

Allgemeine Methoden

Version 3.0 vom 27.05.2008

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

E-Mail: methoden@iqwig.de

ABGELÖST DURCH
"Allgemeine Methoden -
Version 4.0
vom 23.09.2011"

Allgemeine Anmerkungen:

Der erste Entwurf der Methoden des Instituts entstand im Herbst 2004 und wurde mit der Veröffentlichung am 01. November 2004 zur Diskussion gestellt. Nach dem Eingang von Stellungnahmen und Gutachten schloss sich im Februar 2005 eine Gesprächsrunde mit den Stellungnehmenden an, einschließlich der Beteiligung des wissenschaftlichen Beirats des Instituts. Daraus entstand die erste Version 1.0 vom 01. März 2005.

Im Jahr 2006 wurde das Dokument überarbeitet, wobei zwei aufeinanderfolgende Entwürfe zur Diskussion gestellt wurden; einmal ein interner Entwurf vom 27.04.2006, dann ein im Internet veröffentlichter zweiter Entwurf vom 28.09.2006. Unter Berücksichtigung der eingegangenen Stellungnahmen zu beiden Entwürfen entstand die zweite Version der Methoden des Instituts (Version 2.0) vom 19. Dezember 2006.

Unter Berücksichtigung des GKV-Wettbewerbsstärkungsgesetzes wurde das Dokument strukturell und inhaltlich überarbeitet. Da insbesondere für die Bewertung des Kosten-Nutzen-Verhältnisses von Leistungen der gesetzlichen Krankenversicherung neue Methoden entwickelt werden mussten, wurde das ehemalige Kapitel zur Gesundheitsökonomie ausgegliedert. Dieses wird in einem separaten Dokument („Methoden zur Bewertung des Kosten-Nutzen-Verhältnisses“) publiziert. Im Jahr 2007 wurden für die „Allgemeinen Methoden“ zwei Entwürfe zur Diskussion gestellt; einmal ein interner Entwurf vom 05.07.2007, der ausführlich mit dem wissenschaftlichen Beirat des Instituts diskutiert wurde, dann ein im Internet veröffentlichter weiterer Entwurf vom 15.11.2007. Unter Berücksichtigung der eingegangenen Stellungnahmen zu beiden Entwürfen entstand die vorliegende dritte Version der „Allgemeinen Methoden“ des Instituts (Version 3.0) vom 27. Mai 2008. Die „Allgemeinen Methoden“ des Instituts werden ergänzt durch ein Glossar, das auf der Website des Instituts verfügbar ist (www.iqwig.de/methoden.428.html).

Die Methoden des Instituts werden in der Regel jährlich auf eine notwendige Überarbeitung hin überprüft, es sei denn, Fehler im Dokument oder wesentliche Entwicklungen legen eine vorzeitige Aktualisierung nahe. Die projektspezifische Methodik wird auf Basis der zum jeweiligen Zeitpunkt gültigen Methodenversion festgelegt. Ergeben sich im Projektverlauf Änderungen des allgemeinen methodischen Vorgehens, wird geprüft, ob das projektspezifische Vorgehen entsprechend anzupassen ist.

In diesem Dokument wird bei der Angabe von Personenbezeichnungen jeweils die männliche Form angewandt. Dies erfolgt mit dem Ziel einer besseren Lesbarkeit. Somit handelt es sich um eine rein sprachliche Beschränkung. Die zentrale Rolle von „Geschlecht“ als wissenschaftliche Kategorie bleibt davon unberührt.

Präambel

Die „Allgemeinen Methoden“ erläutern die gesetzlichen und wissenschaftlichen Grundlagen des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Die Aufgaben und der organisatorische Aufbau des Instituts werden hier ebenso dargelegt wie die wissenschaftlichen Werkzeuge, die für die Bearbeitung der Institutsprodukte verwendet werden. Die entsprechenden Methoden zur Bewertung des Kosten-Nutzen-Verhältnisses von Leistungen der gesetzlichen Krankenversicherung werden im Dokument „Methoden zur Bewertung des Kosten-Nutzen-Verhältnisses“ dargelegt. Somit leisten die sogenannten Methodenpapiere des Instituts einen wichtigen Beitrag dazu, die Arbeitsweise des Instituts transparent zu machen.

Die „Allgemeinen Methoden“ richten sich vor allem an Fachwissenschaftler. Um aber möglichst allen Interessierten einen Zugang zu Informationen über die Arbeitsweise des Instituts zu verschaffen, haben sich die Autoren um Verständlichkeit bemüht. Wie jeder fachwissenschaftliche Text setzen allerdings auch die „Allgemeinen Methoden“ ein bestimmtes Maß an Vorwissen voraus.

Die „Allgemeinen Methoden“ können die Vorgehensweise nur allgemein beschreiben. Welche konkreten einzelnen Schritte das Institut bei der Bewertung einer bestimmten medizinischen Maßnahme unternimmt, hängt unter anderem von der jeweiligen Fragestellung und von der vorhandenen wissenschaftlichen Evidenz ab. Insofern sind die „Allgemeinen Methoden“ als eine Art Rahmen zu verstehen. Wie der Bewertungsprozess im konkreten Einzelfall ausgestaltet ist, wird projektspezifisch detailliert dargelegt.

Um seine Arbeitsweise kontinuierlich weiterzuentwickeln und zu verbessern, stellt das Institut seine „Allgemeinen Methoden“ öffentlich zur Diskussion. Für die jeweils gültige Fassung gilt das ebenso wie für Entwürfe der folgenden Versionen.

Inhaltsverzeichnis

Allgemeine Anmerkungen	i
Präambel	ii
Inhaltsverzeichnis	iii
1 Das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen	1
1.1 Gesetzliche Aufgaben	1
1.2 Struktur	3
1.3 Evidenzbasierte Medizin	6
1.3.1 <i>Praktische evidenzbasierte Medizin</i>	7
1.3.2 <i>Die Strategien der EbM</i>	8
1.3.3 <i>Bedeutung von EbM für das Institut</i>	9
1.3.4 <i>Evidenzbasierte Entscheidungsfindung im Gesundheitswesen</i>	9
1.3.5 <i>Die Bedeutung der Ergebnissicherheit</i>	9
1.3.6 <i>Die Verbindung von Ergebnissicherheit und Alltagsnähe</i>	11
2 Produkte des Instituts	14
2.1 Produktspezifische Verfahrensabläufe	14
2.1.1 <i>Berichte</i>	16
2.1.2 <i>Rapid Reports</i>	20
2.1.3 <i>Gesundheitsinformationen</i>	22
2.1.4 <i>Arbeitspapiere</i>	25
2.2 Generelle Aspekte bei der Produkterstellung	27
2.2.1 <i>Auswahl externer Sachverständiger</i>	27
2.2.2 <i>Gewährleistung der fachlichen Unabhängigkeit</i>	28
2.2.3 <i>Begutachtung der Produkte des Instituts</i>	29
2.2.4 <i>Veröffentlichung der Produkte des Instituts</i>	30
3 Nutzen- und Schadenbewertung medizinischer Interventionen	31
3.1 Patientenrelevanter medizinischer Nutzen	31
3.1.1 <i>Definition des patientenrelevanten medizinischen Nutzens</i>	31
3.1.2 <i>Surrogate des patientenrelevanten medizinischen Nutzens</i>	33
3.1.3 <i>Ermittlung des Schadenpotenzials medizinischer Interventionen</i>	35
3.1.4 <i>Zusammenfassende Bewertung</i>	37
3.2 Spezielle Aspekte der Nutzenbewertung	38
3.2.1 <i>Ergebnis(un)sicherheit</i>	38
3.2.2 <i>Dramatischer Effekt</i>	39
3.2.3 <i>Studiendauer</i>	40
3.2.4 <i>Patientenberichtete Endpunkte</i>	41
3.2.5 <i>Nutzen in kleinen Populationen</i>	41
3.2.6 <i>Nutzen im Einzelfall</i>	42
3.3 Nutzenbewertung von Arzneimitteln	43
3.3.1 <i>Stellenwert des Zulassungsstatus</i>	43
3.3.2 <i>Studien zur Nutzenbewertung von Arzneimitteln</i>	45
3.4 Nichtmedikamentöse therapeutische Interventionen	46
3.5 Diagnostische Verfahren	48

3.6	Bevölkerungsweite Interventionen	50
3.6.1	<i>Screening</i>	50
3.6.2	<i>Prävention</i>	51
4	Leitlinien und Disease-Management-Programme	53
4.1	Hintergrund der Leitlinienbewertung	53
4.2	Ziele der Leitlinienbewertung	53
4.3	Methoden der Leitlinienbewertung	54
4.3.1	<i>Formale Prüfung</i>	56
4.3.2	<i>Leitlinienvergleich und inhaltliche Bewertung von Kernempfehlungen</i>	56
4.3.3	<i>Verbesserung des Outcomes</i>	58
4.4	Darlegung der Qualitätsprüfung	58
4.5	Abgabe von Empfehlungen zu Disease-Management-Programmen	58
5	Evidenzbasierte Gesundheitsinformation für Bürger und Patienten	60
5.1	Hintergrund und Ziele	60
5.2	Patientenzentrierte Kommunikation	61
5.2.1	<i>Kommunikationsstandards</i>	61
5.2.2	<i>Visuelle Kommunikation und Multimedia</i>	64
5.2.3	<i>Informationszugang und Barrierefreiheit</i>	65
5.3	Themenwahl, Recherche und wissenschaftliche Bewertung	67
5.3.1	<i>Themenwahl</i>	67
5.3.2	<i>Recherche</i>	69
5.3.3	<i>Evidenzbewertung</i>	70
5.3.4	<i>Aktualisierung</i>	72
5.4	Informationsprodukte	73
5.4.1	<i>Kernprodukte</i>	73
5.4.2	<i>Zusätzliche Elemente</i>	76
5.4.3	<i>Erfahrungsberichte</i>	77
5.4.4	<i>Website</i>	78
5.5	Überwachung und Auswertung	79
5.5.1	<i>Routinemonitoring</i>	79
5.5.2	<i>Feedback, Korrekturen und Verbesserungen</i>	81
5.5.3	<i>Evaluation</i>	82
6	Allgemeine methodische Aspekte	84
6.1	Informationsbeschaffung	84
6.1.1	<i>Allgemeine Grundsätze einer themenbezogenen Literaturrecherche</i>	84
6.1.2	<i>Ablauf einer Recherche nach Primärpublikationen</i>	85
6.1.3	<i>Sonstige Datenquellen für die Recherche</i>	86
6.1.4	<i>Auswahl relevanter Publikationen</i>	87
6.1.5	<i>Dokumentation</i>	88
6.1.6	<i>Literaturmonitoring</i>	88
6.1.7	<i>Berücksichtigung datenschutzrechtlicher Aspekte/ Vertraulichkeit</i>	89
6.2	Qualitätsbewertung von Einzelstudien	90
6.2.1	<i>Bezug zwischen Studientyp/-art und Fragestellung</i>	90
6.2.2	<i>Rangordnung verschiedener Studienarten/ Evidenzgrade</i>	91
6.2.3	<i>Aspekte der Bewertung des Verzerrungspotenzials von Ergebnissen</i>	91

6.2.4	<i>Interpretation von Subgruppenanalysen</i>	95
6.2.5	<i>Bewertung von Prognosestudien</i>	96
6.2.6	<i>Bewertung der Konsistenz publizierter Daten</i>	99
6.3	<i>Berücksichtigung von Sekundärliteratur</i>	100
6.3.1	<i>Einordnung systematischer Übersichten</i>	100
6.3.2	<i>Nutzenbewertung auf Basis von systematischen Übersichten und HTA-Berichten</i>	101
6.3.3	<i>Berücksichtigung publizierter Meta-Analysen</i>	104
6.4	<i>Spezielle biometrische Aspekte</i>	105
6.4.1	<i>Darstellung von Effekten und Risiken</i>	105
6.4.2	<i>Beurteilung statistischer Signifikanz</i>	106
6.4.3	<i>Beurteilung klinischer Relevanz</i>	107
6.4.4	<i>Nachweis der Verschiedenheit</i>	108
6.4.5	<i>Nachweis der Gleichheit</i>	109
6.4.6	<i>Adjustierung und multifaktorielle Verfahren</i>	110
6.4.7	<i>Meta-Analysen</i>	111
6.4.8	<i>Umgang mit nicht oder nicht vollständig publizierten Daten</i>	114
6.4.9	<i>Darstellung von Biasarten</i>	114
6.5	<i>Qualitative Methoden</i>	116
6.5.1	<i>Qualitative Studien</i>	116
6.5.2	<i>Konsultationstechniken</i>	118
	Literaturverzeichnis	120

ABGELÖST DURCH
"Allgemeine Methoden -
Version 4.0
vom 23.09.2011"

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

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

1 Das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

1.1 Gesetzliche Aufgaben

Der Gesetzgeber hat sich in der Gesundheitsreform 2004 („Gesetz zur Modernisierung der gesetzlichen Krankenversicherung GKV-Modernisierungsgesetz“ – GMG, in wesentlichen Teilen zum 01.01.2004 in Kraft getreten) für die Etablierung eines fachlich unabhängigen Instituts als Bestandteil der gesetzlichen Krankenversicherung in Deutschland entschieden [1]. Der Gemeinsame Bundesausschuss (G-BA) ist der Empfehlung des Gesetzgebers in § 139a Abs. 1 S. 2 SGB V gefolgt und hat im Jahr 2004 eine Stiftung des privaten Rechts errichtet, die „Stiftung für Qualität und Wirtschaftlichkeit im Gesundheitswesen“. Der ausschließliche Zweck dieser Stiftung ist die Unterhaltung des „Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen“ (IQWiG).

Das Institut wird zu Fragen von grundsätzlicher Bedeutung für die Qualität und Wirtschaftlichkeit von Leistungen der gesetzlichen Krankenversicherung tätig. Die **spezifischen Aufgaben** sind in § 139a Abs. 3 SGB V näher benannt:

- 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. Dabei sind alters-, geschlechts- und lebenslagen-spezifische Besonderheiten zu berücksichtigen.
- Bewertung evidenzbasierter Leitlinien für die epidemiologisch wichtigsten Krankheiten
- Abgabe von Empfehlungen zu Disease-Management-Programmen

Im Rahmen der Gesundheitsreform 2007 („Gesetz zur Stärkung des Wettbewerbs in der gesetzlichen Krankenversicherung GKV-Wettbewerbsstärkungsgesetz – GKV-WSG“) wurden die beiden nachstehenden Aufgaben in § 139a Abs. 3 SGB V wie folgt neu gefasst [2]:

- Bewertung des Nutzens und der Kosten von Arzneimitteln
- Bereitstellung von für alle Bürgerinnen und Bürger verständlichen allgemeinen Informationen zur Qualität und Effizienz in der Gesundheitsversorgung sowie zur Diagnostik und Therapie von Krankheiten mit erheblicher epidemiologischer Bedeutung

Die **Aufgabe der Bewertung des Nutzens und der Kosten von Arzneimitteln** wird in § 35b Abs. 1 SGB V in der Fassung des GKV-WSG näher bestimmt. Danach hat eine Bewertung durch den Vergleich mit anderen Arzneimitteln und Behandlungsformen unter Berücksichtigung des therapeutischen Zusatznutzens für die Patienten im Verhältnis zu den Kosten zu erfolgen. Folgende Kriterien zur Bestimmung des Patientennutzens werden explizit benannt: Verbesserung des Gesundheitszustands, Verkürzung der Krankheitsdauer, Verlängerung der Lebensdauer, Verringerung der Nebenwirkungen, Verbesserung der Lebensqualität. Die genaue Operationalisierung eines patientenrelevanten Nutzens unter Berücksichtigung dieser Kriterien findet sich in Abschnitt 3.1 des vorliegenden Papiers, im Besonderen in Abschnitt 3.1.1. Bei der wirtschaftlichen Bewertung sollen insbesondere die Angemessenheit und Zumutbarkeit einer Kostenübernahme durch die Versichertengemeinschaft angemessen berücksichtigt werden.

Im Übrigen legt der Gesetzgeber in § 35b Abs. 1 und § 139a Abs. 4 SGB V fest, dass die Bewertungen des medizinischen Nutzens nach den international anerkannten Standards der evidenzbasierten Medizin und die ökonomische Bewertung nach den hierfür maßgeblichen international anerkannten Standards, insbesondere der Gesundheitsökonomie, zu erfolgen haben. In Abschnitt 1.3 des vorliegenden Papiers werden der Begriff der evidenzbasierten Medizin, seine Entwicklung und das dahinterliegende Konzept näher erläutert.

Die Modalitäten der **Aufgabendurchführung** werden in § 139b SGB V näher geregelt. Danach kann als *Auftraggeber* der G-BA (Abs. 1) oder das Bundesministerium für Gesundheit (BMG) (Abs. 2) auftreten. Im Falle einer Beauftragung durch das BMG kann das Institut den Auftrag als unbegründet ablehnen, es sei denn, das BMG finanziert die Auftragsbearbeitung gesondert. Die Arbeitsergebnisse der Aufträge nach Abs. 1 und 2 leitet das Institut dem G-BA als Empfehlungen zu, der diese im Rahmen seiner Aufgabenstellung zu berücksichtigen hat.

Das Institut hat gemäß § 139b Abs. 3 S. 1 SGB V darauf zu achten, dass *externe Sachverständige* in die Auftragsbearbeitung einbezogen werden. Zur Sicherung der fachlichen Unabhängigkeit des Instituts haben diese Sachverständigen gemäß § 139b Abs. 3 S. 2 SGB V alle Beziehungen zu

Interessenverbänden, Auftragsinstituten, insbesondere der pharmazeutischen Industrie und der Medizinprodukteindustrie, einschließlich Art und Höhe der Zuwendungen offenzulegen. Dieselbe Verpflichtung gehen gemäß § 139a Abs. 6 SGB V auch alle *Beschäftigten des Instituts* ein.

Eine gesetzliche *Verpflichtung zur Transparenz* von Verfahrensschritten beinhaltet § 139a Abs. 5 in Verbindung mit § 35b Abs. 5 S. 6 SGB V, die im Zuge des GKV-WSG neu gefasst wurden. Nach § 139a Abs. 5 SGB V hat das Institut in allen wichtigen Abschnitten des Bewertungsverfahrens Sachverständigen der medizinischen, pharmazeutischen und gesundheitsökonomischen Wissenschaft und Praxis, den Arzneimittelherstellern, den für die Wahrnehmung der Interessen der Patientinnen und Patienten und die Selbsthilfe chronisch kranker und behinderter Menschen maßgeblichen Organisationen sowie der oder dem Beauftragten der Bundesregierung für die Belange der Patientinnen und Patienten Gelegenheit zur Stellungnahme zu geben. Diese Stellungnahmen sind in die Bewertungen einzubeziehen. Entsprechende Regelungen finden sich in Abschnitt 2.1.1 des vorliegenden Papiers im Zusammenhang mit der Erstellung von Berichtsplänen und Vorberichten.

Eine gesetzliche Verpflichtung zur Transparenz besteht ferner hinsichtlich der Arbeitsmethodik des Instituts. Gemäß § 139a Abs. 4 S. 2 SGB V hat das Institut in regelmäßigen Abständen über die Arbeitsprozesse und -ergebnisse einschließlich der Grundlagen öffentlich zu berichten. Auch dies ist in Abschnitt 2.1.1 sowie in den Abschnitten 2.1.2 bis 2.1.4 des vorliegenden Papiers durch die obligate Veröffentlichung der Produkte über die Internetseiten des Instituts geregelt.

Die **Finanzierung** des Instituts findet ihre gesetzliche Grundlage in § 139c SGB V. Sie erfolgt durch die Erhebung eines besonderen Zuschlags für jeden abzurechnenden Krankenhausfall und durch die Anhebung der Vergütungen der ambulanten vertragsärztlichen und vertragszahnärztlichen Versorgung. Diese Zuschläge werden „Systemzuschlag“ genannt und stammen jeweils zur Hälfte aus dem stationären und dem ambulanten Bereich. Der G-BA legt jährlich die Höhe dieses Systemzuschlags fest. Die Gelder aus der Erhebung dieses Systemzuschlags fließen neben dem Institut auch dem G-BA zu.

1.2 Struktur

Das Institut ist eine Einrichtung der Stiftung für Qualität und Wirtschaftlichkeit im Gesundheitswesen unter wissenschaftlich unabhängiger Leitung. Organe der Stiftung sind der *Stiftungsrat*, der sich je zur Hälfte aus Mitgliedern der Spitzenverbände der gesetzlichen Krankenkassen und aus Mitgliedern der Organisationen der Leistungserbringer (Kassenärztliche Bundesvereinigung, Kassenzahnärztliche Bundesvereinigung, Deutsche Krankenhausgesellschaft) zusammensetzt, und ein vom Stiftungsrat bestellter fünfköpfiger *Stiftungsvorstand*. Ein Mitglied

des Stiftungsvorstands wird vom Bundesministerium für Gesundheit ernannt. Das dritte Organ der Stiftung ist der G-BA, der gemäß § 15 Abs. 1 der Stiftungssatzung mit einer Mehrheit von zwei Dritteln seiner Mitglieder Satzungsänderungen beschließen kann. Aufsichtsbehörde der Stiftung ist die Senatsverwaltung für Justiz des Landes Berlin.

Als beratendes Gremium des Instituts fungiert das *Kuratorium*, das aus 30 Mitgliedern besteht, die sich aus den Trägerorganisationen des G-BA, aus maßgeblichen, nicht im G-BA vertretenen Organisationen von Leistungserbringern und Sozialpartnern sowie sonstigen für das Gesundheitswesen relevanten, nicht im G-BA vertretenen Organisationen rekrutieren. Der *Wissenschaftliche Beirat* bildet das zweite beratende Gremium des Instituts. Seine bis zu 12 Mitglieder werden im Einvernehmen mit der Institutsleitung vom Vorstand der Stiftung bestellt.

Das Institut gliedert sich derzeit in die nachstehenden Ressorts:

- Arzneimittelbewertung
- Medizinische Biometrie
- Nichtmedikamentöse Verfahren
- Gesundheitsökonomie
- Versorgungsqualität
- Kommunikation
- Gesundheitsinformation
- Verwaltung

Die Ressortleiter bilden gemeinsam mit dem Institutsleiter das Steuergremium. Das Steuergremium berät die Institutsleitung. Es befasst sich sowohl mit institutsinternen Aufgaben als auch mit Aufgaben mit externem Bezug. Das Gremium ist für die Einhaltung der Verfahrensweisen und Methoden verantwortlich, die es – gemeinsam mit Institutsmitarbeitern – erstellt und modifiziert. Zudem erarbeitet es die Bedingungen für die externe Auftragsvergabe (entsprechend den Grundsätzen zur Auftragsvergabe des Stiftungsvorstandes).

Der vom Stiftungsvorstand bestellte Institutsleiter vertritt das Institut nach innen und nach außen. Er ist dafür verantwortlich, dass das Institut seine gesetzlichen Aufgaben erfüllt und dabei den vom G-BA gesetzten Prioritäten folgt. Auch für die Einhaltung des Haushaltsplans und der Geschäftsordnung hat er Sorge zu tragen.

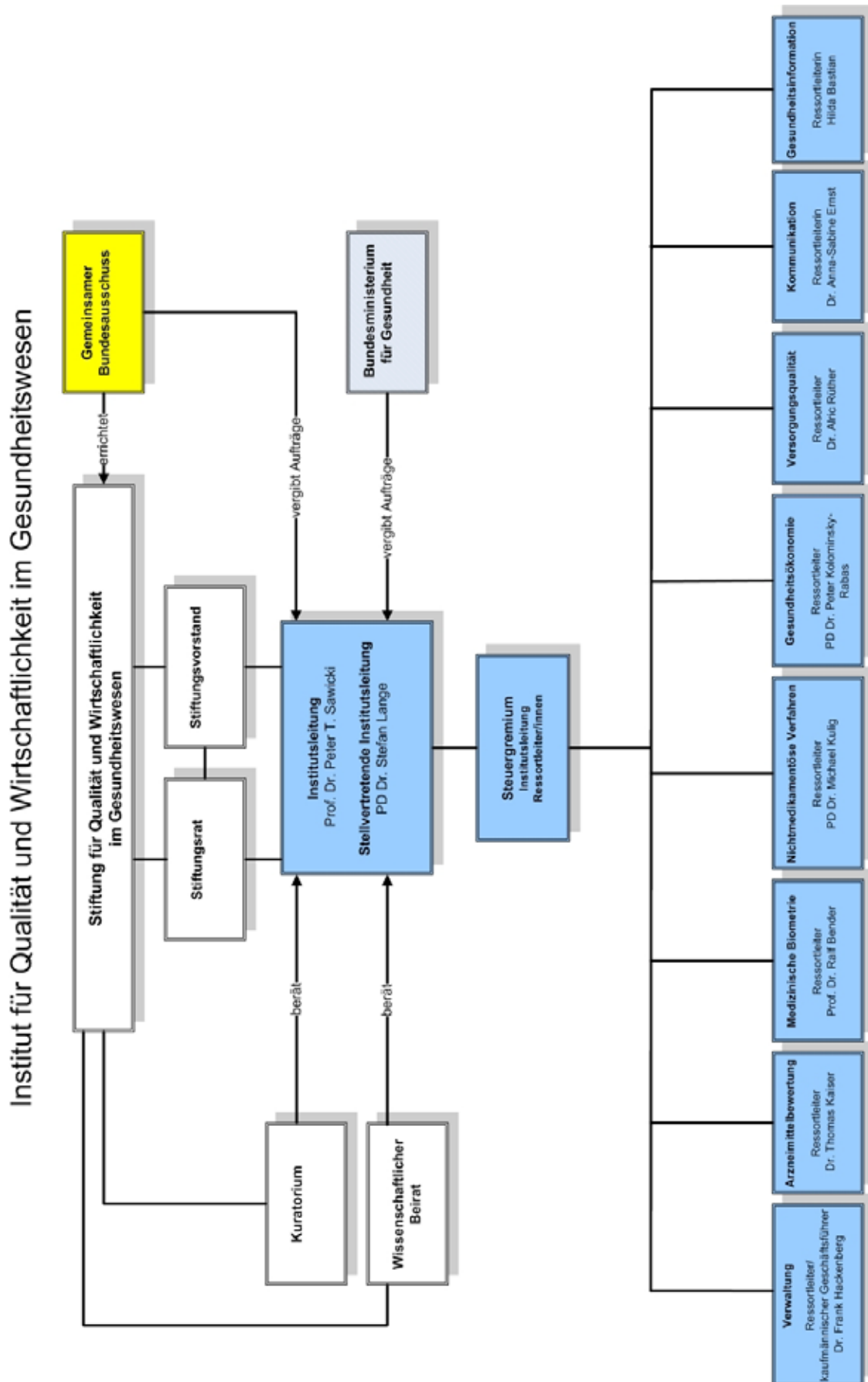


Abbildung 1: Organigramm des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen

1.3 Evidenzbasierte Medizin

Der Begriff „evidenzbasierte Medizin“ (EbM) ist eine etwas unscharfe Eindeutschung des englischen „evidence-based medicine“. Gemeint ist damit eine medizinische Betreuung von Patienten, die sich nicht alleine auf Meinungen und Übereinkünfte stützt, sondern „Evidenz“ einbezieht – Belege, die mit möglichst objektiven wissenschaftlichen Methoden erhoben wurden. Evidenzbasierte Medizin umfasst Werkzeuge und Strategien, die vor Fehlentscheidungen und falschen Erwartungen schützen sollen. Fehlentscheidung kann in diesem Zusammenhang bedeuten, dass nützliche Interventionen nicht oder erst verspätet in die Versorgung kommen oder dass nutzlose oder gar schädliche Interventionen weite Verbreitung finden [14,123,165,168].

Instrumente, die einer subjektiv geprägten und damit häufig verzerrten Bewertung vorbeugen sollen (siehe u. a. Kapitel 6), wurden allerdings nicht erst mit der Einführung des Begriffes „evidenzbasierte Medizin“ erfunden, sondern ihre Wurzeln reichen weit in die Vergangenheit zurück. In Deutschland war es bereits 1932 Paul Martini, der in seiner Monografie „Methodenlehre der therapeutischen Untersuchung“ die wesentlichen Elemente einer fairen Überprüfung der Wirksamkeit von Arzneimitteln beschrieb [258]. Die Methode der zufälligen Zuteilung (Randomisierung) von Probanden auf Vergleichsgruppen wurde mit Beginn der 1960er-Jahre international akzeptierter Standard, um die Wirksamkeit und Sicherheit von medizinischen Interventionen zu prüfen [188]. Beginnend in den USA, wurden zu dieser Zeit entsprechende Studien zur Voraussetzung für die behördliche und in Gesetzen und Verordnungen geregelte Zulassung von Arzneimitteln und (teilweise) Medizinprodukten [31]. Etwa 20 Jahre später bemühten sich klinische Epidemiologen, die Methodenlehre auch in der klinischen Praxis zu etablieren [127]. Begleitet von teilweise heftig geführten Kontroversen gelang dies tatsächlich erst in den 1990er-Jahren, zeitgleich mit der Benennung des Konzepts als „evidenzbasierte Medizin“. Studien an sich und das systematische Auffinden und Bewerten von Studien bilden seit dieser Zeit auch den internationalen wissenschaftlichen Standard für die Technologiebewertung im Gesundheitswesen (Health Technology Assessment) [30].

Das Institut ist gesetzlich verpflichtet, die „in den jeweiligen Fachkreisen anerkannten internationalen Standards der evidenzbasierten Medizin“ anzuwenden (§ 139a Abs. 4 SGB V). Es ist die Aufgabe des vorliegenden Methodenpapiers des Instituts, die Methoden und Strategien zu beschreiben, die diese internationalen Standards definieren. EbM ist kein starres Konzept. Welches Standardinstrument wann eingesetzt werden sollte, hängt von der Frage ab, die es zu beantworten, und von der Entscheidung, die es zu treffen gilt. Trotz der Benutzung von Standards sind bei der Suche, Aufarbeitung und Bewertung von Studien immer wieder Entscheidungen zu treffen, für die es (noch) keine internationalen Festlegungen gibt. EbM beinhaltet auch die Freiheit, in solchen

Situationen eigene Festlegungen zu treffen. Allerdings ist diese Freiheit mit der Verpflichtung verbunden, solche eigenen Festlegungen nach Möglichkeit vorab zu definieren und Bewertungen transparent zu erläutern, sodass die Begründung nachvollziehbar ist. Dieses Kapitel erläutert, dass eine Institution wie das IQWiG bei der Anwendung der EbM und den Definitionen von Festlegungen in einer anderen Situation ist als Ärzte, die Unterstützung bei einer Therapieentscheidung suchen.

1.3.1 Praktische evidenzbasierte Medizin

Evidenzbasierte Medizin ist von der Idee her als Strategie für Ärzte gedacht, die für ihre Patienten unter möglichen alternativen Therapien die vielversprechendsten und deren Bedürfnissen am ehesten entsprechenden Alternativen herausfinden und die Erfolgsaussichten ohne Übertreibung darstellen wollen. Diese Anwendung der evidenzbasierten Medizin in der täglichen Praxis für „individuelle Patienten“ haben im Jahr 1996 David Sackett und Kollegen [93] folgendermaßen definiert: „EbM ist der gewissenhafte, ausdrückliche und vernünftige Gebrauch der gegenwärtig besten externen, wissenschaftlichen Evidenz für Entscheidungen in der medizinischen Versorgung individueller Patienten. Die Praxis der EbM bedeutet die Integration individueller klinischer Expertise mit der bestverfügbaren externen Evidenz aus systematischer Forschung.“

Oft erweist sich die „bestverfügbare Evidenz“ aber als lückenhaft oder nicht zuverlässig. EbM hilft Ärzten und Patienten auch dabei, eine solche Unsicherheit zu erkennen. Arzt und Patient können dann besprechen, wie man mit der Unsicherheit umgehen soll. Gerade in unsicheren Situationen kommt es auf die persönlichen Präferenzen der Patienten an, die darüber entscheiden, welche Option sie wählen. Im Idealfall basieren Entscheidungen außer auf der Evidenz auch auf dem klinischen Zustand und den Umständen eines einzelnen Patienten und auf seinen Präferenzen und Handlungen [180]. Gleichzeitig schafft die Beschreibung der identifizierten Wissenslücken erst die Voraussetzung für zielgerichtete medizinische Forschung, die sich nach den Bedürfnissen von Patienten richtet.

Hinter evidenzbasierter Medizin steht eine hinterfragende Grundhaltung [231]. Wie notwendig Skepsis häufig ist, zeigt sich daran, dass in den letzten Jahrzehnten eine Reihe von unzureichend erprobten, aber bereits breit eingesetzten Therapien mit Methoden der EbM überprüft wurde und diese Überprüfung gezeigt hat, dass vorschneller Optimismus gefährliche Folgen für Patienten haben kann [104,304]. Aufgabe des Institutes ist es, objektiv zu bewerten, mit welcher Sicherheit der Nutzen medizinischer Maßnahmen nachgewiesen ist, um nicht sachgerechten Urteilen entgegenzuwirken.

1.3.2 Die Strategien der EbM

Ein charakteristisches Standardelement der evidenzbasierten Medizin ist die strukturierte und systematische Art, mit der Antworten auf eine medizinische Frage gesucht werden.

1. Die medizinische Frage muss präzise formuliert werden. In der Medizin geht es (fast) immer um die Entscheidung zwischen mindestens zwei Alternativen: Diese können Therapien, Diagnoseverfahren oder komplexe Lebensstiländerungen betreffen. Daraus leitet sich stets die Frage ab: Ist Option A besser als Option B? Dabei kann zum Beispiel auch der Verzicht auf eine Therapie eine Option sein, die ernsthaft überprüft werden muss. Zu betonen ist allerdings, dass Verzicht auf Therapie nicht gleichbedeutend mit „nichts tun“ ist (zum Beispiel kontrolliertes Zuwarten, „watchful waiting“).
2. Es muss definiert sein, woran der Nutzen einer Therapie (Diagnose, Lebensstiländerung) gemessen werden soll. Standardelement der EbM ist die Frage nach relevanten Konsequenzen für Patienten: Kann das Leben verlängert werden, bessern sich Beschwerden und die Lebensqualität?
3. EbM formuliert explizit, dass in der Medizin bezüglich des Nutzens von Therapie, Diagnose und Lebensstil in der Regel nur Wahrscheinlichkeitsaussagen oder Aussagen über Gruppen von Patienten möglich sind. Nutzen wird dadurch nachgewiesen, dass sich die Wahrscheinlichkeit eines günstigen Ausgangs erhöht und/oder das Risiko eines ungünstigen Ausgangs verringert. Um den Nutzen nachweis zu führen, sind Studien an ausreichend großen Gruppen von geeigneten Patienten nötig. Zur Planung, Durchführung und Auswertung solcher Studien haben internationale Wissenschaftler eine Vielzahl von Regeln und Instrumenten entwickelt. Wichtigstes Ziel ist es, Einflüsse, die die Ergebnisse eines Vergleichs verzerren können, zu minimieren oder – falls das unmöglich ist – zumindest zu erfassen. Die Effekte solcher verzerrenden Einflüsse werden auch im Deutschen häufig mit dem englischen Begriff „Bias“ bezeichnet. Diese Regeln und Werkzeuge, die international als Standard akzeptiert sind und laufend weiterentwickelt werden, sind die methodische Basis der EbM und der Arbeit des Instituts.
4. Eine weitere zentrale Strategie der EbM besteht darin, alle qualitativ angemessenen Studien zu einer Frage zu identifizieren und so den Stand des Wissens zusammenzufassen. Zeigen sich dabei große Unterschiede zwischen den Ergebnissen einzelner Studien (sogenannte Heterogenität), muss versucht werden, diese Unterschiede zu erklären. Die Ergebnisse dieser Zusammenfassungen und Bewertungen werden als sogenannte systematische Übersicht bezeichnet, die statistische Auswertung als Meta-Analyse.

1.3.3 Bedeutung von EbM für das Institut

Die Hauptaufgabe des Instituts ist es, die vorgegebene Frage seiner Auftraggeber, ob ein Nutzen oder Schaden durch eine Maßnahme nachzuweisen ist, möglichst zuverlässig zu beantworten. Ziel ist es, ausreichend zuverlässige Belege dafür vorzulegen, dass Therapie A für Patienten mit einer bestimmten Krankheit besser ist als Alternative B. Oder kürzer: Was ist der Nutzen von A im Vergleich zu B?

Der Auftrag des Instituts zielt also ganz bewusst nicht auf die Behandlung einzelner Patienten mit ihren möglichen Besonderheiten, sondern darauf, für welche Gruppen von Patienten es Belege für einen Nutzen gibt.

1.3.4 Evidenzbasierte Entscheidungsfindung im Gesundheitswesen

Die meisten der Gutachten, die das Institut erstellt, sollen dem G-BA als eine Grundlage für spätere Entscheidungen dienen, die für alle gesetzlich Krankenversicherten bindend sind. Die übrigen Gutachten dienen dem G-BA zum Beispiel als Information. Die Art der Entscheidungen, die Institutionen wie der G-BA zu treffen haben, hat Auswirkungen auf den Einsatz der Methoden der EbM.

1.3.5 Die Bedeutung der Ergebnissicherheit

Eine Besonderheit der EbM ist, dass sie eine Einschätzung erlaubt, inwieweit das vorhandene Wissen zuverlässig ist. Entscheidungen des G-BA benötigen eine hohe wissenschaftliche Zuverlässigkeit, da sie für alle gesetzlich Krankenversicherten verbindliche Konsequenzen wie zum Beispiel den Ausschluss von Leistungen haben können.

Das wirkt sich auch auf die Arbeit des Instituts aus: Die Bewertung der Ergebnissicherheit spielt in den Berichten des Instituts eine zentrale Rolle. Zahlreiche Details, wie Studien geplant, ausgeführt, ausgewertet und veröffentlicht wurden, haben einen Einfluss darauf, wie verlässlich die vorhandenen Ergebnisse sind. Es ist internationaler Standard der EbM, diese Aspekte kritisch zu prüfen und zu bewerten. Wie die zur Beantwortung einer Frage nötige Ergebnissicherheit erreicht werden kann, hängt aber auch von der Krankheit ab und von der Größe des Effekts einer Intervention: Wenn zwei Läufer in einem fairen Rennen mit großem Abstand über die Ziellinie gehen, braucht man keine Uhr, um den Sieger zu erkennen. So kann der Nutzen einer neuen Therapie, die zur Heilung einer bislang immer tödlichen Krankheit führt, schon durch eine relativ kleine Zahl von überlebenden Patienten bewiesen sein. Auch hier beruht das Urteil letztlich auf einem Vergleich, doch bei solch dramatischen Effekten kann schon der Vergleich von früheren mit

heutigen Patienten genügend Sicherheit bieten. Allerdings sind Therapien, die einen so dramatischen Nutzen aufweisen, in der heutigen Medizin selten.

Gerade bei chronischen Krankheiten ist es die Regel, dass Unterschiede zwischen zwei Therapiealternativen leicht durch einen variablen und schwankenden Verlauf der Krankheit überlagert werden können. Hier sind sehr präzise Methoden nötig, um unter den Schwankungen einen Effekt der Therapie erkennen zu können.

Es kann davon ausgegangen werden, dass das Institut gerade mit Vergleichen beauftragt wird, bei denen eben nicht auf den ersten Blick zu erkennen ist, welche Alternative Vorteile bietet. Je kleiner aber die zu erwartenden Unterschiede zwischen zwei Alternativen sind, umso präzisere und verlässlichere Studien sind nötig, um hinreichend sicher sein zu können, dass ein beobachteter Unterschied nicht auf Zufall oder Messfehlern beruht. Der 100-Meter-Weltrekord lässt sich heute nicht mehr mit einer Sanduhr messen.

Bei kleinen Unterschieden muss zudem auch die klinische Relevanz des Unterschieds beurteilt werden. Diese Anforderungen an Präzision und Verlässlichkeit bestimmen die Arbeitsweise des Instituts:

1. Es gehört zu den internationalen Standards der evidenzbasierten Medizin, für jede Frage den Typ von Studien als Messinstrument zu benennen, der das Risiko minimiert, dass eine der Alternativen unberechtigterweise benachteiligt wird.
2. Falls sich herausstellt, dass Studien mit der nötigen Qualität und Präzision generell fehlen, ist es die Kernaufgabe des Instituts, den Sachverhalt zu beschreiben, mit der Schlussfolgerung, dass auf Basis der „gegenwärtig besten“ Evidenz keine sicheren Empfehlungen möglich sind.
3. Somit stützt sich das Institut bei der Bewertung des Nutzens bzw. Schadens in der Regel nur auf Studien mit ausreichender Ergebnissicherheit. Damit ist gewährleistet, dass die auf Empfehlungen des Instituts beruhenden Entscheidungen des G-BA auf einer wissenschaftlich gesicherten Basis stehen. Zudem wäre eine Untersuchung, die eine Recherche nach Studien unzureichender Ergebnissicherheit einschließt, zeit- und kostenintensiv.
4. Der G-BA hat die Aufgabe, neben der wissenschaftlichen Sachlage auch andere Aspekte in seine Entscheidungen einzubeziehen, wozu neben der Wirtschaftlichkeit auch die Bedürfnisse und Werte der Bevölkerung gehören [155]. In einer wissenschaftlich unsicheren Situation gewinnen diese Aspekte an Gewicht.

1.3.6 Die Verbindung von Ergebnissicherheit und Alltagsnähe

Die Betonung der Abschätzung der Ergebnissicherheit wird oft kritisiert. Ein Argument ist, dass Studien mit hoher Ergebnissicherheit, insbesondere randomisierte kontrollierte Studien, zwar interne Validität besäßen, jedoch oft die Versorgung von Patienten im Alltag nicht abbildeten und deshalb nicht „übertragbar“ seien, also keine externe Validität vorhanden sei. Diese Kritik ist dann oft mit der Forderung verbunden, andere Studientypen hinzuzunehmen, um Alltagsbedingungen besser einfließen zu lassen.

Diese Kritik vermischt jedoch Argumentationsebenen, die klar voneinander zu trennen sind. Folgende Aspekte müssen Beachtung finden.

1. Grundvoraussetzung für eine Nutzenbewertung ist der Nachweis von Kausalität. Und eine unverzichtbare Bedingung für den Nachweis von Kausalität ist ein vergleichendes Experiment, in dem ein Effekt auf eine einzige Einflussgröße zurückgeführt werden kann. Das macht für klinische Studien erhebliche Anstrengungen nötig, weil es zahlreiche unerwünschte Einflüsse gibt, die einen Effekt vortäuschen oder verbergen. Einer der stärksten Einflüsse ist eine unfaire Zuteilung der Studienteilnehmer auf die Vergleichsgruppen. Es gibt keinen Zweifel daran, dass die Randomisierung das beste gegenwärtig verfügbare Instrument ist, um diese Verzerrung zu minimieren. Die zufällige Zuteilung der Teilnehmer zu den Gruppen stellt sicher, dass es keine systematischen Unterschiede zwischen den Gruppen gibt, und zwar weder bezüglich bekannter Eigenschaften (z. B. Alter, Geschlecht, Schwere der Erkrankung) noch bezüglich unbekannter Faktoren. Damit ist eine grundlegende Voraussetzung für einen Kausalitätsnachweis gegeben. Allerdings ist auch Randomisierung alleine kein Garant für eine hohe Ergebnissicherheit. Dazu sind z. B. auch eine unverzerrte Erhebung, Zusammenfassung und Veröffentlichung der Ergebnisse nötig.
2. Eine der Kernerkenntnisse der EbM ist, dass andere Studientypen als randomisierte kontrollierte Studien (Randomized Controlled Trials, RCTs) in der Regel für einen Kausalitätsbeweis nicht geeignet sind. Es ist zwar möglich, dass Patienten in anderen Studientypen unter Alltagsbedingungen beobachtet werden. Diese Studien können aber zumeist die maßgebliche Frage nicht hinreichend sicher beantworten, ob ein Unterschied auf der Intervention beruht. Nicht randomisierte Studien liefern immer ein potenziell verzerrtes Ergebnis, auch wenn die Auswahl der Teilnehmer wenig selektiert war. In nicht randomisierten Studien kann grundsätzlich nicht von einer Strukturgleichheit der Gruppen ausgegangen werden. Daher bedarf die Verwendung nicht randomisierter Studien als Beleg für die Kausalität einer Intervention einer besonderen Begründung.

3. Es trifft zu, dass randomisierte Studien oft z. B. Patienten mit Begleiterkrankungen ausschließen, die im Alltag häufig sind. Doch das ist keine Konsequenz der Randomisierung, sondern bedingt durch andere Faktoren wie z. B. Ein- und Ausschlusskriterien der Studie. Zudem werden Patienten in randomisierten Studien oft intensiver und enger betreut als im Alltag. Das alles sind aber bewusste Entscheidungen derjenigen, die in einer Studie eine bestimmte Frage beantworten wollen. Auch in nicht randomisierten Studien findet eine Selektion der Teilnehmer durch Designcharakteristika statt, sodass externe Validität ebenso wenig automatisch vorausgesetzt werden kann wie bei RCTs.
4. Aber selbst wenn sich Patientengruppen eines RCTs von Patienten des Versorgungsalltags unterscheiden, muss das die externe Validität des Ergebnisses nicht infrage stellen. Entscheidend ist vielmehr die Frage, ob zu erwarten ist, dass der in einer Population festgestellte Therapieeffekt in einer anderen Population anders ist.
5. Es hängt vom Einzelfall ab, wie sich die Intensität der Betreuung in einer Studie auswirken könnte. Zum Beispiel ist denkbar, dass ein Nutzen tatsächlich nur in den Händen besonders qualifizierter Ärzte vorhanden ist, weil unter Alltagsbedingungen ansonsten zu häufig Komplikationen auftreten. Es ist aber auch möglich, dass eine enge Betreuung Unterschiede zwischen Gruppen eher verringert. Zum Beispiel könnten Unterschiede in der Therapieadhärenz in Studien kleiner ausfallen, in denen Patienten grundsätzlich sehr eng betreut werden.
6. Die Festlegung der Studienbedingungen liegt jedoch in der Hand des Initiators einer klinischen Studie. Er kann die Fragestellung und Endpunkte definieren, die er als so wichtig einschätzt, dass sie in der Studie beobachtet werden sollen. Wenn ein Hersteller eines Arzneimittels z. B. Therapieadhärenz für einen wichtigen Aspekt des Nutzens seines Produktes hält, wäre die naheliegende Konsequenz, dann Studien aufzulegen, die diesen Aspekt mit möglichst großer Ergebnissicherheit und Nähe zu Alltagsbedingungen messen können und gleichzeitig die Patientenrelevanz demonstrieren.

Diese Ausführungen zeigen, dass Ergebnissicherheit und Alltagsnähe (oder interne und externe Validität) nicht fest miteinander verknüpft sind: Hohe Ergebnissicherheit und Alltagsnähe schließen sich nicht aus, sondern erfordern nur eine intelligente Kombination aus Studientyp, Design und Durchführung.

Auch wenn die Kritik mangelnder Alltagsnähe tatsächlich für viele Studien zutreffen mag, wäre nichts dadurch gewonnen, wenn man eine hohe Ergebnissicherheit für eine bessere Alltagsnähe aufgeben würde. Man würde einen Mangel durch einen zweiten, fundamentaleren Mangel zu kompensieren versuchen.

Es sind Studien wünschenswert und machbar, die Alltagsnähe und hohe Ergebnissicherheit vereinigen. RCTs, in denen weder über den Behandlungsalltag hinausgehende Anforderungen an Patienten gestellt noch definierte Studienvisiten vorgegeben werden, sind sehr wohl möglich. Sie werden international als „Real World Trials“, „Practical Trials“ oder „Pragmatic Trials“ diskutiert [138,140,151,257,361].

ABGELÖST DURCH
"Allgemeine Methoden -
Version 4.0
vom 23.09.2011"

2 Produkte des Instituts

Gemäß seinem gesetzlichen Auftrag erstellt das Institut verschiedene Produkte in Form von wissenschaftlichen Berichten und allgemein verständlichen Informationen für Bürger und Patienten. Das nachfolgende Kapitel beschreibt Abläufe und allgemeine Methoden der Erstellung von Institutsprodukten. Dabei werden zunächst die einzelnen Produkte benannt und produktspezifische Verfahrensabläufe dargestellt (Abschnitt 2.1). Im darauf folgenden Abschnitt (Abschnitt 2.2) werden weitere produktunabhängige Aspekte beschrieben.

Die allgemeine Methodik der Produkterstellung ergibt sich aus der Aufgabe des Instituts, systematisch zu einer gegebenen Fragestellung diejenigen Informationen zusammenzutragen und zu bewerten, die dazu geeignet sind, interpretierbare Aussagen zum Nutzen oder Schaden einer medizinischen Intervention zu treffen. Es ist nicht Aufgabe des Instituts, den möglichen Nutzen oder Schaden beziehungsweise das Fehlen eines Nutzens oder Schadens einer medizinischen Intervention *nachzuweisen*.

