM. Evidence base of clinical diagnosis. Evaluation of diagnostic procedures. *BMJ* 2002; 324: 477-480.
- [2] Tatsioni A, Zarin DA, Aronson N, Samson DJ, Flamm CR, Schmid C, Lau J. Challenges in systematic reviews of diagnostic technologies. *Ann Intern Med* 2005; 142: 1048-1055.
- [3] Foerster V, Murtagh J, Lentle BC, Wood RJ, Reed MH, Husereau D, Mensinkai S. CT and MRI for selected clinical disorders: A systematic review of clinical systematic reviews [Technology report no 59]. Ottawa: Canadian Coordinating Office for Health Technology Assessment; 2005.
- [4] Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: Assessing new tests against existing diagnostic pathways. *BMJ* 2006; 332: 1089-1092.
- [5] Köbberling J, Trampisch HJ, Windeler J. Memorandum for the Evaluation of Diagnostic Measures. *J Clin Chem Clin Biochem* 1990; 20: 873-879.
- [6] Richter K, Lange S. Methoden der Diagnoseevaluierung. *Internist* 1997; 38: 325-336.
- [7] Sackett DL, Haynes RB. Evidence base of clinical diagnosis. The architecture of diagnostic research. *BMJ* 2002; 324: 539-541.
- [8] Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, van der Meulen JHP, Bossuyt PMM. Empirical Evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999; 282: 1061-1066.

- [9] Whiting P, Rutjes AW, Dinnes J, Reitsma J, Bossuyt PM, Kleijnen J. Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technol Assess* 2004; 8: iii, 1-234.
- [10] Westwood ME, Whiting PF, Kleijnen J. How does study quality affect the results of a diagnostic meta-analysis. *BMC Med Res Methodol* 2005; 5: 20.
- [11] Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 2006; 174: 469-476.
- [12] Rückmann A, Windeler J. Selektionsbias bei der Schätzung der Sensitivität von Screeningmaßnahmen. In: Trampisch HJ, Lange S (Hrsg.). *Medizinische Forschung – Ärztliches Handeln*. München: MMV Medizin Verlag; 1995. p. 227-231.
- [13] Simel DL, Feussner JR, DeLong ER, Matchar DB. Intermediate, indeterminate and uninterpretable diagnostic test results. *Med Decis Making* 1987; 7:107-114.
- [14] Smidt N, Rutjes AW, van der Windt DA, Ostelo AW, Reitsma JB, Bossuyt PM, Bouter LM, de Vet HC. Quality of reporting of diagnostic accuracy studies. *Radiology* 2005; 235: 347-353.
- [15] Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HC. Standards for Reporting of Diagnostic Accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy. The STARD initiative. *Radiology* 2003; 226: 24-28.
- [16] Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: A tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; 3: 25.
- [17] FDA. Guidance for Industry. Developing Medical and Imaging Drug and Biological Products. Part 2: Clinical Indications. <http://www.fda.gov/cder/guidance/5742prt2.pdf> [Zugriff am 01.12.2006].
- [18] Lord SJ, Irwig L, Simes RJ. When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials. *Ann Intern Med* 2006; 144: 850-855.

2.4 Screening

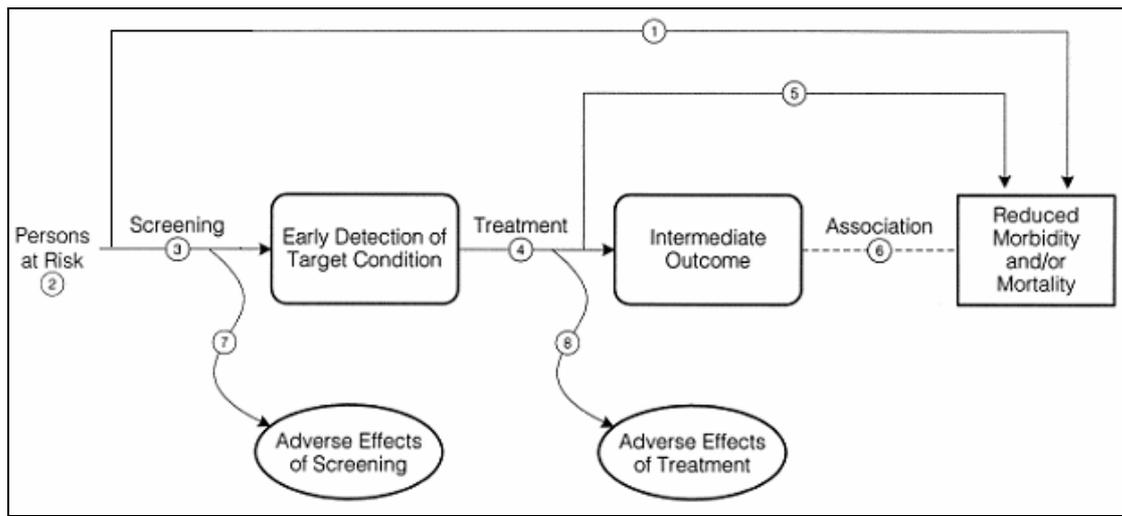
Screeningprogramme setzen sich aus unterschiedlichen „Bausteinen“ zusammen, die entweder als Ganzes oder in Teilen untersucht werden (können). Bei der Bewertung erfolgt eine Orientierung an bereits etablierten und publizierten Kriterien, wie zum Beispiel diejenigen des UK National Screening Committee (UK NSC [1]) und der US Preventive Services Task Force (US PSTF [2]).

Diese Bewertungskriterien umfassen (1.) die Erkrankung, die es zu entdecken gilt, (2.) den einzusetzenden (diagnostischen) Screeningtest selbst, (3.) die Therapie im Falle eines positiven (pathologischen) Ergebnisses des Screeningtests beziehungsweise eine andersartige Konsequenz, die aus einem positiven Ergebnis gezogen wird, sowie (4.) das Screeningprogramm als Ganzes.

Bei der Bewertung sollte differenziert werden zwischen Programmen, bei denen aus dem positiven Ergebnis des Screeningtests eine therapeutische oder eine anderweitige, nicht-therapeutische Kon-

sequenz folgt. Weiterhin sollte unterschieden werden zwischen Situationen, bei denen eine direkte Evidenz für die Wirksamkeit des Screeningprogramms vorliegt (Vergleich von Personen mit und ohne Screening mit einem patientenrelevanten Zielkriterium im Rahmen einer Studie, Pfeil 1 in Abbildung 1), und Situationen, bei denen die Evidenz über Analogieschlüsse indirekt erhoben wird (mehrere „Screeningbausteine“ werden in unterschiedlichen Studien untersucht, zum Beispiel Pfeile 3, 4 und 6 in Abbildung 1).

Abbildung 1: Screeningkette



modifiziert nach [2]

2.4.1 Erkrankung

Es soll bewertet werden, ob es sich um ein bedeutsames Gesundheitsproblem handelt, wobei sich die Beurteilung dieses Aspekts auf verschiedene Indikatoren beziehen kann – z.B. auf die Häufigkeit, die Schwere oder die Kosten einer Erkrankung – bzw. auf verschiedene Ebenen, z.B. auf die Bevölkerungs- oder eine individuelle Ebene.

Damit verbunden ist die Notwendigkeit, die Epidemiologie und den natürlichen Verlauf der Erkrankung genau zu kennen. Datenquellen dafür entstammen typischerweise epidemiologischen Querschnitt-, Register- und Kohortenstudien, in Ausnahmefällen auch Fallserien sowie ökonomischen Studien.

2.4.2 Screeningtest

Es gelten die allgemeinen Anforderungen für die Beurteilung eines diagnostischen Testverfahrens, wie in den Abschnitten 2.3.1 und 2.3.2 formuliert. Ein wichtiger Unterschied zwischen Diagnose und Screening ist aber, dass sich Screeningangebote (zumeist) an gesunde und beschwerdefreie Personen richten, von denen nur wenige die gesuchte Erkrankung aufweisen. Aufgrund der beson-

deren ethischen Implikationen werden in aller Regel höhere Anforderungen an die Testgütekriterien und die Qualität der zu Grunde liegenden Studien zu stellen sein [3]. Darüber hinaus sollte der Test einfach handhabbar sein, und es ist zu prüfen, ob für den Fall eines positiven Testergebnisses eine allgemein akzeptierte Strategie zur weiteren diagnostischen Abklärung (Goldstandard) sowie zu den verfügbaren Alternativen vorliegt.

2.4.3 Therapie

Es ist zu prüfen, ob für Patienten mit positivem Testergebnis und gegebenenfalls nachfolgender Bestätigung im Rahmen einer weiteren diagnostischen Abklärung (Goldstandard) eine wirksame Behandlung oder Intervention existiert (für eine diesbezügliche Bewertung siehe auch Abschnitt 2.2) und ob es Evidenz dafür gibt, dass eine frühzeitige Behandlung zu besseren Ergebnissen führt als eine späte Behandlung. Bei Programmen, bei denen aus dem positiven Ergebnis des Screeningtests keine unmittelbare therapeutische Konsequenz folgt, muss beurteilt werden, ob die Information, die aus dem positiven Testergebnis resultiert, mit einem anderweitigen (nicht-therapeutischen) Nutzen verbunden ist, zum Beispiel dergestalt, dass die Betroffenen in die Lage versetzt werden, besser informiert persönliche (Lebens)Entscheidungen zu treffen (zum Beispiel pränatales Screening auf Down-Syndrom, Screening auf Genträger für unheilbare/nicht beeinflussbare Erkrankungen). Für eine solche Beurteilung kann es gegebenenfalls sinnvoll sein, Methoden der Entscheidungsanalyse zu verwenden.

2.4.4 Screeningprogramm

Idealerweise liegt Evidenz dafür vor, dass das Screeningprogramm (als Ganzes) wirksam ist im Hinblick auf eine Reduktion von Morbidität und/oder Mortalität. Zur Beurteilung entsprechender Studien werden die im Abschnitt 2.2 formulierten Kriterien herangezogen. Hierbei ist zu beachten, dass Evidenz aus nicht randomisierten Studien zur Beurteilung von Screeningprogrammen besonders kritisch zu werten ist, da hier spezielle Verzerrungsmechanismen greifen können wie der Lead-Time Bias oder der Length Bias (siehe auch Abschnitt 1.18).

Liegt keine direkte Evidenz für die Wirksamkeit des Screeningprogramms vor, sondern lediglich Evidenz für die einzelnen Screeningbausteine, muss neben der Beurteilung der einzelnen Bausteine noch zusätzlich eine Beurteilung der Kohärenz und Konsistenz erfolgen. Kohärenz meint in diesem Zusammenhang, dass sich die Bausteine in einem verständlichen Modell zusammenfügen, Konsistenz bedeutet, dass verschiedene Studien beziehungsweise deren Ergebnisse unter unterschiedlichen Bedingungen zur Kohärenz beitragen [2].

Das Screeningprogramm muss einen Nettonutzen erzielen, das heißt der Nutzen, der durch ein Screeningprogramm erreicht werden kann, muss den möglichen physischen oder psychischen

Schaden (hervorgerufen durch den Screeningtest, nachfolgende diagnostische Maßnahmen und/oder die nachfolgende Therapie) übersteigen (zur Bewertung von „unerwünschten Wirkungen“ einer Intervention siehe auch Abschnitt 1.6).

Falls das Screeningprogramm beziehungsweise seine Bausteine nicht in dem Setting geprüft wurden, in dem das Programm zur Anwendung kommen soll, muss geprüft werden, ob Evidenz dafür vorliegt, dass die Ergebnisse verallgemeinerbar beziehungsweise übertragbar sind.

Literatur

- [1] UK National Screening Committee. Criteria for Appraising the Viability, Effectiveness and Appropriateness of a Screening Programme. <http://www.nsc.nhs.uk/pdfs/criteria.pdf> [Zugriff am 28.10.2004].
- [2] Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, Atkins D, Methods Work Group, Third US Preventive Services Task Force. Current methods of the US Preventive Services Task Force: A review of the process. *Am J Prev Med* 2001; 20 (Suppl 3): 21-35.
- [3] Ewart RM. Primum non nocere and the quality of evidence: Rethinking the ethics of screening. *J Am Board Fam Pract* 2000; 13: 188-196.

2.5 Gesundheitsökonomie

2.5.1 Hintergrund

Gemäß § 139a Abs. 3 SGB V nimmt das IQWiG zu Fragen von grundsätzlicher Bedeutung für die Qualität und Wirtschaftlichkeit der im Rahmen der gesetzlichen Krankenversicherung erbrachten Leistungen Stellung. Zur Aufgabe des Ressorts Gesundheitsökonomie gehören damit zum einen die Bearbeitung gesundheitsökonomischer Fragestellungen, die durch die Beauftragung durch den G-BA oder das BMG an das IQWiG herangetragen sowie im Rahmen des Generalauftrags entwickelt werden. Das Institut führt nach derzeitiger Gesetzeslage keine Kosten-Nutzen-Bewertungen von Arzneimitteln durch.

2.5.2 Ziel gesundheitsökonomischer Evaluationen

Ziel gesundheitsökonomischer Evaluationen ist die wissenschaftlich begründete Entscheidungshilfe über die Allokation von Gesundheitsleistungen im Sinne einer „evidence-based health policy“. Die Verbesserung der Gesundheitsversorgung durch eine optimierte Mittelverwendung ist Ziel dieser Entscheidungsprozesse.

2.5.3 Methodik

Damit gesundheitsökonomische Evaluationsstudien valide interpretiert werden können, müssen die Studien eine ausreichende Transparenz und methodische Qualität aufweisen. Die methodische Qualität gesundheitsökonomischer Studien wird durch die Einhaltung unabdingbarer Mindeststandards erreicht. Diese Standards orientieren sich an den international konsentierten, wissenschaftlichen Kriterien des *US Public Health Service Panel on Cost-Effectiveness in Health and Medicine* [1], der *BMJ Economic Evaluation Working Party* [2], der *Policy on Cost-Effectiveness Analyses des New England Journals of Medicine* [3] sowie der *Deutschen Empfehlungen zur gesundheitsökonomischen Evaluation* [4]. Gesundheitsökonomische Evaluationen, die diesen Qualitätskriterien nicht genügen, finden keine Berücksichtigung.

Darüber hinaus müssen gesundheitsökonomische Evaluationen ihre methodischen Ansätze am Kontext der Gesundheitsversorgung in Deutschland ausrichten und die Besonderheiten des deutschen Gesundheitswesens berücksichtigen.

Übertragbarkeit

Die Übertragbarkeit der gesundheitsökonomischen Ergebnisse auf die Rahmenbedingungen des deutschen Gesundheitssystems muss gegeben sein. Die Übertragbarkeit von Studienergebnissen aus einem anderen Land auf Deutschland wird insbesondere durch die Demographie des betrachteten Landes und die Epidemiologie des untersuchten Krankheitsbildes (Inzidenz, Prävalenz), die Verfügbarkeit von Gesundheitsressourcen (Angebot an Therapiemaßnahmen und Gesundheitseinrichtungen, Zugangswege für Patienten) und Variationen in der klinischen Praxis, die Vergütung von Leistungserbringern (Anreizsysteme) und Organisationsstrukturen, die relativen Preise und Kosten und die Wertevorstellungen der betrachteten Population beeinflusst [5,6].

Die Einschätzung der Übertragbarkeit von Ergebnissen von Studien, die in anderen Ländern durchgeführt wurden, setzt voraus, dass die angewendeten Methoden, Datengrundlagen und die Studienergebnisse transparent und nachvollziehbar beschrieben sind [7,8]. Bevor die Übertragbarkeit von Studienergebnissen geprüft wird, sollten folgende Fragen mit „ja“ beantwortet werden können [8]:

1. Ist die betrachtete Technologie vergleichbar mit der in Deutschland verwendeten Technologie?
2. Ist die betrachtete Vergleichsintervention für Deutschland relevant?
3. Hat die Studie eine akzeptable Qualität, d.h. erfüllt sie die geforderten internationalen Standards?

Kann eine dieser Fragen nicht mit „ja“ beantwortet werden, ist die Übertragbarkeit der Studienergebnisse nicht gegeben.

Perspektive

Die darzustellenden Kosten, deren Erhebung sowie die resultierenden Ergebnisse hängen von der jeweilig eingenommenen Perspektive ab [9]. Gesundheitsökonomische Studien können aus den unterschiedlichsten Perspektiven – z.B. derjenigen der Gesellschaft, der Kostenträger, der Leistungserbringer und Patienten – durchgeführt werden. Gemäß seines Auftrags nach § 139a Abs. 3 SGB V wird das IQWiG zu Fragen von grundsätzlicher Bedeutung für die Qualität und die Wirtschaftlichkeit der im Rahmen der GKV erbrachten Leistungen tätig. Für die gesetzliche Aufgabengestaltung des IQWiG nimmt die Kostenbetrachtung aus der Perspektive der gesetzlichen Krankenversicherung (GKV) deshalb eine besondere Stellung ein [9].

Über die GKV-Perspektive hinaus wird das Institut in Abhängigkeit von der Fragestellung die für eine Bewertung der Wirtschaftlichkeit geeignete Perspektive wählen. So kann es z.B. sinnvoll sein, bei der Bewertung von Interventionen, die sich auf Rehabilitation, Pflegebedürftigkeit oder Arbeitsunfähigkeit auswirken, von der Perspektive des Sozialversicherungsträgers oder von einer gesamtgesellschaftlichen Perspektive auszugehen.

Definition des Entscheidungsproblems

Bei einer Wirtschaftlichkeitsuntersuchung muss primär das Ziel der Analyse spezifiziert werden. Dies beinhaltet die Definition der Maßnahme, für die eine Wirtschaftlichkeitsanalyse durchgeführt werden soll, sowie der für die Maßnahme relevanten Patientengruppen und Indikation. Des Weiteren muss beschrieben sein, für welche Region und welchen Entscheidungsträger die Analyse durchgeführt wird [10].

Vergleichsstandards

Im Rahmen gesundheitsökonomischer Evaluationen wird ein Vergleich zweier Interventionen angestrebt, wovon eine die zu bewertende Intervention, die andere die Vergleichsintervention darstellt. Bei der Auswahl der Vergleichsmethode ist zu beachten, dass diese dem derzeit angewandten Versorgungsstandard in der Praxis entspricht [11,12], im deutschen Gesundheitswesen verbreitet und den Beziehern von Gesundheitsleistungen zugänglich ist. Bei mehreren möglichen vergleichbaren Behandlungsverfahren sollten alle genannt werden und mindestens eins davon in der Evaluation als Vergleichsintervention herangezogen werden [13]. Die Auswahl der Vergleichsintervention soll ausreichend begründet und diese exakt beschrieben werden.

Zielpopulation

Die Zielpopulation, die von der betrachteten Intervention profitiert, sollte klar beschrieben werden. So müssen bestimmte Variablen angegeben werden, die die Studienpopulation ausreichend charak-

terisieren, wie z. B. Alter, Geschlecht, sozio-ökonomischer Status, Vorerkrankungen, Risikoprofil etc. [10]. Die Studienpopulation muss repräsentativ für die Zielpopulation in Deutschland sein.

Datenquellen und Studiendesign

Hohe Priorität haben Evaluationen, deren Daten innerhalb von klinischen Studien erhoben werden (*trial-based economic evaluation*), wobei deren Limitierungen hinsichtlich der ökonomischen Aussagen Rechnung getragen wird [6]:

- Die Vergleichstherapie ist nicht notwendigerweise die für gesundheitsökonomische Entscheidungen relevante Vergleichstherapie.
- Die Studienkollektive sind durch die im Studienprotokoll festgelegten Ein- und Ausschlusskriterien einer starken Selektion unterworfen.
- Der Zeithorizont ist meist kurzfristig angelegt, so dass gesundheitsökonomische Langzeitfolgen (ambulante und stationäre Pflege- und Rehabilitationsmaßnahmen, wiederholte Krankenhauseinweisungen), die im Anschluss an die Akutphase und sektorenübergreifend auftreten, nicht hinreichend erfasst werden.
- Durch im Studienprotokoll vorgeschriebene medizinische Leistungen können erhöhte Ressourcenverbräuche entstehen, die sich von der Versorgung unter Alltagsbedingungen erheblich unterscheiden (*protocol-driven costs*).

In Ergänzung zu randomisierten kontrollierten Studien sind deshalb Studiendesigns notwendig, die die Übertragung der Ergebnisse auf die Versorgungsrealität sicherstellen [14]. Diese Anforderungen werden im Rahmen von so genannten pragmatischen Studien (*pragmatic trials*) erfüllt. In pragmatischen Studien werden ebenfalls die Konzepte zu Minimierung von Bias verfolgt, es werden jedoch weniger Restriktionen in Bezug auf die Rekrutierung der Patienten und den Studienverlauf gestellt. Folgende Anforderungen sollen beim Design von pragmatischen Studien erfüllt sein [6]:

1. Es werden Patienten eingeschlossen, die die typische Fallschwere aufweisen.
2. Die Vergleichsintervention entspricht dem derzeit angewandten Versorgungsstandard in der Praxis.
3. Das Setting und die beteiligten Ärzte entsprechen der Versorgungsrealität.
4. Alle eingeschlossenen Patienten werden unter Routinebedingungen behandelt.
5. Eine große Bandbreite an Endpunkten wird gemessen (Wirksamkeit, Ressourcenverbrauch, Lebensqualität, Kosten etc.).

Zur Durchführung derartiger Langzeitstudien eignen sich epidemiologische Verlaufsstudien und Register.

