

Allgemeine Methoden

Version 4.1 vom 28.11.2013

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

ISBN 978-3-9815265-0-9

© 2013 Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, Köln
www.iqwig.de

Anschrift des Herausgebers:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
Im Mediapark 8 (KölnTurm)
50670 Köln

Tel.: +49 (0)221 – 35685-0
Fax: +49 (0)221 – 35685-1
E-Mail: methoden@iqwig.de
Internet: www.iqwig.de

Präambel

Das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) ist eine Einrichtung der Stiftung für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Das IQWiG ist ein fachlich unabhängiges, wissenschaftliches Institut. Informationen zu Struktur und Organisation von Stiftung und Institut finden sich auf der Website www.iqwig.de.

Die „Allgemeinen Methoden“ erläutern die gesetzlichen und wissenschaftlichen Grundlagen des Instituts. Die Aufgaben 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 „Allgemeine Methoden zur Bewertung von Verhältnissen zwischen Nutzen und Kosten“ 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 Fachwissenschaftlerinnen und Fachwissenschaftler. Um aber möglichst allen Interessierten einen Zugang zu Informationen über die Arbeitsweise des Instituts zu verschaffen, haben sich die Autorinnen und Autoren um Verständlichkeit bemüht. Wie jeder fachwissenschaftliche Text setzen allerdings auch die „Allgemeinen Methoden“ ein bestimmtes Maß an Vorwissen voraus.

Die „Allgemeinen Methoden“ sollen die Vorgehensweise des Instituts 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.

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. 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.

Was ist neu?

In der Version 4.1 wurden im Vergleich zur Version 4.0 der „Allgemeinen Methoden“ des Instituts vom 23.09.2011 kleinere Fehler beseitigt und wenige redaktionelle Änderungen durchgeführt. Es haben folgende inhaltliche Änderungen stattgefunden:

- Darstellung des externen Reviews für Vorberichte als optionalen Schritt in den Abschnitten 2.1.1 und 2.2.3
- Teilung des bisherigen Abschnitts 3.1.4 in die neuen Abschnitte 3.1.4 und 3.1.5 und Konkretisierung der Anforderungen an die Beleglage zur Formulierung von Nutzenaussagen mit unterschiedlichen Aussagesicherheiten
- Operationalisierung der Feststellung des Ausmaßes des Zusatznutzens sowie dessen Rationale in Abschnitt 3.3.3 und dem neuen Anhang
- Verwendung von Prädiktionsintervallen für Meta-Analysen mit zufälligen Effekten in Abschnitt 7.3.8

Inhaltsverzeichnis

Präambel	ii
Was ist neu?	iii
Inhaltsverzeichnis.....	iv
Tabellenverzeichnis.....	viii
Abbildungsverzeichnis.....	ix
1 Das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen.....	1
1.1 Gesetzliche Aufgaben.....	1
1.2 Evidenzbasierte Medizin	3
1.2.1 Praktische evidenzbasierte Medizin	4
1.2.2 Bedeutung von EbM für das Institut	5
1.2.3 Evidenzbasierte Entscheidungsfindung im Gesundheitswesen.....	5
1.2.4 Die Strategien der EbM.....	5
1.2.5 Die Bedeutung der Ergebnissicherheit.....	6
1.2.6 Die Verbindung von Ergebnissicherheit und Alltagsnähe	8
1.2.7 Nutzen im Einzelfall.....	10
2 Produkte des Instituts	11
2.1 Produktspezifische Verfahrensabläufe	11
2.1.1 Bericht	13
2.1.2 Rapid Report.....	17
2.1.3 Dossierbewertung	18
2.1.4 Addendum	20
2.1.5 Gesundheitsinformationen.....	21
2.1.6 Arbeitspapier	24
2.2 Generelle Aspekte bei der Produkterstellung	25
2.2.1 Auswahl externer Sachverständiger	25
2.2.2 Gewährleistung der fachlichen Unabhängigkeit	26
2.2.3 Review der Produkte des Instituts	28
2.2.4 Veröffentlichung der Produkte des Instituts.....	28
3 Nutzenbewertung medizinischer Interventionen.....	30
3.1 Patientenrelevanter medizinischer Nutzen und Schaden.....	30
3.1.1 Definition des patientenrelevanten medizinischen Nutzens bzw. Schadens	30
3.1.2 Surrogate patientenrelevanter Endpunkte.....	32
3.1.3 Ermittlung des Schadens medizinischer Interventionen.....	34
3.1.4 Endpunktbezogene Bewertung	36

3.1.5	Zusammenfassende Bewertung	41
3.2	Spezielle Aspekte der Nutzenbewertung	42
3.2.1	Auswirkung nicht publizierter Studienergebnisse auf das Fazit	42
3.2.2	Dramatischer Effekt	44
3.2.3	Studiendauer	45
3.2.4	Patientenberichtete Endpunkte	45
3.2.5	Nutzen und Schaden in kleinen Populationen	46
3.3	Nutzenbewertung von Arzneimitteln	47
3.3.1	Stellenwert des Zulassungsstatus	47
3.3.2	Studien zur Nutzenbewertung von Arzneimitteln	49
3.3.3	Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V	50
3.4	Nichtmedikamentöse therapeutische Interventionen	54
3.5	Diagnostische Verfahren	56
3.6	Früherkennung und Screening	58
3.7	Prävention	59
3.8	Prognosestudien	60
4	Leitlinien und Versorgungsanalyse	63
4.1	Hintergrund	63
4.2	Identifikation von Versorgungsstandards über Leitlinien	63
4.2.1	Versorgungsstandards in Leitlinien	63
4.2.2	Methodische Bewertung von Leitlinien	64
4.2.3	Strukturierte Aufbereitung und Auswertung von Empfehlungen	65
4.3	Empfehlungen zu Disease-Management-Programmen	67
4.4	Versorgungsanalyse	67
4.4.1	Hintergrund	67
4.4.2	Inhaltliche Aspekte einer Versorgungsanalyse	68
4.4.3	Ziele einer Versorgungsanalyse	69
4.4.4	Fragestellung der Versorgungsanalyse	69
4.4.5	Mögliche Versorgungsparameter	70
4.4.6	Ablauf einer Versorgungsanalyse	71
4.4.7	Ebenen der Versorgungsanalyse	73
4.4.8	Methodische Besonderheiten der Versorgungsanalyse	73
4.4.9	Informationsbeschaffung	74
4.4.10	Bewertung der identifizierten Daten	76
4.4.11	Informationssynthese und -analyse	76
4.4.12	Darstellung und Bewertung der Versorgungsqualität	77

4.5	Validität von Leitlinienempfehlungen.....	77
4.5.1	Hintergrund	77
4.5.2	Validität von Leitlinienempfehlungen.....	78
4.5.3	Ziel der Analyse und Bewertung von Leitlinienempfehlungen	78
4.5.4	Vorgehen bei der Analyse und Bewertung der internen Validität	79
4.5.5	Mögliche Fragestellungen	79
5	Evidenzbasierte Gesundheitsinformation für Bürgerinnen und Bürger	80
5.1	Hintergrund und Ziele.....	80
5.2	Patientenzentrierte Kommunikation.....	81
5.2.1	Kommunikationsstandards	81
5.2.2	Methode der Patientenwege	84
5.2.3	Beteiligung von Bürgerinnen und Bürgern	87
5.2.4	Visuelle Kommunikation und Multimedia.....	87
5.2.5	Informationszugang und Barrierefreiheit	88
5.3	Themenwahl, Recherche und wissenschaftliche Bewertung.....	90
5.3.1	Themenwahl	90
5.3.2	Recherche	93
5.3.3	Evidenzbewertung	94
5.3.4	Aktualisierung	95
5.4	Informationsprodukte	99
5.4.1	Informationsberichte, Merkblätter und Kurzantworten.....	99
5.4.2	Zusätzliche Elemente	102
5.4.3	Pressemitteilungen.....	103
5.4.4	Erfahrungsberichte	103
5.4.5	Website.....	104
5.5	Überwachung und Auswertung	106
5.5.1	Routinemonitoring.....	106
5.5.2	Feedback, Korrekturen und Verbesserungen	107
5.5.3	Evaluation.....	108
6	Informationsbeschaffung	110
6.1	Eigene Informationsbeschaffung durch das Institut	110
6.1.1	Ablauf einer Recherche	111
6.1.2	Bibliografische Datenbanken	111
6.1.3	Suche in Studienregistern	112
6.1.4	Leitliniendatenbanken und -anbieter	112
6.1.5	Herstellieranfragen	113
6.1.6	Sonstige Datenquellen für die Recherche.....	114

6.1.7	Auswahl relevanter Publikationen.....	114
6.1.8	Dokumentation der Recherche	115
6.1.9	Ergänzungsrecherchen bei systematischen Übersichten	116
6.2	Überprüfung der Informationsbeschaffung für Dossiers.....	116
7	Informationsbewertung.....	117
7.1	Qualitätsbewertung von Einzelstudien	117
7.1.1	Kriterien zum Einschluss von Studien	117
7.1.2	Zusammenhang zwischen Studientyp/-art und Fragestellung.....	117
7.1.3	Rangordnung verschiedener Studienarten / Evidenzgrade.....	118
7.1.4	Aspekte der Bewertung des Verzerrungspotenzials	119
7.1.5	Interpretation von kombinierten Endpunkten.....	122
7.1.6	Interpretation von Subgruppenanalysen	123
7.1.7	Bewertung der Konsistenz von Daten	125
7.2	Berücksichtigung systematischer Übersichten.....	126
7.2.1	Einordnung systematischer Übersichten	126
7.2.2	Nutzenbewertung auf Basis systematischer Übersichten.....	127
7.2.3	Berücksichtigung publizierter Meta-Analysen.....	129
7.3	Spezielle biometrische Aspekte	130
7.3.1	Darstellung von Effekten und Risiken	130
7.3.2	Beurteilung statistischer Signifikanz.....	131
7.3.3	Beurteilung klinischer Relevanz.....	132
7.3.4	Bewertung subjektiver Endpunkte bei offenen Studiendesigns	134
7.3.5	Nachweis der Verschiedenheit	135
7.3.6	Nachweis der Gleichheit	135
7.3.7	Adjustierung und multifaktorielle Verfahren	136
7.3.8	Meta-Analysen	138
7.3.9	Indirekte Vergleiche	143
7.3.10	Umgang mit nicht oder nicht vollständig publizierten Daten	144
7.3.11	Darstellung von Verzerrungsarten.....	144
7.4	Qualitative Methoden	147
7.4.1	Qualitative Studien	147
7.4.2	Konsultationstechniken	149
Anhang A	- Rationale der Methodik zur Feststellung des Ausmaßes des Zusatznutzens.....	150
8	Literaturverzeichnis	162

Tabellenverzeichnis

	Seite
Tabelle 1 Übersicht über die Produkte des Instituts.....	12
Tabelle 2: Regelhaft abgeleitete Aussagesicherheiten für verschiedene Evidenzsituationen beim Vorliegen von Studien derselben qualitativen Ergebnissicherheit.....	40
Tabelle 3: Szenarien der Datenübermittlung und Konsequenzen für das Fazit einer Nutzenbewertung	43
Tabelle 4: Schwellenwerte zur Feststellung des Ausmaßes eines Effekts	53
Tabelle 5: Beispiele möglicher Versorgungsparameter	70
Tabelle 6: Informationsquellen zur Identifizierung von deutschen Versorgungsstandards	74
Tabelle 7: Informationsquellen zur Identifizierung von Versorgungsdaten	75
Tabelle 8: Verschiedene Dimensionen eines Patientenweges.....	86
Tabelle 9: Aspekt zur Priorisierung der Themenauswahl	93
Tabelle 10: Kategorisierung der Aktualisierungen	99
Tabelle 11: Feststellung des Ausmaßes des Zusatznutzens – Kriterien gemäß AM-NutzenV	151
Tabelle 12: Feststellung des Ausmaßes des Zusatznutzens – Kriterien gemäß AM-NutzenV mit Ergänzungen*	153
Tabelle 13: Feststellung des Ausmaßes des Zusatznutzens – hierarchisierte Kriterien gemäß AM-NutzenV mit Ergänzungen*.....	154
Tabelle 14: Inferenzstatistische Schwellenwerte (Hypothesengrenzen) für relative Effektmaße	158
Tabelle 15: Tatsächliche Effekte für das relative Risiko	161

Abbildungsverzeichnis

	Seite
Abbildung 1: Ablauf der Berichtserstellung	14
Abbildung 2: Ablauf der Erstellung eines Rapid Reports	17
Abbildung 3: Ablauf der Erstellung einer Dossierbewertung	19
Abbildung 4: Ablauf der Erstellung eines Addendums	20
Abbildung 5: Ablauf der Gesundheitsinformationserstellung	22
Abbildung 6: Ablauf der Erstellung eines Arbeitspapiers	24
Abbildung 7: Exemplarischer Ablauf einer Versorgungsanalyse	72
Abbildung 8: Tatsächliche Effekte in Abhängigkeit des Basisrisikos	160

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

Das Institut wurde im Zuge der Gesundheitsreform 2004 [119] als Einrichtung der Stiftung für Qualität und Wirtschaftlichkeit im Gesundheitswesen gegründet. Die gesetzlichen Grundlagen und Aufgaben des Instituts sind im Sozialgesetzbuch V (SGB V) [2] verankert und wurden im Zuge weiterer Gesundheitsreformen mehrfach angepasst und erweitert. Informationen zu Struktur und Organisation des Instituts sind auf der Website www.iqwig.de verfügbar.

Das Institut wird zu Fragen von grundsätzlicher Bedeutung für die Qualität und Wirtschaftlichkeit der im Rahmen der gesetzlichen Krankenversicherung erbrachten Leistungen tätig. Die spezifischen Aufgaben sind in § 139a 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 unter Berücksichtigung alters-, geschlechts- und lebenslagenspezifischer Besonderheiten,
- Bewertungen evidenzbasierter Leitlinien für die epidemiologisch wichtigsten Krankheiten,
- Abgabe von Empfehlungen zu Disease-Management-Programmen,
- 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 Beauftragung und Wahrnehmung der Aufgaben werden in § 139b SGB V näher geregelt. Danach können nur der Gemeinsame Bundesausschuss (G-BA) und das Bundesministerium für Gesundheit (BMG) das Institut beauftragen. Das Institut kann einen Antrag des BMG als unbegründet ablehnen, es sei denn, das Ministerium übernimmt die Finanzierung der Bearbeitung.

Das Institut hat darauf zu achten, dass externe Sachverständige in die Auftragsbearbeitung einbezogen werden. Diese haben zur Sicherstellung der fachlichen Unabhängigkeit des Instituts alle Beziehungen zu Interessenverbänden, Auftragsinstituten, insbesondere der pharmazeutischen Industrie und der Medizinprodukteindustrie, einschließlich der Art und Höhe von Zuwendungen offenzulegen (siehe Abschnitt 2.2.2).

Die Arbeitsergebnisse zu Aufträgen des G-BA leitet das Institut dem G-BA als Empfehlungen zu. Der G-BA hat laut Gesetz die Empfehlungen im Rahmen seiner Entscheidungen zu berücksichtigen.

Das Institut wird zum weitaus überwiegenden Teil aus den Beiträgen der Mitglieder der gesetzlichen Krankenversicherungen finanziert. Dazu werden nach § 139c SGB V vom G-BA Zuschläge festgelegt. Diese Zuschläge werden von allen deutschen Arztpraxen und Krankenhäusern abgeführt, die gesetzlich versicherte Patienten behandeln.

Im Rahmen des Arzneimittelmarktneuordnungsgesetzes (AMNOG) wurden Anfang 2011 die Aufgaben des Instituts um die Bewertung des Nutzens von Arzneimitteln mit neuen Wirkstoffen kurz nach der Markteinführung erweitert [120]. Dazu müssen Hersteller Dossiers einreichen, die die Ergebnisse aus Studien zusammenfassen. Zuständig für die frühe Nutzenbewertung ist der G-BA; er hat jedoch die Möglichkeit, das Institut oder Dritte mit der Prüfung und Bewertung der Dossiers zu beauftragen.

Grundlage sind die neu gefassten Regelungen des § 35a SGB V. Sie werden ergänzt durch eine ebenfalls seit Anfang 2011 geltende Rechtsverordnung des Bundesgesundheitsministeriums [70] und die Verfahrensordnung des G-BA [198].

Verbunden mit einer Nutzenbewertung kann der G-BA das Institut auch mit einer Kosten-Nutzen-Bewertung beauftragen. Dabei werden Nutzen und Kosten einer medizinischen Technologie gegenübergestellt mit dem Ziel, einen angemessenen Preis zu finden. Die Rahmenbedingungen dieser Kosten-Nutzen-Bewertungen geben § 35b SGB V und § 139a SGB V vor. So hat der G-BA in seinem Auftrag zum Beispiel festzulegen, welche Vergleichstherapien berücksichtigt werden sollen und für welche Patientengruppen die Bewertung erfolgen soll.

Die Kosten-Nutzen-Bewertung selbst basiert auf einem Vergleich mit anderen Arzneimitteln oder nichtmedikamentösen Behandlungsformen. Als Kriterien für den Patientennutzen nennt das Gesetz insbesondere die Verbesserung des Gesundheitszustandes, eine Verkürzung der Krankheitsdauer, eine Verlängerung der Lebensdauer, eine Verringerung der Nebenwirkungen sowie eine Verbesserung der Lebensqualität. Aus diesen Vorgaben des Gesetzes leitet sich die für das Institut geltende Definition des patientenrelevanten Nutzens ab (siehe Abschnitt 3.1).

Das Institut bestimmt auftragsbezogen die Methoden und Kriterien für die Erarbeitung von Bewertungen auf der Grundlage der in den jeweiligen Fachkreisen anerkannten inter-

nationalen Standards der evidenzbasierten Medizin und der Gesundheitsökonomie. In Abschnitt 1.2 werden der Begriff der evidenzbasierten Medizin, seine Entwicklung und das dahinterliegende Konzept näher erläutert.

Das Institut gewährleistet während der Bearbeitung seiner Berichte eine hohe Verfahrenstransparenz und eine angemessene Beteiligung. Das Gesetz verpflichtet das Institut dazu, in allen wichtigen Abschnitten der Berichtserstellung Fachexpertinnen und Fachexperten, Herstellern und 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“. Das Institut geht über diese Verpflichtung hinaus, indem es allen interessierten Personen und Institutionen die Möglichkeit zur Stellungnahme zu seinen Berichten einräumt.

Diese Stellungnahmen bezieht das Institut in seine Bewertungen ein. Die Umsetzung dieser Regelungen ist in Abschnitt 2.1.1 im Zusammenhang mit der Erstellung von Berichtsplänen und Vorberichten beschrieben.

Außerdem dokumentiert das Institut seine Arbeitsergebnisse und ergänzende Information auf seiner frei zugänglichen Website. Interessierte können darüber hinaus den E-Mail-Dienst (Infodienst) des Instituts abonnieren. Er bietet Abonnenten die Möglichkeit, selbst festzulegen, welche Informationen des Instituts sie erhalten möchten.

1.2 Evidenzbasierte Medizin

Der Begriff „evidenzbasierte Medizin“ (EbM) ist eine etwas unscharfe Eindeutigung des englischen Begriffs „Evidence-based Medicine“. Gemeint ist damit eine medizinische Betreuung von Patientinnen und 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 [17,164,218,222].

Instrumente, die einer subjektiv geprägten und damit häufig verzerrten Bewertung vorbeugen sollen (siehe u. a. Kapitel 7), 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 [358]. 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 [251]. 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 gemacht [31]. Etwa 20 Jahre später bemühten sich klinische Epidemiologen, die Methodenlehre auch in der klinischen Praxis zu etablieren [170]. Begleitet von teilweise heftig geführten Kontroversen gelang dies tatsächlich erst in den 1990er-Jahren, zeitgleich mit der Benennung des Konzepts als „evidenzbasierte Medizin“. Klinische Studien 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 „Bewertung des medizinischen Nutzens nach den international anerkannten Standards der evidenzbasierten Medizin“ zu gewährleisten (§ 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 Anwendung 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 Ärztinnen und Ärzte, die Unterstützung bei einer Therapieentscheidung suchen.

1.2.1 Praktische evidenzbasierte Medizin

Evidenzbasierte Medizin ist von der Idee her als Strategie für Ärztinnen und Ärzte gedacht, die für ihre Patientinnen und Patienten unter möglichen Interventionen die vielversprechendsten und deren Bedürfnissen am ehesten entsprechenden Alternativen herausfinden und die Erfolgsaussichten neutral darstellen wollen. Diese Anwendung der evidenzbasierten Medizin in der täglichen Praxis für „individuelle Patienten“ haben im Jahr 1996 David Sackett und Kollegen [445] 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 hat Instrumente entwickelt, die Unsicherheit einzuschätzen, oft werden sogenannte Evidenzlevel zur Illustration verwendet. Auf diese Weise hilft EbM Ärztinnen und Ärzten sowie Patientinnen und Patienten dabei, die Art und das Ausmaß der Unsicherheit zu

erkennen. Ärztinnen und Ärzte sowie Patientinnen und Patienten können dann besprechen, wie man mit der Unsicherheit umgehen soll. Gerade in unsicheren Situationen kommt es auf die persönlichen Präferenzen an, die darüber entscheiden, welche Option Patientinnen und Patienten wählen. Im Idealfall basieren Entscheidungen außer auf der Evidenz auch auf dem klinischen Zustand und den Umständen, in denen sich eine Person befindet, und auf ihren Präferenzen und Handlungen [239]. Gleichzeitig schafft die Beschreibung der identifizierten Wissenslücken erst die Voraussetzung für zielgerichtete medizinische Forschung, die sich nach den Bedürfnissen von Patientinnen und Patienten richtet.

Hinter evidenzbasierter Medizin steht eine hinterfragende Grundhaltung [306]. 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 Patientinnen und Patienten haben kann [140,429]. Aufgabe des Instituts ist es, objektiv zu bewerten, mit welcher Sicherheit der Nutzen medizinischer Maßnahmen nachgewiesen ist, um nicht sachgerechten Urteilen entgegenzuwirken.

1.2.2 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 Patientinnen und 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 Personen mit ihren möglichen Besonderheiten, sondern darauf, für welche Gruppen von Patientinnen und Patienten es Belege für einen Nutzen gibt. Der G-BA berücksichtigt dann bei seinen Entscheidungen Aspekte der Patientenversorgung, die außerhalb der Nutzenbewertung liegen [198].

1.2.3 Evidenzbasierte Entscheidungsfindung im Gesundheitswesen

Die Berichte des Instituts sollen dem G-BA als eine Grundlage für Entscheidungen dienen, die im Grundsatz für alle gesetzlich Krankenversicherten gelten. Weitere Institutsprodukte 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.

1.2.4 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 der Verzicht auf eine 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 Patientinnen und 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 Nutznachweis zu führen, sind Studien an ausreichend großen Gruppen von geeigneten Personen nötig. Zur Planung, Durchführung und Auswertung solcher Studien haben internationale Wissenschaftlerinnen und 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 nach der Qualität ihres Designs und ihrer Durchführung angemessenen Studien zu einer Frage zu identifizieren und so den Stand des zuverlässigen 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.

1.2.5 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 gesetzlich Krankenversicherte weitreichende Konsequenzen wie zum Beispiel den Ausschluss von Leistungen haben können.

Die Bewertung der Ergebnissicherheit spielt daher 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 und von der Größe des Effekts einer Intervention ab: 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 Patientinnen und 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 Patientinnen und Patienten genügend Sicherheit bieten. Allerdings sind Therapien, die einen so dramatischen Nutzen aufweisen, in der heutigen Medizin sehr selten.

Gerade bei chronischen Krankheiten sind Unterschiede zwischen zwei Therapiealternativen meist kleiner und leicht durch variable und schwankende Krankheitsverläufe überlagert. Hier sind sehr präzise Methoden und geeignete Studiendesigns 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 verlässlichere Studien sind nötig, um hinreichend sicher sein zu können, dass ein beobachteter Unterschied nicht auf einem 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.

Die Anforderungen an die 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 unberechtigt benachteiligt wird.
- 2) Das Institut stützt sich 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.
- 3) 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.

- 4) Der G-BA hat die Aufgabe, diese Unsicherheit bei seiner Entscheidungsfindung zu beachten. Neben der wissenschaftlichen Sachlage bezieht er auch andere Aspekte in seine Entscheidungen ein, wozu neben der Wirtschaftlichkeit auch die Bedürfnisse und Werte der Menschen gehören [209]. In einer wissenschaftlich unsicheren Situation gewinnen diese Aspekte an Gewicht. Zudem hat der G-BA die Möglichkeit, selbst Studien zu fordern, um die identifizierten Evidenzlücken zu schließen.

1.2.6 Die Verbindung von Ergebnissicherheit und Alltagsnähe

Großen Wert auf die Abschätzung der Ergebnissicherheit zu legen, wird oft kritisiert. Ein Argument ist, dass Studien mit hoher Ergebnissicherheit, insbesondere randomisierte kontrollierte Studien, zwar eine hohe interne Validität besäßen, jedoch oft die Versorgung von Patientinnen und Patienten im Alltag nicht abbildeten und deshalb nicht „übertragbar“ seien, also nur eine geringe externe Validität vorhanden sei. Dabei ist zu prüfen, wie gut die in Studien untersuchte Patientenpopulation, die eingesetzten Interventionen und die Zielkriterien dem Versorgungsalltag entsprechen. Diese Kritik ist dann oft mit der Forderung verbunden, andere Studientypen ohne Randomisierung hinzuzunehmen, um Alltagsbedingungen besser einfließen zu lassen.

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

- 1) Grundlage einer Nutzenbewertung ist der Nachweis von Kausalität. Unverzichtbare Bedingung für den Nachweis von Kausalität ist ein vergleichendes Experiment, das so angelegt sein muss, dass ein Unterschied zwischen Interventionsgruppen – ein Effekt – nur auf eine einzige Einflussgröße – die geprüfte Intervention – zurückgeführt werden kann. Dieses Ziel macht für klinische Studien erhebliche Anstrengungen nötig, weil es zahlreiche unerwünschte Einflüsse gibt, die einen Effekt vortäuschen oder auch verbergen können (Verzerrung). Der stärkste dieser verzerrenden Einflüsse sind ungleiche Ausgangsbedingungen zwischen den Vergleichsgruppen. Die Randomisierung zusammen mit einer sorgfältigen Verdeckung der Zuteilung (Concealment) ist das beste gegenwärtig verfügbare Instrument, 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 in randomisierten kontrollierten Studien (Randomized Controlled Trials, RCTs) eine grundlegende Voraussetzung für einen Kausalitätsnachweis gegeben. Allerdings ist auch eine Randomisierung alleine kein Garant für eine hohe Ergebnissicherheit. Dazu sind z. B. auch eine unverzerrte Erhebung, Zusammenfassung und Veröffentlichung der Ergebnisse nötig.
- 2) Andere Studientypen als RCTs sind in der Regel für einen Kausalitätsnachweis nicht geeignet. In nicht randomisierten vergleichenden Studien kann grundsätzlich nicht von einer Strukturgleichheit der Gruppen ausgegangen werden. Sie liefern daher immer ein

potenziell verzerrtes Ergebnis und können zumeist die maßgebliche Frage nicht hinreichend sicher beantworten, ob ein Unterschied auf der geprüften Intervention beruht. Daher bedarf die Verwendung nicht randomisierter Studien als Nachweis der Kausalität einer Intervention einer besonderen Begründung bzw. besonderer Voraussetzungen und spezieller Qualitätsanforderungen.