Die Bewertung einer Methode beziehungsweise Intervention ist theoretisch zu jeder Zeit möglich. Der Zeitpunkt der Bewertung wird, sofern hierfür ein Auftrag erforderlich ist, insbesondere durch den Zeitpunkt dieses Auftrags an das Institut bestimmt. Generelle Vorgaben dahin gehend, dass die Erstellung einer systematischen Bewertung durch das Institut frühestens nach Verstreichen einer gewissen Zeit nach Zulassung oder Etablierung einer Methode oder Intervention erfolgt, werden nicht getroffen. Sofern im Falle einer frühzeitigen Bewertung aufgrund fehlender Langzeitstudien eine hohe Ergebnisunsicherheit festgestellt wird, wird diese den allgemeinen Arbeitsmethoden folgend dargestellt.

2.1 Produktspezifische Verfahrensabläufe

In der Produktpalette des Instituts lassen sich vier übergeordnete Produkte unterscheiden:

- ausführliche **Berichte**, insbesondere zur Nutzenbewertung oder zur Bewertung des Kosten-Nutzen-Verhältnisses
- Schnellberichte („**Rapid Reports**“)
- **Gesundheitsinformationen** (allgemein verständliche Informationen für Bürger und Patienten)
- **Arbeitspapiere** zu versorgungsrelevanten Entwicklungen im Gesundheitswesen sowie zur methodischen Arbeit des Instituts

Die Erstellung von Berichten und Rapid Reports erfolgt ausschließlich auf Basis eines Einzelauftrags durch den G-BA oder das BMG. Vorrangiges Ziel von Berichten ist es, Empfehlungen für Richtlinienentscheidungen des G-BA auszusprechen. Ziel von Rapid Reports ist die zeitnahe Information zu relevanten aktuellen Themen sowie zu Fragestellungen, die nicht auf Richtlinienentscheidungen des G-BA ausgerichtet sind.

Gesundheitsinformationen können auf Basis eines Einzelauftrags erstellt werden, sie können aber auch Folge eines Auftrags zur Nutzenbewertung sein (allgemein verständliche Version eines Berichts) oder im Rahmen der allgemeinen gesetzlichen Aufgabe zur Bereitstellung von Gesundheitsinformationen erarbeitet werden.

Arbeitspapiere entstehen in Eigenverantwortung des Instituts, ohne dass es hierzu eines gesonderten Auftrags durch den G-BA oder das BMG bedarf. Dies geschieht entweder auf Grundlage des Generalauftrags (siehe Abschnitt 2.1.4) mit dem Ziel, zu versorgungsrelevanten Entwicklungen im Gesundheitswesen Auskunft zu erteilen, oder im Rahmen des gesetzlichen Auftrags zur Entwicklung der Institutsmethoden. Die „Allgemeinen Methoden“ des Instituts sind nicht als Arbeitspapier in diesem Sinne zu verstehen und unterliegen einem gesonderten Verfahren der Erstellung und Aktualisierung, das in den „Allgemeinen Anmerkungen“ auf Seite 2 dieses Dokuments dargestellt ist.

In nachfolgender Tabelle 1 sind die verschiedenen Produkte des Instituts in einer Übersicht dargestellt. Die produktspezifischen Verfahrensabläufe sind in den sich daran anschließenden Abschnitten 2.1.1 bis 2.1.4 beschrieben.

Tabelle 1: Übersicht über die Produkte des Instituts

Produkt	Zielsetzung	Ablauf	Beauftragung durch
Bericht	Empfehlung für Richtlinienentscheidungen des G-BA	beschrieben in Abschnitt 2.1.1	G-BA, BMG
Rapid Report	Zeitnahe Information zu aktuellen Themen sowie Fragestellungen, die nicht auf Richtlinienentscheidungen des G-BA ausgerichtet sind	beschrieben in Abschnitt 2.1.2	G-BA, BMG
Gesundheitsinformation	Allgemein verständliche Information für Bürger und Patienten, breit gefächertes Themenspektrum	beschrieben in Abschnitt 2.1.3	G-BA, BMG; auch Initiierung durch das Institut
Arbeitspapier	Information zu versorgungsrelevanten Entwicklungen in der Medizin oder zu methodischen Aspekten	beschrieben in Abschnitt 2.1.4	Initiierung durch das Institut

2.1.1 Berichte

Der Ablauf der Berichtserstellung ist in Abbildung 2 schematisch dargestellt. Alle Arbeitsschritte werden in Verantwortung des Instituts ggf. unter Einbeziehung externer Expertise getätigt. Je nach Notwendigkeit wird auch der wissenschaftliche Beirat des Instituts einbezogen. Der interne Qualitätssicherungsprozess ist in diesem Flussdiagramm nicht dargestellt.

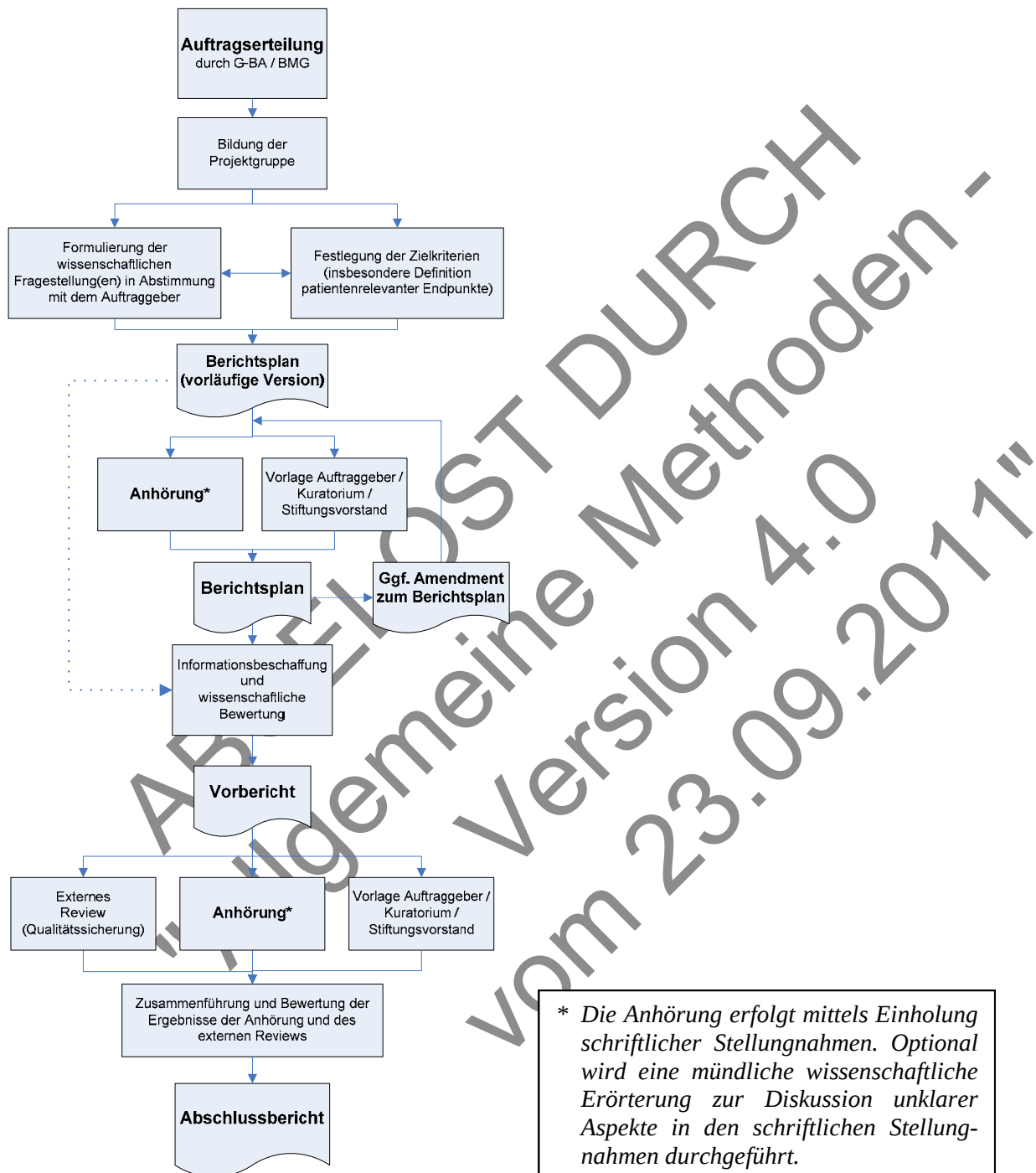


Abbildung 2: Ablauf der Berichtserstellung

Nach der **Auftragsvergabe** durch den G-BA oder das BMG erfolgt die Bildung der internen Projektgruppe des Instituts unter der Leitung des zuständigen Ressorts. Eine Person wird als verantwortlicher Projektleiter bestimmt. Die Zusammensetzung der Projektgruppe ist an dieser Stelle nicht endgültig, da sich durch die folgenden Arbeitsschritte Änderungsbedarf ergeben kann. Die Formulierung der wissenschaftlichen Fragestellung erfolgt in Abstimmung mit den zuständigen Gremien des Auftraggebers, gegebenenfalls unter Einbeziehung externer fachlicher Expertise oder der Meinung einzelner Betroffener. Regelhaft werden dabei insbesondere zur Definition patientenrelevanter Endpunkte entsprechende Patientenorganisationen beteiligt. Anschließend erfolgt die Erstellung des Berichtsplans.

Der **Berichtsplan** beinhaltet, vergleichbar mit dem Studienprotokoll einer geplanten klinischen Studie, die genaue wissenschaftliche Fragestellung einschließlich der Zielkriterien (z. B. patientenrelevante Endpunkte), die Ein- und Ausschlusskriterien der für die Nutzenbewertung zu verwendenden Informationen sowie die Darlegung der projektspezifischen Methodik der Beschaffung und Bewertung dieser Informationen. Der Berichtsplan wird in Verantwortung der Projektgruppe des Instituts, i. d. R. unter Beteiligung externer Sachverständiger, erstellt. Nach Abschluss der internen Qualitätssicherung und der Freigabe durch das Steuergremium des Instituts wird die vorläufige Version des Berichtsplans dann zunächst dem Auftraggeber, auch zur Frage der Vollständigkeit in Bezug auf den ursprünglich erteilten Auftrag, sowie dem Kuratorium und dem Vorstand der Stiftung zugeleitet. Die Veröffentlichung auf den Internetseiten des Instituts zwecks Gelegenheit zur Stellungnahme erfolgt in der Regel fünf Arbeitstage später.

Für eine Frist von mindestens vier Wochen wird der Öffentlichkeit und damit auch den Sachverständigen der medizinischen, pharmazeutischen und gesundheitsökonomischen Wissenschaft und Praxis, den Berufsvertretungen der Apotheker, den Arzneimittelherstellern, den für die Wahrnehmung der Interessen der Patienten und die Selbsthilfe chronisch kranker und behinderter Menschen maßgeblichen Organisationen sowie dem oder der Beauftragten der Bundesregierung für die Belange der Patienten **Gelegenheit zur Stellungnahme (Anhörung)** gegeben (schriftliche Stellungnahmen). Dadurch wird ein offenes und unabhängiges Reviewverfahren zum Berichtsplan ermöglicht. Die Gelegenheit zur Stellungnahme bezieht sich dabei insbesondere auf das projektspezifische methodische Vorgehen zur Beantwortung der Fragestellung einschließlich der Festlegung patientenrelevanter Zielgrößen. Ebenso besteht zu diesem Zeitpunkt die Möglichkeit, qualitativ angemessene Unterlagen jeglicher Art, insbesondere unpublizierte Daten, die aus Sicht des jeweiligen Stellungnehmenden zur Beantwortung der Fragestellung des Berichts geeignet sind, vorzulegen. Falls die in dem Berichtsplan definierte Suchstrategie z. B. auf randomisierte kontrollierte Studien beschränkt ist, können im Rahmen des Stellungnahmeverfahrens trotzdem nicht randomisierte Studien eingereicht werden. In solchen Fällen ist aber zusätzlich eine adäquate

Begründung für die Validität der kausalen Interpretation der in solchen Studien beschriebenen Effekte erforderlich. Die Fragestellung selber ist i. d. R. durch den Auftrag vorgegeben und nicht Gegenstand des Stellungnahmeverfahrens. Optional kann eine mündliche wissenschaftliche Erörterung mit Stellungnehmenden durchgeführt werden. Diese Erörterung dient der ggf. notwendigen Klarstellung von Inhalten der schriftlichen Stellungnahmen mit dem Ziel der Verbesserung der wissenschaftlichen Qualität des Berichtsplans. Um die Arbeit des Instituts nicht unangemessen zu verzögern, müssen die Stellungnahmen bestimmten formalen Anforderungen genügen. Weiterführende Angaben zum Stellungnahmeverfahren zum Berichtsplan in seiner vorläufigen Version sowie zu Amendments zum Berichtsplan finden sich auf den Internetseiten des Instituts in einem entsprechenden Leitfaden (www.iqwig.de/stellungnahmeverfahren.507.html). Es gelten die in der aktuellen Version dieses Leitfadens genannten Bedingungen.

Nach Auswertung der Stellungnahmen wird der Berichtsplan in einer überarbeiteten Version zusammen mit den Ergebnissen der Anhörung (schriftliche Stellungnahmen, Protokoll der wissenschaftlichen Erörterung, sofern eine Erörterung durchgeführt wurde, Würdigung der Stellungnahmen) publiziert. Dieser Berichtsplan ist Grundlage für die Erstellung des Vorberichts. Sofern weitere wesentliche methodische Änderungen im Verlauf der Vorberichtserstellung erforderlich sind, erfolgen diese i. d. R. in Form eines oder mehrerer Amendments zum Berichtsplan. Nach Veröffentlichung eines Amendments besteht i. d. R. ebenfalls Gelegenheit zur Stellungnahme zu diesem Amendment unter den o. g. Bedingungen.

Im **Vorbericht** werden die Ergebnisse der Informationsbeschaffung und der wissenschaftlichen Bewertung dargestellt. Um die Arbeit des Instituts nicht unangemessen zu verzögern, beginnt die Informationsbeschaffung und -bewertung bereits vor Abschluss des Stellungnahmeverfahrens zum Berichtsplan auf Basis der im vorläufigen Berichtsplan formulierten Kriterien. Das Ergebnis des Stellungnahmeverfahrens wird damit jedoch explizit nicht vorweggenommen, da diese Kriterien sich durch das Stellungnahmeverfahren zum Berichtsplan in seiner vorläufigen Version ändern können und dies dann zu einer Ergänzung und/oder Änderung der Informationsbeschaffung und -bewertung führen kann.

Der Vorbericht enthält die vorläufige Empfehlung an den G-BA. Er wird in Verantwortung der Projektgruppe des Instituts, i. d. R. unter Beteiligung externer Sachverständiger, erstellt. Nach Abschluss der internen Qualitätssicherung und der Freigabe durch das Steuergremium des Instituts wird der Vorbericht dann dem Auftraggeber (auch zur Frage der Vollständigkeit in Bezug auf den ursprünglich erteilten Auftrag) sowie dem Kuratorium und dem Vorstand der Stiftung zugeleitet. Zudem wird als weiterer Schritt der Qualitätssicherung der Vorbericht einem oder mehreren externen Gutachtern mit ausgewiesener methodischer und/oder fachlicher Kompetenz vorgelegt.

Die Veröffentlichung auf den Internetseiten des Instituts zwecks Gelegenheit zur Stellungnahme erfolgt i. d. R. fünf Arbeitstage nach Versendung an den Auftraggeber.

Für eine Frist von mindestens vier Wochen wird der Öffentlichkeit und damit auch den Sachverständigen der medizinischen, pharmazeutischen und gesundheitsökonomischen Wissenschaft und Praxis, den Berufsvertretungen der Apotheker, den Arzneimittelherstellern, den für die Wahrnehmung der Interessen der Patienten und die Selbsthilfe chronisch kranker und behinderter Menschen maßgeblichen Organisationen sowie dem oder der Beauftragten der Bundesregierung für die Belange der Patienten **Gelegenheit zur Stellungnahme (Anhörung)** gegeben (schriftliche Stellungnahmen). Dadurch wird ein offenes und unabhängiges Reviewverfahren zum Vorbericht ermöglicht. Gegenstand des Stellungnahmeverfahrens sind insbesondere die im Vorbericht dargestellten Ergebnisse der Informationsbeschaffung und -bewertung. Ebenso besteht zu diesem Zeitpunkt die Möglichkeit, qualitativ angemessene Unterlagen jeglicher Art, die aus Sicht des jeweiligen Stellungnehmenden zur Beantwortung der Fragestellung des Berichts geeignet sind, vorzulegen. Falls die in dem Berichtsplan definierte Suchstrategie z. B. auf randomisierte kontrollierte Studien beschränkt ist, können im Rahmen des Stellungnahmeverfahrens zum Vorbericht trotzdem nicht randomisierte Studien eingereicht werden. In solchen Fällen ist aber zusätzlich eine adäquate Begründung für die Validität der kausalen Interpretation der in solchen Studien beschriebenen Effekte erforderlich. Optional kann eine mündliche wissenschaftliche Erörterung mit Stellungnehmenden durchgeführt werden. Diese Erörterung dient der gegebenenfalls notwendigen Klarstellung von Inhalten der schriftlichen Stellungnahmen mit dem Ziel der Verbesserung der wissenschaftlichen Qualität des Abschlussberichts. Um die Arbeit des Instituts nicht unangemessen zu verzögern, müssen die Stellungnahmen bestimmten formalen Anforderungen genügen. Weiterführende Angaben zum Stellungnahmeverfahren zum Vorbericht finden sich auf den Internetseiten des Instituts in einem entsprechenden Leitfaden (www.iqwig.de/stellungnahmeverfahren.507.html). Es gelten die in der aktuellen Version dieses Leitfadens genannten Bedingungen.

Der **Abschlussbericht**, der auf dem Vorbericht aufbauend die Bewertung der wissenschaftlichen Erkenntnisse unter Berücksichtigung der Ergebnisse der Anhörung zum Vorbericht enthält, stellt das abschließende Produkt der Berichtserstellung dar. Er wird in Verantwortung der Projektgruppe des Instituts, i. d. R. unter Beteiligung externer Sachverständiger, erstellt. Nach Abschluss der internen Qualitätssicherung und der Freigabe durch das Steuergremium des Instituts wird der Abschlussbericht zunächst dem Auftraggeber und anschließend (i. d. R. vier Wochen später) dem Vorstand sowie dem Kuratorium der Stiftung zugeleitet, zusammen mit der Dokumentation der schriftlichen Stellungnahmen, dem Protokoll der wissenschaftlichen Erörterung, sofern eine Erörterung durchgeführt wurde, sowie einer Würdigung der Stellungnahmen. I. d. R. weitere vier

Wochen später erfolgt die Veröffentlichung dieser Dokumente (Abschlussbericht und Dokumentation und Würdigung der Stellungnahmen) auf den Internetseiten des Instituts. Sollten Stellungnahmen zu Abschlussberichten eingehen, die substanzielle nicht berücksichtigte Evidenz enthalten, oder erlangt das Institut auf andere Weise Kenntnis von solcher Evidenz, wird dem Auftraggeber begründet mitgeteilt, ob eine Neubeauftragung zu dem Thema (ggf. Aktualisierung des Berichts) aus Sicht des Instituts erforderlich erscheint oder nicht. Der Auftraggeber entscheidet über die Beauftragung des Instituts. Der Aktualisierungsprozess unterliegt den allgemeinen methodischen und verfahrenstechnischen Anforderungen an Produkte des Instituts.

2.1.2 *Rapid Reports*

Der Ablauf der Erstellung eines **Rapid Reports** ist in Abbildung 3 schematisch dargestellt. Alle Arbeitsschritte werden in Verantwortung des Instituts, ggf. auch unter Einbeziehung externer Expertise, getätigt. Je nach Notwendigkeit wird auch der wissenschaftliche Beirat des Instituts einbezogen. Der interne Qualitätssicherungsprozess ist in diesem Flussdiagramm nicht dargestellt.

Rapid Reports werden vorrangig mit dem Ziel einer zeitnahen Information zu relevanten Entwicklungen im Gesundheitswesen einschließlich neuer Technologien sowie zu Fragestellungen, die nicht auf Richtlinienentscheidungen des G-BA ausgerichtet sind, erstellt. Dabei ist zumeist eine kürzere Erarbeitungszeit notwendig. Der Ablauf der Erstellung von Rapid Reports unterscheidet sich von dem der Berichtserstellung insbesondere in zwei Punkten:

1. Es werden kein Berichtsplan und kein Vorbericht erstellt. Eine Anhörung ist während der Erarbeitung nicht vorgesehen. Arbeitsdokumente (Projektskizze, vorläufige Version des Rapid Reports) werden nicht veröffentlicht.
2. Die Bewertung erfolgt i. d. R. auf Basis bereits publizierter Informationen.

Nach der **Auftragsvergabe** durch den G-BA oder das BMG erfolgt die Bildung der internen Projektgruppe des Instituts unter der Leitung des zuständigen Ressorts. Eine Person wird als verantwortlicher Projektleiter bestimmt. Die Zusammensetzung der Projektgruppe ist an dieser Stelle nicht endgültig, da sich durch die folgenden Arbeitsschritte Änderungsbedarf ergeben kann. Die Formulierung der wissenschaftlichen Fragestellung erfolgt in Abstimmung mit den zuständigen Gremien des Auftraggebers, gegebenenfalls unter Einbeziehung externer fachlicher Expertise oder der Meinung einzelner Betroffener. Regelmäßig werden dabei insbesondere zur Definition patientenrelevanter Endpunkte entsprechende Patientenorganisationen beteiligt. Anschließend erfolgt die Erstellung der Projektskizze.

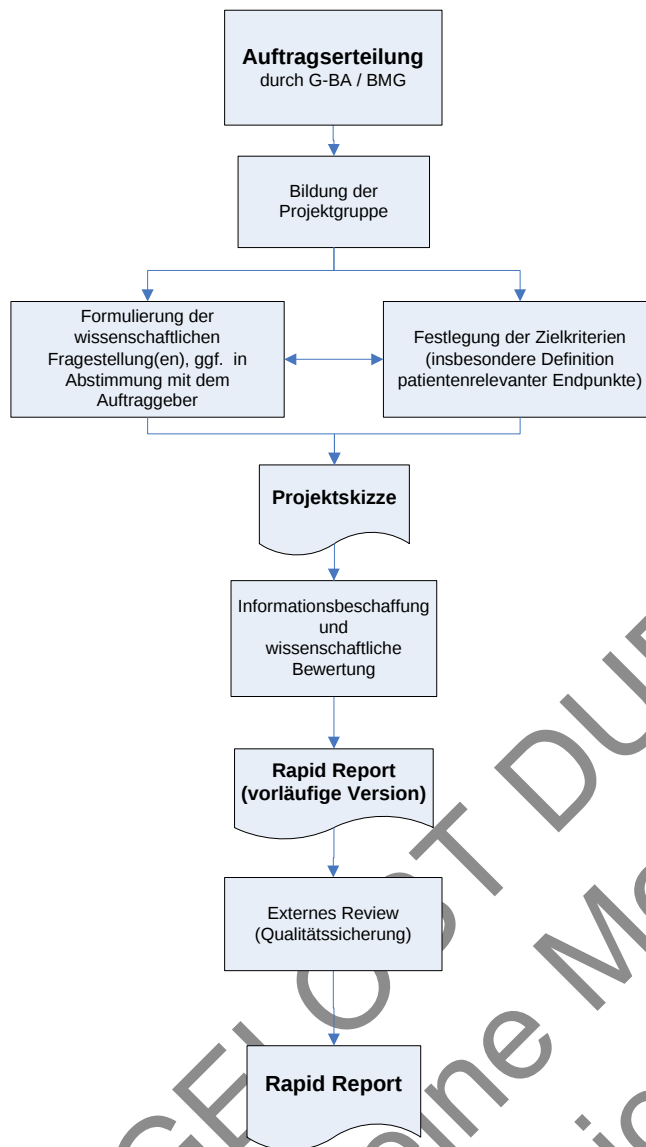


Abbildung 3: Ablauf der Erstellung eines Rapid Reports

In der **Projektskizze** werden die wesentlichen Schritte der Informationsbeschaffung und der wissenschaftlichen Bewertung zusammengefasst. Die Projektskizze wird nicht veröffentlicht.

Der **Rapid Report** wird zunächst in einer vorläufigen Version in Verantwortung der Projektgruppe des Instituts, i. d. R. unter Beteiligung externer Sachverständiger, erstellt. Nach Abschluss der internen Qualitätssicherung wird die vorläufige Version des Rapid Reports auch einem oder mehreren externen Gutachtern mit ausgewiesener methodischer und/oder fachlicher Kompetenz als weiterer Schritt der Qualitätssicherung vorgelegt. Im Anschluss wird der abschließende Rapid Report (Endbericht) erstellt. Nach Abschluss des internen Qualitätssicherungsverfahrens zu diesem Endbericht und der Freigabe durch das Steuergremium des Instituts erfolgt die Versendung an den Auftraggeber. I. d. R. vier Wochen später wird der Rapid Report dem Vorstand sowie dem

Kuratorium der Stiftung übermittelt und über die Internetseiten des Instituts veröffentlicht. Sollten Stellungnahmen zu Rapid Reports eingehen, die substanzielle nicht berücksichtigte Evidenz enthalten, oder erlangt das Institut auf andere Weise Kenntnis von solcher Evidenz, wird dem Auftraggeber begründet mitgeteilt, ob eine Neubeauftragung zu dem Thema (ggf. Aktualisierung des Rapid Reports) aus Sicht des Instituts erforderlich erscheint oder nicht. Der Auftraggeber entscheidet über die Beauftragung des Instituts. Der Aktualisierungsprozess unterliegt den generellen methodischen und verfahrenstechnischen Anforderungen an Produkte des Instituts.

2.1.3 Gesundheitsinformationen

Das Institut erstellt **Gesundheitsinformationen** für Patienten und die Öffentlichkeit in verschiedenen Formaten, die in Kapitel 5 ausführlicher vorgestellt werden. Zu den wichtigsten medizinischen Informationsprodukten des Instituts gehören:

- **Informationsberichte:** umfassende Beiträge, die die Grundlage für ein erweitertes Produktmodul zu einem relevanten gesundheitlichen Thema darstellen
- **Merkblätter:** kurze, leicht verständliche Informationen, wie sie z. B. auch vom G-BA zur Erläuterung von Richtlinien in Auftrag gegeben werden
- **Kurzantworten:** knappe Zusammenfassungen von systematischen Übersichten, HTA-Berichten oder größeren Studien, darunter auch Zusammenfassungen der anderen Produkte des Instituts, sofern sie für Patienten relevant sind

Diese Kernprodukte werden um verschiedene zusätzliche Produkte („zusätzliche Elemente“) ergänzt. Diese zusätzlichen Produkte sollen die Verständlichkeit der Kernaussagen der Gesundheitsinformationen verbessern und das Interesse daran steigern.

Der Ablauf der Erstellung von Gesundheitsinformationen ist in Abbildung 4 schematisch dargestellt. Alle Arbeitsschritte werden in Verantwortung des Instituts, ggf. unter Einbeziehung externer Expertise, getätigt. Der interne Reviewprozess ist in diesem Flussdiagramm *nicht* dargestellt. Der Prozess der selbst initiierten Themenfindung wird in Abschnitt 5.3.1 beschrieben.

Die vom Institut erstellten Patienten- bzw. Verbraucherinformationen gehen hervor:

- aus Direktaufträgen des G-BA oder des BMG zur Erstellung von Patienteninformationen;
- aus Zusammenfassungen der anderen Produkte des Instituts und als „Begleitinformationen“ für diese Produkte;
- aus der gesetzlichen Aufgabe des Instituts zur Bereitstellung von Gesundheitsinformationen für Bürger sowie aus Eigeninitiative des Instituts gemäß dem Generalauftrag des G-BA.

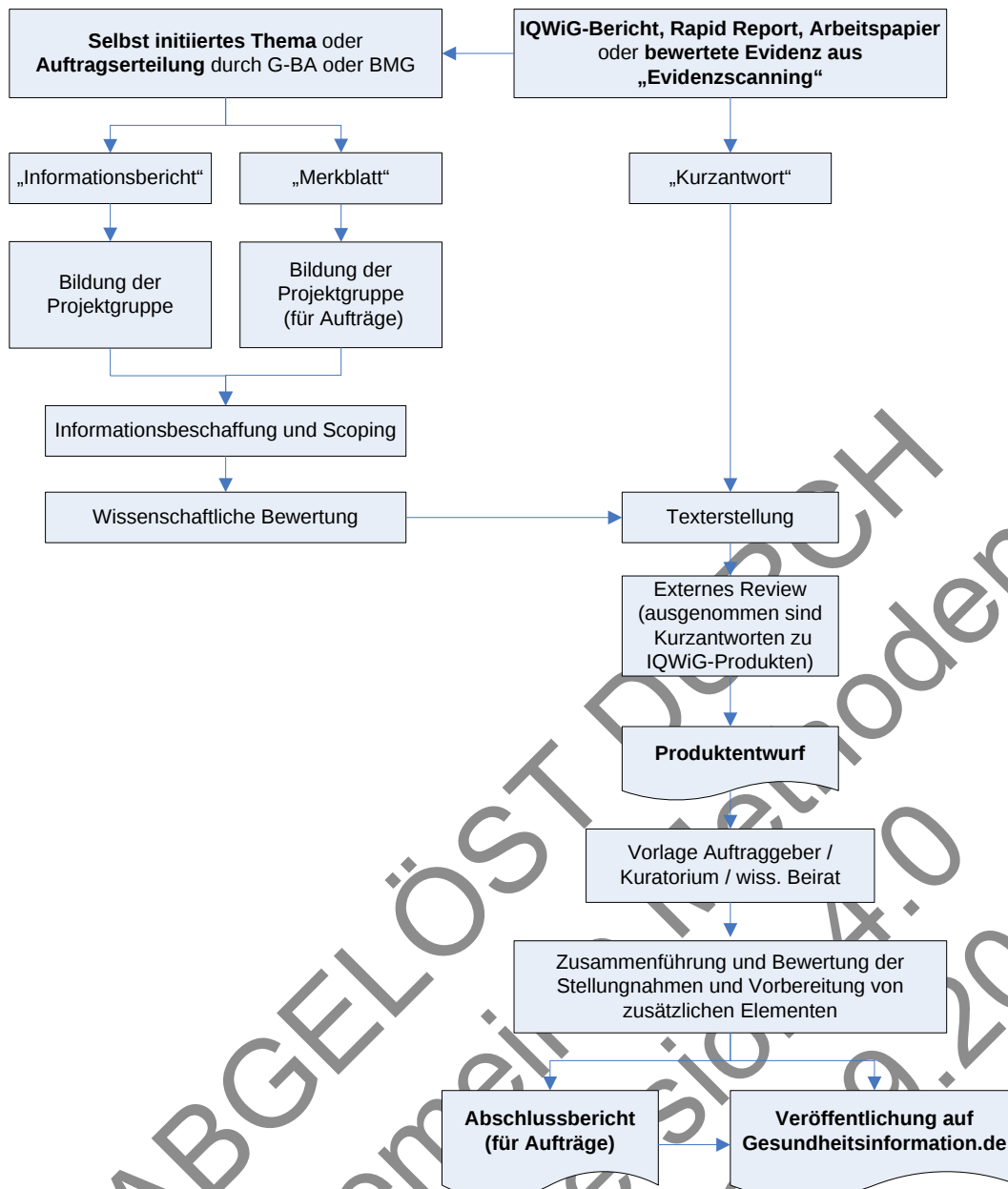


Abbildung 4: Ablauf der Gesundheitsinformationserstellung

Der Generalauftrag des Instituts (siehe Abschnitt 2.1.4) wurde im Juli 2006 konkretisiert, um auch Informationen für Bürger zu erfassen. Der Prozess zur Sichtung der Evidenz („Evidenzscanning“), der vom Institut dazu benutzt wird, potenzielle Themen für diese Informationen herauszuarbeiten, ist in Abschnitt 5.3.1 beschrieben. Kapitel 5 beschreibt zudem die Methodik der Recherche sowie der Einbeziehung von Patienten.

Nach der **Auftragsvergabe** durch den G-BA oder das BMG erfolgt die Bildung der internen Projektgruppe. Eine Person wird als verantwortlicher Projektleiter bestimmt. Den Projektgruppen

gehört mindestens ein Institutsmitglied an, das nicht aus dem Ressort „Gesundheitsinformation“ stammt. Ebenso wird für jeden „Informationsbericht“ eine Projektgruppe gebildet.

Die Begleitinformationen zu den Berichten und zu den anderen Produkten des Instituts werden in enger Zusammenarbeit mit dem Projektleiter und den für das jeweils beauftragte Produkt verantwortlichen Ressorts erstellt. Die Ressorts werden auch bei der Erstellung anderer Gesundheitsinformationen, die in ihre jeweiligen Zuständigkeitsbereiche fallen, konsultiert.

Im Anschluss an die **Texterstellung** und die ressortinterne Qualitätssicherung werden die Entwürfe zur **externen Begutachtung** verschickt. Die Entwürfe zu den „Informationsberichten“ und „Merkblättern“ werden an mindestens einen externen Reviewer in Deutschland und i. d. R. an mindestens einen externen Reviewer im Ausland versandt. „Kurzantworten“ werden dem Autor des wissenschaftlichen Beitrags, der zusammengefasst wurde, präsentiert sowie mindestens einem internen oder externen Reviewer. Die „Kurzantworten“ zu anderen Produkten des Instituts werden gewöhnlich nur von den internen Autoren des jeweiligen Ressorts begutachtet, können aber auch an externe Reviewer, z. B. externe Sachverständige, versandt werden, die an der Berichtserstellung beteiligt waren. Alle diese Kernprodukte werden zudem mindestens einem Patienten bzw. „Nutzer“ zur Prüfung vorgelegt.

Die „zusätzlichen Elemente“ unterliegen denselben internen Reviewprozessen wie die entsprechenden Kernprodukte; auch sie können ggf. extern begutachtet werden. Patientengeschichten („Erfahrungsberichte“, siehe Abschnitt 5.4.3) werden nur veröffentlicht, wenn eine schriftliche Einverständniserklärung des betreffenden Patienten vorliegt.

Der endgültige und vom Steuergremium freigegebene Entwurf zu einem Informationsprodukt wird im Rahmen einer einmonatigen Beratungsperiode zur **Stellungnahme** an den Auftraggeber und an das Kuratorium verschickt. Dieses Gremium repräsentiert relevante „Stakeholder“ einschließlich Patientenvertretern. Im Gegensatz zu den Vorberichten werden Entwürfe zu Patienteninformationen nicht auf der Website des Instituts veröffentlicht.

Die während dieser Beratungsperiode eingegangenen Kommentare werden zusammengefasst und geprüft. Die Zusammenfassungen und die ausführlichen Versionen stehen auch der Projektgruppe – falls vorhanden – und dem Steuergremium zur Verfügung.

Im Falle von Aufträgen zu Patienteninformationen wird dem Auftraggeber auch über die Ergebnisse der Stellungnahmen Bericht erstattet. Abschlussberichte für im Auftrag erstellte Patienteninformationen unterliegen hinsichtlich der Publikation demselben Verfahren wie andere Abschlussberichte. Sie werden zunächst dem Auftraggeber und anschließend (i. d. R. vier Wochen später) dem Vorstand sowie dem Kuratorium der Stiftung zugeleitet. I. d. R. weitere vier Wochen

später erfolgt die Veröffentlichung des Abschlussberichts auf den Internetseiten (www.iqwig.de) des Instituts. Anschließend wird die betreffende Gesundheitsinformation selbst auf der Institutswebsite für Patienten und Bürger unter www.gesundheitsinformation.de publiziert. Die Texte für Patienteninformationen zur Erläuterung von G-BA-Richtlinien werden erst nach Veröffentlichung der Richtlinien selbst unter www.gesundheitsinformation.de veröffentlicht.

„Kurzantworten“ als Begleitung von Produkten des Instituts kommen i. d. R. gleichzeitig mit dem Abschlussbericht unter www.gesundheitsinformation.de zur Veröffentlichung.

Korrekturen, Verbesserungen und Aktualisierungen der Gesundheitsinformationen des Instituts werden intern vorgenommen, bei wichtigeren inhaltlichen Aktualisierungen auch durch externe Gutachter. Eine ausführlichere Beschreibung der Aktualisierungsmechanismen des Instituts findet sich in Kapitel 5.

2.1.4 Arbeitspapiere

Der Ablauf der Erstellung eines **Arbeitspapiers** ist in Abbildung 5 schematisch dargestellt. Alle Arbeitsschritte werden in Verantwortung des Instituts, im Ausnahmefall unter Einbeziehung externer Expertise, getätigt. Je nach Notwendigkeit wird auch der wissenschaftliche Beirat des Instituts einbezogen. Der interne Qualitätssicherungsprozess ist in diesem Flussdiagramm nicht dargestellt.

Die Erstellung von Arbeitspapieren erfolgt u. a. im Rahmen des **Generalauftrags** vom 21.12.2004 durch den G-BA. Mit dem Generalauftrag wurde das Institut beauftragt, „durch die Erfassung und Auswertung des relevanten Schrifttums eine kontinuierliche Beobachtung und Bewertung medizinischer Entwicklungen von grundlegender Bedeutung und ihrer Auswirkungen auf die Qualität und Wirtschaftlichkeit der medizinischen Versorgung in Deutschland vorzunehmen und den G-BA hierüber regelmäßig zu informieren. Der G-BA geht bei diesem Auftrag davon aus, dass das Institut auf den ihm gemäß § 139a Abs. 3 SGB übertragenen Arbeitsfeldern nicht nur Einzelaufträge des G-BA bearbeitet, sondern aus der eigenverantwortlichen wissenschaftlichen Arbeit heraus dem G-BA für dessen gesetzliche Aufgaben notwendige Informationen über versorgungsrelevante Entwicklungen in der Medizin zur Verfügung stellt und konkrete Vorschläge für Einzelaufträge erarbeitet, die aus Sicht des Instituts vor dem Hintergrund dieser Informationen relevant sind.“

Aus dem gesetzlichen Auftrag des Instituts und dem Generalauftrag folgt das Erfordernis eigenständiger wissenschaftlicher Arbeit. Hierunter sind u. a. auch Projekte zur methodischen Weiterentwicklung zu verstehen, deren Ergebnisse ebenfalls als Arbeitspapiere veröffentlicht werden können.

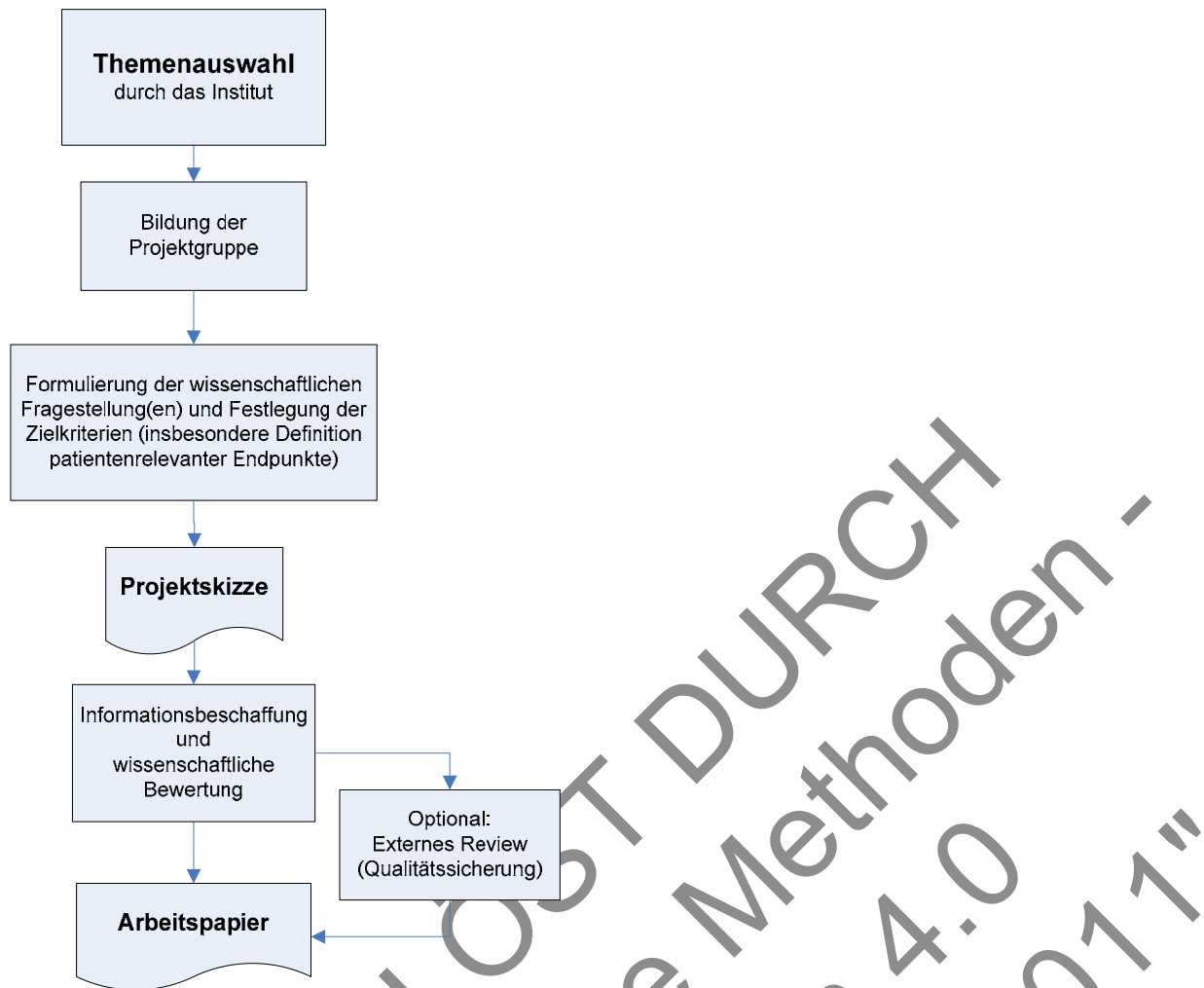


Abbildung 5: Ablauf der Erstellung eines Arbeitspapiers

Die **Themenauswahl** erfolgt i. d. R. durch eine kontinuierliche Beobachtung der medizinisch-wissenschaftlichen Literatur und sonstiger Informationsquellen innerhalb der Ressorts des Instituts. Auf Basis dieser Erkenntnisse wird dem Steuergremium ein Thema zur Bearbeitung vorgeschlagen. Das Steuergremium entscheidet über die Bearbeitung des vorgeschlagenen Themas.

In der **Projektskizze** werden die wesentlichen Schritte der Informationsbeschaffung und der wissenschaftlichen Bewertung zusammengefasst. Dabei können, insbesondere zur Definition patientenrelevanter Endpunkte, auch entsprechende Patientenorganisationen beteiligt werden. Die Projektskizze wird nicht veröffentlicht.

Das **Arbeitspapier** wird in Verantwortung der Projektgruppe des Instituts erstellt. Eine Beteiligung externer Sachverständiger erfolgt nur im Ausnahmefall. Die Qualitätssicherung kann optional ein externes Review beinhalten. Nach Abschluss der internen und ggf. externen Qualitätssicherung und

der Freigabe durch das Steuergremium des Instituts erfolgt die Versendung an den G-BA. I. d. R. 4 Wochen später wird das Arbeitspapier (Endbericht) dem Vorstand sowie dem Kuratorium der Stiftung übermittelt und über die Internetseiten des Instituts veröffentlicht. Sollten Stellungnahmen zu Arbeitspapieren eingehen, die substanzielle nicht berücksichtigte Evidenz enthalten, oder erlangt das Institut auf andere Weise Kenntnis von solcher Evidenz, wird geprüft, ob eine Aktualisierung des Arbeitspapiers aus Sicht des Instituts erforderlich erscheint oder nicht. Der Aktualisierungsprozess unterliegt den generellen methodischen und verfahrenstechnischen Anforderungen an Produkte des Instituts.

2.2 Generelle Aspekte bei der Produkterstellung

In diesem Kapitel werden folgende, für alle Produkte geltende Abläufe und Aspekte dargestellt:

- die Auswahl externer Sachverständiger für die Mitarbeit an der Produkterstellung
- die Gewährleistung der fachlichen Unabhängigkeit bei der Produkterstellung
- die Begutachtung der Produkte
- die Veröffentlichung der Produkte

2.2.1 Auswahl externer Sachverständiger

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

Die Vergabe von wissenschaftlichen Forschungsaufträgen erfolgt nach den eigenen Vergabe- und Beschaffungsrichtlinien des Instituts, denen die Bestimmungen des § 22 der Verordnung zum Haushaltswesen in der Sozialversicherung (SVHV) sowie die gesetzlichen Vorschriften des Vergaberechts zugrunde liegen. Das Institut informiert in Fachzeitschriften darüber, dass es auf seiner Website (www.iqwig.de) in regelmäßigen Abständen die an externe Sachverständige zu vergebenden Projekte veröffentlicht. Die aktuell durch das Institut zu vergebenden Aufträge sind auf einer Internetseite des Instituts (www.iqwig.de/index.174.html) aufgelistet. Aufträge mit einem Auftragsvolumen oberhalb des aktuell geltenden Schwellenwerts der Vergabebestimmungen der Europäischen Union (EU) werden EU-weit ausgeschrieben.

Die konkreten Eignungsanforderungen an die Bewerber sind den jeweiligen Bekanntmachungen bzw. den Verdingungsunterlagen zu entnehmen. Für die ordnungsgemäße Durchführung des Vergabeverfahrens hat das Institut eine Vergabestelle eingerichtet.

2.2.2 Gewährleistung der fachlichen Unabhängigkeit

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 Stiftung verankert.

A) Gewährleistung der Unabhängigkeit externer Sachverständiger

Vor dem Abschluss eines Vertrages zwischen dem Institut und einem externen Sachverständigen oder einer externen Institution zur Erstellung eines Produkts muss entschieden werden, ob Bedenken wegen potenzieller Interessenkonflikte bestehen. Dafür müssen alle externen Sachverständigen und Institutionen alle Tätigkeiten darlegen, die *potenziell* geeignet sind, ihre fachliche Unabhängigkeit zu beeinflussen. Als Interessenkonflikte werden dabei in Anlehnung an entsprechende Regelungen wissenschaftlicher medizinischer Zeitschriften insbesondere folgende Kriterien angesehen: alle finanziellen Abmachungen, Beschäftigungen, Beratungen, Honorare, bezahlten Expertenaussagen, Reiseunterstützungen, Patentanträge sowie Aktienbesitz innerhalb der letzten drei Jahre, die bzw. der die beauftragte Arbeit inhaltlich beeinflussen könnten, sowie alle aktuellen persönlichen Beziehungen zu anderen Personen oder Organisationen, die die beauftragte Arbeit inhaltlich beeinflussen könnten [214]. Diese Kriterienliste findet sich auch auf dem im Internet zur Verfügung gestellten Formblatt, das ggf. aktualisiert wird. Maßgeblich ist immer das jeweils auf den Internetseiten abrufbare Formblatt. Die Namen der externen Sachverständigen, die an der Erstellung von Berichten oder Rapid Reports beteiligt waren, werden in der Regel im Abschlussbericht bzw. Rapid Report veröffentlicht. Auf ausdrücklichen Wunsch der externen Sachverständigen, auf Aufforderung des Auftraggebers oder aufgrund anderer wichtiger Umstände ist es möglich, die Namen externer Sachverständiger zur Gewährleistung ihrer Unabhängigkeit und zur Vermeidung einer interessenbedingten Einflussnahme nicht zu veröffentlichen. Entsprechende Regelungen wurden für die Arbeit des G-BA im SGB beschrieben (§ 35 Abs. 1 S. 8 SGB V). Wenn die Namen der externen Sachverständigen publiziert werden, erfolgt auch eine Veröffentlichung der von ihnen dargelegten potenziellen Interessenkonflikte. Dies geschieht dergestalt, dass für die im oben genannten Formblatt aufgeführten Kriterien berichtet wird, ob zum jeweiligen Kriterium ein potenzieller Interessenkonflikt beschrieben wurde oder nicht. Eine Veröffentlichung z. B. des finanziellen Umfangs von Zuwendungen erfolgt nicht. Der Prozess zur Auswahl von externen Sachverständigen ist in Abschnitt 2.2.1 beschrieben.

B) Gewährleistung der internen fachlichen Unabhängigkeit

Die interne fachliche Unabhängigkeit wird – soweit möglich – durch die Auswahl der Mitarbeiter gewährleistet. Diese müssen im Rahmen ihrer Einstellung sowohl glaubhaft ihr bisheriges Tätigkeitsumfeld darlegen als auch versichern, (Neben-)Tätigkeiten, die ihre fachliche

Unabhängigkeit potenziell infrage 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 bzw. im wiederholten oder schwerwiegenden Fall zu einer Kündigung des Arbeitsverhältnisses führen. Im Einzelfall entscheidet die Institutsleitung nach Beratung durch das Steuergremium, ob ein Mitarbeiter von einer bestimmten Tätigkeit oder einem bestimmten Projekt aus Gründen der Befangenheit ausgeschlossen werden muss.