Modellierung

Entscheidungsanalytische Modelle sind nicht als Substitut zur Erlangung verlässlicher Evidenz für den Nutznachweis zu betrachten [1,14,15]. Klinische Studien stellen jedoch nicht immer alle notwendigen Daten zur Verfügung, die für eine vollständige ökonomische Evaluation benötigt werden. In diesen Situationen, die jeweils spezifisch beurteilt werden müssen, kann eine Modellierung zur Unterstützung des Entscheidungsprozesses nützlich sein [16]. Dazu gehören z. B. die Analyse von Langzeitkosten von Interventionen oder Einsparungen durch die Verhinderung von Ereignissen [17]. Mögliche Fehler, die durch die Extrapolation von Daten aus klinischen Studien mit zu kurzem Zeithorizont bedingt sind, sind dabei zu beachten [1]. Modellierungen zur Extrapolation von Daten sollten nur durchgeführt werden, wenn Daten ausreichender Qualität für die jeweilige Fragestellung vorliegen [14]. Die Entscheidung der Nutzung eines Modells zur Beurteilung der Wirtschaftlichkeit von Interventionen muss ausreichend begründet sein. Des Weiteren müssen sich entscheidungsanalytische Modelle folgenden Qualitätsstandards stellen [2,3,18]:

1. Entscheidungsmodelle müssen transparent und nachvollziehbar beschrieben sein (vgl. dazu „*A suggested checklist for assessing quality in decision analytic models*“ [19]).
2. Die den Entscheidungsmodellen zu Grunde liegenden Annahmen müssen begründet werden
3. Die verwendeten Daten müssen den Anforderungen entsprechen (s.o.).
4. Die Wahl der Variablen für die Sensitivitätsanalyse, der Bereich, in dem die Variablen variiert werden und die Ergebnisse der Sensitivitätsanalyse sind transparent und nachvollziehbar beschrieben.
5. Sollten verschiedene Modelle zur gleichen Fragestellung zu unterschiedlichen Aussagen kommen, müssen die Gründe für diese Unterschiede von den Modellerstellern erklärt werden können (*Cross-Validation*).
6. Entscheidungsmodelle sollten validiert werden: Die mathematischen Berechnungen müssen auf ihre Richtigkeit und Konsistenz mit den Modellspezifikationen hin geprüft werden und es muss sichergestellt sein, dass die Modell-Inputdaten und die Outcomes konsistent mit den verfügbaren Daten sind.

Nutzwerte

In der gesundheitsökonomischen Evaluationsform der Kosten-Nutzwert-Analyse (*Cost-Utility Analysis*) wird der Nutzen der zu bewertenden Alternativen als Produkt der gesundheitsbezogenen Lebensqualität und der Lebensjahre, die man in dieser Lebensqualität verbringt, ausgedrückt. Die

gesundheitsbezogene Lebensqualität wird bewertet mit Hilfe von Nutzwerten (*Utilities*), die sich auf einer auf einer Skala von 0 (Tod) bis 1 (perfekte Gesundheit) befinden. Das Produkt aus dem Gewinn an Lebensqualität und den damit verbundenen Lebensjahren wird in der Anzahl der gewonnenen qualitätsadjustierten Lebensjahre (QALY- *Quality Adjusted Life Years*) dargestellt. Ein QALY ist äquivalent zu einem Lebensjahr in perfekter Gesundheit.

Dieser zunächst plausibel anmutende Ansatz erweist sich bei eingehender Betrachtung als nicht widerspruchsfrei und hat bisher für eine anhaltende Debatte über seine theoretischen und praktischen Probleme gesorgt. Die Methoden zur Messung des Gesundheitszustandes sowie der Generierung von Nutzwerten divergieren erheblich. Die Methoden liefern in Abhängigkeit vom eingesetzten Bewertungsverfahren (*Standard Gamble, Time Trade-Off, Visual Analogue Scale*) sowie der Referenzpopulation jeweils verschiedene Ergebnisse bei der Bewertung eines Gesundheitszustandes. Damit stellt sich die grundsätzliche Frage nach der Vergleichbarkeit. Es gibt bisher keine verlässliche Methode, um die Ergebnisse der verschiedenen Verfahren gleichzustellen [20,21]. Das Problem der Vergleichbarkeit wird durch die Ergebnisse einer jüngst in Deutschland erschienenen Publikation verdeutlicht [22]. In einem systematischen Review von 18 in Deutschland durchgeführten Kosten-Nutzwert-Analysen konnten Schwappach et al. zeigen, dass QALYs nicht universell einsetzbar und vergleichbar sind. Des Weiteren spiegeln die in der Mehrzahl der Studien angesetzten Nutzenwerte nicht die Präferenzen von Patienten in Deutschland oder der deutschen Bevölkerung wider [22].

Der Kosten-Nutzwert-Analyse liegt die Annahme zu Grunde, dass es für die Gesellschaft rational ist, die Summe des individuellen Nutzens - ausgedrückt in der Aggregation individueller QALYs - von Interventionen bei einem gegebenen Budget zu maximieren. Studien zeigen, dass die Verteilung der Ressourcen auf Basis aggregierter individueller QALYs von der Bevölkerung nicht akzeptiert wird [23-25]; insbesondere gibt es den Wunsch, Ressourcen gerecht zu verteilen und Ungleichheiten zu reduzieren [23,24,26]. Das QALY-Konzept steht zudem in der Kritik, betagte, behinderte und chronisch kranke Menschen zu benachteiligen [24,27]. Damit werden unmittelbar Fragen der Verteilungs- und Zugangsgerechtigkeit als zentrale Wertvorstellungen unserer Gesellschaftsordnung berührt [23].

Die Verwendung von QALYs als ein „virtuelles“ Einheitsmaß zur Beschreibung gesundheitlicher Ergebnisse ist als unzulänglich zu bewerten und sollte nicht als Entscheidungsgrundlage herangezogen werden. Auch Institutionen, die QALYs als ein Maß der Kosten-Nutzen-Bewertung verwenden, sind sich deren methodischen Unzulänglichkeiten bewusst [28].

Studientyp

Wenn mehrere alternative Vorgehensweisen im Hinblick auf einen standardisierten Behandlungserfolg vorliegen, sollte eine Kosten-Wirksamkeits-Analyse (*Cost-Effectiveness Analysis*) durchgeführt werden. Kosten-Minimierungs-Analysen (*Cost-Minimization Analysis*) sollten durchgeführt werden, wenn gezeigt werden kann, dass die betrachteten Alternativen zu den gleichen medizinischen Ergebnissen führen [13].

Kosten

Abhängig von der Fragestellung sollten alle für die Wirtschaftlichkeitsanalyse relevanten Kosten identifiziert werden. Der Ressourcenverbrauch sollte sich an der klinischen Praxis in Deutschland orientieren [29].

Der Ressourcenverbrauch (z. B. stationäre Verweildauer, Inanspruchnahme von Heil-/Hilfsmitteln, Arztbesuche etc.) sollte getrennt von den dazugehörigen Kosten pro Einheit dargestellt werden [2]. Die Methoden der Ermittlung der Kosten pro Einheit sind darzulegen. Der Zeitpunkt für die Preisermittlung und jegliche Anpassungen aufgrund von Inflation oder Währungsumrechnung müssen dargestellt werden [2].

Da die gesundheitsökonomischen Evaluationen insbesondere von den Entscheidern innerhalb der Gesundheitspolitik herangezogen werden, sollen die relevanten Kosten als Durchschnittswerte dargestellt werden [2].

Zeithorizont

Der Zeithorizont von gesundheitsökonomischen Studien sollte so groß sein, dass alle Effekte und Kosten die durch die betrachtete Maßnahme entstehen, erfasst werden. Dabei ist zu beachten, dass ökonomisch relevante Daten unter Umständen erst bei einem mittel- bis langfristigen Zeithorizont zu erwarten sind.

Endpunkte

Gesundheitsökonomische Evaluationsstudien sollen grundsätzlich Endpunkte verwenden, die sich am Patientennutzen orientieren (siehe Abschnitt 2.1).

Ergebnisdarstellung

Die Ergebnisse der gesundheitsökonomischen Evaluation sollten so dargestellt werden, dass alle relevanten methodischen Aspekte der Evaluation transparent und verständlich sind [29].

Ausblick

Gesundheitsökonomische Evaluationen basieren auf der Wissenschaft verpflichteten, allgemein gültigen Prinzipien und Methoden. Diese Methoden bedürfen im Hinblick auf deren Theorieentwicklung als auch die Standards der praktischen Umsetzung der stetigen Weiterentwicklung [30]. In dem Maße indem vermehrt Fragen nach der Wirtschaftlichkeit medizinischer Maßnahmen gestellt werden, werden gesundheitsökonomische Evaluationen im deutschen Gesundheitswesen an Bedeutung gewinnen. Für den gesundheitspolitischen Dialog leitet sich daraus ab, dass die gesundheitsökonomischen Methoden und die ihnen zugrunde liegenden Prämissen verstanden und akzeptiert werden.

Für die Gesundheitspolitik im Allgemeinen sowie die Entscheidungsträger im Besonderen sind die kritische Würdigung und die Folgenabschätzung gesundheitsökonomischer Bewertungen unverzichtbar. Dafür ist die vollständige Transparenz der Methoden notwendig.

Innerhalb der Solidargemeinschaft wird die Akzeptanz der Gesundheitsökonomie als Entscheidungshilfe auch davon abhängen, in welchem Maß es gelingt, deren Methoden in Einklang mit den gesellschaftlichen Wertvorstellungen von Verteilungs- und Zugangsgerechtigkeit zu bringen.

Wesentliche Aufgabe des Ressorts wird es deshalb sein, nach Verabschiedung des GKV-Wettbewerbsstärkungsgesetzes einen strukturierten Konsultationsprozess über die neuen Methoden zur Kosten-Nutzen-Bewertung zu führen. Dieser Dialog schließt alle Beteiligten im Gesundheitswesen ein und erfolgt in vollständiger Transparenz.

Literatur

- [1] Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on Cost-effectiveness in Health and Medicine. JAMA 1996; 276: 1253-1258.
- [2] Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. BMJ 1996; 313: 275-283.
- [3] Kassirer JP, Angell M. The Journal's policy on cost-effectiveness analyses. N Engl J Med 1994; 331: 669-670.
- [4] Bestehorn K, Biller M, Brecht JG, Clouth J, Fricke F-U, Glaeske G, Greiner W et al. Deutsche Empfehlungen zur gesundheitsökonomischen Evaluation – Revidierte Fassung des Hannoveraner Konsens. Gesundheitsökon Qualitätsmanage 1999; 4: A62-A65.
- [5] Busse R. Gesundheitsökonomie. Ziele, Methodik und Relevanz. Bundesgesundhbl Gesundheitsforsch Gesundheitsschutz 2006; 49: 3-10.
- [6] Drummond MF, Sculpher MJ, Torrance GW, O'Brien B, Stoddart GL. Methods for the Economic Evaluation of Health Care Programmes. Oxford: Oxford University Press; 2005.
- [7] Boulenger S, Nixon J, Drummond M, Ulmann P, Rice S, de PG. Can economic evaluations be made more transferable? Eur J Health Econ 2005; 6: 334-346.

- [8] Welte R, Feenstra T, Jager H, Leidl R. A decision chart for assessing and improving the transferability of economic evaluation results between countries. *Pharmacoeconomics* 2004; 22: 857-876.
- [9] Luce BR, Elixhauser A. Estimating costs in the economic evaluation of medical technologies. *Int J Technol Assess Health Care* 1990; 6: 57-75.
- [10] Sculpher MJ, Pang FS, Manca A, Drummond MF, Golder S, Urdahl H et al. Generalisability in economic evaluation studies in healthcare: a review and case studies [online]. Letztes Update 2004.
<http://www.ncchta.org/htacd.htm> [Zugriff am: 16.03.2006].
- [11] Canadian Agency for Drugs and Technologies in Health. Guidelines for the Economic Evaluation of Health Technologies: Canada [online]. Letztes Update 2006.
<http://www.cadth.ca> [Zugriff am: 08.11.2006].
- [12] National Institute for Clinical Excellence. Guide to the Methods of Technology Appraisal (Reference N0515) [online]. Letztes Update 07 May 2004.
<http://www.nice.org.uk/download.aspx?o=201973> [Zugriff am: 19.09.2006].
- [13] Members of the Collège des Économistes de la Santé (the French Health Economists Association). French Guidelines for the Economic Evaluation of Health Care Technologies [online]. Letztes Update 2004.
http://www.ces-asso.org/docs/France_Guidelines_HE_Evaluation.PDF
[Zugriff am: 09.02.2006].
- [14] Buxton MJ, Drummond MF, Van Hout BA, Prince RL, Sheldon TA, Szucs T, Vray M. Modelling in economic evaluation: An unavoidable fact of life. *Health Econ* 1997; 6: 217-227.
- [15] Sheldon TA. Problems of using modelling in the economic evaluation of health care. *Health Econ* 1996; 5: 1-11.
- [16] Commonwealth of Australia Department of Health and Ageing. 1995 Guidelines for the Pharmaceutical Industry on Preparation of Submissions to the Pharmaceutical Benefits Advisory Committee: including major submissions involving economic analyses [online]. Letztes Update 2002.
<http://www.health.gov.au/internet/wcms/publishing.nsf/Content/health-pbs-general-pubs-pharmpac-gusubpac.htm> [Zugriff am: 08.11.2006].
- [17] Siebert U. When should decision-analytic modeling be used in the economic evaluation of health care? *Eur J Health Econ* 2003; 4 143-150.
- [18] Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C, Luce BR. Principles of good practice for decision analytic modeling in health-care evaluation: Report of the ISPOR Task Force on Good Research Practices - Modeling Studies. *Value Health* 2003; 6: 9-17.
- [19] Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment [online]. Letztes Update 2004.
<http://www.hta.nhsweb.nhs.uk/> [Zugriff am: 16.08.2006].
- [20] Gafni A, Birch S. Guidelines for the adoption of new technologies: a prescription for uncontrolled growth in expenditures and how to avoid the problem. *CMAJ* 1993; 148: 913-917.
- [21] McGregor M, Caro JJ. QALYs. Are they helpful to decision makers? *Pharmacoeconomics* 2006; 24: 947-952.

- [22] Schwappach DL, Boluarte TA. Wie werden in deutschen Studien qualitätsadjustierte Lebensjahre definiert? Ein systematischer Review deutscher Kosten-Nutzwert-Analysen. Dtsch Med Wochenschr 2006; 131: 2004-2009.
- [23] Coast J. Is economic evaluation in touch with society's health values? BMJ 2004; 329: 1233-1236.
- [24] Dolan P, Shaw R, Tsuchiya A, Williams A. QALY maximisation and people's preferences: a methodological review of the literature. Health Econ 2005; 14: 197-208.
- [25] Nord E. Towards cost-value analysis in health care? Health Care Anal 1999; 7 167-175.
- [26] Loomes G, McKenzie L. The use of QALYs in health care decision making. Soc Sci Med 1989; 28(4): 299-308.
- [27] Hadorn DC. The problem of discrimination in health care priority setting. JAMA 1992; 268: 1454-1459.
- [28] National Institute for Health and Clinical Excellence. Social Value Judgements. Principles for the Development of NICE Guidance [online]. Letztes Update 2005. <http://www.nice.org.uk/> [Zugriff am: 27.11.2006].
- [29] da Silva EA, Pinto CG, Sampaio C, Pereira JA, Drummond MF, Trindade R. Guidelines for Economic Drug Evaluation Studies [online]. Letztes Update 1998. <http://www.infarmed.pt/portal/page/portal/INFARMED> [Zugriff am: 09.02.2006].
- [30] Bridges JF. Future challenges for the economic evaluation of healthcare: Patient preferences, risk attitudes and beyond. Pharmacoeconomics 2005; 23: 317-321.

2.6 Leitlinien und Disease-Management-Programme

2.6.1 Hintergrund der Bewertung von Leitlinien

Medizinische Leitlinien gelten als Schlüsselinstrumente zur Verbesserung und Sicherung der medizinischen Qualität in der Patientenversorgung [1]. Ihr Sinn liegt darin, durch die Formulierung von konkreten Handlungsempfehlungen unangemessene Unterschiede in der Patientenversorgung zu reduzieren und die Versorgung der Patienten zu verbessern. Darüber hinaus werden sie in Deutschland auch als Basis für Entscheidungen zu Steuerungszwecken im Gesundheitswesen wie z.B. im Zusammenhang mit der Formulierung von Anforderungen an Disease-Management-Programme (nach § 137f SGB V) genutzt und beeinflussen so auch auf der strukturellen Ebene zunehmend Entscheidungen im Gesundheitswesen.

Vor diesem Hintergrund muss sichergestellt werden, dass medizinische Leitlinien auf der besten verfügbaren und aktuellen wissenschaftlichen Evidenz basieren und unter Berücksichtigung klinischer Erfahrung formuliert werden.

In vielen Fällen fehlt jedoch der Bezug zur aktuellen wissenschaftlichen Evidenz [2,3] und Leitlinien zu gleichen Themen weisen teilweise erhebliche inhaltliche Unterschiede in ihren Empfehlungen auf [4,5].

Ein wesentlicher Grund hierfür ist, dass die international geforderten Qualitätsstandards für die Entwicklung von Leitlinien nicht durchgängig berücksichtigt werden [6-8].

2.6.2 Ziele der Bewertung von Leitlinien

Die spezielle Beurteilung medizinischer Leitlinien soll zu einer Verbesserung der Versorgung durch mehr Transparenz im Gesundheitswesen führen. Hierzu ist es insbesondere notwendig,

- methodisch und inhaltlich gute von schlechten Leitlinien zu diskriminieren,
- die Evidenzgrundlage, auf der Kernempfehlungen in Leitlinien beruhen, auszuarbeiten und zu überprüfen,
- klare fachliche Aussagen über die Sinnhaftigkeit und Wirksamkeit der Umsetzung verschiedener medizinischer Empfehlungen zu geben,
- dem Gemeinsamen Bundesausschuss oder seinen Gremien eine Entscheidungsgrundlage für die Beratungen über strukturierte Behandlungsprogramme zu bieten,
- sicherzustellen, dass nur geprüfte (qualitätsgesicherte) Leitlinien in die Versorgung eingeführt werden, bei denen Hinweise auf eine Verbesserung des Outcomes bestehen,
- Forschungsbedarf zu benennen und sinnvolle Projekte zur Entwicklung und Implementierung von evidenzbasierten Empfehlungen anzustoßen sowie
- die Einbindung in umfassende Qualitätsmanagement-Prozesse (UQM) zu fördern.

Darüber hinaus ermöglichen die Ergebnisse dieser Arbeit dem Nutzer von Leitlinien (Ärzte, Gesundheitseinrichtungen, gesundheitspolitische Ausschüsse, Entscheider im Gesundheitswesen und Patienten) eine Orientierung an sinnvollen und adäquaten Empfehlungen zu prioritären Versorgungsproblemen.

Für die spezielle inhaltliche Bewertung medizinischer Leitlinien sollen die vorhandene methodische Kompetenz und inhaltliche Expertise von externen Instituten/Einrichtungen/Organisationen soweit wie möglich genutzt und einbezogen werden [9].

2.6.3 Methoden der Bewertung von Leitlinien

Wesentliche Ansätze bei der Bewertung und Prüfung der Qualität medizinischer Leitlinien sind

- die Untersuchung formaler Kriterien, die im Wesentlichen die Transparenz des Entstehungsprozesses abbilden und von der Voraussetzung ausgehen, dass Leitlinien, die diese Kriterien berücksichtigen mit einer größeren Wahrscheinlichkeit zu inhaltlich richtigen Empfehlungen kommen (vergleichbar der internen Validität von Studien) [10,11],
- die genaue inhaltliche Prüfung in Hinblick auf die zugrunde liegende Evidenz,

- die Bewertung der Angemessenheit von Empfehlungen,
- die Evaluation der Effekte durch die Anwendung von Leitlinien (Outcomeevaluation).

Die Ansätze unterscheiden sich nach Aufwand, Herangehensweise und Aussagewert z.T. erheblich und kommen je nach Fragestellung und Auftrag zur Anwendung. Die dabei vom Institut verwendete Methodik wird unter Berücksichtigung aktueller wissenschaftlicher Publikationen sowie nationaler und internationaler Erfahrungen regelmäßig überprüft und ggf. aktualisiert. In der Regel werden verschiedene Aspekte in einem abgestuften Verfahren miteinander kombiniert [9]. Kritisch diskutiert wurde im Vorfeld der Methodenüberarbeitung die Beschränkung auf eher formale und methodische Aspekte der Bewertung, die gut operationalisierbar sind. Diese Beschränkung ist zum einen den derzeitigen gesetzlichen Rahmenbedingungen geschuldet (SGB V § 139a), die bisher (nur) eine Überprüfung der Leitlinien hinsichtlich der zugrunde liegenden Evidenz vorsieht. Andererseits kommen die Methoden der evidenzbasierten (und konsentierten) Leitlinienentwicklung und deren Aktualisierung immer noch unzureichend zur Anwendung [12], weshalb auf eine formale Bewertung derzeit nicht verzichtet werden kann.

Was wissenschaftlich richtig ist, muss jedoch nicht notwendiger Weise auch sinnvoll, praktikabel und angemessen sein. Einzelempfehlungen/Kernpunkte in Leitlinien können in Hinblick auf ihre Relevanz, Angemessenheit und Praktikabilität von den Betroffenen sehr unterschiedlich bewertet werden.