- 3) Es trifft zu, dass viele randomisierte Studien Aspekte des Alltags der Patientenversorgung nicht abbilden, z. B. Patientinnen und Patienten mit Begleiterkrankungen ausschließen, die im Alltag häufig sind. Dies ist jedoch keine Konsequenz der Technik der Randomisierung, sondern die Folge anderer Faktoren wie z. B. der Festlegung enger Ein- und Ausschlusskriterien der Studie, der Auswahl der Interventionen oder der Zielkriterien. Zudem werden Patientinnen und Patienten in randomisierten Studien oft anders (intensiver und enger) betreut als im Alltag. Das alles sind aber bewusste Entscheidungen derjenigen, die in einer Studie eine bestimmte Frage beantworten wollen. Durch einen Verzicht auf die Randomisierung werden diese Entscheidungen nicht geändert. Auch in nicht randomisierten Studien findet eine Selektion der Teilnehmer durch Ein- und Ausschlusskriterien und ggf. andere Designcharakteristika statt, sodass externe Validität ebenso wenig per se gegeben ist wie bei RCTs.
- 4) Aber selbst wenn sich Patientengruppen eines RCTs vom Versorgungsalltag 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 Ärztinnen und 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 Patientinnen und Patienten grundsätzlich sehr eng betreut werden.
- 6) Die Festlegung der Studienbedingungen liegt jedoch in der Hand der Initiatoren einer klinischen Studie. Sie können die Fragestellung und Endpunkte definieren, die sie als so wichtig einschätzen, 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 geeignete 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 größere Alltagsnähe aufgeben würde. Man würde einen Mangel durch Inkaufnahme eines zweiten, noch schwerwiegenderen Mangels zu kompensieren versuchen [237].

Es sind Studien wünschenswert und machbar, die Alltagsnähe und hohe Ergebnissicherheit vereinigen. RCTs, in denen weder über den Behandlungsalltag hinausgehende Anforderungen an Patientinnen und Patienten gestellt noch definierte Studienvisiten vorgegeben werden, sind sehr wohl möglich. Sie werden international als „Real World Trials“, „Practical Trials“ oder „Pragmatic Trials“ diskutiert [185,187,206,357,518]. Allerdings können solche „pragmatischen Studien“ ihrerseits ebenfalls zu Interpretationsproblemen führen. Werden z. B. sehr breite Einschlusskriterien gewählt, stellt sich die Frage, ob die (Gesamt-)Studienergebnisse auf die gesamte Studienpopulation anwendbar sind [550], was letztlich über adäquate Subgruppenanalysen zumindest ansatzweise zu beantworten wäre.

1.2.7 Nutzen im Einzelfall

Inhalt der Nutzenbewertung ist es, anhand der Ergebnisse von Studien, die geeignet sind, kausal begründete Effekte nachzuweisen, belastbare Voraussagen für zukünftige Patientinnen und Patienten zu treffen. Diese Aussagen betreffen immer Gruppen von Patientinnen und Patienten mit bestimmten Eigenschaften. Aussagen über den Nutzen im Sinne von Erfolgsaussagen für den Einzelfall sind grundsätzlich nicht möglich. Umgekehrt sind auf Einzelfällen basierende Erfahrungen – abgesehen von Ausnahmen (z. B. dramatischer Effekt) – für eine Nutzenbewertung ungeeignet, da es nicht möglich ist, Ergebnisse im Einzelfall (d. h. ohne Vergleich) einer Intervention zuzuschreiben.

Für bestimmte Fragestellungen (Therapieoptimierung bei einzelnen Personen) können sogenannte (randomisierte) „n-of-1-Studien“ durchgeführt werden [219,224,290,461]. Diese sind jedoch für die Nutzenbewertung einer Behandlungsmethode für zukünftige Patientinnen und Patienten in aller Regel nicht geeignet.

2 Produkte des Instituts

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

2.1 Produktspezifische Verfahrensabläufe

Zu den Produkten des Instituts gehören:

- Bericht
- Rapid Report (Schnellbericht)
- Dossierbewertung
- Addendum
- Gesundheitsinformation
- Arbeitspapier

Die Erstellung von Berichten und Rapid Reports erfolgt auf Basis eines Einzelauftrags durch den G-BA oder das BMG. Grundlage hierfür sind die in § 139a SGB V beschriebenen Aufgaben des Instituts (siehe auch Abschnitt 1.1). Entsprechend können Berichte und Rapid Reports zur Nutzenbewertung medikamentöser und nichtmedikamentöser Interventionen, zur Kosten-Nutzen-Bewertung, zur Bewertung von Leitlinien usw. erstellt werden. Der wesentliche Unterschied zwischen Berichten und Rapid Reports besteht darin, dass nur bei Berichten, nicht aber bei Rapid Reports Stellungnahmeverfahren (Anhörungen) durchgeführt werden. Entsprechend sind Rapid Reports insbesondere für zeitnahe Empfehlungen gedacht, zu denen aus Sicht des Auftraggebers keine Anhörungen durch das Institut erforderlich sind.

Dossierbewertungen werden vom G-BA beauftragt. Grundlage hierfür ist § 35a SGB V, in dem die Bewertung des Nutzens neuer Wirkstoffe auf Basis eines Dossiers des pharmazeutischen Unternehmers geregelt ist (siehe auch Abschnitt 3.3.3). Für Dossierbewertungen nach § 35a SGB V ist keine Anhörung durch das Institut vorgesehen. Diese erfolgt im weiteren Verfahren durch den G-BA.

Addenda können vom G-BA oder BMG beauftragt werden, wenn sich nach Erstellung eines Berichts, eines Rapid Reports oder einer Dossierbewertung im Zuge der Beratung ein zusätzlicher Bearbeitungsbedarf ergibt.

Gesundheitsinformationen können auf Basis eines Einzelauftrags erstellt werden, sie können aber auch Folge eines Auftrags zu anderen Aufgabenbereichen sein (allgemein verständliche

Version anderer Institutsprodukte, z. B. eines Berichts) oder im Rahmen der allgemeinen gesetzlichen Aufgabe der Bereitstellung von Gesundheitsinformationen erarbeitet werden.

Arbeitspapiere entstehen in Eigenverantwortung des Instituts, ohne dass es hierzu eines Auftrags durch den G-BA oder das BMG bedarf. Dies geschieht entweder auf Grundlage des Generalauftrags (siehe Abschnitt 2.1.6) 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 der Präambel dieses Dokuments dargestellt ist.

In der nachfolgenden Tabelle 1 sind die verschiedenen Produkte des Institutes in einer Übersicht dargestellt. Die produktspezifischen Verfahrensabläufe sind in den sich daran anschließenden Abschnitten 2.1.1 bis 2.1.6 beschrieben.

Tabelle 1 Übersicht über die Produkte des Instituts

Produkt	Zielsetzung	Ablauf	Beauftragung durch
Bericht	Empfehlungen zu den in § 139a SGB V beschriebenen Aufgaben, inklusive Anhörung	beschrieben in Abschnitt 2.1.1	G-BA, BMG
Rapid Report	Empfehlungen zu den in § 139a SGB V beschriebenen Aufgaben, sofern keine Anhörung zu Zwischenprodukten erforderlich ist, insbesondere zeitnahe Informationen zu aktuellen Themen	beschrieben in Abschnitt 2.1.2	G-BA, BMG
Dossierbewertung	Bewertung des Nutzens von Arzneimitteln mit neuen Wirkstoffen gemäß § 35a SGB V	beschrieben in Abschnitt 2.1.3	G-BA
Addendum	Zeitnahe Stellungnahme zu Fragestellungen, die sich im Zuge der Beratung abgeschlossener Institutsprodukte ergeben haben	beschrieben in Abschnitt 2.1.4	G-BA, BMG
Gesundheitsinformation	Allgemein verständliche Informationen für Bürgerinnen und Bürger und Patientinnen und Patienten, breit gefächertes Themenspektrum	beschrieben in Abschnitt 2.1.5	G-BA, BMG; auch Initiierung durch das Institut
Arbeitspapier	Information zu versorgungsrelevanten Entwicklungen in der Medizin oder zu methodischen Aspekten	beschrieben in Abschnitt 2.1.6	Initiierung durch das Institut

2.1.1 Bericht

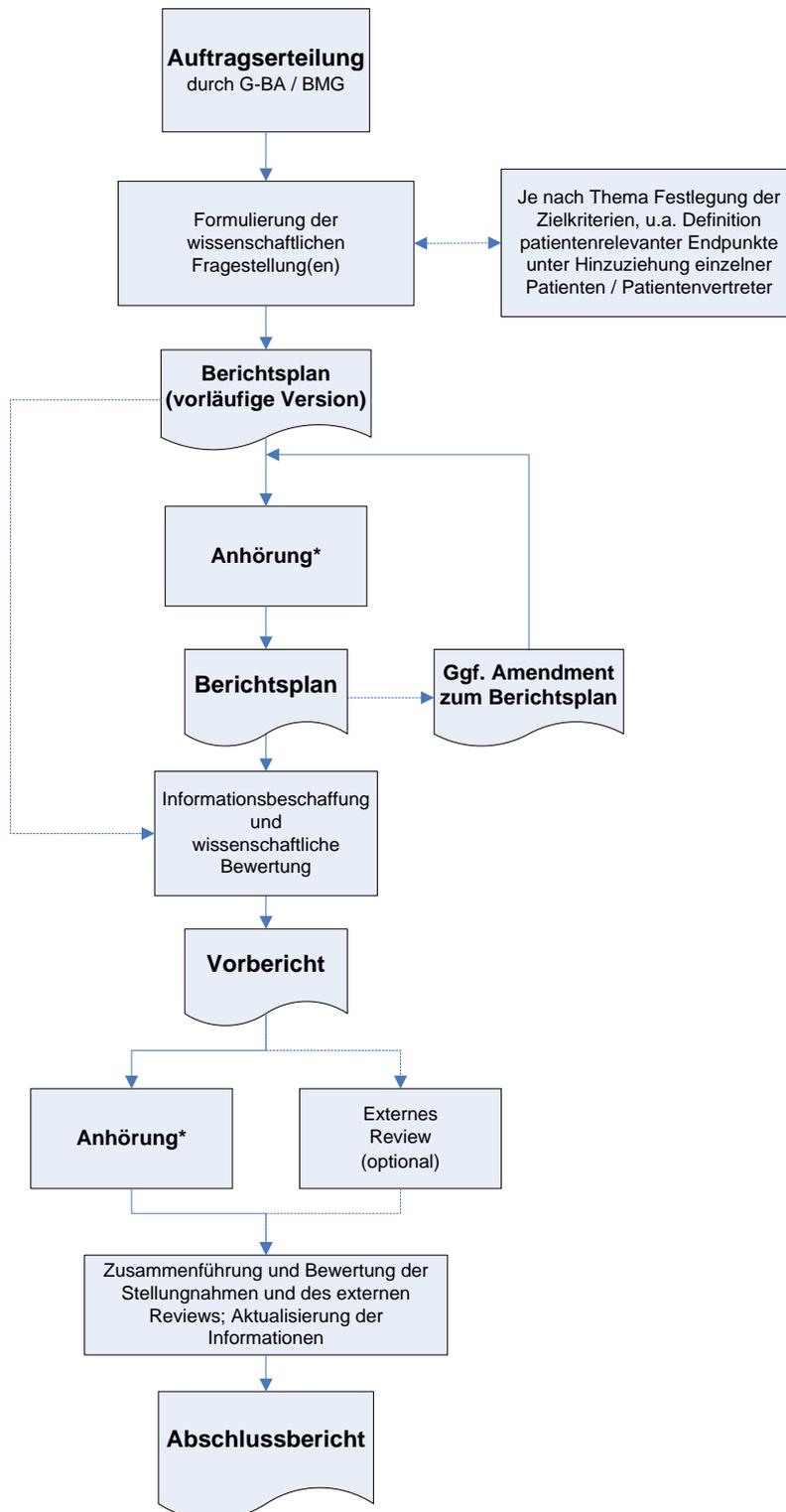
A) Ablauf der Berichterstellung

Der Ablauf der Berichterstellung ist in Abbildung 1 schematisch dargestellt. Alle Arbeitsschritte werden in Verantwortung des Instituts getätigt. Dabei werden regelhaft externe Sachverständige beteiligt (siehe Abschnitt 2.2.1). Bei Bedarf wird auch der wissenschaftliche Beirat des Instituts einbezogen. Der interne Qualitätssicherungsprozess ist in diesem Flussdiagramm nicht dargestellt.

Nach der **Auftragserteilung** durch den G-BA oder das BMG erfolgt die Formulierung der wissenschaftlichen Fragestellung. Je nach Thema ist dabei auch die Festlegung von Zielkriterien erforderlich, z. B. bei Nutzenbewertungen. Regelhaft werden dabei, insbesondere zur Definition patientenrelevanter Endpunkte, entsprechende Patientenorganisationen beteiligt, gegebenenfalls wird auch die Meinung einzelner Betroffener eingeholt. Anschließend wird der Berichtsplan erstellt.

Der **Berichtsplan** beinhaltet die genaue wissenschaftliche Fragestellung einschließlich der Zielkriterien (z. B. patientenrelevante Endpunkte), die Ein- und Ausschlusskriterien der für die Bewertung zu verwendenden Informationen sowie die Darlegung der projektspezifischen Methodik der Beschaffung und Bewertung dieser Informationen. Die vorläufige Version des Berichtsplans wird zunächst dem Auftraggeber, dem Vorstand der Stiftung, dem Stiftungsrat und dem Kuratorium zugeleitet. Die Veröffentlichung auf der Website des Instituts erfolgt i. d. R. fünf Arbeitstage später.

Für eine Frist von mindestens vier Wochen wird dann der Öffentlichkeit **Gelegenheit zur Stellungnahme (Anhörung)** gegeben (schriftliche Stellungnahmen). Die Gelegenheit zur Stellungnahme bezieht sich dabei insbesondere auf das projektspezifische methodische Vorgehen zur Beantwortung der Fragestellung. Die Fragestellung selbst 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 Klärung von Inhalten der schriftlichen Stellungnahmen mit dem Ziel der Verbesserung der wissenschaftlichen Qualität des Berichtsplans.



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

Abbildung 1: Ablauf der Berichtserstellung

Nach Auswertung der Stellungnahmen und ggf. Durchführung der Erörterung wird der Berichtsplan in einer überarbeiteten Version zusammen mit der Dokumentation der Anhörung zum Berichtsplan zunächst dem Auftraggeber, dem Vorstand der Stiftung, dem Stiftungsrat und dem Kuratorium zugeleitet und i. d. R. fünf Arbeitstage später auf der Website des Instituts veröffentlicht. Der überarbeitete 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 der Anhörung zum Berichtsplan auf Basis der im vorläufigen Berichtsplan formulierten Kriterien. Das Ergebnis der Anhörung wird damit jedoch explizit nicht vorweggenommen, da diese Kriterien sich durch die Anhörung zum Berichtsplan in seiner vorläufigen Version ändern können. Dies kann auch zu einer Ergänzung und/oder Änderung der Informationsbeschaffung und -bewertung führen.

Der Vorbericht enthält die vorläufige Empfehlung an den G-BA. Er wird nach Fertigstellung zunächst dem Auftraggeber, dem Vorstand der Stiftung, dem Stiftungsrat und dem Kuratorium zugeleitet. Die Veröffentlichung auf der Website des Instituts erfolgt i. d. R. fünf Arbeitstage nach Versendung an den Auftraggeber.

Für eine Frist von mindestens vier Wochen wird dann der Öffentlichkeit **Gelegenheit zur Stellungnahme (Anhörung)** gegeben (schriftliche Stellungnahmen). Gegenstand des Stellungnahmeverfahrens sind insbesondere die im Vorbericht dargestellten Ergebnisse der Informationsbeschaffung und -bewertung. Optional kann eine mündliche wissenschaftliche Erörterung mit Stellungnehmenden durchgeführt werden. Diese Erörterung dient der ggf. notwendigen Klärung von Inhalten der schriftlichen Stellungnahmen mit dem Ziel der Verbesserung der wissenschaftlichen Qualität des Abschlussberichts.

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. Der Abschlussbericht und die Dokumentation der Anhörung zum Vorbericht werden zunächst dem Auftraggeber, dem Vorstand der Stiftung sowie dem Stiftungsrat und anschließend (i. d. R. vier Wochen später) dem Kuratorium der Stiftung zugeleitet. I. d. R. weitere vier Wochen später erfolgt die Veröffentlichung dieser Dokumente auf der Website des Instituts. Sollten Stellungnahmen zu Abschlussberichten eingehen, die substantielle 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.

Der Auftraggeber entscheidet über die Beauftragung des Instituts. Eine solche Aktualisierung unterliegt den allgemeinen methodischen und verfahrenstechnischen Anforderungen an Produkte des Instituts.

B) Allgemeine Anmerkungen zum Stellungnahmeverfahren (Anhörung)

Stellungnahmeberechtigte Organisationen

Das Institut hat nach § 139a Abs. 5 SGB V zu gewährleisten, dass in allen wichtigen Abschnitten des Bewertungsverfahrens den 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 dem oder der Beauftragten der Bundesregierung für die Belange der Patientinnen und Patienten Gelegenheit zur Stellungnahme gegeben wird. Die Stellungnahmen sind in die Entscheidung einzubeziehen. Diesen Anforderungen wird dadurch Rechnung getragen, dass Anhörungen zum Berichtsplan und zum Vorbericht durchgeführt werden und der Personenkreis der Stellungnahmeberechtigten nicht begrenzt wird. Darüber hinaus werden alle Institutsprodukte nach § 139a SGB V vor der Publikation dem Kuratorium des Instituts zugeleitet. Im Kuratorium sind Patientenorganisationen, der oder die Patientenbeauftragte der Bundesregierung, Organisationen der Leistungserbringer und der Sozialpartner und die Selbstverwaltungsorgane des Gemeinsamen Bundesausschusses vertreten.

Formale Anforderungen

Um die Arbeit des Instituts nicht unangemessen zu verzögern, müssen die Stellungnahmen bestimmten formalen Anforderungen genügen. Weiterführende Angaben zum Stellungnahmeverfahren einschließlich der Bedingungen für die Teilnahme an einer wissenschaftlichen Erörterung finden sich in einem Leitfaden, der auf der Website des Instituts abgerufen werden kann.

Veröffentlichung der Stellungnahmen

Stellungnahmen, die den formalen Anforderungen genügen, werden auf der Website des Instituts in einem gesonderten Dokument (Dokumentation und Würdigung der Anhörung) veröffentlicht. Zur Gewährleistung der Transparenz werden mit den Stellungnahmen eingereichte Unterlagen, die nicht öffentlich zugänglich sind (z. B. Manuskripte), ebenfalls veröffentlicht.

Vorlage von Unterlagen im Rahmen der Anhörung

Im Rahmen der Anhörung sowohl zum Berichtsplan als auch zum Vorbericht besteht die Möglichkeit, qualitativ angemessene Unterlagen jeglicher Art, die aus Sicht des jeweiligen Stellungnehmenden für die 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.

2.1.2 Rapid Report

Der Ablauf der Erstellung eines **Rapid Reports** ist in Abbildung 2 schematisch dargestellt. Alle Arbeitsschritte werden in Verantwortung des Instituts getätigt. Dabei können externe Sachverständige beteiligt werden (siehe Abschnitt 2.2.1). Bei Bedarf wird auch der wissenschaftliche Beirat des Instituts einbezogen. Der interne Qualitätssicherungsprozess ist in diesem Flussdiagramm nicht dargestellt.

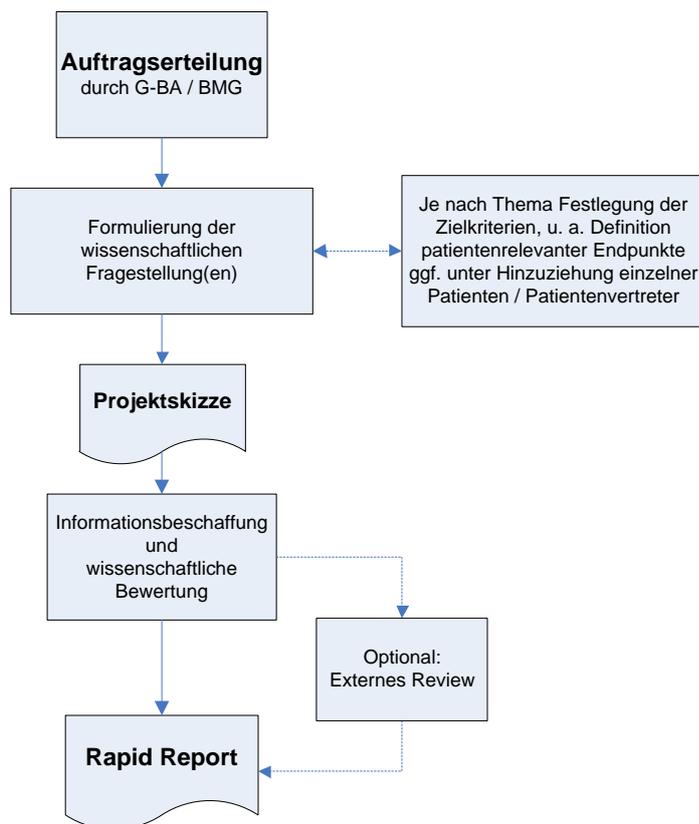


Abbildung 2: Ablauf der Erstellung eines Rapid Reports

Rapid Reports werden vorrangig mit dem Ziel einer zeitnahen Information zu relevanten Entwicklungen im Gesundheitswesen (z. B. neue Technologien, Publikation wichtiger Meilensteinstudien) erstellt. Dabei ist zumeist eine kürzere Erarbeitungszeit notwendig. Zwischenprodukte werden daher nicht veröffentlicht und nicht zur Anhörung gestellt.

Nach der **Auftragserteilung** durch den G-BA oder das BMG erfolgt die Formulierung der wissenschaftlichen Fragestellung. Je nach Thema ist dabei auch die Festlegung von Zielkriterien erforderlich, z. B. bei Nutzenbewertungen. Dabei können auch Patientenorganisationen beteiligt oder die Meinung einzelner Betroffener eingeholt werden, insbesondere

zur Definition patientenrelevanter Endpunkte. Anschließend erfolgt die Erstellung der Projektskizze.

In der **Projektskizze** werden die wesentlichen Schritte der Informationsbeschaffung und der wissenschaftlichen Bewertung zusammengefasst. Sie bildet die Grundlage für die Erstellung des Rapid Report. Die Projektskizze wird nicht veröffentlicht.

Im **Rapid Report** werden die Ergebnisse der Informationsbeschaffung und wissenschaftlichen Bewertung dargestellt. Vor Fertigstellung kann optional ein Entwurf des Rapid Report einem oder mehreren externen Reviewern (siehe Abschnitt 2.2.3) mit ausgewiesener methodischer und/oder fachlicher Kompetenz als weiterer Schritt der Qualitätssicherung vorgelegt werden. Nach Fertigstellung wird der Rapid Report an den Auftraggeber, den Vorstand der Stiftung sowie den Stiftungsrat und anschließend (i. d. R. eine Woche später) an das Kuratorium versandt. Die Veröffentlichung auf der Website des Instituts erfolgt i. d. R. vier Wochen nach Versand an Auftraggeber und Vorstand. 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. Der Auftraggeber entscheidet über die Beauftragung des Instituts. Eine solche Aktualisierung unterliegt den generellen methodischen und verfahrenstechnischen Anforderungen an Produkte des Instituts.

2.1.3 Dossierbewertung

Der Ablauf der Erstellung einer Dossierbewertung ist in Abbildung 3 schematisch dargestellt. Alle Arbeitsschritte werden in Verantwortung des Instituts getätigt. Dabei werden regelhaft externe Sachverständige beteiligt (siehe Abschnitt 2.2.1). Bei Bedarf wird auch der Wissenschaftliche Beirat des Instituts einbezogen. Der interne Qualitätssicherungsprozess ist in diesem Flussdiagramm nicht dargestellt.

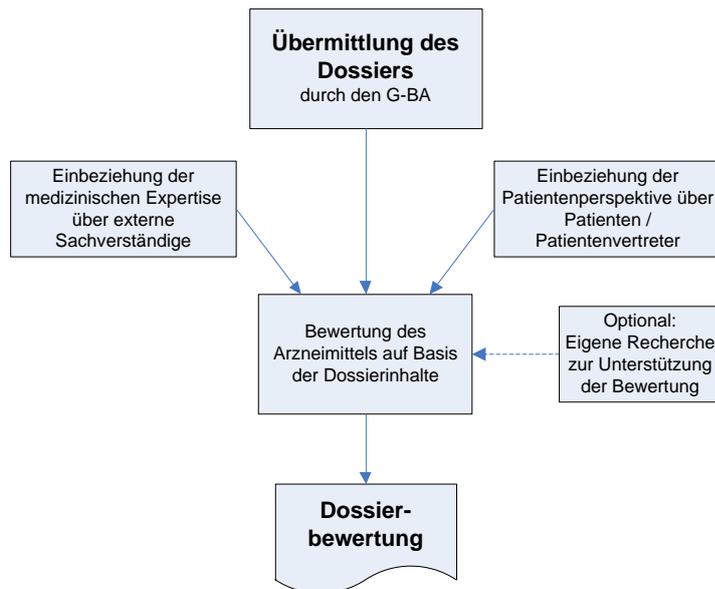


Abbildung 3: Ablauf der Erstellung einer Dossierbewertung

Nach **Übermittlung des Dossiers** durch den G-BA erfolgt die inhaltliche Bewertung der Dossierinhalte in Verantwortung des Instituts. Dabei werden regelhaft die medizinische Expertise über externe Sachverständige sowie die Patientenperspektive über Patientinnen und Patienten bzw. Patientenorganisationen einbezogen.

Die **Einbeziehung der medizinischen Expertise** erfolgt primär auf Basis eines Fragenkatalogs, der zu Beginn der Bewertung an die externen Sachverständigen verschickt wird. Das Institut berücksichtigt bei seiner Bewertung die Rückmeldungen der externen Sachverständigen. Darüber hinaus können die externen Sachverständigen im Bedarfsfall zur Klärung spezieller Fragen im Verlauf der Bewertung herangezogen werden. Externe Sachverständige werden über eine institutseigene Sachverständigendatenbank identifiziert (siehe Abschnitt 2.2.1).

Die **Einbeziehung der Patientenperspektive** erfolgt auf Basis eines Fragebogens, der zu Beginn der Bewertung an Patientinnen und Patienten bzw. Patientenorganisationen verschickt wird. Das Institut berücksichtigt bei seiner Bewertung die in diesem Fragebogen getroffenen Angaben z. B. zu relevanten Endpunkten und zu wichtigen Subgruppen. Patientinnen und Patienten bzw. Patientenorganisationen werden über die in § 140f SGB V benannten maßgeblichen Organisationen identifiziert.

Grundlage für die Bewertung ist das vom pharmazeutischen Unternehmer an den G-BA und dann weiter an das Institut übermittelte Dossier. Optional kann zur Unterstützung der Bewertung eine **eigene Recherche** des Instituts durchgeführt werden.

Den Abschluss des Prozesses bildet die Erstellung einer **Dossierbewertung**. Nach § 35a SGB V muss die Bewertung spätestens drei Monate nach dem für die Einreichung des

Dossiers maßgeblichen Zeitpunkt abgeschlossen werden. Die Dossierbewertung wird nach ihrer Fertigstellung dem G-BA zugeleitet. Zeitnah im Anschluss daran erfolgen die Weiterleitung an den Vorstand der Stiftung, den Stiftungsrat und an das Kuratorium der Stiftung sowie die Publikation der Dossierbewertung auf der Website des Instituts.

2.1.4 Addendum

Der Ablauf der Erstellung eines Addendums ist in Abbildung 4 schematisch dargestellt. Alle Arbeitsschritte werden in Verantwortung des Instituts, bei Bedarf unter Einbeziehung des Wissenschaftlichen Beirats des Instituts, getätigt. Der interne Qualitätssicherungsprozess ist in diesem Flussdiagramm nicht dargestellt.