2.2.3 Begutachtung der Produkte des Instituts

Die Begutachtung der Produkte des Instituts hat insbesondere zum Ziel, eine hohe wissenschaftliche Qualität der Produkte zu gewährleisten. Darüber hinaus können für einzelne Produkte auch andere Ziele wesentlich sein, z. B. die Allgemeinverständlichkeit.

Alle Produkte einschließlich der jeweiligen Zwischenprodukte unterliegen einem umfangreichen mehrstufigen internen Qualitätssicherungsverfahren. Darüber hinaus wird im Verlauf der Erstellung von Berichten, Rapid Reports und z. T. auch von Gesundheitsinformationen ein externes Reviewverfahren als weiterer Schritt der Qualitätssicherung durchgeführt. Bei Arbeitspapieren und bestimmten Gesundheitsinformationen (siehe Abschnitt 2.1.3) ist das externe Reviewverfahren optional. Die Auswahl der internen und externen Gutachter erfolgt primär auf Basis ihrer methodischen und/oder fachlichen Expertise.

Die Frage, wie effektiv bestimmte Verfahrensweisen beim externen Review sind, wird erst in jüngerer Zeit in gezielten Studien untersucht. Bislang gibt es allerdings nur wenige aussagekräftige Untersuchungen dazu [135,300]. Insbesondere gibt es keine ausreichend validen Interventionsstudien. Es gibt keine Evidenz, dass eine bestimmte Vorgehensweise besser als eine andere ist. Es ist beispielsweise unklar, wie viele Personen in ein Review einbezogen werden sollten. Der Stellenwert der üblichen Verfahren in medizinischen Zeitschriften [135,215], einschließlich der Bewertung durch Konsumenten und Patienten [32], ist nach bislang vorliegenden Untersuchungen noch nicht ausreichend geklärt [215]. Weiterhin gibt es verschiedene Methoden, um die Qualität der einzelnen Reviews zu bewerten. Auch diesbezüglich gibt es keine Evidenz, ob und welches Verfahren das beste ist.

Die Identifikation und Auswahl potenzieller externer Reviewer sind abhängig vom Umfang des beauftragten Reviews. Sehr umfangreiche externe Reviews können auch als wissenschaftliche

Forschungsaufträge vergeben werden. Für diese gelten dann die in Abschnitt 2.2.1 genannten Bedingungen. Ansonsten kann die Identifikation externer Reviewer durch eine entsprechende Recherche, durch die Kenntnis der Projektgruppe, durch das Ansprechen von Fachgesellschaften, durch eine Bewerbung im Rahmen der Ausschreibung für die Auftragsbearbeitung usw. erfolgen. Eine Darlegung potenzieller Interessenkonflikte muss aber in jedem Fall erfolgen.

Die Auswahl der externen Reviewer erfolgt durch die Projektgruppe bzw. die an der Produkterstellung beteiligten Personen. Eine Höchstgrenze von Reviewern gibt es nicht. Die externen Gutachten werden hinsichtlich ihrer Relevanz für das jeweilige Produkt geprüft. Eine Veröffentlichung der externen Gutachten erfolgt nicht. Die Namen der externen Gutachter von Berichten und Rapid Reports werden i. d. R. im Abschlussbericht bzw. Rapid Report veröffentlicht, einschließlich einer Darstellung ihrer potenziellen Interessenkonflikte, analog zur Vorgehensweise bei externen Sachverständigen.

Neben dem oben beschriebenen externen Qualitätssicherungsverfahren unter Beteiligung vom Institut ausgewählter und beauftragter Gutachter ist durch die Veröffentlichung der Institutsprodukte und die damit verbundene Möglichkeit zur Stellungnahme ein offenes und unabhängiges Reviewverfahren gewährleistet.

2.2.4 Veröffentlichung der Produkte des Instituts

Wesentliche Aufgabe des Institutes ist es, die vorhandene Evidenz durch eine sorgfältige Prüfung von verfügbaren Informationen zu ergründen und über die Ergebnisse dieser Prüfung zu informieren. Es ist gesetzlich festgelegt, dass das Institut „in regelmäßigen Abständen über die Arbeitsprozesse und -ergebnisse einschließlich der Grundlagen für die Entscheidungsfindung öffentlich zu berichten“ hat (§ 139a Abs. 4 SGB V).

Zur Wahrung der Unabhängigkeit des Institutes muss ausgeschlossen werden, dass die Auftraggeber oder interessierte Dritte Einfluss auf die Inhalte der Berichte nehmen können. Dies könnte zu einer Vermengung der wissenschaftlichen Ergebnisse mit politischen und/oder wirtschaftlichen Aspekten und/oder Interessen führen. Gleichzeitig muss vermieden werden, dass das Institut seinerseits bestimmte Ergebnisse zurückhält. Deshalb werden alle Ergebnisse, die das Institut im Rahmen seiner gesetzlichen Aufgaben erarbeitet, möglichst zeitnah veröffentlicht. Dazu gehört im Falle von Berichten auch der Berichtsplan. Produktspezifische Besonderheiten sind in den jeweiligen Abschnitten, in denen die Verfahrensabläufe beschrieben werden, angegeben. Von den dargestellten Regelfällen (Latenzzeit zwischen Fertigstellung und Veröffentlichung) kann in begründeten Ausnahmefällen abgewichen werden.

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

3 Nutzen- und Schadenbewertung medizinischer Interventionen

3.1 Patientenrelevanter medizinischer Nutzen

3.1.1 Definition des patientenrelevanten medizinischen Nutzens

Zur Definition des patientenrelevanten medizinischen Nutzens ist zunächst eine Abgrenzung der Begriffe „Notwendigkeit“ und „Nutzen“ sinnvoll. Nach § 27 SGB V haben Versicherte Anspruch auf Krankenbehandlung, wenn sie notwendig ist, um eine Krankheit zu erkennen, zu heilen, ihre Verschlimmerung zu verhüten oder Krankheitsbeschwerden zu lindern.

Dies impliziert zunächst einmal zentrale Aspekte des Nutzens, die unter die Oberbegriffe „Mortalität“, „Morbidität“ und „Lebensqualität“ fallen. Der Begriff der Notwendigkeit geht allerdings über den Begriff des Nutzens hinaus. Der Nachweis des Nutzens ist demnach eine notwendige, aber keine hinreichende Bedingung für den Nachweis der Notwendigkeit. Demzufolge kann eine Maßnahme ohne belegten Nutzen nicht als notwendig bezeichnet werden.

Der Begriff „**Nutzenbewertung**“ bezeichnet den gesamten Prozess der Evaluation medizinischer Interventionen hinsichtlich ihrer kausal begründeten positiven und negativen Effekte im Vergleich mit einer klar definierten anderen Therapie, einem Placebo (oder einer andersartigen Scheinbehandlung) oder keiner Behandlung. „Kausal begründet“ meint in diesem Zusammenhang, dass ausreichende Sicherheit besteht, dass beobachtete Effekte allein auf die zu prüfende Intervention zurückgeführt werden können [381].

Mit dem Begriff „**Nutzen**“ werden kausal begründete positive Effekte, mit dem Begriff „**Schaden**“ kausal begründete negative Effekte einer medizinischen Intervention auf patientenrelevante Endpunkte (s. u.) bezeichnet. Die Beschreibung von Nutzen und Schaden erfolgt dabei immer ausgehend von der zu evaluierenden Intervention.

Nutzen bzw. Schaden wird im Vergleich mit Placebo (oder einer andersartigen Scheinbehandlung) oder keiner Behandlung festgestellt. Im Falle eines Vergleichs der zu evaluierenden medizinischen Intervention mit einer anderen eindeutig definierten medizinischen Intervention werden für die vergleichende Bewertung der Nutzen- bzw. Schadenaspekte folgende Begriffe verwendet:

- Nutzenaspekte:
 - Im Falle eines höheren Nutzens wird von „Zusatznutzen“ gesprochen.
 - Im Falle eines geringeren oder vergleichbaren Nutzens wird von einem „geringeren“ bzw. „vergleichbaren Nutzen“ gesprochen.

- Schadenaspekte:
 - Es werden die Begriffe „höherer“, „vergleichbarer“ und „geringerer Schaden“ verwendet; der Begriff „Zusatzschaden“ soll vermieden werden.

Wenn möglich, wird eine Nutzen-Schaden-Abwägung vorgenommen (siehe Abschnitt 3.1.4).

Die Bewertung der Evidenz soll in eine eindeutige Feststellung münden, dass entweder das Vorliegen eines (Zusatz-)Nutzens (bzw. Schadens) einer Maßnahme oder das Fehlen eines (Zusatz-)Nutzens (bzw. Schadens) belegt ist oder das Vorliegen oder Fehlen eines (Zusatz-)Nutzens (bzw. Schadens) nicht belegt und daher unklar ist, ob ein (Zusatz-)Nutzen (bzw. Schaden) durch die Maßnahme erzielt wird. Die Bewertung der medizinischen Notwendigkeit beruht (a) auf der Art und dem Ausmaß des ggf. belegten Nutzens bzw. Zusatznutzens sowie ggf. auch (b) darauf, ob dieser Nutzen bzw. Zusatznutzen allein durch die zu bewertende Intervention erzielt werden kann.

Da sich der Nutzen einer Maßnahme auf den Patienten beziehen soll, beruht diese Bewertung auf Ergebnissen wissenschaftlicher Untersuchungen zur Beeinflussung patientenrelevanter Endpunkte. Als „patientenrelevant“ soll in diesem Zusammenhang verstanden werden, wie ein Patient fühlt, seine Funktionen und Aktivitäten wahrnehmen kann oder überlebt [44]. Dabei werden sowohl die beabsichtigten als auch die unbeabsichtigten Effekte der Interventionen berücksichtigt, die eine Bewertung der Beeinflussung insbesondere folgender patientenrelevanter Zielgrößen zur Feststellung krankheits- und behandlungsbedingter Veränderungen erlauben:

1. Mortalität,
2. Morbidität (Beschwerden und Komplikationen),
3. gesundheitsbezogene Lebensqualität.

Ergänzend können der interventions- und erkrankungsbezogene Aufwand und die Zufriedenheit der Patienten mit der Behandlung berücksichtigt werden. Solche Aspekte kommen allerdings in der Regel nur als sekundäre Zielgrößen in Betracht.

Gemäß § 35b des SGB V sollen beim Patientennutzen insbesondere die Verbesserung des Gesundheitszustandes, eine Verkürzung der Krankheitsdauer, eine Verlängerung der Lebensdauer, eine Verringerung der Nebenwirkungen sowie eine Verbesserung der Lebensqualität angemessen berücksichtigt werden. Diese Nutzendimensionen sind durch die oben aufgeführten Zielgrößen abgebildet. So sind die Verbesserung des Gesundheitszustandes und die Verkürzung der Krankheitsdauer Aspekte der unmittelbaren krankheitsbedingten Morbidität, die Verringerung von Nebenwirkungen ist ein Aspekt der therapiebedingten Morbidität.

Es werden in erster Linie Endpunkte berücksichtigt, die zuverlässig und direkt konkrete Änderungen des Gesundheitszustandes abbilden. Dabei werden insbesondere einzelne Betroffene,

Patientenvertretungs- und/oder Verbraucherorganisationen in die themenbezogene Definition patientenrelevanter Endpunkte einbezogen. Zur Erfassung der Lebensqualität und der Patientenzufriedenheit sollen nur Instrumente eingesetzt werden, die für den Einsatz in klinischen Studien geeignet und entsprechend evaluiert sind [122,365]. Darüber hinaus können für die Nutzenbewertung valide Surrogatendpunkte berücksichtigt werden.

Sowohl Nutzen- als auch Schadenaspekte können eine unterschiedliche Wichtigkeit für die Betroffenen haben, die sich ggf. durch qualitative Erhebungen oder bereits bei der Beratung durch Betroffene, Patientenvertretungs- und/oder Verbraucherorganisationen im Zusammenhang mit der Definition patientenrelevanter Endpunkte abzeichnet. Daher kann in der zusammenfassenden Betrachtung eine unterschiedliche Gewichtung der Zielgrößen vorgenommen werden. Das hat dann zur Folge, dass die Voraussetzung für die vergleichende Würdigung bestimmter (geringer gewichteter) Zielgrößen die vergleichende Würdigung anderer (höher gewichteter) Zielgrößen ist. So kann z. B. eine geringere Ausprägung unerwünschter Effekte nur dann einen Zusatznutzen für die Patienten bedeuten, wenn gleichzeitig die höher gewichteten Zielgrößen in einem ähnlichen bzw. akzeptabel geringeren Ausmaß als dem der Vergleichsintervention günstig beeinflusst werden. Dies setzt allerdings voraus, dass eine a priori Definition dessen, was noch als akzeptabel geringer angesehen werden kann (Irrelevanzbereich, siehe Abschnitt 6.4.5), vorliegt.

Diagnostische Maßnahmen können indirekt nutzen, indem sie eine notwendige Voraussetzung für therapeutische Interventionen sind, durch die das Erreichen einer Wirkung auf die oben genannten patientenrelevanten Zielgrößen möglich wird. Voraussetzung für den Nutzen diagnostischer Maßnahmen sind also die Existenz und der belegte Nutzen einer Behandlung bei Patienten in Abhängigkeit vom Testergebnis. Darüber hinaus können diagnostische Tests in patientenrelevanter Weise das Treffen persönlicher Lebensentscheidungen ermöglichen und damit ggf. einen Nutzen entfalten.

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

3.1.2 *Surrogate des patientenrelevanten medizinischen Nutzens*

Surrogatendpunkte werden in der medizinischen Forschung häufig als Ersatz für patientenrelevante Endpunkte verwendet, meist um Aussagen zum patientenrelevanten (Zusatz-)Nutzen früher und einfacher zu erhalten [13,134,299]. Die meisten Surrogatendpunkte sind jedoch in dieser Hinsicht nicht verlässlich und können bei der Nutzenbewertung irreführend sein [153,161]. Surrogatendpunkte werden im Rahmen der Nutzenbewertung des Institutes in der Regel nur dann

als Beleg des (Zusatz-)Nutzens einer Intervention in Betracht gezogen, wenn zuvor anhand geeigneter statistischer Methoden gezeigt wurde, dass der Effekt von Interventionen mit vergleichbarem Wirkmechanismus auf den zu ersetzenden patientenrelevanten Endpunkt durch den Effekt auf den Surrogatendpunkt in einem ausreichenden Ausmaß erklärt wird [26,378]. In der Regel sind dafür eindeutige Belege aus Interventionsstudien eines plausiblen, starken, konsistenten und gleichgerichteten Zusammenhangs zwischen der Änderung des Surrogats und der Änderung des patientenrelevanten Endpunkts erforderlich. Ein gleichgerichteter Zusammenhang bedeutet, dass eine positive bzw. negative Veränderung des Surrogates mit einer positiven bzw. negativen Veränderung des patientenrelevanten Endpunktes einhergeht.

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

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

Beobachtete Zusammenhänge zwischen einem Surrogat- und dem entsprechenden patientenrelevanten Endpunkt für eine Intervention mit einem bestimmten Wirkmechanismus sind nicht zwangsläufig auf Interventionen zur Behandlung derselben Erkrankung, aber mit einem anderen Wirkmechanismus, anwendbar [153,161].

Surrogatendpunkte von unklarer oder umstrittener Validität können zwar in den Berichten des Institutes dargestellt werden. Derartige Endpunkte sind aber nicht als Belege für den Nachweis des (Zusatz-)Nutzens einer Intervention geeignet. Bei besonders gravierenden Erkrankungen im Sinne von Morbidität und Mortalität ohne Therapiealternative werden Surrogatendpunkte unklarer Validität ggf. als Hinweis auf einen Nutzen akzeptiert werden müssen.

Je nach Nähe zu einem entsprechenden patientenrelevanten Endpunkt werden in der Literatur verschiedene andere Bezeichnungen für Surrogatendpunkte verwendet (z. B. intermediärer Endpunkt). Auf eine diesbezügliche Differenzierung wird an dieser Stelle allerdings verzichtet, da die Frage nach der notwendigen Validität davon unberührt bleibt.

3.1.3 Ermittlung des Schadenpotenzials medizinischer Interventionen

Die Anwendung einer medizinischen Intervention gleich welcher Art (medikamentös, nichtmedikamentös, chirurgisch, diagnostisch, präventiv etc.) birgt das Risiko unerwünschter Wirkungen in sich. „Unerwünschte Wirkungen“ bezeichnet dabei alle diejenigen Ereignisse und Wirkungen, die einen individuell wahrgenommenen oder objektiv feststellbaren physischen oder psychischen Schaden darstellen, der zu einer mehr oder weniger starken kurz- oder langfristigen Verkürzung der Lebenserwartung, Erhöhung der Morbidität oder Beeinträchtigung der Lebensqualität führt oder führen kann. Für eine Begriffsklärung ist anzumerken, dass bei Verwendung des Begriffs „unerwünschte Wirkungen“ ein kausaler Zusammenhang mit der Intervention angenommen wird, während dies bei Verwendung des Begriffs „unerwünschte Ereignisse“ noch offenbleibt [79].

Das Schadenpotenzial beschreibt das Risiko für das Auftreten unerwünschter Wirkungen bei der Anwendung einer medizinischen Intervention. Die Darstellung des Schadenpotenzials ist wesentlicher und gleichberechtigter Bestandteil der Nutzenbewertung einer Intervention. Sie gewährleistet eine informierte, populationsbezogene, aber auch individuelle Nutzen-Schaden-Abwägung [390]. Voraussetzung dafür ist, dass anhand der vorliegenden Daten die Effektstärken einer medizinischen Intervention sowohl für die erwünschten als auch für die unerwünschten Wirkungen im Vergleich z. B. zu therapeutischen Alternativen beschrieben werden können.

Die Darstellung, Analyse und Bewertung des potenziellen Schadens einer medizinischen Intervention im Rahmen einer systematischen Übersicht sind jedoch in vielen Fällen ungleich schwerer als die des (Zusatz-)Nutzens. Dies betrifft insbesondere unerwartet auftretende unerwünschte Ereignisse [79]. Typischerweise sind Studien so konzipiert, dass sie den Einfluss einer medizinischen Intervention auf wenige, vorab definierte Wirksamkeitsendpunkte messen sollen. Die Ergebnisse bezüglich unerwünschter Ereignisse in solchen Studien hängen stark von der zugrunde gelegten Methodik der Erfassung ab [41,206]. Des Weiteren ist die zielgerichtete Forschung zur Erhebung insbesondere der seltenen schwerwiegenden unerwünschten Wirkungen einschließlich der Darstellung eines kausalen Zusammenhangs mit der medizinischen Intervention erheblich unterrepräsentiert [48,109,205]. Darüber hinaus ist die Qualität der Berichterstattung bezüglich unerwünschter Ereignisse in Einzelstudien mangelhaft, was kürzlich zu einer Erweiterung des CONSORT-Statements für randomisierte klinische Studien geführt hat [204]. Schließlich wird die systematische Bewertung unerwünschter Wirkungen einer Intervention auch dadurch erschwert, dass die Kodierung in Literaturdatenbanken diesbezüglich unzureichend ist und die zielgerichtete Suche nach relevanter wissenschaftlicher Literatur deshalb häufig ein unvollständiges Bild liefert [88].

Folge der oben genannten Hindernisse ist, dass in vielen Fällen trotz enormer Anstrengungen die Aussagenunsicherheit bei der Darstellung des Schadenpotenzials höher sein wird als bei der Darstellung positiver Effekte [252]. Notwendig ist hier, eine sinnvolle Balance zwischen der Vollständigkeit der Aufarbeitung und dem Ressourcenaufwand zu finden. Daraus folgend ist es notwendig, die Aufarbeitung und Darstellung auf relevante unerwünschte Wirkungen zu beschränken. Als relevant können dabei insbesondere diejenigen unerwünschten Wirkungen bezeichnet werden, die

- den Nutzen der Intervention teilweise oder ganz aufwiegen könnten,
- sich zwischen zwei oder mehr ansonsten gleichwertigen Therapieoptionen erheblich unterscheiden könnten,
- insbesondere bei denjenigen Therapieoptionen auftreten, die besonders effektiv sein könnten,
- eine Dosis-Wirkungs-Beziehung haben könnten,
- von Patienten als besonders wichtig angesehen werden und die
- von schwerwiegender Morbidität oder gar erhöhter Mortalität begleitet sind oder in Folge mit einer erheblichen Beeinträchtigung der Lebensqualität einhergehen.

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

- Für die Nutzenbewertung wird zunächst angestrebt, eine Auswahl potenziell relevanter unerwünschter Wirkungen, die bei der Entscheidungsfindung für oder gegen den Einsatz der zu evaluierenden Intervention einen wesentlichen Stellenwert haben, zusammenzustellen. Dabei erfolgt eine Auswahl unerwünschter Wirkungen und Ereignisse nach den oben genannten Kriterien. Die Zusammenstellung erfolgt im Rahmen der Vorrecherche zur jeweiligen Fragestellung insbesondere auf Grundlage der Daten kontrollierter Interventionsstudien, in denen zielgerichtet der Nutzen der Intervention untersucht wurde, sowie ggf. auf Basis vorliegender epidemiologischer Studien (zum Beispiel Kohorten- oder Fall-Kontroll-Studien), von Pharmakovigilanzdaten, Informationen von Zulassungsbehörden etc. Im Einzelfall können hier auch Ergebnisse aus Tierexperimenten sowie aus Experimenten zur Überprüfung eines pathophysiologischen Konstrukts hilfreich sein.
- Die oben beschriebene Zusammenstellung potenziell relevanter unerwünschter Wirkungen bildet die Grundlage für die Bewertung des potenziellen Schadens auf Basis der in die Nutzenbewertung eingehenden Studien.

3.1.4 Zusammenfassende Bewertung

Die Nutzenbewertung und die Einschätzung der Stärke der Ergebnis(un)sicherheit (siehe Abschnitt 3.2.1) orientieren sich an internationalen Standards der evidenzbasierten Medizin, wie sie z. B. von der GRADE-Gruppe erarbeitet werden [22,166,328].

Medizinische Maßnahmen werden im Vergleich zu einer anderen klar definierten aktiven Maßnahme oder Scheinmaßnahme (z. B. Placebo) oder keiner Maßnahme bezüglich ihrer Auswirkungen auf definierte patientenrelevante Endpunkte in ihrem (Zusatz-)Nutzen und Schaden zusammenfassend beschrieben. Dafür wird zunächst für jeden vorher definierten patientenrelevanten Endpunkt einzeln aufgrund der Analyse vorhandener wissenschaftlicher Daten genau eine der fünf folgenden bewertenden Aussagen getroffen:

1. Der Beleg für einen (Zusatz-)Nutzen bzw. Schaden liegt vor.
2. Hinweise liegen vor, dass ein (Zusatz-)Nutzen bzw. Schaden vorhanden ist.
3. Der Beleg für das Fehlen eines (Zusatz-)Nutzens bzw. Schadens liegt vor.
4. Hinweise liegen vor, dass kein (Zusatz-)Nutzen bzw. Schaden vorhanden ist.
5. Kein Beleg für und kein Hinweis auf einen (Zusatz-)Nutzen bzw. Schaden liegen vor.

Aussage 1 wird getroffen, wenn ein (Zusatz-)Nutzen oder Schaden wissenschaftlich belegt wurde. Aussage 2 wird getroffen, wenn Hinweise vorliegen, diese aber nicht die Voraussetzungen für einen wissenschaftlichen Beleg erfüllen, dass ein (Zusatz-)Nutzen oder Schaden vorliegt. Aussage 3 wird getroffen, wenn wissenschaftlich belegt wurde, dass diese Maßnahme keinen (Zusatz-)Nutzen hat bzw. mit keinem Schaden verbunden ist. Aussage 4 wird getroffen, wenn Hinweise vorliegen, diese aber nicht die Voraussetzungen für einen wissenschaftlichen Beleg erfüllen, dass kein (Zusatz-)Nutzen oder Schaden vorliegt. Die Voraussetzung für Hinweise auf oder Belege für das Nichtvorliegen eines (Zusatz-)Nutzens bzw. Schadens sind gut begründete Definitionen von Irrelevanzbereichen (siehe Abschnitt 6.4.5). Aussage 5 wird getroffen, wenn kein Beleg für und kein Hinweis auf einen (Zusatz-)Nutzen oder Schaden – zum Beispiel wegen unzureichender oder widersprüchlicher Datenlage – vorliegen.

In der Regel wird an die Aussage eines „Belegs“ die Anforderung zu stellen sein, dass eine Meta-Analyse von Studien mit endpunktbezogen geringer Ergebnisunsicherheit einen entsprechenden statistisch signifikanten Effekt zeigt. Falls eine Meta-Analyse nicht durchführbar ist, sollten mindestens zwei voneinander unabhängig durchgeführte Studien mit endpunktbezogen geringer Ergebnisunsicherheit und entsprechendem statistisch signifikantem Effekt vorliegen, deren Ergebnis nicht durch weitere vergleichbare endpunktbezogen ergebnissichere Studien infrage gestellt wird (Konsistenz der Ergebnisse). Bei den zwei voneinander unabhängig durchgeführten Studien muss es sich nicht um solche mit exakt identischem Design handeln. Welche

Abweichungen im Design zwischen Studien noch akzeptabel sind, hängt von der Fragestellung ab. Eine Meta-Analyse von Studien mit endpunktbezogen hoher Ergebnisunsicherheit oder Einzelergebnisse kann bzw. können trotz statistisch signifikanter Effekte demnach in der Regel allenfalls einen Hinweis liefern. Soll aus lediglich einer Studie im Ausnahmefall ein Nutzenbeleg abgeleitet werden, so sind an eine solche Studie und deren Ergebnisse besondere Anforderungen zu stellen [76].

Diese für jeden patientenrelevanten Endpunkt einzeln getroffenen Aussagen werden anschließend – soweit möglich – in einem bewertenden Fazit in Form einer Nutzen-Schaden-Abwägung zusammengefasst. Beim Vorhandensein von Belegen eines (Zusatz-)Nutzens und/oder eines Schadens bzgl. der Zielgrößen 1 bis 3 aus Abschnitt 3.1.1 stellt das Institut

1. ein Nutzenpotenzial,
2. ein Schadenpotenzial und
3. eine Nutzen-Schaden-Abwägung dar,

soweit dies aufgrund der vorliegenden Daten möglich ist. Hierbei entscheidet sich das Institut, nach dem Prinzip der Risikovorsorge vorgehend, im Zweifel und abhängig vom jeweiligen Kontext für die Annahme eines Schadenpotenzials. Darüber hinaus werden alters-, geschlechts- und lebenslagenspezifische Besonderheiten berücksichtigt.

Die genaue Durchführung der Nutzen-Schaden-Abwägung ist themenspezifisch und sollte – wenn dies prospektiv möglich ist – im Berichtsplan und ansonsten im Vorbericht beschrieben werden. Über eine Nutzen-Schaden-Abwägung im Sinne einer Gegenüberstellung von Nutzen- und Schadenpotenzial hinaus besteht die Möglichkeit, Nutzen und Schaden vergleichend zu gewichten. In diesem Fall würden die Aussagen des IQWiG für jeden einzelnen patientenrelevanten Endpunkt durch die Gewichtung von Nutzen und Schaden z. B. unter Verwendung eines Summenscores berichtet. Die Gewichtung von Nutzen und Schaden ist ebenfalls indikationsspezifisch und sollte prospektiv zum Zeitpunkt der Auswahl der zu untersuchenden Endpunkte erfolgen.

3.2 Spezielle Aspekte der Nutzenbewertung

3.2.1 Ergebnis(un)sicherheit

Grundsätzlich ist jedes Ergebnis einer empirischen Studie oder einer systematischen Übersicht über empirische Studien unsicher. Hierbei ist zu unterscheiden zwischen qualitativer und quantitativer Ergebnisunsicherheit. Die qualitative Unsicherheit wird bestimmt durch das Studiendesign, aus dem sich Evidenzgrade ableiten lassen (siehe Abschnitt 6.2.2), sowie durch (endpunktbezogene) Maßnahmen zur weiteren Vermeidung oder Minimierung möglicher Verzerrungen (z. B.

verblindete Zielgrößenerhebung, Auswertung auf Basis aller eingeschlossenen Patienten, ggf. mithilfe des Einsatzes adäquater Ersetzungsmethoden für fehlende Werte, ggf. Einsatz adäquater, valider Messinstrumente), die in Abhängigkeit vom Studiendesign bewertet werden müssen (siehe Abschnitt 6.2.3). Neben der qualitativen Ergebnisunsicherheit gibt es quantitativ messbare Unsicherheiten aufgrund statistischer Gesetzmäßigkeiten, die wiederum in unmittelbarem Zusammenhang mit dem Stichprobenumfang, d. h. der Anzahl der in einer Studie untersuchten Patienten bzw. der Anzahl der in einer systematischen Übersicht enthaltenen (Primär-)Studien, sowie mit der in bzw. zwischen den Studien beobachteten Variabilität stehen. Falls die zugrunde liegenden Daten dies zulassen, lässt sich die statistische Unsicherheit als Standardfehler bzw. Konfidenzintervall von Parameterschätzungen quantifizieren und beurteilen (Präzision der Schätzung).

Die gemeinsame Würdigung von qualitativer und quantitativer Ergebnissicherheit sowie der Größe beobachteter Effekte und derer Konsistenz bildet die Grundlage für daraus abzuleitende Empfehlungen bzw. deren Graduierung. Beispielsweise kann eine Evidenz aus Studien mit niedrigerer qualitativer Ergebnissicherheit durch große geschätzte Effekte aufgewertet werden. Umgekehrt kann eine geringe quantitative Ergebnissicherheit in Gestalt von wenigen Daten (Studien) trotz möglicherweise hoher qualitativer Ergebnissicherheit zu einer Abwertung führen [328].

3.2.2 Dramatischer Effekt

Ist der Verlauf einer Erkrankung sicher oder nahezu sicher vorhersagbar und bestehen keine Behandlungsoptionen zur Beeinflussung dieses Verlaufs, so kann der Nutzen einer medizinischen Intervention auch durch die Beobachtung einer Umkehr des (quasi)deterministischen Verlaufs bei einer gut dokumentierten Serie von Patienten belegt werden. Wenn z. B. von einer Erkrankung bekannt ist, dass sie nach Diagnosestellung innerhalb kurzer Zeit in sehr vielen Fällen zum Tode führt, und in Fallserien beschrieben wird, dass nach Anwendung einer bestimmten Maßnahme die meisten Betroffenen eine längere Zeitspanne überlebt haben, so wird ein solcher dramatischer Effekt ausreichend für einen Nutzenbeleg sein können. Beispielhaft sei hier die Substitution lebenswichtiger Hormone bei Erkrankungen mit Ausfall der Produktion dieser Hormone (z. B. Gabe von Insulin bei Patienten mit Diabetes mellitus Typ 1) genannt. Wesentliche Voraussetzung für die Einordnung als dramatischer Effekt ist die ausreichend sichere Dokumentation des schicksalhaften Verlaufs der Erkrankung in der Literatur und der Diagnosestellung bei den Patienten in der Fallserie. Dabei sollten auch mögliche Schäden der Maßnahme berücksichtigt werden. Kürzlich wurde von Glasziou et al. [147] der Versuch unternommen, die Einordnung als dramatischer Effekt zu operationalisieren. In einer ersten Näherung wurde vorgeschlagen, einen beobachteten Effekt dann als nicht mehr allein durch die Einwirkung von Störgrößen erklärbar

anzusehen, wenn er auf dem Niveau von 1 % signifikant ist und als relatives Risiko ausgedrückt den Wert 10 übersteigt [147].

Falls im Vorfeld der Bearbeitung einer spezifischen Fragestellung (z. B. aufgrund einer Vorabrecherche) ausreichende Informationen dazu vorliegen, dass ein durch die zu bewertende Intervention bewirkter dramatischer Effekt erwartet werden kann, wird bei der Informationsbeschaffung auch nach solchen Studien gesucht, die aufgrund ihres Designs eine höhere Ergebnisunsicherheit aufweisen.

3.2.3 Studiendauer

Ein wesentliches Kriterium bei der Auswahl der für die Nutzenbewertung relevanten Studien ist deren Dauer. Für die Evaluation einer Intervention zur Behandlung einer akuten Erkrankung, deren primäres Ziel z. B. die Verkürzung der Krankheitsdauer und die Reduktion der Akutsymptomatik in den Vordergrund stellt, ist die Forderung nach Langzeitstudien in der Regel nicht sinnvoll, es sei denn, es sind Spätkomplikationen zu erwarten. Auf der anderen Seite sind für die Evaluation von Interventionen zur Behandlung chronischer Erkrankungen Kurzzeitstudien in der Regel nicht für die vollständige Nutzenbewertung geeignet. Dies trifft insbesondere dann zu, wenn die Behandlung über mehrere Jahre oder sogar lebenslang erforderlich ist. In solchen Fällen sind insbesondere Studien, die einen mehrjährigen Behandlungszeitraum abdecken, sinnvoll und wünschenswert. Da sowohl Nutzenaspekte als auch das Schadenpotenzial über die Zeit unterschiedlich verteilt sein kann bzw. können, ist bei Langzeittherapien eine aussagekräftige Gegenüberstellung von Nutzen und Schaden nur bei Vorliegen von Studien mit einer entsprechend langen Beobachtungsdauer hinreichend sicher möglich. Einzelne Aspekte des Nutzens/Schadens können aber durchaus in kürzeren Studien betrachtet werden.

Bezüglich des Auswahlkriteriums „Mindeststudiendauer“ orientiert sich das Institut primär an Standards bzgl. des Nachweises der Wirksamkeit. Für die Arzneimittelbewertung wird insbesondere auf Angaben in den entsprechenden indikationsspezifischen Leitlinien der Zulassungsbehörden zurückgegriffen (z. B. [119]). Da die Nutzenbewertung auch die Aspekte des Schadenpotenzials beinhaltet, sind bei der Festlegung der Mindeststudiendauer auch die diesbezüglich allgemein konsentierten Anforderungen relevant. Bei Langzeitinterventionen, wie oben beschrieben, wird darüber hinaus auch auf das Kriterium „Langzeitbehandlung“ aus entsprechenden Guidelines zurückgegriffen [201]. Im Einzelfall kann von diesem Vorgehen begründet abgewichen werden, z. B. falls es aus inhaltlichen Gesichtspunkten erforderlich ist, eine Nachbeobachtung über einen längeren Zeitraum zu fordern, spezielle (Unter-)Fragestellungen sich auf einen kürzeren Zeitraum beziehen oder wenn bei der Evaluation neu verfügbarer/zugelassener

Interventionen und/oder der Evaluation von Technologien ohne adäquate Behandlungsalternative auch kurzfristige Wirkungen Gegenstand der Evaluation sind.

3.2.4 Patientenberichtete Endpunkte

Für die unter Abschnitt 3.1.1 beschriebenen patientenrelevanten Nutzendimensionen können auch patientenberichtete Endpunkte (patient reported outcomes, PRO) zum Einsatz kommen. Neben der gesundheitsbezogenen Lebensqualität und der Behandlungszufriedenheit können auch andere Nutzendimensionen mittels PRO erfasst werden, wie z. B. Symptome der Erkrankung. Wie auch für die Erfassung der Lebensqualität und der Behandlungszufriedenheit sind hierfür Instrumente zu fordern, die für den Einsatz in klinischen Studien geeignet sind [122,365]. Bei der Auswahl der zu berücksichtigenden Evidenz (insbesondere Studientypen) für den Nachweis eines Effekts gelten i. d. R. die gleichen Prinzipien wie bei anderen Endpunkten [365]. Das heißt, dass auch im Falle von PRO einschließlich gesundheitsbezogener Lebensqualität und Behandlungszufriedenheit randomisierte kontrollierte Studien am besten für den Nachweis eines Effekts geeignet sind.

Da Angaben zu PRO aufgrund ihrer Natur subjektiv sind, sind offene Studien in diesem Bereich nur von eingeschränkter Validität. Für die Frage, ob sich aus offenen Studien ein Hinweis für einen Nutzen einer Intervention bezüglich PRO ableiten lässt, ist die Größe des beobachteten Effekts ein wichtiges Entscheidungskriterium. Es gibt empirische Evidenz für ein hohes Verzerrungspotenzial bezüglich subjektiver Endpunkte in offenen Studien [386]. Dies ist bei der Interpretation solcher Studien zu beachten (siehe auch Abschnitt 6.2.3). Allerdings sind Situationen denkbar, wo eine Verblindung von Ärzten und Patienten nicht möglich ist. In solchen Situationen sind – sofern möglich – andere Anstrengungen zur Verzerrungsminimierung bzw. -einschätzung (z. B. verblindete Endpunkterhebung und -bewertung) zu fordern. Weitere Aspekte der Qualitätsbewertung von Studien, in denen PRO erhoben wurden, finden sich in [365].

3.2.5 Nutzen in kleinen Populationen

Es gibt kein überzeugendes Argument, bei kleinen Populationen (z. B. Patienten mit seltenen Erkrankungen oder Untergruppen häufiger Erkrankungen) grundsätzlich von der Hierarchie der Evidenzgrade abzuweichen. Auch Patienten mit sehr seltenen Erkrankungen haben einen Anspruch auf eine möglichst sichere Information über Behandlungsoptionen [117]. Nicht randomisierte Studien benötigen aufgrund der Notwendigkeit der Adjustierung nach Störgrößen größere Patientenzahlen als randomisierte Studien. Allerdings kann es aufgrund der Seltenheit einer Erkrankung manchmal unmöglich sein, so viele Patienten in eine Studie einzuschließen, dass diese eine ausreichende statistische Macht besitzt. Eine meta-analytische Zusammenfassung kleinerer

Studien kann in solchen Fällen besonders sinnvoll sein. Im Allgemeinen führen kleinere Stichproben zu einer geringeren Präzision einer Effektschätzung und damit einhergehend zu breiteren Konfidenzintervallen. Aufgrund der Bedeutung des vermuteten Effektes einer Intervention, seiner Größe, der Verfügbarkeit von Alternativen und der Häufigkeit und Schwere potenzieller therapiebedingter Schäden kann es durchaus sinnvoll sein, bei kleinen Stichproben, ähnlich wie für andere problematische Konstellationen empfohlen [121], größere p -Werte als 5 % (z. B. 10 %) für den Nachweis statistischer Signifikanz zu akzeptieren und somit die quantitative Unsicherheit zu erhöhen. Dies muss aber ex ante geschehen und nachvollziehbar begründet werden. Ebenso kann bei kleineren Stichproben eher die Notwendigkeit bestehen, einen zu selten auftretenden patientenrelevanten Endpunkt durch Surrogate zu ersetzen. Solche Surrogatparameter müssen allerdings auch bei kleinen Stichproben valide sein [120].

Für den Fall extrem seltener Erkrankungen oder extrem spezifischer Krankheitskonstellationen kann die Forderung nach (parallel) vergleichenden Studien unangemessen sein [383]. Nichtsdestotrotz sind auch hier zumindest adäquate Dokumentationen des Verlaufs solcher Patienten einschließlich der Erwartung an den Verlauf ohne die ggf. zu prüfende Intervention (z. B. anhand historischer Patientenverläufe) möglich und bewertbar [62]. Die Spezifizierung und explizite Kennzeichnung, dass es sich um eine solche Situation einer extrem seltenen Erkrankung oder extrem spezifischen Krankheitskonstellation handelt, erfolgen im Berichtsplan.

3.2.6 Nutzen im Einzelfall

Ziel der Nutzenbewertung ist es, anhand der Ergebnisse von Studien (mit einer mehr oder weniger großen Zahl von Patienten), die geeignet sind, kausal begründete Effekte nachzuweisen, belastbare Voraussagen für zukünftige Patienten zu treffen. Auf Einzelfällen basierende Erfahrungen können – abgesehen von Ausnahmen (z. B. dramatischer Effekt) – nicht auf zukünftige (andere) Patienten übertragen werden, da hier der kausale Nachweis nicht möglich ist.

Unter bestimmten Voraussetzungen sind allerdings Versuche im Einzelfall möglich, die Voraussagen für genau diesen Einzelfall (aber eben nicht für andere, zukünftige Patienten) zulassen, die auf kausal begründeten Effekten beruhen. Versuche im Einzelfall sind in der praktischen Medizin häufig, werden aber meist unsystematisch durchgeführt und sind dadurch sowie aufgrund der meist ungenügenden Kontrolle der Störgrößen nicht selten nur unzuverlässig interpretierbar [324].

An Versuchen im Einzelfall – den so genannten „n-of-1-Studien“ – nimmt nur ein einzelner Patient teil, wobei die Zielgrößen vollständig und relativ schnell reversibel sein müssen [167,170,219,324]. Reversibilität erfordert deshalb das Vorliegen von chronischen Erkrankungen oder Symptomen, um

eine n-of-1-Studie adäquat durchzuführen. Eine solche Studie erlaubt, wenn sie möglich und adäquat durchgeführt ist, eine Aussage darüber, ob ein einzelner Patient von einer bestimmten Behandlung profitiert. Derartige n-of-1-Studien bestehen aus mehreren Versuchs- und Kontrollstudienperioden, die in zufälliger Reihenfolge bei einem Patienten angewendet werden [170,324]. Bei n-of-1-Studien werden also nicht die Patienten, sondern die Behandlungsperioden randomisiert. Die Intervention sollte verblindet und entweder aktiv- oder placebokontrolliert sein. Leider ist aber in Einzelversuchen die methodische Qualität häufig unzureichend, sodass die Ergebnisse häufig mit Zurückhaltung interpretiert werden müssen. Die Verwendung eines nicht verblindeten Designs bedarf einer nachvollziehbaren Begründung.

3.3 Nutzenbewertung von Arzneimitteln

Ein wesentliches Ziel der Berichte zur Nutzenbewertung von Arzneimitteln ist die Unterstützung von Richtlinienentscheidungen des G-BA zur Erstattungsfähigkeit von Arzneimitteln innerhalb der gesetzlichen Krankenversicherung. Dazu ist es notwendig darzustellen, ob der Nutzen eines Arzneimittels nachgewiesen ist bzw. ob für ein Arzneimittel ein höherer Nutzen (Zusatznutzen) nachgewiesen ist als für eine andere medikamentöse oder nichtmedikamentöse Alternative.

Richtlinienentscheidungen des G-BA betrachten i. d. R. nicht den Einzelfall, sondern den Regelfall. Auch die Berichte des Instituts beziehen sich deshalb i. d. R. nicht auf die Einzelfallentscheidung.

Aufgrund der Zielsetzung der Nutzenbewertung durch das Institut werden in die jeweilige Bewertung nur Studien einer Evidenzstufe eingeschlossen, die zum Nachweis des Nutzens grundsätzlich geeignet ist. Studien, die lediglich Hypothesen generieren können, sind deshalb im Allgemeinen für die Nutzenbewertung nicht relevant. Die Frage, ob eine Studie einen Nachweis eines Nutzens erbringen kann, hängt im Wesentlichen von der Ergebnissicherheit der erhobenen Daten ab.

3.3.1 Stellenwert des Zulassungsstatus

Die Beauftragung des Instituts zur Nutzenbewertung von Arzneimitteln durch den G-BA erfolgt i. d. R. im Rahmen der Zulassung (Indikation, Dosierung, Gegenanzeigen, Begleitbehandlungen etc.) der zu untersuchenden Arzneimittel. Aus diesem Grund bezieht sich auch die Empfehlung des Instituts an den G-BA, die im Fazit des Berichts zur Nutzenbewertung formuliert wird, i. d. R. auf den Gebrauch der bewerteten Arzneimittel innerhalb der bestehenden Zulassung.

Projektspezifisch ist zu klären, wie mit Studien (bzw. der daraus abzuleitenden Evidenz) umgegangen wird, die nicht innerhalb der durch die Zulassung beschriebenen Anwendung eines

Arzneimittels durchgeführt werden. Grundsätzlich ist denkbar, dass Studien, in denen Arzneimittel außerhalb der in den Fachinformationen beschriebenen Zulassung eingesetzt werden, den Nutzen und/oder Schaden des Arzneimittels über- oder unterschätzen. Dies kann zu einer Fehleinschätzung des Nutzens und Schadens für Patienten, die innerhalb der Zulassung behandelt werden, führen. Ist für solche Studien allerdings hinreichend sicher plausibel oder gar nachgewiesen, dass die in diesen Studien gewonnenen Ergebnisse auf Patienten, die gemäß Zulassung behandelt werden, anwendbar sind, können die Ergebnisse dieser Studien in die Nutzenbewertung einbezogen werden.

Daher wird für Studien, die allein deshalb ausgeschlossen werden, weil sie nicht den aktuellen Zulassungsstatus widerspiegeln bzw. weil die Erfüllung des Zulassungsstatus unklar ist, jeweils geprüft, inwieweit die Studienergebnisse auf Patienten anwendbar sind, die innerhalb der Zulassungsbedingungen behandelt werden.

Als „anwendbar“ sind die Ergebnisse von Studien außerhalb des Zulassungsstatus dann anzusehen, wenn hinreichend sicher plausibel oder nachgewiesen ist, dass die Effektschätzer patienten-relevanter Endpunkte nicht wesentlich durch das betreffende Merkmal der Zulassung (z. B. geforderte Vorbehandlung) beeinflusst werden. Die Gleichwertigkeit der Effekte ist i. d. R. mit adäquaten wissenschaftlichen Untersuchungen zu belegen. Diese Untersuchungen sollen auf den Nachweis der Gleichwertigkeit des Effekts zwischen der Gruppe mit und derjenigen ohne Merkmal ausgerichtet sein. Ergebnisse, die auf Patienten, die gemäß der Zulassung behandelt werden, anwendbar sind, können in das Fazit der Bewertung einbezogen werden.

Als „nicht anwendbar“ sind die Ergebnisse aus Studien anzusehen, wenn kein Nachweis der Anwendbarkeit vorliegt und wenn plausible Gründe gegen die Übertragbarkeit der Ergebnisse sprechen. Als in der Regel nicht anwendbar werden Studienergebnisse bspw. dann angesehen, wenn die behandelte Altersgruppe außerhalb der Zulassung liegt, nicht zugelassene Kombinationen mit einem anderen Wirkstoff eingesetzt werden, Patienten mit einem außerhalb der Zulassung liegenden Schweregrad der Erkrankung behandelt werden oder Studien bei Patienten mit einer Kontraindikation gegen die untersuchte Intervention durchgeführt werden. Die Ergebnisse dieser Studien werden im Bericht nicht dargestellt, da sie zur Bewertung des Arzneimittels nicht herangezogen werden können.

Werden Ergebnisse aus Studien, in denen Patienten außerhalb der Zulassung behandelt wurden, als anwendbar angesehen, so wird dies im Berichtsplan spezifiziert. Unabhängig von der Anwendbarkeit von Studienergebnissen auf den durch die Zulassung spezifizierten Gebrauch werden in der Regel die Ergebnisse von Studien, die die Behandlung von Patienten mit dem zu bewertenden Arzneimittel bei der im Auftrag spezifizierten Erkrankung untersuchen und die z. B. aufgrund der Größe, der Dauer oder der untersuchten Endpunkte von besonderer Bedeutung sind, diskutiert.

3.3.2 Studien zur Nutzenbewertung von Arzneimitteln

Die Ergebnisse der Nutzenbewertung von Arzneimitteln durch das Institut können Einfluss auf die Versorgung von Patienten in Deutschland haben. Aus diesem Grund müssen hohe Anforderungen an die Ergebnissicherheit der Studien, die in die Nutzenbewertung einfließen, gestellt werden.

Die Ergebnissicherheit ist definiert als die Sicherheit, mit der sich ein Effekt oder das Fehlen eines Effekts aus einer Studie ableiten lässt. Dies gilt sowohl für „positive“ Aspekte (Nutzen) als auch für „negative“ Aspekte (Schaden). Die Ergebnissicherheit einer einzelnen Studie wird im Wesentlichen durch drei Komponenten beeinflusst:

- das Studiendesign
- die designspezifische und durch die konkrete Durchführung bestimmte interne Validität der Studien
- die Größe eines zu erwartenden bzw. beobachteten Effekts

Im Rahmen der Nutzenbewertung von Arzneimitteln werden nicht nur einzelne Studien betrachtet, sondern die Ergebnisse der Studien werden in eine systematische Übersicht integriert. Die Ergebnissicherheit einer systematischen Übersicht basiert wiederum auf der Ergebnissicherheit der eingeschlossenen Studien. Darüber hinaus wird sie insbesondere durch folgenden Faktor bestimmt:

- die Konsistenz der Ergebnisse mehrerer Studien

Das Studiendesign hat insofern erheblichen Einfluss auf die Ergebnissicherheit, als dass mit Beobachtungsstudien, prospektiv oder retrospektiv, ein kausaler Zusammenhang zwischen Intervention und Effekt in der Regel nicht dargestellt werden kann, während die kontrollierte Interventionsstudie grundsätzlich hierfür geeignet ist [160]. Dies gilt vor allem dann, wenn andere, das Ergebnis beeinflussende Faktoren ganz oder weitgehend ausgeschaltet werden. Aus diesem Grund stellt die randomisierte kontrollierte Studie den Goldstandard bei der Bewertung medikamentöser und nichtmedikamentöser Interventionen dar [241].

