

Methoden

Version 1.0 vom 1. März 2005

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, Kaiser T, Kirchner H, Lange S, Müller de Cornejo G, Noelle G, Sawicki PT

ABGELÖST DURCH
"Methoden Version 2.0
vom 19.12.2006"

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 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 vorliegende erste Version (Version 1.0). Das Dokument wird zukünftig in jährlichem Abstand auf eine notwendige Überarbeitung hin überprüft, es sei denn Fehler im Dokument oder wesentliche Entwicklungen legten eine vorzeitige Aktualisierung nahe. Zu jedem durch das Institut erstellten Produkt wird eine für die Erstellung des Produkts gültige Version der Instituts-Methoden angegeben sowie ggf. Änderungen der Methodik, sofern sie in den Bearbeitungszeitraum des Produkts fallen.

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

ABGELÖST DURCH
"Methoden Version 2.0"
vom 19.12.2006

ABGELÖST DURCH
"Methoden Version 2.0
vom 19.12.2006"

Präambel

Der Gesetzgeber hat mit der Gesundheitsreform 2003 (Gesundheitssystem-Modernisierungsgesetz, GMG) die Etablierung eines neuen staatsunabhängigen Instituts im deutschen Gesundheitswesen vorgesehen. Der Gemeinsame Bundesausschuss (G-BA) hat diese wissenschaftliche Institution in Form einer privaten Stiftung 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 verantwortlicher 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, er besteht aus mindestens sechs und bis zu zwölf Wissenschaftlern. Der Sitz des Institutes für Qualität und Wirtschaftlichkeit im Gesundheitswesen ist bis 2008 in Köln.

Dem Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen obliegt künftig die wissenschaftliche Bewertung des Nutzens, der Qualität und der Wirtschaftlichkeit von medizinischen Leistungen. Dazu gehören unter anderem die Bewertung von Behandlungsleitlinien, die Abgabe von Empfehlungen zu strukturierten Behandlungsprogrammen (DMP), die Nutzenbewertung von Arzneimitteln sowie die Herausgabe von Patienteninformationen.

Das Institut wird zu Fragen grundsätzlicher Bedeutung für die Qualität und Wirtschaftlichkeit der im Rahmen der gesetzlichen Krankenversicherung erbrachten Leistungen als fachlich unabhängige wissenschaftliche Einrichtung der Stiftung unter Berücksichtigung alters-, geschlechts- und lebenslagenpezifischer Besonderheiten 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 (DMP),
- 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, auf diesen Gebieten durch die Abgabe von Empfehlungen den Gemeinsamen Bundesausschuss in der Wahrnehmung seiner gesetzlichen Aufgaben zu unterstützen.

Das Institut soll durch seine Arbeit zu einer fortwährenden Verbesserung der Qualität der medizinischen Versorgung der Bevölkerung beitragen. Das Institut soll eine eigenständige unabhängige wissenschaftliche Kompetenz entwickeln, in der Beantwortung von herangetragenen Fragestellungen, in der Bewertung von versorgungsrelevanten medizinischen Inhalten und Konzepten sowie für patientenrelevanten Forschungsbedarf einschließlich der Weitergabe dieser Information an BMGS, G-BA und die Bevölkerung.

Die Aufgaben des Institutes beziehen sich auf die Erstellung von themenbezogenen Berichten auf Anforderung des G-BA und des BMGS, sowie auf die Initiierung, Koordination und Publikation von wissenschaftlichen Untersuchungen in Bereichen, die einer Vervollständigung des versorgungsrelevanten medizinischen Wissens bedürfen. Das Institut führt ein regelmäßiges Screening der Literatur durch nach versorgungsrelevanten medizinischen Innovationen einschließlich ihrer Bewertung und gibt diese Informationen allgemeinverständlich weiter. Aufgrund der internationalen Literatur und eigener Recherchen kann das Institut Vorschläge für innovative versorgungsrelevante Forschung unterbreiten, diese initiieren, in der Durchführung begleiten und publizieren.

Das Steuergremium des Institutes, das sich aus der Institutsleitung und den Ressortleitern zusammensetzt, erstellt und modifiziert die Methoden und Abläufe der Institutsarbeit. Die Publikation und permanente Diskussion dieser Methoden ist ausdrücklich gewollt und soll neben der Transparenz der Institutsarbeit zu einer fortlaufenden dynamischen Verbesserung dieser Methoden beitragen.

Das Institut ist sich darüber bewusst, dass sich nicht in jedem Fall alle Schritte innerhalb eines Bewertungsprozesses vorab detailliert und prozedural darstellen lassen. Einzelne Vorgehensweisen sind u.a. von der jeweiligen Fragestellung, der vorliegenden wissenschaftlichen Evidenz, eingehenden Stellungnahmen und anderem abhängig. Demzufolge ist das vorliegende Dokument eher als Leitfaden bei der Bewertung einer medizinischen Maßnahme zu verstehen. Die Ausgestaltung und Darlegung des Bewertungsprozesses wird im jeweiligen Berichtsplan fallbezogen auf der Grundlage dieser Methoden erfolgen.

Mit dem Ziel einer sinnvollen und effizienten Ausnutzung der vorhandenen Ressourcen wird das Institut fachliche Vorarbeiten anderer nationaler und internationaler Institutionen im Gesundheitswesen berücksichtigen und ggf. nutzen.

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 Subgruppen-Analysen | 6 |
| 1.5 Ermittlung patientenrelevanter Effekte und Größen | 7 |
| 1.6 Aspekte der Beurteilung der Studienqualität | 9 |
| 1.7 Ermittlung von unerwünschten Wirkungen | 11 |
| 1.8 Beurteilung von Studien mit „überholten“ Methoden | 13 |
| 1.9 Beurteilung verschiedener Studienarten | 13 |
| 1.10 Rangordnung verschiedener Studienarten / Evidenzgrade | 15 |
| 1.11 Bezug zwischen Studientyp / -art und Fragestellung | 16 |
| 1.12 Beurteilung nicht oder nicht vollständig publizierter Daten | 16 |
| 1.13 Umgang mit Rohdaten | 16 |
| 1.14 Beurteilung von Ergebnisunsicherheiten | 17 |
| 1.15 Beurteilung nicht verblindbarer Verfahren | 18 |
| 1.16 Berücksichtigung datenschutzrechtlicher Aspekte / Vertraulichkeit | 18 |
| 1.17 Berücksichtigung ethischer Aspekte | 19 |
| 1.18 Darstellung und Beschreibung von Biasarten | 20 |
| 1.19 Nachweis der Verschiedenheit | 22 |
| 1.20 Nachweis der Gleichheit | 23 |
| 1.21 Meta-Analysen | 24 |
| 1.22 Adjustierung und multifaktorielle Verfahren | 25 |
| 1.23 Bewertung qualitativer Studien | 26 |
| 1.24 Einsatz von Konsultationstechniken | 27 |
| 1.25 Begutachtung (Peer Review) | 30 |
| 2. Spezielle Beurteilung medizinischer und versorgungsrelevanter Inhalte | 32 |
| 2.1 Nutzenbewertung in der Medizin | 32 |
| 2.2 Medikamentöse und nicht-medikamentöse Interventionen | 33 |
| 2.3 Diagnostische Verfahren | 34 |
| 2.4 Screening | 39 |

| | | |
|-----------|---|-----------|
| 2.5 | Gesundheitsökonomie | 42 |
| 2.6 | Leitlinien und Disease Management Programme | 46 |
| 2.7 | Systematische Übersichten und HTA-Berichte | 53 |
| 2.8 | Prognose | 56 |
| 2.9 | Individuelle Risikoabschätzung | 58 |
| 2.10 | Beurteilungsmethoden in Abhängigkeit von der Prävalenz und der Art der Erkrankung | 59 |
| 2.11 | Beurteilung von nicht direkt erkrankungsassoziierten Beschwerden | 60 |
| 2.12 | Beurteilung von komplementären Verfahren | 60 |
| 2.13 | Beurteilung bevölkerungsweiter Interventionen / Prävention | 60 |
| 2.14 | Qualitätsmanagement in der Versorgung | 62 |
| 2.15 | Beschreibung der Art und Größe des Placeboeffektes | 62 |
| 3. | Evidenzbasierte Gesundheitsinformationen für Bürger und Patienten | 64 |
| 3.1 | Ziel | 64 |
| 3.2 | Informationssystem | 64 |
| 3.3 | Entwicklung der Informationsprodukte | 68 |
| 3.4 | Publikation | 76 |
| 3.5 | Evaluation und Aktualisierung | 76 |
| 4. | Berichtserstellung | 80 |
| 4.1 | Produkte | 80 |
| 4.2 | Auswahl externer Sachverständiger | 80 |
| 4.3 | Gewährleistung der fachlichen Unabhängigkeit | 81 |
| 4.4 | Berichtsplan | 82 |
| 4.5 | Struktur der Berichtserstellung | 83 |
| 4.6 | Verkürzte Berichtserstellung | 85 |
| 4.7 | Veröffentlichung von Berichten | 87 |
| 4.8 | Informationsbeschaffung | 87 |
| 4.9 | Fragestellungsbezogene Beweisführung | 92 |
| 4.10 | Prioritätensetzung | 93 |
| 4.11 | Zeitpunkte der Berichtserstellung | 95 |

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.

Bertold 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 sollte zur deskriptiven Darstellung ein Effekt sowohl durch ein absolutes (z.B. absolute Risikoreduktion oder Number Needed to Treat) als auch durch ein relatives Maß (z.B. relatives Risiko oder Odds Ratio) beschrieben werden. 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 (1990,1999) [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 soll ein adäquates Konfidenzintervall mit der Information, ob es sich um ein- oder zweiseitige Konfidenzgrenzen handelt sowie des gewählten Konfidenzniveaus, angegeben werden. 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 (z.B. 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.
- [2] Deeks JJ, Higgins JPT, Altman DG. Analysing and presenting results. In: Alderson P, Green A, Higgins JPT, editors. *Cochrane Reviewers' Handbook 4.2.2* [updated March 2004]; Section 8. In: *The Cochrane Library, Issue 1, 2004*. Chichester: Wiley, 2004.
- [3] Agresti A. *Categorical Data Analysis*. New York: Wiley, 1990.
- [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, auffällige Ergebnisse aufgrund echter Effekte von reinen Zufallsfunden, die auf normale Datenvariabilität zurückführbar sein können, abzugrenzen. Die Konvention, von einem statistisch signifikanten Resultat zu sprechen, falls $p < 0,05$ gilt, macht in vielen Fällen durchaus Sinn. Dennoch sollte man nicht stur auf der Forderung $p < 0,05$ beharren. Je nach Fragestellung und Aussage kann es erforderlich sein einen möglicherweise sehr viel kleineren p -Wert zu fordern. Umgekehrt gibt es Situationen, in denen man auch mit einem höheren Signifikanzniveau arbeiten kann.

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 genau 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 z.B. Nicht-Unterlegenheitsstudien (s.a. 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. Für eine bessere Vergleichbarkeit mit zweiseitigen statistischen Verfahren wird in einigen Guidelines für klinische Studien eine Halbierung des üblichen Signifikanzniveaus von 5% auf

2,5% gefordert [3]. Ein zentral wichtiges Grundprinzip ist in jedem Fall, dass die Hypothesenformulierung (ein- oder zweiseitig) und das Signifikanzniveau a priori klar festzulegen sind.

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].

Ein nicht signifikantes Ergebnis darf nicht per se als Evidenz für das Nicht-Vorhandensein eines Effekts gewertet werden [5]. Zum Nachweis einer Gleichheit sind Methoden für Äquivalenzhypothesen erforderlich (s.a. Abschnitt 1.20).

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.

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 die Höhe des Stichprobenumfangs eingeht. Bei der Interpretation 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.

Eine formale Beurteilung der klinischen Relevanz von Studienergebnissen ist ein noch weitgehend ungelöstes methodisches Problem. Nur in wenigen Leitlinien finden sich Hinweise auf die Definition von relevanten bzw. irrelevanten Unterschieden zwischen Gruppen. Ein erster Ansatz zur Beurteilung der klinischen Relevanz von Studienergebnissen ist die Bewertung des Effektschätzers und des dazugehörigen Konfidenzintervalls (s.a. Abschnitt 1.1) mit Hilfe medizinischer Sachkenntnis. Ein formales Relevanzkriterium kann die Beurteilung der (im Falle von günstigen Effek-

ten) unteren Konfidenzgrenze für den Effektschätzer bzw. die Anwendung eines statistischen Tests mit Verschiebung der Nullhypothese zum Nachweis relevanter klinischer Effekte sein. Eine weitere Möglichkeit liegt darin, ein Relevanzkriterium individuell zu formulieren, z.B. im Sinne einer Responderdefinition. Darüber hinaus spielt die individuelle Einschätzung der Betroffenen eine wesentliche Rolle. Anhaltspunkte hierzu wird im Einzelfall die Darlegung patientenrelevanter Endpunkte liefern (s.a. Abschnitt 1.5).

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.

1.4 Subgruppen-Analysen

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

- Den post-hoc-Charakter, es sei denn Subgruppenanalysen wurden a priori geplant, sind wohl fundiert und sind Bestandteil des Studienprotokolls (ggf. auch von Amendments).
- Im Falle eines statistisch signifikanten Ergebnisses durch das Problem des multiplen Testens (s.a. Abschnitt 1.2), es sei denn, dieses wurde durch Anpassung des Signifikanzniveaus berücksichtigt.
- Im Falle eines statistisch nicht signifikanten Ergebnisses durch die in aller Regel geringe Power, die zumeist auch für das (inferenzstatistische) Aufdecken moderater Unterschiede in aller Regel zu gering ist, es sei denn, es wurde eine Fallzahlplanung mit angemessener Power und ggf. entsprechend erhöhtem Stichprobenumfang zusätzlich für eine Subgruppenanalyse vorgenommen [3].

Die Ergebnisse von Subgruppen-Analysen 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 um so mehr, falls das primäre Studienziel nicht erreicht wurde.

Darüber hinaus sind Subgruppen-Analysen nicht interpretierbar, wenn das Subgruppen-bildende Merkmal nach Therapiebeginn (nach Randomisierung) erhoben wurde (bzw. erhoben werden kann), z.B. sogenannte „Responder-Analysen“.

Die statistische Analyse von Subgruppen-Effekten sollte anhand eines 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 Subgruppen-Effekts interpretiert werden.

Ungeachtet der oben getroffenen Einschränkungen, können für manche Fragestellungen Subgruppen-Analysen 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 Reproduktion der entsprechenden Ergebnisse in einer Validierungsstudie sprechen können.

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.

1.5 Ermittlung patientenrelevanter Effekte und Größen

Klinische Studien werden in der Regel so konzipiert, dass krankheitsbezogene Effekte gemessen werden. Diese Effekte können für Patienten Bedeutung haben oder auch nicht. Darüber hinaus kann das Ausmaß des Nutzens (s.a. Abschnitt 2.1) oder des Risikos von Patienten individuell unterschiedlich eingeschätzt werden. Beispielsweise können Menschen, die eine Erkrankung oder Verletzung (erlebt) haben, und diejenigen, die dies nicht persönlich erlebt haben, Nutzen und Risiken unterschiedlich empfinden und bewerten [1].

Viele Studien messen Surrogate als vermutete Indikatoren für ein patientenrelevantes Ergebnis. Der direkte Rückschluss von einer Veränderung einer Surrogatvariablen auf die Veränderung eines patientenrelevanten Ereignisses ist jedoch nicht ohne weiteres möglich [2-4]. Aus diesem Grund wird das Institut derartige Surrogate nur unter der Voraussetzung der hinreichend sicheren Prüfung der Kausalkette mit patientenrelevanten Effekten für die Bewertung eines medizinischen Verfahrens heranziehen.

Über klinisch relevante Ereignisse wie Mortalität, Morbidität und unerwünschte Wirkungen der medizinischen Intervention hinaus können auch andere Effekte oder Ereignisse für Patienten Bedeutung haben, wie z.B.

- Krankheitsbezogene Lebensqualität, einschließlich der Auswirkung auf die Alltagsaktivitäten und das Wohlbefinden,
- Annehmlichkeiten, Umstände und Aufwand der Behandlung,
- Behandlungszufriedenheit und –präferenzen,
- Auswirkungen auf Angehörige.

Ergebnisse, die für die Patienten wichtig sind, sollten berücksichtigt werden, bevor bewertende Schlussfolgerungen über Interventionen getroffen werden. Daten solcher Ergebnisse können die Aussagen einer systematischen Übersicht wesentlich ändern [5]. Zusätzlich muss berücksichtigt werden, ob Ergebnisse auf Patienten-Untergruppen (z.B. Geschlechtsunterschiede, verschiedene Altersgruppen) übertragbar sind, und ob Ergebnisse aus langfristigen Untersuchungen vorliegen.

Sofern ableitbar, sollte dargestellt werden, ob die Größenordnung des Effekts einer Intervention für Patienten relevant ist [5,6]. Diese Daten werden wenn möglich an Hand der Literaturrecherche inkl. der Auswertung der Ergebnisse qualitativer Forschung und Patienten- und Angehörigenbefragung (z.B. mittels Fokusgruppenbefragung) erlangt.

Die externe Validität von vielen patientenrelevanten Zielgrößen kann von kulturellen und sozialen Gegebenheiten beeinflusst werden und kann sich im Verlauf der Zeit ändern. Darüber hinaus kann die Einschätzung einer Intervention schon alleine dadurch beeinflusst werden, dass es sich um eine Neuentwicklung handelt.

Literatur

- [1] Slevin ML, Stubbs L, Plant HJ et al. Attitudes to chemotherapy: comparing views of patients with cancer with those of doctors, nurses, and general public. *BMJ* 1990; 300: 1458-1460.
- [2] Gøtzsche PC et al. Beware of surrogate outcome measures. *Int J Tech Assess Health Care* 1996; 12: 238-246.
- [3] Liberati A, Sheldon TA, Banta HD et al. 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] The CAST investigators. Preliminary report: Effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. *New Engl J Med* 1989; 321: 406-412.
- [5] Dixon-Woods M et al. Integrative approaches to qualitative and quantitative evidence. London: NHS Health Development Agency; 2004.

- [6] Egger M, Davey Smith G, Phillips AN. Meta-analysis principles and procedures. *BMJ* 1997; 315: 1533-1537.

1.6 Aspekte der Beurteilung der Studienqualität

Bei der Beurteilung der allgemeinen Qualität von Studien spielt eine Reihe von Aspekten eine Rolle, die hier nicht vollständig im Detail wiedergegeben werden können. 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 aus 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 bzw. einer Publikation zum Studiendesign sinnvoll.

Wichtige Stellungnahmen, die zur Verbesserung der Qualität von Publikationen erarbeitet wurden, sind das CONSORT Statement für randomisierte klinische Versuche [4] und das dazugehörige Erläuterungsdokument [5], das CONSORT Statement für Cluster-randomisierte Studien [6], das QUOROM Statement für Meta-Analysen randomisierter Studien [7], das TREND Statement für nicht-randomisierte Interventionsstudien [8], das STARD-Statement für Diagnosestudien [9] und das dazugehörige Erläuterungsdokument [10]. Hält sich eine Publikation nicht an diese Stellungnahmen, so lassen sich hieraus u.U. Hinweise auf Mängel in der entsprechenden Studie ableiten. Weitere wichtige Quellen, die grundlegende Aspekte zur Beurteilung der Studienqualität beschreiben, sind die Standardwerke [11-13].

Zentrale Aspekte zur Beurteilung randomisierter klinischer Studien stellen ein adäquates Concealment, d.h. die Unvorhersehbarkeit und Verheimlichung der Gruppenzuteilung (z.B. durch externe Randomisierung bei nicht verblindbaren Studien), die verblindete Zielgrößenerhebung bei nicht verblindbaren Studien (s.a. Abschnitt 1.15), die adäquate Anwendung des "Intention-to-treat"-Prinzips sowie die Festlegung auf ein eindeutiges Hauptzielkriterium bzw. die angemessene Berücksichtigung eines möglichen Multiplizitätsproblems (s.a. Abschnitt 1.2) dar.

Die Beurteilung formaler Kriterien liefert wesentliche Anhaltspunkte für die Qualität von Studien. Es ist jedoch immer notwendig, eine über rein formale Aspekte hinaus gehende Beurteilung vorzu-

nehmen, um beispielsweise Fehler, Widersprüche und Ungereimtheiten in Publikationen darzustellen und deren Relevanz für die Ergebnisinterpretation zu 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. Oktober 2004.
- [3] Arbeitsgruppe Epidemiologische Methoden der Deutschen Arbeitsgemeinschaft für Epidemiologie (DAE). Leitlinien und Empfehlungen zur Sicherung von Guter Epidemiologischer Praxis (GEP). <http://medweb.uni-muenster.de/institute/epi/dae/Empfehlungen.doc>. Zugriff am 21. Oktober 2004.
- [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. *Br Med J* 2004; 328: 702-708.
- [7] Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DE et al. Improving the quality of reports of meta-analyses of randomised controlled trials: The QUOROM statement. *Lancet* 2000; 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] 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.
- [10] 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.
- [11] Egger M, Davey Smith G, Altman DG, editors. *Systematic Reviews in Health Care: Meta-Analysis in Context*. London: BMJ Books, 2001.
- [12] Alderson P, Green S, Higgins JPT, editors. *Cochrane Reviewers' Handbook 4.2.2* (updated March 2004). In: *The Cochrane Library, Issue 1, 2004*. Chichester: Wiley, 2004.
- [13] Guyatt G, Rennie D, editors. *Users' Guide to the Medical Literature*. Chicago, IL: AMA Press, 2002.