Eine Überprüfung der „inhaltlichen Angemessenheit“, zu der es bisher kein international abgestimmtes Verfahren gibt, wohl aber Ansätze in Deutschland im Nationalen Programm für Versorgungsleitlinien erkennbar sind, muss verschiedene Aspekte berücksichtigen und auch andere Akteure wie z. B. den G-BA, Fachvertreter und Patienten mit einbeziehen.

Aussagen zur inhaltlichen Angemessenheit von Leitlinienempfehlungen beinhalten zwei Kernfragen:

1. Ist die Darstellung und Interpretation der aktuellen Evidenz, aus denen konkrete Einzelempfehlungen generiert werden, angemessen und nachvollziehbar? Dies schließt auch den Fall ein dass Abweichungen vorkommen, diese sind jedoch zu begründen. (Die positive Beantwortung dieser Frage ist eine Voraussetzung, um sich der zweiten Frage zu nähern.)
2. Ist die empfohlene Maßnahme angemessen?

Um diese Fragen zu beantworten, sind folgende Aspekte zu berücksichtigen:

- Für die Maßnahme muss ein Nutzen nachgewiesen sein und die Nutzen/Schaden-Abwägung muss zu einer positiven Entscheidung kommen (siehe Abschnitt 2.1).

- Die Maßnahme muss relevant für das deutsche Gesundheitssystem sein.
- Die Maßnahme muss verfügbar und zugelassen sein.
- Die Maßnahme muss notwendig sein.
- Eine sichere Anwendung (durch Ärzte/Patienten) ist möglich.
- Innerhalb der zur Verfügung stehenden Ressourcen muss die Maßnahme umsetzbar und finanzierbar sein.

Insbesondere der letzte Punkt setzt in einer Solidargemeinschaft eine konsentrierte Entscheidung über den verfügbaren finanziellen Gesamtrahmen sowie über eine Priorisierung von Gesundheitsleistungen voraus. Bisher übernimmt der G-BA zumindest indirekt und in Teilen über den Ein- und Ausschluss von Leistungen in den Leistungskatalog der GKV diese Funktion und stellt unter den verschiedenen Interessen und Interessenvertretern (auch Patienten) einen Konsens her. Es ist vorstellbar, dass der G-BA auch zukünftig im Rahmen der inhaltlichen Bewertung von Leitlinien (in Bezug auf deren Angemessenheit) eine ähnliche Funktion einnehmen könnte und im Anschluss an die Prüfung der Evidenzgrundlage durch das IQWiG entsprechende Nationale Leitlinien herausgibt. Vor dem Hintergrund der Beurteilung der Angemessenheit von Leitlinien wird das IQWiG im nächsten Jahr in einen Diskussionsprozess einsteigen, bei dem die Erwartungen an ein solches Verfahren, die Ziele, Methoden und Verantwortlichkeiten geordnet werden müssen.

I. Formale Prüfung

Eine Annäherung an die Frage der Qualität von Leitlinien gelingt durch eine formale Prüfung mit methodischen Kriterien [5,6,13,14]. Die Autoren dieser Publikationen stellen einen direkten Bezug zur „Validität“ von Leitlinien her, verwenden den Begriff aber sehr undifferenziert, weshalb er im Folgenden im Zusammenhang mit Leitlinien nicht mehr verwendet wird. Die formale Leitlinienbewertung erfolgt in strukturierter Form in Anlehnung an die Methoden des Leitlinien-Clearingverfahrens und unter Bezug auf die Beurteilungskriterien für Leitlinien der Bundesärztekammer und Kassenärztlichen Vereinigung [14,15] anhand des Deutschen Leitlinien-Bewertungsinstrumentes (DELBI). Das DELBI ist die deutsche Adaptation des validierten AGREE-Instrumentes, erweitert um eine Domäne, die sich auf die Anwendbarkeit der jeweiligen Leitlinie im deutschen Gesundheitswesen bezieht. Die formale Bewertung wird durch zwei unabhängige Wissenschaftler durchgeführt. Bei unterschiedlichen Einschätzungen werden die Fragen diskutiert und einer erneuten Bewertung unterzogen. Bleibt der Dissens bestehen, der durch eine Autorenanfrage nicht zu lösen ist, werden die unklaren Aspekte gesondert dokumentiert.

Weltweit existieren zu bestimmten medizinischen Fragestellungen mehrere Leitlinien, die sich methodisch und inhaltlich zum Teil sehr stark voneinander unterscheiden [4]. Eine formale Prüfung hat hier eine wichtige Filterfunktion, um letztlich eine inhaltliche Bewertung von Kernaussagen und spezifischen Empfehlungen der relevanten Leitlinien vornehmen zu können (s.u.).

Zunächst wird zur Identifizierung der aktuellen, existierenden Leitlinien für die jeweilige Fragestellung unter Berücksichtigung der im Abschnitt 4.7 (Informationsbeschaffung) genannten Vorgehensweise eine umfassende Leitlinienrecherche in den relevanten Fachdatenbanken (Leitlinien- und Literaturdatenbanken) durchgeführt. Das Filterverfahren der Leitlinienbewertung ist mehrstufig. Ein- und Ausschlusskriterien sowie die Recherchestrategie (Suchbegriffe, Auswahl der Datenbanken etc.) werden in Abhängigkeit von der Fragestellung vorab festgelegt und dokumentiert. Es folgt ein erster Screeningschritt, in dem die Treffer nach vorab definierten, thematischen Kriterien selektiert werden. In einem zweiten Schritt erfolgt eine methodische Kurzbewertung, die diejenigen Leitlinien auswählt, die einem internationalen Mindeststandard entsprechen. Die verbleibenden Dokumente werden dann inhaltlich bewertet (s.u.).

II. Leitlinienvergleich und inhaltliche Bewertung von Kernempfehlungen

Die Bewertung der Inhalte von Leitlinien ist von besonderer Bedeutung. Die bisher mit den gängigen Instrumenten (ÄZQ-Checkliste^b, AGREE-Instrument^c, DELBI-Instrument^d) abgefragten Kriterien zur Identifizierung und Interpretation der Evidenz sowie zur Formulierung der Leitlinienempfehlungen sind im Wesentlichen Transparenzkriterien^e, mit denen allein die Darstellung des Prozesses z.B. zur Literaturrecherche positiv bewertet wird, ohne z.B. eine Beurteilung der Vollständigkeit und Aktualität der Recherche abzugeben. Zu den wesentlichen Schlüsselempfehlungen einer Leitlinie muss deshalb die Ableitung der Empfehlungen aus der zu Grunde liegenden Evidenz im Einzelnen geprüft werden. Die inhaltliche Prüfung umfasst neben der Bewertung auf Vollständigkeit und Aktualität der einbezogenen Literatur auch die Interpretation und Bewertung der Studienergebnisse. Da dieses Verfahren sehr aufwändig ist, muss sich aus pragmatischen Gründen die inhaltliche Prüfung auf die in Auftrag gegebenen Fragestellungen des G-BA bzw. auf die wesentlichen Kernempfehlungen der Leitlinie beschränken. Die Identifizierung der Schlüsselempfehlungen erfolgt im Kontext jedes speziellen Auftrages in Abstimmung mit den betroffenen Ressortleitern und externen Sachverständigen. Eine Fokussierung auf ausgewählte Kernfragen hat hierbei mehrere Vorteile: Da Leitlinien in ihren Inhalten und Empfehlungen, auch wenn sie sich auf dasselbe

^b In AWMF 2001, Leitlinien-Manual

^c Appraisal of Guidelines Research and Evaluation (AGREE)

^d Deutsches Instrument zur methodischen Leitlinienbewertung (DELBI)

^e Derzeit laufen Bemühungen diese Aspekte methodisch zu verbessern, z.B. durch die GRADE-Arbeitsgruppe [10].

medizinische Thema beziehen, sehr heterogen sind, kann man durch das Reduzieren auf vordefinierte Kernfragen einen guten Vergleich zwischen verschiedenen Leitlinien ziehen. Auch ist in diesem Rahmen eine Überprüfung der zu Grunde gelegten Evidenz und der daraus resultierenden Empfehlung möglich, vorausgesetzt, es liegt diesbezüglich eine ausreichende Transparenz vor. Aber nicht nur zwischen Leitlinien verschiedener Herausgeber, sondern auch zu anderen evidenzbasierten Quellen (systematischen Übersichtsarbeiten, HTA-Berichten, Evidenzberichten des IQWiG u.a.) ist ein inhaltlicher Vergleich möglich bezogen auf die definierte Kernfrage. Einzelne Fragestellungen werden bei fehlenden Angaben in Leitlinien in Abstimmung mit den übrigen Ressorts des Instituts bearbeitet.

Ein synoptischer Vergleich der Leitlinieninhalte kann bei der Identifizierung der Schlüsselempfehlungen hilfreich sein. Insbesondere können Fragestellungen, zu denen ein wissenschaftlicher Dissens besteht, identifiziert werden.

Methodisch stellt der synoptische Vergleich jedoch lediglich eine Erleichterung im Bewertungsprozess dar, eine Überprüfung der Evidenzgrundlagen ist auch bei übereinstimmend empfohlenen Verfahren sinnvoll. Insbesondere sollte auch geprüft werden, ob die in der Leitlinie verwendeten Outcomeparameter patientenrelevant sind und eine Nutzen-Schaden-Abwägung [10] bei der Formulierung der Empfehlung berücksichtigt wurde.

III. Verbesserung des Outcomes

Die zentrale Frage bei der Entwicklung und Implementierung von Leitlinien ist, ob die Umsetzung der Empfehlungen zu messbaren Verbesserungen in der Versorgung führt [1,16,17]. Streng genommen lässt sich dies nur durch eine rigorose Effektevaluation prüfen [2]. Aufgrund des finanziellen und personellen Aufwands ist dies jedoch nicht bei jeder bestehenden Leitlinie zu realisieren. So gibt es zum Themenbereich der Outcome-Evaluation nach Leitlinienimplementierung nur unsystematische Untersuchungen mit sehr heterogenen Fragestellungen und Ergebnissen. Auch Pilotstudien wurden bislang vor der Veröffentlichung nur für wenige Leitlinien durchgeführt [18,19].

Liegen Ergebnisse aus Pilotstudien oder Projekten vor, in denen die Leitlinien getestet wurden, sind diese in die Gesamtbewertung einzubeziehen (z.B. durch Darlegung der Methoden, Qualitätsindikatoren, Ergebnisse und Konsequenzen).

Darüber hinaus ist eine Gegenüberstellung der Leitlinienempfehlungen mit in der Routineversorgung allgemein üblichen Verfahren sinnvoll. Sofern komplexe Veränderungen in den Leitlinien gefordert werden, ist die Implementierung schwieriger und muss von unterstützenden Maßnahmen und Tools begleitet werden [5].

Bei ausländischen Leitlinien ist hierbei insbesondere zu prüfen, ob die Ergebnisse auf das deutsche Gesundheitssystem übertragbar sind und/oder welche strukturellen Voraussetzungen für die erfolgreiche Einführung der Empfehlungen realisiert werden müssen. Das Institut kann im Auftrag des G-BA auch mit der Evaluation von Leitlinien beauftragt werden.

2.6.4 Darlegung der Qualitätsprüfung

Aus den vorliegenden Ergebnissen der Prüfung wird ein strukturierter Bericht (Guideline Assessment Report) erstellt, der dem Gemeinsamen Bundesausschuss als Grundlage weiterer Beratungen zur Verfügung gestellt wird. Die Berichte können weiterhin auch als Basis für die Erstellung von themenbezogenen Informationen für Ärzte und Patienten dienen oder zur Überarbeitung von Leitlinien durch die Fachgesellschaften genutzt werden.

2.6.5 Abgabe von Empfehlungen zu Disease-Management-Programmen

Der Gemeinsame Bundesausschuss nach § 91 SGB V benennt Diagnosen für strukturierte Versorgungsprogramme nach § 137f SGB V und erarbeitet inhaltliche Vorgaben für die Programme. Aufgabe des IQWiG ist nach § 139 a (3) SGB V die Abgabe von Empfehlungen zu Disease-Management-Programmen (DMP). Dies umfasst

- die Unterstützung des G-BA bei der Benennung neuer Diagnosen für strukturierte Behandlungsprogramme,
- die Überarbeitung bestehender Vorgaben für strukturierte Versorgungsprogramme und
- die Neuentwicklung inhaltlicher Vorgaben für strukturierte Versorgungsprogramme.

Die konkreten Möglichkeiten der Unterstützung des Gemeinsamen Bundesausschusses und seiner Gremien außerhalb der Beauftragung bei einzelnen Fragestellungen im Rahmen der Nutzenbewertung werden in Abstimmung mit den zuständigen Gremien derzeit noch geprüft.

Literatur

- [1] Entwicklung einer Methodik für die Ausarbeitung von Leitlinien für optimale medizinische Praxis. Empfehlung Rec. (2001) 13 des Europarates und erläuterndes Memorandum. Z ärztl Fortb Qual sich 2001; 95 (Suppl. III): 96.
- [2] Helou A, Ollenschläger G. Ziele, Möglichkeiten und Grenzen der Qualitätsbewertung von Leitlinien. Ein Hintergrundbericht zum Nutzermanual der Checkliste „Methodische Qualität von Leitlinien“. Z ärztl Fortbild Qual sich 1998; 92: 361-365.
- [3] Savoie I, Kazanjian A, Bassett K. Do clinical practice guidelines reflect research evidence? J Health Serv Res Policy 2000; 5: 76-82.
- [4] Irani J, Brown CT, van der Meulen J, Emberton M. A review of guidelines on benign prostatic hyperplasia and lower urinary tract symptoms: are all guidelines the same? BJU Int 2003; 92: 937-942.

- [5] Burgers JS, Cluzeau FA, Hanna SE, Hunt C, Grol R, and the AGREE Collaboration. Characteristics of high quality guidelines: Evaluation of 86 clinical guidelines developed in ten European countries and Canada. *Int J Technol Assess Health Care* 2003; 19: 148-157.
- [6] Cluzeau F, Littlejohns P, Grimshaw J, Feder G, Moran S. Development and application of a generic methodology to assess the quality of clinical guidelines. *Int J Qual Health Care* 1999; 11: 21-28.
- [7] Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed literature. *JAMA* 1999; 281: 1900-1905.
- [8] On Care Pathways, *Bandolier Forum* 2003; 7: 1-12.
<http://www.jr2.ox.ac.uk/bandolier/Extraforbando/Forum2.pdf> [Zugriff am 17.11.2006].
- [9] Fervers B, Burgers JS, Haugh MC, Latreille J, Mlika-Cabanne N, Paquet L et al. Adaptation of clinical guidelines: Literature review and proposition for a framework and procedure. *Int J Qual Health Care* 2006; 18:167-176.
- [10] Guyatt G, Gutterman D, Baumann MH, Addrizzo-Harris D, Hylek EM, Phillips B et al. Grading strength of recommendations and quality of evidence in clinical guidelines: Report from an American College of Chest Physicians Task Force. *Chest* 2006; 129:174-181.
- [11] Grimshaw J, Eccles M, Russell I. Developing clinically valid practice guidelines. *J Eval Clin Pract* 1995; 1: 37-48.
- [12] Ärztliches Zentrum für Qualität in der Medizin. Das Deutsche Leitlinien-Clearingverfahren 1999-2005, Abschlussbericht. *ÄZQ-Schriftenreihe Band 25; Books on Demand, Norderstedt; 2006.*
- [13] Field MJ, Lohr KN. Clinical practice guidelines. Directions from a new program. Institute of Medicine, Washington D.C.; 1990.
- [14] Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften, Ärztliche Zentralstelle Qualitätssicherung. Das Leitlinien-Manual. *Z ärztl Fortbild Qual sich* 2001; 95 (Suppl. I): 1-84.
- [15] Ärztliche Zentralstelle Qualitätssicherung. Checkliste „Methodische Qualität von Leitlinien“. *Dtsch Arztebl* 1998; 95: A2576-A2578, C1838-C1840
- [16] Grol R, Dalhuijsen J, Thomas S, Veld C, Rutten G, Mokkink H. Attributes of clinical guidelines that influence use of guidelines in general practice: Observational study. *BMJ* 1998; 317: 858-861.
- [17] Worrall G, Chaulk P, Freake D. The effects of clinical practice guidelines on patient outcomes in primary care: A systematic review. *CMAJ* 1997; 156: 1705-1712.
- [18] Ärztliches Zentrum für Qualität in der Medizin. Leitlinien-Clearingbericht „COPD“. *ÄZQ-Schriftenreihe Band 14; Verlag Videel, Niebüll; 2003.*
- [19] Ärztliches Zentrum für Qualität in der Medizin. Leitlinien-Clearingbericht „Depression“. *ÄZQ-Schriftenreihe Band 12; Verlag Videel, Niebüll; 2003.*

2.7 *Systematische Übersichten und HTA-Berichte*

Systematische Übersichten (Reviews) und HTA-Berichte stellen als zusammenfassende Darstellung der aktuellen Studienlage eine wertvolle Entscheidungsgrundlage für Ärzte und Patienten,

aber auch für Entscheidungsträger im Gesundheitswesen dar. Ebenso wird das Institut bei der Erstellung von Berichten und Gesundheitsinformationen auf vorliegende systematische Übersichten und HTA-Berichte zurückgreifen. Notwendige Voraussetzung hierfür ist, dass die Methodik dieser Arbeiten den Anforderungen des Instituts entspricht, das bedeutet insbesondere eine transparente Darstellung der Literaturrecherche- und Auswahlstrategie, und dass diese Strategie und die zur Auswahl der relevanten Arbeiten verwendeten Ein- und Ausschlusskriterien festgeschriebenen Kriterien entsprechen. Ein wesentlicher Aspekt in der Bewertung ist neben der Beschreibung der angewandten Methodik zur Erstellung der Übersicht auch die Frage, ob die einbezogenen Studien einer Qualitätsprüfung unterzogen wurden, was deren Ergebnis war, und ob – und gegebenenfalls wie – die Qualitätsbewertung Einfluss auf eine eventuelle Synthese der Einzelstudienresultate hatte [1,2].

Eine systematische Übersicht und ein HTA-Bericht können, wie jede andere wissenschaftliche Publikation, nicht nur zu fehlerhaften Ergebnissen führen, sondern auch durch verschiedene Verzerrungen in ihrer Aussage gelenkt werden [1]. Grundsätzlich ist zu berücksichtigen, dass es auch bei systematischen Übersichten und HTA-Berichten erhebliche Qualitätsunterschiede gibt, die eine Einstufung solcher Arbeiten auf einer hohen Evidenzstufe nicht immer rechtfertigen.

In den letzten Jahren ist die Anzahl systematischer Übersichten und von HTA-Berichten stark angestiegen; derzeit werden ca. vier systematische Übersichten und HTA-Berichte pro Tag veröffentlicht [3]. Mit dieser Zunahme zeigen sich auch Probleme, zum Beispiel durch systematische Übersichten mit widersprüchlichen Aussagen [4].

Bei der Erstellung von Institutsberichten (siehe Abschnitt 4.4) werden systematische Übersichten und HTA-Berichte primär zur Identifizierung potenziell relevanter (Primär)Studien genutzt (siehe Abschnitt 4.7.2). Während die vom Institut erstellten Gesundheitsinformationen zu einem wesentlichen Teil auf systematischen Übersichten und HTA-Berichten beruhen (siehe Abschnitt 3.3.3), kann ein Institutsbericht nur in zu begründenden Ausnahmefällen allein oder teilweise auf der Basis von solchen Arbeiten erstellt werden, beispielsweise bei besonderer Dringlichkeit oder zur Bearbeitung eines Teilaspekts bei komplexen Fragestellungen.

Notwendige Voraussetzung für die Einbeziehung von systematischen Übersichten und HTA-Berichten in die Gesundheitsinformationen und gegebenenfalls Nutzenbewertungen innerhalb von Berichten des Institutes ist eine methodische Prüfung in Anlehnung an den Qualitätsindex für systematische Übersichten von Oxman und Guyatt [5-7]. Darüber hinaus wird dieser formalen eine inhaltliche Prüfung nachgestellt, zum Beispiel eine stichprobenartige Bewertung der in die systematische Übersicht eingeschlossenen Primärstudien oder eine Überprüfung von eventuell die Übersicht dominierenden Einzelstudien.

Literatur

- [1] Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomized controlled trials: The QUOROM statement. *Lancet* 1999; 354: 1896-1900.
- [2] Chalmers I, Haynes RB. Reporting, updating, and correcting systematic reviews of the effects of health care. In: Chalmers I, Altman DG, eds. *Systematic Reviews*. London: BMJ Publishing Group 1995: 86-95.
- [3] Bastian H, Glasziou P. Four systematic reviews per day: the epidemiology of research synthesis. *Cochrane Colloquium*, Melbourne; Oktober 2005.
- [4] Jadad AR, Cook DJ, Browman G. A guide to interpreting discordant systematic reviews. *CMAJ* 1997; 156: 1411-1416.
- [5] Jadad AR. *Randomised Controlled Trials: A User's Guide*. London: BMJ Books; 1998.
- [6] Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol* 1991; 44: 1271-1278.
- [7] Oxman AD, Guyatt GH, Singer J, Goldsmith CH et al. Agreement among reviewers of review articles. *J Clin Epidemiol* 1991; 44: 91-98.

2.8 Prognose

Eine wesentliche Basis für die Bewertung von Prognosestudien ist eine präzise Fragestellung, da Studien zur Evaluierung prognostischer Merkmale mit unterschiedlichen Zielen wie zum Beispiel Risikofaktorevaluierung, Scoreentwicklung oder Scorevalidierung durchgeführt werden. Die Abgrenzung zu Diagnose- und/oder Screeningstudien kann schwierig sein. Je nach Zielsetzung müssen zur Beurteilung von Prognosestudien unterschiedliche Bewertungsprinzipien zum Zuge kommen.