In der Regel sind RCTs zur Evaluation von Arzneimitteln möglich und praktisch durchführbar. Das Institut lässt daher in der Regel RCTs in die Nutzenbewertung von Arzneimitteln einfließen. Nur in begründeten Ausnahmefällen wird auf nicht randomisierte Interventionsstudien oder Beobachtungsstudien zurückgegriffen. Die Begründung kann einerseits in der Nichtdurchführbarkeit eines RCT liegen, z. B. bei starker Präferenz einer bestimmten Therapiealternative von Therapeuten und/oder Patienten, andererseits darin, dass für die vorliegende Fragestellung auch andere Studienarten eine hinreichende Ergebnissicherheit liefern könnten. Bei Erkrankungen, die ohne Intervention nach kurzer Zeit sicher zum Tod führen, können beispielsweise mehrere konsistente Fallbeobachtungen darüber, dass eine bestimmte Intervention diesen zwangsläufigen Verlauf

verhindert, hinreichende Ergebnissicherheit liefern [249] (dramatischer Effekt, siehe auch Abschnitt 3.2.2). Die besondere Begründungspflicht für ein nicht randomisiertes Design bei der Prüfung von Arzneimitteln findet sich auch im Rahmen des Arzneimittelzulassungsrechts in den Arzneimittelprüfrichtlinien (Richtlinie 2001/83/EG, Abschnitt 5.2.5 [235]).

Das Institut legt im Rahmen der Erstellung des Berichtsplans (siehe auch Abschnitt 2.1.1) vorab fest, welche Studienarten aufgrund der Fragestellung als durchführbar und hinreichend ergebnissicher (bei hoher interner Validität) angesehen werden. Studien, die nicht mindestens diesem Qualitätsanspruch entsprechen (siehe auch Abschnitt 6.2.3), werden in den Bewertungsprozess primär nicht einfließen.

Ausführungen zur Bewertung der internen Validität von Studien sowie zur Konsistenz von Ergebnissen mehrerer Studien und zum Stellenwert der Größe des zu erwartenden Effekts als weiteren Einflussgrößen der Ergebnissicherheit finden sich in den Abschnitten 3.2.1, 3.2.2 und 6.2.

Neben der Charakterisierung der Ergebnissicherheit der berücksichtigten Studien ist es notwendig zu beschreiben, ob und in welchem Umfang sich das Studienergebnis auf lokale Gegebenheiten (zum Beispiel Population, Versorgungsbereich etc.) übertragen lässt beziehungsweise welche lokalen Besonderheiten bei den Studien eine Auswirkung auf die Ergebnisse selbst beziehungsweise deren Interpretation haben oder haben könnten. Unter diesem Gesichtspunkt sind insbesondere Studien relevant, in denen die tatsächliche Versorgungsrealität Deutschlands so weit wie möglich abgebildet wird, ohne dass jedoch die oben beschriebenen Kriterien der Ergebnissicherheit außer Acht gelassen werden dürfen. Die Übertragbarkeit des Studienergebnisses (Generalisierbarkeit bzw. sogenannte externe Validität) selbst muss schließlich in einem gesonderten Prozess überprüft werden, der vom Design und der Qualität der Studie zunächst unabhängig ist.

3.4 Nichtmedikamentöse therapeutische Interventionen

Wenngleich die regulatorischen Voraussetzungen für den Marktzugang von Arzneimitteln und nichtmedikamentösen therapeutischen Interventionen unterschiedlich sind, gibt es dennoch zunächst keinen Grund, an die Bewertung von Nutzen und Schaden einen prinzipiell anderen auf die Ergebnissicherheit bezogenen Maßstab anzulegen. So sieht zum Beispiel die Verfahrensordnung des G-BA die prioritäre Berücksichtigung von RCTs, soweit möglich, unabhängig vom Typ (medikamentös/nichtmedikamentös) der zu bewertenden medizinischen Maßnahme vor (§ 20 [145]). Für Medizinprodukte wird abschwächend im Rahmen der Konformitätsbewertung in der EN-ISO-Norm 14155-2 (Abschnitt 4.7 [92]) zwar nicht die randomisierte

kontrollierte Studie als Design der Wahl dargestellt, allerdings muss die Wahl des Designs überhaupt begründet werden.

Studien im nichtmedikamentösen Bereich sind im Vergleich zu Arzneimittelstudien häufig mit besonderen Herausforderungen und Schwierigkeiten verbunden [261]. Beispielsweise wird oft die Verblindung des die Intervention ausführenden Personals unmöglich und die der Patienten nur schwierig oder ebenfalls nicht zu bewerkstelligen sein. Zudem kann davon ausgegangen werden, dass Präferenzen für bestimmte Therapieoptionen von Therapeuten und Patienten in besonderer Weise die Durchführbarkeit von Studien in diesem Bereich erschweren. Weiterhin kann es insbesondere bei der Evaluierung von komplexen Interventionen notwendig sein, die Möglichkeit von Kontaminationseffekten zu beachten. Auch die Trennung der Effekte durch die zu bewertende Prozedur bzw. das zu bewertende (Medizin-)Produkt auf der einen und der Fähigkeiten und Fertigkeiten des Anwenders bzw. der Anwender auf der anderen Seite ist ggf. zu beachten sowie – im Zusammenhang mit dem Bewertungszeitpunkt – Lerneffekte.

Um den genannten Aspekten gerecht zu werden, bedarf es qualitativ besonders guter Studien, um eine ausreichende Ergebnissicherheit zu erzielen. Paradoxerweise war in der Vergangenheit eher das Gegenteil der Fall, das heißt gerade im nichtmedikamentösen Bereich fehlen oft aussagekräftige randomisierte Studien (z. B. in der Chirurgie [261]). Um überhaupt Aussagen zum Stellenwert einer bestimmten nichtmedikamentösen therapeutischen Intervention treffen zu können, kann es deshalb erforderlich sein, auch nicht randomisierte Studien in die Bewertung einzubeziehen. Auch an diese sind allerdings Qualitätsanforderungen zu stellen, insbesondere bzgl. der Maßnahmen zur Sicherung der Strukturgleichheit. In aller Regel werden solche Studien jedoch aufgrund der ihnen immanenten höheren Ergebnisunsicherheit allenfalls Hinweise auf einen (Zusatz-)Nutzen oder Schaden liefern können. Die Einbeziehung von Studien eines niedrigeren Evidenzgrades steht im Einklang mit der entsprechenden Regelung in § 20 der Verfahrensordnung des G-BA [145], wo aber dafür eine besondere Begründungspflicht betont wird. Dort heißt es: „Die Anerkennung des medizinischen Nutzens einer Methode auf Grundlage von Unterlagen einer niedrigeren Evidenzstufe bedarf jedoch – auch unter Berücksichtigung der jeweiligen medizinischen Notwendigkeit – zum Schutz der Patienten umso mehr einer Begründung, je weiter von der Evidenzstufe 1 abgewichen wird. Dafür ist der potenzielle Nutzen einer Methode insbesondere gegen die Risiken der Anwendung beim Patienten abzuwägen, die mit einem Wirksamkeitsnachweis geringerer Aussagekraft einhergehen“ [145]. Das heißt, allein die Nichtverfügbarkeit von Studien höchster Evidenz kann im Allgemeinen nicht als ausreichende Begründung für eine Nutzenbewertung auf Basis von Studien niedrigerer Evidenz angesehen werden.

Auch bei der Bewertung von nichtmedikamentösen therapeutischen Interventionen kann es notwendig sein, den Zulassungsstatus von zu prüfenden (Medizin-)Produkten, von Komponenten von zu prüfenden komplexen Interventionen oder von Vergleichsinterventionen zu beachten (siehe Abschnitt 3.3.1). Entsprechende Konsequenzen sind dann im Berichtsplan (siehe Abschnitt 2.1.1) zu spezifizieren.

3.5 Diagnostische Verfahren

Allgemein kann der Prozess zur Evaluierung diagnostischer Verfahren in Analogie zur Arzneimittelprüfung in verschiedene, hierarchisch angeordnete Phasen oder Stufen eingeteilt werden [141,232]. Prospektiv geplante, kontrollierte diagnostische Studien der Phase 4 nach Köbberling et al. [232] bzw. Studien der Stufe 5 nach Fryback und Thornbury [141] mit einer (idealerweise zufälligen) Zuteilung von Patienten in eine Strategie mit bzw. ohne Anwendung der zu prüfenden diagnostischen Maßnahme oder in Gruppen mit bzw. ohne Offenlegung der (diagnostischen) Testergebnisse können dabei in Entsprechung zu (Arzneimittel-)Zulassungsstudien der Phase 3 („Wirksamkeitsstudien“) gesehen werden. Dementsprechend wird ihnen auch der höchste Evidenzgrad zugesprochen (siehe z. B. Verfahrensordnung des G-BA [145]). Auch für die Zulassung für bestimmte Indikationen von Arzneimitteln und biologischen Produkten, die im Zusammenhang mit bildgebenden diagnostischen Methoden entwickelt werden, werden von der FDA derartige Studien empfohlen [136] und Beispiele zeigen, dass sie – je nach erwartetem Effekt – auch mit vergleichsweise moderatem Aufwand durchführbar sind [368].

Das Institut folgt dieser Logik und führt Nutzenbewertungen zu diagnostischen Verfahren in erster Linie auf der Grundlage solcher Studien mit patientenrelevanten Endpunkten durch. Die Grundzüge der Bewertung entsprechen dabei den in den Abschnitten 3.1 bis 3.4 gemachten Ausführungen. Unter patientenrelevanten Endpunkten sind in diesem Zusammenhang dieselben Nutzenkategorien zu verstehen wie bei der Bewertung therapeutischer Interventionen, nämlich Mortalität, Morbidität und gesundheitsbezogene Lebensqualität. Deren Beeinflussung durch diagnostische Maßnahmen lässt sich dabei durch die Vermeidung risikobehafteter(er) bzw. komplikationsträchtiger(er) Interventionen oder durch den gezielt(er)en Einsatz von Interventionen erzielen. Allgemein ausgedrückt entfaltet sich der patientenrelevante Nutzen einer diagnostischen Maßnahme dadurch, klinische und/oder persönliche Entscheidungen besser treffen zu können.

Ebenfalls von hoher Evidenz und prioritär für die Nutzenbewertung diagnostischer Verfahren heranzuziehen sind solche Studien, bei denen die Wechselwirkung zwischen diagnostischer Information und therapeutischem Nutzen geprüft wird [319].

Liegen solche Studien nicht oder in nicht ausreichender Quantität und Qualität vor, kann eine Bewertung der sogenannten diagnostischen Kette erfolgen [262]. Dabei wird die Güte des diagnostischen Verfahrens anhand allgemein verwendeter Testgütekriterien, die in ausreichend ergebnissicheren Studien ermittelt wurden (i. d. R. Phase 3 nach Köbberling et al. [232]), bewertet und geprüft, inwieweit belegt ist, dass die aus den Testergebnissen resultierenden Konsequenzen mit einem Nutzen verbunden sind. Für den (zumeist anzunehmenden) Fall therapeutischer Konsequenzen lassen sich solche Belege aus randomisierten Interventionsstudien (mit patientenrelevanten Endpunkten) ableiten, in denen ein bestimmtes (Test-)Ergebnis des zu prüfenden diagnostischen Verfahrens als Einschlusskriterium definiert wurde.

Bei der Bewertung der Ergebnissicherheit von Studien zur diagnostischen Güte orientiert sich das Institut in erster Linie an den QUADAS-Kriterien [380], die allerdings ggf. projektspezifisch angepasst werden. Die Kriterien des STARD-Statements [50,51] werden herangezogen, um im Einzelfall über den Ein- oder Ausschluss nicht im Volltext publizierter Studien zu entscheiden.

Studien der Stufe 3 und 4 nach Fryback und Thornbury [141] sollen den Einfluss der zu prüfenden (diagnostischen) Verfahren auf (differenzial)diagnostische Überlegungen und/oder auf nachfolgende therapeutische (oder andere Management-)Entscheidungen untersuchen, d. h. es wird untersucht, ob das Ergebnis des diagnostischen Tests überhaupt zu einer Veränderung von Entscheidungen führt. Solche Studien bzw. Studienkonzepte haben aber den wesentlichen Nachteil, dass sie nicht scharf definiert sind, und sind daher eher theoretischer Natur. Wesentliches (Qualitäts-)Kennzeichen ist, dass erkennbar geplant die beteiligten Ärzte vor der Durchführung oder dem Bekanntwerden der Ergebnisse des zu prüfenden diagnostischen Verfahrens im Hinblick auf die Wahrscheinlichkeit des Vorliegens der Erkrankung und auf ihr weiteres diagnostisches und/oder therapeutisches Vorgehen befragt werden, um die durch das Ergebnis des zu prüfenden diagnostischen Verfahrens veränderte Einstellung ermitteln zu können. Retrospektive Ein- und theoretische Abschätzungen sind dagegen verzerrungsanfällig [141,171]. Der Stellenwert derartiger, letztlich unkontrollierter Studien im Rahmen von Nutzenbewertungen diagnostischer Verfahren muss als weitgehend unklar angesehen werden.

Für Modifikationen bereits existierender diagnostischer Verfahren, für deren Anwendung ein patientenrelevanter Nutzen gesichert ist oder hinreichend plausibel postuliert werden kann, wird es nicht immer notwendig sein, erneut die gesamte diagnostische Kette zu untersuchen. In solchen Fällen kann es z. B. ausreichend sein, lediglich eine verbesserte oder gleichwertige Intratest-Variabilität zu überprüfen. Für einen Vergleich zweier oder mehrerer diagnostischer Tests hinsichtlich bestimmter Testcharakteristika stellen wiederum Studien mit zufälliger Zuordnung der Reihenfolge der (voneinander unabhängigen und möglichst verblindeten) Testdurchführung bei

denselben Patienten oder mit zufälliger Zuordnung der Tests auf verschiedene Patienten diejenigen mit der höchsten Ergebnissicherheit dar und finden demnach in erster Linie in Berichten des Instituts Berücksichtigung.

Es ist weiterhin vorstellbar, dass ein neues diagnostisches Verfahren in eine bereits existierende diagnostische Strategie eingebettet werden soll, beispielsweise dergestalt, dass ein neuer Test einem etablierten Test vor- (Triage-Test) oder nachgeschaltet (Add-on-Test) werden soll, um den Einsatz des jeweils anderen zu reduzieren [49]. Dabei muss allerdings vor dem Hintergrund der nachfolgenden therapeutischen (oder andersartigen) Konsequenzen bedacht werden, dass sich durch eine solche Kombination von Tests die Patientenpopulationen, die aus den jeweilig kombinierten Testergebnissen resultieren, von denjenigen unterscheiden, die aus den Einzeltestergebnissen resultieren. Dieser Unterschied könnte wiederum Einfluss auf nachfolgende therapeutische (oder andersartige) Konsequenzen und deren Effektivität haben. Ist ein solcher Einfluss nicht hinreichend sicher auszuschließen, sind ggf. vergleichende Studien zu den diagnostischen Strategien mit und ohne neues Verfahren erforderlich [136,253].

Auch bei der Bewertung von Diagnostika kann es notwendig sein, den Zulassungsstatus zu beachten (siehe Abschnitt 3.3.1). Entsprechende Konsequenzen sind dann im Berichtsplan (siehe Abschnitt 2.1.1) zu spezifizieren.

3.6 Bevölkerungsweite Interventionen

3.6.1 Screening

Screeningprogramme setzen sich aus unterschiedlichen „Bausteinen“ zusammen, die entweder als Ganzes oder in Teilen untersucht werden (können). Bei der Bewertung erfolgt eine Orientierung an international akzeptierten Standards bzw. Kriterien, wie zum Beispiel denjenigen des UK National Screening Committee (UK NSC [362]), der US Preventive Services Task Force (US PSTF [177]) oder des National Health Committee (NHC) aus Neuseeland [274].

Gemäß den o. g. Kriterien bewertet das Institut den Nutzen von Screeningmaßnahmen in erster Linie anhand von prospektiv geplanten vergleichenden Interventionsstudien der gesamten Screeningkette mit einer (idealerweise zufälligen) Zuteilung von Personen in eine Strategie mit bzw. ohne Anwendung der Screeningmaßnahme (ggf. auch in unterschiedliche Screeningstrategien) und der Betrachtung von patientenrelevanten Endpunkten. Die Grundzüge der Bewertung entsprechen dabei den in den Abschnitten 3.1 bis 3.4 gemachten Ausführungen.

Liegen solche Studien nicht oder in nicht ausreichender Quantität und Qualität vor, kann eine Bewertung der einzelnen Bausteine der Screeningkette erfolgen. Dabei wird die Güte des

Screeningtests anhand allgemein verwendeter Testgütekriterien, die in ausreichend ergebnssicheren Studien ermittelt wurden (i. d. R. Phase 3 nach Köbberling et al. [232]), bewertet (siehe Abschnitt 3.5) und geprüft, inwieweit belegt ist, dass die aus den Testergebnissen resultierenden Konsequenzen mit einem Nutzen verbunden sind. Für den (zumeist anzunehmenden) Fall therapeutischer Konsequenzen lassen sich solche Belege aus randomisierten Interventionsstudien ableiten, in denen eine frühzeitige(re) mit einer späte(re)n Intervention verglichen wurde. Ggf. kann der Nutzen einer frühzeitige(re)n gegenüber einer späte(re)n Intervention auch anhand von Interventionsstudien geprüft werden, in denen sich eine Wechselwirkung zwischen der Frühzeitigkeit des Interventionsbeginns und dem Interventionseffekt untersuchen lässt. Dies kann entweder direkt innerhalb einer Studie oder indirekt durch den Vergleich von Studien mit unterschiedlichem Interventionsbeginn, aber ansonsten vergleichbarem Design erfolgen. Auch hier entsprechen die Grundzüge der Bewertung den in den Abschnitten 3.1 bis 3.4 gemachten Ausführungen.

3.6.2 Prävention

Prävention zielt darauf ab, eine gesundheitliche Schädigung zu verhindern, weniger wahrscheinlich zu machen oder zu verzögern [376]. Während die Primärprävention alle Maßnahmen vor Eintritt einer fassbaren biologischen Schädigung zur Vermeidung auslösender oder vorhandener Teilursachen umfasst, beinhaltet die sekundäre Prävention Aktivitäten zur Entdeckung klinisch symptomloser Krankheitsfrühstadien und ihre erfolgreiche Frühtherapie (siehe auch Abschnitt 3.6.1). Maßnahmen der Primär- und Sekundärprävention zeichnen sich dadurch aus, dass im Unterschied zu kurativen Maßnahmen häufig ganze Bevölkerungsgruppen im Fokus der Intervention stehen. Tertiärprävention im engeren Sinne bezeichnet spezielle Interventionen zur Verhinderung bleibender, insbesondere sozialer Funktionseinbußen nach dem Eintreten einer Krankheit [376]. Sie ist weniger Gegenstand dieses Abschnitts, sondern findet ihre Entsprechung in den Abschnitten zur Nutzenbewertung von Arzneimitteln und nichtmedikamentösen Verfahren (siehe Abschnitte 3.3 und 3.4).

Nutzenbewertungen zu (anderen als Screening-)Präventionsprogrammen führt das Institut in erster Linie ebenfalls anhand prospektiv geplanter vergleichender Interventionsstudien mit einer (idealerweise zufälligen) Zuteilung von Personen in eine Strategie mit bzw. ohne Anwendung der Präventionsmaßnahme und der Betrachtung von patientenrelevanten Endpunkten durch. Unter anderem wegen einer möglichen Kontamination zwischen Interventions- und Kontrollgruppe kommen alternativ auch Studien, bei denen Cluster den Interventionsarmen zugeteilt werden, infrage [357].

In Einzelfällen ist zu prüfen, inwieweit die Berücksichtigung weiterer Studiendesigns sinnvoll sein kann [209]. So werden beispielsweise massenmediale Kampagnen häufig im Rahmen eines

sogenannten Interrupted Time-Series Designs evaluiert (zum Beispiel bei [370]) und die Anwendung dieses Studiendesigns wird auch für gemeindebezogene Präventionsprogramme empfohlen [43]. Das Institut orientiert sich bei der Qualitätsbewertung solcher Studien an den von der Cochrane Effective Practice and Organisation of Care Review Group entwickelten Kriterien [72].

Für den Nutzen auf Populationsebene ist neben der Effizienz eines solchen Programms auch die Höhe der Teilnahmerate entscheidend. Außerdem ist die Frage relevant, welche Personen erreicht werden; so gibt es Hinweise darauf, dass Präventionsprogramme gerade von Bevölkerungsgruppen mit erhöhtem Krankheitsrisiko weniger in Anspruch genommen werden [242]. Beide Aspekte stehen deshalb im besonderen Fokus von Bewertungen des Instituts.

4 Leitlinien und Disease-Management-Programme

4.1 Hintergrund der Leitlinienbewertung

Medizinische Leitlinien gelten als Schlüsselinstrumente zur Verbesserung und Sicherung der medizinischen Qualität in der Patientenversorgung [118]. Ihr Sinn liegt darin, durch die Formulierung von konkreten Handlungsempfehlungen unangemessene Unterschiede in der Patientenversorgung zu reduzieren und die Versorgung der Patienten zu verbessern. Darüber hinaus werden sie in Deutschland auch als Basis für Entscheidungen zu Steuerungszwecken im Gesundheitswesen wie z. B. im Zusammenhang mit der Formulierung von Anforderungen an Disease-Management-Programme (nach § 137f SGB V) genutzt und beeinflussen so auch auf der strukturellen Ebene zunehmend Entscheidungen im Gesundheitswesen.

Vor diesem Hintergrund sollen 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 [182,321] und Leitlinien zu gleichen Themen weisen teilweise erhebliche inhaltliche Unterschiede in ihren Empfehlungen auf [63,207].

Ein wesentlicher Grund hierfür ist, dass die international geforderten Qualitätsstandards für die Entwicklung von Leitlinien nicht durchgängig berücksichtigt werden [29,70,334].

4.2 Ziele der Leitlinienbewertung

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

- methodisch und inhaltlich gute von schlechten Leitlinien zu diskriminieren,
- die Evidenzgrundlage, auf der Kernempfehlungen in Leitlinien beruhen, auszuarbeiten und zu überprüfen,
- klare fachliche Aussagen zur Sinnhaftigkeit und Wirksamkeit der Umsetzung verschiedener medizinischer Empfehlungen zu geben,
- dem Gemeinsamen Bundesausschuss oder seinen Gremien eine Entscheidungsgrundlage für die Beratungen über strukturierte Behandlungsprogramme zu bieten,
- sicherzustellen, dass nur geprüfte (qualitätsgesicherte) Leitlinien in die Versorgung eingeführt werden, bei denen Hinweise auf eine Verbesserung des Outcomes bestehen,

- Forschungsbedarf zu benennen und sinnvolle Projekte zur Entwicklung und Implementierung von evidenzbasierten Empfehlungen anzustoßen sowie
- die Einbindung in umfassende Qualitätsmanagement-Prozesse (UQM) zu fördern.

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

Zur inhaltlichen Bewertung medizinischer Leitlinien sollen die vorhandene methodische Kompetenz und die inhaltliche Expertise von externen Instituten/Einrichtungen/Organisationen möglichst genutzt und einbezogen werden [132].

4.3 Methoden der Leitlinienbewertung

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

- die Untersuchung formaler Kriterien, die im Wesentlichen die Transparenz des Entstehungsprozesses abbilden und von der Voraussetzung ausgehen, dass Leitlinien, die diese Kriterien berücksichtigen, mit einer größeren Wahrscheinlichkeit zu inhaltlich richtigen Empfehlungen kommen (vergleichbar der internen Validität von Studien) [162,166],
- die genaue inhaltliche Prüfung im Hinblick auf die zugrunde liegende Evidenz,
- die Bewertung der Angemessenheit von Empfehlungen und
- die Evaluation der Effekte durch die Anwendung von Leitlinien (Outcome-Evaluation).

Die Ansätze unterscheiden sich in puncto Aufwand, Herangehensweise und Aussagewert z. T. erheblich und kommen je nach Fragestellung und Auftrag zur Anwendung. Die dabei vom Institut verwendete Methodik wird unter Berücksichtigung aktueller wissenschaftlicher Publikationen sowie nationaler und internationaler Erfahrungen regelmäßig überprüft und ggf. aktualisiert. In der Regel werden verschiedene Aspekte in einem abgestuften Verfahren miteinander kombiniert [132]. Kritisch diskutiert wurde im Vorfeld der Methodenüberarbeitung die Beschränkung auf eher formale und methodische Aspekte der Bewertung, die gut operationalisierbar sind. Diese Beschränkung ist zum einen den derzeitigen gesetzlichen Rahmenbedingungen geschuldet (§ 139a SGB V), die bisher (nur) eine Überprüfung der Leitlinien hinsichtlich der zugrunde liegenden Evidenz vorsehen. Andererseits kommen die Methoden der evidenzbasierten (und konsentierten)

Leitlinienentwicklung und deren Aktualisierung immer noch unzureichend zur Anwendung [20], weshalb auf eine formale Bewertung derzeit nicht verzichtet werden kann.

Einzelempfehlungen/Kernpunkte in Leitlinien können im Hinblick auf ihre Relevanz, Angemessenheit und Praktikabilität von den Betroffenen sehr unterschiedlich bewertet werden.

Zur Bewertung der inhaltlichen Angemessenheit von Leitlinien gibt es bisher kein international abgestimmtes Verfahren. Ein solches Verfahren muss verschiedene Aspekte berücksichtigen und auch andere Akteure wie z. B. den G-BA, Fachvertreter und Patienten mit einbeziehen.

Aussagen zur inhaltlichen Angemessenheit von Leitlinienempfehlungen beinhalten zwei Kernfragen:

1. Ist die Darstellung und Interpretation der aktuellen Evidenz in der Leitlinie, aus der konkrete Einzelempfehlungen generiert werden, angemessen und nachvollziehbar?
2. Ist die empfohlene Maßnahme angemessen?

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

- Für die Maßnahme muss ein Nutzen nachgewiesen sein und die Nutzen-Schaden-Abwägung muss zu einer positiven Entscheidung kommen (siehe Abschnitt 3.1).
- Die Maßnahme muss relevant für das deutsche Gesundheitssystem sein.
- Die Maßnahme muss verfügbar und zugelassen sein.
- Die Maßnahme muss notwendig sein.
- Eine sichere Anwendung (durch Ärzte/ Patienten) ist möglich.
- Innerhalb der zur Verfügung stehenden Ressourcen muss die Maßnahme umsetzbar und finanzierbar sein.

Insbesondere der letzte Punkt setzt in einer Solidargemeinschaft eine konsentrierte Entscheidung über den verfügbaren finanziellen Gesamtrahmen sowie über eine Priorisierung von Gesundheitsleistungen voraus. Bisher übernimmt der G-BA zumindest indirekt und in Teilen über den Ein- und Ausschluss von Leistungen in den Leistungskatalog der GKV diese Funktion und stellt zwischen den verschiedenen Interessen und Interessenvertretern (auch Patienten) einen Konsens her. Es ist vorstellbar, dass der G-BA auch zukünftig im Rahmen der inhaltlichen Bewertung von Leitlinien (in Bezug auf deren Angemessenheit) eine ähnliche Funktion einnehmen könnte und im Anschluss an die Prüfung der Evidenzgrundlage durch das Institut entsprechende nationale Leitlinien herausgibt. Vor dem Hintergrund der Bewertung der Angemessenheit von Leitlinien ist das Institut in einen Entwicklungsprozess eingetreten, bei dem die Ziele und Methoden eines solchen Verfahrens erarbeitet werden.

4.3.1 Formale Prüfung

Eine Annäherung an die Frage nach der Qualität von Leitlinien gelingt durch eine formale Prüfung mit methodischen Kriterien [15,63,70,133]. Die formale Leitlinienbewertung erfolgt in strukturierter Form in Anlehnung an die Methoden des Leitlinien-Clearingverfahrens und in Bezug auf die Bewertungskriterien für Leitlinien der Bundesärztekammer und Kassenärztlichen Vereinigung [15,17] anhand des Deutschen Leitlinien-Bewertungsinstrumentes (DELBI) [16]. Das DELBI ist die deutsche Adaptation des AGREE-Instruments [3], erweitert um eine Domäne, die sich auf die Anwendbarkeit der jeweiligen Leitlinie im deutschen Gesundheitswesen bezieht. Die formale Bewertung wird durch zwei unabhängige Wissenschaftler durchgeführt. Bei unterschiedlichen Einschätzungen werden die Fragen diskutiert und die Leitlinien einer erneuten Bewertung unterzogen. Bleibt der Dissens bestehen, der durch eine Autorenanfrage nicht zu lösen ist, werden die unklaren Aspekte gesondert dokumentiert.

Weltweit existieren zu bestimmten medizinischen Fragestellungen mehrere Leitlinien, die sich methodisch und inhaltlich zum Teil sehr stark voneinander unterscheiden [207]. Eine formale Prüfung hat hier eine wichtige Filterfunktion, um letztlich eine inhaltliche Bewertung von Kernaussagen und spezifischen Empfehlungen der relevanten Leitlinien vornehmen zu können.

Zunächst wird zur Identifizierung der aktuellen existierenden Leitlinien für die jeweilige Fragestellung unter Berücksichtigung der in Abschnitt 6.1 genannten Vorgehensweise eine umfassende Leitlinienrecherche in den relevanten Fachdatenbanken (Leitlinien- und Literaturdatenbanken) durchgeführt. Das Filterverfahren der Leitlinienbewertung ist mehrstufig. Ein- und Ausschlusskriterien sowie die Recherchestrategie (Suchbegriffe, Auswahl der Datenbanken etc.) werden in Abhängigkeit von der Fragestellung vorab festgelegt und dokumentiert. Es folgt ein erster Screeningschritt, bei dem die Treffer nach vorab definierten thematischen Kriterien selektiert werden. In einem zweiten Schritt erfolgt eine methodische Kurzbewertung, die diejenigen Leitlinien auswählt, die einem internationalen Mindeststandard entsprechen. Die verbleibenden Dokumente werden dann inhaltlich bewertet.

4.3.2 Leitlinienvergleich und inhaltliche Bewertung von Kernempfehlungen

Die Bewertung der Inhalte von Leitlinien ist von besonderer Bedeutung. Die bisher mit den gängigen Instrumenten (ÄZQ-Checkliste [15], AGREE-Instrument [3], DELBI [16]) abgefragten Kriterien zur Identifizierung und Interpretation der Evidenz sowie zur Formulierung der

Leitlinienempfehlungen sind im Wesentlichen Transparenzkriterien^a, allein anhand derer die Darstellung des Prozesses z. B. der Literaturrecherche positiv bewertet wird, ohne z. B. eine Beurteilung der Vollständigkeit und Aktualität der Recherche abzugeben. Bei 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 von Vollständigkeit und Aktualität der einbezogenen Literatur auch die Interpretation und Bewertung der Studienergebnisse. Da dieses Verfahren sehr aufwendig ist, muss sich die inhaltliche Prüfung auf die in Auftrag gegebenen Fragestellungen des G-BA bzw. auf die wesentlichen Kernempfehlungen der Leitlinie beschränken. Die Identifizierung der Schlüsselempfehlungen erfolgt im Kontext jedes speziellen Auftrages in Abstimmung mit den betroffenen Ressortleitern und externen Sachverständigen. Eine Fokussierung auf ausgewählte Kernfragen hat hierbei mehrere Vorteile. Da Leitlinien in ihren Inhalten und Empfehlungen, auch wenn sie sich auf dasselbe medizinische Thema beziehen, sehr heterogen sind, kann man durch das Reduzieren auf vordefinierte Kernfragen einen guten Vergleich zwischen verschiedenen Leitlinien ziehen. Auch ist in diesem Rahmen eine Überprüfung der zugrunde gelegten Evidenz und der daraus resultierenden Empfehlung möglich, vorausgesetzt diese ist in der zugrunde gelegten Leitlinie transparent dargelegt. Aber nicht nur zwischen Leitlinien verschiedener Herausgeber, sondern auch mit anderen evidenzbasierten Quellen (systematischen Übersichtsarbeiten, HTA-Berichten, Evidenzberichten des Instituts u. a.) ist ein inhaltlicher Vergleich möglich, bezogen auf die definierte Kernfrage. Einzelne Fragestellungen werden bei fehlenden Angaben in Leitlinien in Abstimmung mit den übrigen Ressorts des Instituts bearbeitet.

Ein synoptischer Vergleich der Leitlinieninhalte kann bei der Identifizierung der Schlüsselempfehlungen hilfreich sein. Insbesondere können Fragestellungen, zu denen ein wissenschaftlicher Dissens besteht, identifiziert werden. Methodisch stellt der synoptische Vergleich jedoch lediglich eine Erleichterung im Bewertungsprozess dar; eine Überprüfung der Evidenzgrundlagen ist auch bei übereinstimmend empfohlenen Verfahren sinnvoll. Weiterhin kann im Rahmen der inhaltlichen Bewertung von Kernempfehlungen geprüft werden, ob die in der Leitlinie verwendeten Outcomeparameter patientenrelevant sind und eine Nutzen-Schaden-Abwägung [166] bei der Formulierung der Empfehlung berücksichtigt wurde.

^a Derzeit laufen Bemühungen, diese Aspekte methodisch zu verbessern, z. B. durch die GRADE Working Group [154].

4.3.3 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 [118,163,387]. Streng genommen lässt sich dies nur durch eine rigorose Effektevaluation prüfen [182]. Dies ist jedoch nicht bei jeder bestehenden Leitlinie zu realisieren. So gibt es zum Themenbereich der Outcome-evaluation nach Leitlinienimplementierung nur unsystematische Untersuchungen mit sehr heterogenen Fragestellungen und Ergebnissen. Auch Pilotstudien wurden bislang nur für wenige Leitlinien vor einer geplanten Veröffentlichung durchgeführt [18,19].

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

Darüber hinaus ist eine Gegenüberstellung der Leitlinienempfehlungen und der 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 [63].

Bei ausländischen Leitlinien ist 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.

4.4 Darlegung der Qualitätsprüfung

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

4.5 Abgabe von Empfehlungen zu Disease-Management-Programmen

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

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

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

ABGELÖST DURCH
"Allgemeine Methoden -
Version 4.0
vom 23.09.2011"

5 Evidenzbasierte Gesundheitsinformation für Bürger und Patienten

5.1 Hintergrund und Ziele

Dem Institut wurde vom Gesetzgeber die Aufgabe zugewiesen, allgemeine Gesundheitsinformationen für Bürger und Patienten zu erstellen. Es zählt nicht zu seinen Aufgaben, Einzelpersonen direkt zu beraten. Ziel des Instituts ist die Verbesserung der Gesundheit und Patientenautonomie durch die Bereitstellung von Gesundheitsinformationen, die die allgemeine Gesundheitskompetenz und die Wissenschaftskenntnisse („Health and Scientific Literacy“) fördern sollen [33,95,228,239]. Die vom Institut erstellten Gesundheitsinformationen sollen daher

- eine aktive und informierte Entscheidungsfindung bei gesundheitlichen Fragen unterstützen,
- die kritische Nutzung gesundheitsbezogener Dienstleistungen fördern,
- das Wissen um körperliche, psychische und emotionale Gesundheit verbessern,
- das Verständnis medizinischer und wissenschaftlicher Informationen verbessern, darunter auch das des Konzepts der evidenzbasierten Medizin, und
- die Unterstützung der Patienten durch ihre Familie und Freunde ermöglichen.

Um diese Ziele zu erreichen, ist es notwendig, dass das Institut als ein vertrauenswürdiger und patientenzentrierter Herausgeber von Informationen wahrgenommen wird. Die Einbindung der Wertvorstellungen von Patienten in die medizinische Entscheidungsfindung gehört zum Kern des Konzepts der evidenzbasierten Medizin [314] und ist damit auch ein wesentlicher Bestandteil evidenzbasierter Gesundheitsinformationen. Es gibt mehrere Definitionen von evidenzbasierten Patienteninformationen [77,115,317,345]. Jede dieser Definitionen verlangt, dass Patienteninformationen Belege für den Nutzen, den Schaden und die Ungewissheiten enthalten müssen, die bezüglich der jeweiligen medizinischen Interventionen bestehen. Doch die Anforderungen an evidenzbasierte Gesundheitsinformationen gehen noch darüber hinaus [345]. Das Institut definiert evidenzbasierte Gesundheitsinformationen als Informationen, in denen

- der Inhalt auf klaren wissenschaftlichen Belegen (strenger Evidenz aus systematischen Übersichten) basiert.
- Informationen nach systematischen Methoden erstellt werden, die helfen sollen, systematische Fehler (Bias) zu minimieren und Neutralität zu wahren.

- evidenzbasierte Kommunikationstechniken verwendet werden, um dem Ziel der Aufklärung, Unterstützung und der Befähigung von Bürgern und Patienten zur selbstbestimmten Entscheidung (sog. Empowerment) gerecht zu werden.
- ungeklärte Sachverhalte sowie der potenzielle Nutzen und Schaden erörtert werden.
- Sprache und Gestaltung neutral und nicht direktiv gehalten sind, sodass die Entscheidungen im Einklang mit den eigenen Wertvorstellungen der Patienten getroffen werden können.
- die Informationen aktualisiert werden, damit sie evidenzbasiert bleiben.

Primäres Kommunikationsmedium des Instituts ist das Internet, über das es Bürger, Patienten und Multiplikatoren zu erreichen sucht. Internetbasierte (Online-) und computergestützte medizinische Offline-Informationen können das Wissen, die Auswahlentscheidungen, die Gesundheit und das Wohlbefinden von Bürgern und Patienten positiv beeinflussen [42,221,246,273,276,341,377,385]. Informationsinterventionen können allerdings auch ineffektiv oder sogar schädlich sein [106,108,159,276]. Überdies sind viele Nutzer mit evidenzbasierten Informationen nicht vertraut, was besondere Anforderungen an die Kommunikation stellt [116,150,346]. Das Institut muss deshalb seine Expertise im Bereich der evidenzbasierten und elektronischen Kommunikation aufrechterhalten und die Anwendung der von ihm erstellten Gesundheitsinformationen überwachen.

5.2 Patientenzentrierte Kommunikation

5.2.1 Kommunikationsstandards

Eine der wichtigsten Herausforderungen im Zusammenhang mit evidenzbasierten Gesundheitsinformationen besteht darin, die Informationen zwar allgemein verständlich, dabei jedoch wissenschaftlich präzise und objektiv zu präsentieren. Die objektiven und leicht handhabbaren Verfahren zur Messung der Lesbarkeit englischer, französischer und spanischer Texte sind in Deutschland aber nur begrenzt anwendbar, und es gibt hier auch kein in ähnlicher Weise validiertes Bewertungsinstrument [146]. Überdies bestehen starke Unterschiede in der Lesekompetenz von Bürgern und Patienten [228].

Um dem unterschiedlichen Bildungsniveau in der Bevölkerung gerecht zu werden, erstellt das Institut seine Informationen auf verschiedenen Verständlichkeitsniveaus – von komplexen Artikeln bis hin zu leichter verständlichen Merkblättern und multimedialen Produkten (siehe unten). Da das Institut seine Informationen sowohl in englischer als auch in deutscher Sprache veröffentlicht, kann es auch englische Bewertungsinstrumente einsetzen, um den Grad der Lesbarkeit seiner

Informationen abzuschätzen. Primäre Mittel einer Qualitätsbewertung im Hinblick auf die Verständlichkeit von Texten sind jedoch die Begutachtung von Entwürfen durch Testleser sowie Leserbewertungen (siehe unten).

Die Evidenz zu erklären und bei der Vermittlung gesundheitsbezogener Informationen objektiv zu bleiben, stellt eine zusätzliche Herausforderung dar [116,327,358]. Zur Gewährleistung eines objektiven und nicht direktiven Vorgehens sollte das Institut in seinen Gesundheitsinformationen das, was als wissenschaftlich bekannt gilt, einerseits nicht überhöhen und andererseits den Menschen nicht vorschreiben, was sie „tun sollen“. Diesem Anspruch wird dadurch Rechnung getragen, dass das Institut keine Empfehlungen ausspricht und eine neutrale Sprache verwendet. Unter Berücksichtigung der wachsenden Evidenz zur Kommunikation von wissenschaftlichen Erkenntnissen verfolgt das Institut das Ziel,

- seine Informationen zum besseren Verständnis in einheitlichen Formaten herauszugeben und sie um „zusätzliche Elemente“ zu ergänzen, die die Verständlichkeit medizinischer Begriffe und numerischer Informationen verbessern sollen,
- die mit der jeweiligen Evidenz verbundene Ungewissheit zu erklären,
- anzugeben, auf welche Personengruppen die Evidenz anwendbar ist,
- 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“) hinzuweisen,
- Informationen zum relativen Risiko um Angaben zum absoluten Risiko und, wann immer möglich, andere Informationen (z. B. Ausgangsrisiko) zu ergänzen und
- verzerrte Informationen für oder gegen die Produkte eines bestimmten Unternehmens zu vermeiden, indem Handelsnamen nur dann verwendet werden, wenn es für das Verständnis unerlässlich ist und/oder wenn alle auf dem deutschen Markt erhältlichen Produkte namentlich aufgeführt werden können.

Evidenz aus Screeningverfahren zeigt, dass die Angabe personalisierter oder individualisierter Risikoschätzer wirksam sein kann [108]. Das Institut setzt Instrumente ein, die Menschen helfen können, ihr persönliches Risiko abzuschätzen. Auch wenn hinsichtlich formaler Entscheidungshilfen – vor allem bei der individuellen Anwendung im Internet [68,113] – eine Ungewissheit bestehen bleibt, behält sich das Institut vor, solche Entscheidungshilfen für manche Themen zu entwickeln. Dabei wird insbesondere auf die Erfahrungen mit spezifischen Entscheidungshilfen zurückgegriffen, deren Wirksamkeit in randomisierten Studien nachgewiesen wurde. Das Institut

orientiert sich bei der Entwicklung von Entscheidungshilfen an den International Patient Decision Aid Standards (IPDAS) [113,189].

Die Bereitstellung von Information ist nicht der einzige Zweck von Gesundheitskommunikation. Ein besonderes Anliegen ist auch, Patienten und Bürgern emotionale Unterstützung zu bieten [129]; ferner kann sie auch bei der Stärkung der Patientenautonomie eine Rolle spielen. Gesundheitskommunikation muss patientenzentriert sein, wenn sie zum Patienten-Empowerment beitragen und emotional unterstützend wirken soll. Nach der Definition der Weltgesundheitsorganisation (WHO) beinhaltet der Begriff „Empowerment“ im Gesundheitswesen die Fähigkeit, Entscheidungen zu treffen und Maßnahmen zu ergreifen, die im Einklang mit den eigenen Zielen stehen [283]. Diese Fähigkeiten bilden die Voraussetzung für autonomes Denken und Handeln aufseiten der Patienten. Gesundheitskommunikation zur Stärkung der Gesundheitskompetenz spricht an, was Patienten wissen möchten, zeigt Interesse an der und Respekt für die Meinung von Patienten und erkennt ihre Kompetenz an [97,227,372].

Historisch gesehen waren Patienteninformationen eher paternalistisch orientiert [97]. Man ging davon aus, dass der Patient unwissend ist und vor Ungewissheit und belastenden Informationen geschützt werden muss: Ihm sollte gesagt werden, was er zu tun hat. Zu einem nicht unbedeutenden Grad ist dies auch heute noch ein Kennzeichen von Patienteninformationen und der Debatten um ihren aktuellen Stellenwert [97]. In Deutschland wird zuweilen argumentiert, die Patienten seien für evidenzbasierte und nicht direktive Informationen noch nicht reif oder würden diese nicht wollen.

Wie Studien jedoch oft zeigen, erhalten Patienten in Deutschland weniger Informationen, als sie gern hätten [95], und – verglichen mit Patienten in mit Deutschland vergleichbaren Ländern – geben in Deutschland mehr Patienten an, dass ihnen nicht genügend Informationen zur Verfügung gestellt würden [322,366]. Deutsche Internetnutzer interessieren sich eventuell sogar stärker für die neuesten Forschungsergebnisse als für die Meinungen von Experten [297]. Dennoch ist zum großen Teil nicht bekannt, welche Wirkungen es hat, wenn Menschen verstärkt über bestehende wissenschaftliche Ungewissheiten aufgeklärt werden. Es wurde argumentiert, dass der Umgang mit Ungewissheit bei Erwachsenen ein integraler Bestandteil der Bewältigung und Selbstverwirklichung ist [137]. Eine potenziell schädliche Wirkung kann allerdings nicht ausgeschlossen werden. Daher wird sich das Institut über die Evidenz zur Verbreitung von Risiken, Ungewissheiten und anderen problematischen Informationen auf dem Laufenden halten und die Wirkungen seiner Gesundheitsinformationen überwachen müssen.

Ebenso, wie sich das Institut um die Verständlichkeit, Objektivität und Genauigkeit seiner Informationen bemüht, hat es sich zum Ziel gesetzt,

- Sensibilität und Respekt vor dem Wissen, den Wertvorstellungen und Sorgen von Bürgern und Patienten, vor ihrer Autonomie, ihren kulturellen Unterschieden sowie gegenüber geschlechts-, alters- und behindertenspezifischen Belangen zu zeigen.
- sich um einen patientenzentrierten, nicht wertenden, nicht direktiven und neutralen Sprachstil zu bemühen.
- auf die Zeit der Leser Rücksicht zu nehmen.

Auf der Grundlage der verfügbaren Evidenz und der Erfahrungen anderer Gruppen entwickelt das Institut für seine Produkte einen Leitfaden zur Textgestaltung. Es wird – im Lichte der Überwachung und Evaluation seiner Produkte sowie neuer Evidenz aus dem Bereich der evidenzbasierten Kommunikation – auch weiterhin an der Entwicklung seiner Kommunikationsstandards arbeiten.

5.2.2 Visuelle Kommunikation und Multimedia

Ein Text allein ist möglicherweise weniger verständlich und einprägsam als Informationen, in denen die Erklärungen durch Bilder ergänzt werden [116,193,236,250,358]. Auch gesprochener Text kann zum besseren Verständnis beitragen [193,318]. Die verständlichste Kommunikationsform – vor allem für Menschen mit geringerem Bildungsstand – könnten Erläuterungen darstellen, in denen Text, Bild und Ton miteinander kombiniert sind [193]. Wo immer möglich, bereitet das Institut seine Texte visuell und akustisch auf, um die Wirksamkeit der Informationen zu verbessern und so ein breiteres Publikum zu erreichen. Dazu zählen anatomische Zeichnungen und kurze Filmsequenzen zu Schlüsselthemen, in denen visuelle Elemente, Text und Ton kombiniert werden (Animationen). Auch Grafiken und Piktogramme helfen vielen Menschen, numerische Daten und andere Aspekte wissenschaftlicher Evidenz zu verstehen [107,250,327]. Visuelle und multimediale Elemente sollen den Text nicht ersetzen, sondern die im Text präsentierten Inhalte unterstützen. Dadurch wird sichergestellt, dass die Informationen auch Menschen mit Seh- oder Hörstörungen zugänglich sind.

Das Internet macht es möglich, dass Gesundheitsinformationen im Multimediaformat präsentiert werden können. Da die Technologie des Internets sich ständig weiterentwickelt und der Zugang zum Internet nicht mehr nur auf Computer beschränkt ist, lässt sich eine effektive Kommunikation mithilfe von Bild und Ton auf Websites für mehr Anwender immer besser realisieren. Das Internet ermöglicht überdies eine Interaktion mit den Nutzern, sodass die Kommunikation nicht nur in Richtung der Nutzer fließen muss. Interesse zu zeigen an dem, was dem Patienten wichtig ist, ist ein entscheidender Faktor in der patientenzentrierten und die Patientenkompetenz stärkenden

Kommunikation [97,227,372]. Auch wenn das Institut keine individuelle Gesundheitsberatung durchführen kann, so verfügt es doch über verschiedene Möglichkeiten, um seinen Nutzern Gelegenheit zu geben, ihre Ansichten und Sorgen mitzuteilen, und zwar u. a. durch

- eine Online-Bewertung der Qualität von Informationen,
- Themenvorschläge und ein allgemeines Online-Kontaktformular,
- ständige Umfragen zur Benutzerfreundlichkeit der Website und
- gelegentliche Online-Umfragen zu spezifischen Gesundheitsthemen.

5.2.3 Informationszugang und Barrierefreiheit

Das Internet hat sowohl besondere Vor- als auch Nachteile, was den Zugang zum Medium betrifft. Die Tatsache, dass das Internet Personen mit dem entsprechenden Anschluss 24 Stunden, also rund um die Uhr, zur Verfügung steht, macht es zu einem sehr leicht zugänglichen Medium. Der Zugang zum Internet breitet sich weiterhin aus: Mehr als die Hälfte der Menschen in Deutschland nutzt das Internet zur Beschaffung von Gesundheitsinformationen, und diese Zahl steigt langsam weiter an [95,325].