1.7 Ermittlung von unerwünschten Wirkungen

Die Anwendung einer medizinischen Intervention gleich welcher Art (medikamentös, nicht-medikamentös, 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 Beeinträchtigung der Lebensqualität und/oder Verkürzung der Lebenserwartung führt oder führen kann.

Die Darstellung potenzieller Risiken einer medizinischen Intervention ist wesentlicher und gleichberechtigter Bestandteil der Berichtserstellung zu dieser Intervention. Sie gewährleistet eine informierte populationsbezogene, aber auch individuelle Nutzen-Risiko-Abwägung [1]. Voraussetzung dafür ist, dass analog zur Angabe der Effektstärke einer medizinischen Intervention auch Wahrscheinlichkeiten für das Eintreten der potenziellen unerwünschten Wirkungen aus den vorliegenden Daten ermittelt und angegeben werden.

Die Darstellung des Risikoprofils einer medizinischen Intervention ist jedoch in den allermeisten Fällen ungleich schwerer als die des Nutzens (s.a. Abschnitt 2.1) derselbigen [2]. Im Gegensatz dazu, dass durch eine medizinische Intervention in der Regel nur wenige patientenrelevante (Wirksamkeits-)Endpunkte positiv beeinflusst werden, ist das Auftreten unerwünschter Wirkungen vielschichtig. Darüber hinaus treten in klinischen Studien schwerwiegende unerwünschte Ereignisse (z.B. Tod, Organversagen) in der Regel eher selten auf. Schließlich ist festzustellen, dass die zielgerichtete Forschung zur Erfassung insbesondere der seltenen, aber schwerwiegenden unerwünschten Wirkungen einschließlich der Darstellung eines kausalen Zusammenhangs mit der medizinischen Intervention erheblich unterrepräsentiert ist [3-5]. Gründe hierfür liegen unter anderem in einer auf diesem Gebiet im Rahmen von klinischen Studien nicht ausreichend entwickelten Methodik und in einem hohen Aufwand an Ressourcen, um eine mehr als irrelevante Erhöhung von seltenen unerwünschten Ereignissen ausschließen zu können. Darüber hinaus verhindern möglicherweise verschiedene Interessenskonflikte eine ausgewogene Darstellung von Nutzen und Risiken.

Folge der oben genannten Hindernisse ist, dass in vielen Fällen trotz enormer Anstrengungen die Aussagenunsicherheit bei der Darstellung unerwünschter Wirkungen höher sein wird als bei der Darstellung positiver Effekte [6]. 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

- dem ersten Anschein nach den Nutzen der Intervention fast oder vollständig oder mehr als aufheben könnten,
- dem ersten Anschein nach sich zwischen zwei oder mehr ansonsten gleichwertigen Therapieoptionen erheblich unterscheiden,
- dem ersten Anschein nach insbesondere bei denjenigen von mehreren Therapieoptionen auftreten, die besonders effektiv sind,
- dem ersten Anschein nach eine Dosis-Wirkungs-Beziehung haben,
- von Patienten als besonders wichtig angesehen werden (s.a. Abschnitt 1.5),
- von schwerwiegender Morbidität oder gar erhöhter Mortalität begleitet sind oder in Folge mit erheblicher Beeinträchtigung der Lebensqualität einhergehen.

Im Interesse der Patientensicherheit 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 im Zusammenhang mit der Durchführung der betreffenden medizinischen Intervention gehäuft gesehen oder berichtet wurden. Die Zusammenstellung erfolgt insbesondere auf Grundlage der Daten kontrollierter Interventionsstudien, in denen zielgerichtet der Nutzen der Intervention untersucht wurde, sowie auf Grundlage epidemiologischer (z.B. Kohorten- oder Fall-Kontroll-Studien) und unkontrollierter Untersuchungen (z.B. aus Anwendungsbeobachtungen, Material von Pharmakovigilanzzentren etc.) sowie Tierexperimenten und Experimenten zur Überprüfung eines pathophysiologischen Konstrukts.
- Liegen aus den oben genannten Datenquellen begründete Verdachtsmomente für das Vorliegen einer unerwünschten Wirkung vor, so wird diese solange als möglich angesehen, bis dies 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 (s.a. die Abschnitte 1.10 und 1.11). Unter „hinreichend“ ist zu verstehen, dass
 - die entsprechende(n) Untersuchung(en) in ihrem Design und ihrer Planung primär auf den Nachweis der Gleichwertigkeit zwischen der medizinischen Intervention und anderen Therapieoptionen oder Placebo bzw. keiner Intervention gerichtet waren und

- die (statistische) Definition der Gleichwertigkeit mit derjenigen vergleichbar ist, die vor Beginn des Bewertungsverfahrens durch das Institut z.B. durch Patientenbefragung ermittelt wurde (s.a. Abschnitt 1.5).^b

Literatur

- [1] 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.
- [2] Derry S, Loke YK, Aronson JK. Incomplete evidence: the inadequacy of databases in tracing published adverse drug reactions in clinical trials. BMC Medical Research Methodology 2001; 1: 7.
- [3] 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.
- [4] Ioannidis JPA, Lau J. Completeness of safety reporting in randomized trials – an evaluation of seven medical areas. JAMA 2001; 285: 437-443.
- [5] Bonhoeffer J, Zumbrunn B, Heininger U. Reporting of vaccine safety data in publications: systematic review. Pharmacoepidemiol Drug Saf 2004 Jun 16 [Epub ahead of print].
- [6] Loke YK, Price D, Herxheimer A, on behalf of the Cochrane Adverse Effects Subgroup. Including adverse effects in your review. <http://www.dsru.org/wwwboard/latestdraft.pdf>. Zugriff am 6. Oktober 2004.

1.8 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 (s.a. Abschnitt 1.9) nicht genügen, weil sie in einer Zeit geplant und durchgeführt wurden, in denen diese Ansprüche noch nicht existierten, so wird dies bei der Beurteilung der Studien berücksichtigt. Es werden die Nachteile und Mängel der Studien aufgezeigt und mögliche Konsequenzen diskutiert, diese Mängel führen aber nicht automatisch zum Ausschluss der Studien aus der Bewertung.

1.9 Beurteilung verschiedener Studienarten

Es werden hier nur die wichtigsten Designs zusammengefasst, die bei der Beurteilung von Studien in der medizinischen Forschung eine Rolle spielen. Zunächst kann man unterscheiden zwischen

^b Bei der (statistischen) Definition der Gleichwertigkeit ist zu beachten, dass die Weite des Bereichs, der als Gleichwertigkeit akzeptiert wird, in der Regel mit der Durchführbarkeit der entsprechenden Studien korreliert. Z.B. wäre die Forderung, dass Gleichwertigkeit nur bei exakter Übereinstimmung der Ereignisraten akzeptiert wird, nicht umsetzbar.

Beobachtungsstudien und Interventionsstudien. Innerhalb der Beobachtungsstudien gibt es häufig den ersten Informationsgewinn aus Fallberichten oder Fall-Serien. Diese sind natürlich anfällig gegen Verzerrungen aller Art, so dass hier, je nach Fragestellung, nur sehr beschränkt eine Evidenz ableitbar ist (s.a. Abschnitt 1.10). Aus bevölkerungsbezogenen Querschnittsstudien lassen sich z.B. Krankheitsprävalenzen schätzen, aber keine Zusammenhänge zwischen veränderbaren Expositionen und Krankheiten ableiten. 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 z.B. Fall-Kontroll-Studien häufig die einzige praktikable Möglichkeit, Informationen über Zusammenhänge zwischen Expositionen und seltenen Erkrankungen zu gewinnen.

Im Bereich der Interventionsstudien ist zunächst 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 (s.a. die Abschnitte 1.6 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 Screening-Studien können sehr unterschiedliche Ziele haben, so dass hiervon die Wahl eines adäquaten Designs abhängt (s.a. 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, auf die hier nicht weiter eingegangen werden kann.

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.

1.10 Rangordnung verschiedener Studienarten / Evidenzgrade

Es gibt verschiedene Ansätze, bestimmten Studientypen bestimmte Evidenzgrade zuzuordnen, um so eine Rangfolge bezüglich der Evidenz verschiedener Studien 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]. In Anlehnung an das System der Scottish Intercollegiate Guidelines Network Grading Review Group [2] lässt sich jedoch eine grobe Hierarchie von Studientypen erstellen, die weitgehend akzeptiert wird. Hiernach wird z.B. für Therapiestudien die höchste Evidenzstufe systematischen Übersichten randomisierter kontrollierter Studien zugeordnet. Auf dem nächsten Platz erscheinen randomisierte kontrollierte Studien. Diese werden in einigen Einteilungen noch in solche hoher oder weniger hoher Qualität eingestuft. Danach folgen nicht-randomisierte Interventionsstudien, prospektive Beobachtungsstudien, retrospektive Beobachtungsstudien, nicht-experimentelle Studien (Fallserien und Fallberichte) und mit niedrigster Evidenzstufe Expertenmeinungen ohne wissenschaftliche Begründung. Dieses grobe System ist der jeweiligen Situation und Fragestellung anzupassen und detaillierter darzustellen.

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 Meth 2004; 4: 22.

1.11 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 diagnostischer Methoden genügt in vielen Fällen eine Querschnittsstudie, z.B. wenn das Ziel der Studie sich auf die Analyse der Diskriminationsfähigkeit beschränkt (s.a. 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 Prognose-Faktoren zusammen mit den aus dem Informationsgewinn resultierenden Konsequenzen (z.B. Einleitung einer Therapie) als Strategie geprüft werden, sind wiederum randomisierte Studien das Design der Wahl (s.a. Abschnitte 2.3, 2.4 und 2.10).

1.12 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 sogenannte "graue" Literatur und Abstracts, aber auch vollständige Publikationen. Das erste Ziel des Instituts ist es in solchen Fällen zumindest zu versuchen, die fehlenden Informationen zu beschaffen, z.B. durch den Kontakt mit den Autoren. Ist es nicht möglich, die notwendigen Daten zu bekommen, so werden bei teilweise vorhandenen Informationen Sensitivitätsanalysen im Sinne von Best-Case- und Worst-Case-Szenarien durchgeführt, dargestellt und diskutiert. Beim vollständigen Fehlen relevanter Informationen, kann eine Publikation nicht beurteilt werden, und es wird lediglich dargestellt, dass es zu einem bestimmten Thema weitere Daten gibt, die aber zur Beurteilung nicht zur Verfügung stehen.

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 nicht ausgeschlossen, dass das Institut für bestimmte Fragestellungen auch bisher noch nicht analysierte Rohdaten auswertet, die von externen Quellen zur Verfügung gestellt werden (z.B. 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 ist. Insbesondere ist darauf zu achten, dass wesentliche Qualitätskriterien eingehalten werden. Z.B. sollten die Daten bei Therapiestudien unter Einhaltung des GCP-Standards entstanden sein (s.a. Abschnitt 1.6). Darüber hinaus wird in den meisten Fällen die Vorlage eines Studienprotokolls zur adäquaten Beurteilung notwendig sein. Im Umgang mit Rohdaten werden datenschutzrechtliche Aspekte berücksichtigt (s.a. Abschnitt 1.16).

1.14 Beurteilung von Ergebnisunsicherheiten

Grundsätzlich ist jedes Ergebnis einer empirischen Studie unsicher. Die statistische Unsicherheit einer Parameterschätzung, die bedingt ist durch die begrenzte Stichprobe, lässt sich in Form von Standardfehler und Konfidenzintervall quantifizieren und beurteilen (s.a. 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]. Es gibt formale Ansätze, auch diese allgemeinen Modellunsicherheiten zu berücksichtigen, z.B. durch Bayes'sche Methoden [2] oder mit Hilfe von Simulationstechniken [3]. Diese Verfahren sind aber noch nicht ausreichend entwickelt und untersucht, um routinemäßig in der Praxis verwendet werden zu können [4,5]. Zu fordern ist in jedem Fall eine qualitative Beurteilung der allgemeinen Ergebnisunsicherheit auf der Basis der aktuellen Literatur zum jeweiligen Thema. Nach wie vor sind die klassischen Kausalitätskriterien von Hill [6] 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] Draper D. Assessment and propagation of model uncertainty (with discussion). J R Stat Soc B 1995; 57: 45-97.
- [3] Phillips CV. Quantifying and reporting uncertainty from systematic errors. Epidemiology 2003; 14: 459-466.
- [4] Hjort NL, Claeskens G. Frequentist model average estimators. J Am Stat Assoc 2003; 98: 879-899.
- [5] 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.
- [6] 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 Verzerrungen (Bias) bei der Effektschätzung (s.a. Abschnitt 1.18) werden kontrollierte Studien wenn möglich randomisiert und doppel-blind 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 eine geeignete Wahl der Zielvariablen. In nicht verblindbaren Studien ist es zentral wichtig, dass ein adäquates Concealment gewährleistet wird (s.a. Abschnitt 1.6), und dass als Zielvariable ein „harter“ objektiver Endpunkt gewählt wird, oder zumindest dass dafür Sorge getragen wird, dass die Zielvariable unabhängig vom (unverblindeten) Behandler ist bzw. unabhängig vom Behandler verblindet erhoben wird (verblindete Zielgrößenerhebung).

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 nach außen zu geben, 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 evt. Vorbehalte zur Weitergabe von Daten an das Institut ausgeräumt werden.

Hinsichtlich der vertraulichen Behandlung von Unternehmensdaten sind ggf. 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 Maximierung des globalen, aber auch des individuellen Patientennutzens sowie die Stärkung der Patientenautonomie durch Gesundheitsbildung und Information im Vordergrund. Die Methode der evidenzbasierten Medizin wird hierbei als notwendiges und hilfreiches Werkzeug verstanden. Das Institut wird dieses Werkzeug gewissenhaft anwenden und seine 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. Das Institut wird sich mit aller Kraft Einflüssen jedweder Herkunft widersetzen, die die Arbeit des Instituts weg von Transparenz und Unabhängigkeit hin zu von Interessen gesteuerten Aussagen lenken wollen.

Das Institut wird sich Fragen der Verteilungsgerechtigkeit nicht verschließen. Begrenzte Ressourcen führen dazu, dass die Wahlfreiheit auf der einen Seite mit einer Einschränkung derselbigen auf der anderen Seite einhergeht. Das Institut wird vermitteln, dass die Entscheidung für oder gegen eine medizinische Methode einem gewissenhaften Abwägungsprozesses des Für und Wider entspringen muss. In diesem Zusammenhang sieht sich das Institut insbesondere auch für Minderheiten und benachteiligte Bevölkerungsgruppen in der Verantwortung stehend.

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

1.18 Darstellung und Beschreibung von Biasarten

Unter Bias versteht man systematische Verzerrungen bei der Schätzung von Effekten aus Studiendaten. Es gibt eine Vielzahl von möglichen 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 verzerrte Zuordnung der Probanden in zu vergleichende Gruppen, so dass systematische Unterschiede zwischen den Gruppen vorhanden sind. Sind dadurch wichtige Confounder in den Gruppen ungleich verteilt, so kommt es zum Selection Bias. Die Randomisierung ist die beste Methode zur Vermeidung eines Selection Bias, da die entstehenden Gruppen sich sowohl bezüglich bekannter als auch unbekannter Confounder nicht systematisch unterscheiden. In nicht randomisierten Studien kann der Effekt bekannter Confounder mit Hilfe multifaktorieller Methoden berücksichtigt werden (s.a. Abschnitt 1.22), es bleibt aber das Risiko eines systematischen Unterschieds zwischen den Gruppen durch nicht bekannte Confounder.

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

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

Bei der Untersuchung von Screening-Programmen 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 bei Krankheiten mit langer präklinischer Phase vorgetäuscht werden (*Length Bias*). Einen Schutz vor diesen Verzerrungsmechanismen bietet die Durchführung einer randomisierten Studie zur Beurteilung der Wirksamkeit einer Screening-Maßnahme (s.a. Abschnitt 2.4).

Ein allgemeines Problem bei der Schätzung von Effekten ist eine Verzerrung durch *Messfehler* und *Misklassifikationen* 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*, d.h. zu einer Verzerrung des Schätzers in Richtung des Nulleffekts. In anderen Modellen und komplexeren Situationen sind jedoch Verzerrungen in alle Richtungen möglich. Daher sollte die Stärke möglicher Messfehler in allen Studien diskutiert werden, und bei Bedarf sollten Methoden zur Adjustierung von Messfehlervverzerrungen 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 bzw. ausgleichen.

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 und/oder statistische Methoden wie die Meta-Regression eingesetzt werden [7].

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. Measurement Error in Nonlinear Models. London: Chapman & Hall, 1995.
- [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. New York: Wiley, 1987.
- [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] Sterne JAC, Egger M, Davey Smith G. Investigating and dealing with publication and other biases in meta-analysis. *BMJ* 2001; 323: 101-105.

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 (s.a. Abschnitt 1.2). Diese Argumentationsebene ist zu trennen von der Beurteilung der klinischen Relevanz eines Unterschieds (s.a. 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 erforderlich. Bei komplexeren Designs oder Fragestellungen sind zusätzlich weitere Angaben wie Korrelationsstruktur, Rekrutierungsschema, Drop-Out-Schema usw. erforderlich [1,2].

Zur abschließenden Darstellung der Ergebnisse gehört neben der Signifikanzangabe einer Aussage ein adäquates Konfidenzintervall für das gewählte Effektmaß (s.a. 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 (s.a. die Abschnitte 1.3 und 1.5).

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 Null-Hypothese 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 z.B. 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, d.h. es ist die Festlegung eines Äquivalenzbereichs erforderlich.

Um sinnvoll auf Äquivalenz schließen zu können, müssen – ähnlich wie beim Nachweis eines Unterschieds (s.a. 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“ bzw. „höchstens irrelevanten Unterlegenheit“ erfolgen. Im letzteren Fall spricht man von einer „Nicht-Unterlegenheits-Hypothese“ [3].

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 [4].

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 [5].

Im Vergleich zu Überlegenheitsstudien besitzen Äquivalenzstudien spezielle methodische Probleme. Zum einen ist es häufig schwierig, Äquivalenzbereiche sinnvoll zu definieren [6]. Zum anderen schützen z.B. die üblichen Designkriterien Randomisierung und Doppelblindheit nicht mehr hinreichend sicher vor Verzerrungen [7]. Auch ohne Kenntnis der Therapiegruppe ist es z.B. 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 um-

zugehen, 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] Roebuck P, Elze M, Hauschke D, Leverkus F, Kieser M. Literaturübersicht zur Fallzahlplanung für Äquivalenzprobleme. *Inform Biom Epidemiol Med Biol* 1997; 28: 51-63.
- [5] ICH E9 Expert Working Group. ICH Harmonised Tripartite Guideline: Statistical Principles for Clinical Trials. *Stat Med* 1999; 18: 1905-1942.
- [6] Lange S, Freitag G. Choice of delta: requirements and reality - results of a systematic review. *Biometrical J* 2005; 47 (in press).
- [7] Senn S. Inherent difficulties with active control equivalence studies. *Stat Med* 1993; 12: 2367-2375.

1.21 Meta-Analysen

Unter Meta-Analyse wird die statistische Zusammenfassung von Ergebnissen mehrerer Studien im Rahmen einer systematischen Übersicht verstanden (s.a. Abschnitt 2.7) [1]. 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 [2]. Für diesen Zweck gibt es spezielle neuere statistische Methoden, wie z.B. das I^2 -Maß [3,4]. Ist die Heterogenität der Studien zu groß, so ist eine statistische Zusammenfassung der Studienergebnisse unter Umständen nicht sinnvoll [5]. 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 Basis-Risiko abhängen [6]. In solchen Fällen, sollte die Datenanalyse über ein relatives Effektmaß erfolgen, zur deskriptiven Darstellung aber absolute Maße hieraus abgeleitet werden.

Literatur

- [1] Cook DJ, Sackett DL, Spitzer WO. Methodological guidelines for systematic reviews of randomized control trials in health care from the Potsdam consultation on meta-analysis. *J Clin Epidemiol* 1995; 48: 167-171.

- [2] Glasziou PP, Sanders SL. Investigating causes of heterogeneity in systematic reviews. *Stat Med* 2002; 21: 1503-1511.
- [3] Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; 21: 1539-1558.
- [4] Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327: 557-560.
- [5] Deeks JJ, Higgins JPT, Altman DG. Analysing and presenting results. In: Alderson P, Green A, Higgins JPT, editors. *Cochrane Reviewers' Handbook 4.2.2* [updated March 2004]; Section 8. In: *The Cochrane Library, Issue 1, 2004*. Chichester: Wiley, 2004.
- [6] 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.