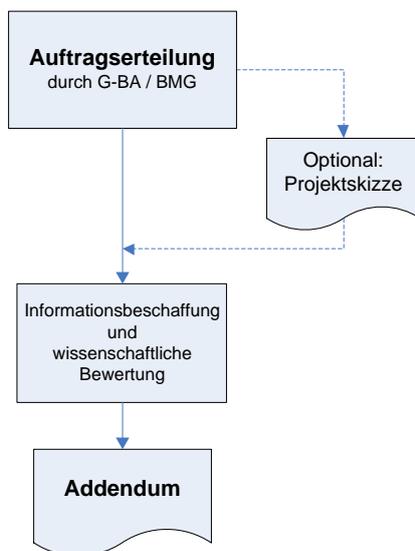


Abbildung 4: Ablauf der Erstellung eines Addendums

Ein Addendum kann in Auftrag gegeben werden, wenn sich im Zuge der Beratungen zu fertiggestellten Institutsprodukten zusätzlicher Bearbeitungsbedarf ergibt. Je nach Art und Umfang der Fragestellung kann es sinnvoll sein, eine **Projektskizze** zu erstellen, in der die wesentlichen Schritte der Informationsbeschaffung und der wissenschaftlichen Bewertung zusammengefasst werden. Die Projektskizze wird nicht veröffentlicht.

Bei dem **Addendum** kann es je nach Art und Umfang der Fragestellung sinnvoll sein, diejenigen externen Sachverständigen zu beteiligen, die an der Erstellung des zugrunde liegenden Institutsprodukts beteiligt waren. Nach Fertigstellung wird das Addendum zunächst an den G-BA sowie an Stiftungsrat und Vorstand versandt. I. d. R. eine Woche später wird es dem Kuratorium der Stiftung übermittelt und weitere drei Wochen später auf der Website des Instituts veröffentlicht.

2.1.5 Gesundheitsinformationen

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

- **Informationsberichte:** umfassende Beiträge, die die Grundlage für ein erweitertes Produkt zu einem relevanten gesundheitlichen Thema darstellen
- **Merkblätter:** kurze, leicht verständliche Artikel
- **Kurzantworten:** Zusammenfassungen von systematischen Übersichten, HTA-Berichten oder größeren Studien, darunter auch Zusammenfassungen der anderen Produkte des Instituts, sofern sie für Patientinnen und Patienten relevant sind

Diese Informationsprodukte 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 im Institut ist in Abbildung 5 schematisch dargestellt. In die Erstellung der Informationsprodukte werden bei mehreren Schritten externe Personen einbezogen. Die je nach Art des Produkts übernommenen Aufgaben sind in Abschnitt 5 genauer beschrieben; auch 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 Bürgerinnen und Bürger sowie aus Eigeninitiative des Instituts gemäß dem Generalauftrag des G-BA.

Der Generalauftrag des Instituts (siehe Abschnitt 2.1.6) wurde im Juli 2006 sowie im März 2008 in Hinblick auf die Erstellung von Gesundheitsinformationen konkretisiert und angepasst, um auch Informationen für Bürgerinnen und 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 Patientinnen und Patienten.

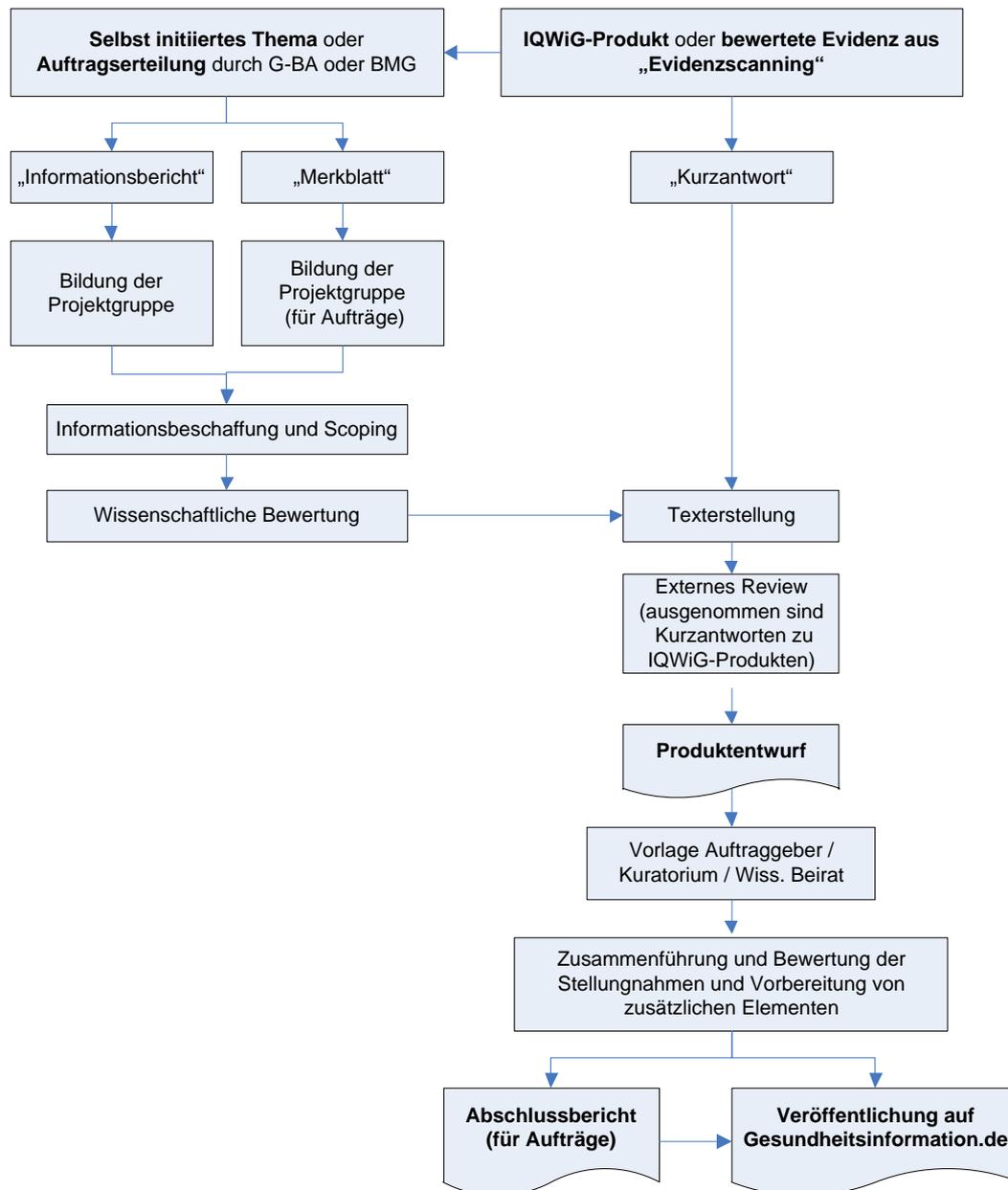


Abbildung 5: Ablauf der Gesundheitsinformationserstellung

Nach der **Auftragsvergabe** durch den G-BA oder das BMG wird die interne Projektgruppe gebildet. Den Projektgruppen gehört mindestens ein Institutsmitglied an, das nicht aus dem Ressort „Gesundheitsinformation“ stammt. Ebenso wird für jeden Informationsbericht eine Projektgruppe gebildet.

Im Anschluss an die **Texterstellung** und die ressortinterne Qualitätssicherung werden die Entwürfe zum **externen Review** verschickt. Die Entwürfe zu den Informationsberichten und Merkblättern werden an mindestens einen externen Reviewer versandt. Kurzantworten werden der Autorin oder dem Autor des zusammengefassten wissenschaftlichen Beitrags präsentiert. Die Kurzantworten zu anderen Produkten des Instituts (Begleitinformationen)

werden gewöhnlich nur von den intern Beteiligten des jeweiligen Ressorts begutachtet, können aber auch an externe Reviewer versandt werden.

Die zusätzlichen Elemente unterliegen denselben internen Reviewprozessen wie die entsprechenden Informationsberichte, Merkblätter und Kurzantworten; auch sie können ggf. einem externen Review unterzogen werden. Patientengeschichten („Erfahrungsberichte“, siehe Abschnitt 5.4.4) werden nur veröffentlicht, wenn eine schriftliche Einverständniserklärung der betreffenden Patientin oder des Patienten vorliegt.

Der endgültige Entwurf zu einem Informationsprodukt wird im Rahmen einer einmonatigen Beratungsperiode zur begrenzten **Stellungnahme** an den Auftraggeber und die Gremien des Instituts verschickt. Eines davon ist das Kuratorium der Stiftung. Im Kuratorium sind Organisationen der Leistungserbringer und der Sozialpartner und die Selbstverwaltungsorgane des Gemeinsamen Bundesausschusses vertreten; diesem Gremium gehören auch Vertreterinnen und Vertreter der für die Wahrnehmung der Interessen der Patientinnen und Patienten und die Selbsthilfe chronisch kranker und behinderter Menschen maßgeblichen Organisationen an sowie die oder der Beauftragte der Bundesregierung für die Belange der Patientinnen und Patienten. Im Gegensatz zu den Vorberichten werden Entwürfe zu Patienteninformationen nicht auf der Website des Instituts veröffentlicht. Alle Informationsberichte, Merkblätter und Kurzantworten durchlaufen – in der Regel zeitgleich mit dem Stellungnahmeverfahren – eine externe Nutzertesting. Bei der Nutzertesting kommentiert eine Gruppe von Patientinnen und Patienten oder möglichen Nutzerinnen und Nutzern die Texte im Hinblick auf ihren Informationsgehalt und ihre Verständlichkeit.

Die während dieser Beratungsperiode eingegangenen Kommentare werden zusammengefasst und geprüft. Die Zusammenfassungen und die ausführlichen Versionen stehen auch der Projektgruppe – falls vorhanden – und der Institutsleitung 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, dem Vorstand der Stiftung sowie dem Stiftungsrat und anschließend (i. d. R. vier Wochen später) dem Kuratorium der Stiftung zugeleitet. I. d. R. weitere vier Wochen später erfolgt die Veröffentlichung des Abschlussberichts auf der Website (www.iqwig.de) des Instituts. Anschließend wird die betreffende Gesundheitsinformation selbst auf der Institutswebsite für Patientinnen und Patienten und Bürgerinnen 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.

Korrekturen, Verbesserungen und Aktualisierungen der Gesundheitsinformationen des Instituts werden intern vorgenommen, bei wichtigeren inhaltlichen Aktualisierungen auch

durch externe Gutachterinnen und Gutachter. Eine ausführlichere Beschreibung der Aktualisierungsmechanismen des Instituts findet sich in Kapitel 5.

2.1.6 Arbeitspapier

Der Ablauf der Erstellung eines **Arbeitspapiers** ist in Abbildung 6 schematisch dargestellt. Alle Arbeitsschritte werden in Verantwortung des Instituts, bei Bedarf unter Einbeziehung externer Expertise bzw. des Wissenschaftlichen Beirats des Instituts, getätigt. Der interne Qualitätssicherungsprozess ist in diesem Flussdiagramm nicht dargestellt.

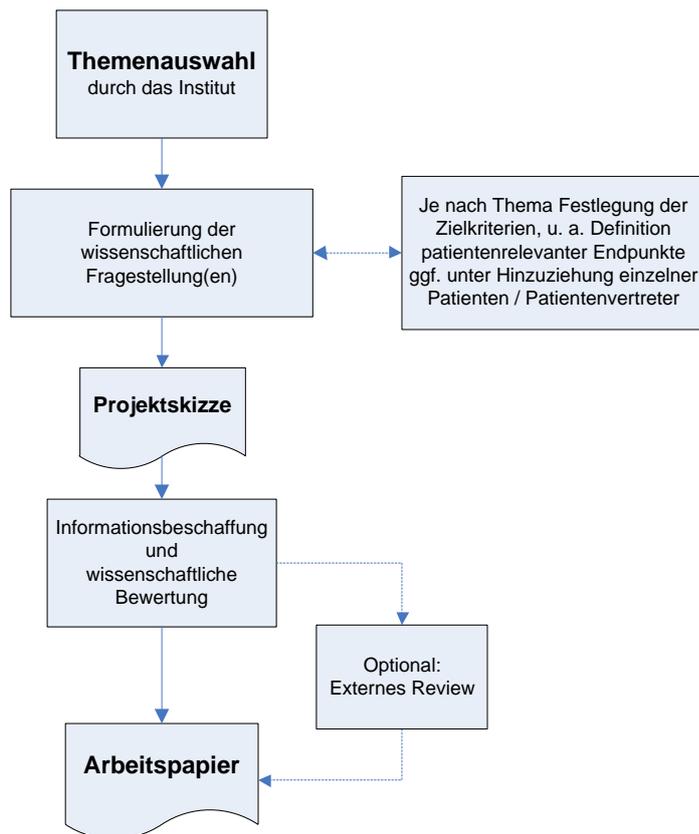


Abbildung 6: Ablauf der Erstellung eines Arbeitspapiers

Die Erstellung von Arbeitspapieren erfolgt u. a. im Rahmen des **Generalauftrags**, der am 21.12.2004 durch den G-BA erteilt und im Juli 2006 sowie im März 2008 in Hinblick auf die Erstellung von Gesundheitsinformationen weiter konkretisiert bzw. angepasst wurde. 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 V 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.

Die **Themenauswahl** erfolgt innerhalb des Instituts, insbesondere auf Basis der im Generalauftrag formulierten Kriterien. An der Formulierung der wissenschaftlichen Fragestellung können auch Patientenorganisationen beteiligt sein oder die Meinung einzelner Betroffener kann eingeholt werden, insbesondere zur Definition patientenrelevanter Endpunkte. Anschließend wird die Projektskizze erstellt.

In der **Projektskizze** werden die wesentlichen Schritte der Informationsbeschaffung und der wissenschaftlichen Bewertung zusammengefasst. Sie bildet die Grundlage für die Erstellung des Arbeitspapiers. Die Projektskizze wird nicht veröffentlicht.

Im Arbeitspapier werden die Ergebnisse der Informationsbeschaffung und wissenschaftlichen Bewertungen dargestellt. Die Qualitätssicherung kann optional ein externes Review beinhalten. Das Arbeitspapier wird nach seiner Fertigstellung zunächst an den G-BA, den Vorstand der Stiftung sowie den Stiftungsrat versandt, i. d. R. eine Woche später dem Kuratorium der Stiftung übermittelt und weitere drei Wochen später auf der Website des Instituts veröffentlicht. Sollten Stellungnahmen zu Arbeitspapieren eingehen, die substantielle 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. Eine solche Aktualisierung 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
- das Review 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 oder Begutachtung von Institutsprodukten vergeben werden oder die das Institut in medizinisch-fachlichen Fragestellungen beraten. Das Institut vergibt diese Aufträge im Rahmen eines den allgemeinen Grundsätzen des Beschaffungswesens unterliegenden Verfahrens.

Die aktuell durch das Institut zu vergebenden Aufträge nach § 139b Abs. 3 SGB V sind auf der Website des Instituts veröffentlicht. 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.

Die Beauftragung externer Sachverständiger bei Nutzenbewertungen nach § 35a SGB V erfolgt auf der Basis von Angaben Interessierter in einer Sachverständigendatenbank. Für die Aufnahme in die Sachverständigendatenbank ist auf der Website des Instituts ein Zugang eingerichtet, über den die Interessentinnen und Interessenten unter Angabe ihres Fachgebiets und ihrer fachlichen Expertise ihr Sachverständigenprofil eingeben können. Für die zu vergebenden Projekte wird anhand einer Kriterienliste die oder der jeweils am besten geeignete Bewerberin oder Bewerber des entsprechenden Fachgebiets aus der Sachverständigendatenbank ausgewählt und beauftragt. Weitere Angaben zum Auswahlverfahren sind auf der Website des Instituts veröffentlicht.

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 internen fachlichen Unabhängigkeit

Den wissenschaftlichen Mitarbeiterinnen und Mitarbeitern des Institutes sind bezahlte Nebentätigkeiten, die grundsätzlich geeignet sind, ihre fachliche Unabhängigkeit infrage zu stellen, untersagt. Alle Nebentätigkeiten sind von der Institutsleitung zu genehmigen. Zu Nebentätigkeiten gehören im weitesten Sinne auch unentgeltliche Ehrenämter wie Aufsichtsrats- oder Vorstandsämter in Vereinen oder Gesellschaften.

B) 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 müssen dem Institut gemäß § 139b SGB V „alle Beziehungen zu Interessenverbänden, Auftragsinstituten, insbesondere der pharmazeutischen Industrie und der Medizinprodukteindustrie, einschließlich Art und Höhe von Zuwendungen“ offengelegt werden.

Aufbauend auf der in der Wissenschaft üblichen Darlegung solcher Beziehungen als potenzielle Interessenkonflikte [339,345] interpretiert das Institut diese Regelung als Aufgabe, im Rahmen der Auswahl von externen Sachverständigen die Offenlegungen im Hinblick auf

die fachliche Unabhängigkeit und Unbefangenheit der Bewerber zu beurteilen. Dabei bewertet das Institut, ob aufgrund der angegebenen finanziellen Beziehungen ein Interessenkonflikt für ein konkretes Auftragsthema besteht. Wenn das der Fall ist, wird im zweiten Schritt bewertet, ob dieser Interessenkonflikt zu gravierenden Bedenken gegen eine sachgerechte Mitarbeit führt. Sollte das der Fall sein, kommt eine Mitarbeit für dieses Auftragsthema i. d. R. nicht oder nur unter besonderen Vereinbarungen infrage. Da diese Bewertung auftragsbezogen stattfindet, ist eine Mitarbeit bei anderen Auftragsthemen sehr wohl möglich. Der weitere Prozess zur Auswahl von externen Sachverständigen ist in Abschnitt 2.2.1 beschrieben.

Die Hauptgrundlage der Bewertung von Interessenkonflikten sind Selbstauskünfte unter Verwendung des „Formblatts zur Offenlegung potenzieller Interessenkonflikte“. Dieses Formblatt steht auf der Website des Instituts zur Verfügung. Die Selbstauskunft bezieht sich auf folgende sechs Arten von finanziellen Beziehungen:

- abhängige Beschäftigung / Anstellung
- Beratungstätigkeit
- Honorare zum Beispiel für Vorträge, Stellungnahmen, Ausrichtung und / oder Teilnahme an Kongressen und Seminaren
- finanzielle Unterstützung für Forschungsaktivitäten, andere wissenschaftliche Leistungen oder Patentanmeldungen
- sonstige finanzielle oder geldwerte Zuwendungen (z. B. Ausrüstung, Personal, Reisekostenunterstützung ohne wissenschaftliche Gegenleistung)
- Aktien, Optionsscheine oder sonstige Geschäftsanteile

Das Institut behält sich vor, zusätzliche Informationen heranzuziehen und Angaben auf ihre Vollständigkeit und Richtigkeit zu überprüfen.

Die Namen der externen Sachverständigen, die an der Erstellung der Produkte des Instituts beteiligt waren, werden in der Regel in den entsprechenden Produkten veröffentlicht. Diese Veröffentlichungen sind grundsätzlich über die Website frei zugänglich. Die Angaben zu Interessenkonflikten werden nur zusammenfassend veröffentlicht. Dabei wird für die im Formblatt erfassten Arten von Beziehungen lediglich genannt, ob diese Art der Beziehung bestand. Konkrete Einzelheiten, wie zum Beispiel Partner oder Höhe von Zuwendungen, werden nicht 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.

2.2.3 Review der Produkte des Instituts

Das Review 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 kann im Verlauf der Produkterstellung ein externes Reviewverfahren als optionaler weiterer Schritt der Qualitätssicherung durchgeführt werden. Die Auswahl der internen und externen Reviewer erfolgt primär auf Basis ihrer methodischen und / oder fachlichen Expertise.

Die Identifikation externer Reviewer kann 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. Auch für die externen Reviewer ist die Darlegung potenzieller Interessenkonflikte erforderlich.

Die Auswahl der externen Reviewer erfolgt durch das Institut. Eine Höchstgrenze von Reviewern gibt es nicht. Die externen Reviews werden hinsichtlich ihrer Relevanz für das jeweilige Produkt geprüft. Eine Veröffentlichung der externen Reviews erfolgt nicht. Die Namen der externen Reviewer 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 Reviewer 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 oder wirtschaftlichen Aspekten 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 Nutzenbewertung medizinischer Interventionen

3.1 Patientenrelevanter medizinischer Nutzen und Schaden

3.1.1 Definition des patientenrelevanten medizinischen Nutzens bzw. Schadens

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. „Kausal begründet“ meint in diesem Zusammenhang, dass ausreichend Sicherheit besteht, dass beobachtete Effekte allein auf die zu prüfende Intervention zurückgeführt werden können [549]. Die Begriffe „Nutzen“ und „Schaden“ beziehen sich auf einen Vergleich mit Placebo (oder einer andersartigen Scheinintervention) oder keiner Behandlung.

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; die Beschreibung erfolgt dabei immer ausgehend von der zu evaluierenden Intervention:

- 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.

Die Bewertung der Evidenz soll nach Möglichkeit 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. Darüber hinaus kann es sinnvoll sein, für den Fall eines nicht eindeutig belegten (Zusatz-)Nutzens (bzw. Schadens) eine weitere Untergliederung vorzunehmen, ob nämlich zumindest „Hinweise“ auf oder auch nur „Anhaltspunkte“ für einen (Zusatz-)Nutzen (bzw. Schaden) bestehen (siehe Abschnitt 3.1.4).

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 ob er 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 Patientinnen und Patienten mit der Behandlung berücksichtigt werden. Ein Nutzen oder Zusatznutzen kann sich allein auf Basis dieser Zielgrößen jedoch nicht ergeben. Für alle genannten Zielgrößen gilt, dass es erforderlich sein kann, diese jeweils nur im Zusammenhang mit Informationen zur vergleichenden Beeinflussung anderer Zielgrößen zu bewerten. Bei besonders schwerwiegenden oder gar lebensbedrohlichen Erkrankungen ist es beispielsweise in der Regel nicht allein ausreichend, eine Verbesserung der Lebensqualität durch die Anwendung einer zu bewertenden Intervention nachzuweisen, wenn nicht gleichzeitig mit ausreichender Sicherheit ausgeschlossen werden kann, dass schwerwiegende Morbidität oder gar Mortalität in einem nicht mehr akzeptablen Ausmaß ungünstig beeinflusst werden. Dies entspricht im Grundsatz der höchstrichterlichen Rechtsprechung, dass bestimmte (Nutzen-)Aspekte erst dann notwendigerweise zu bewerten sind, wenn die therapeutische Wirksamkeit hinreichend belegt ist [71]. Darüber hinaus wird umgekehrt in vielen Bereichen, insbesondere in palliativen Therapiesituationen, eine Beeinflussung der Mortalität nicht ohne Kenntnis damit einhergehender (ggf. ungünstiger) Effekte auf die Lebensqualität adäquat zu bewerten sein.

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 [160]. Darüber hinaus können für die Nutzenbewertung valide Surrogatendpunkte berücksichtigt werden (siehe Abschnitt 3.1.2).

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 (entsprechende Methoden werden am Ende des Abschnittes 3.1.4 exemplarisch aufgeführt). In einer solchen Situation kann es sinnvoll sein, eine Hierarchisierung von Endpunkten vorzunehmen. Allgemeine Nutzen- bzw. Schadenaussagen werden sich dann in erster Linie auf Belege bezüglich der höher gewichteten Zielgrößen stützen. Geplante Subgruppen- und Sensitivitätsanalysen werden dann im Wesentlichen für die höher gewichteten Zielgrößen durchgeführt, während solche Analysen für die verbleibenden Zielgrößen nicht routinemäßig durchlaufen werden.

Diagnostische Maßnahmen können indirekt nutzen, indem sie eine notwendige Voraussetzung für therapeutische Interventionen sind, durch die das Erreichen eines Effekts 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 Patientinnen und Patienten in Abhängigkeit vom Testergebnis.

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.

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. Dabei werden Nutzen- und Schadenaspekte zunächst endpunktbezogen evaluiert und dargestellt. Darüber hinaus ist eine gemeinsame Würdigung der endpunktbezogenen Nutzen- und Schadenaspekte möglich (siehe Abschnitt 3.1.4), sodass beispielsweise ein endpunktbezogener „geringerer Schaden“ (im Sinne einer Verringerung von Nebenwirkungen) bei Betrachtung der Effekte auf alle anderen Endpunkte in die abwägende Feststellung eines „Zusatznutzens“ münden kann.

3.1.2 Surrogate patientenrelevanter Endpunkte

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 [15,180,419]. Die meisten Surrogatendpunkte sind jedoch in dieser Hinsicht nicht verlässlich und können bei der Nutzenbewertung irreführend sein [207,215]. Daher werden Surrogatendpunkte im Rahmen der Nutzenbewertung des Institutes in der Regel nur dann in Betracht gezogen, wenn sie zuvor anhand geeigneter statistischer Methoden innerhalb einer hinreichend eingegrenzten Patientenpopulation und innerhalb von vergleichbaren Interventionen (z. B. Arzneimittel mit vergleichbarem Wirkmechanismus) validiert wurden. Ein Surrogatendpunkt kann dann als valide gelten, wenn der Effekt auf den zu ersetzenden patientenrelevanten Endpunkt durch den Effekt auf den Surrogatendpunkt in einem ausreichenden Ausmaß erklärt wird [27,541].

Die Notwendigkeit für die Betrachtung von Surrogatendpunkten kann im Rahmen der frühen Nutzenbewertung von Arzneimitteln (siehe Abschnitt 3.3.3) eine besondere Bedeutung haben, da in den Zulassungsverfahren primär die Wirksamkeit, aber nicht immer der patientenrelevante Nutzen oder Zusatznutzen untersucht wird.

Für die Validierung eines Surrogatendpunktes gibt es weder ein Standardverfahren noch eine allgemein beste Schätzmethode noch ein allgemein akzeptiertes Kriterium, dessen Erfüllung den Nachweis der Validität bedeuten würde [356]. Allerdings widmet sich die aktuelle methodische Literatur sehr häufig korrelationsbasierten Verfahren zur Surrogatvalidierung mit Schätzung von Korrelationsmaßen auf Studien- und individueller Ebene [268]. In der Nutzenbewertung des Institutes werden daher bevorzugt Validierungen auf Basis eines solchen Verfahrens berücksichtigt. Diese Verfahren bedürfen in der Regel einer Meta-Analyse von mehreren randomisierten Studien, in denen sowohl Effekte auf den Surrogatendpunkt als auch Effekte auf den interessierenden patientenrelevanten Endpunkt untersucht wurden [77,374]. Alternative Methoden [541] werden nur in begründeten Ausnahmefällen in Betracht gezogen.

Bei korrelationsbasierten Verfahren wird für den Nachweis der Validität in der Regel einerseits auf individueller Ebene eine hohe Korrelation zwischen Surrogat und patientenrelevantem Endpunkt sowie andererseits auf Studienebene eine hohe Korrelation zwischen den Effekten für das Surrogat und den patientenrelevanten Endpunkt verlangt [77,78]. Da in der Nutzenbewertung des Institutes Aussagen bezogen auf Gruppen von Patientinnen und Patienten gemacht werden, stützt sich die Bewertung der Validität eines Surrogatendpunktes hier primär auf den Grad der Korrelation auf der Ebene der Behandlungseffekte. Zusätzlich zur Höhe der Korrelation wird für die Bewertung der Validität eines Surrogatendpunktes die Aussagesicherheit der Ergebnisse zur Validierung berücksichtigt. Dazu werden verschiedene Kriterien herangezogen [268]. Zum Beispiel sind beobachtete Zusammenhänge zwischen einem Surrogat- und dem entsprechenden patientenrelevanten Endpunkt für eine Intervention mit einem bestimmten Wirkmechanismus nicht zwangsläufig auf Interventionen zur Behandlung derselben Erkrankung, aber mit einem anderen Wirkmechanismus anwendbar [179,207,215,356]. Die Studien, auf denen die Validierung basiert, müssen daher an Patientenkollektiven und Interventionen durchgeführt worden sein, die Aussagen über das der Nutzenbewertung zugrunde liegende Anwendungsgebiet und die zu bewertende Intervention sowie die Vergleichsintervention erlauben. Zur Beurteilung der Übertragbarkeit sollten bei Validierungsstudien, die verschiedene Krankheitsentitäten oder Interventionen einschließen, mindestens geeignete Untersuchungen zur Heterogenität vorliegen.

Im Falle, dass ein Surrogatendpunkt nicht abschließend validiert werden konnte (zum Beispiel durch eine nicht ausreichend hohe Korrelation), besteht darüber hinaus die Möglichkeit der Anwendung des Konzepts eines sogenannten Surrogate Threshold Effekts (STE) [76,268]. Hierzu wird der Effekt auf das Surrogat, der sich aus den Studien zur Nutzenbewertung ergibt, dem STE gegenübergestellt [78,374].