Denkbar ist, dass das Internet besonders häufig von chronisch Kranken benutzt wird. In Umfragen unter Patienten an orthopädischen Kliniken in Deutschland wurde beispielsweise festgestellt, dass bis zu 70 % der Befragten das Internet zur Beschaffung von Informationen über ihre Krankheit nutzen [290,301]. Mehr als ein Drittel (38 %) der Patienten hatte vor dem Klinikbesuch auf das Internet zugegriffen, um sich auf den bevorstehenden Arztbesuch vorzubereiten [290]. Um von den ins Internet eingestellten Gesundheitsinformationen zu profitieren, muss man nicht unbedingt selbst Zugang zum Internet haben: Häufig durchsuchen Verwandte oder Freunde das Internet für die Betroffenen, und auch Multiplikatoren wie Ärzte, Selbsthilfegruppen und Journalisten nutzen das Internet routinemäßig. Auch im Familienkreis werden Informationen oft in breitem Umfang weitergegeben [329]. Geschlechtsspezifische Unterschiede im Zugang zum Internet waren in den Anfangsjahren des World Wide Web sehr deutlich; mittlerweile scheinen sie sich abzuschwächen [276].

Verschiedene Faktoren können den Zugang zum Internet und seine Nutzung begrenzen, darunter:

- Behinderungen, vor allem (aber nicht nur) Seh- und Hörstörungen
- Leseschwäche
- unzureichende Fertigkeiten im Umgang mit Computern
- technische Ausstattung (die die Geschwindigkeit von und den Zugang zu Multimediaformaten beeinflusst)
- Sprache (die Muttersprache des Nutzers)

Das Institut gewährleistet, dass seine Website die international anerkannten Zugänglichkeitsrichtlinien für Personen mit Behinderungen [374] sowie die deutsche BITV (Barrierefreie Informationstechnik-Verordnung) erfüllt [61]. Es wird sich weiterhin um die Evaluation und Optimierung der Nutzbarkeit seiner Website bemühen und Instrumente entwickeln, die zum Verständnis der vom Institut dargebotenen Informationen beitragen.

Nahezu 10 % der in Deutschland lebenden Menschen sind anderer Nationalität, und weitere 10% der Deutschen kommen aus Migrantenfamilien [344], von denen die größte kulturelle Gruppe aus der Türkei stammt [344]. Menschen mit nicht deutschsprachigem Hintergrund haben als Gruppe womöglich einen größeren Bedarf an Gesundheitsinformationen [84]. Von daher ist es wichtig, dass die Merkblätter in sehr leicht verständlichem Deutsch bleiben. Im Idealfall würden jedoch den Betroffenen kultur- und gendergerechte Gesundheitsinformationen in der Sprache zur Verfügung gestellt, die sie am besten verstehen. Das Institut arbeitet mit externen Partnern zusammen, damit wenigstens einige seiner Gesundheitsinformationen in die am häufigsten in Deutschland gesprochenen Sprachen übersetzt werden können.

Das Institut veröffentlicht seine Gesundheitsinformationen sowohl in deutscher als auch in englischer Sprache und verfügt über die Möglichkeit, auch in weiteren Fremdsprachen zu publizieren. Die Sicherung der bestmöglichen Informationsqualität erfordert eine breite internationale Beteiligung. Da die Gesundheitsinformationen des Instituts auch in englischer Sprache herausgegeben werden, kann das Institut vom Feedback internationaler Wissenschaftler und Reviewer, insbesondere der Autoren systematischer Reviews, profitieren. Die Verfügbarkeit einer englischsprachigen Version erweitert die internationalen Möglichkeiten zur Übersetzung der Texte in andere Sprachen.

Für das Übersetzen von Gesundheitsinformationen sind hohe Standards erforderlich, doch häufig ist die Qualität solcher Übersetzungen unzureichend [144]. Die Qualität von Übersetzungen anhand objektiver Kriterien zu beurteilen, ist aber schwierig. Die Texte können Wort für Wort übersetzt sein, oder die Übersetzung gibt den Inhalt des Ausgangstextes in der Zielsprache sinngemäß wieder [271]. Sinngemäße Übersetzungen bieten die Möglichkeit, den höchsten Grad an Verständlichkeit in der Zielsprache zu erreichen, weshalb diese Form der Übersetzung vom Institut auch bevorzugt wird. Alle Übersetzungen der vom Institut erstellten Gesundheitsinformationen werden nochmals von einer zweiten Person geprüft, die über Kenntnisse in beiden Sprachen verfügt.

5.3 Themenwahl, Recherche und wissenschaftliche Bewertung

5.3.1 Themenwahl

Die vom Institut erstellten Gesundheitsinformationen gehen hervor

- aus Direktaufträgen des G-BA oder des BMG zur Erstellung von Patienteninformationen,
- aus Zusammenfassungen der anderen Produkte des Instituts und Begleitinformationen zu diesen Produkten und
- aus der gesetzlichen Aufgabe des Instituts zur Bereitstellung von Gesundheitsinformationen für Bürger sowie aus Eigeninitiative des Instituts gemäß dem Generalauftrag des G-BA.

Dieser Abschnitt befasst sich mit den Methoden, die der Auswahl der Themen für Begleitinformationen sowie den gemäß dem Generalauftrag in Wahrnehmung der gesetzlichen Aufgaben erstellten Informationen zugrunde liegen. Die Themen für Gesundheitsinformationen sind potenziell unerschöpflich, und natürlich ist es unmöglich, alle über alles zu informieren. Wie bei anderen gesundheitsbezogenen Entscheidungen, bei denen Prioritäten gesetzt werden müssen, beinhaltet auch die Entscheidung über Themen für Gesundheitsinformationen die gleichzeitige Analyse vielfältiger Informationsströme aus unterschiedlichen Quellen [27,28].

Gemäß § 139a SGB V fällt dem Institut folgende Aufgabe zu: „Bereitstellung von für alle Bürger verständlichen allgemeinen Informationen zur Qualität und Effizienz in der Gesundheitsversorgung sowie zu Diagnostik und Therapie von Krankheiten mit erheblicher epidemiologischer Bedeutung“. Der Generalauftrag des Instituts wurde im Juli 2006 konkretisiert, um – nach der „Erfassung und Auswertung des relevanten Schrifttums“ – auch Informationen für Bürger bereitzustellen. Sämtlichen Ressorts des Instituts, darunter auch dem Ressort „Gesundheitsinformation“, obliegt es gemäß Generalauftrag, „eine kontinuierliche Beobachtung und Bewertung medizinischer Entwicklungen von grundlegender Bedeutung“ vorzunehmen und darüber zu berichten.

Eine breit akzeptierte Definition bzw. eine klar umrissene Zusammenstellung der „Krankheiten mit erheblicher epidemiologischer Bedeutung“ konnte nicht ausgemacht werden. Eine praxisnahe Festlegung der epidemiologischen Bedeutung von Krankheiten muss sich auf Faktoren stützen, für die sich Daten zur Krankheitslast identifizieren lassen. Aspekte von epidemiologischer Bedeutung könnten die folgenden Merkmale einschließen:

- Mortalität
- Häufigkeit (Prävalenzen/ Inzidenzen)
- Häufigkeit der Inanspruchnahme medizinischer Leistungen
- Behandlungskosten

- krankheitsbedingte Fehlzeiten am Arbeitsplatz
- Einschränkung der Lebensqualität und andere Folgen, die für die Betroffenen relevant sind

Zur Priorisierung von Themen nutzt das Institut verschiedene Quellen u. a. zu Todesursachen, Diagnosen bei stationären Einweisungen, krankheitsbedingten Fehlzeiten am Arbeitsplatz, häufigen Diagnosen und Arzneimittelverordnungen in Deutschland sowie die Liste der im Rahmen des morbiditätsorientierten Risikostrukturausgleichs ausgewählten Krankheiten [64].

Der Gesetzgeber hat die Informationsaufgaben des Instituts sowohl am Bürger – also am Gesunden – als auch am Patienten orientiert. Von daher muss sein Informationsangebot auch Themen beinhalten, die nicht nur der Perspektive von Betroffenen gerecht werden. Um seine Ziele zu erreichen, muss das Institut Informationen anbieten, die Bürger und Patienten dazu befähigen, Auswahlentscheidungen zu treffen und Maßnahmen zu ergreifen, mit denen sie ihre eigenen gesundheitlichen Ziele verwirklichen können [283], die das Selbstmanagement fördern und das ansprechen, was die Menschen wissen wollen [97,227,372]. Es gibt Hinweise darauf, dass die Beteiligung von Patienten an der Erstellung von Patienteninformationen deren Relevanz steigern kann [282]. Solche „am Verbraucher orientierten“ Patienteninformationen sind eventuell auch besser dazu geeignet, das Wissen der Patienten zu erweitern [282]. Das Institut muss sich bei seiner Themenwahl daher auch danach richten, was Patienten und Bürger interessieren könnte und welche Gegebenheiten sie in der Versorgungswirklichkeit vorfinden.

Das Institut berücksichtigt diverse Quellen, um die Informationswünsche der Bürger und Patienten zu erschließen:

- Umfragen, qualitative Primärstudien und Übersichten von qualitativen Studien zum Informationsbedarf der Bevölkerung
- Erfahrungen von anderen Informationsanbietern, Patientenberatungsstellen und Selbsthilfegruppen
- Anfragen an die Beauftragte der Bundesregierung für die Belange der Patienten [326]
- Themen, die in die Suchmaschine der IQWiG-Website für Bürger und Patienten (www.gesundheitsinformation.de) eingegeben werden, sowie andere Daten zur Nutzung von Internetseiten
- Themenvorschläge von Website-Nutzern
- Ergebnisse der eigenen Online-Umfragen des Instituts zu Prioritäten und Interessen

Das Institut berücksichtigt auch Übersichten über effektive Informationsinterventionen zu spezifischen gesundheits- und krankheitsbezogenen Fragestellungen, um besser entscheiden zu können, welche Gesundheitsinformationen möglicherweise von Nutzen sind.

Die Themenauswahl des Instituts ist nicht ausschließlich an vordefinierten („interessengetriebenen“) Fragestellungen orientiert. Priorität wird solchen Fragen eingeräumt, auf die sich evidenzbasierte Antworten finden lassen. Aus diesem Grund setzt das Institut einen Prozess zur Sichtung der Evidenz („Evidenzscanning“) ein, um potenzielle Themen für diese Informationen herauszuarbeiten. Das Scanningsystem beruht vorwiegend auf der Identifizierung von systematischen Übersichten und HTA-Berichten, darunter auch Reviews zu unerwünschten Wirkungen. In einem Pilotversuch hat das Institut 2006 die Machbarkeit eines Scanningsystems getestet. Dieses System wird in Zusammenarbeit mit dem Centre for Reviews and Dissemination weiterentwickelt, um sich die umfangreichen Erfahrungen des Centre mit der Identifizierung und Bewertung systematischer Übersichten zunutze zu machen [280].

Das Institut erstellt derzeit eine Kriterienliste, die die Auswahl der Themen erleichtern soll. Das Evidenzscanningsystem wird auch eine entscheidende Rolle bei der Aktualisierung der Gesundheitsinformationen des Instituts spielen (siehe unten).

5.3.2 Recherche

Das Institut stützt sich bei der Erstellung seiner Informationen vorwiegend auf systematische Übersichten und qualitative Studien. Wenn ein Thema im Detail recherchiert werden soll, sucht das Institut in der Regel nach folgenden Informationen, um Fragestellungen zu identifizieren, die für Patienten und Bürger von Interesse und Relevanz sind:

- Übersichten von qualitativen Studien sowie Rapid Appraisals qualitativer Primärstudien (siehe Abschnitt 6.5)
- Übersichten zu den Wirkungen von Kommunikation
- Übersichten zur „Adherence“
- kostenlos zugängliche Patienteninformationen im Internet sowie Internetseiten von Selbsthilfegruppen

Das Internet und andere Quellen werden zudem durchsucht, um Interventionen zu identifizieren, die von Patienten und Bürgern genutzt bzw. ihnen angeboten werden.

Anhand der Ergebnisse dieser ersten Auswertung von Patienten- und Informationsbedürfnissen kann sich das Institut ein Bild davon machen, welche Stationen ein Patient bei einem bestimmten gesundheitlichen Problem durchlaufen muss, welche psychologischen und emotionalen Probleme im Zusammenhang mit diesem Thema auftreten können und an welchen Punkten der Einzelne Entscheidungen treffen muss. Ferner werden gewöhnlich auch die Vertreter von Patienten oder

Bürgern befragt, um weitere Aspekte zu identifizieren und um zu erörtern, welche Relevanz den Rechercheergebnissen des Instituts in Deutschland zukommt.

Das Institut führt auch Recherchen nach systematischen Übersichten zu Ursachen, Prognosen, zur Diagnostik, zu Therapien und Nebenwirkungen durch. Üblicherweise lässt sich damit die gesamte Krankheit abdecken, wobei ein später durchzuführendes Scoping dazu dient, auf die Bereiche zu fokussieren, mit denen sich die Gesundheitsinformation befassen wird. Die Recherchen schließen die Datenbank Database of Reviews of Effects (DARE) [54,280] und den systematischen Übersichtenfilter von PubMed [268] ein, sind aber nicht darauf begrenzt. Berücksichtigt werden nur Übersichten mit Suchen, die in den letzten fünf Jahren durchgeführt wurden. In der Regel werden Übersichten dann als aktuell betrachtet, wenn die Suche in den letzten drei Jahren durchgeführt wurde [335,336].

Gelegentlich zieht das Institut eine „Rapid-Recherche“ nach Studien in Betracht. Wenn es beispielsweise zu einem wichtigen Thema keine neuere Übersicht gibt, wird eine Aktualisierungsrecherche dann erwogen, wenn eine hochwertige Übersicht vorliegt, deren Suche mehr als drei Jahre zurückliegt [335]. In einigen Fällen werden auch Aktualisierungen als relevant erwogen, wenn die Zeitspanne diese drei Jahre über- oder unterschreitet. Das richtet sich jeweils nach der Stärke der Evidenz in der betreffenden Übersicht und dem Ausmaß der Forschungsaktivitäten auf dem entsprechenden Gebiet.

Aktualisierungsrecherchen nach Studien, mit denen die Aktualität einer Übersicht überprüft werden soll, werden gewöhnlich im Rahmen des Cochrane Controlled Trials Register und des klinischen Studienfilters von PubMed durchgeführt [179].

„Rapid-Recherchen“ nach qualitativer Forschungsliteratur werden von einem Mitarbeiter durchgeführt, während ein zweiter Mitarbeiter die endgültigen Ergebnisse überprüft. Die Recherche nach anderer Evidenz umfasst ein „Ergebnisscreening“, das mindestens zwei Mitarbeiter unabhängig voneinander durchführen. Über Ausschlüsse müssen sich beide Mitarbeiter verständigen, wobei in Zweifelsfällen – wenn nötig – ein dritter Mitarbeiter hinzugezogen wird.

5.3.3 Evidenzbewertung

Die vom Institut für Patienten und Bürger erstellten Gesundheitsinformationen stützen sich zum großen Teil auf systematische Übersichten (siehe Abschnitt 6.3). Damit das Institut eine systematische Übersicht über Behandlungseffekte für Gesundheitsinformationen verwenden kann, muss diese gewisse Mindestvoraussetzungen erfüllen, d. h. sie darf nach dem Oxman-und-Guyatt-Index nur geringe methodische Mängel aufweisen [210,286,288]. Um sich als Thema einer zusammenfassenden „Kurzantwort“, die einen Behandlungsnutzen nahelegt, zu qualifizieren, sollte

eine systematische Übersicht mindestens eine Studie umfassen, deren Qualität von den Übersichtsautoren als adäquat erachtet wird, und Daten zu mindestens einem patientenrelevanten Endpunkt enthalten. Das Institut berücksichtigt auch die Relevanz und die Übertragbarkeit der Evidenz, insbesondere in Bezug auf Geschlecht und Alter (siehe Abschnitt 6.5).

Unabhängigkeit ist ein entscheidendes Attribut, wenn es um den gesetzlichen Auftrag des Instituts geht, um die Art und Weise, wie das Institut von anderen dargestellt wird [95], und um seine Selbstdarstellung (www.gesundheitsinformation.de: „unabhängig, objektiv und geprüft“). Unabhängigkeit könnte auch dazu beitragen zu gewährleisten, dass den Patienteninformationen des Instituts Vertrauen entgegengebracht wird [226,297,337], selbst wenn dies nicht unbedingt für alle Menschen gilt [34]. Auch die Unabhängigkeit in der Durchführung von Forschungsarbeiten könnte für viele Bürger relevant sein [230]. Systematische Übersichten sollten den aktuellen Wissensstand unvoreingenommen bewerten. Manche Untersuchungen lassen darauf schließen, dass systematische Übersichten, die vom Hersteller eines bestimmten Produkts gesponsert werden, zu Ergebnissen kommen, die im Sinne einer allzu positiven Bewertung der Wirksamkeit des betreffenden Produkts verzerrt sind [218,388]. Solch ein Zusammenhang wurde in einer weiteren Studie beschrieben, auch wenn er statistisch nicht signifikant war [293]. Um den guten Ruf des Instituts als Garant für unabhängige Informationen zu wahren, werden von Herstellern gesponserte Übersichten bei der Erstellung der Patienteninformationen des Instituts daher nicht berücksichtigt.

Wird ein bestimmtes Thema oder Resultat in mehr als einer systematischen Übersicht von angemessener Qualität untersucht, wird eine weitere Qualitätsbewertung durchgeführt, um festzustellen, ob sich unter diesen Übersichten auch solche minderer Qualität befinden. Als Quelle für die in den Gesundheitsinformationen des Instituts benutzten numerischen Daten dienen die Ergebnisse der Übersicht, die im Hinblick auf ein bestimmtes Outcome die höchste Qualität aufweist. Wenn Übersichten zu widersprüchlichen Schlussfolgerungen gelangen, wird nach den möglichen Gründen dafür gesucht [211].

Für die formale Bewertung der Stärke der Primärevidenz in einer ausgewählten systematischen Übersicht kommen die Methoden der GRADE-Arbeitsgruppe infrage [23,24,154,172]. Das GRADE-System bewertet explizit die Qualität der Evidenz und beschreibt den Grad des Vertrauens, das man in die Schätzung spezifischer Therapieeffekte, zum Beispiel in den Schätzer der mit der betreffenden Therapie assoziierten Mortalität, setzen kann.

Bei Fragestellungen wie etwa zur Ätiologie oder Prognose oder bei der qualitativen Beschreibung von Patientenerfahrungen besteht die geeignete Evidenzbasis für eine systematische Übersicht aus anderen Arten von Primärstudien [149]. Bei der Bewertung von solchen systematischen Übersichten wendet das Institut zur Orientierung die Kriterien des Oxford Centre for Evidence-

Based Medicine sowie das Evidenzbewertungssystem der McMaster-Universität an [67,179]. Die Methoden des Instituts zur Bewertung qualitativer Studien werden in Abschnitt 6.5 beschrieben.

5.3.4 Aktualisierung

Ein entscheidender Faktor bei der Erstellung von evidenzbasierten Gesundheitsinformationen ist die Gewährleistung, dass die Schlussfolgerungen nicht veraltet sind. Regelmäßige Aktualisierungen sind eines der Qualitätskriterien, die die Europäische Union für gesundheitsbezogene Internetseiten festgelegt hat [75]. Die Evidenz nimmt exponentiell zu. Das gilt sowohl für Studien [359] als auch für systematische Übersichten [266]. Neue Evidenz kann dazu führen, dass existierende Übersichten als überholt bzw. nicht mehr aktuell gelten [336]; oft kommt es allerdings vor, dass die ursprünglichen Schlussfolgerungen durch die neue Evidenz nicht grundlegend verändert, sondern eher noch bekräftigt werden [203,298].

Eine Studie über Leitlinienempfehlungen kam zu dem Schluss, dass nach drei Jahren mehr als 90 %, nach sechs Jahren jedoch nur noch etwa 50 % der in Leitlinien abgegebenen Empfehlungen aktuell sein dürften [335]. Bei manchen Themen, zu denen beispielsweise sehr starke Evidenz vorliegt, liegt die Halbwertszeit der Evidenz deutlich unter, in anderen Bereichen dagegen wiederum über diesen drei Jahren [336]. Da das exponentielle Wachstum der Evidenz aber weiterhin anhält, ist davon auszugehen, dass die Halbwertszeit der Information sich generell verkürzt, d. h. dass die Informationen schneller veralten. Das Institut sieht daher üblicherweise drei Jahre als Untergrenze für eine Aktualitätsprüfung an und überwacht die Halbwertszeit seiner Gesundheitsinformationen, um so Erkenntnisse für zukünftige Aktualisierungsverfahren zu gewinnen.

Aktualisierungen können sehr ressourcenintensiv sein. Schätzungen zufolge nimmt z. B. die vollständige Aktualisierung einer Leitlinie fast ebenso viel Zeit in Anspruch wie die Erstellung einer neuen Leitlinie [105]. Die herkömmlichen Aktualisierungsmechanismen dienen dazu, einen festen Zeitpunkt für eine Übersicht festzusetzen. Dieses Verfahren kann von Anbietern mehrerer Arten von evidenzbasierten Informationen langfristig aber nur aufrechterhalten werden, wenn auch die Ressourcen kontinuierlich aufgestockt werden: Die durch Aktualisierungen anfallende Arbeitsbelastung nimmt im Laufe der Zeit exponentiell zu. Die Cochrane Collaboration beispielsweise hat sich das Ziel gesetzt, die Recherchen für mindestens 80 % ihrer Übersichten alle zwei Jahre zu aktualisieren [233]. Dieses Ziel konnte bislang allerdings nicht erreicht werden; stattdessen verlieren die Übersichten vermutlich zunehmend an Aktualität [233]. Ginge man bei den Aktualisierungen von diesem Zwei-Jahres-Standard aus, wäre die Kapazität des Instituts im

Hinblick auf die Aktualität seiner bestehenden Informationen und die Erstellung neuer Informationen innerhalb von fünf Jahren erschöpft.

Das Institut geht nach folgendem Modell vor, um seine Informationen auf einen relativ aktuellen Stand zu bringen.

- Als Anstoß für Aktualisierungen („Auffrischung“ der Informationen) dienen die mithilfe des Evidenzscanningsystems und zusammen mit dem Centre for Reviews and Dissemination (siehe Abschnitt 5.3.1) identifizierten neuen Übersichten.
- Die Aktualisierungen einer Schlüsselgruppe anderer Anbieter evidenzbasierter Gesundheitsinformationen sowie Sicherheitshinweise/Warmmeldungen („Safety Alerts“) werden vom Institut mit den Institutsinformationen verlinkt und fungieren ebenfalls als Auslöser für die Aktualisierung der Informationen.
- Die restlichen, nicht automatisch aufgefrischten Informationen werden nach Ablauf eines zuvor festgelegten Zeitraums (in der Regel drei Jahre) überprüft.

Überdies hält sich das Institut über methodologische Entwicklungen auf dem Laufenden, um entscheiden zu können, wann die Evidenz für bestimmte Resultate so überzeugend geworden ist, dass es keiner weiteren Aktualisierungen bedarf.

Die auf der Gesundheitsinformationsseite des Instituts eingestellten Informationen werden daher in der Regel mindestens alle drei Jahre überprüft oder aufgefrischt. Aktualisierungen einzelner Themen werden gewöhnlich nicht häufiger als alle sechs Monate durchgeführt.

Die Daten aller Aktualisierungen und geplanten Überprüfungen werden im Internet bekannt gegeben. Die Informationen zu Aktualisierungen werden an alle Abonnenten des Newsletters von www.gesundheitsinformation.de verschickt und online archiviert. Das Institut wird die Zusammenarbeit zum Thema Aktualisierungsverfahren mit seinen Kooperationspartnern fortsetzen. Zu diesen gehören das Centre for Reviews and Dissemination (NHS England), das Centre for Evidence-Based Medicine (Oxford University), die Haute Autorité de Santé (Paris) sowie die Cochrane Collaboration.

5.4 Informationsprodukte

5.4.1 Kernprodukte

Die Gesundheitsinformationen für Patienten und Bürger werden vom Institut in verschiedenen Formaten erstellt. Diese verschiedenen Formate sollen den Bedürfnissen verschiedener

Adressatengruppen gerecht werden, die über einen unterschiedlichen Informationsbedarf, unterschiedliche Lesefertigkeiten und unterschiedlich viel Zeit zum Lesen verfügen.

Zu den Kernprodukten des Instituts gehören:

- Informationsberichte: umfassende Beiträge, die die Grundlage für ein erweitertes Produktmodul zu einem relevanten gesundheitlichen Thema darstellen;
- Merkblätter: kurze, leicht verständliche Informationen; und
- Kurzantworten: knappe Zusammenfassungen von systematischen Reviews, HTA-Berichten oder größeren Studien, darunter auch die vom Institut erstellten.

Zusammengenommen bilden diese Produkte sowie die zusätzlichen Elemente ein evidenzbasiertes „Nachschlagewerk zur Gesundheit“. Studien zeigen, dass Patienten und Bürger sich mehrheitlich vor allem für Informationen über Behandlungen interessieren und für das, was sie selbst aktiv dazu beitragen können [129,226,267,311,337]. Außerdem fühlen sich deutsche Patienten vor allem in Bezug auf diagnostische Untersuchungen und deren Ergebnisse nicht hinreichend informiert [366]. Darin spiegelt sich wider, dass ein Hauptschwerpunkt auf evidenzbasierten Informationen zu Behandlungseffekten, diagnostischen Untersuchungen und Selbstmanagementstrategien liegt. Aus diesem Grund konzentriert sich das Institut vor allem auf die Publikation von Merkblättern und Kurzantworten.

Das Format der Informationsberichte und Merkblätter ähnelt dem konventioneller Patienteninformationen, während die Kurzantworten eher mit Zeitungsartikeln vergleichbar sind.

Informationsberichte sind Artikel, die in der Regel mehr als 20 Seiten umfassen und sich an all diejenigen richten, die an detaillierteren Informationen zu einem Thema interessiert sind. Der Lesbarkeitsgrad dieser recht umfangreichen Beiträge lässt die Zielgruppe aus stärker interessierten und motivierten Lesern erkennen, die sich durch einen allzu einfachen Sprachstil bevormundet sehen könnten [97].

Die Inhalte der einzelnen Informationsberichte können je nach Thema variieren. Zu jedem Thema werden Informationen aus folgenden Bereichen berücksichtigt:

- Erklärungen zur Krankheit oder zu den Beschwerden, einschließlich
 - Anatomie
 - Physiologie
 - verschiedene Erkrankungsformen
 - Krankheitsursachen
 - Symptomererkennung
 - natürlicher Krankheitsverlauf

- Prognose
- mögliche Komplikationen
- Genesung/ Rehabilitation
- mögliches Wiederauftreten der Krankheit (Rezidiv)
- Rezidiverkennung
- Risikogruppen (einschl. Angehörige)
- Maßnahmen zur Prävention und Gesundheitsförderung, einschließlich
 - Ernährung
 - körperliche Aktivität
 - Screeningverfahren
 - Informationen
- diagnostische Optionen, darunter komplementäre Diagnoseverfahren
- therapeutische Optionen, einschließlich
 - Medikamente
 - chirurgische Eingriffe
 - andere nichtmedikamentöse Maßnahmen
- Rehabilitation
- andere Gesundheitsleistungen
- psychosoziale Aspekte, einschließlich persönlicher Berichte von Patienten mit der betreffenden Erkrankung sowie von anderen Betroffenen, darunter Betreuungspersonen, Familienangehörigen und Freunden

Die Merkblätter sind so abgefasst, dass sie für ein breiteres Publikum leicht verständlich sind. Auch hierbei spiegelt sich im Grad der Lesbarkeit der vermutete Gebrauch dieser Art von Informationen. Merkblätter können im Rahmen der Arzt-Patient-Konsultation eingesetzt werden und richten sich an all diejenigen, die an einem raschen Informationsüberblick interessiert sind.

Unter Kurzantworten kann man sich so etwas vorstellen wie wissenschaftlich fundierte FAQ (häufig gestellte Fragen). Deutsche Internetnutzer interessieren sich am meisten für Informationen zu den neuesten Forschungsergebnissen, und diese interessieren sie eventuell noch mehr als die Meinung von Experten [297]. Die Kurzantworten bieten die Möglichkeit, die Evidenz aus hochwertigen wissenschaftlichen Studien in Deutschland besser zugänglich zu machen.

Der Ablauf der Erstellung von Gesundheitsinformationen ist in Abschnitt 2.1.4 beschrieben. Informationen zur Gesundheitsforschung sollten ein ähnlich hohes Qualitätssicherungsniveau

gewährleisten wie der Forschungsbericht selbst [306]. Die Sicherung der Qualität und Relevanz von Patienteninformationen könnte auch im Rahmen von Patientenreviews erfolgen [282]. Informationsberichte und Merkblätter werden deshalb extern begutachtet, u. a. auch von Patienten. Die Qualitätssicherung für alle Kernprodukte umfasst auch, dass die Autoren der systematischen Übersichten die Gelegenheit erhalten, sich zu diesen Patienteninformationen zu äußern. Die Entwürfe zu den Kernprodukten werden dem Institutskuratorium vorgelegt, das den Patientenvertretern in diesem Gremium ebenfalls die Gelegenheit gibt, diese Entwürfe zu kommentieren. Zusätzlich wird eine Nutzertesting durchgeführt. Überdies sind auch die Patienten, die über ihre Krankengeschichte und ihre Genesung berichten, eingeladen, sich zu den Entwürfen von Patienteninformationen, die zu ihren jeweiligen Geschichten gehören, zu äußern (siehe Abschnitt 5.4.3).

5.4.2 Zusätzliche Elemente

Die Kernprodukte werden um verschiedene zusätzliche Produkte (zusätzliche Elemente) ergänzt. Diese zusätzlichen Elemente sollen die Verständlichkeit der Kernaussagen der Gesundheitsinformationen verbessern und das Interesse daran steigern. Die Integration von Bildern, Ton und Animation kann dazu beitragen, die Verständlichkeit der Website zu erhöhen, vor allem für Menschen mit einem niedrigeren Bildungsstand (siehe Abschnitt 5.2.2). Die Animationen stellen die am leichtesten verständlichen Informationsprodukte des Instituts dar.

Möglicherweise bevorzugen und vertrauen die Nutzer auch eher solchen Internetseiten, die attraktiver gestaltet sind und Multimediaelemente enthalten [77,226,337]. Tatsächlich werden hochwertige Inhalte zuweilen allein aufgrund ihrer mangelhaften Präsentation abgelehnt [337].

Als zusätzliche Elemente sind folgende zu verstehen:

- Grafiken, Fotos und anderes Bildmaterial
- kurze Animationen mit Ton und Bild
- interaktive Quizaufgaben
- ein Online-Wörterbuch (Glossar), das ein- oder ausgeschaltet werden kann
- kurze erklärende Texte zu Themen wie etwa der Erkennung von Krankheitszeichen
- Artikel und interaktive Tools zur Erläuterung evidenzbasierter Medizin, zum besseren Verständnis von Forschung und Zahlenangaben
- (Um-)Rechner
- Online-Umfragen und -Fragebögen
- Evidenztabelle
- Entscheidungshilfen

Diese zusätzlichen Elemente sollen

- das allgemeine Verständnis von gesundheitsbezogenen und medizinischen Fragen fördern;
- Patienten und Bürgern helfen, die potenziellen Vor- und Nachteile medizinischer Interventionen zu verstehen und gegeneinander abzuwägen;
- Selbstmanagementstrategien unterstützen.

Der Inhalt der zusätzlichen Elemente ergibt sich im Allgemeinen aus den Kernprodukten. Bei Originalinhalten oder Inhalten wie einer Evidenztabelle oder Entscheidungshilfe muss dieses zusätzliche Element in der Regel denselben Prozess aus externer Überprüfung und externen Stellungnahmen durchlaufen wie die Kernprodukte.

Interaktive zusätzliche Elemente werden institutsintern auf ihre Brauchbarkeit getestet, die bei den Bewertungen der Website durch die Nutzer ebenfalls von entscheidender Bedeutung sein wird. Eine besondere Bedeutung kommt dabei der Barrierefreiheit zu.

Gelegentlich wird das Institut Entscheidungshilfen entwickeln, vor allem in solchen Bereichen, in denen sich Entscheidungshilfen bereits als wirksam erwiesen haben (siehe Abschnitt 5.2.1). Bei der Erstellung einer Entscheidungshilfe orientiert sich das Institut an den International Patient Decision Aids Standards (IPDAS) [113,189].

5.4.3 *Erfahrungsberichte*

Viele Patienten möchten über die Erfahrungen von anderen Betroffenen mit derselben Erkrankung hören oder lesen [183,351]. Erfahrungsberichte sind sowohl im Journalismus als auch auf dem Gebiet der Patienteninformation eine übliche Form der Informationsvermittlung. Wissenschaftliche Evidenz kann unter anderem anhand von Erfahrungsberichten dargestellt und der Öffentlichkeit zugänglich gemacht werden [150]. Die Bedeutung von Erfahrungsberichten in der medizinischen Praxis und im Gesundheitswesen wird zunehmend gewürdigt [156,347,389]. Patienten haben möglicherweise mehr Vertrauen in Gesundheitsinformationswebsites, wenn diese die Erfahrungen von Patienten beinhalten [337].

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

Ein Beispiel für Erfahrungsberichte in Verbindung mit evidenzbasierten Gesundheitsinformationen im Internet ist die Database of Personal Experiences of Health and Illness (DIPEX), die als

evaluierte Multimediawebsite (www.dipex.org) im Internet kostenfrei nutzbar ist [183,184,389]. Das Institut hat seine Methodik zur Erhebung, Bearbeitung und Veröffentlichung von Erfahrungsberichten an den etablierten Vorgehensweisen von DIPEX orientiert.

Erfahrungsberichte werden durch das Institut folgendermaßen erhoben:

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

Besonderen Wert wird sowohl auf die umfassende Aufklärung vor dem Interview, auf eine jederzeit widerrufbare Einverständniserklärung zur Veröffentlichung, auf eine ausführliche Vorbereitung der Interviews, auf ein an vorher festgelegten Kriterien der Gesprächsführung orientiertes Vorgehen der Interviewer sowie auf ein anonymisiertes Vorgehen gelegt. Zu jedem Informationsbericht sollen, wenn möglich, mindestens zwei Interviews geführt werden.

5.4.4 Website

Die Verbreitung der Gesundheitsinformationen des Instituts erfolgt in erster Linie über die zweisprachige Website www.gesundheitsinformation.de bzw. www.informedhealthonline.org. Ziel des Instituts ist die Gewährleistung eines hohen Webstandard im Hinblick auf

- Benutzerfreundlichkeit und Barrierefreiheit [196,240,281] (siehe Abschnitt 5.2.3)
- Datenschutz [199]
- Transparenz
- Sichtbarkeit für Suchmaschinen [355]
- Attraktivität für Nutzer
- Nutzerinteraktivität

Benutzerfreundlichkeit und Nutzerinteraktion sollen dabei auf verschiedenen Wegen erreicht werden, und zwar über die

- Navigation durch die grafische Darstellung des menschlichen Körpers,
- Verlinkung verwandter Themen,
- Online-Bewertung der Qualität einzelner Informationsprodukte sowie über
- Hilfsfunktionen und Funktionen, die eine Tour durch die Website ermöglichen.

Die Website bietet auch einen kostenlosen elektronischen Newsletter, den man wahlweise wöchentlich oder monatlich abonnieren kann. Er enthält Informationen zu Neuigkeiten auf der Website und zur Aktualisierung von Informationen. Die Website ist auch als Version für Handhelds (Personal Digital Assistant, PDA) verfügbar. Das Institut stellt Kerninformationsdaten im RSS-Format bereit, damit man als Nutzer eines RSS-Readers die Informationen der Website verfolgen kann. Darüber hinaus können diese Inhalte so automatisch in externe Websites eingebunden werden.

Wenn Menschen befragt werden, worauf sie bei einer Website Wert legen, werden schwerpunktmäßig häufig Faktoren wie Glaubwürdigkeit und eindeutig nicht kommerzielle Interessen genannt. Das Nutzerverhalten legt aber nahe, dass ein gutes Design und die Attraktivität der Website tatsächlich in großem Ausmaß das Vertrauen der Nutzer erhöhen [34,78,226].

Die Website beruht auf dem Open Source Web Content Management System Papaya [96]. Um höchsten technischen und Datenschutzstandards gerecht werden zu können, sorgt das DIMDI als IT-Plattform für den technischen Betrieb der IQWiG-Website. DIMDI (www.dimdi.de), das Deutsche Institut für Medizinische Dokumentation und Information, ist eine nachgeordnete Behörde des Bundesministeriums für Gesundheit (BMG). Zu seinen Aufgaben gehört es, der fachlich interessierten Öffentlichkeit Informationen aus dem gesamten Gebiet der Medizin zugänglich zu machen.

Die Website des Instituts ist von der Health on the Net Foundation (HON) zertifiziert und erfüllt die acht Prinzipien des HON Code of Conduct (HONcode) für medizinische und Gesundheitswebsites [181]. Diese Stiftung mit Sitz in der Schweiz hat ein Programm zur Überprüfung und Zertifizierung der Standards von medizinischen Websites hinsichtlich Inhalt, Transparenz und Datenschutz entwickelt. Das Institut hat sich für den HONcode entschieden, weil dieses Gütesiegel international anerkannt ist, mehrere Qualitätsdimensionen erfasst und weil die von HON akkreditierten Websites regelmäßig überprüft werden.

5.5 Überwachung und Auswertung

5.5.1 Routinemonitoring

Das Institut überwacht und wertet die Nutzung seiner Gesundheitsinformationswebsite routinemäßig aus, insbesondere

- die Nutzung der Website, einschl. eines Vergleichs mit ähnlichen Websites,
- Nutzerbewertungen und Nutzerfeedback, einschl. Reaktionen auf die laufenden Nutzerumfragen des Instituts,

- den Rang der Informationen im Rahmen von Google-Suchen und den Google PageRank der Website [57],
- die technische Leistungsfähigkeit der Website,
- An- und Abmeldungen von Newsletter-Abonnements und
- die Übernahme der Informationen des Instituts durch Multiplikatoren.

Die häufig zur Bestimmung der Häufigkeit von Websitebesuchen verwendeten Maßeinheiten wie z. B. Treffer (*Hits*) vermitteln einen übertriebenen Eindruck von der Nutzung der Website. Zahlen im Hunderttausenderbereich spiegeln in Wirklichkeit mitunter eine relativ geringe Anzahl von Personen wider, die die Informationen tatsächlich auch lesen. Begriffe wie „Hits“ sind kein verlässlicher Indikator für die Reichweite der Website, da sie eher technische Aspekte der Vermittlung von Internetinformationen messen als die tatsächliche Leserzahl oder Reichweite der Website [196,281]. Das Institut unterscheidet mehrere Hauptkategorien von Webmetriken:

- Messung des Website-„Verkehrs“ (*Traffic*), also der Anzahl von Personen, die die Website gezielt besuchen oder zufällig darauf stoßen
- Bestimmung der „Herkunft“ der Besucher (Suchmaschinen und Links von anderen Websites)
- Messung der Anzahl der besuchten Informationsseiten
- Reichweite und Interaktion mit der Website, einschließlich Suche

Um den Nutzerverkehr mit dem anderer Websites vergleichen zu können, erhebt und wertet das Institut folgende Daten routinemäßig aus [196,197]:

- die Anzahl der von Nutzern abgerufenen Seiten (*Page Impressions*, auch *Page Views*)
- die Summe der einzelnen zusammenhängenden Nutzungsvorgänge (*Visits*)

Page Impressions und *Visits*, ermittelt durch sogenannte Internetroboter (*Crawler*), werden ebenso ausgeschlossen wie die Nutzung der Website durch das Institut selbst und das mit der Entwicklung der Website beauftragte Team. Es wird sorgfältig darauf geachtet, die Daten in einer Form zu erheben, die den Nutzer nicht identifizieren kann. Die Datenschutzrichtlinien des Instituts sind auf der Website detailliert niedergelegt [199].

Der *Website Traffic* stellt die Zahl der Personen dar, die insgesamt die Website überhaupt erreicht haben, nicht die Zahl der tatsächlichen Leser [196]. Noch kritischer überwacht und wertet das Institut daher die Indikatoren für die Anzahl von Personen aus, die die Informationen anscheinend tatsächlich lesen, beispielsweise:

- Informationssuche
- Navigation durch die Artikel

- Anklicken von Glossarbegriffen und das Weiterbesuchen der zugehörigen Informationen
- Betrachten von Animationen oder Lösen der Quizaufgaben
- Herunterladen von PDF-Dateien
- Besuch der Website, um nach dem Empfang des Newsletters neue Informationen zu lesen

Ferner überwacht das Institut den Rang seiner Website bei Google, darunter die Positionierung seiner Informationen im Rahmen der Google-Recherchen, und den Google PageRank der Website selbst. „PageRank“ ist ein zitatabasiertes Verfahren zur Bewertung der Wichtigkeit einer Website [57].

Außerdem wird das Institut den Umfang abschätzen, in dem seine Gesundheitsinformationen durch Multiplikatoren übernommen werden. Dazu muss ausgewertet werden, wie viele andere Websites – vor allem die der Krankenkassen – einen Link zu www.gesundheitsinformation.de geschaltet haben. Darüber hinaus wird die Anzahl der gedruckten Versionen der Gesundheitsinformationen des Instituts überwacht, vor allem derjenigen, die durch die Krankenkassen erstellt wurden.

5.5.2 Feedback, Korrekturen und Verbesserungen

Ebenso, wie das Institut vor der Veröffentlichung von Informationen eine Qualitätssicherung durchführt, fordert es die Leser auch nach der Publikation auf seiner Website ständig zum Einreichen von Verbesserungsvorschlägen auf: Zwischen März und August 2006 haben mehr als 100 Leser Vorschläge zu Verbesserungen unterbreitet oder darum gebeten, zu bestimmten Informationsprodukten weitere Informationen zur Verfügung zu stellen. Darüber hinaus gingen während dieser sechs Monate über die Website mehr als 3000 E-Mails mit Rückmeldungen und Vorschlägen für weitere Themen ein. Das Institut erteilt zwar keine medizinischen Ratschläge, beantwortet keine individuellen Gesundheitsfragen oder gibt andere als die auf dieser Website oder in den sonstigen Publikationen des Instituts veröffentlichten Informationen weiter; dennoch ist dieses Feedback der Leser ein wichtiges Element im Rahmen der Bemühungen des Instituts, die Patientenfreundlichkeit und Nützlichkeit der Gesundheitsinformationen sicherzustellen.

Bei Vorschlägen zur Verbesserung der Gesundheitsinformationen wird zwischen „geringen“, „mittleren“ und „größeren“ Verbesserungen unterschieden. Als geringfügig werden nur sprachliche Verbesserungen eingestuft, während inhaltliche Änderungen zu den größeren Verbesserungen gerechnet werden. Zu den Standardmaßnahmen gehören die Zuordnung zu diesen Kategorien und die Festlegung der für die Korrekturen und Verbesserungen erforderlichen Qualitätssicherungsstufe. Zu jedem Kernprodukt wird online ein Dokumentationsverlauf erstellt, der über das Datum und den Grad der Änderungen Auskunft gibt, und jede Version wird archiviert. Abgesehen von

geringfügigen Änderungen oder dringenden inhaltlichen Korrekturen werden die Informationen im Normalfall nicht häufiger als ein- oder zweimal pro Jahr geändert.

5.5.3 *Evaluation*

Neben dem Routinemonitoring wird das Ausmaß, in dem die Website ihre Ziele erreicht, durch folgende Maßnahmen bewertet:

- Nutzertests
- Online-Umfragen unter den Nutzern
- Beratung mit Nutzern und Patientenvertretern
- unabhängige Evaluation

Bei einigen interaktiven Produkten ist das Testen einzelner Bestandteile von Informationen durch die Nutzer Bestandteil der Arbeitsabläufe vor der Veröffentlichung. Diese Tests werden von den Mitarbeitern des Ressorts „Gesundheitsinformation“ durchgeführt. Offiziellere Nutzertests der Website werden außerhalb des Ressorts durchgeführt. 2006 wurden zwei solcher Nutzertests durchgeführt, einer davon speziell zur Barrierefreiheit. Jedes Jahr sollen weitere externe Nutzertests ablaufen.

Mit der Freischaltung der Website wurde der Online-Fragebogen, der von der Universität Bielefeld zur Evaluierung von Online-Modellprojekten benutzt wird, gemäß § 65b SGB V mit freundlicher Genehmigung auf der Website des Instituts implementiert [323]. Dadurch soll ein Leistungsvergleich (Benchmarking) zwischen www.gesundheitsinformation.de und den offiziell evaluierten Modellprojekten (15 Websites) ermöglicht werden. Die Auswertung der ersten 2561 ausgefüllten Fragebögen auf www.gesundheitsinformation.de ergab, bezogen auf zahlreiche Kriterien wie z. B. die Verständlichkeit, ein Ranking im mittleren Bereich im Vergleich zu den 15 anderen Websites.

Die Verfahren zur Evaluation von Websites gehören zu den fortlaufenden methodologischen Aufgabenbereichen des Ressorts „Gesundheitsinformation“. Für die Evaluation von Gesundheitsinformationen und Gesundheitswebsites stehen zahlreiche Instrumente und Leitlinien zur Verfügung. Viele davon bewerten den Prozessablauf, nicht aber den Inhalt; oder aber den Inhalt und nicht den Prozess [78,81,237]. Bislang gibt es kein Instrument, das sich als zuverlässiger Indikator für die Qualität von Gesundheitsinformationen oder -websites erwiesen hat [124,143,212,269]. Die gebräuchlichen Instrumente weisen größere Lücken auf: DISCERN beispielsweise befasst sich nicht mit der inhaltlichen Qualität [78,237].

Befragungen von Patienten, darunter auch Patienten in Deutschland [125,366], lassen darauf schließen, dass einige Aspekte, die bei der Evaluation von Gesundheitsinformationen manchmal als

wichtig erachtet werden, den meisten Patienten womöglich gar nicht so wichtig sind. Manche in derlei Instrumenten weitverbreiteten Empfehlungen könnten die Qualität von Gesundheitsinformationen vermutlich sogar mindern. Als Beispiel ließe sich hier etwa nennen, dass Links zu anderen Informationen und zu Selbsthilfegruppen erwartet werden, was möglicherweise aber nur eine Minderheit der Nutzer als tatsächlich wichtig ansieht [297]. Eine minderwertige Qualität der über Links zugänglich gemachten Informationen kann zudem zu Fehlinformationen führen. Solche Links regelmäßig daraufhin zu überprüfen, ob sie weiterhin als hochwertig gelten können, ist eine sehr ressourcenintensive Aufgabe, die realistisch gesehen nur wenige Websites werden leisten können.

Die externe Evaluation und vor allem die qualitative Evaluation durch potenzielle Nutzer nehmen hinsichtlich der ständigen Weiterentwicklung der Informationen und der Website des Instituts einen hohen Stellenwert ein. Dabei übernehmen externe Sachverständige die inhaltliche Evaluation einzelner Informationsprodukte und Informationspakete durch potenzielle Nutzer. Weiterhin sollen, wenn möglich, alle Gesprächspartner bei der Erhebung von Erfahrungsberichten in die Evaluation der Informationsentwürfe einbezogen werden, an denen sie mitgearbeitet haben. Die Ergebnisse der Evaluationen fließen kontinuierlich in den Erstellungs- und Überarbeitungsprozess der Informationsprodukte ein.

ABGELÖST DURCH
"Allgemeine Methoden"
Version 4.0
vom 23.09.2011"

6 Allgemeine methodische Aspekte

In der Forschung bezeichnet der Begriff „Bias“ (Verzerrung) eine systematische Abweichung zwischen den Forschungsergebnissen und der „Wahrheit“ [312]. Dabei könnte es sich beispielsweise um eine irrtümlich zu hohe (oder zu niedrige) Einschätzung eines Behandlungseffekts handeln.

Ein Hauptziel bei der Nutzenbewertung medizinischer Versorgungsleistungen ist es, die tatsächlichen Wirkungen von Therapien und Interventionen so zuverlässig und unverzerrt wie möglich abzuschätzen. Zur Minimierung von Bias bei der Nutzenbewertung der medizinischen Versorgung werden international verschiedenste Ansätze verfolgt; dazu gehören etwa die Anwendung wissenschaftlich robuster Methoden, die Gewährleistung einer breiten Beteiligung an den jeweiligen Studien sowie die Vermeidung von Interessenkonflikten [71]. Alle diese Methoden bilden auch die gesetzliche Grundlage für die Arbeit des Instituts.