Ein prognostisches Merkmal liefert eine Information, die letztlich keinen Selbstzweck haben kann, sondern zu einer Konsequenz führen sollte, die ihrerseits einen prüfbaren Nutzen für den Patienten bedeutet. In diesem Zusammenhang sind an ein Prognoseverfahren ähnliche (allgemeine) Anforderungen wie an einen diagnostischen Test zu stellen (siehe Abschnitt 2.3). Soll ein prognostisches Merkmal im Sinne eines Screening- bzw. Präventionsprogramms eingesetzt werden, sind bei der Beurteilung die in diesen Abschnitten (2.4 bzw. 2.10) formulierten Grundsätze zu berücksichtigen.

Zur Beurteilung von Prognosestudien gibt es keine allgemein akzeptierten Qualitätskriterien [1,2]. Richtlinien zur Planung und Durchführung von Prognosestudien in der Onkologie werden von Simon und Altman [1] beschrieben. Ein allgemeiner Rahmen zur Beurteilung von Prognosestudien wird von Laupacis et al. [3] vorgeschlagen. Zu beachten sind in jedem Fall unten genannte Aspekte, die sich aus der zu Grunde liegenden Datenquelle sowie der verwendeten Datenanalyse ergeben.

Da bei Prognosestudien häufig multifaktorielle Regressionsmodelle eine zentrale Rolle spielen, ist außerdem Abschnitt 1.22 zu beachten. Speziell bedeutsam sind folgende Aspekte:

- klare Formulierung einer Fragestellung und damit verbunden der Studienplanung; dies schließt eine Fallzahlplanung ein, die sich beispielsweise an der gewünschten Präzision der Schätzung (Breite des Konfidenzintervalls) orientieren kann und eine Abschätzung sowohl der Prävalenz der Exposition als auch der Inzidenz im Hinblick auf die interessierende Zielvariable erfordert
- klare Beschreibung der Ziel- und Stichprobenpopulation (zum Beispiel Bevölkerungs- oder Hausarztbezug, Register) und Begründung für deren Wahl; klare Beschreibung der Auswahl und des Vorgehens bei der Rekrutierung der Studienteilnehmer
- Homogenität der betrachteten Population; falls es sich um eine heterogene Population handelt, sollte darauf geachtet werden, dass die prognostische Aussage möglichst konstant über die die Heterogenität verursachenden Subgruppen (beispielsweise unterschiedliche Basis-Risiken für die interessierende Zielvariable) getroffen werden kann
- klare Definition einer (oder mehrerer) Zielvariablen, an der (denen) die sich die prognostische Wertigkeit orientiert werden soll
- klare Definition der prognostischen Merkmale einschließlich der statistischen Handhabung (zum Beispiel Dichotomisierung oder Betrachtung von Terzilen, Quartilen etc. bei einem quantitativen Merkmal) und einer Begründung für das gewählte Vorgehen
- klare Festlegung und Definition von möglichen Störgrößen und Wechselwirkungen, einschließlich deren statistischer Handhabung
- bei Kohortenstudien Vollständigkeit des Follow-up beziehungsweise Maßnahmen zur Erzielung eines möglichst vollständigen Follow-up; Abschätzung möglicher Selektionseffekte bei unvollständigem Follow-up
- Bei der Betrachtung prognostischer Scores ist zu beachten, dass zwischen Scoreentwicklung und Scorevalidierung differenziert wird, zum Beispiel als Entwicklung des Scores innerhalb einer so genannten „Lernstichprobe“ und Validierung an einer „Teststichprobe“. Idealerweise erfolgen Scoreentwicklung und Scorevalidierung in unterschiedlichen Studien.

Typische Studiendesigns für die Evaluierung prognostischer Merkmale im Sinne von Risikofaktoren stellen Kohortenstudien und Fall-Kontroll-Studien dar. In Ausnahmefällen, zum Beispiel bei der Betrachtung unveränderlicher Merkmale, können auch Querschnittstudien eine Rolle spielen. Die grundlegenden Prinzipien für die Bewertung solcher Studien über die oben genannten Aspekte hinaus sind in Abschnitt 1.5 formuliert.

Die Literatursuche bei der Evaluierung von prognostischen Merkmalen (im Rahmen einer systematischen Übersicht) ist schwieriger als beispielsweise bei Therapiestudien und es existiert (noch) keine allgemein akzeptierte optimale Suchstrategie. Darüber hinaus wird angenommen, dass dieser Bereich besonders anfällig für einen Publikationsbias ist [1,2]. Die methodische Qualität von Studien bzw. deren Publikationen zu prognostischen Merkmalen ist häufig unzureichend [4], so dass die Extraktion von benötigten Daten erschwert oder sogar unmöglich wird. Insofern sind Meta-Analysen (nicht jedoch systematische Übersichten an sich) von Prognosestudien oft unangemessen und deren Ergebnisse nur unter Vorbehalt zu verwerten [2]. Einige wichtige Probleme bei Meta-Analysen von Prognosestudien lassen sich vermeiden, wenn individuelle Patientendaten zur Verfügung stehen [2].

Literatur

- [1] Simon R., Altman DG. Statistical aspects of prognostic factor studies in oncology. *Br J Cancer* 1994; 69: 979-985.
- [2] Altman D. Systematic reviews of evaluations of prognostic variables. In: Egger M, Davey Smith G, Altman DG, editors. *Systematic Reviews in Health Care: Meta-analysis in Context*, 2nd Ed. London: BMJ Books; 2001.
- [3] Laupacis A, Wells G, Richardson WS, Tugwell P for the Evidence-Based Medicine Working Group. Users' guides to the medical literature. V. How to use an article about prognosis. *JAMA* 1994; 272: 234-237.
- [4] Pocock SJ, Collier TJ, Dandreo KJ, de Stavola BL, Goldman MB, Kalish LA, Kasten LE, McCormack VA. Issues in the reporting of epidemiological studies: A survey of recent practice. *BMJ* 2004; 329: 883-887.

2.9 Individuelle Risikoabschätzung

Neben Ergebnissen aus Studien, die einzelne oder (zumeist) mehrere prognostische Merkmale untersuchen, werden für die individuelle Risikoabschätzung von Patienten (oder allgemeiner Personen), ein unerwünschtes Ereignis zu erleiden, zunehmend so genannte „risk charts“ (auch als „risk engines“ bezeichnet) eingesetzt, bei denen für das Zusammentreffen mehrerer Risikofaktoren multifaktorielle Schätzwerte angegeben werden (zum Beispiel Sheffield Table [1] oder „Joint British Chart“ [2]). Die Grundlage dieser Risikotabellen sind meist multifaktorielle Regressionsmodelle, deren Ergebnisse zur einfacheren Handhabung in Tabellen oder Punktesysteme überführt werden [3]. Zu beachten ist, dass aus solchen Risikotabellen abgeleitete Risiken keine „persönlichen“ Schätzungen für spezielle Individuen sind, sondern statistische Schätzungen für mittlere Risiken einer Populationsgruppe mit speziellen Risikoprofilen für einen definierten Zeitraum (zum Beispiel zehn Jahre) darstellen. Bei der Bewertung solcher Instrumente ist zu berücksichtigen,

- für welche Population die geschätzten Risiken gelten,

- aus welcher Art von Studien die zu Grunde liegenden Daten stammen,
- ob die für die multifaktorielle Betrachtung herangezogenen Merkmale in diesen zu Grunde liegenden Studien ebenfalls gemeinsam betrachtet wurden,
- ob, und wenn ja, wie eine multifaktorielle statistische Analyse in diesen zu Grunde liegenden Studien erfolgte (siehe Abschnitt 1.22),
- ob diese Instrumente jemals in nachfolgenden Studien validiert wurden (Teststichproben).

Literatur

- [1] Wallis EJ, Ramsay LE, Ul Haq I, Ghahramani P, Jackson PR, Rowland-Yeo K, Yeo WW. Coronary and cardiovascular risk estimation for primary prevention: Validation of a new Sheffield table in the 1995 Scottish health survey population. *BMJ* 2000; 320: 671-676.
- [2] British Cardiac Society, British Hyperlipidaemia Association, British Hypertension Society, endorsed by the British Diabetic Association. Joint British recommendations on prevention of coronary heart disease in clinical practice. *Heart* 1998; 80 (Suppl 2): S1-S29.
- [3] Sullivan LM, Massaro JM, D'Agostino RB. Presentation of multivariate data for clinical use: The Framingham Study risk score functions. *Stat Med* 2004, 23: 1631-1660.

2.10 Beurteilung bevölkerungsweiter Interventionen/Prävention

Prävention zielt darauf ab, eine gesundheitliche Schädigung zu verhindern, weniger wahrscheinlich zu machen oder zu verzögern [1]. Während die Primärprävention alle Maßnahmen vor Eintritt einer fassbaren biologischen Schädigung zur Vermeidung auslösender oder vorhandener Teilursachen umfasst, beinhaltet die sekundäre Prävention Aktivitäten zur Entdeckung klinisch symptomloser Krankheitsfrühstadien und ihre erfolgreiche Frühtherapie (siehe auch Abschnitt 2.4). Tertiärprävention im engeren Sinne bezeichnet spezielle Interventionen zur Verhinderung bleibender, insbesondere sozialer Funktionseinbußen nach dem Eintreten einer Krankheit [1].

Maßnahmen der Primär- und Sekundärprävention zeichnen sich dadurch aus, dass im Unterschied zu kurativen Maßnahmen häufig ganze Bevölkerungsgruppen im Fokus der Intervention stehen. Für den Nutzen auf Populationsebene ist neben der Effizienz eines Programms auch die Höhe der Teilnehmerate entscheidend. Außerdem ist die Frage relevant, welche Personen erreicht werden; so gibt es Hinweise darauf, dass Präventionsprogramme gerade von Bevölkerungsgruppen mit erhöhtem Krankheitsrisiko weniger in Anspruch genommen werden [2].

Eine Herausforderung bei der Nutzenbewertung von Präventionsmaßnahmen besteht darin, dass die Wirkungskette zwischen einer Intervention und einer möglichen Vermeidung des Auftretens von Erkrankungen komplex ist und sich unter Umständen über viele Jahre erstrecken kann, so dass zum Nachweis eines Effektes auf das Krankheitsgeschehen sehr lange Beobachtungszeiträume notwen-

dig sein können [3]. Deshalb wird bei der Evaluation von Präventionsmaßnahmen oft auf Surrogatkriterien – in diesem Zusammenhang häufig „intermediäre Outcome-Parameter“ genannt – wie Veränderungen des Lebensstils (beispielsweise Rauch- oder Ernährungsgewohnheiten) zurückgegriffen. Zum Teil werden auch nur Veränderungen hinsichtlich des gesundheitsrelevanten Wissens oder Veränderungen der Einstellung beziehungsweise Absichten erfasst [3]. Häufig ist allerdings der eindeutige und stets gleichgerichtete Zusammenhang zwischen diesen Surrogatparametern und (günstigen) Effekten im Hinblick auf Morbidität oder gar Mortalität nicht gesichert [4]. In diesem Fall können Studien mit Surrogaten als Zielgrößen nicht allein für eine Nutzenbewertung herangezogen werden.

Zur Beurteilung der Effektivität von Präventionsprogrammen wie generell für Interventionen (siehe auch Abschnitt 1.9) gewährleisteten RCTs die höchste interne Validität [5]. Insbesondere umfassende bevölkerungsbezogene Programme können jedoch spezifische Herausforderungen hinsichtlich des geeigneten Studiendesigns mit sich bringen. Unter anderem wegen einer möglichen Kontamination zwischen Interventions- und Kontrollgruppe werden zunehmend auch Cluster-RCTs durchgeführt [6]. Bei diesem Studiendesign werden nicht Individuen, sondern Gruppen von Personen den verschiedenen Bedingungen zufällig zugeordnet. Eine mögliche Quelle systematischer Verzerrungen, die bei der Bewertung cluster-randomisierter Studien berücksichtigt werden muss, liegt zum Beispiel in der nachträglichen Zuordnung von Personen zu den Clustern [7].

Es ist jeweils zu prüfen, inwieweit in speziellen Fällen die Berücksichtigung weiterer Studiendesigns sinnvoll sein kann [5]. So werden beispielsweise massenmediale Kampagnen häufig im Rahmen eines so genannten „Interrupted Time-Series Designs“ evaluiert (zum Beispiel bei [8]) und die Anwendung dieses Studiendesigns wird auch für gemeindebezogene Präventionsprogramme empfohlen [10]. Qualitätskriterien für die Bewertung solcher Studien wurden von der Cochrane Effective Practice and Organisation of Care Review Group entwickelt [10].

Mehr noch als bei klinischen Interventionen ist zu beachten, dass die Wirksamkeit umfassender Präventionsprogramme wesentlich von (bekannten oder nicht bekannten) Kontextfaktoren abhängen kann [11]. Deshalb sollte eine systematische Übersicht möglichst umfassende Informationen über das jeweilige Setting, in dem die Intervention untersucht worden ist, bereitstellen, damit die Anwendbarkeit auf den eigenen Versorgungskontext beurteilt werden kann [5,12]. Bei der Synthese mehrerer Studien ist jeweils zu prüfen, ob die Studien tatsächlich konzeptionell vergleichbar genug sind, um eine Meta-Analyse zu rechtfertigen (siehe auch Abschnitt 1.21).

Literatur

- [1] Walter U, Schwartz FW. Prävention. In: Schwartz FW, Badura B, Busse R, Leidl R, Raspe H, Siegrist J, Hrsg. Das Public Health Buch. München: Urban & Fischer Verlag; 2003, 189-214.
- [2] Laaser U, Hurrelmann K. Gesundheitsförderung und Krankheitsprävention. In: Hurrelmann K, Laaser U, Hrsg. Handbuch Gesundheitswissenschaften. Weinheim: Juventa; 1998, 395-424.
- [3] Nutbeam D. Health Promotion Effectiveness - The Questions to be Answered. The Evidence of Health Promotion Effectiveness - Shaping Public Health in a New Europe. Brüssel: IUHPE; 2000, 1-11.
- [4] Stacy AW, Bentler PM, Flay BR. Attitudes and health behavior in diverse populations: drunk driving, alcohol use, binge eating, marijuana use, and cigarette use. Health Psychology 1994; 13: 73-78.
- [5] Jackson N, Waters E. Guidelines for systematic reviews in health promotion and public health taskforce. Criteria for the systematic review of health promotion and public health interventions. Health Promot Int 2005; 20: 367-374.
- [6] Torgerson DJ. Contamination in trials: Is cluster randomisation the answer? BMJ 2001; 322: 355-357.
- [7] Hahn S, Puffer S, Torgerson DJ, Watson J. Methodological bias in cluster randomised trials. BMC Med Res Methodol 2005; 5: 10.
- [8] Vidanapathirana J, Abramson MJ, Forbes A, Fairley C. Mass media interventions for promoting HIV testing. Cochrane Database Syst Rev 2005 Jul 20; CD004775.
- [9] Biglan A, Ary D, Wagenaar AC. The value of interrupted time-series experiments for community intervention research. Prev Sci 2000; 1: 31-49.
- [10] Cochrane Effective Practice and Organisation of Care (EPOC) Review Group. The Data Collection Checklist [online], 2002
<http://www.epoc.uottawa.ca/checklist2002.doc> [Zugriff am 07.02.2006].
- [11] Hawe P, Shiell A, Riley T, Gold L. Methods for exploring implementation variation and local context within a cluster randomised community intervention trial. J Epidemiol Community Health 2004; 58: 788-793.
- [12] Briss PA, Zaza S, Pappaioanou M, Fielding J, Wright-De Agüero L, Truman BI, et al. Developing an Evidence-Based Guide to Community Preventive Services - Methods. Am J Prev Med 2000; 18: 35-43.

2.11 Beschreibung der Art und Größe des Placeboeffektes

Der Effekt einer Intervention im Ganzen setzt sich aus verschiedenen Komponenten zusammen: dem eigentlichen, alleine der Intervention zuzuschreibenden Effekt, dem natürlichen Verlauf der Erkrankung und dem so genannten Placeboeffekt [1], der den Kontext, in dem die Behandlung stattfindet, beschreibt. Hierunter werden alle Einflüsse zusammengefasst, die neben einer Intervention begleitend auftreten, wie zum Beispiel die Erwartungshaltung der Patienten und der Behandler, die Suggestion der Behandler oder auch Effekte, die alleine darauf zurückzuführen sind, dass Pati-

enten betreut werden. Placebokontrollierte Studien dienen dazu, diese begleitenden Effekte in beiden Gruppen gleich stark einwirken zu lassen, indem eine Gruppe mitgeführt wird, die eine Placebobehandlung erfährt. Entscheidend hierbei ist die Möglichkeit einer Verblindung von Patienten und Behandlern bezüglich der Intervention. Die Verblindung der Behandler soll gewährleisten, dass die über die Intervention hinausgehende Betreuung, Zuwendung und Suggestion auf die Studienpatienten gleich verteilt wird.

Eine Placebobehandlung ist nicht allein auf medikamentöse Interventionen beschränkt, sondern kann auch bei nichtmedikamentösen Verfahren eingesetzt werden, um zumindest eine Verblindung der Patienten zu erreichen. In diesem Zusammenhang wird auch von so genannten „Sham“-Interventionen gesprochen. Das Ausmaß eines Placeboeffekts kann von der Art der Intervention abhängig sein [2].

Im Laufe einer Studie kann es bei bestimmten Interventionen trotz ursprünglicher Verblindung zu einer Entblindung der Patienten und der Behandler kommen, beispielsweise bei spezifischen Nebenwirkungen der zu prüfenden Therapie. Eine mögliche Entblindung beziehungsweise eine erst gar nicht erfolgte Verblindung kann zu einer Verzerrung der Ergebnisse führen. Das genaue Ausmaß dieser Verzerrung wird in der Regel nicht ermittelbar sein. Dazu wäre eine unverzerrte Schätzung des Ausmaßes des Placeboeffektes notwendig, zum Beispiel durch das Mitführen einer dritten Gruppe ohne jegliche Behandlung. Doch gerade in den Situationen, in denen es zu einer Entblindung kommt, lässt sich der Placeboeffekt nicht unverzerrt schätzen, da auch die Differenz zwischen der Placebogruppe und der Gruppe ohne Behandlung verzerrt ist. Wenn beispielsweise davon ausgegangen wird, dass sich ein möglicher Placeboeffekt bei Entblindung für die Patienten der Placebogruppe substantiell verringert, würde die Differenz zwischen der Placebogruppe und der Gruppe ohne Behandlung eine zu kleine Schätzung für den Placeboeffekt darstellen.

Trotz der beschriebenen Problematik sollte die Validität der Ergebnisse unverblindeter Studien beziehungsweise von Studien mit Entblindung diskutiert werden, falls Anhaltspunkte für einen möglichen großen Placeboeffekt vorliegen.

Es existieren Vorschläge, die Verblindung am Ende einer Studie zu überprüfen, indem verglichen wird, wie viele Patienten und Behandler die zugeordnete Gruppe korrekt erraten haben [3]. Solche Überprüfungen bergen allerdings bisher noch nicht befriedigend gelöste und kontrovers diskutierte methodische Probleme (z.B.: Welche Verfahren sollen für welche Nullhypothesen eingesetzt werden? [4-6]). Im Falle wirksamer Therapien mit für Patienten direkt erfahrbaren Wirkungen ist eine Entblindung in gewissem Umfang möglich oder gar wahrscheinlich. Das heißt, in einer solchen Situation ist es schwierig oder unmöglich, zu beurteilen, ob durch eine Entblindung (welcher Ursache auch immer) eine Verzerrung des Therapieeffekts zustande gekommen ist oder umgekehrt der

Therapieeffekt zu einer Entblindung geführt hat. Ungeachtet dieser Probleme ist die Überprüfung der Verblindung in einer Studie zu begrüßen, da sie ein Indiz dafür liefert, dass das Thema Verblindung während der Planung und Durchführung der Studie angemessen berücksichtigt worden ist.

Die Interpretation der Ergebnisse von unverblindeten Studien beziehungsweise Studien, in denen es (möglicherweise) zu einer Entblindung gekommen ist, muss im Vergleich zu verblindeten Studien vorsichtiger erfolgen und setzt voraus, dass die Endpunkterhebung verblindet vorgenommen wurde (abgesehen vom Endpunkt Gesamtmortalität).

Literatur

- [1] Thompson WG. The Placebo Effect and Health. New York: Prometheus Books, 2005.
- [2] Kaptchuk TJ, Stason WB, Davis RB, Legedza AR, Schnyer RN, Kerr CE, Stone DA, Nam BH, Kirsch I, Goldman RH. Sham device v inert pill: Randomised controlled trial of two placebo treatments. *BMJ* 2006; 332: 391-397.
- [3] Fergusson D, Cranley Glass K, Waring D, Shapiro S. Turning a blind eye: The success of blinding reported in a random sample of randomised, placebo controlled trials. *BMJ* 2004; 328: 432-434.
- [4] Altman DG, Schulz KF, Moher D. Turning a blind eye: Testing the success of blinding and the CONSORT statement (Letter). *BMJ* 2004; 328: 1135.
- [5] Senn SJ. Turning a blind eye: Authors have blinkered view of blinding (Letter). *BMJ* 2004; 328: 1135-1136.
- [6] Sackett DL. Turning a blind eye: Why we don't test for blindness at the end of our trials. *BMJ* 2004; 328: 1136.