Die Schlussfolgerungen für eine Nutzenbewertung des Institutes bezüglich patientenrelevanter Endpunkte können aus den Effekten auf das Surrogat in Abhängigkeit vom Validitätsnachweis des Surrogats bzw. der Betrachtung eines STE gezogen werden. Entscheidend für Ersteres sind der Grad der Korrelation der Effekte auf das Surrogat und den patientenrelevanten Endpunkt und die Aussagesicherheit der Validierung in den Validierungsstudien. Bei Betrachtung eines STE ist die Größe des Effekts auf das Surrogat in den Studien zur Nutzenbewertung im Vergleich zum STE das entscheidende Kriterium. Je nach Konstellation sind bei einem statistisch signifikanten Effekt auf die Surrogatendpunkte alle Abstufungen der Aussagen zum (Zusatz-)Nutzen hinsichtlich des zugehörigen patientenrelevanten Endpunkts gemäß Abschnitt 3.1.4 möglich.

Surrogatendpunkte, die nicht valide sind oder für die kein adäquates Validierungsverfahren durchgeführt wurde, können dennoch in den Berichten des Institutes dargestellt werden. Derartige Endpunkte sind aber unabhängig von den beobachtbaren Effekten nicht als Belege für einen Nachweis des (Zusatz-)Nutzens einer Intervention geeignet.

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. Weiterhin ist zu beachten, dass ein Endpunkt gleichzeitig ein patientenrelevanter Endpunkt sein kann und darüber hinaus als Surrogat (also als Ersatz) für einen anderen patientenrelevanten Endpunkt betrachtet werden kann.

3.1.3 Ermittlung des Schadens medizinischer Interventionen

Die Anwendung einer medizinischen Intervention gleich welcher Art (medikamentös, nichtmedikamentös, chirurgisch, diagnostisch, präventiv etc.) birgt per se das Risiko unerwünschter Wirkungen. „Unerwünschte Wirkungen“ bezeichnet dabei all 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 [92].

Der Begriff „Schaden“ beschreibt das Auftreten unerwünschter Wirkungen bei der Anwendung einer medizinischen Intervention. Die Darstellung des Schadens ist wesentlicher und gleichberechtigter Bestandteil der Nutzenbewertung einer Intervention. Sie gewährleistet eine informierte populationsbezogene, aber auch individuelle Nutzen-Schaden-Abwägung [557]. 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 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 [92]. Typischerweise sind Studien so konzipiert, dass sie den Einfluss einer medizinischen Intervention auf wenige, vorab definierte Endpunkte messen sollen. Zumeist handelt es sich dabei um Endpunkte zur Erfassung der Wirksamkeit, während Nebenwirkungen als unerwünschte Ereignisse begleitend erhoben werden. Die Ergebnisse bezüglich unerwünschter Ereignisse hängen dabei stark von der zugrunde gelegten Methodik der Erfassung ab. So werden z. B. bei expliziten Abfragen definierter unerwünschter Ereignisse i. d. R. höhere Ereigniszahlen ermittelt als bei einer generellen Abfrage [41,279]. Gerade zur Erfassung unerwartet auftretender unerwünschter Ereignisse ist jedoch eine generelle Abfrage des Befindens der Patientinnen und Patienten erforderlich. 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 [48,149,278]. Darüber hinaus ist die Qualität der Berichterstattung unerwünschter Ereignisse in Einzelstudien mangelhaft, was auch zu einer Erweiterung des CONSORT-Statements für randomisierte klinische Studien geführt hat [277]. Schließlich wird die systematische Bewertung unerwünschter Wirkungen einer Intervention auch dadurch erschwert, dass deren Kodierung in Literaturdatenbanken unzureichend ist und die zielgerichtete Suche nach relevanter wissenschaftlicher Literatur deshalb häufig ein unvollständiges Bild liefert [112].

Die oben genannten Hindernisse erschweren häufig die Untersuchung des Schadens. In Fällen, in denen für die Bewertung vollständige Studienberichte zur Verfügung stehen, ist häufig zumindest eine ausreichende Datentransparenz auch für unerwünschte Ereignisse gegeben. Es ist dennoch notwendig, eine sinnvolle Balance zwischen der Vollständigkeit der Aufarbeitung von Schadenaspekten 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 Patientinnen und Patienten als besonders wichtig angesehen werden und die
- von schwerwiegender Morbidität oder gar erhöhter Mortalität begleitet sind oder in der Folge mit einer erheblichen Beeinträchtigung der Lebensqualität einhergehen.

Das Institut beachtet 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 Vorabrecherche 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. Diese Zusammenstellung potenziell relevanter unerwünschter Wirkungen bildet die Grundlage für die Bewertung des Schadens auf Basis der in die Nutzenbewertung eingehenden Studien. Hierbei wird, sofern möglich und inhaltlich sinnvoll, auch auf zusammenfassende Auswertungen (z. B. Gesamtraten schwerwiegender unerwünschter Ereignisse) zurückgegriffen.

3.1.4 Endpunktbezogene Bewertung

Die Nutzenbewertung und die Einschätzung der Stärke der Ergebnis(un)sicherheit orientieren sich an internationalen Standards der evidenzbasierten Medizin, wie sie z. B. von der GRADE-Gruppe erarbeitet werden [23].

Medizinische Interventionen werden im Vergleich zu einer anderen Intervention oder Scheinintervention (z. B. Placebo) oder keiner Intervention 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 eine Aussage zur Beleglage des (Zusatz-)Nutzens und Schadens in 4 Abstufungen bezüglich der jeweiligen Aussagesicherheit getroffen: Es liegt entweder ein Beleg (höchste Aussagesicherheit), ein Hinweis (mittlere Aussagesicherheit), ein Anhaltspunkt (schwächste Aussagesicherheit) oder keine dieser 3 Situationen vor. Der letzte Fall tritt ein, wenn keine Daten vorliegen oder die vorliegenden Daten keine der drei übrigen Aussagen zulassen.

Je nach Fragestellung beziehen sich die Aussagen auf das Vorhandensein oder das Fehlen eines (Zusatz-)Nutzens und Schadens. Die Voraussetzung für Aussagen über das Fehlen eines (Zusatz-)Nutzens bzw. Schadens sind gut begründete Definitionen von Irrelevanzbereichen (siehe Abschnitt 7.3.6).

Ein wichtiges Kriterium zur Ableitung von Aussagen zur Beleglage ist die Ergebnissicherheit. Grundsätzlich ist jedes Ergebnis einer empirischen Studie oder einer systematischen Übersicht über empirische Studien mit Unsicherheit behaftet und daher auf seine Ergebnissicherheit zu prüfen. Hierbei ist zu unterscheiden zwischen qualitativer und

quantitativer Ergebnissicherheit. Die qualitative Ergebnissicherheit wird beeinträchtigt durch systematische Fehler (Verzerrung, siehe Abschnitt 7.3.11), wie z. B. Informationsfehler, Selektionsfehler und Confounding. Die quantitative Ergebnissicherheit wird beeinflusst durch zufällige Fehler verursacht durch die Stichprobenziehung (statistische Unsicherheit).

Die qualitative Ergebnissicherheit wird somit bestimmt durch das Studiendesign, aus dem sich Evidenzgrade ableiten lassen (siehe Abschnitt 7.1.3), sowie durch (endpunktbezogene) Maßnahmen zur weiteren Vermeidung oder Minimierung möglicher Verzerrungen (z. B. verblindete Zielgrößenerhebung, Auswertung auf Basis aller eingeschlossenen Patientinnen und 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 7.1.4).

Die quantitative Ergebnissicherheit steht in unmittelbarem Zusammenhang mit dem Stichprobenumfang, d. h. der Anzahl der in einer Studie untersuchten Patientinnen und Patienten bzw. der Anzahl der in einer systematischen Übersicht enthaltenen (Primär-)Studien, sowie mit der in bzw. zwischen den Studien beobachteten Variabilität. Falls die zugrunde liegenden Daten dies zulassen, lässt sich die statistische Unsicherheit als Standardfehler bzw. Konfidenzintervall von Parameterschätzungen quantifizieren und beurteilen (Präzision der Schätzung).

Das Institut verwendet die folgenden drei Kategorien zur Graduierung des Ausmaßes der qualitativen Ergebnissicherheit auf Einzelstudien- und Endpunktebene:

- **hohe qualitative Ergebnissicherheit:** Ergebnis bezüglich eines Endpunkts einer randomisierten Studie mit niedrigem Verzerrungspotenzial
- **mäßige qualitative Ergebnissicherheit:** Ergebnis bezüglich eines Endpunkts einer randomisierten Studie mit hohem Verzerrungspotenzial
- **geringe qualitative Ergebnissicherheit:** Ergebnis bezüglich eines Endpunkts einer nicht randomisiert vergleichenden Studie

Bei der Ableitung der Beleglage für einen Endpunkt sind die Anzahl der vorhandenen Studien, deren qualitative Ergebnissicherheiten sowie die in den Studien gefundenen Effekte von zentraler Bedeutung. Liegen mindestens 2 Studien vor, wird zunächst unterschieden, ob sich aufgrund der vorhandenen Heterogenität im Rahmen einer Meta-Analyse (siehe Abschnitt 7.3.8) sinnvoll ein gemeinsamer Effektschätzer bilden lässt oder nicht. Im Fall homogener Ergebnisse, die sich sinnvoll poolen lassen, muss der gemeinsame Effektschätzer statistisch signifikant sein, um gemäß der vorhandenen Ergebnissicherheit einen Beleg, einen Hinweis oder einen Anhaltspunkt abzuleiten. Sind die geschätzten Effekte zu heterogen, um sinnvoll einen gepoolten gemeinsamen Effektschätzer zu bilden, wird unterschieden zwischen „nicht gleichgerichteten“, „mäßig gleichgerichteten“ und „deutlich gleichgerichteten“ Effekten, die wie folgt definiert sind.

Falls das Prädiktionsintervall zur Darstellung der Heterogenität in einer Meta-Analyse mit zufälligen Effekten (siehe Abschnitt 7.3.8) dargestellt wird und den Nulleffekt nicht überdeckt, liegen gleichgerichtete Effekte vor. Anderenfalls (keine Darstellung des Prädiktionsintervalls oder dieses überdeckt den Nulleffekt) liegen gleichgerichtete Effekte in folgender Situation vor:

Die Effektschätzer von zwei oder mehr Studien zeigen in eine Richtung. Für diese „gerichteten“ Studien gelten alle folgenden Bedingungen:

- Das Gesamtgewicht dieser Studien ist $\geq 80\%$.
- Mindestens 2 dieser Studien zeigen statistisch signifikante Ergebnisse.
- Mindestens 50 % des Gewichts dieser Studien basieren auf statistisch signifikanten Ergebnissen.

Die Gewichte der Studien kommen hierbei in der Regel aus einer Meta-Analyse mit zufälligen Effekten (siehe Abschnitt 7.3.8). Falls keine Meta-Analyse sinnvoll ist, entspricht die relative Fallzahl dem Gewicht.

Wann gleichgerichtete Effekte mäßig oder deutlich gleichgerichtet sind, wird wenn möglich anhand der Lage des Prädiktionsintervalls entschieden. Da das Prädiktionsintervall in der Regel jedoch nur dargestellt wird, falls mindestens 4 Studien vorliegen (siehe Abschnitt 7.3.8), hängt die Einstufung in mäßig gleichgerichtete und deutlich gleichgerichtete Effekte von der Anzahl der Studien ab.

- 2 Studien: Gleichgerichtete Effekte sind immer deutlich gleichgerichtet.
- 3 Studien:
 - Alle 3 Studien weisen statistisch signifikante Ergebnisse auf: Die gleichgerichteten Effekte sind deutlich gleichgerichtet.
 - Nicht alle 3 Studien weisen statistisch signifikante Ergebnisse auf: Die gleichgerichteten Effekte sind mäßig gleichgerichtet.
- 4 oder mehr Studien:
 - Das Prädiktionsintervall überdeckt nicht den Nulleffekt: Die gleichgerichteten Effekte sind deutlich gleichgerichtet.
 - Das Prädiktionsintervall überdeckt den Nulleffekt: Die gleichgerichteten Effekte sind mäßig gleichgerichtet.

Für den Fall, dass die vorhandenen Studien dieselbe qualitative Ergebnissicherheit aufweisen oder nur eine Studie vorliegt, lassen sich mit diesen Definitionen die regelhaften Anforderungen an die Beleglage zur Ableitung von Aussagen mit unterschiedlichen Aussagesicherheiten definieren. Das Institut unterscheidet – wie oben erläutert – die 3 verschiedenen Aussagesicherheiten „Beleg“, „Hinweis“ und „Anhaltspunkt“.

In der Regel wird an die Aussage eines Belegs die Anforderung zu stellen sein, dass eine Meta-Analyse von Studien mit hoher qualitativer Ergebnissicherheit einen entsprechenden statistisch signifikanten Effekt zeigt. Falls eine Meta-Analyse nicht durchführbar ist, sollten mindestens 2 voneinander unabhängig durchgeführte Studien mit hoher qualitativer Ergebnissicherheit und einem statistisch signifikantem Effekt vorliegen, deren Ergebnis nicht durch weitere vergleichbare ergebnissichere Studien infrage gestellt wird (Konsistenz der Ergebnisse). Bei den 2 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 mäßiger qualitativer Ergebnissicherheit oder eine einzelne Studie mit hoher qualitativer Ergebnissicherheit kann trotz statistisch signifikanten Effekts demnach in der Regel nur einen Hinweis liefern.

Aus lediglich 1 Studie kann im Ausnahmefall für einen Endpunkt ein Beleg für eine spezifische (Teil-)Population abgeleitet werden, wenn ein Studienbericht gemäß ICH-Richtlinien vorliegt, die auch sonst für einen Beleg geforderten Bedingungen erfüllt sind und darüber hinaus die Studie folgende besondere Anforderungen erfüllt:

- Die Studie ist multizentrisch und es gibt mindestens 10 Zentren.
- Der beobachtete Effektschätzer hat einen sehr kleinen zugehörigen p-Wert ($p < 0,001$).
- Das Ergebnis ist innerhalb der Studie konsistent: Für die interessierende (Teil-)Population liegen Auswertungen verschiedener weiterer Teilpopulationen (insbesondere Teilmengen von Studienzentren) vor, die jeweils bewertbare und ausreichend homogene Effektschätzer ergeben. Diese Konsistenzbewertung ist bei binären Daten nur möglich, falls eine gewisse Mindestanzahl an Ereignissen vorliegt.
- Die im vorherigen Punkt adressierten Auswertungen für Teilpopulationen liegen für alle relevanten Endpunkte vor, d. h., diese Auswertungen sind nicht auf einzelne ausgewählte Endpunkte beschränkt.

Es ist möglich, dass sich die Beleglage beim Vorliegen von lediglich 1 Studie, die für sich genommen nur einen Hinweis oder Anhaltspunkt ergibt, durch zusätzliche indirekte Vergleiche ändert. Hierfür sind allerdings hohe methodische Anforderungen an die indirekten Vergleiche zu stellen (siehe Abschnitt 7.3.9). Außerdem ist es bei entsprechend homogener Datenlage möglich, dass sich durch Ergänzung indirekter Vergleiche die Präzision der Effektschätzung erhöht, was bei der Feststellung des Ausmaßes des Zusatznutzens eine wichtige Rolle spielt (siehe Abschnitt 3.3.3).

Eine Meta-Analyse von Studien mit geringer qualitativer Ergebnissicherheit oder eine einzelne Studie mit mäßiger qualitativer Ergebnissicherheit liefert bei statistisch signifikantem Effekt in der Regel nur einen Anhaltspunkt.

Die regelhafte Operationalisierung ist übersichtlich in Tabelle 2 zu finden. In begründeten Fällen beeinflussen weitere Faktoren diese Einschätzungen. Die Betrachtung von Surrogatendpunkten (siehe Abschnitt 3.1.2), das Vorliegen schwerwiegender Designmängel bei einer Studie oder auch begründete Zweifel an der Übertragbarkeit auf die Behandlungssituation in Deutschland können z. B. zu einer Verringerung der Aussagesicherheit führen. Auf der anderen Seite können z. B. große Effekte oder eine eindeutige Richtung eines vorhandenen Verzerrungspotenzials eine Erhöhung der Sicherheit begründen.

Tabelle 2: Regelhaft abgeleitete Aussagesicherheiten für verschiedene Evidenzsituationen beim Vorliegen von Studien derselben qualitativen Ergebnissicherheit

		Anzahl Studien				
		1 (mit statistisch signifikantem Effekt)	≥ 2			
			homogen	heterogen		
			Meta- Analyse statistisch signifikant	gleichgerichtete Effekte ^a		
			deutlich	mäßig	nein	
qualitative Ergebnis- sicherheit	hoch	Hinweis	Beleg	Beleg	Hinweis	–
	mäßig	Anhaltspunkt	Hinweis	Hinweis	Anhaltspunkt	–
	gering	–	Anhaltspunkt	Anhaltspunkt	–	–

a: Zur Erläuterung des Begriffs: siehe Text.

Liegen mehrere Studien mit unterschiedlicher qualitativer Ergebnissicherheit vor, so werden zunächst nur die Studien mit der höherwertigen Ergebnissicherheit betrachtet und auf dieser Grundlage Aussagen zur Beleglage gemäß Tabelle 2 abgeleitet. Bei der Ableitung von Aussagen zur Beleglage für den gesamten Studienpool gelten dann folgende Grundsätze:

- Die Aussagen zur Beleglage bei Beschränkung auf die höherwertigeren Studien werden durch Hinzunahme der übrigen Studien nicht abgeschwächt, sondern allenfalls aufgewertet.
- Die für einen Beleg notwendige Bestätigung (Replikation) eines statistisch signifikanten Ergebnisses einer Studie hoher qualitativer Ergebnissicherheit kann durch eine oder mehrere Ergebnisse mäßiger (jedoch nicht geringer) qualitativer Ergebnissicherheit erbracht werden. Dabei sollte das Gewicht der Studie hoher qualitativer Ergebnissicherheit eine angemessene Größe haben (zwischen 25 und 75 %).
- Ist das meta-analytische Ergebnis für die höherwertigeren Studien nicht statistisch signifikant bzw. liegen für diese Studien keine gleichgerichteten Effekte vor, sind die Aussagen zur Beleglage auf der Grundlage der Ergebnisse des gesamten Studienpools abzuleiten, wobei die Aussagesicherheit durch die minimale qualitative Ergebnissicherheit aller einbezogenen Studien bestimmt wird.

Nach diesen Definitionen und Grundsätzen wird für jeden Endpunkt einzeln eine entsprechende Nutzaussage abgeleitet. Überlegungen zur endpunktübergreifenden Bewertung finden sich im nachfolgenden Abschnitt (siehe Abschnitt 3.1.5).

3.1.5 Zusammenfassende Bewertung

Die im Rahmen der Ableitung von Aussagen zur Beleglage 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) den Nutzen,
- 2) den Schaden und
- 3) ggf. eine Nutzen-Schaden-Abwägung dar,

soweit dies aufgrund der vorliegenden Daten möglich ist. Hierbei werden alters-, geschlechts- und lebenslagenpezifische Besonderheiten berücksichtigt.

Eine Möglichkeit der gleichzeitigen Würdigung von Nutzen und Schaden ist die Gegenüberstellung der endpunktbezogenen Nutzen- und Schadenaspekte. Dabei werden die Effekte auf alle Endpunkte (qualitativ oder semiquantitativ) gegeneinander abgewogen mit dem Ziel, zu einer endpunktübergreifenden Aussage zum Nutzen bzw. Zusatznutzen einer Intervention zu kommen. Eine weitere Möglichkeit der gleichzeitigen Würdigung besteht darin, die verschiedenen patientenrelevanten Endpunkte zu einem einzigen Maß zu aggregieren. In diesem Fall würden die Aussagen des Instituts für jeden einzelnen patientenrelevanten Endpunkt gewichtet z. B. in einen Summenscore einfließen. Die gleichzeitige Würdigung von Nutzen und Schaden wird themenspezifisch konkretisiert und sollte – wenn dies prospektiv möglich ist – im Berichtsplan oder andernfalls im Vorbericht beschrieben werden. Eine quantitative Gewichtung unter Verwendung von Summenscores oder Indizes sollte prospektiv zum Zeitpunkt der Auswahl der zu untersuchenden Endpunkte erfolgen.

Häufig werden sogenannte Nutzwerte für Gesundheitszustände erhoben, die die von den Befragten positiv wie negativ empfundenen Aspekte in einer Indexzahl ausdrücken sollen. Unter Integration der Dauer der entsprechenden Gesundheitszustände können diese Nutzwerte z. B. in sogenannte qualitätsadjustierte Lebensjahre (QALYs = Quality-Adjusted Life Years) überführt werden. Die Erhebung und Berechnung von Nutzwerten wird z. B. dargestellt in: [141,354,464]. In Kosten-Nutzen-Bewertungen kann das Institut auf QALYs als Gesamtmaß des Nutzens zurückgreifen [267].

Die wissenschaftliche Debatte um die ethischen und methodischen Probleme des QALY-Konzepts an sich und deren Lösung oder einer daran geknüpften Zahlungsbereitschafts-

schwelle in Kosten-Nutzen-Bewertungen sowie des Einsatzes des QALY zur reinen Abwägung von Nutzen und Schaden soll hier nicht erneut aufgerollt werden. In diesem Zusammenhang sei auf folgende ausgewählte Publikationen verwiesen: [121,137,195,232,343,364,394,406,420,530].

Neben dem Nutzwert QALY können auch Verfahren der multikriteriellen Entscheidungsfindung oder der Präferenzhebung wie der Analytic Hierarchy Process (AHP) und die Conjoint-Analyse (CA) eingesetzt werden, wenn ein Maß des Gesamtnutzens zum Vergleich von Interventionen erhoben werden soll.

Für das AHP [135,136] wird ein Entscheidungsproblem in sog. Kriterien zerlegt. Diese werden dann in eine Hierarchie gebracht. So kann z. B. ein Arzneimittel nach den Kriterien Wirkung, Nebenwirkung und Lebensqualität beurteilt werden. Das Kriterium Wirkung kann in weitere Subkriterien zerlegt werden, die Endpunkten entsprechen können [260]. Die am AHP Teilnehmenden werden dann jeweils binär zu den Kriterien befragt, d. h., sie müssen auf einer vorgegebenen Skala wählen, wie viel mehr ihnen ein Kriterium als ein anderes Kriterium bedeutet. Mittels eines Verfahrens der Matrizenmultiplikation [440-442] können über den sog. rechten Eigenvektor die Gewichte für die Kriterien bzw. Subkriterien ermittelt werden, die sich zu 1 aufsummieren müssen.

Die CA gehört zur Gruppe der sogenannten Stated-Preference-Techniken [59]. Eine Entscheidung wird ebenfalls in sogenannte Attribute zerlegt. Die Befragten werden anschließend mit einem Set von (theoretischen) Szenarien konfrontiert, die jeweils für alle Attribute eine Effektgröße gegenüberstellen. Aus der Wahl der Szenarien wird dann über ein Regressionsmodell ein Gewichtungsfaktor für jedes Attribut errechnet. Diese Gewichte können wiederum auf 1 normiert werden.

3.2 Spezielle Aspekte der Nutzenbewertung

3.2.1 Auswirkung nicht publizierter Studienergebnisse auf das Fazit

Eine wesentliche Voraussetzung für die Validität einer Nutzenbewertung ist die vollständige Verfügbarkeit der Ergebnisse der durchgeführten Studien. Eine Bewertung, die auf unvollständigen, ggf. sogar selektiv zusammengestellten Daten beruht, kann ein verzerrtes Ergebnis liefern [165,271] (siehe auch Abschnitt 7.3.11).

Die Verzerrung der publizierten Evidenz durch Publikationsbias und Outcome Reporting Bias wurde umfangreich beschrieben [142,363,486]. Um die Auswirkungen dieser Verzerrung zu minimieren, hat das Institut die Informationsbeschaffung über die bibliografische Recherche hinaus erweitert, z. B. durch eine Recherche in Studienregistern. Darüber hinaus kontaktiert das Institut i. d. R. am Beginn einer Bewertung die Hersteller der zu bewertenden Arzneimittel bzw. Medizinprodukte und bittet um Übermittlung vollständiger Informationen zu Studien mit den zu bewertenden Interventionen (siehe auch Abschnitt 6.1.5).

Die Informationsübermittlung durch Hersteller kann das Problem der Verzerrung durch unpublizierte Evidenz nur lösen, wenn sie selbst nicht selektiv, sondern vollständig ist. Eine unvollständige Informationsübermittlung birgt das Risiko, dass das Ergebnis der Nutzenbewertung verzerrt ist. Dieses Risiko soll im Fazit der Nutzenbewertung durch das Institut berücksichtigt werden.

Die folgende Tabelle 3 beschreibt, in welchen Konstellationen ein Risiko für eine Verzerrung von Bewertungsergebnissen besteht und welche Konsequenzen sich daraus für das Fazit der Nutzenbewertung ergeben.

War die Datenübermittlung vollständig und liegt kein Nachweis vor, dass relevante Mengen von Daten fehlen (Szenario 1), erscheint eine Verzerrung unwahrscheinlich. Die Schlussfolgerungen aus der Bewertung der Daten können deshalb ohne Einschränkung in das Fazit übernommen werden.

Tabelle 3: Szenarien der Datenübermittlung und Konsequenzen für das Fazit einer Nutzenbewertung

Szenario	Datenübermittlung durch Dritte (z. B. Daten von Herstellern)	Nachweis, dass relevante Mengen von Daten fehlen	Verzerrung	Bewertung / Einfluss auf das Fazit
1	Vollständig	Nein	unwahrscheinlich	keine Einschränkung des Fazits der Nutzenbewertung
2	Unvollständig	Nein	möglich	das Fazit wird unter Vorbehalt gestellt
3	Unvollständig	Ja	wahrscheinlich	Charakterisierung der vorhandenen und fehlenden Daten; kein Beleg (und kein Hinweis oder Anhaltspunkt) für einen Nutzen bzw. Schaden
4	Vollständig	Ja (z. B. anderer Hersteller, Investigator Initiated Trials)	möglich	das Fazit wird unter Vorbehalt gestellt

War die Datenübermittlung unvollständig, hängt die Konsequenz für das Fazit davon ab, ob durch andere Rechenschritte nachgewiesen werden kann, dass relevante Mengen von Daten fehlen. Liegt dieser Nachweis nicht vor (Szenario 2), ist eine Verzerrung dennoch möglich, da eine selektive Datenübermittlung gegeben sein kann und weitere relevante unpublizierte Daten existieren können, die durch die Rechenschritte nicht zugänglich sind. In solchen Fällen wird das Fazit deshalb unter einen Vorbehalt gestellt. Wurde der Nachweis erbracht, dass relevante Mengen von Daten fehlen (Szenario 3), ist davon auszugehen, dass

die Datenübermittlung selektiv erfolgte. Eine weitere Auswertung der verfügbaren limitierten Daten und eine daraus abgeleitete Aussage zum Nutzen oder Schaden sind in dieser Situation wahrscheinlich hochgradig verzerrt und stellen damit keine valide Entscheidungsgrundlage für den Gemeinsamen Bundesausschuss dar. In dieser Situation kann deshalb kein Beleg (und kein Hinweis oder Anhaltspunkt) für einen Nutzen oder Schaden der zu bewertenden Intervention festgestellt werden, unabhängig davon, ob die vorliegenden Daten einen Effekt der Intervention zeigen oder nicht.

Wird bei vollständiger Datenübermittlung durch den Hersteller durch andere Recherchen der Nachweis erbracht, dass eine relevante Menge von Daten aus Studien fehlt, auf die der Hersteller keinen Zugriff hat (Szenario 4), liegt keine selektive Datenübermittlung durch den Hersteller vor. In dieser Situation ist eine Verzerrung durch die fehlenden Daten dennoch möglich. Das Fazit wird deshalb unter Vorbehalt gestellt.