6.1 Informationsbeschaffung

Die Informationen, die die Grundlage der Institutsberichte darstellen, sind vielfältiger Art (z. B. Ergebnisse wissenschaftlicher Studien zur medizinischen Wirksamkeit und Sicherheit einer Intervention, Registerdaten und andere Datensammlungen oder Unterlagen von Zulassungsbehörden). Im vorliegenden Abschnitt wird der Ablauf einer themenbezogenen Recherche nach wissenschaftlicher Literatur beschrieben.

6.1.1 Allgemeine Grundsätze einer themenbezogenen Literaturrecherche

Eine themenbezogene Literaturrecherche hat zum Ziel, alle für die jeweilige Fragestellung **relevanten**, das heißt zum Erkenntnisgewinn beitragenden, Publikationen zu identifizieren. Die Methodik der systematischen Literaturrecherche folgt daher dem generellen Grundsatz, dass die jeweilige themenbezogene Recherche die folgenden verbindlichen Aspekte berücksichtigen muss, die nach den Ergebnissen diesbezüglicher Forschung einen maßgeblichen Einfluss auf die Beantwortung der Fragestellung haben:

- die Auswahl der Datenquellen (zum Beispiel öffentliche, bibliografische Datenbanken, nicht öffentliche Datenbanken, Handsuche in ausgewählten Fachzeitschriften, Kontaktaufnahme mit Fachexperten/ Industrie/ Patientenorganisationen etc.)
- die Suchtechnik in Bezug auf die Auswahl des Studientyps (zum Beispiel RCT, Fallbeobachtungen etc.)

- die Suchtechnik in Bezug auf die durch die Fragestellung vorgegebenen medizinischen Kriterien (zum Beispiel Zielpopulation, Art der Intervention, Endpunkte etc.)
- die Suchtechnik in Bezug auf formale Merkmale der Publikation (zum Beispiel Abstractpublikationen, Sprache etc.)

Beispiele und Untersuchungen hierzu sind in einer Reihe von Publikationen zu finden [112,126,131,152,158,190,191,220,255,260,264,294,308,315,316,340].

Die Relevanz dieser Kriterien ist für verschiedene Fragestellungen unterschiedlich. Auch die Art des zu erarbeitenden Produkts (zum Beispiel Bericht, Rapid Report, Arbeitspapier) und der sich daraus ergebende Zeitrahmen haben Einfluss auf das Vorgehen bei der Informationsbeschaffung.

Die für das jeweilige Produkt erfolgende Datenbankauswahl orientiert sich an den inhaltlichen, methodischen und regionalen Schwerpunkten der bibliografischen Datenbanken. Die großen biomedizinischen Datenbanken wie beispielsweise MEDLINE werden dabei immer ausgewählt (immer mindestens zwei große biomedizinische Datenbanken). Das Ausmaß der Handsuche in Fachzeitschriften wird bestimmt durch die Existenz relevanter Fachzeitschriften (z. B. Key Journals) und durch das Ausmaß der Indexierung dieser Fachzeitschriften in den berücksichtigten Datenbanken.

Der mindestens auszuwählende Studientyp wird in den Einschlusskriterien definiert. Für die Informationsbeschaffung werden dieser und alle höherwertigen Studientypen ausgewählt.

Die medizinischen Kriterien der Fragestellung orientieren sich themenabhängig an dem Ziel einer hohen Sensitivität und dem resultierenden Grundsatz „so viele Recherchekomponenten wie nötig, so wenige wie möglich“. Suchstrategien für Arzneimittelbewertungen enthalten tendenziell weniger Komponenten als solche für nichtmedikamentöse Verfahren.

6.1.2 Ablauf einer Recherche nach Primärpublikationen

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

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

8. Speicherung der Ergebnisse in Textdateien und Import in ein Literaturverwaltungsprogramm
9. Dokumentation der Recherche

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

Für die Identifizierung der Suchwörter und die Formulierung der Suchstrategie werden i. d. R. durch die Vorrecherche identifizierte themenrelevante Publikationen herangezogen. Im Sinne einer Qualitätssicherung wird geprüft, ob die so entwickelte Suchstrategie bekannte relevante Primärpublikationen („Testset“) hinreichend sicher identifiziert. Das Testset wird durch Rückgriff auf Vorarbeiten anderer Arbeitsgruppen (HTA-Berichte, systematische Übersichten zum Thema) generiert.

Gegebenenfalls können die identifizierten HTA-Berichte und systematischen Übersichten auch herangezogen werden, um die vom Institut durchgeführte Suche auf den nicht von diesen Arbeiten abgedeckten Bereich (z. B. Suchzeitraum) zu beschränken. Notwendige Voraussetzung hierfür ist, dass die jeweilige Suche der Methodik des Instituts nicht widerspricht und die Übertragbarkeit der Ergebnisse auf die jeweilige Fragestellung, insbesondere unter Berücksichtigung der im Berichtsplan genannten Ein- und Ausschlusskriterien, möglich ist.

6.1.3 Sonstige Datenquellen für die Recherche

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

Zusätzlich sind je nach Fragestellung weitere Datenquellen von erheblicher Bedeutung, wie zum Beispiel Studienregister oder Abstractbände wissenschaftlicher Kongresse. Im Falle einer Arzneimittelbewertung, aber auch für die Bewertung bestimmter (nichtmedikamentöser) Medizinprodukte sind öffentlich zugängliche Zulassungsdatenbanken bzw. -korrespondenzen weitere mögliche Quellen für die Informationsgewinnung. Darüber hinaus wird in der Regel bei den Herstellern der zu bewertenden Technologie(n) bezüglich bislang nicht veröffentlichter Informationen angefragt. Ziel dieser Anfrage ist die Identifizierung aller für die Nutzenbewertung relevanten Studien bzw. Informationen unabhängig vom Publikationsstatus. Im Rahmen der Arzneimittelbewertung wird diese Anfrage in der Regel in einem zweistufigen Verfahren durchgeführt. In der ersten Stufe fragt das Institut beim Hersteller eine vollständige Übersicht über alle Studien, die vom Hersteller mit dem zu bewertenden Arzneimittel durchgeführt wurden, an. Gegebenenfalls definiert das Institut projektspezifische Einschlusskriterien für diese Übersicht. Aus der Übersicht der Studien

identifiziert das Institut die Studien, die für die Nutzenbewertung relevant sind, und fragt zu diesen Studien Detailinformationen an. Dabei kann es sich um eine Anfrage nach unveröffentlichten Studien oder nach ergänzenden, bislang unveröffentlichten Informationen zu veröffentlichten Studien handeln. Auch bisher unveröffentlichte Informationen, die in die Nutzenbewertung einfließen, werden in den Berichten des Institutes publiziert, um die Transparenz zu gewährleisten. Basis für die Einbeziehung bislang unveröffentlichter Informationen in die Nutzenbewertung ist der Abschluss einer Vereinbarung zur Übermittlung und Veröffentlichung von Studieninformationen, die vor der Einreichung von Daten zwischen dem Institut und dem beteiligten Hersteller getroffen wird (siehe Mustervertrag [198]). Diese Vereinbarung spezifiziert den Verfahrensablauf, die Anforderungen an die zu übermittelnden Dokumente sowie die vertraulichen bzw. nicht vertraulichen Anteile der übermittelten Dokumente. Falls das jeweilige Unternehmen einem Abschluss dieser Vereinbarung und damit insbesondere der vollständigen Übermittlung aller vom Institut angefragten Informationen nicht zustimmt, oder der vollständigen Übermittlung der angefragten Informationen trotz Abschluss einer Vereinbarung nicht nachkommt, werden keine weiteren Anfragen an den Hersteller gerichtet. Hierdurch soll eine Verzerrung der Ergebnisse durch selektive Bereitstellung von Informationen vermieden werden.

Sofern vom jeweiligen Auftraggeber (GBA, BMG) Unterlagen zur Berücksichtigung übermittelt werden, werden diese als ein Bestandteil der Informationsgewinnung angesehen. Mit den Unterlagen wird dann im weiteren Prozess den sonstigen Grundsätzen der Informationsrecherche und -bewertung folgend verfahren.

6.1.4 Auswahl relevanter Publikationen

Die Literaturrecherche in bibliografischen Datenbanken führt aufgrund des primär sensitiven Vorgehens dazu, dass eine große Zahl der identifizierten Zitate nicht für die Bewertung relevant ist. Die Auswahl der relevanten Publikationen erfolgt über mehrere Selektionsschritte.

- Anhand des Titels und, sofern vorhanden, Abstracts werden zunächst sicher nicht relevante Publikationen (d. h. Publikationen, die nicht die Ein- / Ausschlusskriterien des Berichtsplans bzw. der Projektskizze erfüllen) ausgeschlossen. Dieser Schritt kann für die Trennung gänzlich irrelevanter von themenbezogenen, aber nicht den Ein- / Ausschlusskriterien genügenden Publikationen in zwei Arbeitsschritte unterteilt werden. Themenbezogen sind z. B. Studien eines anderen Studientyps oder Studien mit einer anderen Laufzeit als im Berichtsplan bzw. der Projektskizze gefordert wurde.

- Für die verbleibenden, potenziell relevanten Publikationen wird der Volltext beschafft, auf dessen Basis dann die Entscheidung über den Einschluss in die jeweilige Bewertung getroffen wird.

Alle Selektionsschritte werden grundsätzlich von zwei Personen unabhängig voneinander durchgeführt. Diskrepanzen werden durch Diskussionen aufgelöst. Beim ersten Selektionsschritt wird im Zweifelsfall der zugehörige Volltext bestellt und bewertet. Auch hier können gänzlich irrelevante von themenbezogenen Publikationen unterschieden werden.

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

6.1.5 Dokumentation

Alle Arbeitsschritte der bibliografischen Literaturrecherche werden dokumentiert. Hierzu gehören insbesondere:

- die Suchstrategien für die ausgewählten Datenbanken
- das Datum der Suche
- die Suchoberfläche
- die Trefferzahl
- eine Auflistung der nach Sichtung der Ergebnisse als für die Fragestellung relevant erachteten Publikationen (Zitate)
- eine Auflistung der nach Sichtung des Volltextes als nicht relevant erachteten Literaturstellen, alternativ eine Auflistung der themenbezogenen, aber für den Bericht nicht relevanten Publikationen, jeweils mit Angabe einer Begründung für die Nichtberücksichtigung

Alle anderen Schritte der Informationsbeschaffung werden ebenfalls dokumentiert (z. B. Korrespondenz mit Autoren, Anfragen an Hersteller etc.).

6.1.6 Literaturmonitoring

Neben der themenbezogenen, rückschauenden Recherche ist das frühzeitige Erkennen und Bewerten aktueller und relevanter Publikationen notwendig; Grundlage hierfür bildet die

systematische diesbezügliche Beobachtung wesentlicher wissenschaftlicher 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 Prioritäten- und Projektliste berührt.

Die einzelnen Fachressorts werten 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, aber auch die Erstellung eines Arbeitspapiers initiieren. Hierüber entscheidet zuvor das Steuergremium des Instituts.

6.1.7 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 zur Einhaltung der gesetzlichen Bestimmungen abzugeben; andernfalls ist vorab jeweils im Einzelfall sorgfältig zu prüfen, inwieweit die notwendigen gesetzlichen Voraussetzungen vorliegen (Einverständniserklärungen, Aufklärung etc.).

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

Werden dem Institut Daten übermittelt, die nicht publiziert werden dürfen, so können diese nicht inhaltlich in Nutzenbewertungen des Instituts einfließen, da dies dem Transparenzgebot widerspricht.

Hinsichtlich der vertraulichen Behandlung von Unternehmensdaten sind gegebenenfalls geeignete Erklärungen gegenüber Dritten abzugeben, die eine Wahrung der Vertraulichkeit gewährleisten

(siehe auch Abschnitt 6.1.3). 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 eingehen.

6.2 Qualitätsbewertung von Einzelstudien

6.2.1 *Bezug zwischen Studientyp/-art und Fragestellung*

Es werden hier nur die wichtigsten Designs zusammengefasst, die bei der Nutzenbewertung in der medizinischen Forschung in Abhängigkeit von der Fragestellung eine Rolle spielen.

Für die Nutzenbewertung von Interventionen ist an erster Stelle eine Kontrollgruppe zu fordern. Aus einem reinen Vorher-nachher-Vergleich in einem Design mit abhängigen Stichproben ohne Kontrollgruppe lässt sich in der Regel kein Beleg für einen Effekt einer Intervention ableiten. Ausnahmen bilden Krankheitsbilder mit einem deterministischen (oder nahezu deterministischen) Verlauf (z. B. diabetisches ketoazidotisches Koma; siehe Abschnitt 3.2.2). Gütekriterien, die die Aussagekraft kontrollierter Studien erhöhen, sind Randomisierung und Verblindung. Häufige Designs bei klinischen Studien bilden parallele Gruppen [295], Crossover-Studien [217] und clusterrandomisierte Studien [102]. Sind in einer Studie Zwischenauswertungen geplant, so muss dies mithilfe eines adäquaten sequenziellen Designs berücksichtigt werden [379].

Den ersten Informationsgewinn gibt es häufig aus Fallberichten oder Fallserien. Diese sind anfällig für Verzerrungen aller Art, sodass hier, je nach Fragestellung, nur sehr beschränkt eine verlässliche Evidenz ableitbar ist. Bevölkerungsbezogene Querschnittstudien sind geeignet zur Schätzung von Krankheitsprävalenzen. Weitere grundlegende klassische Studientypen der Epidemiologie sind Fall-Kontroll-Studien [55] zur Untersuchung des Zusammenhangs zwischen Expositionen und seltenen Erkrankungen und Kohortenstudien [56] zur Erforschung des Effekts von Expositionen im Zeitverlauf. Kohortenstudien sind in diesem Sinne prospektiv angelegt, allerdings gibt es auch retrospektive Kohortenstudien, in denen die Exposition aus der Vergangenheit erfasst wird (häufig in der Arbeits- oder auch Pharmakoepidemiologie anzutreffen). Grundsätzlich sind prospektive Studien retrospektiven Designs vorzuziehen. Allerdings sind zum Beispiel Fall-Kontroll-Studien häufig die einzige praktikable Möglichkeit, Informationen über Zusammenhänge zwischen Expositionen und seltenen Erkrankungen zu gewinnen. Neuere Studiendesigns der modernen Epidemiologie enthalten Elemente sowohl von Fall-Kontroll-Studien als auch von Kohortenstudien und sind nicht mehr eindeutig als retrospektiv oder prospektiv zu klassifizieren [223].

Diagnose- und Screeningstudien können sehr unterschiedliche Ziele haben, sodass die Bewertung von der Wahl eines adäquaten Designs abhängt (siehe Abschnitte 3.5 und 3.6).

6.2.2 Rangordnung verschiedener Studienarten/ Evidenzgrade

Es gibt verschiedene Ansätze, bestimmten Studientypen bestimmte Evidenzgrade zuzuordnen, um so eine Rangfolge bezüglich der Validität der Evidenz aus verschiedenen Studienarten im Rahmen einer systematischen Übersicht oder für die Entwicklung von Leitlinien zu erstellen [169,173]. Es gibt jedoch kein System der Evidenzbewertung, das allgemein akzeptiert und universell in allen systematischen Übersichten einsetzbar ist [224]. Aufgrund der Komplexität der Studienbewertung lässt sich anhand einer Evidenzhierarchie auch kein abschließendes Qualitätsurteil bilden. Das Institut orientiert sich im Allgemeinen an folgender grober Hierarchie von Studientypen, die weitgehend akzeptiert wird und auch weitgehend mit der Evidenzklassifizierung des G-BA übereinstimmt [145]. Zumindest im Rahmen von Therapiestudien wird der höchste Evidenzgrad RCTs und systematischen Übersichten von RCTs zugeordnet. Einzelne RCTs werden in einigen Einteilungen noch in solche hoher oder weniger hoher Qualität eingestuft, wobei die Vermengung von Konzept- und Ergebnisqualität von anderen Autoren wiederum kritisiert wird [384]. Auf den nächsten Plätzen folgen nicht randomisierte Interventionsstudien, prospektive Beobachtungsstudien, retrospektive Beobachtungsstudien, nicht experimentelle Studien (Fallserien und Fallberichte) und – mit niedrigstem Evidenzgrad – Expertenmeinungen ohne wissenschaftliche Begründung. Dieses grobe System wird das Institut der jeweiligen Situation und Fragestellung anpassen und bei Bedarf detaillierter darstellen [173].

6.2.3 Aspekte der Bewertung des Verzerrungspotenzials von Ergebnissen

Ein wesentlicher Gesichtspunkt bei der Interpretation von Studienergebnissen ist die Einschätzung des Verzerrungspotenzials (vgl. qualitative Ergebnisunsicherheit, Abschnitt 3.2.1). Dabei spielen die Fragestellung, die Art, das Design und die Durchführung der Studie sowie die Verfügbarkeit der Informationen eine Rolle. Das Verzerrungspotenzial wird von der Qualität der Studien zwar maßgeblich beeinflusst, dessen Bewertung ist jedoch nicht gleichbedeutend mit einer Qualitätseinschätzung von Studien. Beispielsweise können einzelne Endpunkte auch in einer qualitativ hochwertigen Studie möglicherweise bedeutsam verzerrt sein. Andererseits können Studien trotz niedriger Qualität im Einzelfall für bestimmte Endpunkte Resultate hoher Ergebnissicherheit liefern. Das Institut wird daher i. d. R. für alle relevanten Ergebnisse das Ausmaß des Verzerrungspotenzials problemorientiert sowohl für die Studie als auch endpunktspezifisch einschätzen.

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) [202,234] und bei epidemiologischen Studien die Leitlinien und Empfehlungen zur Sicherung *Guter Epidemiologischer Praxis* (GEP) [91]. Ein zentrales Kriterium zur Vermeidung von Verzerrungen ist dabei, ob die Studie auch so ausgewertet wurde wie in der Planung vorgesehen. Dies ist Publikationen in der Regel nur sehr unzuverlässig zu entnehmen. Ein Abschnitt über die Fallzahlplanung kann hierbei jedoch zumindest Hinweise liefern. Darüber hinaus ist ein Vergleich mit dem eventuell zuvor veröffentlichten Studienprotokoll bzw. einer Publikation zum Studiendesign sinnvoll.

Wichtige Stellungnahmen, die zur Verbesserung der Qualität von Publikationen erarbeitet wurden, sind das CONSORT (*Consolidated Standards of Reporting Trials*)-Statement für randomisierte klinische Versuche [265] und das dazugehörige Erläuterungsdokument [12], ein Vorschlag zur Erweiterung des CONSORT-Statements für randomisierte Studien bei nichtmedikamentösen Verfahren [53] und das zugehörige Erläuterungsdokument [52], das CONSORT-Statement für clusterrandomisierte Studien [65], das QUOROM (*Quality of Reporting of Meta-analyses*)-Statement für Meta-Analysen randomisierter Studien [263], das TREND (*Transparent Reporting of Evaluations with Non-randomized Designs*)-Statement für nicht randomisierte Interventionsstudien [89], das STROBE (*Strengthening the Reporting of Observational Studies in Epidemiology*)-Statement für Beobachtungsstudien in der Epidemiologie [373] und das dazugehörige Erläuterungsdokument [369], die MOOSE (*Meta-analysis of Observational Studies in Epidemiology*)-Checkliste für Meta-Analysen von Beobachtungsstudien in der Epidemiologie [349], das STARD (*Standards for Reporting of Diagnostic Accuracy*)-Statement für Diagnosestudien [50] und das dazugehörige Erläuterungsdokument [51]. Werden in einer Publikation diese Standards nicht eingehalten, so kann dies ein Indikator für ein erhöhtes Verzerrungspotenzial der Ergebnisse der entsprechenden Studie sein. Weitere wichtige Quellen, die grundlegende Aspekte zur Bewertung des Verzerrungspotenzials beschreiben, sind die Standardwerke [110,168,185].

Zentrale Aspekte zur Bewertung des Verzerrungspotenzials der Ergebnisse randomisierter kontrollierter Studien (RCTs) durch das Institut stellen ein adäquates Concealment, das heißt die Unvorhersehbarkeit und Verheimlichung der Gruppenzuteilung (zum Beispiel durch externe Randomisierung bei nicht verblindbaren Studien), die verblindete Zielgrößenerhebung bei Studien, in denen Arzt und Patient nicht verblindbar sind, die adäquate Anwendung des *Intention-to-Treat*-Prinzips sowie die Festlegung auf ein eindeutiges Hauptzielkriterium beziehungsweise die angemessene Berücksichtigung eines möglichen Multiplizitätsproblems dar.

Der Effekt einer Intervention im Ganzen setzt sich aus verschiedenen Komponenten zusammen: dem eigentlichen, alleine der Intervention zuzuschreibenden Effekt, dem natürlichen Verlauf der Erkrankung und dem sogenannten Placeboeffekt [354], der den Kontext, in dem die Behandlung stattfindet, beschreibt. Hierunter werden alle Einflüsse zusammengefasst, die bei einer Intervention begleitend auftreten, wie zum Beispiel die Erwartungshaltung der Patienten und der Behandler, die Suggestion der Behandler oder auch Effekte, die alleine darauf zurückzuführen sind, dass Patienten betreut werden. Placebokontrollierte Studien dienen dazu, diese begleitenden Effekte in beiden Gruppen gleich stark einwirken zu lassen, indem eine Gruppe mitgeführt wird, die eine Placebobehandlung erfährt. Entscheidend hierbei ist die Möglichkeit einer Verblindung von Patienten und Behandlern bezüglich der Intervention. Die Verblindung der Behandler soll gewährleisten, dass die über die Intervention hinausgehende Betreuung, Zuwendung und Suggestion auf die Studienpatienten gleich verteilt werden.

Eine Placebobehandlung ist nicht allein auf medikamentöse Interventionen beschränkt, sondern kann auch bei nichtmedikamentösen Verfahren eingesetzt werden, um zumindest eine Verblindung der Patienten zu erreichen. In diesem Zusammenhang wird auch von sogenannten Sham-Interventionen gesprochen. Das Ausmaß eines Placeboeffekts kann von der Art der Intervention abhängig sein [222].

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

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

Es existieren Vorschläge, die Verblindung am Ende einer Studie zu überprüfen, indem verglichen wird, wie viele Patienten und Behandler die zugeordnete Gruppe korrekt erraten haben [130]. Solche Überprüfungen bergen allerdings bisher noch nicht befriedigend gelöste und kontrovers diskutierte methodische Probleme (z. B.: Welche Verfahren sollen für welche Nullhypothesen eingesetzt werden? [11,313,333]). Im Falle wirksamer Therapien mit für Patienten direkt erfahrbaren Wirkungen ist eine Entblindung in einem gewissen Umfang möglich oder gar wahrscheinlich. Das heißt in einer solchen Situation ist es schwierig oder unmöglich, zu beurteilen, ob durch eine Entblindung (welcher Ursache auch immer) eine Verzerrung des Therapieeffekts zustande gekommen ist oder umgekehrt der Therapieeffekt zu einer Entblindung geführt hat. Ungeachtet dieser Probleme ist die Überprüfung der Verblindung in einer Studie zu begrüßen, da sie ein Indiz dafür liefert, dass das Thema Verblindung während der Planung und Durchführung der Studie angemessen berücksichtigt wurde.

Die Interpretation der Ergebnisse von unverblindeten Studien beziehungsweise Studien, in denen es (möglicherweise) zu einer Entblindung kam, muss im Vergleich zu verblindeten Studien vorsichtiger erfolgen. Wichtige Instrumente zur Vermeidung von Bias in Studien, in denen eine Verblindung der Intervention nicht möglich ist, sind wiederum die Randomisierung und die Wahl geeigneter Zielvariablen. In nicht verblindbaren Studien ist es zentral wichtig, dass eine adäquat verdeckte Zuteilung (Allocation Concealment) der Patienten zu den zu vergleichenden Gruppen gewährleistet ist. Weiterhin ist es erforderlich, dass die Zielvariable unabhängig vom (unverblindeten) Behandler ist bzw. unabhängig vom Behandler erhoben wird (verblindete Zielgrößenerhebung). Falls eine verblindete Zielgrößenerhebung nicht möglich ist, sollte ein möglichst objektiver Endpunkt gewählt werden, der in seiner Ausprägung und in der Stringenz der Erfassung so wenig wie möglich durch denjenigen, der den Endpunkt (unverblindet) erhebt, beeinflusst werden kann.

Tritt der Fall ein, dass bei einer Projektarbeit des Instituts ältere Studien bewertet werden müssen, die den heute gültigen Qualitätsansprüchen nicht genügen, weil sie in einer Zeit geplant und durchgeführt wurden, zu der diese Ansprüche noch nicht existierten, so zeigt das Institut die Nachteile und Mängel dieser Studien auf und diskutiert mögliche Konsequenzen. Ein unterschiedlicher Umgang im Vergleich zu neueren Studien, die ähnliche Qualitätsmängel haben, ist jedoch nur dann nötig, wenn dies aus der Fragestellung oder anderen Umständen der Bewertung heraus klar begründbar ist.

Die Bewertung formaler Kriterien liefert wesentliche Anhaltspunkte für das Verzerrungspotenzial der Ergebnisse von Studien. Das Institut nimmt jedoch immer eine über rein formale Aspekte hinausgehende Bewertung des Verzerrungspotenzials vor, um beispielsweise Fehler, Widersprüche

und Ungereimtheiten in Publikationen darzustellen, und prüft deren Relevanz für die Ergebnisinterpretation.

6.2.4 Interpretation von Subgruppenanalysen

Subgruppenanalysen werden in der methodischen Literatur sehr kritisch diskutiert [21,287]. Die Interpretation ihrer Ergebnisse wird im Wesentlichen durch drei Aspekte erschwert:

- **Kein Beweischarakter:** Subgruppenanalysen werden selten a priori geplant und sind selten Bestandteil des Studienprotokolls (ggf. auch von Amendments). Wenn Subgruppenanalysen post hoc stattfinden, können ihre Ergebnisse nicht als methodisch korrekte Prüfung einer Hypothese betrachtet werden.
- **Multiples Testen:** Bei der Analyse zahlreicher Subgruppen besteht eine manchmal recht hohe Wahrscheinlichkeit, dass Ergebnisse irgendeiner Subgruppe statistische Signifikanz erreichen, obwohl es sich in Wahrheit um ein zufälliges Ergebnis handelt.
- **Geringe Power:** Oft erreichen Subgruppen nicht die Stichprobengröße, die für das (inferenzstatistische) Aufdecken moderater Unterschiede nötig ist, sodass auch beim Vorliegen tatsächlicher Effekte mit nicht signifikanten Ergebnissen zu rechnen ist. Anders ist die Situation, wenn schon bei der Fallzahlkalkulation eine für die Subgruppenanalyse angemessene Power berücksichtigt und ein entsprechend erhöhter Stichprobenumfang geplant wurde [60].

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

Darüber hinaus sind Subgruppenanalysen nicht interpretierbar, wenn das subgruppenbildende Merkmal nach Therapiebeginn (nach Randomisierung) erhoben wurde, z. B. bei sogenannten Responderanalysen.

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

Ungeachtet der oben getroffenen Einschränkungen können für manche Fragestellungen Subgruppenanalysen die auf absehbare Zeit beste wissenschaftliche Evidenz für die Bewertung von Effekten in eben diesen Subgruppen darstellen [139], da z. B. ethische Überlegungen gegen eine

Reproduktion der entsprechenden Ergebnisse in einer Validierungsstudie sprechen können. Einen Überblick über Indikationen zur Anwendung von Subgruppenanalysen gibt Rothwell [305].

Eine wichtige Indikation für Subgruppenanalysen stellt eine mögliche Heterogenität des Effekts in verschiedenen, klar voneinander abgrenzbaren Patientenpopulationen dar [238,305]. Gibt es a priori Kenntnis von einem möglichen Effektmodifikator (z. B. Alter, Pathologie), so ist es sogar erforderlich, eine mögliche Heterogenität bezüglich des Effekts in den verschiedenen Patientengruppen vorab zu untersuchen. Ist eine solche vorhanden, so ist der über alle Patienten hinweg geschätzte Gesamteffekt nicht sinnvoll interpretierbar [238]. Daher ist es wichtig, dass die Kenntnis einer möglichen Heterogenität der Patientengruppen beim Studiendesign adäquat berücksichtigt wird. Möglicherweise ist sogar die Durchführung mehrerer Studien notwendig [164]. Im Rahmen von systematischen Übersichten ist die Analyse der Heterogenität zwischen den einzelnen Studien eine wissenschaftliche Notwendigkeit.

Den Goldstandard für Subgruppenanalysen stellen Auswertungen bezüglich a priori gebildeter Subgruppen dar, wobei eine anhand der Subgruppen stratifizierte Randomisierung und eine für die Datenanalyse adäquate statistische Methode (Homogenitätstest, Interaktionstest) vorgenommen bzw. angewandt werden [80].

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

6.2.5 Bewertung von Prognosestudien

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

Ein prognostisches Merkmal liefert eine Information, die letztlich keinen Selbstzweck haben kann, sondern zu einer Konsequenz führen sollte, die ihrerseits einen prüfbaren Nutzen für den Patienten bedeutet. In diesem Zusammenhang sind an ein Prognoseverfahren ähnliche (allgemeine) Anforderungen wie an einen diagnostischen Test zu stellen. Soll ein prognostisches Merkmal im Sinne eines Screening- bzw. Präventionsprogramms eingesetzt werden, sind bei der Bewertung die in Abschnitt 3.6 formulierten Grundsätze zu berücksichtigen.

Für die Bewertung von Prognosestudien gibt es keine allgemein akzeptierten Qualitätskriterien [8,178,339]. Richtlinien zur Planung und Durchführung von Prognosestudien in der Onkologie werden von Simon und Altman [339] beschrieben. Ein allgemeiner Rahmen zur Qualitätsbewertung von Prognosestudien wird von Laupacis et al. [244] vorgeschlagen. Hayden et al. [178] entwickeln Richtlinien zur Qualitätsbewertung von Prognosestudien in Bezug auf potenzielle Biasquellen. Zu beachten sind in jedem Fall unten genannte Aspekte, die sich aus der zugrunde liegenden Datenquelle sowie der verwendeten Datenanalyse ergeben. Da bei Prognosestudien häufig multifaktorielle Regressionsmodelle eine zentrale Rolle spielen, ist außerdem Abschnitt 6.4.6 zu beachten. Speziell bedeutsam sind folgende Aspekte:

- klare Formulierung einer Fragestellung und damit verbunden der Studienplanung; dies schließt eine Fallzahlplanung ein, die sich beispielsweise an der gewünschten Präzision der Schätzung (Breite des Konfidenzintervalls) orientieren kann und eine Abschätzung sowohl der Prävalenz der Exposition als auch der Inzidenz im Hinblick auf die interessierende Zielvariable erfordert
- klare Beschreibung der Ziel- und Stichprobenpopulation (zum Beispiel Bevölkerungs- oder Hausarztbezug, Register) und Begründung für deren Wahl
- klare Beschreibung der Auswahl und des Vorgehens bei der Rekrutierung der Studienteilnehmer
- Homogenität der betrachteten Population; falls es sich um eine heterogene Population handelt, sollte darauf geachtet werden, dass die prognostische Aussage möglichst konstant über die die Heterogenität verursachenden Subgruppen (beispielsweise unterschiedliche Basisrisiken für die interessierende Zielvariable) getroffen werden kann
- klare Definition einer (oder mehrerer) Zielvariablen, an der (denen) sich die prognostische Wertigkeit orientieren soll
- klare Definition der prognostischen Merkmale einschließlich der statistischen Handhabung (zum Beispiel Dichotomisierung oder Betrachtung von Terzilen, Quartilen etc. bei einem quantitativen Merkmal) und einer Begründung für das gewählte Vorgehen
- klare Festlegung und Definition von möglichen Störgrößen und Wechselwirkungen einschließlich deren statistischer Handhabung
- bei Kohortenstudien Vollständigkeit des Follow-ups beziehungsweise Maßnahmen zur Erzielung eines möglichst vollständigen Follow-ups; 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 stellen Kohortenstudien und Fall-Kontroll-Studien dar. In Ausnahmefällen, zum Beispiel bei der Betrachtung unveränderlicher Merkmale, können auch Querschnittstudien eine Rolle spielen. Die grundlegenden Prinzipien für die Bewertung solcher Studien über die oben genannten Aspekte hinaus sind in Abschnitt 6.2.3 formuliert.

Die Literatursuche bei der Evaluierung von prognostischen Merkmalen (im Rahmen einer systematischen Übersicht) ist schwieriger als beispielsweise bei Therapiestudien und es existiert (noch) keine allgemein akzeptierte optimale Suchstrategie. Darüber hinaus wird angenommen, dass dieser Bereich besonders anfällig für einen Publikationsbias ist [8,339]. Die methodische Qualität von Studien bzw. derer Publikationen zu prognostischen Merkmalen ist häufig unzureichend [296], sodass die Extraktion von benötigten Daten erschwert oder sogar unmöglich wird. Insofern sind Meta-Analysen (nicht jedoch systematische Übersichten an sich) von Prognosestudien oft unangemessen und deren Ergebnisse nur unter Vorbehalt zu verwerten [8]. Einige wichtige Probleme bei Meta-Analysen von Prognosestudien lassen sich vermeiden, wenn individuelle Patientendaten zur Verfügung stehen [8].

Neben Ergebnissen aus Studien, die einzelne oder (zumeist) mehrere prognostische Merkmale untersuchen, werden für die individuelle Risikoabschätzung von Patienten (oder allgemeiner Personen), ein unerwünschtes Ereignis zu erleiden, zunehmend sogenannte *Risk Charts* (auch als *Risk Engines* bezeichnet) eingesetzt, bei denen für das Zusammentreffen mehrerer Risikofaktoren multifaktorielle Schätzwerte angegeben werden (z. B. *Sheffield Table* [375] oder *Joint British Chart* [58]). Die Grundlage dieser Risikotabellen bilden meist multifaktorielle Regressionsmodelle, deren Ergebnisse zur einfacheren Handhabung in Tabellen oder Punktesysteme überführt werden [350]. Zu beachten ist, dass aus solchen Risikotabellen abgeleitete Risiken keine „persönlichen“ Schätzungen für spezielle Individuen sind, sondern statistische Schätzungen für mittlere Risiken einer Populationsgruppe mit speziellen Risikoprofilen für einen definierten Zeitraum (z. B. zehn Jahre) darstellen. Bei der Bewertung solcher Instrumente ist zu berücksichtigen,

- für welche Population die geschätzten Risiken gelten,
- aus welcher Art von Studien die zugrunde liegenden Daten stammen,
- ob die herangezogenen Merkmale in diesen Studien gemeinsam betrachtet wurden,

- ob und, wenn ja, wie eine multifaktorielle statistische Analyse in diesen zugrunde liegenden Studien erfolgte und
- ob diese Instrumente jemals in nachfolgenden Studien validiert wurden (Teststichproben).

6.2.6 Bewertung der Konsistenz publizierter Daten

Zur Einschätzung der Aussagekraft publizierter Ergebnisse überprüft das Institut die Daten auf Konsistenz (Plausibilität, Vollständigkeit). Unplausible Daten entstehen zum einen durch eine fehlerhafte Darstellung der Ergebnisse (Tipp-, Formatierungs- oder Rechenfehler), aber auch durch eine mangel- beziehungsweise fehlerhafte Beschreibung der Methodik oder sogar durch gefälschte oder erfundene Daten [7]. Inkonsistenzen können innerhalb einer Publikation bestehen, aber auch zwischen verschiedenen Publikationen zu ein und derselben Studie.

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

Es ist daher angezeigt, mögliche Inkonsistenzen in den Daten aufzudecken. Dazu überprüft das Institut beispielsweise Rechenschritte und vergleicht die Angaben im Text und in den Tabellen und Grafiken. Ein in der Praxis häufiges Problem bei Überlebenszeitanalysen sind Inkonsistenzen zwischen den Angaben zu Lost-to-Follow-up-Patienten und denen zu Patienten unter Risiko bei der grafischen Darstellung von Überlebenskurven. Bei bestimmten Endpunkten, zum Beispiel der Gesamtmortalität, kann die Anzahl der Lost-to-Follow-up-Patienten berechnet werden, wenn die Kaplan-Meier-Schätzer mit den Patienten unter Risiko zu einem Zeitpunkt vor der minimalen Follow-up-Zeit abgeglichen werden. Bei der Aufdeckung erfundener oder gefälschter Daten können statistische Techniken hilfreich sein [7].

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

beeinflussen, wird dieser Sachverhalt dargestellt und die Ergebnisse werden sehr zurückhaltend interpretiert.

6.3 Berücksichtigung von Sekundärliteratur

Unter dem Begriff „Sekundärliteratur“ werden hier Arbeiten bzw. deren Publikationen verstanden, die auf systematische, reproduzierbare und transparente Weise Ergebnisse von Primärstudien zusammenfassend darstellen und bewerten. Dies sind sowohl systematische Übersichten als auch HTA-Berichte, in denen normalerweise versucht wird, eine klinische und/ oder eine für Patienten relevante Frage zu beantworten. HTA-Berichte versuchen oftmals auch, Antworten auf zusätzliche Fragen zu geben, die für die Auftraggeber/ Entscheidungsträger von Interesse sind [103,245,291]. Da im Folgenden nicht zwischen systematischen Übersichten und HTA-Berichten unterschieden werden muss, bezieht der Begriff „systematische Übersicht“ in diesem Abschnitt HTA-Berichte mit ein.

6.3.1 Einordnung systematischer Übersichten

Auf die Ergebnisse einzelner wissenschaftlicher Studien zu vertrauen, kann irreführend sein. Lediglich eine oder nur manche Studien isoliert von anderen, ähnlichen Studien zu derselben Fragestellung zu betrachten, kann dazu führen, dass Therapien mehr oder weniger nützlich erscheinen, als sie es tatsächlich sind [213]. Das Ziel von hochwertigen systematischen Übersichten ist es, diese Art von Verzerrungen dadurch in den Griff zu bekommen, dass die Evidenz nicht selektiv, sondern systematisch identifiziert, bewertet und zusammengefasst wird [103,110,149,291].

In systematischen Übersichten wird die Evidenz aus einem oder mehreren Studientypen, der/die die beste Antwort auf eine spezifische und klar formulierte Frage geben kann/können, identifiziert, bewertet und zusammengefasst. Zur Identifizierung, Auswahl und kritischen Bewertung der für die interessierende Frage relevanten Studien werden systematische und explizite Methoden angewendet. Hat man Studien identifiziert, werden diese Daten systematisch erhoben und ausgewertet. Systematische Übersichten sind nicht experimentelle Studien, deren Methodik darauf abzielen muss, systematische Fehler (Verzerrungen) auf jeder Ebene des Reviewprozesses zu minimieren [110,185,213].

Im Falle von systematischen Übersichten über die Wirkungen medizinischer Behandlungen geben randomisierte kontrollierte Studien die zuverlässigsten Antworten. Bei anderen Fragestellungen wie etwa zur Ätiologie oder Prognose oder bei der qualitativen Beschreibung von Patientenerfahrungen besteht die geeignete Evidenzbasis für eine systematische Übersicht aus anderen Arten

von Primärstudien [149]. Systematische Übersichten über Diagnose- und Screeningtests weisen im Vergleich zu Übersichten über Behandlungsmaßnahmen einige methodische Unterschiede auf [85].

Bei der Erstellung der Berichte des Instituts dienen systematische Übersichten in erster Linie dazu, potenziell relevante (Primär-)Studien zu identifizieren. Ein IQWiG-Bericht kann aber auch teilweise oder sogar ausschließlich auf systematischen Übersichten beruhen (siehe Abschnitt 6.3.2). Die vom Institut für Patienten und Bürger erstellten Gesundheitsinformationen stützen sich zum großen Teil auf systematische Übersichten. Dazu zählen systematische Übersichten über Therapien sowie Übersichten, in denen es um andere Fragen wie Ursachen, unerwünschte Wirkungen und um die Synthese qualitativer Forschungsarbeiten geht (siehe Abschnitt 5.3.3).

Damit das Institut eine systematische Übersicht über Behandlungseffekte verwenden kann, muss diese gewisse Mindestvoraussetzungen erfüllen, d. h. sie darf nach dem Oxman-Guyatt-Index nur geringe methodische Mängel aufweisen [210,286,288]. Neben der Stärke der in systematischen Übersichten untersuchten Evidenz berücksichtigt das Institut auch die Relevanz und die Übertragbarkeit der Evidenz. Dazu gehört auch die Untersuchung der Frage, ob die Ergebnisse in verschiedenen Populationen und Untergruppen sowie in verschiedenen Gesundheitsbereichen einheitlich sind. Berücksichtigt werden i. d. R. folgende Faktoren: die Population der Studienteilnehmer in den eingeschlossenen Studien (darunter Geschlecht und Erkrankungsrisiko zu Beginn der Studie), der medizinische Kontext (darunter die medizinischen Versorgungsbereiche und die medizinischen Leistungserbringer) sowie die Anwendbarkeit und wahrscheinliche Akzeptanz der Intervention in der Form, in der sie bewertet wurde [47,83].

6.3.2 Nutzenbewertung auf Basis von systematischen Übersichten und HTA-Berichten

Eine Nutzenbewertung auf Basis von systematischen Übersichten und HTA-Berichten (beide im Folgenden teilweise als „Sekundärliteratur“ bezeichnet) kann – sofern bestimmte Voraussetzungen erfüllt sind – eine ressourcensparende und verlässliche Evidenzbasis für Empfehlungen an den G-BA bzw. das BMG liefern. Die Anwendbarkeit einer Nutzenbewertung auf Basis von systematischen Übersichten und HTA-Berichten hängt von der Verfügbarkeit der Sekundärliteratur ab, die

- von ausreichend hoher Qualität ist und nur minimales Biaspotenzial aufweist,
- unmittelbar relevant ist und
- qualitativ hochwertige Primärstudien enthält,

um eine klare Schlussfolgerung ziehen zu können [22,286].

Die angewendete Methodik muss eine hinreichende Sicherheit dafür bieten, dass eine neue Nutzenbewertung auf Basis der Primärliteratur nicht zu anderen Schlussfolgerungen käme als die Bewertung auf Basis von systematischen Übersichten und HTA-Berichten. Dies ist z. B. regelhaft dann nicht der Fall, wenn eine relevante Menge bislang unpublizierter Daten zu erwarten ist.

A) Fragestellungen

Prinzipiell ist diese Methode für alle Fragestellungen geeignet, sofern o. g. Kriterien erfüllt sind. Bei der Entwicklung der Fragestellung sind insbesondere die folgenden Punkte zu berücksichtigen:

- Definition der interessierenden Population
- Definition der interessierenden Intervention und Vergleichsintervention
- Definition aller relevanten Endpunkte
- ggf. das fokussierte Gesundheitssystem bzw. der geografische Bezug (z. B. Deutschland, Europa etc.)

Die so definierte Frage bildet auch die Grundlage für die Festlegung der für die Nutzenbewertung geltenden Ein- und Ausschlusskriterien und damit für die Feststellung der inhaltlichen und methodischen Relevanz der gefundenen Publikationen. Auf Basis der Fragestellung wird auch entschieden, auf welchem Primärstudientyp die verwendeten systematischen Übersichten und HTA-Berichte beruhen müssen. Je nach Fragestellung ist es möglich, dass bestimmte Teile eines Auftrags auf Basis von systematischen Übersichten und HTA-Berichten beantwortet werden, während für andere Teile auf die Primärliteratur zurückgegriffen wird.

B) Recherchestrategie einschließlich Aktualisierung/ Ergänzung der Recherche

Vollständigkeit im Sinne einer vollständigen Berücksichtigung aller verfügbaren Primärstudien wird nicht angestrebt. Erforderlich ist jedoch die Gewährleistung einer Robustheit des Ergebnisses, d. h. eine ausreichende Sicherheit, dass sich das Ergebnis durch das Einbeziehen einer weiteren Studie nicht wesentlich verändern würde.

Erstellung einer Recherche:

Zum grundsätzlichen Ablauf einer Recherche siehe Abschnitt 6.1.

Datenbankauswahl:

Bei der Informationsgewinnung sind teilweise andere Quellen zu berücksichtigen als bei der Suche nach Primärliteratur. In erster Linie wird regelhaft in Datenbanken, die ausschließlich oder überwiegend Sekundärliteratur einschließlich HTA-Berichten enthalten, recherchiert. Darüber

hinaus wird in einer Auswahl biomedizinischer Datenbanken, die auch – aber nicht primär – Sekundärliteratur enthalten (z. B. MEDLINE und EMBASE), recherchiert.

Ob und welche Datenbanken oder sonstige Quellen (z. B. Websites einzelner HTA-Institutionen) zusätzlich relevant und zu durchsuchen sind, ist themenabhängig zu entscheiden. Leitlinien werden als Informationsquelle nicht grundsätzlich ausgeschlossen. Es erfolgt jedoch i. d. R. keine systematische Recherche nach Leitlinien. Kostenpflichtige HTAs werden in Ausnahmefällen berücksichtigt, wenn z. B. anzunehmen ist, dass zusätzliche Informationen daraus gewonnen werden können, oder ansonsten keine Informationen verfügbar sind.

Auswahl relevanter Publikationen:

Die Auswahl relevanter Publikationen erfolgt analog der in Abschnitt 6.1.4 beschriebenen Auswahl von Primärpublikationen.

Mindestzahl relevanter systematischer Übersichten:

Alle qualitativ ausreichenden und thematisch relevanten systematischen Übersichten und HTA-Berichte werden berücksichtigt. In der Regel sollten mindestens zwei Arbeiten von hoher Qualität, die unabhängig voneinander durchgeführt wurden, als Grundlage für die Erstellung eines Berichts auf Basis von Sekundärliteratur vorhanden sein, um die Konsistenz der Ergebnisse überprüfen zu können. Wenn nur eine Arbeit von hoher Qualität vorliegt und dementsprechend berücksichtigt werden kann, ist eine Begründung für die Bewertung allein anhand dieser einen systematischen Übersicht/ dieses einen HTA-Berichts erforderlich.

Aktualisierung/ Ergänzung der Recherche:

In den meisten Fällen wird eine ergänzende Literaturrecherche nach aktuellen Primärstudien erforderlich sein (Update), die den Zeitraum zwischen den in der Sekundärliteratur durchgeführten Recherchen und der Erstellung des IQWiG-Berichts abdeckt. Auf eine ergänzende aktualisierende Primärliteraturrecherche kann bei der Nutzenbewertung anhand von Sekundärliteratur nur in begründeten Ausnahmefällen verzichtet werden. Zudem kann es für spezifische, nicht in der Sekundärliteratur behandelte Fragestellungen notwendig sein, ergänzende Recherchen nach Primärliteratur durchzuführen.

C) Qualitätsbewertung der Publikationen einschließlich Mindestanforderungen

Die Bewertung der allgemeinen Qualität der Sekundärliteratur erfolgt anhand des validierten Qualitätsindex für systematische Übersichten von Oxman und Guyatt [285,286,288]. Systematische Übersichten gelten nach dem Oxman-Guyatt-Index dann als methodisch ausreichend, wenn

diese von zwei voneinander unabhängigen Reviewern in der Gesamtbewertung mit mindestens fünf von sieben möglichen Punkten bewertet wurden. Darüber hinaus werden regelhaft die Sponsoren der systematischen Übersichten und HTA-Berichte sowie die Interessenkonflikte der Autoren dokumentiert und diskutiert. Die Checkliste von Oxman und Guyatt [285,286,288] kann je nach Anforderung des Projekts um weitere Items ergänzt werden (Vollständigkeit der Recherche, zusätzliche Aspekte bei diagnostischer Sekundärliteratur etc.).

D) Ergebnisse

Die Ergebnisse einer Nutzenbewertung anhand von systematischen Übersichten und HTA-Berichten werden zu jeder Fragestellung, wenn möglich, in tabellarischer Form zusammengefasst. Bei Vorliegen inkonsistenter Ergebnisse aus mehreren Arbeiten zu einem Endpunkt werden mögliche Erklärungen für diese Heterogenität beschrieben [211].

Liefert die Zusammenstellung der systematischen Übersichten und HTA-Berichte Hinweise darauf, dass eine neu erstellte Nutzenbewertung auf Basis von Primärstudien zu anderen Ergebnissen führen könnte, so wird eine Nutzenbewertung auf Basis von Primärstudien durchgeführt.

E) Schlussfolgerung/ Empfehlung

Berichte auf Basis von Sekundärliteratur fassen die Ergebnisse zugrunde liegender systematischer Übersichten und HTA-Berichte zusammen, ergänzen sie ggf. um die Zusammenfassung der aktuelleren Primärstudien oder Primärstudien zu Fragestellungen, die nicht durch die Sekundärliteratur abgedeckt werden, und ziehen daraus ein eigenständiges Fazit.

Die Empfehlungen auf Basis von Sekundärliteratur beruhen nicht auf der Zusammenfassung der Empfehlungen bzw. Schlussfolgerungen der zugrunde liegenden systematischen Übersichten oder HTA-Berichte. In HTA-Berichten werden diese häufig vor dem Hintergrund spezifischer sozialpolitischer und ökonomischer Rahmenbedingungen eines spezifischen Gesundheitssystems formuliert und sind daher nur selten auf die Versorgungssituation in Deutschland übertragbar.