3. Evidenzbasierte Gesundheitsinformationen für Bürger und Patienten

3.1 Ziel

Das Institut verfolgt das Ziel, ein effektiver, zuverlässiger, vertrauenswürdiger und populärer Herausgeber von evidenzbasierten Gesundheitsinformationen für Bürger und Patienten zu sein. Die Gesundheitsinformationen sollen einer allgemeinen gesundheitlichen Aufklärung dienen. Im Rahmen einer ganzheitlichen und umfassenden Vorgehensweise werden dabei u.a. relevante Informationen aus Institutsberichten in der Bevölkerung primär über die Internetseite "<http://www.gesundheitsinformation.de>" verbreitet.

Die Gesundheitsinformationen sollen zielgerichtet und objektiv sein sowie intensiv genutzt werden. Sie sollen zudem an die psychologischen Bedürfnisse der Patienten angepasst und auch ohne medizinisches Fachwissen verständlich sein.

Zum Ziel einer allgemeinen gesundheitlichen und wissenschaftlichen Aufklärung zählt:

- das Verständnis von physischer, psychischer und seelischer Gesundheit zu verbessern,
- das Verständnis medizinischer und wissenschaftlicher Informationen, einschließlich der Konzepte evidenzbasierter Medizin, zu verbessern,
- gesundheitsrelevante Verhaltensweisen zu fördern,
- die Unterstützung durch Angehörige und Freunde zu fördern,
- die kritische Nutzung von Gesundheitsdienstleistungen zu fördern,
- partizipative Entscheidungen in Gesundheitsfragen (zum Beispiel Arzt-Patient-Beziehung) zu unterstützen, unter Berücksichtigung der individuellen Bedürfnisse und Wertvorstellungen der Patienten bei der Anwendung medizinischer Tests und Therapien.

Das Institut ist kein direkter Ratgeber für Bürger und Patienten. Vielmehr ist es sein Anliegen, eigenständige und eigenverantwortliche Entscheidungen in Gesundheitsfragen zu fördern. Dabei steht die Autonomie der Bürger und Patienten im Vordergrund [1].

3.2 Informationssystem

Durch internetbasierte [2,3] und „offline“ elektronisch verfügbare [4] Gesundheitsinformationen können der Kenntnisstand und die Entscheidungsfähigkeit der Bürger und Patienten und darüber

deren physische, psychische und seelische Gesundheit positiv beeinflusst werden. Jedoch können Informationen und Aufklärungsmaßnahmen auch ineffektiv oder schädlich sein, und einige Techniken sind effektiver als andere [3,5-8]. Die Website www.gesundheitsinformation.de ist das erste Instrument zur Verbreitung der evidenzbasierten Gesundheitsinformationen des Instituts.

Die Website wird zu einem umfassenden und vielseitigen Nachschlagewerk ausgebaut, das die individuellen und vielfältigen Bedürfnisse nach Gesundheitsinformationen der Bürger und Patienten auf verschiedenen Ebenen befriedigen soll [5,6]. Hierfür kommen unterschiedliche Informationsprodukte zum Einsatz, die im folgenden Abschnitt 3.2.1 beschrieben werden. Die Informationsprodukte sind untereinander vernetzt. Darunter ist zu verstehen, dass durch Links Begriffserklärungen, Erläuterungen und ergänzende Themen zugänglich sind.

Der Website werden ein elektronischer Newsletter sowie zum Beispiel elektronisch herunterladbare Texte beziehungsweise sonstige Dateien mit gesundheitsrelevanten Themen als Transportwege für die Gesundheitsinformationen angegliedert, und es können elektronische Informationen auf anderen Websites reproduziert werden. Darüber hinaus besteht die Möglichkeit, Printversionen der Gesundheitsinformationen herunterzuladen.

3.2.1 Informationsprodukte

Zu den Informationsprodukten zählen ausführliche Informationen (auf der Website als „Artikel“ benannt), Merkblätter, Kurzinformationen (auf der Website als „Kurzantworten“ benannt) und sonstige Produkte.

Ausführliche Informationen

Die ausführlichen Informationen werden auf der Website als „Artikel“ mit einem zusammenfassenden Merkblatt veröffentlicht und bilden den Schwerpunkt des entstehenden Nachschlagewerkes, das umfassend über eine Vielzahl von gesundheitsrelevanten Themen informiert. Ausführliche Informationen können zudem aus wissenschaftlichen Berichten des Instituts erstellt werden. Die ausführlichen Informationen beziehen sich auf

- Erklärungen zu Erkrankungen und Beschwerden, einschließlich
 - Anatomie
 - Physiologie
 - verschiedene Erkrankungsformen
 - Erkrankungsursachen
 - Symptomerkennung
 - normaler Verlauf von Erkrankungen

- Prognose
- mögliche Komplikationen
- Erkennung von Komplikationen
- Genesung
- mögliches Wiederkehren der Erkrankung
- Erkennung einer wiederkehrenden Erkrankung
- Risikogruppen (einschließlich Angehörige)
- Maßnahmen zur Prävention und Gesundheitsförderung, einschließlich
 - Ernährung
 - körperliche Aktivität
 - Screeningverfahren
 - Informationen
- diagnostische Maßnahmen, einschließlich
 - komplementäre Diagnoseverfahren
- therapeutische Maßnahmen, einschließlich
 - Arzneimittel
 - Chirurgie
 - komplementäre Therapien
 - andere, auch nichtmedikamentöse Verfahren
- Maßnahmen im Rahmen der Rehabilitation
- weitere Leistungen im Gesundheitssystem sowie weitere Angebote
- psychosoziale Aspekte sowie Erfahrungen von Patienten und Angehörigen mit verschiedenen Erkrankungen

Merkblätter

Neben den ausführlichen Informationen werden Merkblätter als weitere Basis des Nachschlagewerkes entwickelt.

Merkblätter fassen einerseits die wesentlichen Aussagen von ausführlichen Informationen zusammen. Andererseits basieren Merkblätter auf einzelnen oder mehreren Kurzinformationen, ergänzt von weiterführenden Informationen. Weiterhin werden Merkblätter als Zusammenfassung einer oder mehrerer systematischer Übersichten entwickelt.

Kurzinformationen

Neben den ausführlichen Informationen und den Merkblättern werden in weit größerer Anzahl Kurzinformationen („kurze Antworten auf wissenschaftliche Fragen“) erstellt. Diese Kurzinformationen ergänzen das Nachschlagewerk. Zudem sollen sie evidenzbasierte Informationen, die bisher häufig nur in englischer Sprache angeboten wurden, einer breiten Öffentlichkeit zugänglich machen.

Die Kurzinformationen sind allgemeinverständliche Zusammenfassungen von wichtigen, interessanten und/oder aktuellen Gesundheitsthemen. Zudem werden Kurzinformationen zu wissenschaftlichen Berichten des Instituts erstellt.

Sonstige Produkte

Unter sonstigen Produkten sind visuelle und interaktive Tools wie komplexe Schaubilder, Online-Rechner (zum Beispiel für Zigarettenkosten) und Glossare (zum Beispiel Online-Wörterbuch) zu verstehen.

Die Tools sollen zum Beispiel

- das allgemeine medizinische Verständnis fördern,
- das Verständnis von Erkrankungen verbessern, zum Beispiel Kenntnisse über den normalen Verlauf von Erkrankungen, die Symptomererkennung, die möglichen Komplikationen, die Genesung und ein mögliches Wiederkehren der Erkrankung vermitteln,
- Möglichkeiten der Risikoabwägung aufzeigen,
- eine Hilfe zum Selbstmanagement sein, zum Beispiel bei chronischen Erkrankungen.

3.2.2 Redaktionssystem

Das Ressort „Gesundheitsinformation“ wird in Absprache mit der Institutsleitung und den entsprechenden Ressorts sicherstellen, dass die Inhalte der Gesundheitsinformations-Website sowie weiterer Gesundheitsinformationen

- evidenzbasiert sind und dem jeweils aktuellen wissenschaftlichen Kenntnisstand entsprechen,
- sich am Informationsbedarf der Bürger und Patienten orientieren und soweit wie möglich an deren psychologische Bedürfnisse angepasst sind,
- mit allen anderen vom Institut veröffentlichten Informationen kongruent sind.

Interne und externe Kritik an Inhalt und Qualität der Gesundheitsinformationen wird an den zuständigen Redakteur der Gesundheitsinformationen weitergeleitet. Diesem obliegt es, angemessen auf die Kritik zu reagieren und das Steuergremium regelmäßig darüber in Kenntnis zu setzen. Bei

Problemen größeren Umfangs oder bei Dringlichkeit wird die Institutsleitung unmittelbar eingeschaltet.

Eine enge Zusammenarbeit zwischen dem zuständigen Redakteur der Gesundheitsinformationen, der Institutsleitung, dem Ressort „Kommunikation“ und den jeweils betroffenen anderen Ressorts gewährleistet die inhaltliche Konsistenz aller Veröffentlichungen des Instituts. Die endgültige Entscheidung über die Veröffentlichung liegt in der Verantwortung des Steuerungsausschusses.

3.2.3 Multilingualität

Das Institut ist bemüht, seine Gesundheitsinformationen sowohl in deutscher als auch in englischer Sprache zu publizieren und beide Versionen auf dem aktuellen Stand zu halten. Erst durch die englische Version sind ein internationaler weitreichender Austausch und damit eine bestmögliche Qualitätssicherung zu erreichen. Zudem können die Gesundheitsinformationen vom Feedback internationaler Forscher und Reviewer einschließlich der Autoren systematischer Übersichten profitieren.

Das Institut wird darüber hinaus mit Partnern zusammenarbeiten, um einige seiner Gesundheitsinformationen in diejenigen Sprachen übersetzen zu lassen, die in Deutschland am häufigsten gesprochen werden.

Da es sehr schwierig ist, die Übersetzungsqualität nach objektiven Kriterien zu beurteilen, kann hierfür kein spezifischer Standard definiert werden. Es kann wortwörtlich oder sinngemäß (mit dem Ziel, die Intention des Originals in der Zielsprache exakt zu treffen) übersetzt werden. Beim Institut wird häufig auf die sinngemäße Übersetzungsmethode zurückgegriffen.

3.3 *Entwicklung der Informationsprodukte*

3.3.1 Themenauswahl

Die Auswahl der Themen für die Gesundheitsinformationen soll weitestgehend dem öffentlichen Interesse entsprechen. Sie soll ausgewogen und unparteiisch, transparent, klinisch relevant sowie begründet sein.

Da die Erstellung und die Pflege der Gesundheitsinformationen viele Ressourcen binden werden, sind stringente Methoden zur Ressourcenbündelung und Prioritätensetzung erforderlich [9-11]. Die Prioritätensetzung erfolgt systematisch unter Berücksichtigung interner und externer Beratung nach folgendem Schema:

a. Entwicklung von Filterkriterien (Haupt- und Zusatzkriterien)

Zu den Hauptkriterien zählen beispielsweise die Quantität und Qualität wissenschaftlicher Forschung zu dem Thema sowie das öffentliche Interesse an dem Thema. Zusatzkriterien sind beispielsweise die Größe der betroffenen Bevölkerungsgruppe, allgemeine Informations- und Bildungsaspekte, der aktuelle Wissensstand in der Bevölkerung sowie Annahmen über Auswirkungen auf den Gesundheitsstatus und mögliche Risiken (sowohl auf individueller als auch auf Bevölkerungsebene).

b. interne und externe Beratung der Filterkriterien im Laufe des Jahres 2006

c. Festlegung der Filterkriterien

d. Evaluation.

Anhand dieser Filterkriterien erfolgt die systematische Auswahl der Themen durch die Anwendung der Filterkriterien auf eine möglichst umfassende Themensammlung anhand eines Drei-Stufen-Modells:

1. Stufe: Anwendung der Hauptkriterien (grober Filter)
2. Stufe: Anwendung der Zusatzkriterien für ausführliche Informationen auf die in der ersten Stufe ausgewählten Themen
3. Stufe: interne und externe Beratung über die in der zweiten Stufe ausgewählten Themen für ausführliche Informationen

3.3.2 Fragestellung und Inhalte

Es gibt eine Reihe von Methoden, die als Entscheidungshilfe für die im Rahmen einer Gesundheitsinformation zu berücksichtigenden Fragestellungen dient. Diese Methoden variieren bezüglich ihrer verursachenden Kosten, ihrer Praktikabilität sowie der Möglichkeiten, ihre Ergebnisse auf andere Gesundheitsinformationen zu übertragen [12,13]. Sofern möglich, wird das Institut qualitativ hochwertige Daten, Umfragen und Studien nutzen, die mitunter um Telefoninterviews mit Key Informants und/oder Fokusgruppen ergänzt werden (siehe auch Abschnitt 1.24). Die Entscheidungsfindung des Instituts erfolgt dabei stets unter der Prämisse, dass die Fragestellungen im Interesse der allgemeinen Öffentlichkeit sind. Besondere Aufmerksamkeit gilt den Bedürfnissen benachteiligter Bevölkerungsgruppen.

Zur Bearbeitung der einzelnen Gesundheitsinformationen können interne Projektgruppen gebildet werden. Diese werden sich aus Mitarbeitern des Ressorts „Gesundheitsinformation“ sowie anderer Ressorts zusammensetzen. Die Bildung der Projektgruppe sowie die interne Projektkoordination werden vom Ressort „Gesundheitsinformation“ übernommen.

- a. In der ersten Sitzung der Projektgruppe werden die Fragestellungen und die relevanten thematischen Inhalte der zu erstellenden Gesundheitsinformation erarbeitet.
- b. Es erfolgt eine Bestandsaufnahme des derzeitigen Versorgungs- und Informationssystems für Bürger und Patienten zu diesem Thema (zum Beispiel Diagnoseverfahren und Therapien). Hierfür werden eine Literaturrecherche, die Recherche von so genannten Schlüssel-Websites (die wichtigsten Websites mit Informationen zu Gesundheitsthemen), Telefoninterviews sowie die Befragung von Key Informants, z.B. Patientenvertretern und klinischen Experten, durchgeführt.
- c. Die Bedürfnisse der Bürger und Patienten, ihr aktueller Kenntnisstand sowie ihre potenziellen Interessen hinsichtlich des Themas werden so weit wie möglich analysiert. Hierfür werden wiederum eine Literaturrecherche einschließlich der Recherche nach qualitativer Forschung und nach Websites mit Informationen/Daten über Patientenerfahrungen [14], Telefoninterviews mit Key Informants und gegebenenfalls Fokusgruppen durchgeführt.
- d. Daraus erwächst ein Entwurf hinsichtlich der Inhalte der Gesundheitsinformation, der in der zweiten Sitzung der Projektgruppe besprochen wird. Der Entwurf enthält die zentralen Fragen sowie die Gliederung. Neben der Besprechung der Literaturrecherche wird das weitere Vorgehen festgelegt.

3.3.3 Erstellung

Grundsätzlich werden für die Erstellung der Gesundheitsinformationen dieselben relevanten evidenzbasierten Methoden angewendet wie im Institut insgesamt (siehe Abschnitte 1.23 und 2.7). Entsprechend dem Entwurf werden die einzelnen Informationen wie folgt entwickelt.

Ausführliche Informationen

Die Erstellung von ausführlichen Informationen erfolgt entsprechend den folgenden Schritten:

- a. Literaturrecherche nach publizierten systematischen Übersichten und publizierter qualitativer Forschung:
Die Validität und Aktualität der systematischen Übersichten werden diskutiert und bewertet. Erstellung einer vorläufigen Version der ausführlichen Informationen aus den Ergebnissen der Literaturrecherche.
- b. Internes Peer Review:
Mehrfache Durchsicht der vorläufigen Version. Das resultierende Feedback bzw. Kritik wird diskutiert, gegebenenfalls erfolgen eine weitere Analyse und Durchsicht.

c. Externes Peer Review:

Expertengutachten werden von Patienten/Patientenvertretern sowie klinischen Experten eingeholt. Leitende Autoren von wichtigen systematischen Übersichten erhalten zudem die Möglichkeit, das Manuskript zu kommentieren (üblicherweise in der englischen Version). Bei Bedarf werden auch relevante verantwortliche Institutionen herangezogen.

d. Falls erforderlich, werden die Lesbarkeit und Verständlichkeit der ausführlichen Informationen durch Vertreter der Zielgruppe getestet (deutsche Version).

e. Die überarbeitete Version der ausführlichen Informationen wird in deutscher und englischer Sprache verfasst und an das Steuergremium des Instituts geleitet. Die Informationen werden dann zu einer zeitlich begrenzten externen Stellungnahme verschickt. Anschließend werden sie als Internettestversion aufbereitet.

Die Anwendbarkeit der (deutschen) Online-Version kann mit Freiwilligen einschließlich mindestens einer Patientin oder eines Patienten bzw. einer Patientenvertreterin oder eines Patientenvertreters getestet werden [15-17]. Danach wird die englische Version als Internettestversion (offline) aufbereitet.

f. Werden anhand des Quellenmaterials mehrere Informationsversionen oder -produkte entwickelt, ist es notwendig, dass diese auch einen Qualitätssicherungsprozess durchlaufen. Dies trifft nur zu, wenn der Inhalt in wesentlichen Teilen verändert werden soll.

g. Das Steuergremium gibt die endgültige Version frei oder schlägt alternativ eine weitere Beratung oder Überarbeitung vor.

Kurzinformationen

Die Erstellung der Kurzinformationen erfolgt anhand einer Zusammenfassung von kurzen Einzelarbeiten, einer einzelnen systematischen Übersicht oder wichtigen Studie oder einiger weniger Übersichten oder Studien. Sie durchlaufen einen Entwicklungsprozess, wie er in den Schritten c bis h (siehe oben) beschrieben worden ist.

Dem Ressort „Gesundheitsinformation“ obliegt die Verantwortung für die Pflege und Aktualisierung der Kurzinformationen.

Erfahrungsberichte

Viele Patienten möchten über die Erfahrungen von anderen Betroffenen mit derselben Erkrankung hören oder lesen [18]. Erfahrungsberichte sind sowohl im Journalismus als auch auf dem Gebiet der Patienteninformation eine übliche Form der Informationsvermittlung. Die Bedeutung von Er-

fahrungsberichten in der medizinischen Praxis und im Gesundheitswesen wird auch zunehmend gewürdigt [19,20].

Mit der Erhebung von Erfahrungsberichten von Patienten sowie deren Angehörigen sollen individuelle Erfahrungen und Erlebnisse von Krankheit und Pflegebedürftigkeit in ihren verschiedenen Dimensionen anderen Betroffenen beziehungsweise Interessierten als Ergänzung zu den evidenzbasierten Gesundheitsinformationen zur Verfügung gestellt werden. Die Erfahrungsberichte sollen jedoch keinen Widerspruch zu den evidenzbasierten Gesundheitsinformationen darstellen.

Die Erhebung der Erfahrungsberichte erfolgt auf der Basis folgender Schritte:

1. Suche nach Interviewpartnern
2. Einholen der Einverständniserklärung der Interviewpartner bezüglich der Durchführung und Nutzung des Interviews
3. Durchführung der Interviews
4. Dokumentation und Einholen der Einverständniserklärung für die Veröffentlichung der endgültigen Version
5. Veröffentlichung auf der Website mit Einwilligung der Interviewpartner.

Besonderer Wert wird sowohl auf die umfassende Aufklärung vor dem Interview, auf eine jederzeit widerrufbare Einverständniserklärung zur Veröffentlichung, auf eine ausführliche Vorbereitung der Interviews, auf ein an vorher festgelegten Kriterien der Gesprächsführung orientiertes Vorgehen der Interviewer sowie auf eine anonymisierte Transkription gelegt.

3.3.4 Evidenzbasis und Kommunikationsstandards

Das Institut hat bei der Kommunikation der Gesundheitsinformationen folgenden Anspruch:

- respektvoll und effektiv mit der deutschen Bevölkerung zu kommunizieren, so dass diese dem Institut als zuverlässige und allgemeinverständliche Informationsquelle vertraut
- die Gesundheitsinformationen sollen allgemeinverständlich sein, ohne dabei die wissenschaftliche Genauigkeit zu vernachlässigen
- der Kommunikationsstil soll so neutral und so eindeutig wie möglich sein
- Sensibilität und Respekt gegenüber dem Wissen der Bürger und Patienten, ihren Wertvorstellungen und Anliegen, ihrem Selbstbestimmungsrecht und gegenüber kulturellen Besonderheiten sollen klar und deutlich erkennbar sein
- die Patientenkompetenz (Patient Empowerment) soll gestärkt werden

- die Fähigkeit von Bürgern und Patienten, wissenschaftliche Literatur zu verstehen, soll gefördert werden
- dem Einzelnen soll geholfen werden, den Stand der Wissenschaft (Evidenz) auf seine eigene, individuelle Situation zu übertragen
- respektvoller Umgang mit der Zeit der Leser.

Das Institut wird einen behindertengerechten Zugang zur Website entsprechend den Standards der Accessibility Initiative gewährleisten [18].

Die Gesundheitsinformationen sollen einerseits nicht übertrieben wissenschaftlich klingen, andererseits aber auch nicht den Leser mit direktiven Ratschlägen konfrontieren. Wie die Menschen darauf reagieren, wenn sie sich der mit vielen medizinischen Maßnahmen verbundenen wissenschaftlichen Unsicherheit bewusst werden, ist weitgehend unbekannt. Außerdem ist die Allgemeinheit an eine direktivere Art von Gesundheitsinformation gewöhnt, die häufig darauf zielt, ihre Meinung oder ihr Verhalten direkt zu ändern. Das Institut strebt an, die Informationen auf vielfältige Weise zu präsentieren, um möglichst vielen Bürgern und Patienten Zugang zu den Informationen zu ermöglichen [7].