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 Patientinnen und Patienten belegt werden. Wenn z. B. von einer Erkrankung bekannt ist, dass sie nach Diagnosestellung innerhalb kurzer Zeit mit hoher Wahrscheinlichkeit 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 Patientinnen und 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 Patientinnen und Patienten in der zu bewertenden Studie. Dabei sollten auch mögliche Schäden der Maßnahme berücksichtigt werden. Glasziou et al. [202] versuchen, die Einordnung als dramatischen Effekt zu operationalisieren. In einer ersten Näherung wird 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 [202]. Diese Größenordnung dient dem Institut als Orientierung und bedeutet keine starre Grenze. Glasziou et al. [202] haben ihrer Empfehlung Ergebnisse von Simulationsstudien zugrunde gelegt, wonach ein beobachtetes relatives Risiko von 5 bis 10 nicht mehr plausibel allein durch Störgrößeneinflüsse erklärt werden könne. Damit wird deutlich, dass eine entsprechende Grenze auch von den Begleitumständen (u. a. Qualität der Studien, die für das Vorliegen eines dramatischen Effekts bemüht werden) abhängig ist. Diese Abhängigkeit schlägt sich auch in Empfehlungen anderer Arbeitsgruppen (z. B. der GRADE-Gruppe) nieder [321].

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 Nutzen als auch Schaden über die Zeit unterschiedlich verteilt sein 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 bzw. 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. [162]). Da die Nutzenbewertung auch die Aspekte des Schadens 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 [263]. Im Einzelfall kann von diesem Vorgehen begründet abgewichen werden, z. B. falls es unter 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 in Abschnitt 3.1.1 beschriebenen patientenrelevanten Nutzendimensionen können auch patientenberichtete Endpunkte (Patient Reported Outcomes, PROs) 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 [160]. 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 [183]. D. h., dass auch im Falle von PROs einschließlich gesundheitsbezogener Lebensqualität und Behandlungszufriedenheit randomisierte kontrollierte Studien am besten für den Nachweis eines Effekts geeignet sind.

Da Angaben zu PROs 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 auf 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 [555]. Dies ist bei der Interpretation solcher Studien zu beachten (siehe auch Abschnitte 7.1.4 und 7.3.4). Allerdings sind Situationen denkbar, in denen eine Verblindung von Ärztinnen und Ärzten und Patientinnen 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 [183].

3.2.5 Nutzen und Schaden in kleinen Populationen

Es gibt kein überzeugendes Argument dafür, bei kleinen Populationen (z. B. Patientinnen und Patienten mit seltenen Erkrankungen oder Untergruppen häufiger Erkrankungen) grundsätzlich von der Hierarchie der Evidenzgrade abzuweichen. In diesem Zusammenhang ist problematisch, dass keine international einheitliche Definition dessen existiert, was unter einer „seltenen“ Erkrankung zu verstehen ist [552]. Davon unberührt, haben auch Patientinnen und Patienten mit seltenen Erkrankungen einen Anspruch auf eine möglichst sichere Information über Behandlungsoptionen [157]. 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 Patientinnen und 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 [159], 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 Surrogatendpunkte müssen allerdings auch bei kleinen Stichproben valide sein [161].

Für den Fall extrem seltener Erkrankungen oder extrem spezifischer Krankheitskonstellationen kann die Forderung nach (parallel) vergleichenden Studien unangemessen sein [552]. Nichtsdestotrotz sind auch hier zumindest adäquate Dokumentationen des Verlaufs solcher Patientinnen und Patienten einschließlich des erwarteten Verlaufs ohne die ggf. zu prüfende Intervention (z. B. anhand historischer Patientenverläufe) möglich und bewertbar [72]. Die Spezifizierung und explizite Kennzeichnung, dass es sich um eine solche Situation einer extrem seltenen Erkrankung oder extrem spezifischen Krankheitskonstellation handelt, erfolgen im Berichtsplan.

3.3 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 Patientinnen und 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 Patientinnen und Patienten, die gemäß der 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 Patientinnen und 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 patientenrelevanter 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 Patientinnen und 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 z. B. dann angesehen, wenn die behandelte Altersgruppe außerhalb der Zulassung liegt, nicht zugelassene Kombinationen mit einem anderen Wirkstoff eingesetzt werden, Patientinnen und Patienten mit einem außerhalb der Zulassung liegenden Schweregrad der Erkrankung behandelt werden oder Studien bei Patientinnen und 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 Patientinnen und 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 Patientinnen und 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 Patientinnen und 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 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 [214]. 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 [322].

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 [338] (dramatischer Effekt, siehe auch Abschnitt 3.3.2). 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 [311]).

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 7.1.4), 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 weitere Einflussgrößen der Ergebnissicherheit finden sich in den Abschnitten 3.1.4 und 7.1.

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 von der Qualität der Studie zunächst unabhängig ist.

3.3.3 Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V

Eine Nutzenbewertung eines Arzneimittels gemäß § 35a SGB V basiert auf einem Dossier des pharmazeutischen Unternehmers. In diesem Dossier legt der pharmazeutische Unternehmer folgende Angaben vor:

- 1) zugelassene Anwendungsgebiete
- 2) medizinischer Nutzen
- 3) medizinischer Zusatznutzen im Verhältnis zur zweckmäßigen Vergleichstherapie
- 4) Anzahl der Patientinnen und Patienten und Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht
- 5) Kosten der Therapie für die gesetzliche Krankenversicherung
- 6) Anforderung an eine qualitätsgesicherte Anwendung

Die Anforderungen an die Form und den Inhalt des Dossiers sind in Dossievorlagen beschrieben, die Bestandteil der Verfahrensordnung des G-BA sind [198]. Im Dossier ist vom pharmazeutischen Unternehmer unter Angabe der Aussagekraft der Nachweise darzulegen, mit welcher Wahrscheinlichkeit und in welchem Ausmaß ein Zusatznutzen des zu bewertenden Arzneimittels im Vergleich zur zweckmäßigen Vergleichstherapie vorliegt. Diese Angaben sollen sowohl bezogen auf die Anzahl der Patientinnen und Patienten als auch bezogen auf die Größe des Zusatznutzens gemacht werden. Die Kosten sind für das zu bewertende Arzneimittel und die zweckmäßige Vergleichstherapie anzugeben (gemessen am Apothekenabgabepreis und unter Berücksichtigung der Fach- und Gebrauchsinformation).

Die Wahrscheinlichkeit des Zusatznutzens beschreibt die Aussagesicherheit zum Zusatznutzen. Das Ausmaß des Zusatznutzens soll im Dossier gemäß den Kategorien der Arzneimittel-Nutzenbewertungsverordnung beschrieben werden (erheblicher, beträchtlicher, geringer, nicht quantifizierbarer Zusatznutzen, kein Zusatznutzen belegt, Nutzen des zu bewertenden Arzneimittels geringer als Nutzen der zweckmäßigen Vergleichstherapie) [70].

Mit der Nutzenbewertung werden die Validität und die Vollständigkeit der Angaben im Dossier geprüft. Dabei wird auch geprüft, ob die vom pharmazeutischen Unternehmer gewählte Vergleichstherapie als zweckmäßig im Sinne des § 35a SGB V und der Arzneimittel-Nutzenbewertungsverordnung gelten kann. Darüber hinaus bewertet das Institut die in den vorgelegten Unterlagen beschriebenen Effekte unter Berücksichtigung ihrer Ergebnissicherheit. In dieser Bewertung werden die qualitative und die quantitative Ergebnissicherheit der vorgelegten Nachweise sowie die Größe der beobachteten Effekte und deren Konsistenz gewürdigt. Die Nutzenbewertung erfolgt auf Basis der im vorliegenden Methodenpapier beschriebenen Standards der evidenzbasierten Medizin, die Bewertung der Kosten auf Basis der Standards der Gesundheitsökonomie. Als Ergebnis der Bewertung legt das Institut eigene Schlussfolgerungen vor, die die Schlussfolgerungen des pharmazeutischen Unternehmers bestätigen oder begründet von diesen abweichen können.

Die Operationalisierung der Feststellung des Ausmaßes des Zusatznutzens umfasst 3 Schritte:

- 1) Im 1. Schritt gilt es, für jeden Endpunkt separat die Wahrscheinlichkeit für das Vorliegen eines Effekts zu prüfen (qualitative Aussage). Dazu werden die Kriterien zur Ableitung von Aussagen zur Beleglage (siehe Abschnitt 3.1.4) angewendet. Je nach Güte der Evidenz wird die Wahrscheinlichkeit demnach als Anhaltspunkt, Hinweis oder Beleg eingestuft.
- 2) Im 2. Schritt ist für die Endpunkte, für die im ersten Schritt zumindest ein Anhaltspunkt für das Vorliegen eines Effekts attestiert wurde, jeweils separat das Ausmaß der Effektstärke festzustellen (quantitative Aussage). Folgende quantitative Aussagen sind möglich: erheblich, beträchtlich, gering, nicht quantifizierbar.
- 3) Im 3. und letzten Schritt gilt es, anhand aller Endpunkte unter Würdigung der Wahrscheinlichkeit und des Ausmaßes auf Endpunktebene im Rahmen einer Gesamtschau

die Gesamtaussage zum Zusatznutzen entsprechend den 6 vorgegebenen Kategorien festzustellen: erheblicher, beträchtlicher, geringer, nicht quantifizierbarer Zusatznutzen, kein Zusatznutzen belegt, Nutzen des zu bewertenden Arzneimittels geringer als Nutzen der zweckmäßigen Vergleichstherapie.

Zur Feststellung des Ausmaßes auf Endpunktebene im 2. Schritt sind die Qualität der Zielgröße sowie die Effektstärke maßgeblich. Die Rationale für diese Operationalisierung findet sich im Anhang „Rationale der Methodik zur Feststellung des Ausmaßes des Zusatznutzens“. Das grundsätzliche Konzept sieht vor, für relative Effektmaße Schwellenwerte für Konfidenzintervalle in Abhängigkeit von anzustrebenden Effekten abzuleiten, die wiederum von der Qualität der Zielgrößen und den Ausmaßkategorien abhängen.

Das Ausmaß auf Endpunktebene wird nicht in jedem Fall zu quantifizieren sein. Falls z. B. ein statistisch signifikanter Effekt für ein ausreichend valides Surrogat vorliegt, eine verlässliche Schätzung für den jeweiligen Effekt der patientenrelevanten Zielgröße jedoch nicht möglich ist, lässt sich der (patientenrelevante) Effekt nicht quantifizieren. In solchen und ähnlichen Situationen wird mit entsprechender Begründung ein Effekt nicht quantifizierbaren Ausmaßes attestiert.

Vom Fall eines quantifizierbaren Effekts ausgehend, richtet sich das weitere Vorgehen nach der Skala der Zielgröße. Es werden folgende Skalen unterschieden:

- binär (Analysen von Vierfeldertafeln)
- Zeit bis Ereignis (Überlebenszeitanalysen)
- stetig oder quasi-stetig mit jeweils vorliegenden Responderanalysen (Analysen von Mittelwerten und Standardabweichungen)
- sonstige (z. B. Analysen von nominalen Daten)

Im Folgenden wird zunächst das Verfahren für binäre Zielgrößen beschrieben. Die übrigen Skalen werden nachfolgend auf dieses Verfahren zurückgeführt.

Vom Effektmaß relatives Risiko ausgehend werden Zähler und Nenner immer so gewählt, dass sich der Effekt (sofern vorhanden) als Wert < 1 realisiert. D. h., ein Effekt ist umso stärker, je niedriger der Wert ist.

A) Binäre Zielgrößen

Zur Feststellung des Ausmaßes des Effekts bei binären Zielgrößen wird das zweiseitige 95 %-Konfidenzintervall für das relative Risiko – ggf. selbst berechnet – herangezogen. Falls mehrere Studien quantitativ zusammengefasst wurden, findet das meta-analytische Ergebnis für das relative Risiko Anwendung.

Je nach Qualität der Zielgröße muss das Konfidenzintervall vollständig unterhalb eines bestimmten Schwellenwertes liegen, um das Ausmaß als gering, beträchtlich oder erheblich

anzusehen. Entscheidend ist also, dass die obere Grenze des Konfidenzintervalls kleiner als der jeweilige Schwellenwert ist.

Es werden folgende drei Kategorien für die Qualität der Zielgröße gebildet:

- Gesamtmortalität
- schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen sowie gesundheitsbezogene Lebensqualität
- nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen

Die Schwellenwerte sind für jede Kategorie separat festgelegt und umso größer (im Sinne näher an 1), je schwerwiegender das Ereignis ist. Die Schwellenwerte sind umso kleiner (im Sinne weiter entfernt von 1), je höher das Ausmaß ist. Die folgende Tabelle 4 gibt die für die 3 Ausmaßkategorien (gering, beträchtlich, erheblich) zu unterschreitenden Schwellenwerte für jede der 3 Kategorien der Qualität der Zielgrößen wieder.

Tabelle 4: Schwellenwerte zur Feststellung des Ausmaßes eines Effekts

		Zielgrößenkategorie		
		Gesamt-mortalität	Schwerwiegende (bzw. schwere) Symptome (bzw. Folgekompli-kationen) und Nebenwirkungen und gesundheitsbezogene Lebensqualität ^a	Nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen
Ausmaßkategorie	Erheblich	0,85	0,75 und Risiko $\geq 5\%$ ^b	Nicht besetzt
	Beträchtlich	0,95	0,90	0,80
	Gering	1,00	1,00	0,90
a: Voraussetzung ist wie für alle patientenberichtete Endpunkte die Verwendung eines validierten bzw. etablierten Instruments sowie eines validierten bzw. etablierten Responsekriteriums. b: Risiko muss für mindestens eine der beiden zu vergleichenden Gruppen mindestens 5 % betragen.				

Das relative Risiko kann generell auf zwei Arten berechnet werden, je nachdem ob sich das Risiko auf Ereignisse oder Gegenereignisse bezieht (z. B. Tod vs. Überleben, Response vs. Non-Response). Für die vorgelagerte Signifikanzaussage ist dies unerheblich, da der p-Wert diesbezüglich bei einer Einzelstudie invariant ist und bei einer Meta-Analyse eine untergeordnete Rolle spielt. Das gilt jedoch nicht für den Abstand der Konfidenzintervallgrenzen zum Nulleffekt. Daher muss zur Festlegung des Ausmaßes des Effekts für jede binäre Zielgröße anhand inhaltlicher Kriterien unter Berücksichtigung der Art des Endpunkts und der zugrunde liegenden Erkrankung entschieden werden, welches Risiko betrachtet wird – das für das Ereignis oder das für das Gegenereignis.

B) Zeit bis Ereignis

Zur Feststellung des Ausmaßes des Effekts bei Zielgrößen „Zeit bis zu einem Ereignis“ wird das zweiseitige 95 %-Konfidenzintervall für das Hazard Ratio benötigt. Falls mehrere Studien quantitativ zusammengefasst wurden, wird das meta-analytische Ergebnis für das Hazard Ratio herangezogen. Liegt das Konfidenzintervall für das Hazard Ratio nicht vor, wird es anhand der zur Verfügung stehenden Angaben approximiert, sofern möglich [512]. Für die Ausmaßfeststellung werden dieselben Grenzen wie für das relative Risiko angelegt (siehe Tabelle 4).

Liegt kein Hazard Ratio vor und ist dies auch nicht berechenbar oder das vorliegende Hazard Ratio ist nicht sinnvoll interpretierbar (z. B. wegen wesentlicher Verletzung der Proportional-Hazard-Annahme), ist zu eruieren, ob sich aus den Angaben ein relatives Risiko (bezogen auf einen sinnvollen Zeitpunkt) berechnen lässt. Auch bei transienten (vorübergehenden) Ereignissen, für die als Zielgröße „Zeit bis zum Ereignis“ gewählt wurde, ist zu eruieren, ob diese Operationalisierung adäquat ist. Ggf. ist auch hier die Berechnung eines relativen Risikos zu einem Zeitpunkt angezeigt.

C) Stetige oder quasi-stetige Zielgrößen mit jeweils vorliegenden Responderanalysen

Zur Feststellung des Ausmaßes des Zusatznutzens bei stetigen oder quasi-stetigen Zielgrößen werden Responderanalysen herangezogen. Dazu bedarf es eines validierten bzw. etablierten Responsekriteriums bzw. Cut-off-Werts. Anhand der Responderauswertungen (Vierfeldertafeln) werden die relativen Risiken daraus direkt berechnet. Entsprechend Tabelle 4 wird anschließend das Ausmaß des Effekts festgestellt.

D) Sonstige Zielgrößen

Für den Fall sonstiger Zielgrößen, für die auch keine Responderauswertungen mit daraus ableitbaren relativen Risiken vorliegen, ist im Einzelfall zu eruieren, ob relative Risiken approximiert werden können [101], um die entsprechenden Schwellenwerte zur Ausmaßfeststellung anzulegen. Anderenfalls ist das Ausmaß als nicht quantifizierbar festzustellen.

Für den dritten Schritt der Operationalisierung der Gesamtaussage zum Ausmaß des Zusatznutzens bei gemeinsamer Betrachtung aller Endpunkte, ist eine strenge Formalisierung nicht möglich, da für die hierzu zutreffenden Werturteile gegenwärtig keine ausreichende Abstraktion bekannt ist. Das Institut wird im Rahmen seiner Nutzenbewertung die Aussagen zur Wahrscheinlichkeit und zum Ausmaß der Effekte vergleichend gegenüberstellen und einen begründeten Vorschlag für eine Gesamtaussage unterbreiten.

3.4 Nichtmedikamentöse therapeutische Interventionen

Wenngleich die regulatorischen Voraussetzungen für den Marktzugang von Arzneimitteln und nichtmedikamentösen therapeutischen Interventionen unterschiedlich sind, gibt es dennoch keinen Grund, an die Bewertung von Nutzen und Schaden einen bezüglich der Ergebnissicherheit prinzipiell anderen Maßstab anzulegen. So sieht zum Beispiel die

Verfahrensordnung des G-BA [198] die prioritäre Berücksichtigung von RCTs, soweit möglich, unabhängig vom Typ (medikamentös bzw. nichtmedikamentös) der zu bewertenden medizinischen Maßnahme vor. Für Medizinprodukte wird abschwächend im Rahmen der Konformitätsbewertung in der EN-ISO-Norm 14155-2 (Abschnitt 4.7 [122]) 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 [362]. Beispielsweise wird oft die Verblindung des die Intervention ausführenden Personals unmöglich und die der Patientinnen und Patienten nur schwierig oder ebenfalls nicht zu bewerkstelligen sein. Zudem kann davon ausgegangen werden, dass Präferenzen für bestimmte Therapieoptionen von Therapeutinnen und Therapeuten und Patientinnen 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 die Fähigkeiten und Fertigkeiten des Anwenders bzw. der Anwender auf der anderen Seite sowie – im Zusammenhang mit dem Bewertungszeitpunkt – Lerneffekte sind ggf. zu beachten.

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, d. h., gerade im nichtmedikamentösen Bereich fehlen oft aussagekräftige randomisierte Studien (z. B. in der Chirurgie [362]). 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 geringeren Ergebnissicherheit allenfalls Anhaltspunkte auf einen (Zusatz-)Nutzen oder Schaden liefern können. Die Einbeziehung von Studien eines niedrigeren Evidenzgrades steht im Einklang mit der entsprechenden Regelung in der Verfahrensordnung des G-BA [198], 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 Patientinnen und Patienten umso mehr einer Begründung, je weiter von der Evidenzstufe 1 abgewichen wird. Dafür ist der potentielle Nutzen einer Methode insbesondere gegen die Risiken der Anwendung beim Patienten abzuwägen, die mit einem Wirksamkeitsnachweis geringerer Aussagekraft einhergehen“ [198]. D. h., 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, die Verkehrsfähigkeit bzw. CE-Kennzeichnung (gemäß MPG) und den Zulassungsstatus von Arzneimitteln (gemäß AMG) zu beachten, sofern die zu prüfenden Interventionen oder Vergleichsinterventionen den Einsatz von Medizinprodukten oder Arzneimitteln beinhalten (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 [190,307]. Prospektiv geplante, kontrollierte diagnostische Studien der Phase 4 nach Köbberling et al. [307] bzw. Studien der Stufe 5 nach Fryback und Thornbury [190] mit einer (idealerweise zufälligen) Zuteilung von Patientinnen und Patienten zu einer Strategie mit bzw. ohne Anwendung der zu prüfenden diagnostischen Maßnahme oder zu 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 [198]). 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 [182], und Beispiele zeigen, dass sie auch mit vergleichsweise moderatem Aufwand durchführbar sind [16,525].

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. Sofern die Diagnostik selbst risikobehaftet oder komplikationsträchtig ist, kann ein weniger belastendes diagnostisches Verfahren patientenrelevante Vorteile haben, nämlich wenn bei vergleichbarer Testgüte die Testdurchführung selbst weniger Mortalität, Morbidität oder Einschränkungen der gesundheitsbezogenen Lebensqualität verursacht.

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 [455].

Liegen solche Studien nicht oder in nicht ausreichender Quantität und Qualität vor, kann eine Bewertung der sogenannten diagnostischen Kette erfolgen [366]. 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. [307]), 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. Ggf. können allein solche Studien, auch ohne spezielle Bewertung der diagnostischen Güte, für einen Nutznachweis eines diagnostischen Verfahrens herangezogen werden [340], wenn sich daraus mit ausreichender Sicherheit Aussagen zur Wechselwirkung zwischen diagnostischer Information und (zumeist therapeutischer) Konsequenz ableiten lassen.

Bei der Bewertung der Ergebnissicherheit von Studien zur diagnostischen Güte orientiert sich das Institut in erster Linie an den QUADAS-Kriterien [546], die allerdings ggf. projektspezifisch angepasst werden. Die Kriterien des STARD-Statements [51,52] 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 [190] 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 Ärztinnen und 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 [190,225]. Der Stellenwert derartiger, letztlich unkontrollierter Studien im Rahmen von Nutzenbewertungen diagnostischer Verfahren muss als weitgehend unklar angesehen werden. Informationen zu Managementänderungen allein können deshalb nicht für einen Nutznachweis herangezogen werden, solange keine Informationen über die patientenrelevanten Auswirkungen solcher Änderungen vorliegen.

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 Patientinnen und Patienten oder mit zufälliger Zuordnung der Tests auf verschiedene Patientinnen und 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 [50]. 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 [182,346].

Auch bei der Bewertung von Diagnostika kann es notwendig sein, das Ergebnis des Konformitätsbewertungsverfahrens zur CE-Kennzeichnung und den Zulassungsstatus zur Diagnostik eingesetzter Arzneimittel zu beachten (siehe Abschnitt 3.3.1). Entsprechende Konsequenzen sind dann im Berichtsplan (siehe Abschnitt 2.1.1) zu spezifizieren.

3.6 Früherkennung und Screening

Screeningprogramme setzen sich aus unterschiedlichen „Bausteinen“ zusammen, die entweder als Ganzes oder in Teilen untersucht werden (können) [104,479]. Bei der Bewertung erfolgt eine Orientierung an international akzeptierten Standards bzw. Kriterien, wie zum Beispiel denjenigen des UK National Screening Committee (UK NSC [521]), der US Preventive Services Task Force (US PSTF [233,410,458]) oder des National Health Committee (NHC) aus Neuseeland [384].

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 zu einer Strategie mit bzw. ohne Anwendung der Screeningmaßnahme (ggf. auch zu unterschiedlichen 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 ergebnissicheren Studien ermittelt wurden (i. d. R. Phase 3 nach Köbberling et al. [307]), 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.

3.7 Prävention

Prävention zielt darauf ab, eine gesundheitliche Schädigung zu verhindern, weniger wahrscheinlich zu machen oder zu verzögern [537]. 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). 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 [238]. 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 zu einer 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 [513].

In Einzelfällen ist zu prüfen, inwieweit die Berücksichtigung weiterer Studiendesigns sinnvoll sein kann [283]. So werden beispielsweise massenmediale Kampagnen häufig im Rahmen eines sogenannten Interrupted Time Series Designs evaluiert (z. B. bei [529]) 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 [90].

Für den Nutzen auf Populationsebene ist neben der Effektivität eines solchen Programms auch die Höhe der Teilnehmerate entscheidend. Außerdem ist die Frage relevant, welche Personen erreicht werden; so gibt es Hinweise darauf, dass Präventionsprogramme gerade von Bevölkerungsgruppen mit erhöhtem Krankheitsrisiko weniger in Anspruch genommen werden [323]. Beide Aspekte stehen deshalb im besonderen Fokus von Bewertungen des Instituts.

3.8 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 die Patientin und 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 [11,236,481]. Richtlinien zur Planung und Durchführung von Prognosestudien in der Onkologie werden von Simon und Altman [481] beschrieben. Ein allgemeiner Rahmen zur Qualitätsbewertung von Prognosestudien wird von Laupacis et al. [328] vorgeschlagen. Hayden et al. [236] entwickeln Richtlinien zur Qualitätsbewertung von Prognosestudien in Bezug auf potenzielle Biasquellen. Eine gute Übersicht über die Bedeutung prognostischer Studien und die verschiedenen Ansätze der Entwicklung, Validierung und Anwendung prognostischer Modelle für die klinische Praxis gibt eine Serie im *BMJ* [14,376,377,437]. Die Entwicklung und Validierung prognostischer Modelle im Fall von fehlenden Daten bei den Prädiktoren wird von Vergouwe et al. [528] beschrieben. Da systematische Übersichten von Prognosestudien oftmals durch Mängel in der Planung, Auswertung und Berichterstattung dieser Studien limitiert sind, machen Hemingway et al. [241] Vorschläge zur Verbesserung der Prognoseforschung.

Bei der Bewertung von Prognosestudien sind in jedem Fall die folgenden Aspekte zu beachten, die sich aus der zugrunde liegenden Datenquelle sowie der verwendeten Datenanalyse ergeben:

- 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 möglicher Störgrößen und Wechselwirkungen einschließlich der statistischen Handhabung
- klare Beschreibung der statistischen Modellbildung
- bei Kohortenstudien Vollständigkeit des Follow-ups beziehungsweise Maßnahmen zur Erzielung eines möglichst vollständigen Follow-ups; Abschätzung möglicher Selektionseffekte bei unvollständigem Follow-up
- klare Beschreibung des Umgangs mit fehlenden Daten
- bei der Betrachtung prognostischer Scores Differenzierung zwischen Scoreentwicklung und Scorevalidierung, z. B. als Entwicklung des Scores innerhalb einer sogenannten Lernstichprobe und Validierung an einer Teststichprobe

Da bei Prognosestudien häufig multifaktorielle Regressionsmodelle eine zentrale Rolle spielen [377], ist außerdem Abschnitt 7.3.7 zu beachten. Typische Studiendesigns für die Evaluierung prognostischer Merkmale im Sinne von Risikofaktoren stellen Kohortenstudien und ggf. auch Fall-Kontroll-Studien dar [377]. 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 7.1.4 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 [11,377,481]. Die methodische Qualität von Studien bzw. derer Publikationen zu prognostischen Merkmalen ist häufig unzureichend [241,415], 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 [11]. Einige wichtige Probleme bei Meta-Analysen von Prognosestudien lassen sich vermeiden, wenn individuelle Patientendaten zur Verfügung stehen [11].

Neben Ergebnissen aus Studien, die einzelne oder (zumeist) mehrere prognostische Merkmale untersuchen, werden für die individuelle Risikoabschätzung von Patientinnen und Patienten (oder allgemein 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 [536] oder Joint British Chart [60]). Die Grundlage dieser Risikotabellen bilden meist multifaktorielle Regressionsmodelle, deren Ergebnisse zur einfacheren Handhabung in Tabellen oder Punktesysteme überführt werden [498]. 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).