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 (s.a. 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, ist 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 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 für alle erklärenden Variablen

Besonders kritisch ist eine unzureichende Darstellung der Ergebnisse multifaktorieller Verfahren dann, wenn es durch die (intransparent dargestellte) statistische Modellierung zu einer Verschiebung von Effekten in einen "gewünschten" Bereich kommt, der bei unifaktorieller Betrachtung so nicht erkennbar ist. Ausführliche Erläuterungen über die Anforderungen zum Umgang mit multi-

faktoriellen Verfahren sind in verschiedenen Übersichten und Guidelines in der Literatur zu finden [1,8,9]

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.

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]. Für das Institut besteht das Ziel bei der Verwendung qualitativer Studien in der Generierung von Hypothesen für die Arbeit des Instituts und der Unterstützung bei der Interpretation von Daten.

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. Das Institut wird die gleichen Standards bei der Berücksich-

tigung von quantitativen und qualitativen Studien anwenden. Die Bewertung der einbezogenen Studien erfolgt entsprechenden der jeweiligen wissenschaftlichen Kriterien.

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 für die Bewertung qualitativer Studien nutzen. Dabei wird auf evidenzbasierte Forschungskriterien für die Bewertung qualitativer Studien im Rahmen systematischer Übersichten und Gesundheitstechnologiebewertungen zurückgegriffen [1].

Literatur

- [1] Dixon-Woods M et al. Integrative approaches to qualitative and quantitative evidence. London: NHS Health Development Agency; 2004.
- [2] Murphy E et al. Qualitative research methods in health technology assessment: a review of the literature. NHS R&D HTA Programme. Health Technology Assessment, 1998; Vol 2: No 16.
- [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. Oktober 2004.
- [4] Harden A et al. 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 et al. 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 z.B. 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 Standards bzw. Qualitätskriterien für die Auswahl, Durchführung und Analyse dieser Techniken orientieren. Dabei werden ethische Aspekte insbesondere bei den Verfahren berücksichtigt, bei denen die Möglichkeit besteht, dass Teilnehmer einen Schaden erleiden könnten (z.B. bei Fokusgruppen).

Dem Einsatz von Konsultationstechniken wird eine Suche nach relevanten qualitativen Daten vorausgehen. Die erzielten 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 z.B. 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 Repräsentativität. Deshalb wird das Institut darauf achten, dass die Sichtweisen von benachteiligten Bevölkerungsgruppen angemessen in diesem Prozess berücksichtigt werden.

Das Ressort „Patienteninformation – Erstellung und Methodik“ verfolgt u.a. das Ziel, die allgemeinen Gesundheits- und Wissenschaftskennnisse der Bevölkerung (Health Literacy) zu fördern. Dabei soll einerseits das Verständnis für wissenschaftliche Begriffe 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 [11] der Konsultation, der Konsensentwicklung und der öffentlichen Entscheidungsfindung einsetzen und weiterentwickeln. Dabei werden die Institutswebseite und gegebenenfalls zusätzliche Multi-Media-Strategien einbezogen.

Methoden, welche sich auf die Bevölkerungsebene beziehen wie beispielsweise Online-Befragungen, Bürgerjurs [11] und öffentliche Diskussionen, 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. Washington: USAID Center for Development Information and Evaluation; 1996 No 2. http://www.usaid.gov/pubs/usaid_eval/pdf_docs/pnabs541.pdf. Zugriff am 22. Oktober 2004.
- [2] USAID Center for Development Information and Evaluation. Conducting focus group interviews. Washington: USAID Center for Development Information and Evaluation; 1996 No 10. http://www.usaid.gov/pubs/usaid_eval/pdf_docs/pnaby233.pdf. Zugriff am 22. Oktober 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. Oktober 2004.
- [4] Dixon-Woods M et al. 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. Oktober 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/pdf/r22.pdf>. Zugriff am 22. Oktober 2004.
- [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. Oktober 2004.
- [8] Liberati A, Sheldon TA, Banta HD et al. 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. Washington: USAID Center for Development Information and Evaluation; 1996 No 1. http://www.usaid.gov/pubs/usaid_eval/pdf_docs/pnabs539.pdf. Zugriff am 22. Oktober 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. Oktober 2004.
- [11] Ryan M et al. Eliciting public preferences for healthcare: A systematic review of techniques. *Health Technology Assessment*, 2001; Vol 5 No 5.
- [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 (Peer 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.

Es ist eine Zunahme der Evidenzbasierung von Peer Reviews zu verzeichnen. Zur Frage der Effektivität bestimmter Verfahrensweisen beim Peer Review liegen nur wenig aussagekräftige Untersuchungen vor [1,2]. Insbesondere gibt es nicht ausreichend valide Interventionsstudien. Die üblichen Verfahren in medizinischen Zeitschriften [2,3], einschließlich der Bewertung durch Konsumenten und Patienten [4], sind nach bislang vorliegenden Untersuchungen nur von geringem Nutzen [3].

Es gibt nicht genügend gesicherte wissenschaftliche Erkenntnisse über das genaue Vorgehen bei einem Peer Review. Dies beinhaltet u.a. die Frage nach der erforderlichen Anzahl von Gutachtern (Reviewer), und ob ein Peer Review offen (mit Nennung der Namen von Autoren und Gutachtern) oder geschlossen (Namen der Autoren und Gutachter werden nicht erwähnt) sein soll. Die jeweiligen Projekt- bzw. Ressortleiter werden entscheiden, ob ein oder mehrere Reviewer an einem Bericht beteiligt werden sollen. Gemäß dem Anspruch des Instituts nach Transparenz wird es mit einem offenen Peer-Review-System arbeiten.

Das Institut wird einen Fragebogen für Reviewer entwickeln. In diesem Fragebogen werden zum Beispiel Informationen über die Interessen der Reviewer und die Anzahl der Berichte, die sie bewerten können, erhoben. Die gewonnenen Daten werden in einer Datenbank erfasst [4].

Jeder medizinische Verlag, der mehrere Artikel jährlich veröffentlicht, muss ein Peer-Review-System entwickeln, anhand dessen die Qualität der Peer Reviewer beurteilt werden kann [5]. Das Institut wird dafür auf Grundlage validierter Systeme ein Punktesystem entwickeln [5].

Allen Reviewern wird eine Kopie der weiteren Reviews im Rahmen des Peer-Review-Prozesses sowie der veröffentlichte Abschlussbericht zugesandt.

Das Institut wird sein Peer-Review-System evaluieren und aktualisieren bzgl. neuer Entwicklungen auf dem Gebiet der Peer-Review-Methodik und -Evidenz. Die Ressorts können neben den traditionell in Zeitschriften genutzten Methoden (individuelle Bewertungen von Manuskripten) auch unterschiedliche Rapid-Appraisal-Methoden anwenden (s.a. Abschnitt 1.24).

Literatur

- [1] Rennie D. Editorial peer review: Its development and rationale. In Godlee F, Jefferson T (eds). Peer Review in Health Sciences (2nd ed). London: BMJ Books, 2003; 248-262.

- [2] Fletcher RH, Fletcher SW. The effectiveness of journal peer review. In Godlee F, Jefferson T (eds). Peer Review in Health Sciences (2nd ed). London: BMJ Books, 2003; 248-262.
- [3] Jefferson TO, Alderson P, Davidoff F, Wager E. Editorial peer-review for improving the quality of reports of biomedical studies. The Cochrane Database of Methodology Reviews 2001, Issue 3.
- [4] Bastian H. Non-peer review: consumer involvement in research review. In Godlee F, Jefferson T (eds). Peer Review in Health Sciences (2nd ed). London: BMJ Books, 2003; 248-262.
- [5] Smith J. How to set up a peer review system. In Godlee F, Jefferson T (eds). Peer Review in Health Sciences (2nd ed). London: BMJ Books, 2003; 151-163.

ABGELÖST DURCH
"Methoden Version 2.0
vom 19.12.2006"

2. Spezielle Beurteilung medizinischer und versorgungsrelevanter Inhalte

Die Extraktion von Daten aus den bei der Literaturrecherche als relevant angesehenen Publikationen (s.a. Abschnitt 4.8) und deren Bewertung wird strukturiert erfolgen und dokumentiert werden. Zur Dokumentation können noch zu entwickelnde 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.

Extraktionsbögen werden u.a. für

- HTA-Berichte
- Systematische Übersichten
- Leitlinien
- Interventionsstudien
- Diagnosestudien zur Evaluierung von Testgütekriterien
- Prognosestudien

zur Verfügung gestellt.

2.1 *Nutzenbewertung in der Medizin*

Für den Begriff des „Nutzens“ oder der „Nützlichkeit“ (von medizinischen Maßnahmen) existiert bislang weder national noch international eine einheitliche und allgemein akzeptierte Definition.

Die allgemeine Einschätzung welche medizinischen Maßnahmen für die Bürger notwendig und nützlich sind, hängt nicht nur von ihrer jeweiligen objektiven Wirksamkeit ab, sondern auch von den gesellschaftlich akzeptierten Werten und den wirtschaftlichen Möglichkeiten, also Ressourcen eines Staates ab. Sind die Gesundheitsressourcen nicht grundsätzlich gesetzlich limitiert und bezieht sich der Begriff des Notwendigen sowohl auf Heilung einer Krankheit als auch auf Prävention möglicher zukünftiger unerwünschter Ereignisse für Kranke und für Gesunde, dann konzentriert sich die Entscheidung auf die Abwägung zwischen den potentiell günstigen und schädigenden Ergebnissen einer medizinischen Maßnahme, sowie der Unsicherheit dieser Annahmen. Bei der Beurteilung des möglichen Nutzens und Schadens sollten primär patientenrelevante Ereignisse und nicht ihre Surrogate (krankheitsrelevante Aspekte) berücksichtigt werden. Weiterhin ist das Vorlie-

gen aussagekräftiger medizinisch-wissenschaftlicher Untersuchungen zur Abschätzung der Effektgröße dieser Ereignisse wesentlich. Die Wahrscheinlichkeit einer positiven Beeinflussung des Gesundheits- bzw. Krankheitszustandes wird dabei dem diagnostischen oder therapeutischen Aufwand und dem möglichen Schaden, die jede medizinische Maßnahme zwingend mit sich bringt, gegenüber gestellt. Die Bewertung dieser drei Aspekte und die sich daraus ergebende Entscheidung sind individuell verschieden, sie können also nur auf der individuellen Ebene subjektiv befriedigend getroffen werden.

Die Beschreibung des Begriffs „Nutzen“ wird sich bei der Arbeit des Instituts in der jeweiligen wissenschaftlichen Fragestellung und den Angaben im Berichtsplan widerspiegeln. Dabei wird, sofern auftragsbezogen möglich, die Relevanz für die Betroffenen im Vordergrund stehen

2.2 *Medikamentöse und nicht-medikamentöse Interventionen*

Die Beurteilung einer Studie zu einer medikamentösen oder nicht-medikamentösen 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 (z.B. Population, Versorgungsbereich etc.) übertragen lässt bzw. welche lokalen Besonderheiten Auswirkung auf die Ergebnisse selbst bzw. 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 Design-spezifische interne Validität der Studie,
- die Konsistenz der Ergebnisse mehrerer Studien gleicher Güte,
- die zu Grunde liegende Erkrankung, auf die die medizinische Intervention therapeutisch oder präventiv ausgerichtet ist.

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 nicht-medikamentösen Interventionen entsprechend zur Anwendung.

Das Studiendesign hat insofern erheblichen Einfluss auf die Ergebnissicherheit, als dass mit epidemiologischen Studien, prospektiv oder retrospektiv, ein kausaler Zusammenhang zwischen In-

tervention und Effekt nicht dargestellt werden kann, während die experimentelle Forschung hierzu grundsätzlich geeignet ist [1]. Dies gilt zumindest dann, wenn andere das Ergebnis beeinflussende Faktoren ganz oder weitgehend ausgeschaltet werden, weshalb die randomisierte, kontrollierte Studie (RCT) den Goldstandard bei der Beurteilung medikamentöser und nicht-medikamentöser 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 Nicht-Durchführbarkeit einer RCT liegen, andererseits darin, dass für die vorliegende Fragestellung auch andere Studienarten eine hinreichende Ergebnissicherheit liefern könnten. Diesbezüglich kommt insbesondere der dritte oben genannte Punkt zum Tragen. Bei Erkrankungen, die ohne Intervention nach kurzer Zeit sicher zum Tod führen, liefern beispielsweise mehrere, konsistente Fallbeobachtungen darüber, ob eine bestimmte Intervention diesen zwangsläufigen Verlauf verhindert, hinreichende Ergebnissicherheit [3].

Das Institut wird im Rahmen der Erstellung des Berichtsplans (s.a. Abschnitt 4.4) daher vorab festlegen, welche Studienarten auf Grund der Fragestellung als theoretisch hinreichend ergebnissicher (bei hoher interner Validität) angesehen werden. Studien, die diesem Qualitätsanspruch nicht mindestens entsprechen (s.a. die Abschnitte 1.9 und 1.10), 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 et al. 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.

2.3 Diagnostische Verfahren

Die Methodik zur Evaluierung diagnostischer Verfahren hat bis heutzutage bei weitem nicht die gleiche Durchdringung erfahren wie diejenige zur Prüfung von therapeutischen Verfahren [1]. So-

mit wird damit zu rechnen sein, dass nur in Ausnahmefällen die im Folgenden aufgelisteten Informationen (s.a. Abschnitte 2.3.1 bis 2.3.3) zur Bewertung eines diagnostischen Verfahrens komplett vorliegen. 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 relevant sind. Die Formulierung der Fragestellung bzw. der Ablauf, der zu einer endgültigen Fragestellung führt, folgt den im Abschnitt 4.5 niedergelegten Grundsätzen.

2.3.1 Allgemeine Aspekte

Für die Bewertung eines diagnostischen Verfahrens sind insbesondere die folgenden Informationen relevant:

- Klare Definition der zu diagnostizierenden / entdeckenden Erkrankung(en) bzw., allgemeiner, des zu entdeckenden Gesundheitsstatus (z.B. Fitness).
- Angaben zur Prävalenz der zu diagnostizierenden / entdeckenden Erkrankung(en) in der zu untersuchenden Population und in Sub-Populationen.
- Eindeutige Definition des „Goldstandard“, d.h. 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, die mit der Anwendung des diagnostischen Verfahrens verbunden sind, und der Akzeptanz sowie Zumutbarkeit sowohl für die Patienten als auch für das medizinische Personal als auch für die Allgemeinheit (z.B. Umweltrisiken).
- Angaben zu den weiteren Konsequenzen, die sich aus den jeweiligen Befunden ergeben (z.B. weiterführende Diagnostik, Therapie, andersartige nicht-therapeutische Intervention, Monitoring, Lifestyle, Basis für informierte Entscheidungen), mit Darstellung inwieweit sich aus diesen Konsequenzen ein Nutzen für die Patientin oder den Patienten ergibt.
- Angaben zu alternativen diagnostischen Verfahren und ggf. 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)
- Intra-Test-/Beurteiler-Variabilität
 - Inter-Test-/Beurteiler-Variabilität
 - Intra-Patient-Variabilität (Kurz- und Langzeit)
- Ggf. Angaben zur Übereinstimmung (von Messwerten) mit bereits etablierten Standardverfahren
- Angaben zu möglichen Störfaktoren, vor allem systematisch verzerrende Einflüsse

II. Diskriminationsfähigkeit

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

III. Vorhersage

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

2.3.3 Wirksamkeitsnachweis

- Nachweis, dass sich durch die Anwendung des diagnostischen Verfahrens eine Verbesserung im Outcome für die Patienten ergibt.

Bei Studien zu den technischen Voraussetzungen (2.3.2.I) 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.

Bei Studien zur Diskriminationsfähigkeit (2.3.2.II) lassen sich prinzipiell zwei Vorgehensweisen unterscheiden: Zum einen die Untersuchung bei ausgewählten Personen mit bekanntem Krankheitsstatus, zum anderen die Prüfung in der Anwendungssituation bei unselektierten Personen und unbekanntem Krankheitsstatus [2-4]. 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 [5,6], was bei einer Bewertung berücksichtigt werden sollte.

Folgende methodische Grundprinzipien sollten darüber hinaus bei der Bewertung von Studien zur Evaluierung der Diskriminationsfähigkeit (2.3.2.II) und Vorhersagefähigkeit (2.3.2.III) diagnostischer Verfahren beachtet werden:

- 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) und/oder dem Nachweis des Überschreitens einer Mindestschwelle orientieren kann.
- Durchführung in einer Population von Patienten bzw. Personen, bei denen das Verfahren auch zur Anwendung kommen soll (angemessenes Patientenspektrum bzw. Vermeidung eines Spectrum-Bias).
- Voneinander unabhängige und verblindete Beurteilung von zu evaluierendem diagnostischen Verfahren und Goldstandard (wechselseitige Blindheit).
- Die Diagnosesicherung (Goldstandard) bzw. 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^c.
- Angemessene Berücksichtigung von Patienten mit unklaren oder nicht interpretierbaren Testergebnissen (kein simpler Ausschluss!).
- 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).

^c Dies löst das Problem allerdings auch nicht vollständig, insbesondere in Situationen mit niedrigen a priori Wahrscheinlichkeiten [7].

- Falls das zu evaluierende diagnostische Verfahren Bestandteil des Goldstandard ist, treten ebenfalls besondere methodische Probleme auf, die einer genauen Erörterung und Berücksichtigung bedürfen.

Die Erfahrung zeigt, dass sich diese Prinzipien in publizierten Diagnose-Studien häufig nicht wiederfinden. 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 unlängst in Analogie zum CONSORT-Statement für Therapiestudien eine ähnliche Empfehlung für Studien zu diagnostischen Verfahren publiziert [8]. Whiting et al. haben eine Checkliste zur Qualitätsbeurteilung von Diagnosestudien im Rahmen von systematischen Übersichten zusammengestellt [9].

Studien zum Wirksamkeitsnachweis (2.3.3) von diagnostischen Tests^d 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 z.B. 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, d.h. es werden Personen miteinander verglichen, bei denen das Testergebnis bekannt ist, mit solchen, bei denen das Ergebnis verblindet bleibt [4]. 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 die Wechselwirkung zwischen diagnostischer Information und therapeutischer Wirksamkeit

^d Es sei angemerkt, dass Studien zum Wirksamkeitsnachweis von diagnostischen Verfahren wie hier beschrieben in der Literatur nur selten anzutreffen sind.

prüfen, ob also Patienten in Abhängigkeit vom Ergebnis des diagnostischen Tests einen unterschiedlichen therapeutischen Nutzen erfahren [3].

Literatur

- [1] Knottnerus JA, van Weel C, Muris JWM. Evidence base of clinical diagnosis. Evaluation of diagnostic procedures. *BMJ* 2002; 324: 477-480.
- [2] Köbberling J, Trampisch HJ, Windeler J. Memorandum for the Evaluation of Diagnostic Measures. *J Clin Chem Clin Biochem* 1990; 20: 873-879.
- [3] Richter K, Lange S. Methoden der Diagnoseevaluierung. *Internist* 1997; 38: 325-336.
- [4] Sackett DL, Haynes RB. Evidence base of clinical diagnosis. The architecture of diagnostic research. *BMJ* 2002; 324: 539-541.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.

2.4 Screening

Wesentliche Basis für die Bewertung ist eine präzise Fragestellung, da Screening-Programme mit unterschiedlichen Zielen durchgeführt werden, und sie sich aus unterschiedlichen „Bausteinen“ zusammensetzen, die entweder als Ganzes oder in Teilen untersucht werden (können). Die Formulierung der Fragestellung bzw. der Ablauf, der zu einer endgültigen Fragestellung führt, folgt den im Abschnitt 4.5 niedergelegten Grundsätzen.

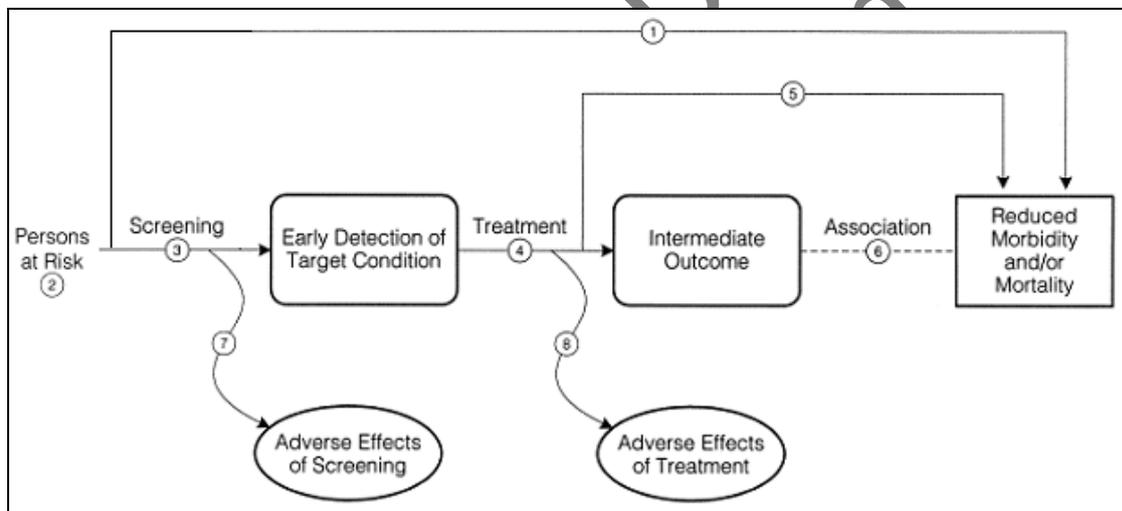
Bei der Bewertung erfolgt eine Orientierung an bereits etablierten und publizierten Kriterien, wie z.B. des UK National Screening Committee (UK NSC [1]) und der US Preventive Services Task Force (US PSTF [2]).