Gesundheitsinformationen, die allein relative Risiken angeben, sollen in der Regel vermieden werden. Dennoch kann die Angabe relativer Risiken gelegentlich notwendig sein, um den Einzelnen in die Lage zu versetzen, Behandlungen vergleichen zu können. Relative Risiken werden dann aber nicht alleine aufgeführt, sondern wenn möglich zusammen mit absoluten Risiken oder der Anzahl der Patienten, die behandelt werden müssen, damit eine Patientin oder ein Patient von der Behandlung profitiert bzw. durch die Behandlung einen Schaden erleidet.

Personalisierte bzw. individualisierte Risikoabschätzungen sind als effektive Kommunikationsformen für Gesundheitsinformationen nachgewiesen [7,22]. Das Ressort „Gesundheitsinformation“ kann Tools entwickeln oder adaptieren, mit deren Hilfe die Bürger und Patienten ihr persönliches Risiko einschätzen können, wenn sehr verlässliche Daten für deren Entwicklung vorliegen.

„Patient Decision Aids“ (z.B. visualisierte Risikodarstellungen) sind effektive Maßnahmen zur Vermittlung von Gesundheitsinformationen, und sie unterstützen Patientenentscheidungen [23]. Das Institut wird gegebenenfalls Decision Aids entwickeln oder adaptieren und versuchen, effektive Elemente der Decision Aids in die Gesundheitsinformationen einzubauen.

Zusätzlich wird das Ressort „Gesundheitsinformation“

- Gesundheitsinformationen in konsistenter Weise präsentieren,
- den Grad der Unsicherheit erklären, der mit den Informationen verbunden sein kann,

- aufzeigen, für wen (das heißt, für welche Bevölkerungsgruppe/n) die dargestellten wissenschaftlichen Nachweise gelten,
- versuchen, für die Website höchstmögliche Standards für Benutzerfreundlichkeit (inklusive Navigation) zu erfüllen,
- sehr klar und sorgfältig auf den Unterschied zwischen „fehlendem wissenschaftlichen Nachweis eines Unterschieds“ (absence of evidence) und dem „wissenschaftlichen Nachweis, dass kein Unterschied existiert“ (evidence of no effect) hinweisen,
- unbedingt vermeiden, Informationen bezüglich der Produkte eines bestimmten Unternehmens voreingenommen darzustellen; es werden grundsätzlich Substanznamen für die Produkte verwendet, und nur falls notwendig durch die Markennamen erhältlicher Produkte ergänzt,
- das durch Hypertext verlinkte Online-Glossar ausbauen, so dass die Verständlichkeit der Gesundheitsinformationen auf der Website erhöht wird.

3.4 Publikation

Die Distribution der Gesundheitsinformationen wird mit der Institutsleitung, fachlich zuständigen Ressortleitern und dem Ressort „Kommunikation“ besprochen. Änderungsvorschläge werden an das Steuergremium weitergeleitet. Die Institutsleitung, die Leiterin oder der Leiter des Ressorts „Gesundheitsinformation“ und die Sprecherin oder der Sprecher des Institutes stellen die inhaltliche Konsistenz bei der Kommunikation nach außen sicher.

Die Gesundheitsinformationen werden als ausführliche Information, Kurzinformation oder sonstiges Produkt vorwiegend auf der Gesundheitsinformations-Website publiziert (siehe auch Abschnitt 3.2).

3.5 Evaluation und Aktualisierung

Es gibt zahlreiche Instrumente und Leitlinien zur Qualitätsbewertung und -sicherung von Gesundheitsinformationen im Internet. Es gibt jedoch keine verlässlichen Daten über die Validität solcher Instrumente und Leitlinien [21-24]. Somit steht kein Instrument zur Erstellung von Gesundheitsinformationen zur Verfügung, das als ein zuverlässiger Qualitätsindikator angesehen werden kann [22,24]. Zudem ist keine evidenzbasierte Aussage zur Kosteneffizienz bei der Erstellung von Gesundheitsinformationen möglich.

Einige Sachverhalte, die im Rahmen der wissenschaftlichen Bewertung von Gesundheitsinformationen als wichtig erachtet werden, haben in der Praxis nur geringe Relevanz. Dies zeigen Studien

aus der Patientenforschung – auch mit deutschen Patienten [25,26]. Einige der in Instrumenten und Leitlinien zur Bewertung von Informationen geläufigen Empfehlungen können die wissenschaftliche Qualität der Informationen sogar herabsetzen.

Die meisten Bewertungsinstrumente konzentrieren sich auf die Darstellung nachgewiesener Therapieeffekte und bieten keine vollständige Bewertung eines gesamten Krankheitsbildes einschließlich Ätiologie, Prognose, Screening und diagnostischer Testverfahren. Gesundheitsinformationen zu Diagnose- und Screeningverfahren beinhalten jedoch komplexere Entscheidungsprozesse und Kommunikationsaspekte als Informationen zu Behandlungen [6,7,19].

Aus diesem Grunde werden sich die Gesundheitsinformationen des Instituts nicht auf die gegenwärtig vorliegenden Bewertungsinstrumente stützen. Es wird auch kein eigenes Bewertungsinstrument entwickelt. Das Institut wird jedoch eigene Studien über bestimmte Aspekte, die die Entscheidungen von Patienten und nachweislich die Qualität von Informationen beeinflussen könnten, durchführen und die Gesundheitsinformationen an den jeweiligen Ergebnissen ausrichten.

Das Ressort „Gesundheitsinformation“ wird neue wissenschaftliche Erkenntnisse über die Möglichkeiten der Qualitätsbewertung verfolgen und seine Methoden gegebenenfalls anpassen. Des Weiteren werden die Gesundheitsinformationen des Instituts so weit wie möglich von repräsentativen Bevölkerungsgruppen überprüft.

Zur Gewährleistung der Aktualität wird die für die einzelnen Gesundheitsinformationen verwendete Literatur kodiert. Dadurch wird es möglich, dass die Veröffentlichung wichtiger neuer Evidenz sowie Änderungen in Cochrane-Übersichten offenkundig werden. Die Aktualität der Gesundheitsinformationen wird auch dadurch gewährleistet, dass es einen kontinuierlichen Informationsaustausch zwischen den Schlüsselpersonen der Cochrane Collaboration, des Centre for Reviews and Dissemination und der Zeitschrift „Evidence Based Medicine“ gibt.

Die Aktualität der Gesundheitsinformationsprodukte wird des Weiteren durch eine Überarbeitung mindestens alle zwei Jahre sichergestellt.

Die Gesundheitsinformations-Website wird für die Leser Feedbackmechanismen bereitstellen. Jegliche Art von internem und externem Feedback kann zu unmittelbarer Überarbeitung bereits veröffentlichter beziehungsweise zu einer Überarbeitung der Entwürfe von Gesundheitsinformationen führen.

Die enge inhaltliche Kooperation des Ressorts „Gesundheitsinformation“ mit fachlich relevanten Ressorts gewährleistet, dass die erstellten Gesundheitsinformationen dem wissenschaftlichen Stand entsprechen. Seinem Anspruch an den Qualitätsprozess wird das Ressort „Gesundheitsinformation“ dadurch gerecht, dass es auf interne Fachkompetenz baut, die Key Informants einbezieht und den

Autoren systematischer Übersichten die Möglichkeit bietet, zur Interpretation ihrer Arbeiten Stellung zu beziehen.

Literatur

- [1] Hope T. Evidence-Based Patient Choice. London: King's Fund; 1996.
- [2] Bessell TL, McDonald S, Silagy CA, Anderson JN, Hiller JE, Sansom LN. Do internet interventions for consumers cause more harm than good? A systematic review. *Health Expect* 2002; 5: 28-37.
- [3] National Institute of Clinical Studies. The Impact of the Internet on Consumers Health Behaviour. Melbourne: NICS; 2003.
<http://www.nicsl.com.au/asp/download.asp?media=/data/portal/00000005/content/15746001153874589315.pdf> [Zugriff am 17.11.2006].
- [4] Lewis D. Computer-based approaches to patient education: A review of the literature. *J Am Med Inform Assoc* 1999; 6: 272-282.
- [5] Coulter A, Entwistle V and Gilbert D. Informing Patients: An Assessment of the Quality of Patient Information Materials. London: King's Fund Publishing; 1998.
- [6] Entwistle VA, Watt IS, Davis H, Dickson R, Pickard D, Rosser J. Developing information materials to present the findings of technology assessments to consumers: The experience of the NHS Centre for Reviews and Dissemination. *Int J Tech Assess Health Care* 1998; 14: 47-70.
- [7] Edwards A, Bastian H. Risk communication – making evidence part of patient choices? In: Edwards A, Elwyn G, editors. Evidence-Based Patient Choice: Inevitable or Impossible? Oxford: Oxford University Press; 2001: 144-160.
- [8] Eysenbach G, Jadad AR. Consumer health informatics in the Internet age. In: Edwards A, Elwyn G, editors. Evidence-Based Patient Choice: Inevitable or Impossible? Oxford: Oxford University Press; 2001: 289-307.
- [9] Henshall C, Oortwijn W, Stevens A, Granados A, Banta D. Priority setting for health technology assessment. Theoretical considerations and practical approaches. A paper produced for the EUR-ASSESS project. *Int J Technol Assess Health Care* 1997; 13: 144-185.
- [10] Townsend J, Buxton M, Harper G. Prioritisation of health technology assessment. The PATHS model: Methods and case studies. *Health Technol Assess* 2003; 7: iii, 1-82.
- [11] Ghaffar A, de Francisco A, Matlin S. The Combined Approach Matrix: A Priority-setting Tool for Health Research. Geneva: Global Forum for Health Research; 2004.
<http://www.globalforumhealth.org/filesupld/90.pdf> [Zugriff am 17.11.2006].
- [12] Liberati A, Sheldon TA, Banta HD. EUR-ASSESS Project Subgroup report on Methodology. Methodological guidance for the conduct of health technology assessment. *Int J Tech Assess Health Care*, 1997; 13: 186-219.
- [13] Sixma HJ, Kerssens JJ, Campen CV, Peters L. Quality of care from the patients' perspective: from theoretical concept to a new measuring instrument. *Health Expec* 1998; 1: 82-95.
- [14] DiPEX. The Database of Patients' Experiences.
<http://www.dipex.org> [Zugriff am 18.10.2004].
- [15] Krug S. Don't make me think: A common sense approach to web usability. Indiana: New Riders; 2000.

- [16] Nielsen J. Designing web usability. Indiana: New Riders; 2000.
- [17] Inan H. Measuring the success of your website: A customer-centric approach to website management. Sydney: Pearson Educational Australia; 2002.
- [18] Swift TL, Dieppe PA. Using expert patients' narratives as an educational resource. *Patient Educ Couns* 2005; 57: 115-121.
- [19] Greenhalgh T, Hurwitz B. Narrative based medicine: Why study narrative? *BMJ* 1999, 318: 48-50.
- [20] Steiner JF. The use of stories in clinical research and health policy. *JAMA* 2005; 294: 2901-2904.
- [21] W3C. Web Accessibility Initiative. <http://www.w3.org> [Zugriff am 26. Januar 2005]
- [22] Edwards A, Unigwe S, Elwyn G, Hood K. Personalised risk communication for informed decision making about entering screening programs (Cochrane Review). In: *The Cochrane Library*, Issue 3, 2004. Chichester: Wiley; 2004.
- [23] O'Connor AM, Stacey D, Entwistle V, Llewellyn-Thomas H, Rovner D, Holmes-Rovner M, Tait V, Tetroe J, Fiset V, Barry M, Jones J. Decision aids for people facing health treatment or screening decisions (Cochrane Review) (last updated October 2003). In: *The Cochrane Library*, Issue 3, 2004. Chichester: Wiley; 2003.
- [24] Jadad AR, Gagliardi A. Rating health information on the internet: navigating to knowledge or to Babel? *JAMA* 1998; 279: 611-614.
- [25] Eysenbach G. Consumer health informatics. *BMJ* 2000; 320: 1713-1716.
- [26] Gagliardi A, Jadad AR. Examination of instruments used to rate quality of health information on the internet: chronicle of a voyage with an unclear destination. *BMJ* 2002; 324: 569-573.
- [27] Moulton B, Franck LS, Brady H. Ensuring quality information for patients: development and preliminary validation of a new instrument to improve the quality of written health information. *Health Expect* 2004; 7: 165-175.
- [28] van den Brink-Muinen A, Verhaak PF, Bensing JM, Bahrs O, Deveugele M, Gask L, Leiva F, Mead N, Messerli V, Oppizzi L, Peltenburg M, Perez A. Doctor-patient communication in different European health care systems: Relevance and performance from the patients' perspective. *Patient Educ Counsel* 2000; 39: 115-127.
- [29] Eysenbach G, Kohler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *BMJ* 2002; 324: 573-577.

4. Berichtserstellung

4.1 Produkte

Gemäß seinem gesetzlichen Auftrag erstellt das Institut verschiedene Produkte in Form von wissenschaftlichen Berichten und allgemeinverständlichen Informationen für Bürger und Patienten (siehe auch Präambel). Bei der Erstellung dieser Produkte wird die in diesem Dokument beschriebene Methodik zu Grunde gelegt. Für die wissenschaftlichen Arbeiten zur Nutzenbewertung sind zwei Formate möglich:

- ausführliche **Berichte** zur Nutzenbewertung,
- Schnellberichte („**Rapid Reports**“), die zwei Zielsetzungen haben:
 - kurzfristige Information über relevante (zum Beispiel neue) Technologien oder
 - zeitnahe Information zu tagesaktuellen Themen zu liefern; hierbei erfolgt bei Bedarf eine Abstimmung mit weiteren verantwortlichen öffentlichen Institutionen (zum Beispiel dem Bundesinstitut für Arzneimittel und Medizinprodukte, BfArM).

Darüber hinaus ist das Institut vom G-BA mit einem Generalauftrag beauftragt worden:

„Durch die Erfassung und Auswertung des relevanten Schrifttums eine kontinuierliche Beobachtung und Bewertung medizinischer Entwicklungen von grundlegender Bedeutung und ihrer Auswirkungen auf die Qualität und Wirtschaftlichkeit der medizinischen Versorgung in Deutschland vorzunehmen und den G-BA hierüber regelmäßig zu informieren. Der G-BA geht bei diesem Auftrag davon aus, dass das Institut auf den ihm gemäß § 139a Abs. 3 SGB übertragenen Arbeitsfeldern nicht nur Einzelaufträge des G-BA bearbeitet, sondern aus der eigenverantwortlichen wissenschaftlichen Arbeit heraus dem G-BA für dessen gesetzliche Aufgaben notwendige Informationen über versorgungsrelevante Entwicklungen in der Medizin zur Verfügung stellt und konkrete Vorschläge für Einzelaufträge erarbeitet, die aus Sicht des Instituts vor dem Hintergrund dieser Informationen relevant sind.“

Wissenschaftliche Arbeiten im Rahmen dieses Generalauftrags werden als „**Arbeitspapiere**“ veröffentlicht.

Die Abläufe für die Erstellung der verschiedenen Berichtstypen sind den verschiedenen Zielsetzungen der Dokumente angepasst und werden in den Abschnitten 4.4 und 4.5 beschrieben.

Arbeitspapiere werden analog zum Ablauf der Rapid Reports erstellt. Für die Erstellung von allgemeinverständlichen Informationen („**Gesundheitsinformationen**“) für Bürger und Patienten gelten die in Kapitel 3 dargestellten methodischen Anforderungen.

In nachfolgender Tabelle sind die verschiedenen Produkte des Institutes in einer Übersicht dargestellt.

Übersicht der Produkte des Instituts

Produkt	Zielsetzung	Ablauf	Beauftragung durch
Bericht	ausführliche Nutzenbewertung einschließlich öffentlicher schriftlicher Anhörung, u.a. für Richtlinienentscheidungen	beschrieben in Abschnitt 4.4	G-BA, BMG
Rapid Report	kurzfristige Information bei besonderer Dringlichkeit; Sofortinformation zu tagesaktuellen Themen	beschrieben in Abschnitt 4.5	G-BA, BMG
Arbeitspapier	Information über versorgungsrelevante Entwicklungen in der Medizin	beschrieben in Abschnitt 4.5	Selbstinitiierung durch das Institut
Gesundheitsinformation	allgemeinverständliche Information für Bürger und Patienten, breit gefächertes Themenspektrum	beschrieben in Kapitel 3	G-BA, BMG; auch Selbstinitiierung durch das Institut

4.2 Auswahl externer Sachverständiger

Gemäß seinem gesetzlichen Auftrag bezieht das Institut externe Sachverständige in seine Arbeit ein. Als externe Sachverständige gelten dabei Personen, an die wissenschaftliche Forschungsaufträge im Rahmen der Erstellung von Institutsprodukten oder der Begutachtung von Institutsprodukten vergeben werden.

Die Vergabe von wissenschaftlichen Forschungsaufträgen erfolgt nach den eigenen Vergabe- und Beschaffungsrichtlinien des Institutes, denen die Bestimmungen des § 22 der Verordnung zum Haushaltswesen in der Sozialversicherung (SVHV) sowie die gesetzlichen Vorschriften des Vergaberechts zugrunde liegen. Das Institut informiert in Fachzeitschriften darüber, dass es auf seiner Internetseite (<http://www.iqwig.de>) in regelmäßigen Abständen die an externe Sachverständige zu vergebenden Projekte veröffentlicht. Die aktuell durch das IQWiG zu vergebenden Aufträge sind

auf einer Internetseite des Instituts (<http://www.iqwig.de/index.174.html>) aufgelistet. Aufträge mit einem Auftragsvolumen von über 200.000 EUR werden europaweit ausgeschrieben.

Das Institut gibt folgende Eignungskriterien vor, die interessierte Personen oder Arbeitsgruppen zwingend erfüllen und darlegen müssen:

- Die fachliche Unabhängigkeit, das heißt, die Bieter müssen auftragsbezogen ihre potenziellen Interessenkonflikte darlegen (vgl. § 139 b Abs. 3, S. 2 SGB V). Dies geschieht anhand des Formulars „Darlegung potenzieller Interessenkonflikte“ und der „Verpflichtungserklärung“.
- die Beherrschung der deutschen Sprache
- jeweils auftragsbezogen Referenzen bzw. Erfahrungen medizinisch-fachlicher Art
- jeweils auftragsbezogen Referenzen bzw. Erfahrungen methodischer Art, d.h. insbesondere Darlegungen zur Fähigkeit, nach den Grundsätzen der aktuellen Methoden des Institutes zu arbeiten.

Die notwendigen Formulare („Anschreiben“, „Darlegung potenzieller Interessenskonflikte“ und „Verpflichtungserklärung“) sind im Internetauftritt des IQWiG (<http://www.iqwig.de/index.267.html>) abrufbar.

Die eingegangenen Unterlagen werden durch ein Gremium des Instituts geprüft, das sich aus

- einem Mitglied der Institutsleitung,
- einem Beschäftigten des Ressorts „Verwaltung“ (Ressortleiter/in oder Justitiar/in),
- einem Ressortleiter/einer Ressortleiterin aus einem wissenschaftlichen Ressort, das mit dem konkret zu vergebenden Auftrag nicht befasst ist, und
- dem Ressortleiter/der Ressortleiterin, in dessen/deren Ressort der zu vergebende Auftrag fällt

zusammensetzt.

Die fachliche Unabhängigkeit der potenziellen Auftragnehmer wird durch ein dreiköpfiges Gremium, bestehend aus

- einem Mitglied der Institutsleitung,
- dem Leiter/der Leiterin eines wissenschaftlichen Ressorts und
- einem wissenschaftlichen Mitarbeiter/einer wissenschaftlichen Mitarbeiterin,

projektbezogen beurteilt.

Dabei werden diejenigen Bieter ausgewählt, die zur weiteren Teilnahme am Verhandlungsverfahren aufgefordert werden. Mit der Aufforderung werden Kriterien benannt, nach denen der Bieter ausgewählt wird, der die bestmögliche Leistung erwarten lässt. Die Bewertung dieser Kriterien

kann unter anderem durch die Darlegung der entsprechenden Vorgehensweise zu einem Referenzprojekt oder durch konkrete Ausführungen zu wesentlichen Teilbereichen des zu vergebenden Projektes erfolgen. Darüber hinaus wird sich das Auswahlgremium in einem Verhandlungsgespräch durch die Vorstellung des jeweiligen Projektleiters beziehungsweise gegebenenfalls Projektteams ein persönliches Bild von den ausgewählten Bietern machen. Externe Sachverständige, die die vereinbarte Arbeit oder Teile davon durch Dritte erfüllen lassen wollen, müssen diese verbindlich benennen. Für diese gelten insbesondere hinsichtlich ihrer potenziellen Interessenkonflikte dieselben Maßgaben wie für die externen Sachverständigen selbst.

Anschließend wird derjenige Bieter ausgewählt, der die bestmögliche Leistung erwarten lässt. Die einzelnen Schritte dieses Bewertungsprozesses werden dokumentiert.

Über die Vertragsbestandteile, insbesondere über die wissenschaftlichen Inhalte oder Arbeits(teil)ergebnisse ist bis zur Abnahme gegenseitiges Stillschweigen zu vereinbaren. Die externen Sachverständigen verpflichten sich zusätzlich zur Einhaltung gegebenenfalls anzuwendender datenschutzrechtlicher Bestimmungen.