4 Leitlinien und Versorgungsanalyse

4.1 Hintergrund

Leitlinien sind systematisch entwickelte Entscheidungshilfen für Leistungserbringer und Patientinnen und Patienten zur angemessenen Vorgehensweise bei speziellen Gesundheitsproblemen. Sie haben zum Ziel, die Patientenversorgung zu verbessern. Ihren Empfehlungen liegen eine systematische Überprüfung der Evidenz und eine Bewertung des Nutzens und Schadens der alternativen Behandlungsoptionen zugrunde. [177,208] Leitlinien können normativ Standards in allen Bereichen der Versorgungskette beschreiben, sei es Diagnostik, Behandlung, Rehabilitation oder Nachsorge. Diese Versorgungsstandards beinhalten wesentliche Informationen über die in einem Gesundheitssystem angestrebte Versorgungsqualität. Die Bestimmung eines Versorgungsstandards ist eine zentrale Voraussetzung, um Aussagen über die Versorgungsqualität in einem Gesundheitssystem treffen zu können.

Die Identifikation und Beschreibung von Versorgungsstandards auf Basis hochwertiger Leitlinien dient als Grundlage für verschiedene wissenschaftliche Analysen, z. B. als Ausgangspunkt für die Entwicklung oder Aktualisierung von Disease-Management-Programmen (DMPs) (siehe Abschnitt 4.3). Ebenso können durch den Abgleich dieser Standards mit konkreten Versorgungsstrukturen, -prozessen und -ergebnissen Versorgungslücken und Verbesserungspotenziale aufgedeckt werden (siehe Abschnitt 4.4). Dies wird im Folgenden als Versorgungsanalyse bezeichnet. Die Versorgungsanalyse ermöglicht so Aussagen zu Fragen der Qualität und Wirtschaftlichkeit der im Rahmen der gesetzlichen Krankenversicherung erbrachten Leistungen (vgl. § 139a Abs. 3, Nr. 2 SGB V).

Der Fokus liegt auf der überblicksartigen Darstellung gesamter Krankheitsbilder. Darüber hinaus können auch einzelne Verfahren bzw. Technologien untersucht werden, z. B. als Grundlage für die weiterführende Bewertung in systematischen Übersichten.

Ziel ist es, für Entscheidungsträger und Akteure im Gesundheitssystem aktuelle Versorgungsstandards darzustellen oder deren Fehlen zu dokumentieren. Weiterhin sollen die Versorgungsstandards je nach Fragestellung mit der konkreten Versorgungssituation abgeglichen werden, um so fundierte Entscheidungen zur Verbesserung der Versorgungsqualität im Gesundheitssystem zu ermöglichen.

4.2 Identifikation von Versorgungsstandards über Leitlinien

4.2.1 Versorgungsstandards in Leitlinien

Ärztliches Handeln, das nach medizinisch-wissenschaftlicher Erkenntnis und / oder ärztlicher Erfahrung in der Profession akzeptiert ist, stellt den medizinischen Standard dar [234]. Eine Form, einen medizinischen Standard wissenschaftlich und institutionell festzusetzen, ist die Leitlinie.

Zur Beantwortung von Fragestellungen zum Versorgungsstandard werden in den Berichten des Ressorts in der Regel evidenzbasierte Leitlinien herangezogen. Unter evidenzbasierten Leitlinien werden Leitlinien verstanden, deren Empfehlungen auf einer systematischen Literaturrecherche beruhen, deren Empfehlungen grundsätzlich mit einer Evidenz- und / oder Empfehlungseinstufung (Level of Evidence [LoE] und / oder Grade of Recommendation [GoR]) und deren Empfehlungen mit den Referenzen der ihnen zugrunde liegenden Primär- und / oder Sekundärliteratur verknüpft sind (modifiziert nach AGREE [4]). Die Evidenzbasierung einer Leitlinie setzt nicht voraus, dass jede in ihr enthaltene Einzelempfehlung mit einem hohen Evidenzlevel verknüpft ist. Auch systematisch und transparent erstellte und damit evidenzbasierte Leitlinien enthalten in der Regel Empfehlungen, die auf einer schwachen Evidenzgrundlage beruhen [515].

4.2.2 Methodische Bewertung von Leitlinien

Die Informationsbeschaffung erfolgt entsprechend den in Kapitel 6 beschriebenen Vorgehensweisen.

International werden verschiedene Instrumente zur methodischen Leitlinienbewertung eingesetzt [533]. Das von einem Netzwerk von Forschern und Gesundheitspolitikern entwickelte und validierte AGREE-Instrument (Appraisal of Guidelines Research and Evaluation in Europe) [4,352] und dessen Weiterentwicklung in Form des AGREE-II-Instruments [5,65,66] sind international am weitesten verbreitet. Auch das deutschsprachige DELB-Instrument der AWMF und des ÄZQ basiert auf dem Bewertungsinstrument der AGREE Collaboration. Um ggf. einen Vergleich der Ergebnisse der Leitlinienbewertung des Instituts mit den in anderen Studien veröffentlichten Leitlinienbewertungen zu vereinfachen, wird für die methodische Bewertung von Leitlinien im Institut regelhaft das AGREE-Instrument verwendet. An der Weiterentwicklung des DELB-Instrumentes arbeitet das Institut aktiv mit.

Das Institut legt im Rahmen der Erstellung des Berichtsplans vorab fest, ob aufgrund der Fragestellung eine methodische Bewertung der Leitlinien mit dem AGREE-Instrument erfolgen soll. Das AGREE-Instrument [4] enthält 23 Einzelfragen (Items), die in sechs Domänen dargestellt sind und anhand einer Skala bewertet werden. Jede Domäne deckt eine separate Dimension der Leitlinienqualität ab:

- Domäne 1: Geltungsbereich und Zweck
- Domäne 2: Beteiligung von Interessengruppen
- Domäne 3: Methodologische Exaktheit der Leitlinienentwicklung
- Domäne 4: Klarheit der Präsentation
- Domäne 5: Anwendbarkeit
- Domäne 6: Redaktionelle Unabhängigkeit

Jede Leitlinienbewertung wird durch zwei Reviewer unabhängig voneinander durchgeführt.

A) Standardisierte Domänenwerte

Die Domänen sind voneinander unabhängig. Daher werden für jede Leitlinie Summenwerte für die einzelnen Domänen separat berechnet. Zur besseren Vergleichbarkeit der Domänen untereinander erfolgt, wie im AGREE-Instrument vorgegeben, eine Standardisierung der errechneten Domänenwerte. Diese werden in den Berichten tabellarisch dargestellt.

Das Instrument lässt keine Festsetzung von Schwellenwerten für die Beurteilung der Domänen zu. Jedoch können die einzelnen standardisierten Domänenwerte für den Vergleich von Leitlinien genutzt werden.

B) Gesamtbewertung der methodischen Leitlinienqualität

Ergänzend zur Berechnung der standardisierten Domänenwerte kann entsprechend dem im Berichtsplan festgelegten Vorgehen mit dem AGREE-Instrument eine Gesamtbewertung der Leitlinienqualität durchgeführt werden [3]. Wie im Vorschlag der AGREE Collaboration werden drei Kategorien unterschieden: „Zu empfehlen“, „Zu empfehlen unter Vorbehalt / nach Änderung“ und „Nicht zu empfehlen“.

4.2.3 Strukturierte Aufbereitung und Auswertung von Empfehlungen

A) Leitlinienempfehlungen, Evidenz- und Empfehlungsstufen

Eine Empfehlung wird als ein Handlungsvorschlag für die klinische Entscheidung bzw. für Systementscheidungen definiert. Adressat ist in der Regel der professionell Handelnde. Als Empfehlungen werden in den Leitlinien grundsätzlich jene Aussagen identifiziert, die von den Autorinnen und Autoren einer Leitlinie formal als Empfehlung gekennzeichnet sind. Darüber hinaus können Empfehlungen, wenn sie nicht formal gekennzeichnet sind und in Abhängigkeit von der Fragestellung, aufgrund einer sprachlichen Kennzeichnung (z. B. „wird empfohlen, muss, soll, sollte, kann, könnte, in Erwägung ziehen“ inklusive Verneinungen oder Negativempfehlungen) identifiziert werden.

Die Ersteller von evidenzbasierten Leitlinien verwenden unterschiedliche Systeme zur Evidenz- und Empfehlungseinstufung ihrer Empfehlungen [24,143,221,320,468]. Evidenzeinstufungen (LoE) sollen die Leserin oder den Leser bzw. die Anwenderin oder den Anwender einer Leitlinie in Kurzform über die Stärke (Qualität und Quantität) der der Empfehlung zugrunde liegenden Evidenz informieren. Unter „Evidenz“ wird hierbei die von den Leitlinienerstellern systematisch recherchierte und ausgewertete Primär- und Sekundärliteratur verstanden. Evidenzeinstufungen im Hinblick auf die (Nutzen-)Bewertung medizinischer Interventionen basieren grundsätzlich auf einer Hierarchie der Evidenz / Studientypen.

Empfehlungseinstufungen (GoR) geben der Leserin oder dem Leser bzw. der Anwenderin oder dem Anwender einer Leitlinie Informationen über die Stärke einer Leitlinienempfehlung. Sie gehen über die Evidenzeinstufungen hinaus, da sie nicht nur die Stärke der zugrunde gelegten Evidenz berücksichtigen, sondern in der Regel auch eine Abwägung des

medizinischen, sozialen, patientenspezifischen und / oder ökonomischen Nutzens und der entsprechenden Risiken einer Empfehlung beinhalten [24,221,468]. Ebenso können sie Bezug auf die konkrete Versorgungssituation in einem Gesundheitssystem nehmen.

B) Strukturierte Aufbereitung von Empfehlungen

Im Rahmen der strukturierten Aufbereitung erfolgt zunächst eine tabellarische Auflistung der Empfehlungen aus den eingeschlossenen Leitlinien in der Originalsprache, separat für die Versorgungsaspekte Prävention, Diagnosestellung, Therapie und / oder Rehabilitation bzw. Nachsorge. Darüber hinaus werden in den Extraktionstabellen die Evidenzeinstufungen und / oder Empfehlungsgrade für eine Empfehlung aufgeführt, sofern diese von den Leitlinienerstellern vergeben wurden. Je nach Fragestellung werden ggf. weitere Informationen dargestellt.

Da international eine konsenterte Vereinheitlichung von Evidenz- und Empfehlungsgraduierungssystemen bislang nicht erfolgt ist, werden grundsätzlich die von den einzelnen Leitlinienerstellern verwendeten Evidenz- bzw. Empfehlungsstufen angegeben und die entsprechenden Graduierungssysteme dokumentiert. Um die Systeme der verschiedenen Ersteller besser miteinander vergleichen zu können, werden, wenn möglich bzw. vorgesehen, vergleichbare Evidenz- / Empfehlungseinstufungen verschiedener Ersteller in übergreifenden Evidenz- bzw. Empfehlungskategorien zusammengefasst.

C) Auswertung der extrahierten Empfehlungen

Zusammenfassung der Empfehlungen

Die Auswertung der Extraktionstabellen besteht zunächst in einer inhaltlichen Zusammenfassung der extrahierten Empfehlungen zu den einzelnen Versorgungsaspekten. Dabei wird – sofern in den Leitlinien angegeben – in den Berichten sowohl GoR als auch LoE der jeweiligen Empfehlungen dargestellt.

Synthese von Kernaussagen

Wenn im Berichtsplan vorgesehen, können die inhaltlichen Informationen aus den Empfehlungen verschiedener Leitlinien zu demselben Versorgungsaspekt in einer übergreifenden „Kernaussage“ zusammengefasst werden. Kernaussagen werden tabellarisch dargestellt und mit der Angabe verbunden, welche Leitlinie diese Aussage mit welcher Evidenz- und / oder Empfehlungskategorie belegt.

D) Identifizierung von Lücken oder Divergenzen in der Darstellung der Versorgungsstandards

Die strukturierte Aufbereitung und Auswertung von Empfehlungen aus Leitlinien ermöglicht die Identifizierung von Lücken und Abweichungen bzw. Übereinstimmungen in der Darstellung bestehender Versorgungsstandards.

Leitlinien adressieren je nach Zielsetzung bestimmte Versorgungsbereiche bis hin zur gesamten Versorgungskette einer Erkrankung. Fehlen in einzelnen Leitlinien Empfehlungen zu adressierten Bereichen der Versorgungskette (z. B. zur Rehabilitation oder zur Therapie), liegt eine Lücke in der Darstellung von Versorgungsstandards vor. Dies kann mehrere Ursachen haben. Zentral sind Besonderheiten der Evidenzlage (z. B. fehlend, mangelhaft, unzureichend übertragbar etc.). Wesentlich sind auch Aspekte der Versorgung wie beispielsweise die Zulassungs- oder Erstattungssituation in einem Gesundheitssystem oder die im entsprechenden Kontext zur Verfügung stehenden Möglichkeiten.

Abweichungen in der Darstellung von Versorgungsstandards sind Unterschiede in den Leitlinienempfehlungen bzw. den vorgenommenen Evidenz- oder Empfehlungseinstufungen. Diese können zum Beispiel dann auftreten, wenn Empfehlungen sich inhaltlich unterscheiden oder inhaltlich nahezu identische Empfehlungen oder Kernaussagen zum gleichen Versorgungsaspekt mit sehr unterschiedlichen Evidenz- oder Empfehlungsgraden hinterlegt sind. Die Identifikation von Abweichungen kann unter anderem auf eine unsichere Evidenz- oder Konsenslage in einem umschriebenen Aspekt oder den Einfluss kontextspezifischer Faktoren hinweisen.

4.3 Empfehlungen zu Disease-Management-Programmen

Die über das in Abschnitt 4.2 beschriebene Vorgehen identifizierten Versorgungsstandards können als Grundlage für die Erstellung von DMPs dienen. Durch die Gegenüberstellung von Versorgungsstandards mit vorliegenden DMP-Empfehlungen kann zudem ein potenzieller Überarbeitungsbedarf für DMPs ermittelt werden. Als Grundlage für die Erstellung oder Überarbeitung von DMPs eignen sich insbesondere inhaltlich konsistente Empfehlungen, denen in den zugrunde gelegten Leitlinien ein hoher Empfehlungsgrad zugewiesen wurde. Bei einem Fehlen von Empfehlungsgraden wird auf einen hohen Evidenzgrad zurückgegriffen.

4.4 Versorgungsanalyse

4.4.1 Hintergrund

A) Versorgung

Versorgung ist definiert durch die medizinische und psychosoziale Betreuung kranker Menschen sowie Maßnahmen zur Prävention und Gesundheitsförderung durch medizinische und nicht medizinische Anbieter von Gesundheitsleistungen. Die medizinische Betreuung beinhaltet Diagnose, Behandlung, Pflege, Rehabilitation und Nachsorge. Die Versorgung umfasst alle Maßnahmen innerhalb des Gesundheitssystems, die direkt oder indirekt darauf ausgerichtet sind, den Gesundheitszustand (Mortalität, Morbidität, Lebensqualität) bestimmter Individuen oder Populationen zu verbessern oder zu erhalten [20].

B) Versorgungsstandard

Ärztliches Handeln, das nach medizinisch-wissenschaftlicher Erkenntnis und / oder ärztlicher Erfahrung in der Profession akzeptiert ist, stellt den medizinischen Standard dar [234]. Dieser

medizinische Standard wird als Versorgungsstandard bezeichnet. Er kann durch Gesetze, Verordnungen und Richtlinien vorgegeben sein oder in Leitlinien identifiziert werden (siehe auch Abschnitt 4.2). Auch die Referenzwerte von Qualitätsindikatoren können als Versorgungsstandards interpretiert werden [176].

C) Versorgungsqualität

Für die Beurteilung der Versorgungsqualität wird die tatsächliche Versorgungssituation, bezogen auf Strukturen, Prozesse und Ergebnisse, den jeweiligen Versorgungsstandards, die durch Normen, Richt- und Leitlinien vorgegeben werden, gegenübergestellt [235,273]. Durch den Abgleich von Soll- und Ist-Zustand werden Aussagen zur aktuellen Versorgungsqualität möglich. Hierbei bildet die aktuelle Versorgung das „Ist“ und ein aktueller Versorgungsstandard das „Soll“ ab. Das Soll beschreibt dabei die angestrebten Ziele der Versorgung, d. h. die „ideale“ Versorgung. Diese Analyse / Bewertung erfolgt für einen durch die Fragestellung des Auftrags umgrenzten Versorgungsbereich. Voraussetzung für die Bestimmung der Versorgungsqualität ist, dass Daten über die Versorgung, die systematisch erhoben und ausgewertet wurden, vorliegen und ein entsprechender Versorgungsstandard ermittelt werden kann. Dabei wird unter „systematisch“ die geplante Datenerhebung mit einheitlichen Dokumentationsvorschriften (Kodierungsvorschriften, z. B. Version der ICD), mit standardisierten Erhebungsbögen, mit vollzähliger Erfassung und, sofern möglich, mit flächendeckender Erhebung (abhängig von der Fragestellung) verstanden.

4.4.2 Inhaltliche Aspekte einer Versorgungsanalyse

Die Versorgungsanalyse umfasst die – je nach Auftrag unterschiedlich detaillierte – aktuelle und systematische Beschreibung, Analyse und Bewertung von Versorgungsaspekten einer definierten Bevölkerungsgruppe zu einer konkreten medizinischen oder systembezogenen Fragestellung (vgl. § 139a Abs. 3, Nr. 1 und 2 SGB V).

In der Regel untersucht die Analyse die deutsche Versorgungssituation, eventuell ergänzt um einen internationalen Vergleich. Die Versorgungsanalyse erlaubt die Untersuchung komplexer Interventionen, bezogen sowohl auf patientenrelevante Endpunkte als auch auf gesundheitssystembezogene Zielgrößen. Für die Versorgungsanalyse können in einem modularen System verschiedene individualmedizinische, populations- und gesundheitssystembezogene Daten bzw. Studien zusammengestellt werden. In den Gesundheitswissenschaften wird der Begriff „Individualmedizin“ für die „klassische“ Medizin (am Patienten) benutzt; damit soll eine Abgrenzung zur Bevölkerungsmedizin geschaffen werden. Letztere ist ein Teil von Public Health.

Die Versorgungsanalyse kann verschiedene Ebenen und / oder mehrere Versorgungsaspekte beschreiben und bewerten. Grundsätzlich werden dabei zwei Bereiche unterschieden, ein epidemiologischer und ein Bereich der sozialen Organisation medizinischer Versorgung. Der erste beschreibt die Verteilung und Häufigkeit von Krankheiten in der Bevölkerung. Wenn man sich mit einem Versorgungsproblem beschäftigt, ist der erstgenannte Bereich wichtig,

um abschätzen zu können, welche und wie viele Patientinnen und Patienten davon betroffen sind und ob bei der Lösung des Versorgungsproblems ein Augenmerk auf bestimmte Subgruppen gelegt werden muss, z. B. Seniorinnen und Senioren oder sozial benachteiligte Personen. Der zweite Bereich beschäftigt sich z. B. mit Fragen der versorgungsbezogenen Strukturen und Prozesse.

Die Versorgungsanalyse kann verschiedene Ressourcen des Gesundheitssystems (Input), Strukturen und Prozesse (Throughput), Versorgungsleistungen (Output) und / oder Ergebnisse (Outcome) untersuchen [411]. Zur Bewertung der Versorgungsqualität wird die Versorgungssituation mit einem normativen Standard, dem Versorgungsstandard, abgeglichen, sofern ein solcher vorhanden ist.

4.4.3 Ziele einer Versorgungsanalyse

Das übergeordnete Ziel einer Versorgungsanalyse ist die Beurteilung der Versorgungsqualität.

Folgende Punkte können Teilziele der Versorgungsanalyse sein:

- Prüfung der Umsetzung von Standards innerhalb der Versorgung und Identifizierung möglicher Verbesserungspotenziale,
- Untersuchung der Auswirkungen von Versorgungsmodellen oder Maßnahmen zur Qualitätssicherung auf die Bevölkerung bzw. Patienten- / Bevölkerungsgruppen,
- Bereitstellung von (Hintergrund-)Informationen zur Entwicklung von Qualitätsindikatoren oder zur Priorisierung von Fragestellungen,
- Aufzeigen von Hinweisen auf eine mögliche Über-, Unter- oder Fehlversorgung [443] und ggf. Formulierung von Verbesserungsvorschlägen im Sinne einer optimierten Nutzung vorhandener Ressourcen,
- Identifizierung eines potenziellen Forschungsbedarfs (z. B. klinische Forschung, HTA, Gesundheitssystemforschung).

Im Rahmen eines Projekts wird aus Machbarkeitsgründen in der Regel auf einzelne oder einige wenige der o. g. Ziele mit Bezug auf eine bestimmte Erkrankung fokussiert.

4.4.4 Fragestellung der Versorgungsanalyse

Voraussetzung für die systematische Beschreibung, Untersuchung und Bewertung der Versorgungsbereiche ist die Formulierung einer konkreten Fragestellung. Zur Definition der Fragestellung gehört die Festlegung folgender Punkte:

- Population (Alter, Geschlecht, Erkrankung, falls relevant, Subgruppe bzw. Schweregrad der Erkrankung),
- die zu untersuchenden Interventionen (z. B. Betreuung von Patientinnen und Patienten mit Diabetes mellitus in der Allgemeinarztpraxis),

- Zielgrößen / patientenrelevante Endpunkte (z. B. Strukturmerkmale oder gesundheitsbezogene Lebensqualität),
- Setting, in dem die Versorgung stattfindet (z. B. ambulante Versorgung, akutstationäre Versorgung oder sektorenübergreifende Versorgung).

Bei der Formulierung der Fragestellung ist festzulegen, aus welcher Perspektive (z. B. Patientinnen und Patienten, Gesellschaft, Kostenträger usw.) die Versorgung beschrieben und bewertet werden soll, denn in Abhängigkeit von der Perspektive können sich der Schwerpunkt der Untersuchung und die Auswahl der Zielgrößen ändern. Besonderes Augenmerk kann dabei auf die Interessen vulnerabler Gruppen gerichtet werden.

Regionale Variationen (Disparitäten), internationale Vergleiche wie auch zeitliche Entwicklungen (Trends) können entsprechend der Fragestellung ergänzend behandelt werden.

4.4.5 Mögliche Versorgungsparameter

Im Rahmen der Versorgungsanalyse können verschiedene Parameter herangezogen werden. Unter Versorgungsparameter werden z. B. epidemiologische Maßzahlen oder Indikatoren verstanden, mit deren Hilfe verschiedene Bereiche des Gesundheitssystems beschrieben werden (siehe Tabelle 5).

Tabelle 5: Beispiele möglicher Versorgungsparameter

Beispiele möglicher Versorgungsparameter	
Indikatoren	Versorgungsparameter
Inzidenz, Prävalenz, Morbidität	Krankheitslast
Case-Fatality-Rate	Schweregrad der Erkrankung
Einschränkungen und Behinderungen nach der International Classification of Functioning (ICF), Frühberentungen, Mortalität	Folgen der Erkrankung
Anzahl der Ärztinnen und Ärzte pro 1000 Einwohner, Anzahl der Leistungserbringer pro Raumeinheit, Anzahl der Krankenhausbetten pro 1000 Einwohner usw.	Strukturen des Gesundheitssystems (z. B. Deutschland)
Inanspruchnahme von Leistungen oder Leistungserbringung	Leistungsmengen
Qualitätsindikatoren für den stationären / ambulanten Bereich, z. B. für Patientensicherheit, leitlinienkonforme Betreuung der Patientinnen und Patienten	Qualität der medizinischen Versorgung
z. B. Säuglings- und / oder Müttersterblichkeit, Durchimpfungsraten, Verweildauern im Krankenhaus	Strukturen, Prozesse und Ergebnisse der Versorgung im internationalen Vergleich

Um einen Überblick über das Ausmaß des Versorgungsproblems zu erhalten, können epidemiologische Maßzahlen, z. B. die Prävalenz einer Erkrankung, herangezogen werden. Sie geben Auskunft über die Häufigkeit der Erkrankung [326]. Anhand der Case-Fatality-Rate kann der Schweregrad einer Krankheit eingeschätzt werden [242]. Die Folgen einer Erkrankung können anhand der Daten gemäß der International Classification of Functioning (ICF) und der Daten der Rentenversicherung, z. B. zu Erwerbsminderungsrenten, beurteilt werden [117,490]. Versorgungsstudien sowie Daten der Kostenträger bzw. Leistungserbringer (Krankenkassen Kassenärztliche Vereinigungen u. a.) können das Inanspruchnahmeverhalten von Patientinnen und Patienten aufzeigen. Sie geben somit Auskunft darüber, wie häufig medizinische Leistungen nachgefragt, in Anspruch genommen bzw. erbracht werden. Qualitätsindikatoren für die Struktur-, Prozess- und Ergebnisqualität der stationären und/oder ambulanten Versorgung können den Datenpool ergänzen. Sie dienen der Qualitätssicherung und können Hinweise auf konkrete Versorgungsprobleme bezogen auf einzelne Strukturmerkmale, Prozessschritte oder Einzelergebnisse geben. Darüber hinaus können Daten zur Patientensicherheit aus Qualitätsberichten der Krankenhäuser, Registern sowie klinischen Studien und qualitativen Studien – soweit verfügbar – in eine Versorgungsanalyse einfließen. Sie decken z. B. vermeidbare unerwünschte Ereignisse auf. Evaluationsberichte zu Modellvorhaben nach § 63 SGB V können Hinweise auf mögliche neue Versorgungswege aufzeigen. Auf Systemebene können weitere Parameter zur Beschreibung der Versorgungssituation eingesetzt und international verglichen werden. Beispiele sind die Durchimpfungsraten, die krankheitsspezifische Lebenserwartung, die Anzahl der Krankenhausbetten bezogen auf 1000 Einwohner und der Anteil der Ausgaben für Gesundheitsleistungen bezogen auf das Bruttoinlandsprodukt [297,303,542].

In Abhängigkeit von der Fragestellung lassen sich die o. g. Parameter und ggf. weitere kombinieren und ermöglichen so einen umfassenden Überblick über einzelne Versorgungsbereiche. Die diesen Versorgungsbereichen zugeordneten Versorgungsstandards werden – wie in Abschnitt 4.4.8 beschrieben – identifiziert.

4.4.6 Ablauf einer Versorgungsanalyse

Der Ablauf einer Versorgungsanalyse wird in Abbildung 7 beispielhaft dargestellt.

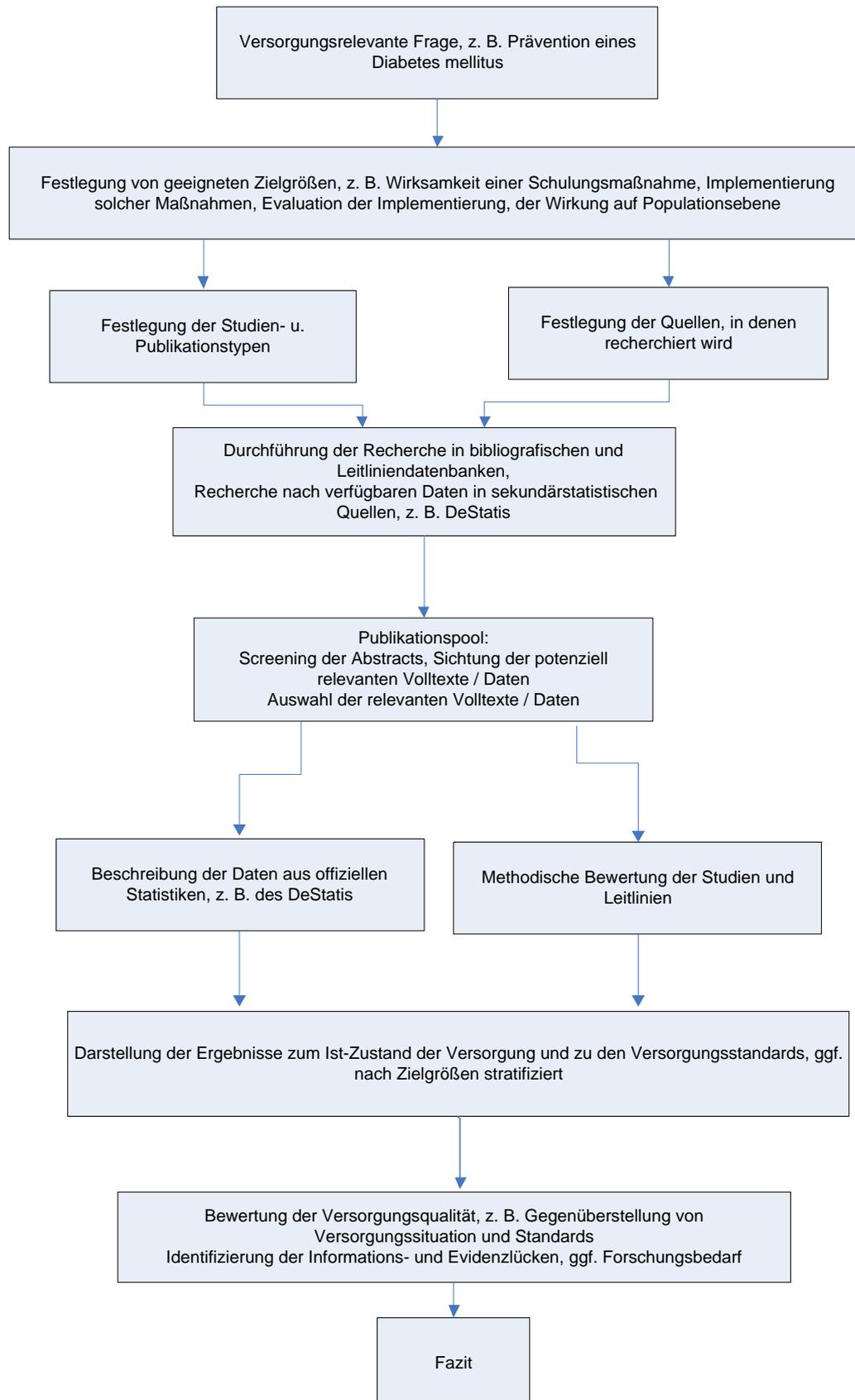


Abbildung 7: Exemplarischer Ablauf einer Versorgungsanalyse

4.4.7 Ebenen der Versorgungsanalyse

Die Beschreibung der Versorgung anhand der o. g. Parameter kann bezogen auf drei unterschiedliche Ebenen erfolgen: auf der Ebene der Individualmedizin, auf der Ebene der Bevölkerungsmedizin und auf der Ebene des Gesundheitssystems.