Diese Bewertungskriterien umfassen (1.) die Erkrankung, die es zu entdecken gilt, (2.) den einzusetzenden (diagnostischen) Screening-Test selbst, (3.) die Therapie im Falle eines positiven (patho-

logischen) Ergebnisses des Screening-Tests bzw. eine andersartige Konsequenz, die aus einem positiven Ergebnis gezogen wird, sowie (4.) das Screening-Programm als Ganzes.

Bei der Bewertung sollte differenziert werden in Programme, bei denen aus dem positiven Ergebnis des Screening-Tests eine therapeutische oder eine anderweitige, nicht-therapeutische Konsequenz folgt. Weiterhin sollte unterschieden werden in Situationen, bei denen direkte Evidenz für die Wirksamkeit des Screening-Programms vorliegt (Vergleich von Personen mit und ohne Screening mit einem patientenrelevanten Zielkriterium im Rahmen einer Studie, Pfeil 1 in Abbildung 1), und in Situationen, bei denen die Evidenz über Analogieschlüsse indirekt erhoben wird (mehrere „Screening-Bausteine“ werden in unterschiedlichen Studien untersucht, z.B. 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 Screening-Test

Es gelten die allgemeinen Anforderungen für die Beurteilung eines diagnostischen Testverfahrens, wie in den Abschnitten 2.3.1 und 2.3.2 formuliert. Es werden jedoch aufgrund der besonderen ethischen Implikationen in aller Regel höhere Anforderungen an die Testgütekriterien und die Qualität der zugrundeliegenden 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 s.a. 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 Screening-Tests 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, z.B. dergestalt, dass die Betroffenen in die Lage versetzt werden, besser informiert persönliche (Lebens-) Entscheidungen zu treffen (z.B. pränatales Screening auf Down-Syndrom, Screening auf Genträger für unheilbare / nicht beeinflussbare Erkrankungen). Für eine solche Beurteilung kann es ggf. sinnvoll sein, Methoden der Entscheidungsanalyse zu verwenden.

2.4.4 Screening-Programm

Idealerweise liegt Evidenz dafür vor, dass das Screening-Programm (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 Screening-Programmen besonders kritisch zu werten ist, da hier spezielle Verzerrungsmechanismen greifen können, wie der Lead-Time-Bias oder der Length-Bias (s.a. Abschnitt 1.18).

Liegt keine direkte Evidenz für die Wirksamkeit des Screening-Programms vor, sondern lediglich Evidenz für die einzelnen Screening-Bausteine, 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 bzw. deren Ergebnisse unter unterschiedlichen Bedingungen zur Kohärenz beitragen [2].

Das Screening-Programm muss einen Netto-Nutzen erzielen, das heißt der Nutzen, der durch ein Screening-Programm erreicht werden kann, muss den möglichen physischen oder psychischen Schaden (hervorgerufen durch den Screening-Test, nachfolgende diagnostische Maßnahmen und / oder die nachfolgende Therapie) übersteigen (zur Bewertung von „unerwünschten Wirkungen“ einer Intervention s.a. Abschnitt 1.7).

Falls das Screening-Programm bzw. 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 bzw. 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. Oktober 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(Suppl3): 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.

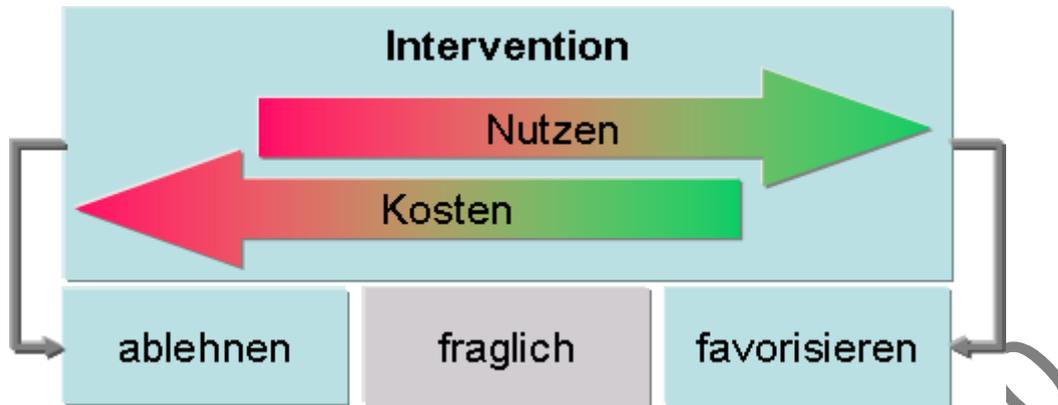
2.5 Gesundheitsökonomie

2.5.1 Inhalte

Das Ressort Gesundheitsökonomie erfüllt mittel- und unmittelbare Aufgaben innerhalb des Instituts. Mittelbar im Zusammenhang mit wissenschaftlichen Fragestellungen aus den Bereichen der übrigen Ressorts, bei denen auch ökonomische Aspekte eine Rolle spielen, unmittelbar in Bezug auf direkte Beauftragung durch den Gemeinsamen Bundesausschuss oder das BMGS. Darüber hinaus können auch eigene Fragestellungen für eine wissenschaftliche Aufarbeitung innerhalb des Institutes oder des Ressorts nach Absprache mit dem Steuergremium entwickelt werden.

Im Sinne einer bedarfsgerechten Versorgung hat die Gesundheitsökonomie als Zielsetzung die Vermeidung einer Fehl-, Unter- oder Überversorgung als wichtige Aspekte von Qualitätsdefiziten im Sinne einer unausgewogenen Kosten-Nutzen-Allokation. Kosten-Nutzen-Bewertungen können sich dabei auf verschiedene Aspekte bei der Bewertung vorhandener und zukünftiger medizinischer und gesundheitspolitischer Interventionen beziehen.

Abbildung 2: Kosten-Nutzen-Relation und Entscheidungsspektrum



Während Konstellationen mit geringem Nutzen und hohen Kosten oder solche mit großem Nutzen und geringen Kosten relativ eindeutig beurteilt werden können, ergeben sich für die sicher häufiger anzutreffenden übrigen Konstellationen komplexere Entscheidungsprozesse [1], in die insbesondere auch weitere Rahmenbedingungen mit einfließen müssen.

Im Unterschied zu anderen Wirtschaftsbereichen ergeben sich dabei insbesondere auch ethische Fragestellungen, etwa im Zusammenhang einer Güterabwägung zwischen dem Prinzip der Solidarität und des einzelnen Individuums. Dass die Ökonomie letztendlich nicht Vorrang vor der medizinischen Indikation oder der Wirksamkeit einer medizinischen Maßnahme haben darf, gilt dabei als international akzeptiert [2-6], so dass letztlich weniger die Ausgabengrößen im Einzelnen von Bedeutung sind als vielmehr die gesellschaftlichen Maßstäbe und deren ethische Grundlagen, nach denen die Mittelverteilung im Gesundheitswesen erfolgt.

2.5.2 Zielgrößen

Die Beurteilung von Kosten-Nutzen-Relationen ist auch abhängig von der verfolgten gesellschaftlichen Zielsetzung, wie etwa der Maximierung der durchschnittlichen Lebenserwartung der Gesamtbevölkerung versus eine Maximierung der durchschnittlichen Lebenserwartung von sozialen Einzelgruppen bei oder mit vermuteter geringer Lebenserwartung. Gleiches gilt etwa bei der Verfolgung von Strategien zur Steigerung der Lebensqualität oder dem Vergleich Lebenserwartung versus Lebensqualität. Hier muss zwangsläufig zunächst in Bezug auf eine einzelne Fragestellung die jeweilige Zielgröße festgelegt werden, auf die dann weitere Betrachtungen abzielen können. Bei der Definition mehrerer Zielgrößen müssen die Auswirkungen auf die jeweiligen Zielgrößen

beschrieben werden, um letztlich eine Abwägung der Zielgrößen, insbesondere auch in Hinsicht auf ethische Aspekte, ermöglichen zu können.

Dabei ist aber auch zu beachten, dass verschiedene in den neunziger Jahren veröffentlichte Vorschläge zu einer adäquaten Kosten-Nutzen-Relation gemessen an quality-adjusted-live-years (QALY) sich bislang nicht durchsetzen konnten in Hinblick auf daraus ableitbare standardisierbare Empfehlungen. Dennoch werden häufig 50.000 € pro gewonnenem QALY als noch akzeptabel bzw. kosteneffektiv bezeichnet. Die UK Health Authorities haben in Großbritannien ein Matrixmodell vorgestellt, das eine Kosten-Nutzen-Bewertung auf unterschiedlichen Evidenzklassen abbildet. Damit wird zumindest der Aspekt unterstrichen, dass eine reine Kosten-Nutzen-Betrachtung nur eine Dimension der Versorgungsqualität darstellt. Eine kombinierte Betrachtung aus EBM und KNA zur Ressourcenallokation und Priorisierung medizinischer Maßnahmen erscheint somit sinnvoll, wenn sie weitere Dimensionen wie Solidarität und Subsidiarität angemessen und ausgewogen berücksichtigt.

2.5.3 Methodik

Für die Heranziehung und valide Interpretation von Studienergebnissen und Publikationen im Bereich Gesundheitsökonomie müssen diese definierten Standards genügen, die sich allgemein an den Begriffen Transparenz, Vergleichbarkeit und Qualität orientieren [7]. Dabei haben sich in den letzten Jahren unterschiedliche Bewertungssysteme für die Planung und die Bewertung von Studien etabliert. Hierunter fallen unter anderem die des Commonwealth of Australia (1995), des Ontario Ministry of Health (1994), des Canadian Co-ordinating Office for Health Technology Assessment (1994), der Task Force on Principles of Economic Analysis of Health Care Technology (1996), des US Public Health Service Panel on Cost-Effectiveness in Health and Medicine [8] sowie die Empfehlungen zur gesundheitsökonomischen Evaluation [9], die sich nah an den Empfehlungen der BMJ-Checkliste von Drummond orientieren [10].

Es wird eine der wesentlichen ersten Aufgabe des Ressorts sein, die unterschiedlichen Instrumente ihrerseits zu evaluieren und zu bewerten, um einen institutseigenen Standard zu erarbeiten, zu publizieren und verbindlich anzuwenden.

Bei Kostenanalysen sind die unterschiedlichen Arten veranschlagter Interventionskosten transparent darzustellen. Hier sind direkte und indirekte Kosten, fixe und variable Kosten, Gemeinkosten und Grenzkosten abzugrenzen. Die Berechnung und der Vergleich von Opportunitätskosten ist dagegen in der Praxis schwer oder nur aufwendig umsetz- und darstellbar [11]. Die ebenfalls kritisch zu betrachtende Berechnung indirekter Kosten muss transparent nachvollziehbar sein und kann sich am Humankapitalansatz oder der Friktionskostenmethode orientieren. Ebenfalls den indi-

rekten Kosten zuzurechnen sind so genannte intangible Kosten, die sich nicht oder nur indirekt monetär bewerten lassen.

Auf Seiten der Nutzenanalyse wird dagegen die gesundheitsbezogene Lebensqualität zu betrachten sein. Diese zeigt sich darin, wie zufrieden Menschen mit ihrem physischen, psychischen und sozialen Gesundheitszustand sind bzw. wie sie diese drei Aspekte ihrer Gesundheit bewerten [12]. Dazu stehen so genannte nutzentheoretische Messverfahren psychometrischen Verfahren gegenüber. Während erstere eher eine allgemeine Herangehensweise darstellen, sind letztere konkreter in Bezug auf Vorstellungen, was Lebensqualität beinhaltet. Insgesamt ist die Diskussion darüber, welches Verfahren am besten geeignet ist, um Lebensqualität messbar zu machen, nicht abgeschlossen. Dies gilt auch für die Bewertung von Validität und Reliabilität aus methodologischer Sicht. Hier gilt es, die weitere empirische Forschung kritisch zu beobachten und zu begleiten.

Bei der Betrachtung von Kosten und Nutzen gelten als wichtigste Formen die Kostenminimierungs-Analyse (CMA), die Kosten-Effektivitäts- oder –Wirksamkeits-Analyse (CEA), Kosten-Nutzwert-Analyse (CUA) sowie die Kosten-Nutzen-Analyse im engeren Sinn (CBA). Die Kostenanalyse erfolgt bei allen Formen in ähnlicher Form, unterschiedlich geht dagegen die Nutzenbewertung ein. Die CMA geht von einer unterschiedlichen Kostensituation bei unterstelltem gleichem Nutzen aus und ist daher naturgemäß kritisch einzuschätzen. Bei der CEA wird der Nutzen in natürlichen Einheiten gemessen, so dass sich ein abschließender Kosten-Nutzen-Quotient ergibt, der eine Vergleichbarkeit mit anderen Interventionen nur dann ermöglicht soweit einheitliche Nutz- bzw. Effektivitätsparameter verwendet werden. Die Wahl geeigneter Effektivitätsparameter ist häufig schwierig und sollte sich an definierten patientenrelevanten Endpunkten orientieren. Hierzu sind in der Regel große randomisiert-kontrollierte epidemiologische Langzeitstudien notwendig. Da es häufig gilt, Zusatzkosten und –nutzen unterschiedlicher Interventionen zu vergleichen, bedient man sich hier der inkrementellen Analyse, die das Verhältnis von Zusatzkosten zu Zusatznutzen bestimmt. Bei der CBA wird auch der Nutzen in monetären Einheiten gemessen, z. B. als Einspareffekt. Die dabei häufig angewandte Umrechnung von Lebensqualität und –dauer in monetäre Einheiten, etwa über die Zahlungsbereitschaft („willingness to pay“ - WTP) ist dabei grundsätzlich als problematisch einzustufen. Bei der CUA [13] werden die Kosten zumeist (gewonnenen) qualitätsadjustierten Lebensjahren (QALYs) gegenübergestellt. Sie ermöglicht den Vergleich unterschiedlicher Interventionsstrategien und eignet sich insbesondere für Fragestellungen bezüglich der Lebensqualität, Mortalität und Morbidität. Nach den Empfehlungen des US Panel on Cost-Effectiveness in Health and Medicine soll eine gesundheitsökonomische Studie grundsätzlich als Kosten-Nutzwert-Analyse durchgeführt werden.

Literatur

- [1] Haycox A, Bagust A, Walley T. Clinical guidelines – the hidden costs. *BMJ* 1999; 318: 391-393.
- [2] Pellegrino ED. The commodification of medical and health care: the moral consequences of a paradigm shift from a professional to a market ethic. *J Med Phil* 1999; 3: 243-266.
- [3] Schwartz P. Medical ethics under managed care. *Int J Fert* 1996; 2: 124-128.
- [4] Ulsenheimer K. Qualitätssicherung und Risk-Management im Spannungsverhältnis zwischen Kostendruck und medizinischem Standard. *MedR* 1995; 11: 438-442.
- [5] Weltärztebund. Erklärung des Weltärztebundes zur Verantwortung des Arztes hinsichtlich der Qualität der medizinischen Versorgung. Oktober 1996.
- [6] Westhofen M. Operative Hochleistungsmedizin. Handlungszwang zwischen ärztlicher Ethik, wirtschaftlichem Erfolg und Qualitätskontrolle. In: Brudermüller G (Hrsg.) *Angewandte Ethik und Medizin*. Würzburg: Schriften des Instituts für Angewandte Ethik e.V. 1999, Bd. 1; 171-184.
- [7] Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the *BMJ*. *BMJ Economic Evaluation Working Party*. *BMJ* 1996; 313: 275-283.
- [8] Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendation of the Panel on Cost-Effectiveness in Health and Medicine. *JAMA* 1996; 276: 1253-1258.
- [9] Hannoveraner Konsensus Gruppe. Deutsche Empfehlungen zur gesundheitsökonomischen Evaluation – revidierte Fassung des Hannoveraner Konsens. *Gesundheitsökonomie u Qualitätsmanagement* 1999; 4: A 62-65.
- [10] Kurscheid T, Schrappe M, Lauterbach KW. Kritische Bewertung gesundheitsökonomischer Studien. In: Lauterbach, Schrappe (Hrsg). *Gesundheitsökonomie, Qualitätsmanagement und Evidence-based Medicine*. 2. Aufl. Stuttgart, New York: Schattauer 2004; 114-126.
- [11] Schumann J. *Mikroökonomie*. 6. Aufl. Berlin Heidelberg New York: Springer 1992.
- [12] Konecny U. Gesundheitsbezogene Lebensqualität In: Lauterbach, Schrappe (Hrsg). *Gesundheitsökonomie, Qualitätsmanagement und Evidence-based Medicine*. 2. Aufl. Stuttgart, New York: Schattauer 2004; 160-182.
- [13] Sinclair C, Frankel M. the effect of quality assurance activities on the quality of mental health services. *Qual Rev Bull* 1982; 8: 7-15.

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 von Disease-Management-Programmen (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,
- 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 über sinnvolle und adäquate Empfehlungen zu prioritären Versorgungsproblemen.

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

2.6.3 Methoden der Bewertung von Leitlinien

Wesentliche Aspekte bei der Bewertung und Prüfung von medizinischen Leitlinien sind

- Interne Validität,
- Relevanz, Angemessenheit und Praktikabilität der Empfehlungen in der Versorgung sowie
- Externe Validität.

Zentrales Qualitätskriterium einer medizinischen Leitlinie ist die Validität. Eine Leitlinie gilt dann als valide, wenn durch ihre Anwendung tatsächlich der erwartete Nutzen (Outcome) (medizinisch/ökonomisch) erzielt werden kann [1,9,10]. Streng genommen lässt sich die Validität einer Leitlinie nur durch eine rigorose Effektevaluation prüfen [2]. Aufgrund des finanziellen und personellen Aufwands ist dies jedoch nicht bei jeder bestehenden Leitlinien zu realisieren. Auch Pilotstudien wurden bislang vor der Veröffentlichung nur für wenige Leitlinien durchgeführt [11,12].

Zur abschließenden Bewertung wird deshalb ein mehrstufiger Prozess, der verschiedene Aspekte der formalen und inhaltlichen Prüfung kombiniert und in einem Gesamtbericht zusammenfasst, durchgeführt. Die dabei vom Institut verwendete Methodik wird unter Berücksichtigung aktueller wissenschaftlicher Publikationen sowie nationaler und internationaler Erfahrungen regelmäßig überprüft und ggf. aktualisiert.

I. Formale Prüfung

Eine Annäherung an die Frage der Gültigkeit von Leitlinien gelingt durch eine formale Prüfung mit methodischen Kriterien, für die ein großer Einfluss auf die Validität von Leitlinien nachgewiesen wurde [5,6,13]. Die formale Leitlinienbewertung erfolgt in Anlehnung an die Methoden des Leitlinien-Clearingverfahrens und unter Bezug auf die Beurteilungskriterien für Leitlinien von Bundesärztekammer und Kassenärztlicher Vereinigung [14,15] in strukturierter Form. Sie wird durch zwei unabhängige Bewerter durchgeführt. Bei unterschiedlichen Einschätzungen werden die Fragen erneut diskutiert und einer erneuten Bewertung unterzogen. Bleibt der Dissens bestehen, 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]. Die formale Prüfung hat eine wichtige Filterfunktion für weitere Schritte in der Bewertung einzelner Empfehlungen.

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^e, AGREE-Instrument^f) abgefragten Kriterien zur Identifizierung und Interpretation der Evidenz sowie zur Formulierung der Leitlinienempfehlungen sind im Wesentlichen Transparenzkriterien, in 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 zugrunde 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 aufwendig 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. Einzelne Fragestellungen werden bei fehlenden Angaben in Leitlinien in Abstimmung mit den übrigen Ressorts des Instituts bearbeitet.

Ein synoptischer Vergleich der Leitlinieninhalte kann zur 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 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].

Zur Identifizierung der aktuellen, existierenden Leitlinien für die jeweilige Fragestellung wird unter Berücksichtigung der im Abschnitt 4.8 (Informationsbeschaffung) genannten Vorgehensweise eine umfassende Leitlinienrecherche in den relevanten Fachdatenbanken (Leitlinien- und Literaturdatenbanken) durchgeführt. Die Recherchestrategie (Suchbegriffe, Auswahl der Datenbanken etc.) sowie Ein- und Ausschlusskriterien werden vorab unter optionaler Hinzuziehung einer oder eines Sachverständigen festgelegt und dokumentiert.

^e In AWMF 2001, Leitlinien-Manual

^f Appraisal of Guidelines Research and Evaluation (AGREE)

III. Bewertung der Relevanz, Angemessenheit und Praktikabilität von Empfehlungen

Was wissenschaftlich richtig ist, muss nicht notwendiger Weise auch sinnvoll, praktikabel und angemessen sein. Einzelempfehlungen/Kernpunkte von Leitlinien können in Hinblick auf ihre Relevanz, Angemessenheit und Praktikabilität von den Betroffenen sehr unterschiedlich bewertet werden. Dies lässt sich durch eine formale Prüfung nicht erfassen, sondern sollte mittels fachlicher Expertise in Form eines Peer Reviews und durch Patienten bewertet werden. Dazu können beispielsweise unabhängige Fokusgruppen aus Anwendern unterschiedlicher Versorgungsbereiche [11,16,17] und Patienten eingerichtet werden, die die inhaltliche Relevanz der Empfehlungen prüfen und ggf. wichtige fehlende, versorgungsrelevante Aspekte benennen.