6.3.3 Berücksichtigung publizierter Meta-Analysen

Den Bewertungen des IQWiG liegt, dem internationalen Standard der evidenzbasierten Medizin folgend, in der Regel eine fragestellungsbezogene systematische Recherche nach relevanten Primärstudien zugrunde. Sofern sinnvoll und möglich, werden die Ergebnisse der identifizierten Einzelstudien mittels Meta-Analysen zusammenfassend bewertet. Das Institut hat dadurch allerdings in der Regel lediglich Zugriff auf aggregierte Daten der Primärstudien auf Basis der jeweiligen Publikation oder des zur Verfügung gestellten Studienberichts. Es gibt Situationen, in

denen Meta-Analysen, die auf Basis der individuellen Patientendaten der relevanten Studien durchgeführt werden, eine höhere Aussagekraft haben (siehe Abschnitt 6.4.7). Dies ist insbesondere dann der Fall, wenn neben dem alleinigen Interventionseffekt die Evaluation weiterer Faktoren von Interesse ist, die den Interventionseffekt möglicherweise beeinflussen (Interaktion zwischen Interventionseffekt und Kovariablen). Hier liefern Meta-Analysen mit individuellen Patientendaten im Vergleich zu Meta-Regressionen, die auf aggregierten Daten beruhen, im Allgemeinen eine höhere Ergebnissicherheit, d. h. präzisere und nicht durch ökologischen Bias verzerrte Ergebnisse [338]. Solche Analysen können daher im Einzelfall zu genaueren Schlussfolgerungen führen, insbesondere bei Vorliegen heterogener Ergebnisse, die möglicherweise auf unterschiedliche Ausprägungen von Patientenmerkmalen zurückgeführt werden können. Von einer erhöhten Validität von Meta-Analysen basierend auf individuellen Patientendaten kann man allerdings nur dann ausgehen, wenn solche Analysen auch tatsächlich auf die Fragestellung der Bewertung des Instituts ausgerichtet sind und zudem eine hohe Ergebnissicherheit aufweisen. Für die Beurteilung der Ergebnissicherheit solcher Analysen ist eine maximale Transparenz Grundvoraussetzung. Dies bezieht sich sowohl auf die Planung als auch auf die Durchführung der Analysen. Diesbezüglich wesentliche, für die Durchführung von Meta-Analysen allgemeingültige Aspekte sind z. B. in einem Dokument der Europäischen Zulassungsbehörde EMA beschrieben [76]. Das Institut berücksichtigt bei der Nutzenbewertung publizierte Meta-Analysen auf Basis individueller Patientendaten, sofern sie (Unter-)Fragestellungen des Berichts, die durch Meta-Analysen auf Basis aggregierter Daten nicht ausreichend sicher beantwortet werden können, adressieren und von einer hohen Ergebnissicherheit der jeweiligen Analyse ausgegangen werden kann.

6.4 Spezielle biometrische Aspekte

6.4.1 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 mithilfe von Mittelwerten sowie Differenzen von Mittelwerten – unter Umständen nach geeigneter Gewichtung – dargestellt werden. Bei kategoriellen Zielvariablen kommen die üblichen Effekt- und Risikomaße aus Vierfeldertafeln infrage [36]. Eine übersichtliche Zusammenfassung der Vor- und Nachteile der üblichen Effektmaße gibt Kapitel 9 des *Cochrane*

Handbook for Systematic Reviews of Interventions [86]. Besondere Aspekte, die beim Vorliegen ordinaler Daten berücksichtigt werden müssen, beschreibt Agresti [4,5].

Zwingend erforderlich ist es, bei jeder Effektschätzung den Grad der statistischen Unsicherheit zu beschreiben. Häufige Methoden hierfür stellen die Berechnung des Standardfehlers sowie die Angabe eines Konfidenzintervalls dar. Wann immer möglich gibt das Institut adäquate Konfidenzintervalle für Effektschätzungen mit der Information, ob es sich um ein- oder zweiseitige Konfidenzgrenzen handelt, sowie das gewählte Konfidenzniveau an. 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. [10].

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

6.4.2 Beurteilung statistischer Signifikanz

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

Bei der Interpretation von p -Werten ist eine Reihe von Aspekten zu betrachten. Es muss im Detail klar sein, auf welche Fragestellung und welche Datensituation sich die Signifikanzangabe bezieht und wie die statistische Hypothesenformulierung lautet. Insbesondere muss klar sein, ob es sich um eine ein- oder zweiseitige Hypothese handelt [45] und ob die betrachtete Hypothese als Teil eines Mehrhypothesenproblems zu sehen ist [360]. Beide Aspekte – ob eine Hypothese ein- oder zweiseitig zu formulieren ist und ob für multiples Testen adjustiert werden muss – werden in der wissenschaftlichen Literatur immer wieder kontrovers diskutiert.

Bezüglich der Hypothesenformulierung ist traditionell von einem zweiseitigen Testproblem auszugehen. Ausnahmen hiervon bilden zum Beispiel Nichtunterlegenheitsstudien. Die Formulierung eines einseitigen Hypothesenproblems ist grundsätzlich immer möglich, bedarf jedoch einer genauen Begründung. Im Fall einer einseitigen Hypothesenformulierung sind die

Anwendung einseitiger Signifikanztests sowie die Berechnung einseitiger Konfidenzgrenzen angebracht. Zur besseren Vergleichbarkeit mit zweiseitigen statistischen Verfahren wird in einigen Guidelines für klinische Studien eine Halbierung des üblichen Signifikanzniveaus von 5 % auf 2,5 % gefordert [200]. Das Institut folgt dem zentralen Grundprinzip, dass die Hypothesenformulierung (ein- oder zweiseitig) und das Signifikanzniveau a priori klar festzulegen sind, und begründet Abweichungen von den üblichen Festlegungen (einseitige anstelle zweiseitiger Hypothesenformulierung, Signifikanzniveau ungleich 5 %) bzw. achtet auf entsprechende Begründungen in der Primärliteratur.

Ist eine betrachtete Hypothese eindeutig Teil eines Mehrhypothesenproblems, so ist eine adäquate Adjustierung für multiples Testen erforderlich, wenn der Fehler 1. Art für das gesamte Mehrhypothesenproblem kontrolliert werden soll. Eine Übersicht, in welchen Situationen dies der Fall ist und welche Methoden hierfür zur Verfügung stehen, geben Bender und Lange [40]. Die besonderen Aspekte beim Umgang mit Multiplizität in systematischen Übersichten beschreiben Bender et al. [37]. Wenn sinnvoll und möglich, wendet das Institut Methoden zur Adjustierung für multiples Testen an. Im Rahmen der Nutzenbewertung (siehe Abschnitt 3.1) versucht das Institut, den Fehler 1. Art bezüglich der Aussagen zu jedem einzelnen Nutzenaspekt getrennt zu kontrollieren. Die zusammenfassende Bewertung wird in der Regel nicht auf quantitative Weise durchgeführt, sodass hier auch keine formalen Methoden zur Adjustierung für multiples Testen angewendet werden können.

Das Institut wertet ein statistisch nicht signifikantes Ergebnis nicht als Evidenz für das Nichtvorhandensein eines Effekts (Abwesenheit bzw. Äquivalenz) [9]. Zum Nachweis einer Gleichheit wendet das Institut adäquate Methoden für Äquivalenzhypothesen an.

Als prinzipielle Alternative zu statistischen Signifikanztests können Bayes'sche Methoden angesehen werden [342,343]. Das Institut wird in Abhängigkeit von der Fragestellung nach Bedarf auch Bayes'sche Methoden anwenden.

6.4.3 Beurteilung klinischer Relevanz

Die klinische Relevanz eines Effekts oder Risikos ist grundsätzlich nicht am p -Wert ablesbar. Die statistische Signifikanz ist eine Wahrscheinlichkeitsaussage, in die neben der Stärke eines möglichen Effekts auch die Variabilität der Daten und der Stichprobenumfang eingehen. Bei der Interpretation der Bedeutsamkeit von p -Werten muss insbesondere der Stichprobenumfang der zugrunde liegenden Studie berücksichtigt werden [307]. 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 [128,195].

Folglich lässt sich die klinische Relevanz eines Studienergebnisses auf keinen Fall am p -Wert ablesen.

Für die Beurteilung der klinischen Relevanz von Studienergebnissen gibt es noch kein breit akzeptiertes methodisches Vorgehen. Nur in wenigen Leitlinien finden sich Hinweise auf die Definition von relevanten beziehungsweise irrelevanten Unterschieden zwischen Gruppen. Ein erster Ansatz zur Beurteilung der klinischen Relevanz von Studienergebnissen ist die Bewertung des Effektschätzers und des dazugehörigen Konfidenzintervalls mithilfe medizinischer Sachkenntnis. Ein formales Relevanzkriterium kann die Beurteilung der (im Falle von günstigen Effekten) unteren Konfidenzgrenze für den Effektschätzer beziehungsweise die Anwendung eines statistischen Tests mit Verschiebung der Nullhypothese zum statistischen Nachweis relevanter klinischer Effekte sein [382]. Eine weitere Möglichkeit liegt darin, ein Relevanzkriterium individuell zu formulieren, zum Beispiel im Sinne einer Responderdefinition [229]. Darüber hinaus spielt die individuelle Einschätzung der Betroffenen eine wesentliche Rolle. Anhaltspunkte hierfür wird im Einzelfall die Darlegung patientenrelevanter Endpunkte liefern. Das Institut nimmt die Bewertung der klinischen Relevanz problemorientiert unter Berücksichtigung dieser Aspekte vor.

6.4.4 Nachweis der Verschiedenheit

Beim empirischen Nachweis, dass sich gewisse Gruppen bezüglich eines bestimmten Merkmals unterscheiden, sind verschiedene Aspekte zu beachten. Zunächst ist festzuhalten, dass „Nachweis“ hier nicht als „Beweis“ im mathematischen Sinne zu verstehen ist. Mithilfe empirischer Studiendaten lassen sich Aussagen nur unter Zulassung gewisser Irrtumswahrscheinlichkeiten treffen. Durch die 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. Diese Argumentationsebene ist zu trennen von der Beurteilung der klinischen Relevanz eines Unterschieds. 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 [46,90].

Zur abschließenden Darstellung der Ergebnisse gehören neben der Signifikanzangabe für eine Aussage auch ein mit adäquaten Methoden berechnetes Konfidenzintervall für das gewählte Effektmaß, 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.

6.4.5 Nachweis der Gleichheit

Einer der häufigen schwerwiegenden Fehler bei der Interpretation medizinischer Daten ist es, ein nicht signifikantes Ergebnis eines traditionellen Signifikanztests als Evidenz für die Richtigkeit der Nullhypothese zu werten [9]. Für den Nachweis einer „Gleichheit“ ist die Anwendung von Methoden zur Untersuchung von Äquivalenzhypothesen erforderlich [216]. Wichtig ist hierbei zunächst das Verständnis, dass eine exakte „Gleichheit“ – also zum Beispiel, dass die Differenz der Mittelwerte zweier Gruppen exakt null beträgt – mithilfe statistischer Methoden nicht nachweisbar 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. die Festlegung eines Äquivalenzbereichs ist erforderlich.

Um sinnvoll auf Äquivalenz schließen zu können, müssen – ähnlich wie beim Nachweis eines Unterschieds – a priori die wissenschaftliche Fragestellung und darauf basierend die Zielvariable, das Effektmaß und die statistische Hypothesenformulierung festgelegt werden. Zusätzlich ist bei Äquivalenzstudien der Äquivalenzbereich klar zu definieren. Dies kann entweder zweiseitig, was zu einem Äquivalenzintervall führt, oder einseitig im Sinne einer „höchstens irrelevanten Unterschiedlichkeit“ beziehungsweise „höchstens irrelevanten Unterlegenheit“ erfolgen. In letzterem Fall spricht man von einer „Nichtunterlegenheitshypothese“ [82,200,303].

Wie bei Überlegenheitsstudien ist es auch bei Äquivalenzstudien erforderlich, den benötigten Stichprobenumfang vor Beginn der Studie zu berechnen. Welche Methode hierfür infrage kommt, hängt von der genauen Hypothese und von der gewählten Analyse-methode ab [302].

Für die 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 Konfidenzintervalls zum Niveau 90 % [216].

Im Vergleich zu Überlegenheitsstudien besitzen Äquivalenzstudien spezielle methodische Probleme. Zum einen ist es häufig schwierig, Äquivalenzbereiche sinnvoll zu definieren [243]. Zum anderen schützen zum Beispiel die üblichen Designkriterien Randomisierung und Verblindung nicht mehr hinreichend sicher vor Verzerrungen [332]. Auch ohne Kenntnis der Therapiegruppe ist es zum Beispiel möglich, die Differenz der Behandlungsunterschiede zur Null und damit zur gewünschten Alternativhypothese hin zu verschieben. Des Weiteren ist mit dem Intention-to-Treat-Prinzip sehr vorsichtig umzugehen, da dessen inadäquate Anwendung eine falsche Äquivalenz vortäuschen kann [216]. Somit ist bei der Bewertung von Äquivalenzstudien besondere Vorsicht geboten.

6.4.6 Adjustierung und multifaktorielle Verfahren

Vor allem in nicht randomisierten Studien spielen multifaktorielle Verfahren, die es ermöglichen, den Effekt von Confoundern auszugleichen, eine zentrale Rolle [225]. Ein weiteres wichtiges Anwendungsgebiet für multifaktorielle Methoden sind Studien mit mehreren Behandlungen [259]. Die Darstellung von Ergebnissen multifaktorieller Verfahren ist jedoch in der medizinischen Fachliteratur leider oft unzureichend [38,270]. Um die Qualität einer solchen Analyse bewerten zu können, sind eine Darstellung der wesentlichen Aspekte im Rahmen der statistischen Modellbildung notwendig [176,309] sowie Angaben zur Güte des gewählten Modells (Goodness-of-Fit) [192]. Die wichtigsten Informationen hierbei sind in der Regel:

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

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

Besonders kritisch ist eine unzureichende Darstellung der Ergebnisse multifaktorieller Verfahren dann, wenn es durch die (nicht ausreichend transparent dargestellte) statistische Modellierung zu einer Verschiebung von Effekten in einen „gewünschten“ Bereich kommt, der bei unifaktorieller Betrachtung so nicht erkennbar ist. Ausführliche Erläuterungen zu den Anforderungen an den Umgang mit multifaktoriellen Verfahren sind in verschiedenen Übersichten und Leitlinien in der Literatur zu finden [25,39,225].

In eigenen regressionsanalytischen Berechnungen setzt das Institut moderne Verfahren ein [175]. Sehr vorsichtig zu interpretieren sind hierbei Ergebnisse multifaktorieller Modelle, die aus einem Variablenselektionsprozess hervorgegangen sind. Falls bei der Modellwahl solche Variablenselektionsverfahren unumgänglich sind, wird eine Form der Backward-Elimination eingesetzt, da diese dem Verfahren der Forward-Selektion vorzuziehen ist [175]. Wichtig ist hierbei eine sorgfältige Vorauswahl der potenziellen erklärenden Variablen unter sachwissenschaftlicher Kenntnis [87]. Bei Bedarf werden auch moderne Verfahren wie die Lasso-Methode eingesetzt [356]. Bei der Modellierung stetiger Einflussgrößen greift das Institut im Bedarfsfall auf flexible Modellansätze wie zum Beispiel das Verfahren der Fractional Polynomials zurück [310,320], um eine adäquate Beschreibung nicht monotoner Zusammenhänge zu ermöglichen.

6.4.7 Meta-Analysen

In der Literatur verwendete Begriffe wie „Literaturübersicht“, „systematische Übersicht“, „Meta-Analyse“, „gepoolte Analyse“ oder „Forschungssynthese“ sind häufig unterschiedlich definiert und nicht klar voneinander abgegrenzt [110]. Das Institut verwendet folgende Begriffe und Definitionen: Bei einer „nicht systematischen Übersicht“ handelt es sich um eine Beschreibung und Bewertung von Studienergebnissen zu einer definierten Thematik ohne eine ausreichend systematische und reproduzierbare Identifikationsmethode der relevanten Forschungsergebnisse zu dieser Thematik. Eine quantitative Zusammenfassung von Daten mehrerer Studien wird als „gepoolte Analyse“ bezeichnet. Wegen der fehlenden Systematik und der inhärenten subjektiven Komponente sind Übersichten und Auswertungen, die nicht auf einer systematischen Literaturrecherche basieren, sehr anfällig für Verzerrungen. Eine „systematische Übersicht“ beruht auf einer umfassenden systematischen Vorgehensweise und Studienbewertung, um mögliche Biasquellen zu minimieren. Sie kann – muss aber nicht – eine quantitative Zusammenfassung der Studienergebnisse beinhalten. Eine „Meta-Analyse“ wird verstanden als eine statistische Zusammenfassung von Ergebnissen mehrerer Studien im Rahmen einer systematischen Übersicht. Sie basiert in den meisten Fällen auf aggregierten Studiendaten aus Publikationen. Dabei wird aus den in einzelnen Studien gemessenen Effektstärken, unter Berücksichtigung der Fallzahlen und der Varianzen, ein Gesamteffekt berechnet. Effizientere Auswertungsverfahren sind möglich, wenn

individuelle Patientendaten aus den Studien verfügbar sind. Unter einer „Meta-Analyse mit individuellen Patientendaten“ (IPD = Individual Patient Data) wird die Auswertung von Daten auf Patientenebene im Rahmen eines allgemeinen statistischen Modells mit festen oder zufälligen Effekten verstanden, in das die Studie als Effekt und nicht als Beobachtungseinheit eingeht. Unter einer „prospektiven Meta-Analyse“ versteht das Institut die a priori geplante statistische Zusammenfassung der Ergebnisse mehrerer prospektiv gemeinsam geplanter Studien. Sollte es zur jeweiligen Fragestellung auch noch andere Studien geben, so müssen diese jedoch auch in der Auswertung berücksichtigt werden, um den Charakter einer systematischen Übersicht zu bewahren.

Die übliche Darstellung der Ergebnisse einer Meta-Analyse erfolgt mittels Forest-Plots, in denen die Effektschätzer der einzelnen Studien und des Gesamteffekts inklusive Konfidenzintervallen grafisch aufgetragen werden [247]. Es kommen zum einen Modelle mit festen Effekten zum Einsatz, die (z. B. durch die Inverse der Varianz) gewichtete Mittelwerte der Effektstärken liefern. Es werden zum anderen aber auch häufig Modelle mit zufälligen Effekten gewählt, bei denen eine Schätzung der Varianz zwischen den einzelnen Studien (Heterogenität) berücksichtigt wird. Die Frage, in welchen Situationen welches Modell eingesetzt werden soll, wird seit Langem kontrovers diskutiert [114,330,371]. Liegen Anhaltspunkte dafür vor, dass die Effekte der einzelnen Studien homogen sind, ist eine Meta-Analyse unter der Annahme fester Effekte ausreichend. Diese Anhaltspunkte werden jedoch häufig nicht vorliegen, sodass zur Evaluierung der Studien in ihrer Gesamtheit die Annahme zufälliger Effekte hilfreich ist [331]. Des Weiteren ist zu beachten, dass die aus einem Modell mit festen Effekten berechneten Konfidenzintervalle für den erwarteten Gesamteffekt selbst bei Vorhandensein einer geringen Heterogenität im Vergleich zu Konfidenzintervallen aus einem Modell mit zufälligen Effekten eine substanziell kleinere Überdeckungswahrscheinlichkeit aufweisen können [59]. Das Institut verwendet daher vorrangig Modelle mit zufälligen Effekten und weicht nur in begründeten Ausnahmefällen auf Modelle mit festen Effekten aus. Dabei ist zu beachten, dass sich die meta-analytischen Ergebnisse von Modellen mit zufälligen und festen Effekten bei homogener Datenlage allenfalls marginal unterscheiden. Des Weiteren wird das Institut, wie im Folgenden beschrieben, nur dann stark heterogene Studienergebnisse meta-analytisch zusammenfassen, wenn plausible Gründe für die Heterogenität ersichtlich sind, die eine Zusammenfassung trotzdem rechtfertigen.

Bevor eine Meta-Analyse durchgeführt wird, muss zunächst überlegt werden, ob die Zusammenfassung der betrachteten Studien überhaupt sinnvoll ist, da die Studien bezüglich der Fragestellung vergleichbar sein müssen. Darüber hinaus werden sich in den zusammenzufassenden Studien trotz Vergleichbarkeit häufig heterogene Effekte zeigen [187]. In dieser Situation ist es erforderlich, die Heterogenität der Studien bezüglich der Ergebnisse zu untersuchen [148]. Das Vorliegen von Heterogenität kann statistisch getestet werden, wobei diese Verfahren in der Regel

eine sehr niedrige Power haben. Daher wird für diese Tests empfohlen, ein Signifikanzniveau von 0,1 bis 0,2 zu wählen [208]. Wichtiger jedoch ist es, das Ausmaß der Heterogenität zu quantifizieren. Zu diesem Zweck gibt es spezielle neuere statistische Methoden wie zum Beispiel das I^2 -Maß [186]. Für dieses Maß existieren Untersuchungen, die eine grobe Einschätzung der Heterogenität zulassen (z. B. niedrig/ mittelmäßig/ hoch für I^2 -Werte im Bereich 25/50/75 %) [187]. Ist die Heterogenität der Studien zu groß, so ist eine statistische Zusammenfassung der Studienergebnisse unter Umständen nicht sinnvoll [86]. Dabei spielt auch die Lage der Effekte eine Rolle. Zeigen die einzelnen Studien einen deutlichen und gleichgerichteten Effekt, dann können auch heterogene Ergebnisse mittels eines Modells mit zufälligen Effekten geeignet zusammengefasst werden. In den übrigen Situationen führt das Institut jedoch keine Meta-Analyse durch. In diese Entscheidung müssen jedoch neben statistischen Maßzahlen in jedem Fall auch inhaltliche Gründe einfließen, die nachvollziehbar darzustellen sind. In diesem Zusammenhang spielt auch die Wahl des Effektmaßes eine Rolle. Es kann sein, dass die Wahl eines bestimmten Maßes zu großer Studienheterogenität führt, ein anderes Maß jedoch nicht. Bei binären Daten sind relative Effektmaße häufig stabiler als absolute, da sie nicht so stark vom Basisrisiko abhängen [142]. In solchen Fällen sollte die Datenanalyse über ein relatives Effektmaß erfolgen; für die deskriptive Darstellung sollten aber absolute Maße für spezifische Basisrisiken hieraus abgeleitet werden.

Bei großer Heterogenität der Studien ist es notwendig, mögliche Ursachen hierfür zu untersuchen. Unter Umständen lassen sich mittels Meta-Regressionen Faktoren finden, die die Heterogenität der Effektstärken erklären können [353,367]. In einer Meta-Regression wird die statistische Assoziation zwischen den Effektstärken der einzelnen Studien und den Studiencharakteristika untersucht, sodass möglicherweise Studiencharakteristika gefunden werden können, die einen Erklärungswert für die unterschiedlichen Effektstärken, also die Heterogenität, haben. Wichtig ist jedoch, dass man bei der Interpretation der Ergebnisse die Einschränkungen solcher Analysen berücksichtigt. Selbst wenn eine Meta-Regression auf randomisierten Studien basiert, kann aus ihr nur die Evidenz einer Beobachtungsassoziation abgeleitet werden, nicht jedoch ein kausaler Zusammenhang [353]. Besonders schwierig zu interpretieren sind Meta-Regressionen, die versuchen, eine Beziehung zwischen den unterschiedlichen Effektstärken und den durchschnittlichen Patientencharakteristika der einzelnen Studien aufzuzeigen. Solche Analysen unterliegen den gleichen Beschränkungen wie die Ergebnisse ökologischer Studien der Epidemiologie [157]. Aufgrund der hohen Anfälligkeit für Verzerrungen, die bei auf aggregierten Daten basierenden Analysen auch nicht durch Adjustierung ausgeglichen werden können, sind hier gesicherte Schlussfolgerungen nur auf der Basis individueller Patientendaten möglich [338,353] (siehe auch Abschnitt 6.3.3).

6.4.8 Umgang mit nicht oder nicht vollständig publizierten Daten

In der Praxis ergibt sich häufig das Problem, dass zur Qualitätsbewertung von Publikationen notwendige Daten oder Informationen ganz oder teilweise fehlen. Dies betrifft in erster Linie sogenannte graue Literatur und Abstracts, aber auch vollständige Publikationen. Darüber hinaus ist es möglich, dass Studien zum Zeitpunkt der Bewertung einer Technologie durch das Institut (noch) nicht publiziert sind.

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

Ein häufig auftretendes Problem besteht darin, dass wichtige Angaben zur Durchführung einer Meta-Analyse (z. B. Varianzen der Effektschätzer) fehlen. In vielen Fällen lassen sich jedoch die fehlenden Daten aus den vorhandenen Angaben berechnen oder zumindest schätzen [94,194,289]. Das Institut wendet nach Möglichkeit solche Verfahren an.

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

6.4.9 Darstellung von Biasarten

Unter Bias versteht man eine systematische Abweichung der Schätzung eines Effekts aus Studien-daten vom wahren Effekt. Es gibt vielfältige mögliche Ursachen, die einen Bias hervorrufen können [312]. Nur die wichtigsten Biasarten werden im Folgenden beschrieben; eine ausführliche Übersicht über verschiedene Biasarten in unterschiedlichsten Situationen gibt Feinstein [127].

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

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

Protokollverletzungen und Studienaustritte können das Studienergebnis systematisch verzerren, was als „Attrition Bias“ bezeichnet wird. Zur Vermeidung von Attrition Bias kann in Studien 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 an 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. Um diesen Bias zu vermeiden, muss der diagnostische Test in einem repräsentativen Patientenkollektiv untersucht werden.

Bei der Untersuchung von Screeningprogrammen besteht die Gefahr, dass durch eine Vorverlagerung der Diagnose eine nur scheinbare Verlängerung der Überlebenszeit beobachtet wird, die jedoch in Wahrheit auf nicht vergleichbare Anfangszeitpunkte zurückzuführen ist („Lead Time Bias“). Eine verlängerte Überlebenszeit kann ebenso vorgetäuscht werden, wenn eine

Screeningmaßnahme bevorzugt milder oder langsamer verlaufende Frühformen einer Krankheit erkennen kann („Length Bias“). Einen Schutz vor diesen Verzerrungsmechanismen bietet die Durchführung einer randomisierten Studie zur Beurteilung der Wirksamkeit einer Screeningmaßnahme.

Ein allgemeines Problem bei der Schätzung von Effekten ist eine Verzerrung durch Messfehler und Fehlklassifikation in den erhobenen Studiendaten [66,69]. In der Praxis sind Messfehler kaum zu vermeiden und es ist bekannt, dass auch nicht differenzielle Messfehler zu einem Bias in der Effektschätzung führen können. Im Fall eines einfachen linearen Regressionsmodells mit klassischem 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. Je nach Fragestellung sollte die Stärke möglicher Messfehler diskutiert werden, und bei Bedarf sollten Methoden zur Adjustierung von Messfehlerverzerrungen angewendet werden.

Ein ähnliches Problem stellen fehlende Werte dar. Fehlende Werte, die nicht durch einen Zufallsmechanismus zustande kommen, können ebenso einen Bias im Ergebnis verursachen [251]. Die möglichen Ursachen und Auswirkungen fehlender Werte sind daher im Einzelfall zu diskutieren und bei Bedarf sind statistische Verfahren erforderlich, die einen Bias berücksichtigen beziehungsweise ausgleichen.

Im Rahmen von systematischen Übersichten spielt noch der „Publication Bias“ eine wichtige Rolle [35]. 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. Für die Entdeckung und Berücksichtigung von Publication Bias können grafische Methoden wie der Funnel Plot [111] und statistische Methoden wie die Meta-Regression eingesetzt werden [254,292,348].

6.5 Qualitative Methoden

6.5.1 Qualitative Studien

Verschiedene Quellen von Informationen können die Erarbeitung systematischer Übersichten unterstützen [99,248,352]. Eine dieser Quellen können Forschungsergebnisse aus qualitativen Studien sein [74,174,248,279,352]. Qualitative Studien scheinen sich in systematischen Übersichten zur Nutzenbewertung medizinischer Leistungen zu etablieren [98,99,279].

Qualitative Forschungsmethoden werden eingesetzt, um subjektives Erleben, subjektive Erfahrungen und individuelles Handeln sowie die soziale Realität zu erforschen und zu verstehen

[98,174,256,272]. Sie können Zugang zu den Sichtweisen und Erfahrungen von Patienten, Angehörigen und des medizinischen Personals mit einem bestimmten Krankheitsbild oder einer Intervention schaffen.

Qualitative Forschung kann Informationen zur Akzeptanz und Eignung von Interventionen in der Praxis liefern [73,98]. Die Ergebnisse qualitativer Forschung können bei der Interpretation einer systematischen Übersicht hilfreich sein [73,352].

Quantitative Forschung arbeitet primär mit Zahlen unterschiedlicher Größenordnung und ist durch eine starke Standardisierung gekennzeichnet. Es können jedoch auch persönliche und soziale Erfahrungen, zum Beispiel bei Untersuchungen zur Lebensqualität, berücksichtigt werden. Im Gegensatz dazu stehen bei der qualitativen Forschung subjektive Daten im Vordergrund, zum Beispiel bei der Durchführung von Fokusgruppen mit Teilnehmern einer randomisierten kontrollierten Studie. Qualitative Daten können auch über Interviews, Beobachtungen und schriftliche Dokumente wie Tagebücher erhoben werden. Das analytische Vorgehen hat zumeist das Ziel, übergreifende Themen und Konzepte der erhobenen Daten zu identifizieren und zu analysieren.

Qualitative Methoden können unter anderem als eigenständige Forschungsmethode, als Vorbereitung von oder in Ergänzung zu quantitativen Studien, im Rahmen einer Triangulation bzw. eines Mixed-Method-Ansatzes oder nach der Durchführung quantitativer Studien zur Erklärung von Prozessen oder Ergebnissen eingesetzt werden. Qualitative Forschung wird als Mittel angesehen, um die Verbindung zwischen Evidenz und Praxis zu fördern [100].

Qualitative Methoden können u. a. im Rahmen von Primärstudien oder systematischen Übersichten zur Ermittlung patientenrelevanter Endpunkte eingesetzt werden [73,98,100,272,279].

Trotz zunehmender Bedeutung qualitativer Methoden gibt es derzeit keine allgemein akzeptierte Vorgehensweise zur Synthese qualitativer Studien und zur Kombination qualitativer und quantitativer Daten [73,100,101].

Das Institut kann qualitative Forschungsergebnisse zur Ermittlung patientenrelevanter Endpunkte, im Rahmen der Darstellung von Hintergrundinformationen zu den Erfahrungen von Patienten und der Bedeutung des zu prüfenden Behandlungsverfahrens für Patienten sowie für die Diskussion und Interpretation der Ergebnisse einer systematischen Übersicht nutzen.

Im Rahmen der Entwicklung von Gesundheitsinformationen nutzt das Institut vorhandene qualitative Forschungsergebnisse zur Ermittlung von Informationsbedürfnissen sowie zur Eruierung von Erfahrungen mit einem bestimmten Krankheitsbild bzw. mit einer Intervention.

Die Recherche nach qualitativen Studien erfolgt im Anschluss an die Konkretisierung der Fragestellung systematisch in der Regel in folgenden Datenbanken: PubMed, CINAHL und PsycInfo. Es erfolgen eine Auswahl relevanter Publikationen anhand vorher festgelegter Ein- und Ausschlusskriterien sowie eine Abschätzung der Studienqualität anhand vorher definierter Kriterien.

Die Hauptaufgabe des Institutes bei der Bewertung qualitativer Studien besteht darin, festzustellen, ob das Forschungsdesign, die Studienqualität und die Reliabilität für die untersuchte Fragestellung angemessen sind. Bezüglich der Validität der Kriterien für die Durchführung, die Bewertung und die Synthese qualitativer Studien gibt es im Vergleich zu anderen Forschungsrichtungen einen schwächeren allgemeinen Konsens [98,101,174,272].

Die Ergebnisse der berücksichtigten Studien werden extrahiert, thematisch geordnet und für die Verwendung bei der Entwicklung von Gesundheitsinformationen deskriptiv zusammengefasst. Dieses Verfahren kann das Institut gegebenenfalls bei der Erarbeitung von Berichten nutzen.

Das Institut verfolgt die weitere methodische Entwicklung bei der Kombination qualitativer und quantitativer Methoden bzw. der jeweiligen Forschungsergebnisse sowohl in Primärstudien als auch in systematischen Übersichten zur Nutzenbewertung medizinischer Leistungen und passt die eigenen Verfahren gegebenenfalls an.

6.5.2 Konsultationstechniken

Die Bearbeitung der an das Institut herangetragenen Fragen und Aufgaben erfordert oft die Konsultation von Patienten, Patientenvertretern sowie nationalen und internationalen Experten. Dazu setzt das Institut verschiedene Konsultationstechniken ein.

Bei der Erarbeitung von Berichten nutzt das Institut diese Techniken zur Ermittlung patientenrelevanter Endpunkte, zur Einbindung nationaler und internationaler Experten sowie im Rahmen der Stellungnahmeverfahren.

Im Rahmen der Entwicklung von Gesundheitsinformationen dienen Konsultationstechniken dazu, Patienten und Patientenvertreter bei der Eruiierung von Informationsbedürfnissen, bei der Evaluation von Gesundheitsinformationen und im Rahmen des Stellungnahmeprozesses einzubeziehen.

Folgende Konsultationstechniken werden vom Institut eingesetzt:

- Interviews mit Key Informants (Schlüsselinformanten) [364], zum Beispiel Interviews mit Patientenvertretern zur Ermittlung patientenrelevanter Endpunkte

- Gruppenmeetings und Konsultationen [275,277,278], zum Beispiel im Rahmen von Erörterungen
- Gruppeninterviews und Fokusgruppen [98,277,363], zum Beispiel mit Patienten bezüglich der Evaluation von Gesundheitsinformationen
- Erfassung von Meinungen und Surveys (einschließlich Online-Umfragen und Feedbackmechanismen), zum Beispiel bei der Ermittlung der Informationsbedürfnisse der Leser von www.gesundheitsinformation.de

Wenn ein tieferes Verständnis von Erfahrungen oder Meinungen notwendig ist, dann soll das Institut die wissenschaftlichen Erkenntnisse qualitativer Forschung nutzen.

Die Anwendung von Konsultationstechniken und die Einbindung von Experten sind mit einem Mehraufwand verbunden. Die Einbeziehung von Patienten in wissenschaftliche Prozesse ermöglicht jedoch die Berücksichtigung von und die Ausrichtung des Forschungsprozesses auf Fragen und Bedürfnisse von Patienten [284].

Literaturverzeichnis

1. Gesetz zur Modernisierung der gesetzlichen Krankenversicherung (GKV-Modernisierungsgesetz - GMG). 19.11.2003. Bundesgesetzblatt 2003; Teil 1(55): 2190-2258.
2. Gesetz zur Stärkung des Wettbewerbs in der gesetzlichen Krankenversicherung (GKV-Wettbewerbsstärkungsgesetz - GKV-WSG). 26.03.2007. Bundesgesetzblatt 2007; Teil 1(11): 378-473.
3. AGREE Collaboration. Appraisal of guidelines for research & evaluation, AGREE Instrument [Online]. 2001 [Zugriff am 20.06.2007]. URL: <http://www.agreecollaboration.org/pdf/agreeinstrumentfinal.pdf>.
4. Agresti A. Modelling ordered categorical data: recent advances and future challenges. Stat Med 1999; 18(17-18): 2191-2207.
5. Agresti A. Categorical data analysis. Hoboken: Wiley; 2002.
6. Agresti A. Dealing with discreteness: making 'exact' confidence intervals for proportions, differences of proportions, and odds ratios more exact. Stat Methods Med Res 2003; 12(1): 3-21.
7. Al-Marzouki S, Evans S, Marshall T, Roberts I. Are these data real? Statistical methods for the detection of data fabrication in clinical trials. BMJ 2005; 331(7511): 267-270.
8. Altman DG. Systematic reviews of evaluations of prognostic variables. In: Egger M, Davey Smith G, Altman DG (Ed). Systematic reviews in health care: Meta-analysis in context. London: BMJ Publishing Group; 2001: 228-247.
9. Altman DG, Bland JM. Absence of evidence is not evidence of absence. BMJ 1995; 311(7003): 485.
10. Altman DG, Machin D, Bryant TM, Gardner MJ. Statistics with confidence: confidence intervals and statistical guidelines. London: BMJ Publishing Group; 2000.
11. Altman DG, Schulz K, Moher D. Turning a blind eye: testing the success of blinding and the CONSORT statement. BMJ 2004; 328(7448): 1135.
12. Altman DG, Schulz K, 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(8): 663-694.
13. American Society of Clinical Oncology. Outcomes of cancer treatment for technology assessment and cancer treatment guidelines. J Clin Oncol 1996; 14(2): 671-679.
14. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. JAMA 1992; 268(2): 240-248.
15. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF), Ärztliche Zentralstelle Qualitätssicherung (ÄZQ). Das Leitlinien-Manual von AWMF und ÄZQ. Z Ärztl Fortbild Qualitätssich 2001; 95(Suppl 1): 1-84.
16. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF), Ärztliche Zentralstelle Qualitätssicherung (ÄZQ). Deutsches Instrument zur methodischen Leitlinienbewertung (DELBI) [Online]. 2005 [Zugriff am 20.06.2007]. URL: <http://www.versorgungsleitlinien.de/methodik/delbi/pdf/delbi05.pdf>.
17. Ärztliche Zentralstelle Qualitätssicherung (ÄZQ). Checkliste "Methodische Qualität von Leitlinien". Dtsch Arztebl 1998; 95(41): A2576-A2578.
18. Ärztliches Zentrum für Qualität in der Medizin. Leitlinien-Clearingbericht „COPD“: Leitlinien-Clearingverfahren von Bundesärztekammer und Kassenärztlicher Bundesvereinigung in Kooperation mit Deutscher Krankenhausgesellschaft, Spitzenverbänden der Krankenkassen und Gesetzlicher Rentenversicherung. Köln: ÄZQ; 2003.

19. Ärztliches Zentrum für Qualität in der Medizin. Leitlinien-Clearingbericht „Depression“: Leitlinien-Clearingverfahren von Bundesärztekammer und Kassenärztlicher Bundesvereinigung in Kooperation mit Deutscher Krankenhausgesellschaft, Spitzenverbänden der Krankenkassen und Gesetzlicher Rentenversicherung. Köln: ÄZQ; 2003.
20. Ärztliches Zentrum für Qualität in der Medizin. Das Deutsche Leitlinien-Clearingverfahren 1999-2005: Hintergrund, Zielsetzung, Ergebnisse. Abschlussbericht. Berlin: ÄZQ; 2006.
21. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet* 2000; 355(9209): 1064-1069.
22. Atkins D, Best D, Briss PA, Eccles MP, Falck-Ytter Y, Flottorp S et al. Grading quality of evidence and strength of recommendations. *BMJ* 2004; 328(7454): 1490.
23. Atkins D, Briss PA, Eccles MP, Flottorp S, Guyatt GH, Harbour RT et al. Systems for grading the quality of evidence and the strength of recommendations II: pilot study of a new system. *BMC Health Serv Res* 2005; 5(1): 25.
24. Atkins D, Eccles MP, Flottorp S, Guyatt GH, Henry D, Hill S et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. *BMC Health Serv Res* 2004; 4(1): 38.
25. 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(10): 979-985.
26. Baker SG. Surrogate endpoints: wishful thinking or reality? *J Natl Cancer Inst* 2006; 98(8): 502-503.
27. Baltussen R, Niessen L. Priority setting of health interventions: the need for multi-criteria decision analysis. *Cost Eff Resour Alloc* 2006; 4: 14.
28. Baltussen R, Stolk E, Chisholm D, Aikins M. Towards a multi-criteria approach for priority setting: an application to Ghana. *Health Econ* 2006; 15(7): 689-696.
29. Bandolier. On care pathways. *Bandolier Forum* 2003; 1-12.
30. Banta D. The development of health technology assessment. *Health Policy* 2003; 63(2): 121-132.
31. Barron BA, Bukantz SC. The evaluation of new drugs. Current Food and Drug Administration regulations and statistical aspects of clinical trials. *Arch Intern Med* 1967; 119(6): 547-556.
32. Bastian H. Non-peer review: consumer involvement in research review. In: Godlee F, Jefferson T (Ed). *Peer review in health sciences*. London: BMJ Publishing Group; 2003: 248-262.
33. Bastian H, Kaiser T, Matschewsky S. Förderung allgemeiner Gesundheits- und Wissenschaftskenntnisse mittels Bürger- und Patienteninformationen: die Rolle des IQWiG. *Z Ärztl Fortbild Qual Gesundh wes* 2005; 99(6): 379-385.
34. Bates BR, Romina S, Ahmed R, Hopson D. The effect of source credibility on consumers' perceptions of the quality of health information on the internet. *Med Inform Internet Med* 2006; 31(1): 45-52.
35. Begg CB, Berlin JA. Publication bias: a problem in interpreting medical data. *J R Stat Soc [Ser A]* 1988; 151(3): 419-463.
36. Bender R. Interpretation von Effizienzmaßen der Vierfeldertafel für Diagnostik und Behandlung. *Med Klin* 2001; 96(2): 116-121.
37. Bender R, Bunce C, Clarke M, Gates S, Lange S, Pace NL et al. Attention should be given to multiplicity issues in systematic reviews. *J Clin Epidemiol* 2008; 61: (in press).
38. Bender R, Grouven U. Logistic regression models used in medical research are poorly presented. *BMJ* 1996; 313(7057): 628.
39. Bender R, Grouven U. Ordinal logistic regression in medical research. *J R Coll Physicians Lond* 1997; 31(5): 546-551.

40. Bender R, Lange S. Adjusting for multiple testing - when and how? *J Clin Epidemiol* 2001; 54(4): 343-349.
41. Bent S, Padula A, Avins AL. Brief communication: better ways to question patients about adverse medical events. A randomized, controlled trial. *Ann Intern Med* 2006; 144(4): 257-261.
42. Bessell TL, McDonald S, Silagy CA, Anderson JN, Hiller JE, Sansom LN. Do internet interventions for consumers cause more harm than good? A systematic review. *Health Expect* 2002; 5(1): 28-37.
43. Biglan A, Ary D, Wagenaar AC. The value of interrupted time-series experiments for community intervention research. *Prev Sci* 2000; 1(1): 31-49.
44. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001; 69(3): 89-95.
45. Bland JM, Altman DG. One and two sided tests of significance. *BMJ* 1994; 309(6949): 248.
46. Bock J, Toutenburg H. Sample size determination in clinical research. In: Rao CR, Chakraborty R (Ed). *Statistical Methods in Biological and Medical Sciences*. Amsterdam: Elsevier; 1991: 515-538.
47. Bonell C, Oakley A, Hargreaves J, Strange V, Rees R. Assessment of generalisability in trials of health interventions: suggested framework and systematic review. *BMJ* 2006; 333(7563): 346-349.
48. Bonhoeffer J, Zumbunn B, Heininger U. Reporting of vaccine safety data in publications: systematic review. *Pharmacoepidemiol Drug Saf* 2005; 14(2): 101-106.
49. Bossuyt PM, Irwig LM, Craig J, Glasziou PP. Comparative accuracy: assessing new tests against existing diagnostic pathways. *BMJ* 2006; 332(7549): 1089-1092.
50. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD Initiative. *Ann Intern Med* 2003; 138(1): 40-44.
51. 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(1): W1-W12.
52. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Intern Med* 2008; 148(4): 295-309.
53. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. Methods and processes of the CONSORT Group: example of an extension for trials assessing nonpharmacologic treatments. *Ann Intern Med* 2008; 148(4): W60-W66.
54. Boynton J, Glanville J, McDaid D, Lefebvre C. Identifying systematic reviews in MEDLINE: Developing an objective approach to search strategy design. *J Inf Sci* 1998; 24(3): 137-155.
55. Breslow NE, Day NE. *Statistical methods in cancer research, Vol. I - The analysis of case-control studies*. Lyon: International Agency for Research on Cancer; 1980.
56. Breslow NE, Day NE. *Statistical methods in cancer research, Vol. II - The design and analysis of cohort studies*. Lyon: International Agency for Research on Cancer; 1987.
57. Brin S, Page L. The anatomy of a large-scale hypertextual web search engine [Online]. 2002 [Zugriff am 20.06.2007]. URL: <http://www.stanford.edu/class/cs240/readings/google.pdf>.
58. British Cardiac Society, British Hyperlipidaemia Association, British Hypertension Society. Joint british recommendations on prevention of coronary heart disease in clinical practice. *Heart* 1998; 80(Suppl 2): S1-S29.
59. Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. *Stat Med* 2001; 20(6): 825-840.
60. Brookes ST, Whitely E, Egger M, Smith GD, 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(3): 229-236.

61. Bundesministerium des Innern. Verordnung zur Schaffung barrierefreier Informationstechnik nach dem Behindertengleichstellungsgesetz. Berlin: 2002.
62. Bundesverfassungsgericht. Leitsatz zum Beschluss des Ersten Senats. 06.12.2005: 1 BvR 347/98.
63. Burgers JS, Cluzeau FA, Hanna SE, Hunt C, Grol R. 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(1): 148-157.
64. Busse R, Drösler S, Glaeske G, Greiner W, Schäfer T, Schrappe M. Wissenschaftliches Gutachten für die Auswahl von 50 bis 80 Krankheiten zur Berücksichtigung im morbiditätsorientierten Risikostrukturausgleich. Wissenschaftlicher Beirat zur Weiterentwicklung des Risikostrukturausgleichs; 2007.
65. Campbell MK, Elbourne DR, Altman DG, CONSORT Group. CONSORT statement: extension to cluster randomised trials. *BMJ* 2004; 328(7441): 702-708.
66. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. Measurement error in nonlinear models: a modern perspective. London: Chapman & Hall; 2006.
67. Centre for Evidence-based Medicine. Levels of evidence [Online]. 2001 [Zugriff am 07.05.2007]. URL: http://www.cebm.net/levels_of_evidence.asp.
68. Charles C, Gafni A, Whelan T, O'Brien MA. Treatment decision aids: conceptual issues and future directions. *Health Expect* 2005; 8(2): 114-125.
69. Cheng CL, Van Ness JW. Statistical regression with measurement error. London: Arnold; 1999.
70. Cluzeau FA, Littlejohns P, Grimshaw JM, Feder G, Moran SE. Development and application of a generic methodology to assess the quality of clinical guidelines. *Int J Qual Health Care* 1999; 11(1): 21-28.
71. Cochrane Collaboration. The Cochrane Collaboration - Principles [Online]. 2006 [Zugriff am 01.05.2006]. URL: <http://www.cochrane.org/docs/tenprinciples.htm>.
72. Cochrane Effective Practice and Organisation of Care Review Group. The data collection checklist. Ottawa: EPOC; 2002.
73. Cochrane Qualitative Research Methods Group. Methodological issues arising from the inclusion of qualitative evidence in systematic reviews [Online]. 2007 [Zugriff am 30.04.2007]. URL: <http://www.lancs.ac.uk/fass/ihr/research/public/cochrane.htm>.
74. Cochrane Qualitative Research Methods Group. What is qualitative research? What is its role in evidence review? [Online]. 2007 [Zugriff am 30.04.2007]. URL: <http://www.joannabriggs.edu.au/cqrng/role.html>.
75. Commission of the European Communities. eEurope 2002: quality criteria for health related websites. Brussels: 2002.
76. Committee for Proprietary Medicinal Products (CPMP). Points to consider on application with 1. Meta-analyses; 2. One pivotal study. London: EMEA; 2001.
77. Coulter A. Evidence based patient information is important, so there needs to be a national strategy to ensure it. *BMJ* 1998; 317(7153): 225-226.
78. Coulter A, Ellins J, Swain D, Clarke A, Heron P, Rasul F et al. Assessing the quality of information to support people in making decisions about their health and healthcare. Oxford: Picker Institute Europe; 2006.
79. Council of Europe: Committee of Experts on Management of Safety and Quality in Health Care (SP-SQS) EGoSMP. Glossary of terms related to patient and medication safety. Strasbourg: Council of Europe; 2005.
80. Cui L, Hung HM, Wang SJ, Tsong Y. Issues related to subgroup analysis in clinical trials. *J Biopharm Stat* 2002; 12(3): 347-358.