Die erste Ebene betrifft individuelle Patientinnen und Patienten bzw. Patientengruppen in einem klinischen Setting. Typische Zielgrößen auf dieser Ebene sind patientenrelevante Endpunkte wie Mortalität, Morbidität und gesundheitsbezogene Lebensqualität.

Die zweite Ebene bezieht sich auf populationsbezogene Untersuchungen im Sinne der Evidence Based Health Care (Bevölkerungsmedizin) [245,326]. Zielgrößen auf dieser Ebene sind ebenfalls patientenrelevante Endpunkte wie Mortalität und Morbidität allerdings mit dem Bezugspunkt allgemeine Bevölkerung [326]. Darüber hinaus können weitere Zielgrößen untersucht werden, wie z. B. die Teilnehmerrate oder die Gründe für die Teilnahme an Screening- oder Impfprogrammen oder an Versorgungsmodellen wie den DMPs.

Die dritte Ebene ist die des Gesundheitssystems [21,110]. Zielgrößen dieser Ebene können die Inanspruchnahme von medizinischen Leistungen oder die Leistungserbringung in verschiedenen Settings (ambulante / stationäre Versorgung) oder durch verschiedene Berufsgruppen / Anbieter sein.

Nach der Beschreibung von Bereichen der deutschen Versorgung kann ein internationaler Vergleich sinnvoll sein. Je nach Fragestellung kann sich die Beschreibung der Versorgung in einem modularen System auf Informationen aus allen drei Ebenen (Individualmedizin, Bevölkerungsmedizin und Gesundheitssystem) beziehen. Auch können auf allen Ebenen zeitliche Entwicklungen und regionale Variationen (Disparitäten) untersucht werden [115]; dazu können u. a. auch geografische Informationssysteme zum Einsatz kommen.

4.4.8 Methodische Besonderheiten der Versorgungsanalyse

Mit Bezug auf die Komplexität des Gesundheitssystems und die o. g. Ebenen (siehe Abschnitte 4.4.5 und 4.4.7) können im Rahmen der Versorgungsanalyse unterschiedliche Studien- und Publikationstypen herangezogen werden.

Darüber hinaus kann es notwendig sein, unterschiedliche Fragestellungen zur Versorgung mit verschiedenen quantitativen und qualitativen Methoden zu bearbeiten (Methodenpluralismus). Auch werden Daten mehrerer Quellen herangezogen (siehe Abschnitt 4.4.9) und mit verschiedenen Methoden bearbeitet. Die methodische Bewertung erfolgt – soweit möglich – mit geeigneten Instrumenten (siehe Abschnitt 4.4.9).

Zusätzlich kann die Berücksichtigung von soziokulturellen und ethischen Aspekten bei der Bewertung der Versorgungsqualität bestimmter Patientengruppen notwendig sein, z. B. der Zugang zur Versorgung.

4.4.9 Informationsbeschaffung

In Abhängigkeit von der Fragestellung wird ggf. in unterschiedlichen Quellen recherchiert. Die Recherche wird entsprechend den Anforderungen der Quelle entwickelt. Eine Literaturrecherche wird wie auch die Leitlinienrecherche nach den Methoden des Instituts durchgeführt (siehe Abschnitt 6.1).

A) Bestimmung des Versorgungsstandards

Aus der Fragestellung für die Versorgungsanalyse wird die Art des Versorgungsstandards abgeleitet. In erster Präferenz werden Versorgungsstandards über evidenzbasierte Leitlinien identifiziert. Das systematische Vorgehen zur Identifizierung von Versorgungsstandards über Leitlinien wird in Abschnitt 4.2 beschrieben. Gesetze, Verordnungen und Richtlinien stecken rechtsverbindlich den Rahmen der gesundheitlichen / medizinischen Versorgung ab.

Strukturen und Prozesse werden meist mittels Qualitätsindikatoren geprüft. Qualitativ hochwertige Leitlinien weisen u. a. Qualitätsindikatoren aus. Dabei handelt es sich um Maße, die die Güte der Versorgung indirekt abbilden. Sie können auf Struktur-, Prozess- und Ergebnisqualität angewendet werden. Der Referenzbereich des Qualitätsindikators gibt das Versorgungsziel, d. h. den Standard vor. Ein Indikator bezieht sich immer nur auf einen Versorgungsbereich, deshalb ist die Kombination mehrerer Indikatoren zur Prüfung der Qualität sinnvoll [10]. Tabelle 6 gibt einen Überblick über mögliche Quellen zur Identifizierung von Versorgungsstandards.

Tabelle 6: Informationsquellen zur Identifizierung von deutschen Versorgungsstandards

Informationen zu	Beispiele für Datenhalter
Versorgungs- bzw. medizinischen Standards (Leitlinien)	AWMF, G-I-N, NGC
Gesetze (SGB) und Verordnungen	BMJ, BMG
Richtlinien	G-BA BÄK
Indikatoren für die Struktur-, Prozess- und Ergebnisqualität	KBV, z. B. AQUIK BQS, AQUA

B) Identifikation von Datenquellen für Versorgungsdaten

Alle für die jeweilige Fragestellung relevanten Datenquellen sollen zur Beschreibung der Versorgung identifiziert und, soweit möglich, genutzt werden. Entsprechend den allgemeinen Grundsätzen einer themenbezogenen Informationsgewinnung (siehe Abschnitt 6.1) wird die Auswahl der Datenquellen (z. B. bibliografische Datenbanken, Datenbanken von Organisationen, die offizielle Statistiken vorhalten, Morbiditätsregister, Handsuche in ausgewählten Fachzeitschriften, Kontaktaufnahme mit Fachexpertinnen und Fachexperten, Patientenorganisationen, ggf. Industrie etc.) im Berichtsplan verbindlich festgelegt. Im

Folgenden werden mögliche Datenquellen zur Identifizierung von Versorgungsdaten benannt (siehe Tabelle 7).

Tabelle 7: Informationsquellen zur Identifizierung von Versorgungsdaten

Informationen zu	Beispiele für Quellen	
Morbidität und Mortalität, z. B. Inzidenzen und Prävalenzen (Populationsebene)	<ul style="list-style-type: none"> ▪ Gesundheitsberichterstattung des Bundes und der Länder (z. B. Kinder- und Jugendgesundheitssurvey des RKI) ▪ Berichte des Statistischen Bundesamts (z. B. Krankenhausentlassungsdiagnosen, Todesursachenstatistik) ▪ Morbiditätsregister (z. B. epidemiologische Krebsregister) ▪ Routinedaten, z. B. von Krankenkassen oder Kassenärztlichen Vereinigungen 	
Versorgungsbedarf (z. B. regionale Bedarfsanalysen)	<ul style="list-style-type: none"> ▪ Versorgungsstudien 	
Inanspruchnahme- und Verordnungsverhalten	<ul style="list-style-type: none"> ▪ Arzneiverordnungs-Report (WidO) ▪ Krankenhaus-Report (WidO) ▪ Heilmittel-Report (WidO) ▪ ICD-10-Schlüsselnummern nach Fachgruppen (Zentralinstitut) ▪ Routinedaten, z. B. der Krankenkassen oder der Kassenärztlichen Vereinigungen 	
Patientensicherheit	<ul style="list-style-type: none"> ▪ Schiedsstellen der Landesärztekammern ▪ Qualitätsindikatoren der OECD ▪ weitere Publikationen der GKV 	
Messen der Versorgungsqualität mit Indikatoren	<ul style="list-style-type: none"> ▪ Qualität der Versorgung auf Systemebene ▪ Qualität in der ambulanten ärztlichen Versorgung ▪ Qualität in der stationären Versorgung ▪ Qualität in der Pflege 	<ul style="list-style-type: none"> ▪ OECD (z. B. Zugang zur Versorgung) ▪ Qualitätsberichte der Kassenärztlichen Vereinigungen ▪ Qualitätsberichte der Krankenhäuser nach § 137 ▪ Veröffentlichungen der BQS / AQUA ▪ Pflegeberichte des Medizinischen Dienstes der Krankenversicherung (MDK)
▪ DMPs	<ul style="list-style-type: none"> ▪ Evaluationsberichte von DMPs 	
Gesundheitssystem / Gesundheitssystemvergleich	<ul style="list-style-type: none"> ▪ Z. B. Publikationen der WHO (z. B. World Health Report) 	

C) Ablauf einer Recherche nach Versorgungsdaten

Der Ablauf der Recherche folgt dem institutsüblichen Vorgehen. Dies ist für die Leitlinien- und für die bibliografische Recherche an anderer Stelle beschrieben (siehe Abschnitt 6.1). Nach Versorgungsdaten, z. B. aus offiziellen Statistiken oder Morbiditätsregistern, wird gezielt gesucht. Die Suche und das Rechercheergebnis werden dokumentiert. Bei relevanten Daten, die ausschließlich auf Internetseiten publiziert werden, richtet sich die anzuwendende Suchstrategie nach dem Aufbau und den Möglichkeiten der jeweiligen Internetseiten.

In Abhängigkeit von der konkreten Fragestellung werden zur Beschreibung der Versorgung verschiedene Datenquellen / Studientypen herangezogen.

4.4.10 Bewertung der identifizierten Daten

Bei der Bewertung der identifizierten Daten sind folgende Aspekte zu beachten:

- Bewertung der Studien- und Publikationsqualität der eingeschlossenen Studien: Die Studien werden nach den Allgemeinen Methoden des Instituts bewertet. Ergänzungen, z. B. bei Evaluations- oder qualitativen Studien [201], werden begründet.
- Bewertung von Studien mit Konstrukten als Endpunkten: Bei patientenrelevanten Endpunkten, die Konstrukte sind, wie z. B. die gesundheitsbezogene Lebensqualität, wird die Validität des Erhebungsinstruments überprüft. Nicht validierte Instrumente eignen sich nicht für einen Vergleich.
- Bewertung der offiziellen Statistiken: Die offiziellen Statistiken, z. B. des DeStatis, werden nicht methodisch bewertet, weil es oft nicht möglich ist, solche Statistiken – z. B. die Todesursachenstatistik – methodisch zu prüfen. Darüber hinaus unterliegen sie bereits den strengen Qualitätskriterien ihrer herausgebenden Organisation [68,403]. Publikationen dieser Daten durch Dritte, z. B. als Artikel in Zeitschriften, werden nach den Allgemeinen Methoden des Instituts bewertet.
- Bewertung von Registerdaten: Wenn Endpunkte mittels Registerdaten dargestellt werden, sollte nach der Validität des Registers (Datenqualität, d. h. Vollständigkeit und Plausibilität der Datensätze, Vollzähligkeit, Aktualität) gefragt werden [403,520].
- Abschließende Bewertung der Studien- und Publikationsqualität: Die Einschätzung des Verzerrungspotenzials der zu beurteilenden Studien bzw. Publikationen erfolgt nach den Allgemeinen Methoden des Instituts.
- Bewertung der methodischen Leitlinienqualität: Dies ist in Abschnitt 4.2.3 beschrieben.

4.4.11 Informationssynthese und -analyse

Die Informationssynthese und -analyse läuft wie folgt ab: Zunächst wird die vorhandene Literatur auf relevante Angaben zu den im Berichtsplan vorgegebenen Zielgrößen geprüft und nach den Allgemeinen Methoden des Instituts bewertet und beschrieben. Anschließend

werden die Ergebnisse zusammengefasst. Basierend auf den Ergebnissen der Versorgungsanalyse wird eine Bewertung der Versorgungsqualität durchgeführt.

4.4.12 Darstellung und Bewertung der Versorgungsqualität

Die Bewertung der Versorgungsqualität besteht aus drei Schritten:

Schritt 1: Beschreibung des Ist-Zustands

Die Beschreibung des Ist-Zustandes der Versorgung erfolgt wie im Berichtsplan festgelegt. Folgende Fragen sind dabei zu beachten:

- Gibt es Daten, Studien oder Publikationen zu den Zielen bzw. Versorgungsaspekten der Untersuchung?
- Wie zuverlässig sind die gefundenen Ergebnisse?

Schritt 2: Beschreibung des Soll-Zustands

Im nächsten Schritt werden Versorgungsstandards identifiziert und dargelegt. Auch hier wird nach dem Vorliegen und der methodischen Qualität der Standards gefragt:

- Gibt es einen Versorgungsstandard für die im Berichtsplan genannten Ziele bzw. Versorgungsaspekte?
- Wie zuverlässig sind die gefundenen Ergebnisse?

Schritt 3: Abgleich zwischen Ist- und Soll-Zustand

Danach wird der Ist-Zustand der Versorgung den Versorgungsstandards gegenübergestellt und abgeglichen. Unter Berücksichtigung der folgenden Fragen ergibt sich daraus die Bewertung der Versorgungsqualität:

- Wird der Versorgungsstandard in der alltäglichen Versorgung umgesetzt?
- Wie groß sind diese Abweichungen von Ist- und Soll-Zustand? In welche Richtung weichen Ist und Soll ab (Über-, Unterversorgung)?
- Welche Schlussfolgerungen ergeben sich aus dem o. g. Vergleich?

Eine abschließende Beurteilung erfolgt im Fazit des Berichts. Die Bewertung ermöglicht eine Einschätzung, ob Informations- und / oder Evidenzlücken vorliegen, ob Forschungsbedarf besteht und / oder ob es Verbesserungspotenziale gibt.

4.5 Validität von Leitlinienempfehlungen

4.5.1 Hintergrund

Obwohl die Methodik zur Leitlinienentwicklung zunehmend weiterentwickelt wird [19,220], weisen Leitlinien immer noch erhebliche Unterschiede in ihrem Erstellungsprozess, ihrer Darstellung, ihrer methodischen Qualität und nicht zuletzt in ihren Inhalten auf

[74,75,80,253,262,359,361,365]. Eine hohe methodische Qualität von Leitlinien korreliert zudem nicht notwendigerweise mit der inhaltlichen Qualität der darin enthaltenen Empfehlungen [539]. Für viele Fragestellungen reicht die Überprüfung der methodischen Qualität einer Leitlinie daher nicht aus, um die Wertigkeit einzelner Empfehlungen einzuschätzen [220]. Hieraus ergibt sich die Notwendigkeit, dass Inhalte von Leitlinien, insbesondere in Bezug auf die Validität einzelner Empfehlungen, analysiert und überprüft werden müssen.

4.5.2 Validität von Leitlinienempfehlungen

Es wird zwischen „interner Validität“ und „externer Validität“ von Leitlinienempfehlungen unterschieden. Diese werden wie folgt definiert:

Interne Validität von Leitlinienempfehlungen: Sicherstellung der Minimierung potenziell verzerrender Einflussfaktoren bei der Entwicklung von Leitlinienempfehlungen.

Externe Validität von Leitlinienempfehlungen: Beschreibung der Anwendbarkeit einer Empfehlung unter den Bedingungen der in der Leitlinie beschriebenen Versorgungssituation. Dies kann die klinische Versorgung ebenso umfassen wie die Nutzung einer Leitlinie auf Systemebene.

Unter der Bewertung der internen Validität von Leitlinienempfehlungen werden die Bewertung des Umgangs mit der der Empfehlung zugrunde liegenden Literatur sowie die Bewertung des Konsensusprozesses verstanden. Unter der Bewertung der externen Validität von Leitlinienempfehlungen wird die Berücksichtigung von Kontextaspekten (z. B. Verfügbarkeit, Patientenpräferenzen, ethische Aspekte) oder Kosten bei der Generierung und Formulierung der Empfehlung gefasst.

Von der externen Validität wird der Begriff der „Übertragbarkeit“ abgegrenzt, der beschreibt, in welchem Maße eine Empfehlung auf einen anderen Kontext übertragbar ist. Dies kann eine Übertragbarkeit zwischen verschiedenen Gesundheitssystemen ebenso betreffen wie die Übertragbarkeit innerhalb eines Systems (z. B. unterschiedliches Setting, unterschiedliche Patientenzielgruppen).

4.5.3 Ziel der Analyse und Bewertung von Leitlinienempfehlungen

Das Ziel des methodischen Vorgehens ist die Bewertung der internen Validität einzelner Leitlinienempfehlungen. Aspekte der externen Validität werden nur berücksichtigt, wenn sie für die Bewertung der internen Validität hilfreich sind. So muss beispielsweise der Kontext, für den die Leitlinie bestimmt ist, oder die Anwenderzielgruppe ggf. bei der Bewertung der zugrunde liegenden Evidenz berücksichtigt werden. Eine darüber hinausgehende tiefer gehende Überprüfung der externen Validität erfolgt nicht.

4.5.4 Vorgehen bei der Analyse und Bewertung der internen Validität

Die Bewertung der internen Validität einzelner Leitlinienempfehlungen erfolgt anhand der:

- 1) Identifikation und Dokumentation potenziell verzerrender Faktoren, die die interne Validität einer Leitlinienempfehlung beeinträchtigen könnten. Potenziell verzerrende Faktoren werden auf der Ebene der Charakteristika und Struktur der Leitlinie bzw. Empfehlung der Leitlinie, der Evidenzbasis der Empfehlung und des Konsensusprozesses identifiziert und dokumentiert.
- 2) Identifikation von Anpassungsbedarf für die Leitlinienempfehlung. Dieser resultiert aus dem unter Punkt 1 identifizierten Verzerrungspotenzial. Hierbei ist je nach Schwere der festgestellten Mängel zwischen einem möglichen und einem zwingenden Anpassungsbedarf zu unterscheiden. Eine Empfehlung wird als „nicht valide“ eingestuft, wenn ein zwingender Anpassungsbedarf für diese Empfehlung festgestellt wird.

4.5.5 Mögliche Fragestellungen

Die Methoden zur Analyse und Bewertung der internen Validität von Leitlinienempfehlungen sind auf verschiedene Fragestellungen anwendbar. Mögliche Fragestellungen sind:

- Bewertung einzelner Empfehlungen mehrerer Leitlinien zu einer Erkrankung bzw. zu einem mehrere Interventionen umfassenden Versorgungsaspekt: z. B. Bewertung der internen Validität von Leitlinienempfehlungen evidenzbasierter Leitlinien zur präoperativen Diagnostik.
- Bewertung einzelner Empfehlungen mehrerer Leitlinien zu einer Intervention: z. B. Bewertung der internen Validität von Leitlinienempfehlungen evidenzbasierter Leitlinien zur Behandlung des Diabetes mellitus Typ 2 mit langwirksamen Insulinanaloga.
- Bewertung einzelner Empfehlungen einer konkreten Leitlinie..

Die Analyse und Bewertung der internen Validität von Leitlinienempfehlungen kann darüber hinaus auch einen Beitrag zur Bewertung evidenzbasierter Leitlinien für die epidemiologisch wichtigsten Krankheiten (vgl. § 139a Abs. 2 Nr. 3 SGB V) leisten.

5 Evidenzbasierte Gesundheitsinformation für Bürgerinnen und Bürger

5.1 Hintergrund und Ziele

Dem Institut wurde vom Gesetzgeber die Aufgabe zugewiesen, allgemeine Gesundheitsinformationen für Bürgerinnen und Bürger 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 [34,88,127,300,315]. 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 Patientinnen und 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 Patientinnen und Patienten in die medizinische Entscheidungsfindung gehört zum Kern des Konzepts der evidenzbasierten Medizin [446] und ist damit auch ein wesentlicher Bestandteil evidenzbasierter Gesundheitsinformationen. Es gibt mehrere Definitionen von evidenzbasierten Patienteninformationen [96,123,155,453,492]. 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 [123,492]. 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 Befähigung der Nutzerinnen und Nutzer 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 getroffen werden können.
- die Informationen aktualisiert werden, damit sie evidenzbasiert bleiben.

Primäres Kommunikationsmedium des Instituts ist das Internet, über das es Bürgerinnen und Bürger, Patientinnen und Patienten und Multiplikatoren zu erreichen sucht. Internetbasierte (Online-) und computergestützte medizinische Offline-Informationen können das Wissen, die Auswahlentscheidungen, die Gesundheit und das Wohlbefinden von Nutzerinnen und Nutzern positiv beeinflussen [42,293,334,383,387,487,538,554]. Informationsinterventionen können allerdings auch ineffektiv oder sogar schädlich sein [146,148,212,387]. Überdies sind viele Menschen mit evidenzbasierten Informationen nicht vertraut, was besondere Anforderungen an die Kommunikation stellt [156,205,493].

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 [199]. Überdies bestehen starke Unterschiede in den Lesekompetenzen von Menschen [300].

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. Das Institut versucht, eine Lesbarkeit der Informationen unterhalb eines Universitätsniveaus zu erreichen. Primäre Mittel einer Qualitätsbewertung im Hinblick auf die Verständlichkeit von Texten sind jedoch die Begutachtung von Entwürfen durch Testleserinnen und Testleser sowie Leserbewertungen (siehe unten).

Die Evidenz zu erklären und bei der Vermittlung gesundheitsbezogener Informationen objektiv zu bleiben, stellt eine zusätzliche Herausforderung dar [156,300,467,514]. 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.

Es gibt zunehmend mehr Evidenz zur wirksamen Kommunikation mit Patientinnen und Patienten sowie der Öffentlichkeit [492,514]. Daraus lassen sich wichtige Erkenntnisse für die Erstellung von Gesundheitsinformationen ableiten. Allerdings sind damit einige Einschränkungen verbunden [145]; beispielsweise werden in den Untersuchungen häufig vorrangig kognitive Zielgrößen berücksichtigt. Erkenntnisse hinsichtlich der Auswirkungen auf psychologische, verhaltensbezogene oder gesundheitsrelevante Outcomes können häufig nicht in gleicher Weise abgeleitet werden. Entsprechende Studien werden häufig auch in Bevölkerungsgruppen mit einem bestimmten Bildungsniveau durchgeführt, dies erschwert eine bevölkerungsbezogene Verallgemeinerung [145]. Darüber hinaus vergleichen die Studien häufig sehr unterschiedliche Interventionen, die in sich vielschichtig sind, sodass unklar bleibt, welcher Bestandteil der Intervention letztendlich ausschlaggebend für die Wirkung auf einen bestimmten Outcome ist.

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 vermitteln,
- unter Berücksichtigung der Übertragbarkeit anzugeben, auf welche Personengruppen die Evidenz anwendbar ist,
- sehr klar und sorgfältig auf den Unterschied zwischen dem fehlenden wissenschaftlichen Nachweis eines Unterschieds („Absence of Evidence“) und dem wissenschaftlichen Nachweis, dass kein Unterschied existiert („Evidence of no Effect“), hinzuweisen,
- wenn es verlässliche und entscheidungsrelevante Daten gibt, werden Risiken mit dem absoluten Risiko angegeben, ggf. ergänzt durch andere Informationen wie dem relativen Risiko, und
- verzerrte Informationen zugunsten oder zuungunsten der 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 von Risikoschätzern für Patientinnen und Patienten zur persönlichen Orientierung bei ihren Entscheidungen hilfreich sein kann [148]. Wenn es verlässliche Zahlen gibt, die Menschen helfen können, ihr Risiko abzuschätzen, wird das Institut diese Informationen darstellen. Eine Technik, die Patientinnen und Patienten bei der persönlichen Abwägung des Nutzens und Schadens helfen soll, ist die Bereitstellung von Entscheidungshilfen. Auch wenn hinsichtlich formaler Entscheidungshilfen – vor allem bei der individuellen Anwendung im Internet [85,153] – eine Ungewissheit bestehen bleibt [397], 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) [153,254].

Menschen zu informieren ist nicht der einzige Zweck von Gesundheitskommunikation. Ein besonderes Anliegen ist es auch, den Nutzerinnen und Nutzern emotionale Unterstützung zu bieten [173]; ferner kann sie auch bei der Stärkung der Patientenautonomie eine Rolle spielen. Gesundheitskommunikation muss patientenzentriert sein, wenn sie zum 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 [396]. Diese Fähigkeiten bilden die Voraussetzung für autonomes Denken und Handeln aufseiten der Bürgerinnen und Bürger. Gesundheitskommunikation zur Stärkung der Gesundheitskompetenz spricht an, was Bürgerinnen und Bürger wissen möchten, zeigt Interesse und Respekt vor ihrer Meinung und erkennt ihre Kompetenz an [129,299,532].

Historisch gesehen waren Patienteninformationen eher paternalistisch orientiert [129]. Man ging davon aus, dass Patientin und Patient unwissend sind und vor Ungewissheit und belastenden Informationen geschützt werden müssen: Ihnen sollte gesagt werden, was sie zu tun haben. Obwohl heute feststeht, dass dies nicht zielführend ist [417], ist dies auch heute noch ein Kennzeichen von Patienteninformationen und der Debatten um ihren aktuellen Stellenwert [129].

Patientinnen und Patienten in Deutschland erhalten weniger Informationen, als sie gern hätten [127], und – mit Blick auf vergleichbare Länder – geben in Deutschland mehr Patientinnen und Patienten an, dass ihnen nicht genügend Informationen zur Verfügung gestellt würden [459,523]. Deutsche Internetnutzerinnen und -nutzer interessieren sich eventuell sogar stärker für die neuesten Forschungsergebnisse als für Ratschläge von Expertinnen und Experten [416]. Dennoch ist zum großen Teil nicht bekannt, welche Wirkungen es hat, wenn Menschen verstärkt über bestehende wissenschaftliche Ungewissheiten aufgeklärt werden. Als Argument für eine Information auch über Unsicherheiten wird vorgebracht, dass der Umgang mit Ungewissheit bei Erwachsenen ein integraler Bestandteil der Bewältigung und Selbstverwirklichung sei [184]. Zur Frage der Messung der Wirkung unterschiedlicher Gesundheitsinformationen in der Öffentlichkeit besteht Forschungsbedarf.

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 der Nutzerinnen und Nutzer, vor ihrer Autonomie, ihren kulturellen Unterschieden sowie gegenüber geschlechts-, alters- und behindertenspezifischen Belangen zu zeigen [191],

- sich um einen patientenzentrierten, nicht wertenden, nicht direktiven und neutralen Sprachstil zu bemühen,
- auf die Zeit der Leserinnen und Leser Rücksicht zu nehmen.