Die Zusammenstellung der Fokusgruppen erfolgt durch das Institut.

Stellt die Fokusgruppe fest, dass wesentliche Fragestellungen für die Verbesserung der Versorgung nicht durch die vorliegenden Leitlinien abgedeckt werden oder dass aktuell vorliegende Evidenz nicht berücksichtigt wurde, so muss zu diesen Schlüsselfragen die fehlende Evidenz recherchiert und aufgearbeitet werden.

IV. 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. 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). 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 erstellt, der dem Gemeinsamen Bundesausschuss als Grundlage weiterer Beratungen zur Verfügung gestellt wird. In der Frage der DMPs werden auf der Grundlage der Bewertungen konkrete Vorschläge für die Strukturierung von DMPs erarbeitet.

Die Berichte können weiterhin auch als Basis für die Erstellung von themenbezogenen Informationen für Ärzte und Patienten dienen.

2.6.5 Abgabe von Empfehlungen zu Disease-Management-Programmen

Der Gemeinsame Bundesausschuss nach §91 SGB V gibt Empfehlungen für strukturierte Behandlungsprogramme ab. Dies betrifft die Auswahl der zu empfehlenden chronischen Erkrankungen, orientiert an:

1. der Zahl der von der Krankheit betroffenen Versicherten,
2. den Möglichkeiten zur Verbesserung der Qualität der Versorgung,
3. der Verfügbarkeit von evidenzbasierten Leitlinien,
4. dem sektorenübergreifenden Behandlungsbedarf,
5. der Beeinflussbarkeit des Krankheitsverlaufs durch Eigeninitiative des Versicherten und
6. der Höhe des finanziellen Aufwands für die Behandlung.

Für die ausgewählten chronischen Erkrankungen empfiehlt der Gemeinsame Bundesausschuss dem Bundesministerium für Gesundheit und soziale Sicherung Anforderungen an die Ausgestaltung der Programme. Dies betrifft insbesondere:

1. die Behandlung nach dem aktuellen Stande der medizinischen Wissenschaft unter Berücksichtigung von evidenzbasierten Leitlinien oder nach der jeweils besten, verfügbaren Evidenz sowie unter Berücksichtigung des jeweiligen Versorgungssektors,
2. durchzuführende Qualitätssicherungsmaßnahmen,
3. Voraussetzungen und Verfahren für die Einschreibung des Versicherten in ein Programm, einschließlich der Dauer der Teilnahme,
4. Schulungen der Leistungserbringer und der Versicherten,
5. die Dokumentation und
6. die Bewertung der Wirksamkeit und der Kosten (Evaluation) und die zeitlichen Abstände zwischen den Evaluationen eines Programms sowie der Dauer seiner Zulassung nach §137g SGB V.

Ziele und Rolle des Institutes bei der Abgabe von Empfehlungen zu Disease-Management-Programmen (DMPs)

Ziel ist es, für den Gemeinsamen Bundesausschuss wissenschaftliche Erkenntnisse zu einzelnen Aspekten der DMPs aufzubereiten und als Grundlage für seine Entscheidungen zur Verfügung zu stellen. Da es bei den Anforderungen nicht alleine um die Benennung von evidenzbasierten Leitlinien geht, sondern auch um die Konkretisierung der Leistungspflicht für die teilnehmenden Leis-

tungserbringer, können die Anforderungen Bestandteil von aktuellen evidenzbasierten Leitlinien sein, sie können darüber hinaus aber auch Empfehlungen enthalten, die auf der Basis von aktuellen Ergebnissen aus systematischer Literaturrecherche und Bewertung erstellt wurden. Darüber hinaus können auf dieser Grundlage auch Qualitätsindikatoren entwickelt werden, die die Verbesserung der Versorgung abbilden.

Weitere Möglichkeiten der Unterstützung des G-BA und seiner Gremien bei der Entwicklung, Prüfung oder Evaluation von DMPs werden in Abstimmung mit den zuständigen Gremien geprüft.

Literatur

- [1] Europarat. Methodology for drawing up guidelines on best medical practice – recommendation No R(01)13. Europarat, Straßburg, Eigenverlag, <http://www.coe.int>. Deutsche Version: Entwicklung einer Methodik für die Ausarbeitung von Leitlinien für optimale medizinische Praxis. Empfehlung Rec. (2001) 13 des Europarates und erläuterndes Memorandum. Z ärztl Fortb Quallsich 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 Quallsich 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] Bandolier Forum, Independent evidence-based health care, On Care Pathways, July 2003, p. 1-12. <http://www.ebandolier.com>; Zugriff am 10. Oktober 2004.
- [9] Grol R, Dalhuijsen J, Thomas S, in t'Veeld C, Rutten G, Mokkink H. Attributes of clinical guidelines in general practice: observational study. BMJ 1998; 317: 858-861.
- [10] 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.
- [11] Ärztliches Zentrum für Qualität in der Medizin. Leitlinien-Clearingbericht „COPD“. ÄZQ-Schriftenreihe Band 14; Verlag videel, Niebüll; 2003.
- [12] Ärztliches Zentrum für Qualität in der Medizin. Leitlinien-Clearingbericht „Depression“. ÄZQ-Schriftenreihe Band 12; Verlag Videel, Niebüll; 2003.

- [13] Field MJ, Lohr KN. Clinical practice guidelines. Directions from a new program. Institute of Medicine, Washington D.C. 1990.
- [14] Ärztliche Zentralstelle Qualitätssicherung. Checkliste „Methodische Qualität von Leitlinien“. Dtsch Arztebl 1998; 95: A2576-A2578, C1838-C1840.
- [15] Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften, Ärztliche Zentralstelle Qualitätssicherung. Das Leitlinien-Manual. Z ärztl Fortbild Quallsich 2001; 95 (Suppl. I): 1-84.
- [16] Kitzinger J. Introducing focus groups. BMJ 1995; 311: 299-302.
- [17] Kean S. Focus Group Interviews: Ein qualitativer Forschungsansatz in der Pflege. Pflege 2000; 13: 145-151.

2.7 Systematische Übersichten und HTA-Berichte

Systematische Übersichten (Reviews) und HTA-Berichte stellen als zusammenfassende Darstellung der aktuellen Studienlage eine wertvolle Entscheidungsgrundlage für Arzt und Patient, aber auch für Entscheidungsträger im Gesundheitswesen dar. Ebenso wird das Institut bei der Erstellung von Berichten und Patienteninformationen auf vorliegende systematische Übersichten und HTA-Berichte zurückgreifen. Notwendige Voraussetzung hierfür ist, dass die Methodik dieser Arbeiten den Anforderungen des Instituts entspricht, das heißt insbesondere eine transparente Darstellung der Literaturrecherche- und Auswahl-Strategie, und dass diese Strategie und die zur Auswahl der relevanten Arbeiten verwendeten Ein- und Ausschlusskriterien denen im jeweiligen (Instituts-) Berichtsplan festgeschriebenen Kriterien entsprechen.

Eine systematische Übersicht kann wie jede andere wissenschaftliche Publikation nicht nur zu fehlerhaften Ergebnissen führen, sondern auch durch systematische Verzerrungen in ihrer Aussage gelenkt werden [1]. Grundsätzlich ist zu berücksichtigen, dass es auch bei systematischen Übersichten erhebliche Qualitätsunterschiede gibt, die eine generelle Einstufung auf eine hohe Evidenzstufe nicht immer rechtfertigen.

In den letzten Jahren ist die Anzahl der systematischen Übersichten stark angestiegen [2]. Mit der Zunahme von publizierten systematischen Übersichten zeigen sich auch Probleme durch systematische Übersichten mit widersprüchlichen Aussagen und Unterschieden zwischen Ergebnissen von systematischen Übersichten [3] und randomisierten kontrollierten Studien zu vergleichbaren Fragestellungen [4].

Notwendige, aber nicht hinreichende Voraussetzung für die Einbeziehung von systematischen Übersichten in die Berichte des Instituts ist eine methodische Prüfung in Anlehnung an das Quorum Statement [1] und die Methoden des Centre for Reviews and Dissemination in York, nach denen systematische Übersichten geprüft werden, bevor sie in die Datenbank DARE aufgenommen wer-

den (s.a. <http://www.york.ac.uk/inst/crd/>). Darüber hinaus werden diesen formalen Prüfungen inhaltliche Prüfungen nachgestellt.

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. Die systematische Übersicht kann nicht besser sein, als die Einzelstudien, auf deren Basis sie erstellt wurde. Die Einbeziehung von Studien schlechter Qualität kann zu einer deutlichen Verzerrung des Ergebnisses [5] führen.

Kritische Bewertung von HTA-Berichten

Im Gegensatz zu systematischen Übersichten werden HTA-Berichte nicht primär als Arbeitshilfen für den klinischen Arbeitsalltag geschrieben. Ihr Ziel ist es, entscheidungsrelevante Informationen auf verschiedenen Ebenen der Steuerung des Gesundheitswesens zur Verfügung zu stellen [6]. Hierzu zählen beispielsweise Investitionsentscheidungen, Kostenübernahmeentscheidungen und die Gestaltung des Leistungskatalogs. Darüber hinaus werden HTA-Berichte oft als wichtige Grundlage zur Erstellung medizinischer Leitlinien und von Patienteninformationen heran gezogen.

Ein standardisiertes Format für HTA-Berichte, wie es beispielsweise von der EUR-ASSESS-Gruppe vorgeschlagen wurde [7], hat sich bisher nicht durchgesetzt, was deutlich macht, dass der lokale Bezug immer einen wesentlichen Einfluss auf das Berichtsformat hat. Qualitätskriterien, nach denen HTA-Berichte geprüft werden können, beziehen sich deshalb primär auf die ordnungsgemäße, transparente und reproduzierbare Durchführung und Interpretation der Einzelbestandteile (z.B. gesundheitsökonomischer Übersichten oder der gesundheitsökonomischen Analyse) [8].

Das Institut wird sich zur Prüfung von HTA-Berichten an den international akzeptierten Standards (z.B. INAHTA^g oder ECHTA^h) für die Erstellung und Bewertung [9,10] orientieren.

Insbesondere werden folgende Aspekte [6] näher betrachtet:

- die Darstellung des Hintergrunds der Technologiebewertung,
- die Formulierung der spezifischen Fragestellung,
- Angaben zum Status quo der Technologie,
- technische Charakteristika,

^g INAHTA – International Network of Agencies for health Technology Assessment

^h ECHTA – European Collaboration on Health Technology Assessment

- die systematische Evaluation der Sicherheit und klinischen Wirksamkeit,
- die gesundheitsökonomische Evaluation,
- der Zusammenhang von Organisationsstrukturen und Abläufen mit der Technologie,
- die Diskussion der Generalisierbarkeit und Übertragbarkeit, Einschätzung ethischer, sozialer und juristischer Implikationen.

Für das deutsche Gesundheitssystem hat sich in Hinblick auf die Akzeptanz der HTA-Berichte u.a. das Vorgehen des ehemaligen Bundesausschuss Ärzte und Krankenkassen bewährt, Stellungnahmen interessierter Parteien [8] einzubeziehen. Das Institut wird im Rahmen der Bewertung von HTA-Berichten auch berücksichtigen, ob solche Stellungnahmen einbezogen wurden.

Der Ein- oder Ausschluss bereits vorhandener HTA-Berichte wird nachvollziehbar dokumentiert. Entsprechende Kriterien sind abhängig vom zu behandelnden Thema. Sie sind vorab mit der Projektleitung abzustimmen und im Berichtsplan (s.a. Abschnitt 4.4) zu dokumentieren.

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 QUORUM 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] Jadad AR, Cook DJ, Browman G. A guide to interpreting discordant systematic reviews. *Can med Assoc J* 1997; 156: 1411-1416.
- [4] LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med* 1997; 337: 536-542.
- [5] Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273: 408-412.
- [6] Perleth M. Kritische Bewertung von HTA-Berichten. In: Kunz R, Ollenschläger G, Raspe H, Jonitz G, Kolkman FW eds. *Lehrbuch Evidenzbasierte Medizin in Klinik und Praxis*. Dt Ärzteverl 2000: 147-155.
- [7] Liberati A, Sheldon TA, Banta HD. EUR-Assess project Subgroup report on methodology: methodological guidance for the conduct of health technology assessment. *Int J Techn Assess Health Care* 1997; 13: 186-219.
- [8] Gibis B, Rheinberger P. Erfahrungen mit Health Technology Assessment im Bundesausschuss der Ärzte und Krankenkassen. *Z ärztl Fortbild Quallsich* 2002; 96: 82-90
- [9] Busse R, Velasco M, Perleth M, Orvain J. Best practice in undertaking and reporting HTA. ECHTA working group 4 report: 1-104.

- [10] Hailey D. Toward transparency in health technology assessment. A checklist for HTA reports. *Int J Techn Assess Health Care*. 2003; 19: 1-7.

2.8 *Prognose*

Wesentliche Basis für die Bewertung von Prognosestudien ist eine präzise Fragestellung, da Studien zur Evaluierung prognostischer Merkmale mit unterschiedlichen Zielen (Risikofaktorevaluierung, Scoreentwicklung / -validierung, sog. „Responder-Analysen“) durchgeführt werden, bzw. die Abgrenzung zu Diagnose- und/oder Screeningstudien schwierig sein kann, und je nach Zielsetzung unterschiedliche Bewertungsprinzipien zum Zuge kommen. Die Formulierung der Fragestellung bzw. der Ablauf, der zu einer endgültigen Fragestellung führt, folgt den im Abschnitt 4.5 niedergelegten Grundsätzen.

Ein prognostisches Merkmal liefert eine Information, die letztlich keinen Selbstzweck haben kann, sondern zu einer Konsequenz führen sollte, die ihrerseits einen prüfbaren Nutzen für die Patientin und den Patient bedeutet. In diesem Zusammenhang sind an ein Prognoseverfahren ähnliche (allgemeine) Anforderungen wie an einen diagnostischen Test zu stellen (s.a. die Abschnitte 2.3.1 und 2.3.2).

Soll ein prognostisches Merkmal im Sinne eines Screenings bzw. eines Präventionsprogramms eingesetzt werden, sind bei der Beurteilung die in diesen Abschnitten (2.4 bzw. 2.13) formulierten Grundsätze zu berücksichtigen.

Speziell bedeutsam bei der Bewertung von Studien zu prognostischen Merkmalen sind folgende Aspekteⁱ:

- Klare Festlegung auf *ein* zu evaluierendes Merkmal (in epidemiologischen Studien typischerweise als Exposition oder Risiko- [bzw. protektiver] Faktor bezeichnet). Sollen mehrere Merkmale gleichzeitig innerhalb einer Studie untersucht werden, ist prinzipiell das Problem des multiplen Testens zu beachten (s.a. Abschnitt 1.2).
- Klare Definition des zu evaluierenden Merkmals^j einschließlich der statistischen Handhabung (z.B. Dichotomisierung oder Betrachtung von Terzilen, Quartilen, etc. bei einem quantitativen Merkmal) und einer Begründung für das gewählte Vorgehen.

ⁱ Es sei allerdings an dieser Stelle darauf hingewiesen, dass es für die Beurteilung von Prognose-Studien keine allgemein akzeptierten Qualitätskriterien gibt [2].

^j Auch wenn, wie oben beschrieben, auch mehrere Merkmale gleichzeitig innerhalb einer Studie untersucht werden können, wird hier und im Folgenden – soweit nicht anders ausgewiesen – der besseren Lesbarkeit willen nur die Einzahl verwendet.

- Klare Festlegung und Definition von möglichen Störgrößen und Effektmodifikatoren, einschließlich deren statistischer Handhabung.
- Klare Definition des Outcome, an dem die prognostische Wertigkeit orientiert werden soll.
- 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 den interessierenden Outcome erfordert.
- Klare Beschreibung der Ziel- und Stichprobenpopulation (z.B. Bevölkerung, Register, Hausarztbezogen) und Begründung für deren Wahl. Klare Beschreibung für die Auswahl und das Vorgehen bei der Rekrutierung der Studienteilnehmer (Zufallsauswahl, Frage der Repräsentativität).
- 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 den interessierenden Outcome) getroffen werden kann.
- Bei Kohortenstudien Vollständigkeit des follow-up bzw. 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, z.B. als Entwicklung des Scores innerhalb einer sogenannten „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 umfassen Fall-Kontroll- und Kohortenstudien, in Ausnahmefällen (bei der Betrachtung unveränderlicher Merkmale) auch Querschnittstudien. Die grundlegenden Prinzipien für die Bewertung solcher Studien über die oben genannten Aspekte hinaus sind im Abschnitt 1.6 formuliert.

Die methodische Qualität von Studien (bzw. deren Publikationen) zu prognostischen Merkmalen ist häufig unzureichend [1]. Insofern sind Meta-Analysen (nicht systematische Übersichten!) von Prognose-Studien oft unangemessen und deren Ergebnisse nur unter Vorbehalt zu verwenden [2]. 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 Publikations-Bias ist [2].

Literatur

- [1] 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] Altman D. Systematic reviews of evaluations of prognostic variables. In: Egger M, Davey Smith G, Altman DG (Hrsg.). *Systematic reviews in health care: meta-analysis in context*, 2nd ed. London: BMJ Books, 2001.

2.9 Individuelle Risikoabschätzung

Neben Ergebnissen aus Studien, die einzeln oder (zumeist) mehrere prognostische Merkmale untersuchen, werden für die individuelle Risikoabschätzung von Patienten (oder allgemeiner Personen), ein unerwünschtes Ereignis zu erleiden, zunehmend sogenannte „risk charts“ (auch als „risk engines“ bezeichnet) eingesetzt, bei denen für das Zusammentreffen mehrerer Risikofaktoren multifaktorielle Schätzwerte angegeben werden (z.B. Sheffield table [1] oder „joint British chart“ [2]). Bei der Bewertung solcher Instrumente ist zu berücksichtigen,

- aus welcher Art von Studien die zugrundeliegenden Daten stammen,
- ob die für die multifaktorielle Betrachtung herangezogenen Merkmale in diesen zugrundeliegenden Studien ebenfalls gemeinsam betrachtet wurden,
- ob, und wenn ja, wie eine multifaktorielle statistische Analyse in diesen zugrundeliegenden Studien erfolgte,
- ob bei einer eventuell durchgeführten multifaktorielle Analyse in den zugrundeliegenden Studien adäquate statistische Modelle (z.B. log-lineares Modell) mit Berücksichtigung von Wechselwirkungstermen verwendet wurden, und
- 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(Suppl2): S1-S29.

2.10 Beurteilungsmethoden in Abhängigkeit von der Prävalenz und der Art der Erkrankung

Jede generalisierende Aussage einer Studie ist mit einem Fehler behaftet und kann daher zu Fehleinschätzungen führen. Die mögliche Größe des Fehlers ist unter anderem abhängig von dem Studiendesign, der Studiendurchführung, der Güte und Vollständigkeit der Evaluation der Daten und der Qualität des Ergebnisberichts. Bei der Interpretation der Studienergebnisse und der Übertragung in die Versorgungsrealität ist daher die mögliche Größe dieses Fehlers zu berücksichtigen. Die potentielle Möglichkeit der Fehleinschätzung ist dabei in Beziehung zu setzen unter anderem zur Art und Schwere der Erkrankung, der Größe und Art des diagnostischen oder des Therapieeffektes und zur Qualität der Evidenz über bereits vorhandene diagnostische und therapeutische Verfahren ähnlicher Zielsetzung.

Bei sehr seltenen Krankheiten kann es mitunter sehr schwer oder gar unmöglich sein, genügend Patienten in eine klinische Studie einzuschließen, um mit ausreichender Power selbst moderate Effekte statistisch aufdecken zu können. Dies begründet allerdings nicht den Verzicht auf die wesentlichen methodischen Instrumente für das Design klinischer Studien wie Randomisierung, Doppelblindheit, Intention-to-treat-Prinzip, etc. Um dem Powerproblem zu begegnen, kann es bei der Bewertung von Studien in solchen Indikationsgebieten sinnvoll sein, a priori – also in Unkenntnis der Datenlage – ein größeres Irrtumsniveau zu akzeptieren als das in der medizinischen Forschung üblicherweise verwendete Niveau von (zumeist zweiseitig) 5% (s.a. Abschnitt 1.2). Allerdings sollte darauf geachtet werden, dass das Kriterium „sehr seltene Erkrankung“ nicht durch eine inadäquate Subgruppenbildung bei eigentlich nicht sehr seltenen Krankheiten künstlich erzeugt wurde (sogenanntes „disease slicing“).