4.3 Gewährleistung der fachlichen Unabhängigkeit

4.3.1 Zielsetzung

Die wissenschaftliche und fachliche Unabhängigkeit des Institutes und der von ihm vertretenen und veröffentlichten Inhalte ist gesetzlich in § 139a SGB V sowie in der Satzung der Institutsstiftung verankert. Der Begriff „Unabhängigkeit“ kann dabei naturgemäß nur annähernd real Anwendung finden, da die Einschätzung der fachlichen Unabhängigkeit individuell unterschiedlich sein kann. Insofern „Unabhängigkeit“ nur ein relatives Ziel sein kann, gilt es gleichzeitig, den Begriff „Transparenz“ einzuführen, um etwaige Entscheidungsprozesse und deren Ergebnisse vor dem Hintergrund einer „relativen Unabhängigkeit“ nach außen und innen nachvollziehbar zu gestalten.

4.3.2 Gewährleistung der externen fachlichen Unabhängigkeit

Vor dem Abschluss eines Vertrages zwischen dem Institut und einer oder einem externen Sachverständigen oder einer externen Institution zur fachlichen Beratung, Studiendurchführung oder Erstellung eines Berichtes muss entschieden werden, ob Bedenken wegen potenzieller Interessenkonflikte bestehen. Dafür müssen alle externen Sachverständigen und Institutionen alle Tätigkeiten darlegen, die potenziell geeignet sein könnten, ihre fachliche Unabhängigkeit zu beeinflussen (Interessenkonflikte, siehe auch Abschnitt 4.2). Als Interessenkonflikte werden dabei in Anlehnung an entsprechende Regelungen wissenschaftlicher medizinischer Zeitschriften insbesondere folgende

Kriterien angesehen: alle finanziellen Abmachungen, Beschäftigungen, Beratungen, Honorare, bezahlten Expertenaussagen, Reiseunterstützungen, Patentanträge sowie der Aktienbesitz innerhalb der letzten drei Jahre, die bzw. der die beauftragte Arbeit inhaltlich hätte/n beeinflussen können, sowie alle aktuellen persönlichen Beziehungen mit anderen Personen oder Organisationen, die die beauftragte Arbeit inhaltlich beeinflussen könnten [1]. Diese Kriterienliste findet sich auch in dem auf den Internetseiten des Instituts zur Verfügung gestellten Formblatt (siehe auch Abschnitt 4.2), das ggf. aktualisiert wird. Maßgeblich ist immer das jeweils auf den Internetseiten abrufbare Formblatt. Von allen externen Sachverständigen, die an der Erstellung von Berichten oder Rapid Reports beteiligt waren, erfolgt unter namentlicher Nennung mit dem Abschlussbericht bzw. Rapid Report eine Veröffentlichung der von ihnen dargelegten potenziellen Interessenkonflikte. Dies geschieht dergestalt, dass für die im oben genannten Formblatt aufgeführten Kriterien berichtet wird, ob zum jeweiligen Kriterium ein potenzieller Interessenkonflikt beschrieben wurde oder nicht. Eine Veröffentlichung z.B. des finanziellen Umfangs von Zuwendungen erfolgt nicht. Der Prozess zur Auswahl von externen Sachverständigen ist in Abschnitt 4.2 beschrieben.

4.3.3 Gewährleistung der internen fachlichen Unabhängigkeit

Die interne fachliche Unabhängigkeit wird – soweit möglich – durch die Auswahl der Mitarbeiter gewährleistet. Diese müssen im Rahmen ihrer Einstellung glaubhaft ihr bisheriges Tätigkeitsumfeld darlegen als auch versichern, (Neben)Tätigkeiten, die ihre fachliche Unabhängigkeit potenziell in Frage stellen könnten, im Rahmen einer Tätigkeit im Institut einzustellen. Den wissenschaftlichen Mitarbeitern des Institutes sind bezahlte Nebentätigkeiten, die im weitesten Sinne in einem beruflichen Kontext stehen können, untersagt. Grundsätzlich sind von allen Mitarbeitern alle Nebentätigkeiten gegenüber der Institutsleitung oder -verwaltung anzeigepflichtig. Zu Nebentätigkeiten gehören im weitesten Sinne auch unentgeltliche Ehrenämter wie Aufsichtsrats- oder Vorstandsämtler in Vereinen oder Gesellschaften. Zuwiderhandlungen können im Einzelfall zu einer Abmahnung bzw. im wiederholten oder schwerwiegenden Fall zu einer Kündigung des Arbeitsverhältnisses führen. Im Einzelfall entscheidet die Institutsleitung nach Beratung durch das Steuergremium, ob eine Mitarbeiterin oder ein Mitarbeiter von einer bestimmten Tätigkeit oder einem bestimmten Projekt aus Gründen der Befangenheit ausgeschlossen werden muss.

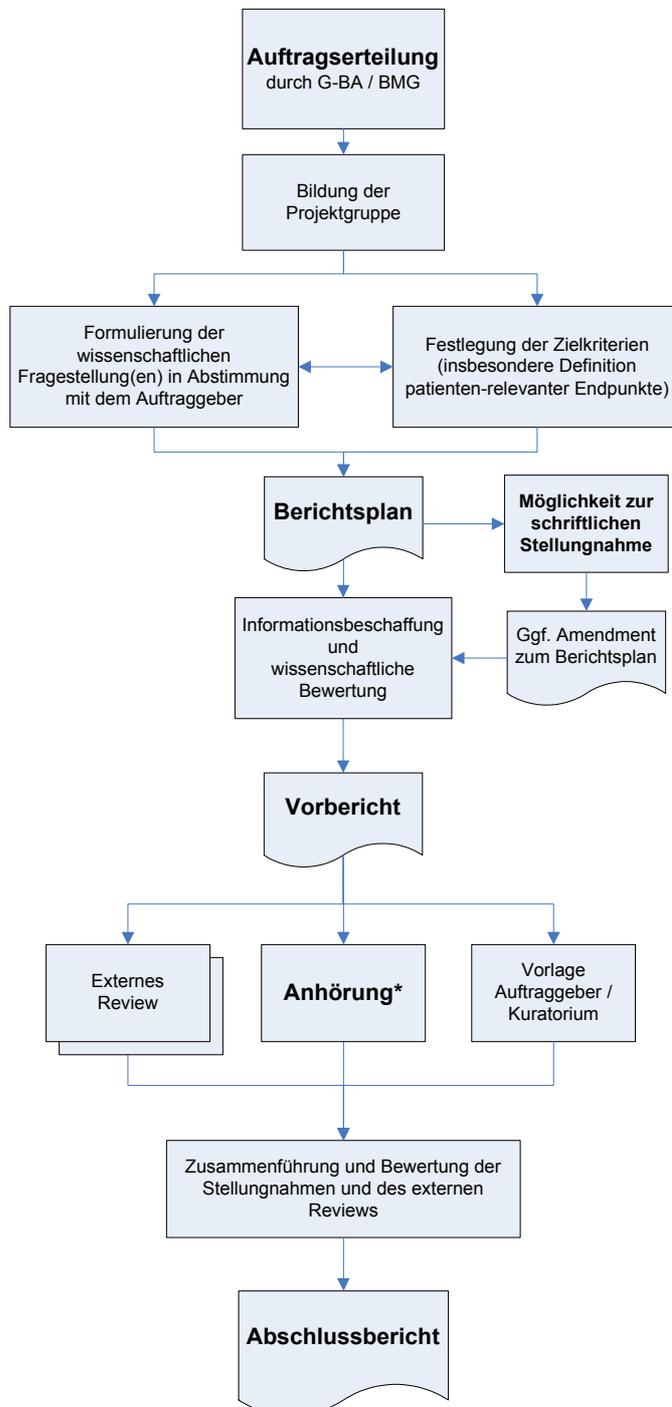
Literatur

- [1] A. James, R. Horton. The Lancet's policy on conflicts of interest. Lancet 2003; 361: 8-9.

4.4 Erstellung von Berichten

Der Ablauf der Berichtserstellung ist in Abbildung 2 schematisch dargestellt.

Abbildung 2: Ablauf der Berichtserstellung als Flussdiagramm



* Die Anhörung erfolgt mittels Einholung schriftlicher Stellungnahmen. Optional wird eine mündliche wissenschaftliche Erörterung zur Diskussion unklarer Aspekte in den schriftlichen Stellungnahmen durchgeführt.

Alle Arbeitsschritte werden in Verantwortung des Instituts getätigt. Dies kann, je nach Notwendigkeit, unter Einbeziehung externer Expertise einschließlich des wissenschaftlichen Beirats des Instituts erfolgen. Der interne Reviewprozess ist in diesem Flussdiagramm nicht dargestellt. Weitere Erläuterungen siehe nachfolgenden Text.

Nach **Auftragsvergabe** durch den G-BA oder das BMG erfolgt die Bildung einer Projektgruppe unter Leitung des beauftragten Ressorts. Eine Person wird als verantwortliche Projektleiterin oder verantwortlicher Projektleiter bestimmt. Die Zusammensetzung der Projektgruppe ist an dieser Stelle nicht fix, da sich durch die folgenden Arbeitsschritte Änderungsbedarf ergeben kann. Sofern notwendig, erfolgt anschließend eine Konkretisierung des Auftrags in Abstimmung mit den zuständigen Gremien des Auftraggebers. Dies beinhaltet auch die (grobe) Definition der Zielkriterien, insbesondere der patientenrelevanten Endpunkte. Sofern notwendig, wird im Weiteren durch die Projektgruppe, gegebenenfalls unter Einbeziehung externer Expertise und/oder Meinung, diese Definition verfeinert. Hierbei werden regelhaft einzelne Betroffene, Patientenvertretungs- und/oder Verbraucherorganisationen für die themenbezogene Definition patientenrelevanter Endpunkte einbezogen. Anschließend erfolgt die Erstellung des Berichtsplans. Dieser ist Grundlage für die darauf folgende Informationsbeschaffung und wissenschaftliche Bewertung.

Der **Berichtsplan** beinhaltet, vergleichbar mit dem Studienprotokoll einer geplanten klinischen Studie, die genaue wissenschaftliche Fragestellung einschließlich der Zielkriterien (z.B. patientenrelevante Endpunkte), die Ein- und Ausschlusskriterien der für die Nutzenbewertung zu verwendenden Informationen sowie die Darlegung der projektspezifischen Methodik der Beschaffung und Bewertung dieser Informationen. Darüber hinaus wird der Zeitplan bis zur Publikation des Abschlussberichts dargestellt. In die Erstellung des Berichtsplans können externe Sachverständige eingebunden werden. Der Berichtsplan wird in Verantwortung der Projektgruppe des IQWiG erstellt und dem Steuergremium zum internen Review vorgelegt. Nach Abschluss des internen Reviewverfahrens wird der Berichtsplan dann in der Regel zeitgleich

- dem Auftraggeber und Kuratorium, auch zur Frage der Vollständigkeit in Bezug auf den ursprünglich erteilten Auftrag, zugeleitet und
- auf den Internetseiten des Instituts mit dem Ziel einer Anhörung veröffentlicht.

Für die **Anhörung** mittels Einholung schriftlicher Stellungnahmen sind die im Berichtsplan getroffenen Angaben insbesondere zu den Ein- und Ausschlusskriterien relevanter Informationen (z.B. der einzubeziehenden wissenschaftlichen Literatur) von zentraler Bedeutung. Die Stellungnahmefrist beträgt in der Regel vier Wochen nach Veröffentlichung des Berichtsplans. Weiterführende Angaben zum Stellungnahmeverfahren finden sich auf den Internetseiten des Instituts in einem entsprechenden Leitfaden. Es gelten die in der aktuellen Version dieses Leitfadens genannten Bedingungen. Nach Auswertung der eingegangenen schriftlichen Stellungnahmen kann ein Amendement zum Berichtsplan erstellt und publiziert werden. Alle Änderungen des Berichtsplans werden im Vorbericht und im Abschlussbericht dargestellt.

Im **Vorbericht** werden die Ergebnisse der Informationsbeschaffung und der wissenschaftlichen Bewertung dargestellt. Er kann, zumindest in Teilen, durch externe Sachverständige erstellt werden, die unter Berücksichtigung der in Abschnitt 4.2 genannten Auswahlkriterien ihre Eignung dargelegt haben. Der Vorbericht wird in Verantwortung der Projektgruppe des IQWiG erstellt und dem Steuergremium zum internen Review vorgelegt. Nach Abschluss des internen Reviewverfahrens wird der Vorbericht dann in der Regel zeitgleich

- einem oder mehreren externen Gutachtern mit ausgewiesener methodischer und/oder fachlicher Kompetenz zugeleitet,
- dem Auftraggeber und Kuratorium, auch zur Frage der Vollständigkeit in Bezug auf den ursprünglich erteilten Auftrag, zugeleitet und
- auf den Internetseiten des Instituts mit dem Ziel einer Anhörung veröffentlicht.

Die **Anhörung** wird mittels Einholung schriftlicher Stellungnahmen durchgeführt. Hierbei sind die im Berichtsplan einschließlich etwaiger Amendments getroffenen Angaben insbesondere zu den Ein- und Ausschlusskriterien relevanter Informationen (z.B. der einzubeziehenden wissenschaftlichen Literatur) von zentraler Bedeutung. Die Stellungnahmefrist beträgt mindestens vier Wochen nach Veröffentlichung des Vorberichts.

Optional kann eine mündliche wissenschaftliche Erörterung mit Stellungnehmenden durchgeführt werden. Diese Erörterung dient der gegebenenfalls notwendigen Klarstellung von in den schriftlichen Stellungnahmen vorgebrachten Inhalten mit dem Ziel der Verbesserung der wissenschaftlichen Qualität des Abschlussberichts.

Weiterführende Angaben zum Stellungnahmeverfahren finden sich auf den Internetseiten des Instituts in einem entsprechenden Leitfaden. Es gelten die in der aktuellen Version dieses Leitfadens genannten Bedingungen.

Der **Abschlussbericht**, der auf dem Vorbericht aufbauend die Bewertung der wissenschaftlichen Erkenntnisse unter Berücksichtigung der Ergebnisse der schriftlichen Anhörung enthält, stellt das abschließende Produkt der Auftragsbearbeitung dar. Er wird in Verantwortung der Projektgruppe des Instituts erstellt und dem Steuergremium zum internen Review vorgelegt. Der Abschlussbericht wird vor der Publikation dem Auftraggeber sowie dem Vorstand und dem Kuratorium der Stiftung zugeleitet.

Stellungnahmen zu Abschlussberichten können bezüglich des Vorhandenseins substanzieller nicht berücksichtigter Evidenz und wesentlicher Interpretation der berücksichtigten Evidenz vom Institut beurteilt werden. Dem Auftraggeber wird dann ggf. begründet mitgeteilt, ob aufgrund solcher Stellungnahmen eine Neubeauftragung zu dem Thema erforderlich ist oder nicht. Erscheint aufgrund

neuer Erkenntnisse eine **Aktualisierung** des Abschlussberichts erforderlich, so wird dies dem Auftraggeber mitgeteilt. Der Auftraggeber entscheidet über die Beauftragung des Instituts bezüglich der Aktualisierung des Abschlussberichts. Der Aktualisierungsprozess unterliegt den gleichen methodischen und verfahrenstechnischen Anforderungen wie die übliche Berichterstattung.

4.5 Erstellung von „Rapid Reports“ und Arbeitspapieren

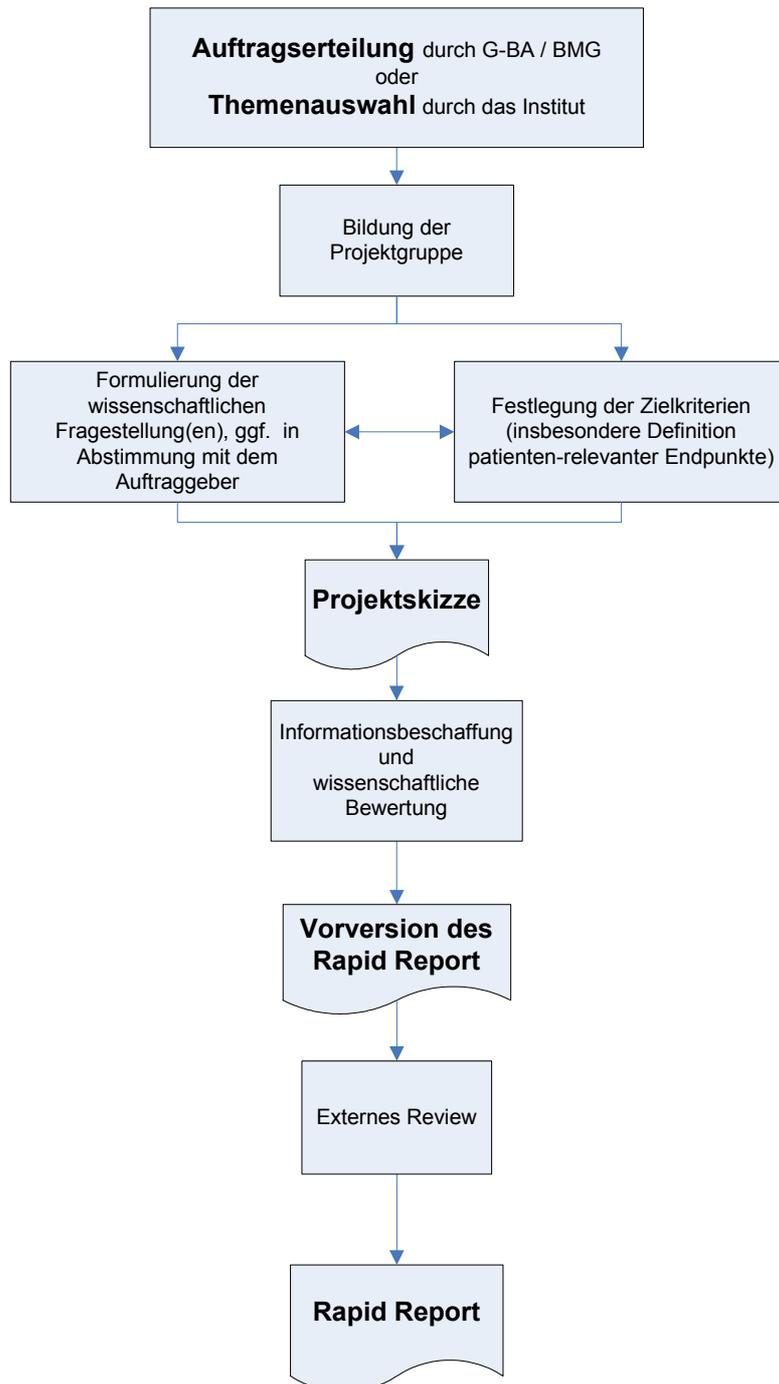
Der Ablauf der Erstellung eines Rapid Reports ist in Abbildung 3 schematisch dargestellt. Rapid Reports werden vorrangig mit dem Ziel einer kurzfristigen Information über relevante Entwicklungen im Gesundheitswesen einschließlich neuer Technologien erstellt und sind nicht ohne weiteres als Grundlage für Richtlinienentscheidungen geeignet. Dabei ist zumeist eine kürzere Erarbeitungszeit notwendig. Der Ablauf der Erstellung von Rapid Reports unterscheidet sich von dem der Berichtserstellung in zwei wesentlichen Punkten:

1. Zwischenprodukte werden nicht veröffentlicht,
2. eine schriftliche Anhörung für alle Interessierten erfolgt nicht.

Die Erstellung eines Rapid Reports kann entweder im Auftrag des G-BA oder BMG erfolgen oder vom Institut selbst initiiert werden. Analog zum „Berichtsplan“ bei der Berichtserstellung wird für einen Rapid Report eine Projektskizze erstellt, in der die wesentlichen Schritte der Informationsbeschaffung und der wissenschaftlichen Bewertung zusammengefasst werden. Abhängig vom Projekt und vom notwendigen Zeitrahmen kann die Informationsbeschaffung für einen Rapid Report begrenzt werden, beispielsweise durch den Verzicht auf die Einbeziehung bisher unveröffentlichter Daten. Der Umfang der Informationsbeschaffung wird im Rapid Report beschrieben. Eine Vorversion des Rapid Reports wird einer externen Begutachtung unterzogen und dem Steuergremium zum internen Review vorgelegt. Im Anschluss wird unter Berücksichtigung der Ergebnisse des internen und externen Reviews der Rapid Report erstellt.

An der Erstellung der Projektskizze und/oder des Rapid Reports können externe Sachverständige beteiligt werden. Arbeitspapiere werden analog zu dem Vorgehen bei Rapid Reports erstellt, wobei ein externes Review optional erfolgen kann. Über den Zeitablauf bei der Weiterleitung und Publikation von Rapid Reports und Arbeitspapieren wird in Abhängigkeit von der Dringlichkeit individuell entschieden.

Abbildung 3: Ablauf der Erstellung eines Rapid Reports als Flussdiagramm



Alle Arbeitsschritte werden in Verantwortung des Instituts getätigt. Dies kann, je nach Notwendigkeit, unter Einbeziehung externer Expertise einschließlich des wissenschaftlichen Beirats des Instituts erfolgen. Der interne Reviewprozess ist in diesem Flussdiagramm nicht dargestellt. Weitere Erläuterungen siehe vorangehenden Text.