81. Cullen R. Health information on the internet: a study of providers, quality, and users. Westport: Praeger; 2006.
82. D'Agostino RB, Sr., Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues. The encounters of academic consultants in statistics. *Stat Med* 2003; 22(2): 169-186.
83. Dans AL, Dans LF, Guyatt GH, Richardson S, Evidence-Based Medicine Working Group. Users' guides to the medical literature: XIV. How to decide on the applicability of clinical trial results to your patient. *JAMA* 1998; 279(7): 545-549.
84. David M, Borde T, Kentenich H. Knowledge among German and Turkish women about specifically female bodily functions, contraception, preventative medical examinations and menopause. *Ethn Health* 2000; 5(2): 101-112.
85. Deeks JJ. Systematic reviews in health care: systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001; 323(7305): 157-162.
86. Deeks JJ, Higgins JPT, Altman DG, Eds. Chapter 9: Analysing and presenting results. In: Higgins JPT, Green S, Eds. *Cochrane handbook for systematic reviews of interventions*, Version 5.0.0 (updated February 2008) [Online]. 2008 [Zugriff am 13.05.2008]. URL: <http://www.cochrane-handbook.org>.
87. Derksen S, Keselman HJ. Backward, forward, and stepwise automated subset selection algorithms: frequency of obtaining authentic and noise variables. *Br J Math Statist Psych* 1992; 45: 265-282.
88. Derry S, Loke YK, Aronson JK. Incomplete evidence: the inadequacy of databases in tracing published adverse drug reactions in clinical trials. *BMC Med Res Methodol* 2001; 1: 7.
89. Des Jarlais DC, Lyles C, Crepaz N, TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Public Health* 2004; 94(3): 361-366.
90. Desu MM, Raghavarao D. Sample size methodology. Boston: Academic Press; 1990.
91. Deutsche Gesellschaft für Epidemiologie (DGEpi). Leitlinien und Empfehlungen zur Sicherung von guter epidemiologischer Praxis (GEP). Deutsche Gesellschaft für Epidemiologie (DGEpi); 2008.
92. Deutsches Institut für Normung. Klinische Prüfung von Medizinprodukten an Menschen. Teil 2: klinische Prüfpläne (ISO 14155-2:2003). Berlin: Beuth; 2003.
93. Deutsches Netzwerk Evidenzbasierte Medizin e.V. Definitionen [Online]. 2007 [Zugriff am 25.09.2007]. URL: <http://www.ebm-netzwerk.de/grundlagen/grundlagen/definitionen>.
94. Di Pietrantonj C. Four-fold table cell frequencies imputation in meta analysis. *Stat Med* 2006; 25(13): 2299-2322.
95. Dierks ML, Seidel G, Schwartz FW, Horch K. Bürger- und Patientenorientierung im Gesundheitswesen: Gesundheitsberichterstattung des Bundes - Heft 32. Berlin: Robert Koch Institut; 2006.
96. Dimensional. Papaya CMS - pflegen Sie Ihre Website ohne Programmierkenntnisse [Online]. 2007 [Zugriff am 27.05.2007]. URL: http://www.dimensional.de/index_ndim85.php.
97. Dixon-Woods M. Writing wrongs? An analysis of published discourses about the use of patient information leaflets. *Soc Sci Med* 2001; 52(9): 1417-1432.
98. Dixon-Woods M, Agarwal S, Young B, Jones D, Sutton A. Integrative approaches to qualitative and quantitative evidence. London: NHS Health Development Agency; 2004.
99. Dixon-Woods M, Fitzpatrick R. Qualitative research in systematic reviews. Has established a place for itself. *BMJ* 2001; 323(7316): 765-766.
100. Dixon-Woods M, Fitzpatrick R, Roberts K. Including qualitative research in systematic reviews: opportunities and problems. *J Eval Clin Pract* 2001; 7(2): 125-133.

101. Dixon-Woods M, Sutton A, Shaw R, Miller T, Smith J, Young B et al. Appraising qualitative research for inclusion in systematic reviews: a quantitative and qualitative comparison of three methods. *J Health Serv Res Policy* 2007; 12(1): 42-47.
102. Donner A, Klar J. Design and analysis of cluster randomization trials in health research. London: Arnold; 2000.
103. Draborg E, Gyrd-Hansen D, Poulsen PB, Horder M. International comparison of the definition and the practical application of health technology assessment. *Int J Technol Assess Health Care* 2005; 21(1): 89-95.
104. Drazen JM. COX-2 inhibitors - a lesson in unexpected problems. *N Engl J Med* 2005; 352(11): 1131-1132.
105. Eccles MP, Rousseau N, Freemantle N. Updating evidence-based clinical guidelines. *J Health Serv Res Policy* 2002; 7(2): 98-103.
106. Edwards AG, Bastian H. Risk communication: making evidence part of patient choices? In: Edwards AG, Elwyn GJ (Ed). *Evidence-based patient choice: Inevitable or impossible?* Oxford: Oxford University Press; 2001: 144-160.
107. Edwards AG, Elwyn GJ, Mulley A. Explaining risks: turning numerical data into meaningful pictures. *BMJ* 2002; 324(7341): 827-830.
108. Edwards AG, Evans R, Dundon J, Haigh S, Hood K, Elwyn GJ. Personalised risk communication for informed decision making about taking screening tests. *Cochrane Database Syst Rev* 2006; 4: CD001865.
109. 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(6): 427-437.
110. Egger M, Davey Smith G, Altman DG. *Systematic reviews in health care: meta-analysis in context*. London: BMJ Publishing Group; 2001.
111. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple graphical test. *BMJ* 1997; 315(7109): 629-634.
112. Egger M, Jüni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technol Assess* 2003; 7(1): 1-76.
113. Elwyn GJ, O'Connor A, Stacey D, Volk R, Edwards AG, Coulter A et al. Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. *BMJ* 2006; 333(7565): 417.
114. Engels EA, Schmid C, Terrin N, Olkin I, Lau J. Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. *Stat Med* 2000; 19(13): 1707-1728.
115. Entwistle VA, Sheldon TA, Sowden AJ, Watt I. Evidence-informed patient choice. Practical issues of involving patients in decisions about health care technologies. *Int J Technol Assess Health Care* 1998; 14(2): 212-225.
116. Epstein RM, Alper BS, Quill TE. Communicating evidence for participatory decision making. *JAMA* 2004; 291(19): 2359-2366.
117. Europäisches Parlament, Rat der Europäischen Union. Verordnung (EG) Nr. 141/2000 des Europäischen Parlaments und des Rates vom 16. Dezember 1999 über Arzneimittel für seltene Leiden. *Amtsbl Eur Gemeinschaften* 2000; L18: L18-1-L18/5; 22. Jan. 2000.
118. Europarat. Entwicklung einer Methodik für die Ausarbeitung von Leitlinien für optimale medizinische Praxis: Empfehlung Rec(2001)13 des Europarates und Erläuterndes Memorandum. Strasbourg: Council of Europe; 2002.
119. European Medicines Agency. Note for guidance on clinical investigation of medicinal products in the treatment of diabetes mellitus. London: EMEA; 2002.

120. European Medicines Agency. Guideline on clinical trials in small populations. London: EMEA; 2005.
121. European Medicines Agency. Guideline on the choice of the non-inferiority margin. London: EMEA; 2005.
122. European Medicines Agency. Reflection paper on the regulatory guidance for the use of Health Related Quality of Life (HRQL) measures in the evaluation of medicinal products. London: EMEA; 2005.
123. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA 1992; 268(17): 2420-2425.
124. Eysenbach G. Recent advances: consumer health informatics. BMJ 2000; 320(7251): 1713-1716.
125. 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(7337): 573-577.
126. Eysenbach G, Tuische J, Diepgen TL. Evaluation of the usefulness of Internet searches to identify unpublished clinical trials for systematic reviews. Med Inform Internet Med 2001; 26(3): 203-218.
127. Feinstein AR. Clinical epidemiology: the architecture of clinical research. Philadelphia: Saunders; 1985.
128. Feinstein AR. Invidious comparisons and unmet clinical challenges. Am J Med 1992; 92(2): 117-120.
129. Feldman-Stewart D, Brennenstuhl S, Brundage MD. A purpose-based evaluation of information for patients: an approach to measuring effectiveness. Patient Educ Couns 2007; 65(3): 311-319.
130. Fergusson D, Glass KC, Waring D, Shapiro S. Turning a blind eye: the success of blinding reported in a random sample of randomised, placebo controlled trials. BMJ 2004; 328(7437): 432-434.
131. 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(4): 1109-1119.
132. Fervers B, Burgers JS, Haugh MC, Latreille J, Mlika-Cabanne N, Paquet L et al. Adaptation of clinical guidelines: literature review and proposition for a framework and procedure. Int J Qual Health Care 2006; 18(3): 167-176.
133. Field MJ, Lohr KN. Clinical practice guidelines: directions for a new program. Washington, D. C.: National Academy Press; 1990.
134. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? Ann Intern Med 1996; 125(7): 605-613.
135. Fletcher RH, Fletcher SW. The effectiveness of journal peer review. In: Godlee F, Jefferson T (Ed). Peer review in health sciences. London: BMJ Publishing Group; 2003: 62-75.
136. Food and Drug Administration. Guidance for industry: developing medical imaging drug and biological products. Part 2: clinical indications. Rockville: FDA; 2004.
137. Forde OH. Is imposing risk awareness cultural imperialism? Soc Sci Med 1998; 47(9): 1155-1159.
138. Fransen GA, van Marrewijk CJ, Mujakovic S, Muris JW, Laheij RJ, Numans ME et al. Pragmatic trials in primary care. Methodological challenges and solutions demonstrated by the DIAMOND-study. BMC Med Res Methodol 2007; 7: 16.
139. 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(7292): 989-991.
140. Freemantle N, Blonde L, Bolinder B, Gerber RA, Hobbs FD, Martinez L et al. Real-world trials to answer real-world questions. Pharmacoeconomics 2005; 23(8): 747-754.
141. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. Med Decis Making 1991; 11(2): 88-94.

142. 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(1): 72-76.
143. 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(7337): 569-573.
144. Garcia-Castillo D, Feters MD. Quality in medical translations: a review. *J Health Care Poor Underserved* 2007; 18(1): 74-84.
145. Gemeinsamer Bundesausschuss. Verfahrensordnung des Gemeinsamen Bundesausschusses vom 20. September 2005. Siegburg: G-BA; 2006.
146. Gemoets D, Rosemblat G, Tse T, Logan R. Assessing readability of consumer health information: an exploratory study. *Medinfo* 2004; 11(Pt 2): 869-873.
147. Glasziou PP, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. *BMJ* 2007; 334(7589): 349-351.
148. Glasziou PP, Sanders SL. Investigating causes of heterogeneity in systematic reviews. *Stat Med* 2002; 21(11): 1503-1511.
149. Glasziou PP, Vandenbroucke JP, Chalmers I. Assessing the quality of research. *BMJ* 2004; 328(7430): 39-41.
150. Glenton C, Nilsen ES, Carlsen B. Lay perceptions of evidence-based information - a qualitative evaluation of a website for back pain sufferers. *BMC Health Serv Res* 2006; 6: 34.
151. Godwin M, Ruhland L, Casson I, MacDonald S, Delva D, Birtwhistle R et al. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. *BMC Med Res Methodol* 2003; 3: 28.
152. Golder S, McIntosh HM, Duffy S, Glanville J. Developing efficient search strategies to identify reports of adverse effects in MEDLINE and EMBASE. *Health Info Libr J* 2006; 23(1): 3-12.
153. Gøtzsche PC, Liberati A, Torri V, Rossetti L. Beware of surrogate outcome measures. *Int J Technol Assess Health Care* 1996; 12(2): 238-246.
154. GRADE Working Group. The GRADE Working Group [Online]. 2007 [Zugriff am 06.05.2007]. URL: <http://www.gradeworkinggroup.org/index.htm>
155. Gray JAM. How to get better value healthcare. Oxford: Offox Press; 2007.
156. Greenhalgh T, Hurwitz B. Narrative based medicine: why study narrative? *BMJ* 1999; 318(7175): 48-50.
157. Greenland S, Morgenstern H. Ecological bias, confounding, and effect modification. *Int J Epidemiol* 1989; 18(1): 269-274.
158. 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(1): 159-163.
159. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. *J Med Internet Res* 2006; 8(2): e10.
160. Grimes DA, Schulz K. An overview of clinical research: the lay of the land. *Lancet* 2002; 359(9300): 57-61.
161. Grimes DA, Schulz K. Surrogate end points in clinical research: hazardous to your health. *Obstet Gynecol* 2005; 105(5 Pt 1): 1114-1118.
162. Grimshaw JM, Eccles MP, Russell I. Developing clinically valid practice guidelines. *J Eval Clin Pract* 1995; 1(1): 37-48.
163. Grol R, Dalhuijsen J, Thomas S, Veld C, Rutten G, Mookink H. Attributes of clinical guidelines that influence use of guidelines in general practice: observational study. *BMJ* 1998; 317(7162): 858-861.

164. Grouin JM, Coste M, Lewis J. Subgroup analyses in randomized clinical trials: statistical and regulatory issues. *J Biopharm Stat* 2005; 15(5): 869-882.
165. Guyatt GH. Evidence-based medicine. *ACP J Club* 1991; 14 (Suppl 2): A-16.
166. Guyatt GH, Gutterman D, Baumann MH, Addrizzo-Harris D, Hylek EM, Phillips B et al. Grading strength of recommendations and quality of evidence in clinical guidelines: report from an american college of chest physicians task force. *Chest* 2006; 129(1): 174-181.
167. Guyatt GH, Jaeschke R, Roberts R. N-of-1 randomized clinical trials in pharmacoepidemiology. In: Strom BL (Ed). *Pharmacoepidemiology*. Chichester: Wiley; 2005: 665-680.
168. Guyatt GH, Rennie D. *Users' guides to the medical literature: a manual for evidence-based clinical practice*. Chicago: American Medical Association; 2002.
169. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ. *Users' guides to the medical literature. IX: a method for grading health care recommendations*. *JAMA* 1995; 274(22): 1800-1804.
170. Guyatt GH, Sackett DL, Taylor DW, Chong J, Roberts R, Pugsley S. Determining optimal therapy - randomized trials in individual patients. *N Engl J Med* 1986; 314(14): 889-892.
171. Guyatt GH, Tugwell P, Feeny DH, Drummond MF, Haynes RB. The role of before-after studies of therapeutic impact in the evaluation of diagnostic technologies. *J Chronic Dis* 1986; 39(4): 295-304.
172. Guyatt GH, Vist GE, Falck-Ytter Y, Kunz R, Magrini N, Schünemann HJ. An emerging consensus on grading recommendations? *ACP J Club* 2006; 144(1): A8-A9.
173. Harbour RT, Miller J. A new system for grading recommendations in evidence based guidelines. *BMJ* 2001; 323(7308): 334-336.
174. Harden A, Garcia J, Oliver S, Rees R, Shepherd J, Brunton G et al. Applying systematic review methods to studies of people's views: an example from public health research. *J Epidemiol Community Health* 2004; 58(9): 794-800.
175. Harrell FE. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. New York: Springer; 2001.
176. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; 15(4): 361-387.
177. Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM et al. Current methods of the US Preventive Services Task Force: a review of the process. *Am J Prev Med* 2001; 20(3 Suppl): 21-35.
178. Hayden JA, Cote P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. *Ann Intern Med* 2006; 144(6): 427-437.
179. Haynes RB, Cotoi C, Holland J, Walters L, Wilczynski N, Jedraszewski D et al. Second-order peer review of the medical literature for clinical practitioners. *JAMA* 2006; 295(15): 1801-1808.
180. Haynes RB, Devereaux PJ, Guyatt GH. Clinical expertise in the era of evidence-based medicine and patient choice. *Evid Based Med* 2002; 7: 36-38.
181. Health on the Net (HON). HON Code of Conduct (HONcode) für medizinische und Gesundheits-Websites [Online]. 1997 [Zugriff am 25.09.2007]. URL: <http://www.hon.ch/HONcode/German>.
182. 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 Qualitätssich* 1998; 92(5): 361-365.
183. Herxheimer A, McPherson A, Miller R, Shepperd S, Yappe J, Ziebland S. Database of patients' experiences (DIPEX): a multi-media approach to sharing experiences and information. *Lancet* 2000; 355(9214): 1540-1543.

184. Herxheimer A, Ziebland S. DIPEX: fresh insights for medical practice. *J R Soc Med* 2003; 96(5): 209-210.
185. Higgins JPT, Green S, Eds. *Cochrane handbook for systematic reviews of interventions*, Version 5.0.0 (updated February 2008) [Online]. 2008 [Zugriff am 13.05.2008]. URL: <http://www.cochrane-handbook.org>.
186. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; 21(11): 1539-1558.
187. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327(7414): 557-560.
188. Hill AB, Ed. *Controlled clinical trials*. Oxford: Blackwell; 1960.
189. Holmes-Rovner M. International patient decision aid standards (IPDAS): Beyond decision aids to usual design of patient education materials. *Health Expect* 2007; 10(2): 103-107.
190. Hopewell S, Clarke M, Lefebvre C, Scherer R. Handsearching versus electronic searching to identify reports of randomized trials. [Cochrane Review]. *Cochrane Database Syst Rev* 2007; 2: MR000001.
191. Hopewell S, McDonald S, Clarke M, Egger M. Grey literature in meta-analyses of randomized trials of health care interventions. [Cochrane Review]. *Cochrane Database Syst Rev* 2007; 2: MR000010.
192. Hosmer DW, Taber S, Lemeshow S. The importance of assessing the fit of logistic regression models: a case study. *Am J Public Health* 1991; 81(12): 1630-1635.
193. Houts PS, Doak CC, Doak LG, Loscalzo MJ. The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. *Patient Educ Couns* 2006; 61(2): 173-190.
194. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol* 2005; 5(1): 13.
195. Hung HM, O'Neill RT, Bauer P, Kohne K. The behavior of the P-value when the alternative hypothesis is true. *Biometrics* 1997; 53(1): 11-22.
196. Inan H. *Measuring the success of your website: a customer-centric approach to website management*. Frenchs Forest: Pearson Education Australia; 2002.
197. Informationsgemeinschaft zur Feststellung der Verbreitung von Werbeträgern e.V. (IVW). Definitionen und technische Erläuterungen Version 1.9, Anlage 1 zu den IVW-Richtlinien für Online-Angebote. Berlin: IVW; 2007.
198. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Vereinbarung über die vertrauliche Behandlung von Unterlagen [Online]. 2005 [Zugriff am 30.11.2006]. URL: http://www.iqwig.de/download/VertraulichkeitIQWiG-VFA_19-8-05_final_2.pdf.
199. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Datenschutzerklärung [Online]. 2007 [Zugriff am 25.09.2007]. URL: <http://www.gesundheitsinformation.de/nutzungsbedingungen.51.de.html>.
200. International Conference on Harmonisation E9 Expert Working Group. ICH harmonised tripartite guideline. Statistical principles for clinical trials. *Stat Med* 1999; 18(15): 1905-1942.
201. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH harmonised tripartite guideline. The extent of population exposure to assess clinical safety: for drugs intended for long-term treatment of non-life-threatening conditions. Geneva: ICH; 1994.
202. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. Official web site for ICH [Online]. 2004 [Zugriff am 20.10.2004]. URL: <http://www.ich.org>.
203. Ioannidis JPA. Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 2005; 294(2): 218-228.

204. Ioannidis JPA, Evans S, Gøtzsche PC, O'Neill RT, Altman DG, Schulz K et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med* 2004; 141(10): 781-788.
205. Ioannidis JPA, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. *JAMA* 2001; 285(4): 437-443.
206. Ioannidis JPA, Mulrow CD, Goodman SN. Adverse events: the more you search, the more you find. *Ann Intern Med* 2006; 144(4): 298-300.
207. Irani J, Brown CT, Van der Meulen JH, Emberton M. A review of guidelines on benign prostatic hyperplasia and lower urinary tract symptoms: are all guidelines the same? *BJU Int* 2003; 92(9): 937-942.
208. Jackson D. The power of the standard test for the presence of heterogeneity in meta-analysis. *Stat Med* 2006; 25(15): 2688-2699.
209. Jackson N, Waters E. Criteria for the systematic review of health promotion and public health interventions. *Health Promot Int* 2005; 20(4): 367-374.
210. Jadad AR. Randomised controlled trials: a user's guide. London: BMJ Publishing Group; 1998.
211. Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. *CMAJ* 1997; 156(10): 1411-1416.
212. Jadad AR, Gagliardi A. Rating health information on the internet: navigating to knowledge or to Babel? *JAMA* 1998; 279(8): 611-614.
213. James Lind Library. Editorial commentary: Avoiding biased comparisons [Online]. 2007 [Zugriff am 01.05.2007]. URL: <http://www.jameslindlibrary.org/essays/bias/avoiding-biased-comparisons.pdf>.
214. James A, Horton R, Collingridge D, McConnell J, Butcher J. The Lancet's policy on conflicts of interest. *Lancet* 2003; 363(9402): 8-9.
215. Jefferson T, Rudin M, Brodny FS, Davidoff F. Editorial peer-review for improving the quality of reports of biomedical studies [Cochrane Methodology Review]. *Cochrane Database Methodol Rev* 2006; 1: MR000016.
216. Jones B, Jarvis P, Lewis J, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. *BMJ* 1996; 313(7048): 36-39.
217. Jones B, Kenward MG. Design and analysis of cross-over trials. London: Chapman & Hall; 1989.
218. Jørgensen AW, Hilden J, Gøtzsche PC. Cochrane reviews compared with industry supported meta-analyses and other meta-analyses of the same drugs: Systematic review. *BMJ* 2006; 333(7572): 782.
219. Jull A, Bennett D. Do n-of-1 trials really tailor treatment? *Lancet* 2005; 365(9476): 1992-1994.
220. 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(1): 115-123.
221. Kaltenthaler E, Brazier J, De Nigris E, Tumor I, Ferriter M, Beverley C et al. Computerised cognitive behaviour therapy for depression and anxiety update: a systematic review and economic evaluation. *Health Technol Assess* 2006; 10(33): iii, xi-iii, 168.
222. Kaptchuk TJ, Stason WB, Davis RB, Legedza AR, Schnyer RN, Kerr CE et al. Sham device v inert pill: randomised controlled trial of two placebo treatments. *BMJ* 2006; 332(7538): 391-397.
223. Kass PH, Gold EB. Modern epidemiologic study designs. In: Ahrens W, Pigeot I (Ed). *Handbook of Epidemiology*. Berlin: Springer; 2005: 321-344.
224. Katrak P, Bialocerkowski AE, Massy-Westropp N, Kumar S, Grimmer KA. A systematic review of the content of critical appraisal tools. *BMC Med Res Methodol* 2004; 4: 22.
225. Katz MH. Multivariable analysis: a primer for readers of medical research. *Ann Intern Med* 2003; 138(8): 644-650.

226. Kerr CE, Murray E, Stevenson F, Gore C, Nazareth I. Internet interventions for long-term conditions: patient and caregiver quality criteria. *J Med Internet Res* 2006; 8(3): e13.
227. Kettunen T, Liimatainen L, Villberg J, Perko U. Developing empowering health counseling measurement. Preliminary results. *Patient Educ Couns* 2006; 64(1-3): 159-166.
228. Kickbusch IS. Health literacy: addressing the health and education divide. *Health Promot Int* 2001; 16(3): 289-297.
229. Kieser M, Röhm J, Friede T. Power and sample size determination when assessing the clinical relevance of trial results by 'responder analyses'. *Stat Med* 2004; 23(21): 3287-3305.
230. Kim SY, Millard RW, Nisbet P, Cox C, Caine ED. Potential research participants' views regarding researcher and institutional financial conflicts of interest. *J Med Ethics* 2004; 30(1): 73-79.
231. Köbberling J. Der Zweifel als Triebkraft des Erkenntnisgewinns in der Medizin. In: Kunz R, Ollenschläger G, Raspe H, Jonitz G, Donner-Banzhoff N (Ed). *Lehrbuch evidenz-basierte Medizin in Klinik und Praxis*. Köln: Deutscher Ärzteverlag; 2007: 3-14.
232. Köbberling J, Trampisch HJ, Windeler J. Memorandum for the evaluation of diagnostic measures. *J Clin Chem Clin Biochem* 1990; 28(12): 873-879.
233. Koch G. No Improvement - still less than half of the Cochrane reviews are up to date. In: Cochrane Collaboration (Ed). *XIV Cochrane Colloquium. Programme and Abstract Book*. Dublin: Cochrane Collaboration; 2006: 104.
234. Kolman J, Meng P, Scott G. *Good clinical practice: standard operating procedures for clinical researchers*. Chichester: Wiley; 1998.
235. Kommission der Europäischen Gemeinschaften. Richtlinie 2003/63/EG der Kommission vom 25. Juni 2003 zur Änderung der Richtlinie 2001/83/EG des Europäischen Parlaments und des Rates zur Schaffung eines Gemeinschaftskodexes für Humanarzneimittel, Abschnitt 5.2.5.1: Berichte über kontrollierte klinische Studien zur angegebenen Indikation. *Amtsbl Eur Gemeinschaften* 2005; L 159: 76.
236. Kools M, Van de Wiel MW, Ruiter RA, Kok G. Pictures and text in instructions for medical devices: effects on recall and actual performance. *Patient Educ Couns* 2006; 64(1-3): 104-111.
237. Köpke S, Berger B, Steckelberg A, Meyer G. In Deutschland gebräuchliche Bewertungsinstrumente für Patienteninformationen: eine kritische Analyse. *Z Ärztl Fortbild Qual Gesundh wes* 2005; 99(6): 353-357.
238. Kraemer HC, Frank E, Kupfer DJ. Moderators of treatment outcomes: clinical, research, and policy importance. *JAMA* 2006; 296(10): 1286-1289.
239. Kranich C. Patientenkompetenz: was müssen Patienten wissen und können? *Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz* 2004; 47(10): 950-956.
240. Krug S. *Don't make me think: a common sense approach to web usability*. Indianapolis: New Riders Publishing; 2000.
241. Kunz R, Vist GE, Oxman AD. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database Syst Rev* 2007; 2: MR000012.
242. Laaser U, Hurrelmann K. Gesundheitsförderung und Krankheitsprävention. In: Hurrelmann K, Laaser U (Ed). *Handbuch Gesundheitswissenschaften*. Weinheim: Juventa; 1998: 395-424.
243. Lange S, Freitag G. Choice of delta: requirements and reality. Results of a systematic review. *Biom J* 2005; 47(1): 12-27.
244. Laupacis A, Wells G, Richardson WS, Tugwell P, Evidence-Based Medicine Working Group. Users' guides to the medical literature. V: how to use an article about prognosis. *JAMA* 1994; 272(3): 234-237.
245. Lehoux P, Tailliez S, Denis JL, Hivon M. Redefining health technology assessment in Canada: diversification of products and contextualization of findings. *Int J Technol Assess Health Care* 2004; 20(3): 325-336.

246. Lewis D. Computer-based approaches to patient education: a review of the literature. *J Am Med Inform Assoc* 1999; 6(4): 272-282.
247. Lewis S, Clarke M. Forest plots: trying to see the wood and the trees. *BMJ* 2001; 322(7300): 1479-1480.
248. Leys M. Health care policy: qualitative evidence and health technology assessment. *Health Policy* 2003; 65(3): 217-226.
249. Liberati A, Sheldon TA, Banta HD. EUR-ASSESS project subgroup report on methodology: methodological guidance for the conduct of health technology assessment. *Int J Technol Assess Health Care* 1997; 13(2): 186-219.
250. Lipkus IM, Hollands JG. The visual communication of risk. *J Natl Cancer Inst Monogr* 1999; (25): 149-163.
251. Little RJA, Rubin DB. Statistical analysis with missing data. Hoboken: Wiley; 2002.
252. Loke YK, Price D, Herxheimer A, Eds. Chapter 14: Adverse effects. In: Higgins JPT, Green S, Eds. *Cochrane handbook for systematic reviews of interventions*, Version 5.0.0 (updated February 2008) [Online]. 2008 [Zugriff am 13.05.2008]. URL: <http://www.cochrane-handbook.org>.
253. Lord SJ, Irwig LM, Simes RJ. When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? *Ann Intern Med* 2006; 144(11): 850-855.
254. Macaskill P, Walter SD, Irwig LM. A comparison of methods to detect publication bias in meta-analysis. *Stat Med* 2001; 20(4): 641-654.
255. 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(1): 44-51.
256. Malterud K. The art and science of clinical knowledge: evidence beyond measures and numbers. *Lancet* 2001; 358(9279): 397-400.
257. March JS, Silva SG, Compton S, Shapiro M, Califf R, Krishnan R. The case for practical clinical trials in psychiatry. *Am J Psychiatry* 2005; 162(5): 836-846.
258. Martini P. Methodenlehre der therapeutischen Untersuchung. Berlin: Julius Springer; 1932.
259. McAlister FA, Straus SE, Sackett DL, Altman DG. Analysis and reporting of factorial trials: a systematic review. *JAMA* 2003; 289(19): 2545-2553.
260. 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(9237): 1228-1231.
261. McCulloch P, Taylor I, Sasako M, Loyett B, Griffin D. Randomised trials in surgery: problems and possible solutions. *BMJ* 2002; 324(7351): 1448-1451.
262. Medical Services Advisory Committee. Guidelines for the assessment of diagnostic technologies. 2005.
263. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet* 1999; 354(9193): 1896-1900.
264. Moher D, Pham B, Klassen TP, Schulz K, Berlin JA, Jadad AR et al. What contributions do languages other than English make on the results of meta-analyses? *J Clin Epidemiol* 2000; 53(9): 964-972.
265. Moher D, Schulz K, Altman DG, CONSORT Group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann Intern Med* 2001; 134(8): 657-662.
266. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med* 2007; 4(3): e78.

267. Molenaar S, Sprangers M, Oort F, Rutgers E, Luiten E, Mulder J et al. Exploring the black box of a decision aid: what information do patients select from an interactive cd-rom on treatment options in breast cancer? *Patient Educ Couns* 2007; 65(1): 122-130.
268. Montori VM, Wilczynski N, Morgan D, Haynes RB. Optimal search strategies for retrieving systematic reviews from Medline: analytical survey. *BMJ* 2005; 330(7482): 68.
269. 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 care information. *Health Expect* 2004; 7(2): 165-175.
270. Müllner M, Matthews H, Altman DG. Reporting on statistical methods to adjust for confounding: a cross-sectional survey. *Ann Intern Med* 2002; 136(2): 122-126.
271. Munday J. *Introducing translation studies*. London: Routledge; 2001.
272. Murphy E, Dingwall R, Greatbatch D, Parker S, Watson P. Qualitative research methods in health technology assessment: a review of the literature. *Health Technol Assess* 1998; 2(16)
273. Murray E, Burns J, See TS, Lai R, Nazareth I. Interactive health communication applications for people with chronic disease [Cochrane Review]. *Cochrane Database Syst Rev* 2005; 4: CD004274.
274. National Advisory Committee on Health and Disability. Screening to improve health in New Zealand. Criteria to assess screening. 2003.
275. National Health and Medical Research Council. Statement on consumer and community participation in health and medical research. Canberra: 2002.
276. National Institute of Clinical Studies. The impact of the internet on consumers health behaviour. Melbourne: NICS; 2003.
277. National Resource Centre for Consumer Participation in Health. Feedback, participation and diversity: a literature review. Canberra: Brown & Wilton Integrated Publishing Services; 2000.
278. National Resource Centre for Consumer Participation in Health. Methods and models of consumer participation. Canberra: 2004.
279. NHS Centre for Reviews and Dissemination. Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews. CRD Report 4 (2nd edition) [Online]. 2001 [Zugriff am 07.05.2007]. URL: <http://www.york.ac.uk/inst/crd/report4.htm>.
280. NHS Centre for Reviews and Dissemination. Effectiveness matters. The database of abstracts of reviews of effects (DARE). 2002; 6(2): 1-4.
281. Nielsen J. *Designing web usability*. Indianapolis: New Riders Publishing; 2000.
282. Nilsen ES, Myrhaug HT, Johansen M, Oliver S, Oxman AD. Methods of consumer involvement in developing healthcare policy and research, clinical practice guidelines and patient information material. *Cochrane Database Syst Rev* 2006; 3: CD004563.
283. Nutbeam D. Health promotion glossary. *Health Promot Int* 1998; 13(4): 349-364.
284. Oliver S, Clarke-Jones L, Rees R, Milne R, Buchanan P, Gabbay J et al. Involving consumers in research and development agenda setting for the NHS: developing an evidence-based approach. *Health Technol Assess* 2004; 8(15): 1-IV.
285. Oxman AD, Guyatt GH. Guidelines for reading literature reviews. *CMAJ* 1988; 138(8): 697-703.
286. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol* 1991; 44(11): 1271-1278.
287. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med* 1992; 116(1): 78-84.
288. Oxman AD, Guyatt GH, Singer J, Goldsmith CH, Hutchison BG, Milner RA et al. Agreement among reviewers of review articles. *J Clin Epidemiol* 1991; 44(1): 91-98.

289. Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Stat Med* 1998; 17(24): 2815-2834.
290. Pennekamp PH, Diedrich O, Schmitt O, Kraft CN. Prävalenz und Stellenwert der Internetnutzung orthopädischer Patienten. *Z Orthop* 2006; 144(5): 459-463.
291. Perleth M, Jakubowski E, Busse R. What is 'best practice' in health care? State of the art and perspectives in improving the effectiveness and efficiency of the European health care systems. *Health Policy* 2001; 56(3): 235-250.
292. Peters JL, Sutton A, Jones D, Abrams KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. *JAMA* 2006; 295(6): 676-680.
293. Petticrew M, Song F, Wilson P, Wright K, NHS CRD Review DaIT. Quality-assessed reviews of health care interventions and the database of abstracts of reviews of effectiveness (DARE). *Int J Technol Assess Health Care* 1999; 15(4): 671-678.
294. 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(2): 109-125.
295. Pocock SJ. *Clinical trials: a practical approach*. Chichester: Wiley; 1983.
296. Pocock SJ, Collier TJ, Dandreo KJ, De Stavola BL, Goldman MB, Kalish LA et al. Issues in the reporting of epidemiological studies: a survey of recent practice. *BMJ* 2004; 329(7471): 883-887.
297. Poensgen A, Larsson S. *Patients, physicians and the internet: myth, reality, and implications*. Boston: Boston Consulting Group; 2001.
298. Poynard T, Munteanu M, Ratzu V, Benhamou Y, Di Martino V, Taieb J et al. Truth survival in clinical research: an evidence-based requiem? *Ann Intern Med* 2002; 136(12): 888-895.
299. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med* 1989; 8(4): 431-440.
300. Rennie D. Editorial peer review: its development and rationale. In: Godlee F, Jefferson T (Ed). *Peer review in health sciences*. London: BMJ Publishing Group; 2003: 1-13.
301. Richter JG, Becker A, Specker C, Monser R, Schneider M. Krankheitsbezogene Internetnutzung bei Patienten mit entzündlichen-rheumatischen Systemerkrankungen. *Z Rheumatol* 2004; 63(3): 216-222.
302. Roebruck P, Elze M, Hauschke D, Leverkus F, Kieser M. Literaturübersicht zur Fallzahlplanung für Äquivalenzprobleme. *Inform Biom Epidemiol Med Biol* 1997; 28(2): 51-63.
303. Röhm J, Hauschke D, Koch A, Pigeot I. Biometrische Verfahren zum Wirksamkeitsnachweis im Zulassungsverfahren: Nicht-Unterlegenheit in klinischen Studien. *Bundesgesundheitsbl Gesundheitsforsch Gesundheitschutz* 2005; 48: 562-571.
304. Rossouw JE. Estrogens for prevention of coronary heart disease. Putting the brakes on the bandwagon. *Circulation* 1996; 94(11): 2982-2985.
305. Rothwell PM. Treating individuals 2: subgroup analysis in randomised controlled trials. Importance, indications, and interpretation. *Lancet* 2005; 365(9454): 176-186.
306. Royal Society. *Science and the public interest: communicating the results of new scientific research to the public*. London: Royal Society; 2006.
307. Royall RM. The effect of sample size on the meaning of significance tests. *Am Stat* 1986; 40(4): 313-315.
308. Royle P, Waugh N. Literature searching for clinical and cost-effectiveness studies used in health technology assessment reports carried out for the National Institute for Clinical Excellence appraisal system. *Health Technol Assess* 2003; 7(34): iii, ix-51.
309. Royston P. A strategy for modelling the effect of a continuous covariate in medicine and epidemiology. *Stat Med* 2000; 19(14): 1831-1847.

310. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Appl Stat* 1994; 43(3): 429-467.
311. Rutten LJ, Arora NK, Bakos AD, Aziz N, Rowland J. Information needs and sources of information among cancer patients: a systematic review of research (1980-2003). *Patient Educ Couns* 2005; 57(3): 250-261.
312. Sackett DL. Bias in analytic research. *J Chronic Dis* 1979; 32(1-2): 51-63.
313. Sackett DL. Turning a blind eye: why we don't test for blindness at the end of our trials. *BMJ* 2004; 328(7448): 1136.
314. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine: how to practice and teach EBM. London: Churchill-Livingstone; 2000.
315. Sampson M, Barrowman NJ, Moher D, Klassen TP, Pham B, Platt R et al. Should meta-analysts search Embase in addition to Medline? *J Clin Epidemiol* 2003; 56(10): 943-955.
316. Sampson M, McGowan J. Errors in search strategies were identified by type and frequency. *J Clin Epidemiol* 2006; 59(10): 1057-1063.
317. Sanger S, Lang B, Klemperer D, Thomeczek C, Dierks ML. Manual Patienteninformation - Empfehlungen zur Erstellung evidenzbasierter Patienteninformationen. Berlin: ZQ; 2006.
318. Santo A, Laizner AM, Shohet L. Exploring the value of audiotapes for health literacy: a systematic review. *Patient Educ Couns* 2005; 58(3): 235-243.
319. Sargent DJ, Conley BA, Allegra C, Collette L. Clinical trial designs for predictive marker validation in cancer treatment trials. *J Clin Oncol* 2005; 23(9): 2020-2027.
320. Sauerbrei W, Royston P. Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials. *J R Stat Soc [Ser A]* 1999; 162(1): 71-94.
321. Savoie I, Kazanjian A, Bassett K. Do clinical practice guidelines reflect research evidence? *J Health Serv Res Policy* 2000; 5(2): 76-82.
322. Sawicki PT. Qualitat der Gesundheitsversorgung in Deutschland: ein randomisierter Sechs-Lander-Vergleich aus Patientensicht. *Med Klin* 2005; 100(11): 755-768.
323. Schaeffer D, Dierks ML, Hurrelmann K, Keller A, Krause H, Schmidt-Kaehler S et al. Evaluation der Modellprojekte zur Patienten- und Verbraucherberatung nach § 65b Sozialgesetzbuch V - Abschlussbericht der wissenschaftlichen Begleitforschung fur die Spitzenverbande der GKV. Bielefeld: 2004.
324. Schluter PJ, Ware RS. Single patient (n-of-1) trials with binary treatment preference. *Stat Med* 2005; 24(17): 2625-2636.
325. Schmidt-Kaehler S. Patienteninformation Online. Theoretische Grundlagen, Planung und Entwicklung eines Konzeptes fur die Patientenschulung im Internet. Bern: Hans Huber; 2004.
326. Schneider N, Dierks ML, Seidel G, Schwartz FW. The federal government commissioner for patient issues in Germany: initial analysis of the user inquiries. *BMC Health Serv Res* 2007; 7: 24.
327. Schunemann HJ, Best D, Vist GE, Oxman AD. Letters, numbers, symbols and words: how to communicate grades of evidence and recommendations. *CMAJ* 2003; 169(7): 677-680.
328. Schunemann HJ, Jaeschke R, Cook DJ, Bria WF, El-Solh AA, Ernst A et al. An official ATS statement: Grading the quality of evidence and strength of recommendations in ATS guidelines and recommendations. *Am J Respir Crit Care Med* 2006; 174(5): 605-614.
329. Scott JT, Harmsen M, Prictor MJ, Entwistle VA, Sowden AJ, Watt I. Recordings or summaries of consultations for people with cancer [Cochrane Review]. *Cochrane Database Syst Rev* 2003; 2: CD001539.
330. Senn S. The many modes of meta. *Drug Inf J* 2000; 34: 535-549.
331. Senn S. Trying to be precise about vagueness. *Stat Med* 2007; 26(7): 1417-1430.

332. Senn SJ. Inherent difficulties with active control equivalence studies. *Stat Med* 1993; 12(24): 2367-2375.
333. Senn SJ. Turning a blind eye: authors have blinkered view of blinding. *BMJ* 2004; 328(7448): 1135-1136.
334. Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA* 1999; 281(20): 1900-1905.
335. Shekelle PG, Ortiz E, Rhodes S, Morton SC, Eccles MP, Grimshaw JM et al. Validity of the agency for healthcare research and quality clinical practice guidelines: how quickly do guidelines become outdated? *JAMA* 2001; 286(12): 1461-1467.
336. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. *Ann Intern Med* 2007; 147(4): 224-233.
337. Sillence E, Briggs P, Harris PR, Fishwick L. How do patients evaluate and make use of online health information? *Soc Sci Med* 2007; 64(9): 1853-1862.
338. Simmonds MC, Higgins JP. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. *Stat Med* 2007; 26(15): 2982-2999.
339. Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. *Br J Cancer* 1994; 69(6): 979-985.
340. Song F, Eastwood AJ, Gilbody S, Duley L, Sutton A. Publication and related biases. *Health Technol Assess* 2000; 4(10): 1-115.
341. Spek V, Cuijpers P, Nyklicek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. *Psychol Med* 2007; 37(3): 319-328.
342. Spiegelhalter DJ, Freedman LS, Parmar MK. Bayesian approaches to randomized trials. *J R Stat Soc [Ser A]* 1994; 157(3): 357-416.
343. Spiegelhalter DJ, Myles JP, Jones D, Abrams KR. Methods in health service research: an introduction to bayesian methods in health technology assessment. *BMJ* 1999; 319(7208): 508-512.
344. Statistisches Bundesamt. Leben in Deutschland: Haushalte, Familien und Gesundheit - Ergebnisse des Mikrozensus 2005. Wiesbaden: Statistisches Bundesamt; 2005.
345. Steckelberg A, Berger B, Kopke S, Heesen C, Muhlhauser I. Kriterien für evidenzbasierte Patienteninformationen. *Z Ärztl Fortbild Qual Gesundh wes* 2005; 99(6): 343-351.
346. Steckelberg A, Kasper J, Redegeld M, Muhlhauser I. Risk information - barrier to informed choice? a focus group study. *Soz Präventivmed* 2004; 49(6): 375-380.
347. Steiner JF. The use of stories in clinical research and health policy. *JAMA* 2005; 294(22): 2901-2904.
348. Sterne J, Egger M, Smith GD. Systematic reviews in health care: investigating and dealing with publication and other biases in meta-analysis. *BMJ* 2001; 323(7304): 101-105.
349. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA* 2000; 283(15): 2008-2012.
350. Sullivan LM, Massaro JM, D'Agostino RB, Sr. Presentation of multivariate data for clinical use: the Framingham Study risk score functions. *Stat Med* 2004; 23(10): 1631-1660.
351. Swift TL, Dieppe PA. Using expert patients' narratives as an educational resource. *Patient Educ Couns* 2005; 57(1): 115-121.
352. Thomas J, Harden A, Oakley A, Oliver S, Sutcliffe K, Rees R et al. Integrating qualitative research with trials in systematic reviews. *BMJ* 2004; 328(7446): 1010-1012.
353. Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002; 21(11): 1559-1573.

354. Thompson WG. The placebo effect and health: combining science and compassionate care. Amherst: Prometheus Books; 2005.
355. Thurow S. Search engine visibility. Indianapolis: New Riders; 2003.
356. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc [Ser B]* 1996; 58(1): 267-288.
357. Torgerson DJ. Contamination in trials: is cluster randomisation the answer? *BMJ* 2001; 322(7282): 355-357.
358. Trevena LJ, Davey HM, Barratt A, Butow P, Caldwell P. A systematic review on communicating with patients about evidence. *J Eval Clin Pract* 2006; 12(1): 13-23.
359. Tsay MY, Yang YH. Bibliometric analysis of the literature of randomized controlled trials. *J Med Libr Assoc* 2005; 93(4): 450-458.
360. Tukey JW. Some thoughts on clinical trials, especially problems of multiplicity. *Science* 1977; 198(4318): 679-684.
361. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA* 2003; 290(12): 1624-1632.
362. UK National Screening Committee. Criteria for appraising the viability, effectiveness and appropriateness of a screening programme. Edinburgh: NSC; 2003.
363. US Agency for International Development. Conducting focus group interviews. Washington: U.S. AID; 1996.
364. US Agency for International Development. Conducting key informant interviews. Washington: U.S. AID; 1996.
365. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Center for Devices and Radiological Health (CDRH). Patient-reported outcome measures: use in medical product development to support labeling claims. Draft guidance. Rockville: CDER; 2006.
366. Van den Brink-Muinen A, Verhaak PF, Bensing JM, Bahrs O, Deveugele M, Gask L et al. Doctor-patient communication in different European health care systems: relevance and performance from the patients' perspective. *Patient Educ Couns* 2000; 39(1): 115-127.
367. Van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med* 2002; 21(4): 589-624.
368. van Tinteren H, Hoekstra OS, Smit EF, van den Bergh JH, Schreurs AJ, Stallaert RA et al. Effectiveness of positron emission tomography in the preoperative assessment of patients with suspected non-small-cell lung cancer: the PLUS multicentre randomised trial. *Lancet* 2002; 359(9315): 1388-1393.
369. Vandenbroucke JP, Von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ et al. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. *Ann Intern Med* 2007; 147(8): W163-W194.
370. Vidanapathirana J, Abramson MJ, Forbes A, Fairley C. Mass media interventions for promoting HIV testing [Cochrane Review]. *Cochrane Database Syst Rev* 2005; 3: CD004775.
371. Villar J, Mackey ME, Carroli G, Donner A. Meta-analyses in systematic reviews of randomized controlled trials in perinatal medicine: comparison of fixed and random effects models. *Stat Med* 2001; 20(23): 3635-3647.
372. Virtanen H, Leino-Kilpi H, Salanterä S. Empowering discourse in patient education. *Patient Educ Couns* 2007; 66(2): 140-146.
373. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007; 147(8): 573-577.

374. W3C Web Accessibility Initiative. Web Accessibility Initiative (WAI) [Online]. 2006 [Zugriff am 21.09.2006]. URL: <http://www.w3.org/WAI/>.
375. Wallis EJ, Ramsay LE, Ul Haq I, Ghahramani P, Jackson PR, Rowland-Yeo K et al. Coronary and cardiovascular risk estimation for primary prevention: validation of a new sheffield table in the 1995 Scottish health survey population. *BMJ* 2000; 320(7236): 671-676.
376. Walter U, Schwartz FW. Prävention. In: Schwartz FW (Ed). *Das Public Health Buch: Gesundheit und Gesundheitswesen*. München: Urban und Fischer; 2003: 189-214.
377. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of web-based vs. non-web-based interventions: a meta-analysis of behavioral change outcomes. *J Med Internet Res* 2004; 6(4): e40.
378. Weir CJ, Walley RJ. Statistical evaluation of biomarkers as surrogate endpoints: a literature review. *Stat Med* 2006; 25(2): 183-203.
379. Whitehead J. *The design and analysis of sequential clinical trials*. Chichester: Ellis Horwood; 1983.
380. 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.
381. Windeler J. Bedeutung randomisierter klinischer Studien mit relevanten Endpunkten für die Nutzenbewertung. In: Gesundheitsforschungsrat (GFR) des Bundesministeriums für Bildung und Forschung (BMBF) (Ed). *Begriffsdefinitionen und Einführung - Dokumentation des ersten gemeinsamen Workshops von GFR und IQWiG am 4. September 2007 in Berlin*. Bonn: Gesundheitsforschungsrat (GFR) des Bundesministeriums für Bildung und Forschung (BMBF); 2007: 26-31.
382. Windeler J, Conradt C. Wie können "Signifikanz" und "Relevanz" verbunden werden? *Med Klin* 1999; 94(11): 648-651.
383. Windeler J, Lange S. Nutzenbewertung in besonderen Situationen – Seltene Erkrankungen. *Z Evid Fortb Qual Gesundh wes* 2008; 102(1): 25-30.
384. Windeler J, Ziegler S. EBM-Splitter: Evidenzklassifizierungen. *Z Ärztl Fortbild Qualitätssich* 2003; 97: 513-514.
385. Wofford JL, Smith ED, Miller DP. The multimedia computer for office-based patient education: a systematic review. *Patient Educ Couns* 2005; 59(2): 148-157.
386. Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008; 336(7644): 601-605.
387. Worrall G, Chaulk P, Freake D. The effects of clinical practice guidelines on patient outcomes in primary care: a systematic review. *CMAJ* 1997; 156(12): 1705-1712.
388. Yank V, Rennie D, Bero LA. Financial ties and concordance between results and conclusions in meta-analyses: retrospective cohort study. *BMJ* 2007; 335(7631): 1202-1205.
389. Ziebland S, McPherson A. Making sense of qualitative data analysis: an introduction with illustrations from DIPEX (personal experiences of health and illness). *Med Educ* 2006; 40(5): 405-414.
390. 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(5): 706-713.