Auch bei besonders schwerwiegenden Erkrankungen, also z.B. solchen, die neu beschrieben sind und bei einem großen Anteil von Patienten innerhalb kurzer Frist zu irreversibler Morbidität oder gar zum Tod führen (z.B. AIDS in den 80er Jahren des vorigen Jahrhunderts oder SARS), kann eine Anpassung des (statistischen) Irrtumsniveaus (nach oben) in Erwägung gezogen werden.

Bei Krankheiten mit nahezu deterministischem Verlauf (z.B. Glioblastoma multiforme) können für eine positive Nutzenbewertung auch Studien niedriger Evidenzklasse (z.B. Fallserien) ausreichend sein, wenn solche Studien den für die jeweilige Studienart gültigen Qualitätsansprüchen genügen, und wenn die Ergebnisse solcher Studien einen „dramatischen“, dem natürlichen Verlauf quasi entgegengesetzten Effekt zeigen. Dies gilt selbstverständlich auch in umgekehrter Weise für die Beobachtung und Bewertung von schwerwiegenden (unerwarteten) unerwünschten Ereignissen bei Zustandsbildern mit bekanntem benignem Verlauf.

2.11 Beurteilung von nicht direkt erkrankungsassoziierten Beschwerden

Unter „nicht direkt erkrankungsassoziierten Beschwerden“ sollen Symptomkomplexe verstanden werden, die sich nicht eindeutig den gängigen, zumeist pathophysiologisch begründeten Krankheitsentitäten zuordnen lassen (z.B. sogenanntes „chronisches Müdigkeitssyndrom“). Für die Bewertung von therapeutischen Interventionen solcher Symptomkomplexe lassen sich prinzipiell die gleichen Methoden anwenden wie bei anderen Erkrankungen auch. Bei der Bewertung von diagnostischen und prognostischen Verfahren zur Erkennung solcher Symptomkomplexe ist speziell darauf zu achten, dass diese Verfahren nicht Bestandteil des Symptomkomplex sind, da dies eine Evaluierung im Hinblick auf Diskriminationsfähigkeit und Vorhersagegenauigkeit in aller Regel ausschließt.

2.12 Beurteilung von komplementären Verfahren

Komplementäre Verfahren gleich welcher Art (im Wesentlichen therapeutisch oder diagnostisch) lassen sich prinzipiell anhand der gleichen Methoden evaluieren und bewerten wie (natur-) wissenschaftlich begründete Verfahren. Dies ist durch zahlreiche Beispiele belegt (z.B. Homöopathie [1], Akupunktur [2]). Ein Verweis auf die Nicht-Prüfbarkeit solcher Verfahren durch wissenschaftliche Methoden aufgrund von Spezifika wie besonderes Arzt- (bzw. Therapeuten-) Patienten-Verhältnis, besondere Krankheits- (bzw. Gesundheits-) Konstrukte, etc. ist nicht akzeptabel. In aller Regel wird es sinnvoll sein, Vertreter solcher Verfahren auf der Grundlage der Methodik des Instituts in die Protokollplanung für die Erstellung von Berichten und in den Reviewprozess von Berichten einzubeziehen.

Literatur

- [1] Jonas WB, Anderson RL, Crawford CC, Lyons JS. A systematic review of the quality of homeopathic clinical trials. BMC Complement Altern Med 2001; 1 :12.
- [2] Haake M, Muller HH, Schade-Brittinger C et al. The German multicenter, randomized, partially blinded, prospective trial of acupuncture for chronic low-back pain: a preliminary report on the rationale and design of the trial. J Altern Complement Med 2003; 9: 763-770.

2.13 Beurteilung bevölkerungsweiter Interventionen / Prävention

Eine methodische Prüfung der vorliegenden Studien zu bevölkerungsweiten Interventionen erfolgt nach den vorab genannten Kriterien zur Prüfung der einzelnen Studientypen.

Die zentrale Anforderung, die darüber hinaus an präventive Maßnahmen und Maßnahmen der Gesundheitsförderung gestellt wird, ist, dass sie wirksame Strategien beinhalten, die die Gesundheit der Bevölkerung erhalten und damit Lebensqualität, Mobilität und Leistungsfähigkeit der Bevölkerung nachhaltig verbessern.

Die Vielzahl der möglichen präventiven Maßnahmen und die Begrenztheit der Mittel machen eine Priorisierung der Maßnahmen erforderlich. Dabei sind die allgemeinen Grundsätze zur Priorisierung von Gesundheitsthemen (s.a. Abschnitt 4.10) zu berücksichtigen.

Bisherige Präventionsprogramme setzen mit Hinweis auf die Beeinflussbarkeit von Defiziten die Schwerpunkte oft im jüngeren Lebensalter. Für das höhere Lebensalter [1], die Gesundheitsförderung in Betrieben, Schulen, Einrichtungen des Gesundheitswesens oder in sozialen Brennpunkten bestehen jedoch Angebotslücken [2].

Die präventiven Maßnahmen müssen klar an spezifischen Zielgruppen ausgerichtet sein. Dem muss eine genaue Analyse der Bedürfnisse, Erwartungen und Gewohnheiten der Zielgruppe vorausgehen. Unter der Voraussetzung, dass es sich um eine sinnvolle Intervention handelt, ist zu klären, welche Voraussetzungen zu schaffen sind, damit alle Mitglieder der Zielgruppe einen gleichberechtigten Zugang zu der Intervention haben.

Eine Nutzenbewertung hat sowohl zeitliche (mittelfristige oder langfristige Verbesserung der Gesundheit) als auch perspektivische Dimensionen (aus Sicht des Einzelnen, gesellschaftliche Aspekte, ökonomische Aspekte). Zu Beurteilung des Nutzens müssen klare Wirkungsziele der Intervention angegeben sein.

Programme, die auf langfristige Wirkungen ausgelegt sind, sollten zu einer Senkung der vorzeitigen Mortalität und/oder Morbidität und zur Verbesserung und/oder zum Erhalt der Lebensqualität führen.

Zur Beurteilung mittelfristiger Wirkungen und zur Abschätzung des Erfolgs solcher Programme dienen insbesondere folgende Parameter:

- gesundheitliche Outcome-Parameter
- gesundheitsbezogene Lebensqualität [3]
- Bewertung der Kompetenzentwicklung (z. B. Wissen, Selbstbewusstsein, Einstellung)
- Beurteilung der Zugangswege
- Zufriedenheit der Zielgruppe mit der Struktur und Durchführung des Programms
- Strukturbildung (capacity building) [4].

Um abzuschätzen, ob und unter welchen Umständen der Nutzen stabil ist, sind auch Sensitivitätsanalysen zu fordern.

Zur Bewertung des Netto-Nutzens für den Einzelnen oder die Gesellschaft, sind gesundheitsökonomische Analysen zu fordern (s.a. Abschnitt 2.5), ggf. auch unter Einbezug von Studienergebnissen aus betrieblichen oder gemeindebezogenen Projekten.

Literatur

- [1] World Health Organization. The Jakarta Declaration on Leading Health Promotion into the 21st Century. WHO 1997, Geneva.
- [2] Badura B. Gesundheitsförderung und Prävention – Schritte in eine „gesunde Gesellschaft“. Public Health Forum 2000; 28: 5-7.
- [3] Bullinger M. Erfassung der gesundheitsbezogenen Lebensqualität mit dem SF-36. Health Survey. Rehabilitation 1996; 35: 17-27.
- [4] Hawe P, Noort M, King L, Jordaens C. Multiplying health gains: the critical role of capacity-building within health promotion programs. Health policy 1997; 39: 29-42.

2.14 Qualitätsmanagement in der Versorgung

Zahlreiche Paragraphen des SGB V sowie zahlreiche Initiativen im ambulanten und im stationären Bereich zielen auf die Verbesserung der *Qualität* der gesundheitlichen Versorgung. Ebenso befassten sich viele neue Versorgungsmodelle mit einer Steigerung der Qualität und einer höheren Kosteneffizienz. Das Institut könnte in Abstimmung mit dem G-BA und anderen Gremien sowie den Ressorts „Gesundheitsökonomie“ und „Biometrie“ die Aufgabe einer Erfassung und Bewertung der verschiedenen Qualitätsaspekte und -modelle übernehmen.

2.15 Beschreibung der Art und Größe des Placeboeffektes

Die Planung, Durchführung und Auswertung einer Studie können durch schwer kontrollierbare Einflüsse einseitig beeinflusst werden und dadurch ihre Ergebnisse verändern. Diese Beeinflussung geschieht vor allem, wenn die aktiven und passiven Studienteilnehmer wissen, wer welcher Gruppe zugeordnet ist [1]. Die Verblindung einer Studie und die Verwendung einer Substanz, die keine pharmakologischen Effekte ausweist, haben dabei zum Ziel, diese Einflüsse zu minimieren. Darüber hinaus hat Placebo möglicherweise eigene Einflüsse auf medizinische Parameter, vor allem solche die einer subjektiven Einschätzung unterliegen wie Schmerz [2]. Bei der Interpretation von placebokontrollierten Studien ist es daher notwendig zu überprüfen, ob durch die Verwendung eines Placebo tatsächlich eine Verblindung gewährleistet wurde. Der Effekt von Placebo scheint mit der

Größe der Studie abzunehmen [2]. Daher ist eine Fehlinterpretation von nicht-placebokontrollierten Studien vor allem bei relativ kleinen Fallzahlen zu befürchten. Bei einigen Interventionen kann eine Verblindung aber unmöglich sein, zum Beispiel bei therapeutischen Interventionen mit subjektiv eindeutig identifizierbaren erwünschten oder unerwünschten Wirkungen. Dabei ist es in einer Studie sinnvoll, zu ermitteln, wie viele Patienten und Ärzte korrekt die zugeordnete Gruppe erraten haben. Ist bei der Intervention die Größe des Placeboeffektes bekannt, kann daraus ggf. der Grad der Verzerrung der Ergebnisse durch fehlende bzw. unmögliche Verblindung ermittelt werden. Das Fehlen einer Placebogruppe kann dadurch bedingt sein, dass bereits eine wirksame Therapie für eine Störung oder Erkrankung vorliegt, und es mit der Helsinki Konvention unvereinbar wäre, eine nicht-aktive Therapie in der Kontrollgruppe durchzuführen [3]. In einem solchen Fall ist ggf. eine Verblindung der Therapie auch ohne eine Placebogruppe notwendig.

Literatur

- [1] Gillespie R. Manufacturing knowledge: a history of the Hawthorne experiments. Cambridge, England: Cambridge University Press, 1991.
- [2] Hróbjartsson A and Gøtzsche PC. Is the placebo powerless? N Engl J Med 2001; 344: 1594-1602.
- [3] Emanuel EJ and Miller FG. The ethics of placebo-controlled trials. N Engl J Med 2001; 345: 915-919.

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 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 (z.B. Arzt-Patient-Beziehung) zu unterstützen, unter Berücksichtigung der individuellen Bedürfnisse und Wertvorstellungen der Patienten bei der Anwendung medizinischer Tests und Therapien.

Das Institut versteht sich nicht als 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 Webseite wird das erste Instrument für die Verbreitung der evidenzbasierten Gesundheitsinformationen des Instituts sein.

Die Webseite wird zu einem umfassenden und vielseitigen Nachschlagewerk ausgebaut, welches 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 Webseite wird ein elektronischer Newsletter sowie z.B. elektronisch herunterladbare Texte bzw. sonstige Dateien mit gesundheitsrelevanten Themen als Transportwege für die Gesundheitsinformationen angegliedert, und es können elektronische Informationen auf anderen Webseiten reproduziert werden. Darüber hinaus wird die Möglichkeit bestehen, Printversionen der Gesundheitsinformationen herunter zu laden.

3.2.1 Informationsprodukte

Zu den Informationsprodukten zählen ausführliche Informationen, Kurzinformationen und sonstige Produkte.

Ausführliche Informationen

Die ausführlichen Informationen bilden den Schwerpunkt des entstehenden Nachschlagewerkes, welches umfassend über eine Vielzahl von gesundheitsrelevanten Themen informiert. Ausführliche Informationen können zudem aus wissenschaftlichen Berichten des Instituts entwickelt werden. Die ausführlichen Informationen beziehen sich auf

- Erklärungen zu Erkrankungen und Beschwerden, einschließlich
 - Anatomie
 - Physiologie
 - Erkrankungsursache
 - Symptomerkennung
 - normaler Verlauf von Erkrankungen
 - Prognose
 - mögliche Komplikationen
 - Erkennung von Komplikationen

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

Kurzinformationen

Neben den ausführlichen Informationen werden in weit größerer Anzahl Kurzinformationen erstellt. Diese Kurzinformationen sollen das Nachschlagewerk ergänzen. 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 von wissenschaftlichen Berichten des Instituts erstellt.

Sonstige Produkte

Unter sonstigen Produkten sind visuelle und interaktive Tools wie komplexe Schaubilder, Online-Rechner (z.B. für Zigarettenkosten) und Glossar (z.B. Online-Wörterbuch) zu verstehen.

Die Tools sollen z.B.

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

3.2.2 Redaktionssystem

Das Ressort „Patienteninformation - Erstellung und Methodik“ wird in Absprache mit der Institutsleitung und den entsprechenden Ressorts sicher stellen, dass die Inhalte der Gesundheitsinformations-Webseite 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 psychologischen Bedürfnissen angepasst sind,
- mit allen anderen vom Institut veröffentlichten Informationen konsistent sind.

Die Leiterin oder der Leiter des Ressorts „Patienteninformation - Erstellung und Methodik“ wird Chefredakteurin bzw. Chefredakteur der Gesundheitsinformationen sein.

Interne und externe Kritik über Inhalt und Qualität der Gesundheitsinformationen werden an die Chefredakteurin bzw. den Chefredakteur weitergeleitet. Dieser bzw. 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 Chefredakteurin bzw. Chefredakteur, Institutsleitung, Ressort „Öffentlichkeitsarbeit und Qualitätsmanagement“ 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 Steuergremiums.

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 (z.B. 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. Die Qualität der Übersetzungen wird stichprobenartig überprüft.

Das Ressort „Patienteninformation – Erstellung und Methodik“ wird für die Übersetzung auf hausinterne bilinguale Mitarbeiter mit entsprechendem Fachwissen zurückgreifen. Darüber hinaus werden die Fremdsprachenkenntnisse der Mitarbeiter des Ressorts „Patienteninformation - Erstellung und Methodik“ in Weiterbildungen gefördert. Zusätzlich werden Programme für eine computergestützte Übersetzung angewandt. Diese enthalten verschiedene Computertools, die Phrasen und Terminologien sammeln und vereinheitlichen und dadurch die Übersetzung durch den Menschen erleichtern und standardisieren können.

3.3 *Entwicklung der Informationsprodukte*

Bei der Entwicklung der Informationsprodukte wird als wesentliches Merkmal zur Qualitätssicherung neben interner auf externe Beratung zurückgegriffen. Ist im Folgenden von „Beratungsgruppe“ die Rede, so sind darunter Patientenvertreter und die Patientenbeauftragte oder der Patientenbeauftragte im Kuratorium sowie der Patienteninformationsausschuss im G-BA zu verstehen.

3.3.1 Themenauswahl

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

Da die Erstellung und Pflege der Gesundheitsinformationen viele Ressourcen binden wird, 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

c. Festlegung der Filterkriterien

d. Evaluation.

Anhand dieser Filterkriterien erfolgt die systematische Auswahl der Themen nach folgenden Schritten:

a. 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 auf die in der 1. Stufe ausgewählten Themen

3. Stufe: interne und externe Beratung über die in der 2. Stufe ausgewählten Themen

b. abschließende Themenauswahl durch das Steuergremium

c. Evaluation der ausgewählten Themen

Der gesamte Prozess zur Erstellung von Gesundheitsinformationen ist in der Abbildung 3 im Abschnitt 3.3.3 zusammenfassend dargestellt.

3.3.2 Fragestellung und Inhalte

Es gibt eine Reihe von Methoden, die als Entscheidungshilfen für die im Rahmen einer Gesundheitsinformation zu berücksichtigenden Fragestellungen dienen. 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 qua-

litativ hochwertige Daten, Umfragen und Studien nutzen, die mitunter durch Telefoninterviews mit Experten (key informants) und/oder Fokus-Gruppen sowie Beratungsgruppen (s.a. 1. Absatz zu Abschnitt 3.3) ergänzt werden (s.a. 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 werden interne Projektgruppen gebildet. Diese werden sich aus Mitarbeitern des Ressorts „Patienteninformation - Erstellung und Methodik“ sowie anderer Ressorts zusammensetzen. Die Bildung der Projektgruppe sowie die interne Projektkoordination werden vom Ressort „Patienteninformation - Erstellung und Methodik“ 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 (z.B. Diagnoseverfahren und Therapien). Hierfür werden eine Literaturrecherche, die Recherche von sog. Schlüssel-Webseiten (die wichtigsten Webseiten mit Informationen zu Gesundheitsthemen), Telefoninterviews sowie die Befragung von key informants, z.B. Patientenvertreter und klinische Experten, durchgeführt.
- c. Die Bedürfnisse der Bürger und Patienten, ihres aktuellen Kenntnisstandes sowie ihrer potenziellen Interessen hinsichtlich des Themas werden so weit wie möglich analysiert. Hierfür werden wiederum eine Literaturrecherche einschließlich der Recherche von Webseiten 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.

Der gesamte Prozess zur Erstellung von Gesundheitsinformationen ist in der Abbildung 3 in Abschnitt 3.3.3 zusammenfassend dargestellt.

3.3.3 Erstellung

Grundsätzlich werden für die Erstellung der Gesundheitsinformationen dieselben evidenzbasierten Methoden angewendet wie im Institut insgesamt. Entsprechend dem Entwurf (s.a. Abschnitt 3.3.2) werden die einzelnen Informationen wie folgt entwickelt.

Ausführliche Informationen

Die Erstellung von ausführlichen Informationen erfolgt entsprechend der folgenden Schritte.

- a. Literaturrecherche nach publizierten systematischen Übersichten. Die Validität und Aktualität der gefundenen systematischen Übersichten werden diskutiert und bewertet, sowie ggf. nachfolgend publizierter Studien.
Stehen keine Systematischen Übersichten, sondern nur einzelne Studien zur Verfügung, bilden diese die Grundlage für die Bewertung und Verwendung.
- b. Erstellung einer vorläufigen Version der ausführlichen Informationen aus den Ergebnissen der Literaturrecherche.
- c. Internes Peer Review:
mehrfache Durchsicht der vorläufigen Version. Das resultierende Feedback bzw. Kritik werden diskutiert, gegebenenfalls erfolgt eine weitere Analyse und Durchsicht.
- d. Externes Peer Review:
Expertengutachten werden von Patientenvertretern, Vertretern aus dem Bereich der Gesundheitsförderung, klinischen Experten sowie der Beratungsgruppe (s.a. 1. Absatz im Abschnitt 3.3) 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.
- e. Falls erforderlich, wird die Lesbarkeit und Verständlichkeit der ausführlichen Informationen durch Vertreter der Zielgruppe getestet (deutsche Version). Feedback und erforderliche Korrekturen werden mit den entsprechenden Ressorts des Instituts sowie der Institutsleitung diskutiert.
- f. Die überarbeitete Version der ausführlichen Informationen wird in deutscher und englischer Sprache verfasst und an das Steuergremium des Instituts geleitet. Anschließend wird sie als Internet-Testversion aufbereitet.
Die Anwendbarkeit der (deutschen) online-Version wird mit 3-5 Freiwilligen einschließlich mindestens einer Patientin oder eines Patienten bzw. einer Patientenvertreterin oder eines Patientenvertreters getestet [15-17]. Danach wird die englische Version als Internet-Testversion (offline) aufbereitet. Schließlich werden alle weiteren Kommunikationsinstrumente (z.B. Download-Versionen) entwickelt, getestet und fertig gestellt.
- g. Werden anhand des Quellenmaterials mehrere Informationsversionen oder -produkte entwickelt, ist es notwendig, dass diese einen vergleichbaren Qualitätssiche-

rungsprozess durchlaufen. Dies trifft nur zu, wenn der Inhalt in wesentlichen Teilen verändert werden soll.

- h. Vor ihrer Veröffentlichung werden die ausführlichen Informationen mit der Institutsleitung und den fachlich zuständigen Ressortleitern abgesprochen. Das Steuergremium kann die endgültige Version freigeben oder alternativ eine weitere Beratung oder Überarbeitung vorschlagen (s.a. Abschnitt 3.2.2).

Kurzinformationen

Die Erstellung der Kurzinformationen erfolgt anhand von:

- kurzen Einzelartikeln, die eine einzelne systematische Übersichtsarbeit oder wichtige Studie, oder einige wenige Übersichten oder Studien zusammenfassen
- Artikeln auf der Basis von Leitfäden, die speziell für die Erstellung von Kurzinformationen (vom Institut) entwickelt wurden
- Artikeln, die entweder durch das Ressort „Patienteninformation – Erstellung und Methodik“ und gegebenenfalls gemeinsam mit fachlich relevanten Ressorts erstellt und mit dem Steuergremium abgesprochen werden oder Artikeln, die von den fachlich relevanten Ressorts erstellt und vom Ressort „Patienteninformation – Erstellung und Methodik“ bezüglich Inhalt und Kommunikationsstandards redaktionell überarbeitet werden
- Artikeln, die entweder mittels der gleichen Vorgehensweise, die auch bei den ausführlichen Informationen angewendet wird, überprüft werden oder Artikeln, die nur intern in Zusammenarbeit mit den Autoren von berücksichtigten Übersichten und Studien evaluiert werden. (Den Autoren wird dadurch die Möglichkeit geboten, die Interpretation ihrer Arbeiten zu kommentieren)

Dem Ressort „Patienteninformation – Erstellung und Methodik“ obliegt die Verantwortung für die Pflege und Aktualisierung der Kurzinformationen.