4.6 *Veröffentlichung von Berichten*

Wesentliche Aufgabe des Institutes ist es, die vorhandene Evidenz durch eine sorgfältige Prüfung von verfügbaren Informationen zu ergründen und die Ergebnisse dieser Prüfung so nutzbar zu machen, dass sie zur Verbesserung der individuellen und allgemeinen Gesundheitsversorgung beitragen. Das Institut hat das Selbstverständnis, alle Ergebnisse, die im Institut oder im Namen des Institutes erarbeitet wurden, transparent zu machen und zu veröffentlichen.

Zur Wahrung der Unabhängigkeit des Institutes muss ausgeschlossen werden, dass die Auftraggeber oder andere interessierte Dritte Einfluss auf die Inhalte der Berichte nehmen können. Dies könnte zu einer Vermengung der wissenschaftlichen Ergebnisse mit politischen und/oder wirtschaftlichen Aspekten und/oder Interessen führen. Gleichzeitig muss vermieden werden, dass das Institut seinerseits bestimmte Ergebnisse zurückhält. Deshalb werden alle Ergebnisse, die das Institut im Rahmen seiner gesetzlichen Aufgaben erarbeitet, möglichst zeitnah veröffentlicht. Dazu gehören auch der Berichtsplan und eine Zusammenfassung der Materialien, die den Ergebnissen zu Grunde liegen.

Die Rechte an der Publikation liegen, falls nicht anders vereinbart, beim Institut.

Der Zeitrahmen für die Veröffentlichung von Vorbericht und Abschlussbericht ist Bestandteil des Berichtsplans.

4.7 *Informationsbeschaffung*

Die Informationen, die einschließlich ihrer wissenschaftlichen Bewertung die Grundlage der Institutsberichte darstellen, können vielfältiger Art sein, zum Beispiel wissenschaftliche Studien, Datensammlungen etc. Der Umgang mit Rohdaten wird im entsprechenden Abschnitt (1.13) beschrieben. Im vorliegenden Abschnitt wird der Ablauf einer themenbezogenen Recherche nach wissenschaftlicher Literatur beschrieben. Die Ausführungen beziehen sich sowohl auf von Mitarbeitern des Institutes selbst durchgeführte Recherchen als auch auf Recherchen, die von externen Sachverständigen für das Institut durchgeführt werden.

4.7.1 **Allgemeine Grundsätze einer themenbezogenen Literaturrecherche**

Eine themenbezogene Literaturrecherche hat zum Ziel, alle für die jeweilige Fragestellung **relevanten**, das heißt zum Erkenntnisgewinn beitragenden, Publikationen zu identifizieren. Die Methodik der systematischen Literaturrecherche folgt daher dem generellen Grundsatz, dass die jeweilige themenbezogene Recherche alle diejenigen Kriterien erfüllen soll, die nach den Ergebnissen diesbezüglicher Forschung einen maßgeblichen Einfluss auf das Ergebnis, also auf die Beantwortung

der Fragestellung, haben oder haben könnten. Unter Kriterien sind dabei insbesondere zu verstehen:

- die Auswahl der Datenquellen (zum Beispiel öffentliche Datenbanken, nicht öffentliche Datenbanken, Handsuche in ausgewählten Fachzeitschriften, Kontaktaufnahme mit Fachexperten/Industrie/Patientenorganisationen etc.)
- die Suchtechnik in Bezug auf die Auswahl des Studientyps (zum Beispiel RCTs, Fallbeobachtungen etc.)
- die Suchtechnik in Bezug auf die durch die Fragestellung vorgegebenen medizinischen Kriterien (zum Beispiel Zielpopulation, Art der Intervention, Endpunkte etc.)
- die Suchtechnik in Bezug auf formale Merkmale der Publikation (zum Beispiel Abstract-Publikationen, Sprache etc.).

Beispiele hierzu liefern die Publikationen [1-8].

Die Relevanz dieser Kriterien kann für verschiedene Fragestellungen unterschiedlich sein. Auch die Art des zu erarbeitenden Produkts (zum Beispiel Bericht, Rapid Report, Arbeitspapier) und der sich daraus ergebende Zeitrahmen haben Einfluss auf das Vorgehen bei der Informationsbeschaffung. Es wird im Einzelfall geprüft, welche Kriterien zur Anwendung kommen. Dabei wird auch geprüft, ob etablierte Strategien internationaler oder nationaler Arbeitsgruppen (zum Beispiel der Cochrane Collaboration) zur Anwendung kommen können.

4.7.2 Ablauf einer Recherche nach Primärpublikationen

Die Recherche nach Primärpublikationen in bibliographischen Datenbanken gliedert sich in die folgenden neun Aspekte:

1. ggf. Konkretisierung der Fragestellung
2. Umwandlung der Fragestellung in eine recherchierbare Fragestellung
3. Formulierung eines Recherchekonzepts (z.B. Sprache, Zeitraum)
4. Auswahl von Datenquellen
5. Identifizierung von Suchwörtern je Konzeptkomponente
6. Formulierung der Suchstrategien
7. Durchführung der Suche
8. Speicherung der Ergebnisse in einem Literaturverwaltungsprogramm
9. Dokumentation der Recherchen.

Dabei wird im Einzelfall geprüft, ob die Hinzuziehung von Sachverständigen sinnvoll ist. Dies kann insbesondere dann der Fall sein, wenn institutsintern keine fachliche Expertise in Bezug auf die genannte Fragestellung gegeben ist.

Die *Durchführung der Suche* (Datenbankabfrage) kann durch eine einzelne Person erfolgen.

Dabei ist es zumeist sinnvoll, zunächst nach Vorarbeiten anderer Arbeitsgruppen zu suchen, zum Beispiel mittels einer Suche in spezialisierten Datenbanken (Cochrane Database of Systematic Reviews, HTA-Datenbanken, DARE etc.). Sofern sich hierdurch relevante Übersichten identifizieren lassen, kann eventuell auf die für diese Übersichten durchgeführte Suche nach relevanten Publikationen zurückgegriffen werden. Notwendige Voraussetzung hierfür ist, dass die jeweilige Suche der Methodik des Instituts nicht widerspricht und die Übertragbarkeit der Ergebnisse auf die jeweilige Fragestellung, insbesondere unter Berücksichtigung der im Berichtsplan genannten Ein- und Ausschlusskriterien, möglich ist.

Zumeist wird aber auch bei Rückgriff auf Vorarbeiten eine ergänzende Suche nach Primärpublikationen notwendig sein, die sich dann auf den nicht von den Übersichten berücksichtigten Publikationszeitraum, nicht berücksichtigte Sprachen oder nicht berücksichtigte inhaltliche Aspekte beziehen wird. Wurden keine relevanten Übersichten identifiziert, so wird die Recherche für den gesamten für die Fragestellung relevanten Publikationszeitraum durchgeführt.

4.7.3 Sonstige Datenquellen für die Recherche

Neben der Suche in bibliographischen Datenbanken kann je nach Fragestellung eine Handsuche in ausgewählten Fachzeitschriften sinnvoll sein. Dies wird im Einzelfall entschieden.

Zusätzlich sind je nach Fragestellung weitere Datenquellen von erheblicher Bedeutung, wie zum Beispiel Studienregister oder Abstractbände wissenschaftlicher Kongresse. Im Falle einer Arzneimittelbewertung, aber auch für die Bewertung bestimmter (nichtmedikamentöser) Medizinprodukte sind öffentlich zugängliche Zulassungsdatenbanken bzw. -korrespondenzen weitere mögliche Quellen für die Informationsgewinnung. Darüber hinaus werden in der Regel die Hersteller der zu bewertenden Technologie(n) bezüglich bislang nicht veröffentlichter Informationen angefragt. Ziel dieser Anfrage ist die Identifizierung aller für die Nutzenbewertung relevanten Studien bzw. Informationen unabhängig vom Publikationsstatus.

Im Rahmen der Arzneimittelbewertung wird diese Anfrage in der Regel in einem zweistufigen Verfahren durchgeführt. In der ersten Stufe fragt das Institut beim Hersteller eine vollständige Übersicht aller Studien, die mit dem zu bewertenden Arzneimittel durchgeführt wurden, an. Ggf. definiert das Institut projektspezifische Einschlusskriterien für diese Übersicht. Aus der Übersicht der Studien identifiziert das Institut die Studien, die für die Nutzenbewertung relevant sind, und

fragt zu diesen Studien Detailinformationen an. Dabei kann es sich um eine Anfrage nach unveröffentlichten Studien oder nach ergänzenden, bislang unveröffentlichten Informationen zu veröffentlichten Studien handeln (siehe auch Abschnitt 1.11). Bisher unveröffentlichte Informationen, die in die Nutzenbewertung einfließen, werden in den Berichten des Institutes publiziert, um die Transparenz zu gewährleisten. Basis für die Einbeziehung bislang unveröffentlichter Informationen in die Nutzenbewertung ist der Abschluss einer Vereinbarung zur Übermittlung und Veröffentlichung von Studieninformationen, die vor der Einreichung von Daten zwischen dem Institut und dem beteiligten Hersteller abgeschlossen wird (siehe Mustervertrag [9]). Diese Vereinbarung spezifiziert den Verfahrensablauf, die Anforderungen an die zu übermittelnden Dokumente sowie die vertraulichen bzw. nicht vertraulichen Anteile der übermittelten Dokumente.

Die vom Auftraggeber zur Berücksichtigung vorgelegten Unterlagen werden als ein Bestandteil der Informationsgewinnung angesehen. Mit diesen Unterlagen wird dann im weiteren Prozess den sonstigen Grundsätzen der Informationsrecherche und -bewertung folgend verfahren.

4.7.4 Auswahl relevanter Publikationen

Die Auswahl der relevanten Publikationen aus den Ergebnissen der Recherche erfolgt zumeist über die zwei Arbeitsschritte

- Sichtung der Titel und Abstracts mit dem Ziel des Ausschlusses *sicher* irrelevanter Publikationen und
- Sichtung der Volltexte der daraus verbleibenden potenziell relevanten Publikationen.

Beide Arbeitsschritte sind grundsätzlich von zwei Personen unabhängig voneinander durchzuführen.

In der Regel erfolgt eine Beschränkung auf Publikationen in einer westeuropäischen Sprache. Es können aber auch Publikationen in einer anderen Sprache eingeschlossen werden, wenn aus vorliegenden Informationen zu diesen Publikationen ein für die Beantwortung der Fragestellung zusätzlicher und relevanter Erkenntnisgewinn zu erwarten ist.

4.7.5 Dokumentation

Alle Arbeitsschritte der Literaturrecherche sind in ihrer Gesamtheit zu dokumentieren. Hierzu gehören insbesondere

- die Suchstrategien für die ausgewählten Datenbanken,
- bei Abfrage bibliographischer Datenbanken zusätzlich
 - ein Recherchenachweis

- eine Auflistung der Anzahl der durch Anwendung der Suchstrategie erlangten Ergebnisse
- eine Auflistung der Ergebnisse der Recherchen in einer Literatur-Verwaltungsdatenbank (soweit möglich vollständige Datensätze)
- eine Auflistung der nach Sichtung der Ergebnisse als für die Fragestellung relevant erachteten Publikationen (Zitate)
- eine Auflistung der in der Vollsichtung als nicht relevant erachteten Literaturstellen einschließlich einer Begründung für die Nichtberücksichtigung
- bei persönlicher Kontaktaufnahme mit Produktherstellern, Fachleuten, Fachgesellschaften etc.
 - Korrespondenznachweise
- das Datum und der Zeitraum der Recherche.

Literatur

- [1] Pham B, Platt R, McAuley L, Klassen TP, Moher D. Is there a “best” way to detect and minimize publication bias? An empirical evaluation. *Eval Health Prof* 2001; 24: 109-125.
- [2] McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *Lancet* 2000; 356: 1228-1231.
- [3] Moher D, Pham B, Klassen TP et al. What contributions do languages other than English make on the results of meta-analyses? *J Clin Epidemiol* 2000; 53: 964-972
- [4] Gregoire G, Derderian F, Le Lorier J. Selecting the language of the publications included in a meta-analysis: Is there a Tower of Babel bias? *J Clin Epidemiol* 1995; 48: 159-163
- [5] Jüni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: Empirical study. *Int J Epidemiol* 2002; 31: 115-123.
- [6] Fergusson D, Laupacis A, Salmi LR, McAlister FA, Huet C. What should be included in meta-analyses? An exploration of methodological issues using the ISPO meta-analyses. *Int J Technol Assess Health Care* 2000; 16: 1109-1119.
- [7] Sampson M, Barrowman NJ, Moher D et al. Should meta-analysts search Embase in addition to Medline? *J Clin Epidemiol* 2003; 56: 943-955.
- [8] MacLean CH, Morton SC, Ofman JJ, Roth EA, Shekelle PG, Southern California Evidence-Based Practice Center. How useful are unpublished data from the Food and Drug Administration in meta-analysis? *J Clin Epidemiol* 2003; 56: 44-51.
- [9] IQWiG. Vereinbarung über die vertrauliche Behandlung von Unterlagen. Mustervertrag, Entwurf IQWiG/VFA vom 19.08.2005.
http://www.iqwig.de/download/VertraulichkeitIQWiG-VFA_19-8-05_final_2.pdf [Zugriff am 13.12.2006].

4.7.6 Literaturscreening

Neben der themenbezogenen, rückschauenden Recherche ist das frühzeitige Erkennen und Bewerten aktueller und relevanter Publikationen notwendig; Grundlage hierfür bildet die systematische diesbezügliche Beobachtung der wesentlichen wissenschaftlichen Datenquellen. Der Begriff „Datenquellen“ umfasst dabei nicht nur Fachzeitschriften, sondern auch die Laienpresse, Tages-, Wochen- und Monatspresse, elektronische Medien etc.

Als „relevant“ ist in diesem Zusammenhang eine Publikation insbesondere dann einzustufen, wenn

- die Publikation die aktuelle Versorgungslage mutmaßlich erheblich beeinflussen wird,
- die Publikation als „Meilensteinstudie“ zu werten ist,
- in der Publikation ein in der Öffentlichkeit vorrangig präsent Thema behandelt wird oder
- die Publikation ein Thema der internen Prioritätenliste berührt (siehe auch Abschnitt 4.9).

Das Steuergremium legt fest, welche Datenquellen (Fachzeitschriften, Internetseiten etc.) wesentlich sind und daher regelmäßig beobachtet werden. Die Ressorts werten die ihnen zugeteilten Datenquellen aus und leiten den jeweiligen anderen Ressorts für sie relevante Publikationen zu. Eine Publikation kann eine Ad-hoc-Bewertung einschließlich einer offiziellen Stellungnahme des Instituts nach sich ziehen. Das diesbezügliche Vorgehen folgt festgelegten Standards (siehe auch Abschnitt 4.5).

4.8 Fragestellungsbezogene Beweisführung

Es ist nicht Aufgabe des Instituts, den möglichen Nutzen oder Schaden beziehungsweise das Fehlen eines Nutzens oder Schadens einer medizinischen Intervention *nachzuweisen*. Es ist vielmehr Aufgabe des Instituts, systematisch Informationen zusammenzutragen und zu bewerten, die geeignet sind, interpretierbare Aussagen zum Nutzen oder Schaden einer medizinischen Intervention zu treffen, und allgemein verfügbar sind und/oder – falls noch nicht allgemein verfügbar – fragestellungsbezogen dem Institut zur Verfügung gestellt und damit allgemein verfügbar gemacht werden. Letzteres impliziert, dass es im Interesse derjenigen Institutionen oder Personen ist, die die Anwendung einer medizinischen Intervention propagieren oder deren Anwendung ablehnen, alle ihnen zur Verfügung stehenden Daten, die geeignet sind, interpretierbare Aussagen zum Nutzen oder Schaden einer medizinischen Intervention zu treffen, und noch nicht allgemein verfügbar sind, dem Institut zu überlassen.

4.9 *Prioritätensetzung*

Ein Großteil der Arbeitskapazität des Instituts wird durch die Beantwortung externer Aufträge seitens des G-BA und des BMG in Anspruch genommen werden. Die Prioritätensetzung bezüglich des Beginns der Auftragsbearbeitung obliegt in diesem Fall dem Auftraggeber in Absprache mit dem Institut. Darüber hinaus wird das Institut bei der Auftragsbearbeitung intern Schwerpunkte festlegen müssen. Zudem wird das Institut eigene Prioritäten setzen bezogen auf:

- interne Arbeitsprojekte der einzelnen Ressorts,
- die Erstellung von Gesundheitsinformationen durch das Ressort „Gesundheitsinformation“,
- die Projektvergabe an externe Sachverständige und Institutionen,
- institutseigene methodische und wissenschaftliche Arbeiten.

Bei der Festlegung der Prioritäten müssen mitunter die Besonderheiten einzelner Bereiche berücksichtigt werden. Jedes Ressort verfügt über die Möglichkeit, spezifische Verfahren für die Prioritätensetzung zu entwickeln, die auf die ressortspezifischen Aktivitäten zugeschnitten sind. Insgesamt haben jedoch die Prozesse der Prioritätensetzung durch das Steuergremium und die einzelnen Ressorts den Anspruch, geradlinig und mit den nachstehend beschriebenen allgemeinen Grundprinzipien, Werten und Methoden konsistent zu sein.

Die Prioritätensetzung erfolgt nach keinem starren Verfahren. Dennoch ist es bei der Aufteilung der geistigen und finanziellen Ressourcen einer öffentlichen Einrichtung wie des Instituts unerlässlich, Gerechtigkeit und Transparenz anzustreben.

Die Festlegung der Prioritäten für Forschungsaktivitäten sowie für die systematische Bewertung medizinischer Verfahren und Technologien erfordert transparente Mechanismen, und sie muss mit den Grundprinzipien, Werten, Methoden und Kriterien des Instituts in Einklang stehen. Auf diese Weise wird gewährleistet, dass die Aktivitäten und Schwerpunkte des Instituts von Außenstehenden nachvollzogen und beurteilt werden können.

4.9.1 **Hintergründe für die Prioritätensetzung des Instituts**

Das Institut ist darum bemüht, durch seine Aktivitäten die Gesundheit der Bevölkerung in Deutschland günstig zu beeinflussen. Des Weiteren sollen seine Aktivitäten zur Entwicklung eines öffentlichen und wissenschaftlichen Verständnisses von Gesundheit beitragen.

Das Institut wird bei den Entscheidungsprozessen über die von ihm durchzuführenden Arbeiten folgende Aspekte bedenken:

- Bevölkerungsanteil, der von der (durch das Institut gewählten) Arbeit profitieren könnte,

- Krankheitslast (einschließlich der Kosten), die gegenwärtig und zukünftig sowohl auf individueller als auch gesellschaftlicher Ebene und insbesondere für benachteiligte Gruppen anfällt,
- Gesundheitsprioritäten auf nationaler Ebene,
- Möglichkeiten des Instituts, Einfluss auf die klinische Praxis, auf klinische Entscheidungen sowie auf den Gesundheitsstatus der Bevölkerung auszuüben,
- die vom Institut benötigten Ressourcen, um die Aktivität effektiv durchführen zu können,
- die Möglichkeiten, auf die Gesellschaft beziehungsweise das Gesundheitswesen einen günstigen Einfluss auszuüben, unter Berücksichtigung des Gleichheitsprinzips,
- potenzielle Schäden,
- Erstellung von Beiträgen, die ausschließlich das Institut gewährleisten kann, einschließlich der Einschätzung, wozu andere Arbeiten bzw. wer ebenfalls in der Lage sein könnte, bestimmte Aktivitäten zu übernehmen,
- Beitrag der Aktivität zur Qualitätssicherung der Institutsarbeit,
- potenzieller Beitrag der Aktivität zur Evaluation von wissenschaftlichen Kenntnissen.

4.9.2 Prozess der Prioritätensetzung

Es gibt eine Vielfalt von Modellen und Prozeduren zur Prioritätensetzung, die bei verschiedenen, dem Institut ähnlichen Institutionen angewendet werden und üblicherweise einige Gemeinsamkeiten aufweisen. Diese bilden die Grundlage für das vom Institut verwendete Modell zur Prioritätensetzung und umfassen:

- die Sammlung von Daten und Meinungen (siehe auch Abschnitt 1.24), die der Entscheidungsfindung dienen, sowie deren fortlaufende Dokumentation;
- die Anwendung der von dem Institut oder seinen Ressorts entwickelten relevanten Kriterien;
- die Bereitstellung eines Berichts über die Datenlage und vorherrschenden Meinungen gemeinsam mit einer Empfehlung für das Steuergremium. Das Steuergremium trifft die endgültige Entscheidung und dokumentiert die Gründe der Entscheidungsfindung.

4.10 Zeitpunkte der Berichtserstellung

Die Bewertung einer Methode beziehungsweise Intervention ist theoretisch zu jeder Zeit möglich. Jedoch hat das IQWiG die Aufgabe, die Nutzenbewertung nur von verordnungsfähigen, nach AMG zugelassenen und auf dem deutschen Markt verfügbaren Medikamenten vorzunehmen. Im Weiteren wird der Zeitpunkt der Bewertung durch den Auftrag an das Institut bestimmt. Generelle Vor-

gaben dahingehend, dass die Erstellung einer systematischen Bewertung durch das Institut frühestens nach Verstreichen einer gewissen Zeit nach Zulassung oder Etablierung einer Methode oder Intervention erfolgt, werden nicht getroffen. Sofern im Falle einer frühzeitigen Bewertung aufgrund fehlender Langzeitstudien eine hohe Ergebnisunsicherheit festgestellt wird, wird diese den allgemeinen Arbeitsmethoden folgend dargestellt.