Sonstige Produkte

Visuelle und interaktive Tools werden selber erstellt, für Deutschland adaptiert, käuflich erworben oder extern in Auftrag gegeben. Sie sollen sowohl den Kommunikationsstandards des Ressorts „Patienteninformation – Erstellung und Methodik“ als auch den Ansprüchen der Gesundheitsförderung entsprechen.

Der gesamte Prozess zur Erstellung von Gesundheitsinformationen ist in Abbildung 3 (folgende Seite) zusammenfassend dargestellt.

3.3.4 Evidenzbasierte Kommunikationsstandards

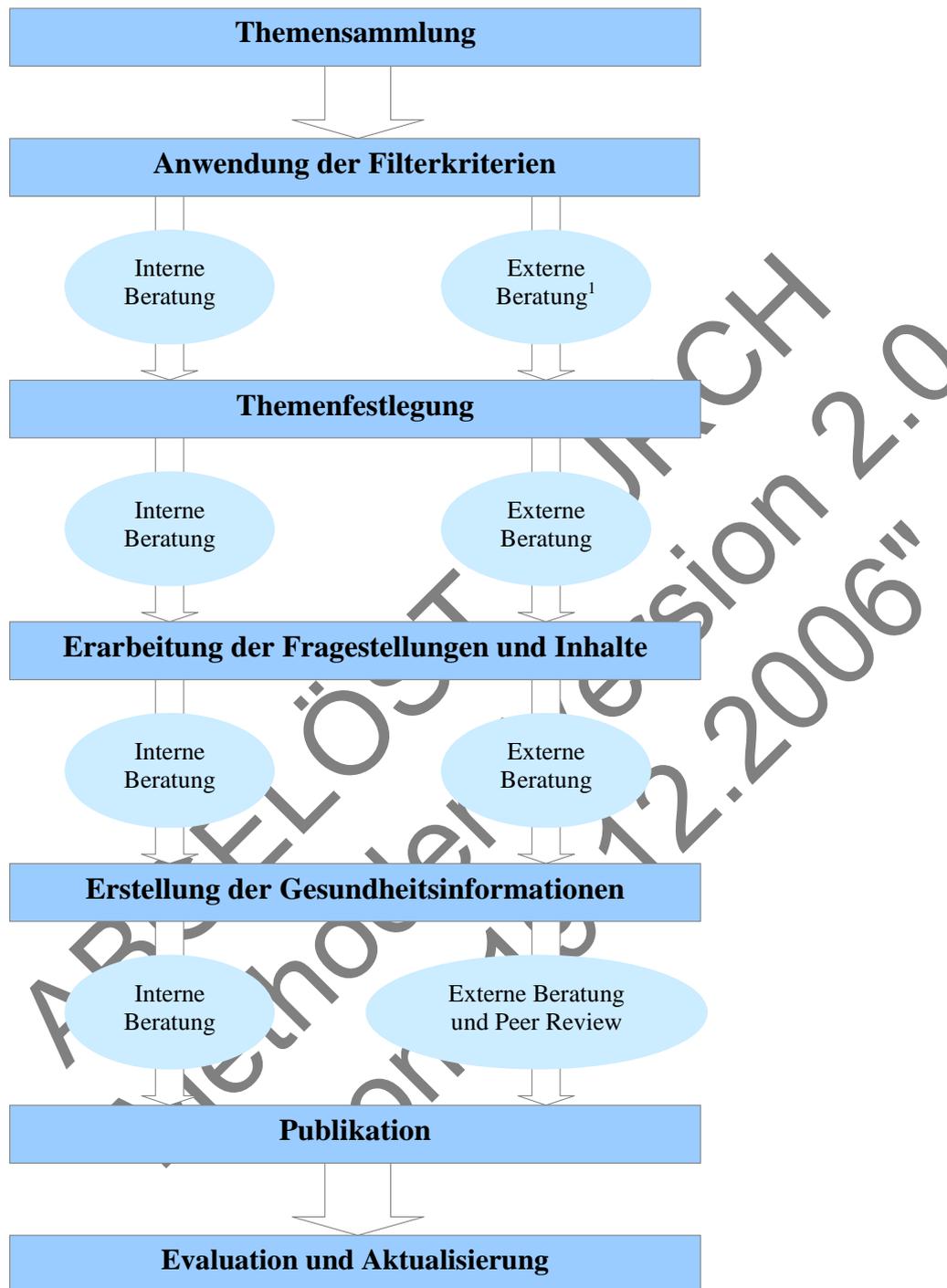
Das Institut hat bei der Kommunikation der Gesundheitsinformationen folgenden Anspruch:

- respektvoll und effektiv mit der deutschen Bevölkerung zu kommunizieren, so dass diese dem Institut als zuverlässige und allgemeinverständliche Informationsquelle vertraut;
- die Gesundheitsinformationen sollen leicht und unterhaltsam zu lesen 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, ihrer Wertvorstellungen und Anliegen, ihrem Selbstbestimmungsrecht und gegenüber kulturellen Besonderheiten sollen klar und deutlich erkennbar sein;
- die Patienten-Kompetenz (patient empowerment) soll gestärkt,
- 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 Webseite entsprechend den Standards der Accessibility Initiative gewährleisten [18].

Das Institut wird seine Wertvorstellungen durch seine Professionalität und hohe Qualität der von ihm erstellten Gesundheitsinformationen sowie durch die verwendete Sprache und die verwendeten Illustrationen klar zu erkennen geben. Für jedes Informationsprodukt wird ein institutsinterner Leitfaden (style guide) entwickelt. Dieser Leitfaden wird so weit wie möglich auf Grundlage wissenschaftlich nachgewiesener Effektivität erstellt. Darüber hinaus muss er unter der Bevölkerung Deutschlands populär sein.

Abbildung 3: Entwicklung der Gesundheitsinformationsprodukte



¹ s.a. 1. Absatz im Abschnitt 3.3.

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 wissen-

schaftlichen 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 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 zusammen mit absoluten Risiken oder der Anzahl der Patienten, die behandelt werden muss, bevor ein Patient von der Behandlung profitiert bzw. durch sie einen Schaden erleidet.

Personalisierte bzw. individualisierte Risikoabschätzungen sind als effektive Kommunikationsformen für Gesundheitsinformationen nachgewiesen [7,19]. Das Ressort „Patienteninformation – Erstellung und Methodik“ kann Tools entwickeln oder adaptieren, durch die Bürger und Patienten ihr persönliches Risiko einschätzen können, wenn sehr verlässliche Daten für deren Entwicklung vorliegen.

„Patient decision aids“ (Entscheidungshilfen für Patienten) sind effektive Maßnahmen zur Vermittlung von Gesundheitsinformationen, und sie unterstützen Patientenentscheidungen [20]. Das Institut wird gegebenenfalls decision aids entwickeln oder adaptieren und versuchen, effektive Elemente der decision aids in die Gesundheitsinformationen einzubauen.

Zusätzlich wird das Ressort „Patienteninformation – Erstellung und Methodik“

- Gesundheitsinformationen in konsistenter Weise präsentieren,
- den Grad der Unsicherheit erklären, der mit den Informationen verbunden sein kann,
- aufzeigen, für wen (d.h. für welche Bevölkerungsgruppe/n) die dargestellten wissenschaftlichen Nachweise gelten,
- versuchen, für die Webseite höchst mögliche Standards für Benutzerfreundlichkeit (inklusive Navigation) zu erfüllen,
- sehr klar und sorgfältig auf den Unterschied zwischen „fehlendem wissenschaftlichem 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 Pro-

dukte verwendet, und nur falls notwendig durch die Markennamen erhältlicher Produkte ergänzt,

- ein durch Hypertext verlinktes Online-Glossar erstellen, das die Verständlichkeit der Gesundheitsinformationen auf der Webseite erhöht.

3.4 Publikation

Die Distribution der Gesundheitsinformationen wird mit der Institutsleitung, fachlich zuständigen Ressortleitern und dem Ressort „Öffentlichkeitsarbeit und Qualitätsmanagement“ besprochen. Änderungsvorschläge werden an das Steuergremium weitergeleitet. Die Institutsleitung, die Leiterin oder der Leiter des Ressorts „Patienteninformation – Erstellung und Methodik“ 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-Webseite publiziert (s. a. 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]. Des Weiteren bleibt unklar, ob sie tatsächlich das messen, was sie zu messen vorgeben. Somit steht kein zuverlässiger Qualitätsindikator für die Erstellung von Gesundheitsinformationen zur Verfügung [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 über Diagnose- und Screeningverfahren beinhalten jedoch komplexere Entscheidungsprozesse und Kommunikationsaspekte als Informationen über 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 nach den jeweiligen Ergebnissen ausrichten.

Das Ressort „Patienteninformation - Erstellung und Methodik“ wird neue wissenschaftliche Erkenntnisse über die Möglichkeiten der Qualitätsbewertung verfolgen und seine Methoden gegebenenfalls anpassen. Des Weiteren werden die Gesundheitsinformationen des Instituts soweit 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-Übersichtsarbeiten 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-Webseite wird für die Leser Feedback Mechanismen bereitstellen. Jegliche Art von internem und externem Feedback kann zu unmittelbarer Überarbeitung bereits veröffentlichter bzw. zu einer Überarbeitung der Entwürfe von Gesundheitsinformationen führen.

Die enge inhaltliche Kooperation des Ressorts „Patienteninformation - Erstellung und Methodik“ mit fachlich relevanten Ressorts gewährleistet, dass die erstellten Gesundheitsinformationen dem wissenschaftlichen Stand entsprechen. Seinem Anspruch an den Qualitätsprozess wird das Ressort „Patienteninformation - Erstellung und Methodik“ 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 et al. Do internet interventions for consumers cause more harm than good? A systematic review. Health Expectations 2002; 5: 28-37.
- [3] National Institute of Clinical Studies. The impact of the internet on consumers' health behaviour. Prepared by the Centre for General Practice and the Centre for Evidence Based Practice, University of Queensland. NICS, Melbourne: 2003.

- [4] Lewis D. Computer-based approaches to patient education: a review of the literature. *J Am Informatics Association* 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 et al. 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 (eds). *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 (eds). *Evidence-Based Patient Choice: Inevitable or Impossible?* Oxford: Oxford University Press, 2001; 289-307.
- [9] Henshall C, Oortwijn W, Stevens A et al. 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 Technology Assessment* 2003; Vol 7: No 20. NHS R&D HTA Programme.
- [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, June 2004.
- [12] Liberati A, Sheldon TA, Banta HD et al. 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, Campben CV, Peters L. Quality of care from the patients' perspective: from theoretical concept to a new measuring instrument. *Health Expectations* 1998; 1: 82-95.
- [14] DiPEX. The Database of Patients' Experiences. <http://www.dipex.org>. Zugriff am 18. Oktober 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] W3C. Web Accessibility Initiative. <http://www.w3.org> [Zugriff am 26. Januar 2005]
- [19] 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, UK: John Wiley & Sons, Ltd.
- [20] O'Connor AM, Stacey D, Rovner D et al. Decision aids for people facing health treatment or screening decisions (Cochrane Review) (last updated October 2003). In: *The Cochrane Library*, Issue 3, 2004. Chichester, UK: John Wiley & Sons, Ltd.
- [21] Jadad AR, Gagliardi A. Rating health information on the internet: navigating to knowledge or to Babel? *JAMA* 1998; 279: 611-614.
- [22] Eysenbach G. Consumer health informatics. *BMJ* 2000; 320: 1713-1716.

- [23] 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.
- [24] 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 Expectations* 2004; 7: 165-175.
- [25] van den Brink-Muinen A, Verhaak PF, Bensing JM et al. Doctor-patient communication in different European health care systems: Relevance and performance from the patients' perspective. *Patient Education Counseling* 2000; 39: 115-127.
- [26] 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.

ABGELÖST DURCH
"Methoden Version 2.0
vom 19.12.2006"

4. Berichtserstellung

4.1 *Produkte*

Gemäß seinem gesetzlichen Auftrag erstellt das Institut verschiedene Produkte in Form von wissenschaftlichen Berichten und allgemein verständlichen Informationen für Bürger und Patienten (s.a. Präambel). Die Erstellung dieser Produkte wird sich dabei nicht in der in diesem Dokument niedergelegten Methodik unterscheiden, sondern allenfalls im Hinblick auf den Einbezug bestimmter Abläufe. Wissenschaftliche Berichte folgen den in Abschnitt 4.5 aufgeführten Schritten. In besonderen Situationen kann es erforderlich sein, auf tagesaktuelle Themen zeitnah und ohne großen zeitlichen Verzug zu reagieren. Hier erfolgt in Abstimmung mit dem G-BA und ggf. weiteren verantwortlichen öffentlichen Institutionen (z.B. dem Bundesinstitut für Arzneimittel und Medizinprodukte) ein verkürzter Prozess, der in Abschnitt 4.6 beschrieben ist. Für die Erstellung von allgemein verständlichen Informationen für Bürger und Patienten gelten die gleichen methodischen Prinzipien wie für die Erstellung von wissenschaftlichen Berichten, es sei denn im Kapitel 3 werden darüber hinausgehende methodische Anforderungen formuliert.

4.2 *Auswahl externer Sachverständiger*

Das Institut wird über seine Internetseiten aktuell über die zur Bearbeitung anstehenden Aufträge bzw. geplanten Projekte informieren. Gemäß seinem gesetzlichen Auftrag wird das Institut externe Sachverständige in seine Arbeit einbeziehen, z.B. bei der Berichtserstellung oder beim Review von Berichten. Zur Mitarbeit bei der Berichtserstellung ist es für interessierte Personen oder Arbeitsgruppen fragestellungsbezogen erforderlich

- erstens dem Institut ihre potentiellen Interessenskonflikte darzulegen,
- zweitens dem Institut gegenüber ihre fachlich-medizinische und methodische Expertise in der praktischen Anwendung, Forschung und Lehre auszuweisen (einschließlich eventueller Vorarbeiten) und
- drittens glaubhaft zu versichern, dass sie über ausreichend personelle Ressourcen verfügen, die zur Bearbeitung anstehenden Arbeitsschritte in einer angemessenen Zeit zu erledigen.

Ein Formblatt zur Darlegung potentieller Interessenskonflikte wird über die Internetseiten des Instituts zur Verfügung gestellt. Dieses Formblatt muss von jeder an der wissenschaftlichen Arbeit zu beteiligenden Person ausgefüllt und dem Institut zur Verfügung gestellt werden. Die Erklärung

über die fachlich-medizinische und methodische Expertise kann dagegen gemeinsam für die zu beteiligenden Personen (z.B. als Arbeitsgruppe) erfolgen.

Das Steuergremium wird aus dem Kreis der Interessierten unter Abwägung der oben aufgeführten Kriterien (potentielle Interessenskonflikte, fachlich-medizinische und methodische Expertise sowie ausreichende personelle Ressourcen) fragestellungsbezogen Kooperationspartner, ggf. auch zur Bearbeitung von Teilarbeitsschritten, auswählen.

Externe Sachverständige, die die vereinbarte Arbeit oder Teile hiervon durch Dritte erfüllen lassen wollen, müssen diese verbindlich benennen. Für diese gelten insbesondere hinsichtlich ihrer potentiellen Interessenskonflikte die gleichen Maßgaben wie für die externen Sachverständigen selbst (s.a. Abschnitt 4.3).

Über die Vertragsbestandteile, insbesondere jedoch über die wissenschaftlichen Inhalte oder Arbeits(teil)ergebnisse ist bis zur Abnahme gegenseitiges Stillschweigen zu vereinbaren. Die externen Sachverständigen verpflichten sich zusätzlich zur Einhaltung ggf. 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 das Steuergremium entscheiden, ob Bedenken wegen potentieller Interessenskonflikte bestehen. Dafür müssen alle externen Sachverständigen und Institutionen eine Liste mit allen Tätigkeiten, die potentiell geeignet sein könnten, ihre fachliche Unabhängigkeit zu beeinflussen (Interessenskonflikte), dem Steuergremium darlegen (s.a. Abschnitt 4.2). Als Interes-

senskonflikte werden dabei in Anlehnung an entsprechende Regelungen wissenschaftlicher medizinischer Zeitschriften insbesondere folgende Kriterien angesehen: Alle finanziellen Abmachungen, Beschäftigungen, Beratungen, Honorare, bezahlte Experten-Aussagen, Reise-Unterstützungen, Patent-Anträge sowie der Aktienbesitz innerhalb der letzten 3 Jahre, die bzw. der die beauftragte Arbeit inhaltlich hätten beeinflussen können, sowie alle aktuellen persönlichen Beziehungen mit anderen Personen oder Organisationen, die die beauftragte Arbeit inhaltlich beeinflussen könnten [1]. Diese Kriterien-Liste findet sich auch in dem auf den Internetseiten des Instituts zur Verfügung gestellten Formblatt (s.a. Abschnitt 4.2). Das Formblatt wird gegebenenfalls aktualisiert. Maßgeblich ist immer das jeweils auf den Internetseiten abrufbare Formblatt.

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 potentiell 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 Vorstandsämter in Vereinen oder Gesellschaften. Zuwiderhandlungen können im Einzelfall zu einer Abmahnung führen bzw. im wiederholten oder schwerwiegenden Fall zu einem Kündigungsgrund. Im Einzelfall entscheidet 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 *Berichtsplan*

Der Berichtsplan beinhaltet, vergleichbar mit dem Studienprotokoll einer geplanten klinischen Studie, die genaue wissenschaftliche Fragestellung einschließlich der Zielkriterien (z.B. patientenrelevante Endpunkte) sowie die Darlegung der Methodik der Beschaffung und Bewertung der Informationen (einschließlich der Ein- und Ausschlusskriterien der zu beschaffenden und zu bewertenden Informationen). Darüber hinaus werden Angaben über die sich anschließenden Schritte bei

der Erstellung des Abschlussberichts, wie z.B. die Durchführung des Peer Review, Zeitpunkt der Veröffentlichung des Abschlußberichts etc. dargestellt, sofern diese zum Zeitpunkt der Erstellung des Berichtsplans bereits definiert wurden bzw. definiert werden können. Sofern externe Sachverständige in die Auftragsbearbeitung einbezogen werden, wird diesen Gelegenheit gegeben, einzelne Details des Berichtsplans in Abstimmung mit der Projektgruppe des Instituts zu präzisieren. Der Berichtsplan wird anschließend über die Internetseiten des Instituts veröffentlicht. Begründete Änderungen am Berichtsplan sind, wie auch bei einer klinischen Studie, möglich in Form einer Nachbesserung (Amendment), die ebenfalls über die Internetseiten des Instituts veröffentlicht wird.

Die im Berichtsplan getroffenen Angaben insbesondere zu den Ein- und Ausschlusskriterien relevanter Informationen (z.B. wissenschaftlicher Literatur) sind von zentraler Bedeutung für die Kriterien, denen spätere Stellungnahmen genügen müssen. Grundsätzlich werden Stellungnahmen möglich sein, die sich auf

- die Vollständigkeit der Informationsbeschaffung und
- die Richtigkeit der wissenschaftlichen Bewertung dieser Informationen

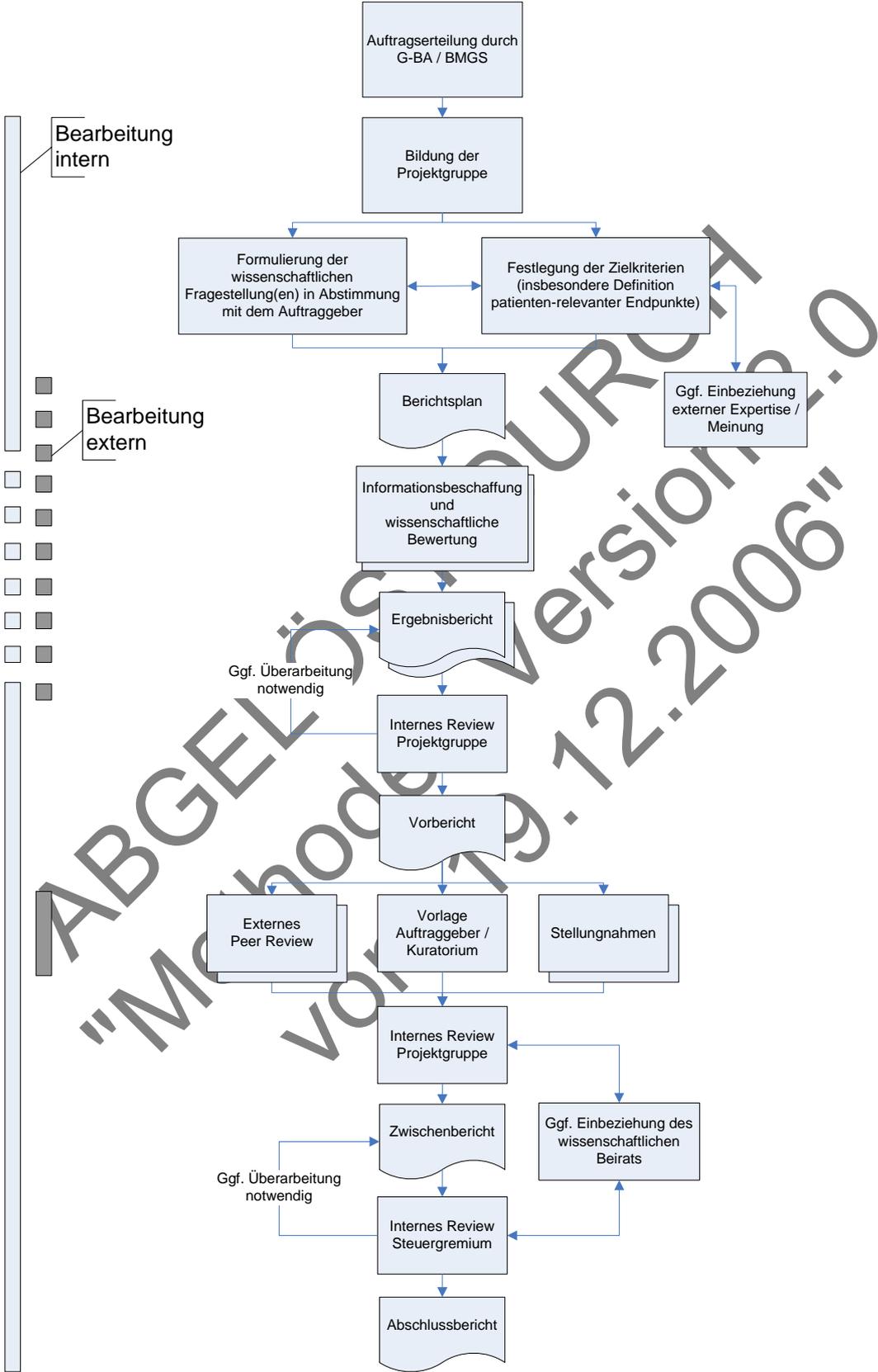
beziehen.

4.5 Struktur der Berichtserstellung

Der Ablauf der Berichtserstellung ist in der Abbildung 4 schematisch dargestellt.

Nach Auftragsvergabe durch den G-BA oder das BMGS erfolgt die Bildung einer Projektgruppe unter Leitung des beauftragten Ressorts. Eine Person wird als verantwortliche Projektleiterin oder 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, ggf. unter Einbeziehung externer Expertise und/oder Meinung, diese Definition verfeinert. Anschließend erstellt die Projektgruppe den Berichtsplan (s.a. Abschnitt 4.4). Dieser ist Grundlage für die darauf folgende Informationsbeschaffung und wissenschaftliche Bewertung. Die Informationsbeschaffung und wissenschaftliche

Abbildung 4: Ablauf der Berichtserstellung als Flussdiagramm



Bewertung kann, zumindest in Teilen, durch einen oder mehrere externe Sachverständige ausgeführt werden, die unter Berücksichtigung der unter Abschnitt 4.2 genannten Auswahl-Kriterien ihre Eignung dargelegt haben. Im Einzelfall wird eine Abstimmung mit der oder dem bzw. den externen Sachverständigen über die im Berichtsplan aufgeführten Arbeitsschritte einschließlich einer Präzisierung der Fragestellungen notwendig sein. Der Berichtsplan wird anschließend über die Internetseiten des Instituts veröffentlicht. Aus der Informationsbeschaffung und der sich daran anschließenden wissenschaftlichen Bewertung gehen je nach Zahl der externen Sachverständigen ein oder mehrere Ergebnisberichte hervor, die nach internem Review durch die Projektgruppe zum Vorbericht zusammengefasst werden. Dieser Vorbericht wird dann in der Regel zeitgleich

- einem oder mehreren externen Peer Reviewern 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,
- auf den Internetseiten des Instituts veröffentlicht, um Stellungnahmen, die den dort genannten Kriterien genügen müssen, einzuholen.

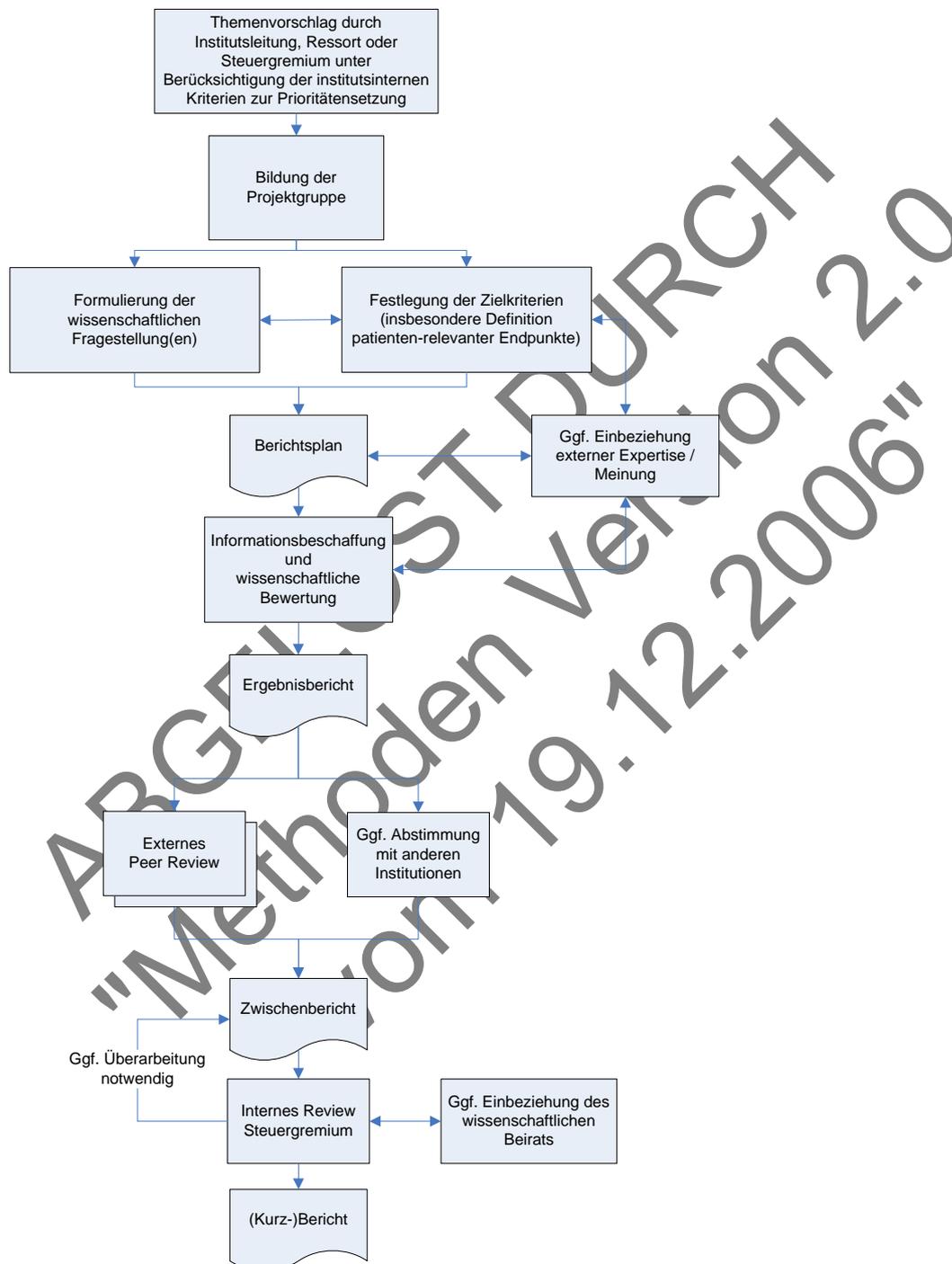
Je nach Art und Umfang der eingegangenen Stellungnahmen einschließlich der Ergebnisse des Peer Reviews sowie der Stellungnahmen des Auftraggebers und des Kuratoriums kann anschließend optional eine Diskussion unter Teilnahme der externen Sachverständigen durchgeführt werden, sofern dies der Verbesserung der wissenschaftlichen Qualität des späteren Abschlussberichts dienlich erscheint. Daran anschließend erstellt die Projektgruppe einen Zwischenbericht, der dem Steuergremium zum internen Review vorgelegt wird. Der sich daran anschließende Abschlussbericht wird dann dem Auftraggeber zugeleitet und nach dessen Zustimmung, spätestens aber nach einer im Berichtsplan vorab festgelegten Latenzzeit, durch das Institut veröffentlicht.

4.6 Verkürzte Berichtserstellung

In besonderen Situationen kann es notwendig werden, auf tagesaktuelle Themen zeitnah und ohne großen zeitlichen Verzug zu reagieren. Dies setzt eine verkürzte Beratungsphase und eine schnelle Entscheidungsfindung innerhalb des Steuergremiums voraus. Darüber hinaus können auch Abstimmungen mit externen Einrichtungen erforderlich werden. Aufgrund des hohen zeitlichen Drucks und der damit verbundenen Gefahr einer ungenauen oder fehlerhaften Aussage in solchen Fällen ist die Orientierung an vorab definierten Prozessabläufen zur Fehlerminimierung notwendig.

Der Ablauf einer verkürzten Berichtserstellung ist in Analogie zum üblichen Vorgehen (s.a. Abschnitt 4.5) in der Abbildung 5 schematisch dargestellt.

Abbildung 5: Ablauf einer verkürzten Berichtserstellung als Flussdiagramm



4.7 Veröffentlichung von Berichten

Wesentliche Aufgabe des Institutes ist es, vorhandene Evidenz einer sorgfältigen Prüfung zu unterziehen und die Ergebnisse dieser Prüfung mit dem Ziel der Verbesserung der individuellen und globalen Gesundheitsversorgung nutzbar zu machen.

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 einerseits und politischen oder interessengesteuerten Aspekten andererseits führen. Gleichzeitig muss vermieden werden, dass das Institut seinerseits bestimmte Ergebnisse zurückhält. Um dies auszuschließen, werden alle Ergebnisse, die das Institut im Rahmen seiner gesetzlichen Aufgaben erarbeitet, einschließlich des Berichtsplans und einer Zusammenfassung der zugrunde liegenden Materialien, die zur Ergebnisfindung beigetragen haben, zeitnah veröffentlicht.

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

Der Zeitrahmen für die Veröffentlichung ist Bestandteil des Berichtsplans.

4.8 Informationsbeschaffung

Die Informationen, die einschließlich ihrer wissenschaftlicher Bewertung die Grundlage der Institutsberichte darstellen, können vielfältiger Art sein, z.B. 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.

4.8.1 Allgemeine Grundsätze einer themenbezogenen Literaturrecherche

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

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

- die Auswahl der Datenquellen (z.B. öffentliche Datenbanken, nicht öffentliche Datenbanken, Handsuche in ausgewählten Fachzeitschriften, Kontaktaufnahme mit Fachexperten / Industrie / Patientenorganisationen etc.),
- die Suchtechnik in Bezug auf die Auswahl des Publikations- bzw. Studientyps (z.B. randomisierte, kontrollierte Studien, Fallbeobachtungen etc.),
- die Suchtechnik in Bezug auf die durch die Fragestellung vorgegebenen medizinischen Kriterien (z.B. Zielpopulation, Art der Intervention, Endpunkte etc.),
- die Suchtechnik in Bezug auf formale Merkmale der Publikation (z.B. Abstract-Publikationen, Sprache etc.),
- die Durchführung der Recherche selber (z.B. von einer Person oder von zwei Personen unabhängig voneinander).

Beispiele hierzu liefern die Publikationen [1-8].

Da die Relevanz dieser Kriterien für verschiedene Fragestellungen unterschiedlich sein kann, wird im Einzelfall geprüft, welche Kriterien zur Anwendung kommen. Liegen dabei für die einzelnen Kriterien in Bezug auf die Fragestellung keine validen Forschungsergebnisse vor, oder sind diese durch das Institut selbst noch nicht gesichtet und bewertet worden, so wird primär geprüft, ob etablierte Strategien internationaler oder nationaler Arbeitsgruppen (z.B. der Cochrane Collaboration) zur Anwendung kommen können, für die bereits weit reichende Erfahrungen vorliegen.

4.8.2 Ablauf einer Recherche nach Primärpublikationen in bibliographischen Datenbanken

Die Literaturrecherche gliedert sich in die drei Teile

1. Formulierung der Suchstrategie,
2. Anwendung der Suchstrategie (Primärrecherche),
3. Auswahl der relevanten Publikationen aus den Ergebnissen der Primärrecherche.

Die *Formulierung der Suchstrategie* erfolgt auf Basis der mit dem Auftraggeber abgestimmten Fragestellung und der definierten Zielkriterien. Dazu gehören gegebenenfalls auch die Ergebnisse, die im Ermittlungsverfahren patientenrelevanter Effekte und Größen erlangt wurden (s.a. Abschnitt 1.5). Sie beinhaltet insbesondere

- a) die Auswahl der Datenquellen,
- b) die Auswahl der Suchwörter und -kriterien bei Datenbankrecherchen,
- c) die Festlegung, welche Studien als relevant erachtet werden unter Angabe von Ein- und Ausschlusskriterien.

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 vorläufig formulierte Suchstrategie ist der Projektleiterin oder dem Projektleiter des Instituts darzulegen und mit dieser bzw. diesem abzustimmen.

Die *Anwendung der Suchstrategie* (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-Datenbank etc.). Sofern sich hierdurch relevante Übersichten identifizieren lassen, kann evtl. auf die für diese Übersichten durchgeführte Suche nach Primärpublikationen 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, sicher möglich ist. Zumeist wird auch in solchen Fällen eine Suche nach Primärpublikationen notwendig sein, die sich dann auf den nicht von den Übersichten berücksichtigten Publikationszeitraum beziehen wird. Wurden keine relevanten Übersichten identifiziert, so wird die Suche nach Primärpublikationen für den gesamten für die Fragestellung relevanten Publikationszeitraum durchgeführt.

4.8.3 Sonstige Datenquellen für die Primärrecherche

Neben der Suche in bibliographischen Datenbanken kann je nach Fragestellung eine Handsuche in ausgewählten Fachzeitschriften sinnvoll sein. Dies wird im Einzelfall entschieden. Darüber hinaus sind die vom Auftraggeber zur Berücksichtigung vorgelegten Unterlagen als ein Bestandteil der Primärrecherche anzusehen. Mit diesen Unterlagen wird dann im weiteren Prozess den sonstigen Grundsätzen der Informationsrecherche und –bewertung folgend verfahren.

Zusätzlich sind je nach Fragestellung weitere Datenquellen von erheblicher Bedeutung, wie z.B. Studienregister, Abstractbände wissenschaftlicher Kongresse oder, im Falle einer Arzneimittelbewertung, Zulassungsdatenbanken bzw. –korrespondenz, sofern diese dem Institut zugänglich sind. Sofern sich aus diesen Datenquellen Hinweise auf nicht oder nicht vollständig publizierte Studien ergeben, wird das Institut die jeweiligen Studienleiter und/oder Studiensponsoren kontaktieren und

um Bereitstellung aller für die Bewertung der Studien benötigten Informationen bitten. Dabei wird das Institut keine Geheimhaltungsverpflichtung akzeptieren. Wissenschaftler, Institutionen, Unternehmen und andere Personen oder Gruppierungen, die in diesem Zusammenhang die vom Institut benötigten Informationen nicht bereitstellen, werden in Konsequenz im weiteren Berichtsverfahren keine weitere Gelegenheit zur Abgabe von Stellungnahmen erhalten.

4.8.4 Auswahl relevanter Publikationen

Die *Auswahl der relevanten Publikationen* aus den Ergebnissen der Primärrecherche erfolgt zu meist über die zwei Arbeitsschritte

- Sichtung der 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. Nach Abschluss dieses Arbeitsschrittes ist vor Bewertung der Literatur das Rechercheergebnis der Projektleiterin oder dem Projektleiter des Instituts vorzulegen, im Einzelfall kann zur Einschätzung der Vollständigkeit der Literatur erneut eine Fachexpertin oder ein Fachexperte hinzugezogen werden. Sofern sich hieraus die Anforderung einer Modifikation und/oder Erweiterung der Literaturrecherche ergibt, ist diese nach den genannten Grundsätzen durchzuführen und das Ergebnis vor Bewertung der Literatur wiederum vorzulegen.

4.8.5 Dokumentation

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

- die Suchstrategie einschließlich der Begründung für die Auswahl der Datenquellen und der angewandten Suchtechniken,
- bei Abfrage bibliographischer Datenbanken zusätzlich
 - ein Recherche-Nachweis, vorzugsweise in Form einer Bildschirmkopie der Recherche-History,
 - eine Auflistung der durch Anwendung der Suchstrategie erlangten Primärergebnisse (Zitate, Abstracts),
 - eine Auflistung der nach Sichtung der Primärergebnisse als für die Fragestellung relevant erachteten Publikationen,

- eine Auflistung der nicht als relevant erachteten aber im Primäregebnis enthaltenen Literaturstellen einschließlich einer Begründung für die Nicht-Berücksichtigung,
- bei persönlicher Kontaktaufnahme mit Produktherstellern, Fachleuten, Fachgesellschaften etc.
 - Korrespondenznachweise,
- das Datum bzw. der Zeitraum der Recherche.

Die Form der Dokumentation (Papierkopie, Datenbankformat etc.) ist vorab mit der Projektleiterin oder dem Projektleiter des Instituts zu klären und sollte sich grundsätzlich an allgemeinen Vorgaben des Instituts orientieren.

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.

4.8.6 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 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 (s.a. Abschnitt 4.10).

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 (s.a. Abschnitt 4.6).

4.9 Fragestellungsbezogene Beweisführung

Der Nachweis für einen behaupteten günstigen Effekt von medizinischen Verfahren ist von denjenigen Institutionen oder Personen zu erbringen, die die Anwendung dieser Verfahren propagieren. Als "günstiger Effekt" soll in diesem Zusammenhang das Überwiegen von patientenrelevanten Vorteilen gegenüber Risiken durch die Anwendung des infrage stehenden medizinischen Verfahrens verstanden werden. Das heißt, die Nachweise für das Vorhandensein eines patientenrelevanten erwünschten Effekts und den Ausschluss von relevanten Risiken^k muss von denjenigen erbracht werden, die das Verfahren propagieren. Somit ist es im Interesse dieser Institutionen und Personen, für eine Bewertung des Verfahrens durch das Institut alle ihnen zur Verfügung stehenden Daten dem Institut ebenfalls zur Verfügung zu stellen. Es ist *nicht* Aufgabe des Instituts, einen fehlenden patientenrelevanten Effekt^l oder das Vorhandensein von relevanten Risiken nachzuweisen. Das Institut *kann* allerdings – soweit die verfügbare Datenbasis dies zulässt – in seinen Bewertungen solche Nachweise (nach dem jeweils gültigen Stand der Wissenschaft) erbringen.

^k Dies wird in aller Regel nur in einem inferenzstatistischen Sinn möglich sein, und zwar als die mit einem Irrtumsvorbehalt (Signifikanzniveau) getroffene Ablehnung der Hypothese, ein unerwünschter Effekt übersteige eine mehr als irrelevant angesehene Größenordnung.

^l s.a. vorherige Fußnote.

4.10 *Prioritätensetzung*

Ein Großteil der Arbeitskapazität des Instituts wird durch die Beantwortung externer Aufträge seitens des G-BA und des BMGS 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;
- Erstellung von Gesundheitsinformationen durch das Ressort Patienteninformation;
- 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 dem Institut - unerlässlich, Gerechtigkeit und Transparenz anzustreben.

Die Festlegung der Prioritäten für Forschungsaktivitäten sowie für die systematische Bewertung medizinischer Verfahren und Technologien (Health Technology Assessment, HTA) 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.

4.10.1 **Hintergründe für die Prioritätensetzung des Instituts**

Das Institut ist darum bemüht, durch seine Aktivitäten die Gesundheit der deutschen Bevölkerung 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:

- Anteil der Bevölkerung, 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 bzw. das Gesundheitswesen einen günstigen Einfluss auszuüben, unter Berücksichtigung des Gleichheitsprinzips;
- potentielle Risiken;
- Erstellung von Beiträgen, die ausschließlich das Institut gewährleisten kann, einschließlich der Einschätzung, woran andere arbeiten bzw. wer ebenfalls in der Lage sein könnte, bestimmte Aktivitäten zu übernehmen;
- Beitrag der Aktivität zur Qualitätssicherung der Institutsarbeit;
- potentieller Beitrag der Aktivität zur Evaluation von wissenschaftlichen Kenntnissen.

4.10.2 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 (s.a. 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.

4.11 Zeitpunkte der Berichtserstellung

Die Bewertung einer Methode bzw. Intervention ist zu jeder Zeit möglich. Generelle Vorgaben 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 auf Grund fehlender Langzeitstudien eine hohe Ergebnisunsicherheit festgestellt wird, wird diese den allgemeinen Arbeitsmethoden folgend dargestellt.

ABGELÖST DURCH
"Methoden Version 2.0
vom 19.12.2006"