Ein neutraler Sprachstil muss gewährleisten, dass Informationen sowohl Frauen als auch Männer erreichen und sich beide Geschlechter gleichermaßen angesprochen fühlen. Die durchgehende Verwendung maskuliner Personenbezeichnungen (generisches Maskulinum) führt zu einer gedanklichen Unterrepräsentation von Frauen, worin eine sprachliche Benachteiligung von Frauen zu sehen ist [280]. In den Texten der Gesundheitsinformationen verwendet das Institut daher einen weitestgehend geschlechterneutralen Sprachstil, der möglichst vollständig auf das generische Maskulinum verzichtet. Stattdessen werden, wo immer möglich, beide Geschlechter explizit genannt, wenn auch beide gemeint sind, oder geschlechtsneutrale Formulierungen gewählt.

Auf der Grundlage der verfügbaren Evidenz und der Erfahrungen anderer Gruppen entwickelt das Institut für seine Produkte einen Leitfaden zur Textgestaltung sowie Sensibilisierungs- und andere Maßnahmen für die redaktionellen Mitarbeiterinnen und Mitarbeiter [386]. 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 Methode der Patientenwege

Patientenzentrierte Gesundheitsinformationen orientieren sich an den Fragen der Nutzerinnen und Nutzer, berücksichtigen die Erfahrungen und Sichtweisen von Patientinnen und Patienten und erkennen ihre Kompetenz an. Patientenzentrierte Gesundheitsinformationen sollen nicht nur medizinisch-fachliche Fragen beantworten und eine informierte Entscheidung ermöglichen, sondern auch emotionale Unterstützung bieten. Dafür ist es einerseits erforderlich zu wissen, welche Fragen für die Nutzerinnen und Nutzer interessant sein könnten. Andererseits ist es wichtig, dass sich die Autorinnen und Autoren beim Schreiben der Informationen an die Perspektive der Patientinnen und Patienten oder Angehörigen annähern und ein Verständnis dafür entwickeln, was es bedeutet, mit einer bestimmten Krankheit zu leben. Das Institut bedient sich hierzu einer Methode, die die möglichen Wege von Patientinnen und Patienten mit einer bestimmten Erkrankung nachzeichnet. Sie wird im Folgenden als „Methode der Patientenwege“ bezeichnet.

Mithilfe von Patientenwegen sollen die verschiedenen sozialen, emotionalen, kognitiven und klinischen Dimensionen, die mit einer Krankheit einhergehen können, kurz zusammengefasst und dann grafisch dargestellt werden. Die Methode lehnt sich an die medizinsoziologischen Modelle der „Illness Trajectory“ [93] und der „Patientenkarriere“ [200,319] sowie verschiedene Modelle der „Patient Journey“ [325] an.

Die Medizinsoziologie beschäftigte sich schon früh mit den Auswirkungen von Erkrankungen auf das Leben von Patientinnen und Patienten. In diesem Zusammenhang wurde in Deutschland unter anderem der Begriff der „Patientenkarriere“ geprägt. Beispielsweise sind hier die Entwicklungen von Goffman, Gerhardt und Dörner zu nennen [200,319]. Ein weiterer Ansatz ist die Krankheitsverlaufskurve oder „Illness Trajectory“ nach Corbin und Strauss [93].

Aus diesen verschiedenen Ansätzen hat das Institut seine Methode der Patientenwege abgeleitet. Auch wenn Patientinnen und Patienten zu Expertinnen und Experten des Lebens mit einer bestimmten Erkrankung werden und damit in gewisser Weise eine „Karriere“ beschreiten können, bevorzugt das Institut den Begriff „Patientenwege“. Dieser Begriff erfasst möglichst viele der potenziellen Verläufe von Patientinnen und Patienten. Es soll herausgefunden werden, welche verschiedenen Wege es bei einer bestimmten Erkrankung gibt und welche unterschiedlichen Herausforderungen und Entscheidungen auf Patientinnen und Patienten zukommen.

Die Erstellung von Patientenwegen soll unter anderem helfen, den inhaltlichen Rahmen der Gesundheitsinformationen abzustecken. Dazu sollen folgende Fragen beantwortet werden:

- Wer könnte die Information lesen?
- Welche inhaltlichen Fragen könnten Leserinnen und Leser haben?
- In welchem emotionalen Zustand könnten sich die Leserinnen und Leser befinden?
- Zu welchem Zeitpunkt im Krankheitsverlauf werden welche Informationen möglicherweise angewendet?
- Vor welchen Entscheidungen stehen die Patientinnen und Patienten und wann müssen diese Entscheidungen getroffen werden?
- Was sind die möglichen Auswirkungen von Gesundheitsinformationen zu diesem Thema?

Ziel dieser Methode ist es primär, dass die Autorinnen und Autoren der Gesundheitsinformationen des Instituts ein gutes Verständnis für Patientinnen und Patienten sowie Angehörige und ihre Interaktion mit Information systematisch entwickeln; dies wird durch die Orientierung an den in Tabelle 8 angegebenen Dimensionen unterstützt.

Tabelle 8: Verschiedene Dimensionen eines Patientenweges

Leben / Alltag	Auswirkungen der Erkrankung auf soziale Beziehungen und Rollen: die Familie und Partnerschaft, den Beruf, die Lebensqualität, die „Funktionsfähigkeit“ etc.
Machen / Bewältigen	Jegliche Tätigkeiten im Zusammenhang mit der Erkrankung, zum Beispiel Arztbesuche, Einnahme von Medikamenten, Informationssuche, Selbsthilfe etc.
Fühlen	Empfindungen, die im Krankheits- und Behandlungsverlauf auftreten, wie Trauer, Ängste, Sorgen etc.
Wissen	Was wissen Bürgerinnen und Bürger bereits? Wo bestehen potenzielle Informationsbedürfnisse?
Entscheidungen	Welche Entscheidung muss der oder die Betroffene in der jeweiligen Phase treffen?
Klinik	Beschreibung der medizinischen Phasen, zum Beispiel Risikofaktoren, Symptome, Diagnose, Behandlung, Rehabilitation etc.
Kontaktstelle im Gesundheitswesen	Wer ist in der jeweiligen Phase Ansprechpartnerin oder Ansprechpartner im Gesundheits- oder Sozialsystem, zum Beispiel Ärztinnen und Ärzte, Pflegekräfte, Physiotherapeutinnen und Physiotherapeuten, Psychotherapeutinnen und Psychotherapeuten, Sozialarbeiterinnen und Sozialarbeiter, Beratungsstellen, Versicherungen?

Patientenwege können für die umfangreicheren Produkte (Merkblätter und Informationsberichte) erstellt werden. Es soll systematisch erfasst werden, welche Auswirkungen eine Erkrankung auf das Leben von Patientinnen und Patienten haben kann. Die Methode soll nachvollziehbar und reproduzierbar sein sowie den Versorgungsalltag widerspiegeln. Als Quellen hierfür dienen unter anderem Rapid Reviews qualitativer Studien, Erfahrungsberichte aus der DIPEX-Datenbank [128], Literatur zu Einflussfaktoren auf die Adhärenz, Literatur zu Informationsbedürfnissen von Patientinnen und Patienten, systematische Übersichten zur Kommunikation sowie Informationen zu Versorgungsfragen.

Zur Identifizierung wichtiger Diagnose- und Therapieschritte sowie anderer „Meilensteine“ auf dem Weg einer Patientin oder eines Patienten können (klinische) Behandlungspfade dienen. Behandlungspfade sind multidisziplinäre Versorgungs- und Behandlungspläne. Sie beschreiben, wie grundlegende diagnostische und therapeutische Schritte bei typischen Patientinnen und Patienten mit einer bestimmten Diagnose oder Erkrankung optimal aufeinander abgestimmt und organisiert werden. Ihr Ziel ist, die jeweilige Evidenz mit der Praxis zu verbinden sowie Patientenerwartungen und -präferenzen zu erkennen, um letztendlich eine optimale Versorgung zu ermöglichen [379,431].

5.2.3 Beteiligung von Bürgerinnen und Bürgern

Es gibt Hinweise darauf, dass die Beteiligung von Betroffenen an der Erstellung von Gesundheitsinformationen deren Relevanz steigern kann [393]. Die Orientierung an der Sichtweise und an den Informationsbedürfnissen der Bürgerinnen und Bürger zählt zu den Anforderungen an evidenzbasierte Gesundheitsinformationen [123] und ist für das Institut ein zentraler Bestandteil bei der Erstellung von Gesundheitsinformationen [558]. Dabei kommen verschiedene Maßnahmen zum Einsatz. Im Rahmen der Themenwahl werden etwa Themenvorschläge der Websitenutzerinnen und -nutzer, Erfahrungen im Rahmen von Konsultationen mit Selbsthilfegruppen und die Ergebnisse von Online-Umfragen auf der Website www.gesundheitsinformation.de berücksichtigt (siehe Abschnitt 5.3.1). Mit der Erhebung von Erfahrungsberichten von Patientinnen und Patienten sowie deren 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 (siehe Abschnitt 5.4.4). Im Rahmen der Qualitätssicherung werden ggf. Gutachten von Patientinnen und Patienten oder Patientenvertreterinnen und -vertretern zu Textentwürfen eingeholt. Eine weitere Maßnahme hinsichtlich der Einbindung der Öffentlichkeit erfolgt im Rahmen des begrenzten Stellungnahmeverfahrens. Dabei wird u. a. dem Kuratorium die Möglichkeit zur Kommentierung der Entwürfe aller Informationsberichte, Merkblätter und Kurzanworten gegeben. Dem Kuratorium gehören auch Vertreterinnen und Vertreter der für die Wahrnehmung der Interessen der Patientinnen und Patienten und von Selbsthilfe-Organisationen chronisch kranker und behinderter Menschen an. In der Regel zeitgleich mit dem Stellungnahmeverfahren durchlaufen alle Informationsberichte, Merkblätter und Kurzanworten eine externe Nutzertestung. Dabei kommentiert eine Gruppe von Betroffenen oder potenzieller Nutzerinnen und Nutzer die Texte im Hinblick auf ihren Informationsgehalt und ihre Verständlichkeit. Darüber hinaus haben die Nutzerinnen und Nutzer der Website www.gesundheitsinformation.de verschiedene Möglichkeiten der Rückmeldung gegenüber dem Institut. Die einzelnen Informationsprodukte können bewertet und kommentiert werden. Weiterhin erfolgt eine kontinuierliche Nutzerbefragung auf der Website (siehe Abschnitte 5.2.4 und 5.5.2)

5.2.4 Visuelle Kommunikation und Multimedia

Ein Text allein ist möglicherweise weniger verständlich und einprägsam als Informationen, in denen die Erklärungen durch Bilder ergänzt werden [156,258,312,342,514]. Auch gesprochener Text kann zum besseren Verständnis beitragen [258,454]. Die verständlichste Kommunikationsform – vor allem für Menschen mit geringerem Bildungsstand – könnten Erläuterungen darstellen, in denen Text, Bild und Ton miteinander kombiniert sind [258]. Wo immer es sinnvoll erscheint, 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 [147,342,467]. 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 Anwenderinnen und Anwender immer besser realisieren. Das Internet ermöglicht überdies eine Interaktion mit den Nutzerinnen und Nutzern, sodass die Kommunikation nicht nur in deren Richtung fließen muss. Interesse zu zeigen an dem, was den Patientinnen und Patienten wichtig ist, ist ein entscheidender Faktor in der patientenzentrierten und die Patientenkompetenz stärkenden Kommunikation [129,299,532]. Auch wenn das Institut keine individuelle Gesundheitsberatung durchführen kann, so verfügt es doch über verschiedene Möglichkeiten, seinen Nutzerinnen und Nutzern die Gelegenheit zu geben, ihre Ansichten und Sorgen mitzuteilen, und zwar u. a. durch:

- eine Online-Bewertung der Informationen,
- Themenvorschläge und ein allgemeines Online-Kontaktformular,
- ständige Umfragen zur Benutzerfreundlichkeit der Website und
- gelegentliche Online-Umfragen zu spezifischen Gesundheitsthemen [304].

5.2.5 Informationszugang und Barrierefreiheit

Das Internet hat sowohl besondere Vor- als auch Nachteile, was den Zugang zum Medium betrifft [32]. 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 [127,462]. 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 weiter an [329].

Denkbar ist, dass das Internet besonders häufig von chronisch Kranken benutzt wird. In Umfragen unter Patientinnen und 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 [407,424]. Mehr als ein Drittel (38 %) der Befragten hatte vor dem Klinikbesuch auf das Internet zugegriffen, um sich auf den bevorstehenden Arztbesuch vorzubereiten [407]. 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 Ärztinnen und Ärzte, Selbsthilfegruppen sowie Journalistinnen

und Journalisten nutzen das Internet routinemäßig. Auch im Familienkreis werden Informationen oft in breitem Umfang weitergegeben [469]. Geschlechtsspezifische Unterschiede im Zugang zum Internet waren in den Anfangsjahren des World Wide Web sehr deutlich, Ähnliches gilt für Unterschiede in Bezug auf das Alter und die formale Bildung; mittlerweile scheinen sie sich abzuschwächen [32,266].

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 der Nutzerinnen und des Nutzers)

Das Institut gewährleistet, dass seine Website die international anerkannten Zugänglichkeitsrichtlinien für Personen mit Behinderungen [540] sowie die deutsche BITV (Barrierefreie Informationstechnik-Verordnung) erfüllt [69]. 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.

Die Veröffentlichung von Pressemitteilungen trägt u. a. dazu bei, dass Inhalte der Gesundheitsinformationen über andere Medien Menschen erreichen, die das Internet nicht für die Suche nach Informationen zu Gesundheitsthemen nutzen.

Nahezu 10 % der in Deutschland lebenden Menschen sind anderer Nationalität, und weitere 10 % der Deutschen kommen aus Migrantenfamilien [491], von denen die größte kulturelle Gruppe aus der Türkei stammt [491]. Menschen mit nicht deutschsprachigem Hintergrund haben als Gruppe womöglich einen größeren Bedarf an Gesundheitsinformationen [105]. 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 veröffentlicht seine Gesundheitsinformationen sowohl in deutscher als auch in englischer Sprache. 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 Wissenschaftlerinnen und Wissenschaftler, insbesondere auch von der Begutachtung durch die Autorinnen und 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 [194]. Die Qualität von Übersetzungen anhand objektiver Kriterien zu beurteilen, ist aber schwierig. Daher ist es notwendig, dass die Übersetzungen von fachlich und sprachlich qualifizierten Personen vorgenommen und überprüft werden. Grundsätzlich können Texte auf verschiedene Arten übersetzt werden. Dies kann beispielsweise Wort für Wort geschehen oder die Übersetzung gibt den Inhalt des Ausgangstextes in der Zielsprache sinngemäß wieder [381]. Ziel der Übersetzung unserer Gesundheitsinformationen ist die Erstellung eines funktionalen und zielgruppengerechten Informationsangebotes. Dabei werden die Ausgangstexte sinngemäß in die Zielsprache übertragen. 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 Aufträgen 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 Bürgerinnen und Bürger sowie aus Eigeninitiative des Instituts gemäß dem Generalauftrag des G-BA.

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 Informationsquellen [28,29].

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. Demnach ist „eine kontinuierliche Beobachtung und Bewertung medizinischer Entwicklungen von grundlegender Bedeutung“ vorzunehmen und darüber zu berichten. Dieser Generalauftrag wurde für die Patienteninformationen des Instituts 2008 angepasst [197].

Mit dem Evidenzscanningsystem setzt das Institut die kontinuierliche Beobachtung und Bewertung der medizinischen Entwicklung um; dies wird ausführlicher in Abschnitt 5.3.4 beschrieben. 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 (Prävalenz/Inzidenz)
- 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 [73].

Der Gesetzgeber hat die Informationsaufgaben des Instituts an der Bürgerin und am Bürger orientiert. Darunter sind sowohl Gesunde als auch Patientinnen und Patienten zu verstehen. Von daher muss sein Informationsangebot auch Themen beinhalten, die nicht nur der Perspektive von Betroffenen, sondern auch von nicht direkt Betroffenen gerecht werden. Um seine Ziele zu erreichen, muss das Institut Informationen anbieten, die dazu befähigen, Auswahlentscheidungen zu treffen und Maßnahmen zu ergreifen, mit denen die Nutzerinnen und Nutzer ihre eigenen gesundheitlichen Ziele verwirklichen können [396], die das Selbstmanagement fördern und das ansprechen, was die Menschen wissen wollen [129,299,532]. Das Institut muss sich bei seiner Themenwahl daher auch danach richten, was Bürgerinnen und Bürger interessieren könnte und welche Gegebenheiten sie in der Versorgungswirklichkeit vorfinden.

Das Institut berücksichtigt diverse Quellen, um die Informationswünsche der Bürgerinnen und Bürger, gesund oder erkrankt, zu erschließen:

- Umfragen, qualitative Primärstudien und Übersichten von qualitativen Studien zum Informationsbedarf der Bevölkerung
- Anfragen an Callcenter der gesetzlichen Krankenkassen
- Erfahrungen von anderen Informationsanbietern, Patientenberatungsstellen und Selbsthilfegruppen
- Anfragen an die Beauftragte oder den Beauftragten der Bundesregierung für die Belange der Patientinnen und Patienten [463]

- Themen, die in die Suchmaschine der IQWiG-Website www.gesundheitsinformation.de eingegeben werden, sowie andere Daten zur Nutzung von Internetseiten
- Themenvorschläge von Websitenutzerinnen und -nutzern
- Ergebnisse der eigenen Online-Umfragen des Instituts zu Informationsbedürfnissen 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 bei der Beobachtung wird solchen Fragen eingeräumt, auf die es evidenzbasierte Antworten gibt. Die bei der Priorisierung von Themen berücksichtigten Aspekte sind in Tabelle 9 dargestellt. Zur Abschätzung dieser Aspekte werden unterschiedliche Quellen herangezogen. Nach der Themenauswahl wird der Institutsleitung eine interne Produktskizze für eine Gesundheitsinformation vorgelegt, die möglichst Informationen zu folgenden Sachverhalten enthält:

- das Produktvorhaben und die geplanten Botschaften
- Ergebnisse der Evidenzbewertung
- Begründungen zur epidemiologischen Bedeutung und Krankheitslast
- Angaben zur Relevanz aus dem McMaster Online Rating of Evidence (MORE) [238]
- Hinweis auf die Relevanz auf Grundlage deutscher Leitlinien

Die Methoden zur Themenauswahl werden zukünftig weiterentwickelt. Dabei wird auch die Evaluation der Vorgehensweise eine Rolle spielen. Die Aktualisierung der Gesundheitsinformationen bietet die Möglichkeit zu einer Überprüfung des Verfahrens bei der Themenauswahl. Im Zuge der Aktualisierung werden Besucherzahlen und Rückmeldungen der Nutzerinnen und Nutzer berücksichtigt (siehe Abschnitt 5.3.1). So können über die Jahre hinweg Themen, die auf weniger Resonanz stoßen, archiviert werden, was wiederum Kapazitäten für Themen von größerem Nutzerinteresse schafft. Ein weiterer wichtiger Indikator für eine erfolgreiche Themenauswahl wird sein, inwieweit das gewählte Themenspektrum auf der Website www.gesundheitsinformation.de die häufigsten Anfragen an Callcenter der gesetzlichen Krankenkassen abdeckt.

Tabelle 9: Aspekt zur Priorisierung der Themenauswahl

Evidenz	Redaktionelle Erwägungen	Patienten- / Nutzerinteresse
Systematische Übersichten zum Nutzen gesundheitsbezogener Interventionen	Ausgeglichenes Themenangebot	Patienten- / Nutzeranliegen
Evidenz zur Wirkung eines Informationsangebots zum Thema	Aktualität des Themas	Von Nutzerinnen und Nutzern gesuchte Informationen
	Mögliche unerwünschte Wirkungen der Gesundheitsinformation	Leser- / Nutzerinteresse weckendes Thema
	Prioritäten von Auftraggeberinnen und Auftraggebern	Nicht erfüllte Informationsbedürfnisse
	Arbeitsbelastung und Ressourcen	Krankheitslast
		Informationsbedürfnisse aus Expertensicht

5.3.2 Recherche

Das Institut stützt sich bei der Erstellung seiner Gesundheitsinformationen 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 Nutzerinnen und Nutzer von Interesse und Relevanz sind:

- Übersichten von qualitativen Studien sowie Rapid Appraisals qualitativer Primärstudien (siehe Abschnitt 7.4)
- Übersichten zu den Wirkungen von Kommunikation
- Übersichten zur Adhärenz
- kostenlos zugängliche Patienteninformationen im Internet sowie Internetseiten von Selbsthilfegruppen

Das Internet und andere Quellen werden zudem durchsucht, um Interventionen zu identifizieren, die von Bürgerinnen 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 Betroffene bei einem bestimmten gesundheitlichen Problem durchlaufen müssen, welche psychologischen und emotionalen Probleme im Zusammenhang mit diesem Thema auftreten

können und an welchen Punkten Entscheidungen getroffen werden müssen. Ferner werden gewöhnlich auch die Patientenvertreterinnen und -vertreter 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 ressortinternes Scoping dazu dient, auf die Bereiche zu fokussieren, mit denen sich die Gesundheitsinformation befassen wird. Bei der Erstellung von Informationsberichten werden die Scoping-Entscheidungen mit der Projektgruppe diskutiert und die Vorgaben zur Informationsbeschaffung berücksichtigt (siehe Kapitel 6). Die Recherchen schließen die Database of Reviews of Effects (DARE) [55,390], die Cochrane Database of Systematic Reviews (CDSR), die HTA Database des International Network of Agencies for Health Technology Assessment (INAHTA) sowie MEDLINE 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 [476,477].

Gelegentlich zieht das Institut eine Recherche nach Studien in Betracht. Wenn es beispielsweise zu einem wichtigen Thema keine neuere Übersicht gibt, wird eine Aktualisierungsrecherche erwogen, wenn eine hochwertige Übersicht vorliegt, deren Suche mehr als drei Jahre zurückliegt [476]. In einigen Fällen werden auch Aktualisierungen relevant, wenn die Zeitspanne diese drei Jahre 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 Cochrane Controlled Trials Register, in MEDLINE und in EMBASE durchgeführt [238]. Ergänzend können weitere Datenbanken hinzugezogen werden. Weiterhin werden Recherchen nach qualitativer Primärliteratur durchgeführt. Das Screening der Ergebnisse der Recherche erfolgt durch zwei Mitarbeiterinnen oder Mitarbeiter.

5.3.3 Evidenzbewertung

Die vom Institut erstellten Gesundheitsinformationen stützen sich zum großen Teil auf systematische Übersichten (siehe Abschnitt 7.2). Damit das Institut eine systematische Übersicht über die Wirkung einer Intervention für Gesundheitsinformationen verwenden kann, muss diese gewisse Mindestvoraussetzungen erfüllen, d. h., sie darf nach dem Oxman-Guyatt-Index nur geringe methodische Mängel aufweisen [284,400,402]. Um sich als Thema einer zusammenfassenden Kurzantwort, die einen Nutzen nahelegt, zu qualifizieren, sollte eine systematische Übersicht mindestens eine Studie umfassen, deren Qualität von den Autorinnen und Autoren der Übersicht 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 7.4).

Wird ein bestimmtes Thema oder Resultat in mehr als einer systematischen Übersicht von angemessener methodischer Qualität untersucht, wird eine weitere Qualitätsbewertung durchgeführt. Hierbei soll festgestellt werden, ob sich zwischen den Übersichten noch Qualitätsunterschiede zeigen oder einzelne Übersichten weniger geeignet sind. Dazu werden unter anderem folgende Aspekte verglichen:

- inhaltlicher Schwerpunkt des Reviews, insbesondere auch die Relevanz der Fragestellung für eine Patienteninformation
- Ausführlichkeit und Aktualität der Suche
- Durchführung von Sensitivitätsanalysen und Umgang mit Heterogenität
- Adressierung und Umgang mit ggf. vorhandenem Verzerrungspotenzial
- Angabe von Aktualisierungsfristen / regelmäßige Aktualisierung vorgesehen (beispielsweise Cochrane oder Agency for Healthcare Research and Quality [AHRQ])

Als Quelle für die in den Gesundheitsinformationen des Instituts benutzten numerischen Daten dienen die Ergebnisse der Übersicht, die im Hinblick auf ein Thema die höchste Qualität aufweist. Wenn Übersichten zu widersprüchlichen Schlussfolgerungen gelangen, wird nach den möglichen Gründen dafür gesucht [285].

Für die formale Bewertung der Stärke der Primärevidenz in einer ausgewählten systematischen Übersicht kommen die Methoden der GRADE-Arbeitsgruppe infrage [23]. 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 [204]. 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 [83,238]. Die Methoden des Instituts zur Bewertung qualitativer Studien werden in Abschnitt 7.4 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 [91] und die das deutsche Grundsatzpapier „Gute Praxis Gesundheitsinformation“ [123] vorsieht. Die Evidenz nimmt exponentiell zu. Das gilt sowohl

für Studien [33,516] als auch für systematische Übersichten [33,371]. Neue Evidenz kann dazu führen, dass existierende Übersichten als überholt bzw. nicht mehr aktuell gelten [189,451,477,534]; oft kommt es allerdings vor, dass die ursprünglichen Schlussfolgerungen durch die neue Evidenz nicht grundlegend verändert, sondern eher noch bekräftigt werden [276,418,502].

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 [476]. Bei manchen Themen, zu denen beispielsweise sehr starke Evidenz vorliegt, kann die Halbwertszeit der Evidenz deutlich über, in anderen Bereichen dagegen wiederum unter diesen drei Jahren liegen [477]. 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 Zeitintervall für eine Aktualitätsprüfung an und überwacht die Halbwertszeit seiner Gesundheitsinformationen, um so Erkenntnisse für zukünftige Aktualisierungsverfahren zu gewinnen.

Aktualisierungen können sehr ressourcenintensiv sein [287,372]. Schätzungen zufolge nimmt z. B. die vollständige Aktualisierung einer Leitlinie fast ebenso viel Zeit in Anspruch wie die Erstellung einer neuen Leitlinie [144]. 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 [309]. Dieses Ziel konnte bislang allerdings nicht erreicht werden; stattdessen verlieren die Übersichten vermutlich zunehmend an Aktualität [33]. 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 bereits erschöpft.

Das Institut geht nach folgendem Modell vor, um seine Informationen auf einem relativ aktuellen Stand zu halten:

A) Vorbereitung der regelmäßigen Aktualisierung

Vor der Veröffentlichung eines Informationsberichts, Merkblatts oder einer Kurzantwort werden im Rahmen einer Redaktionskonferenz die wichtigsten Erkenntnisse aus dem Erstellungsprozess festgehalten, die bei einer zukünftigen Aktualisierung zu beachten sind. Dabei wird das Datum festgelegt, zu dem die Information spätestens aktualisiert sein soll. In der Regel wird dieses Zeitintervall drei Jahre betragen.

B) Das Evidenzscanningsystem

„Evidenzscanning“ bedeutet, dass kontinuierlich alle neuen deutsch- oder englischsprachigen systematischen Übersichten, die bereits veröffentlichte Informationsprodukte oder Produkte im Erstellungsprozess betreffen könnten, identifiziert werden. Dazu werden folgende Quellen regelmäßig von zwei Mitarbeiterinnen oder Mitarbeitern gesichtet: Cochrane Database of Systematic Reviews (CDSR), Database of Abstracts of Reviews of Effects (DARE), HTA Database des International Network of Agencies for Health Technology Assessment (INAHTA), McMaster Online Rating of Evidence (MORE) und PubMed. Ferner werden Zulassungsbehörden aus Deutschland, Großbritannien und den USA sowie der European Medicines Agency (EMA) ständig bezüglich der Veröffentlichung von Warnhinweisen überwacht. Darüber hinaus werden Angebote von regelmäßig aktualisierten evidenzbasierten Informationen für die Ärzteschaft berücksichtigt; hierzu zählen Clinical Evidence und EBM Guidelines.

Jede potenziell relevante Übersicht, Studie oder Meldung, die eine oder mehrere Gesundheitsinformationen des Instituts betreffen könnte, wird für das weitere Vorgehen berücksichtigt und eingestuft, je nach Bedeutung und redaktionellem Interesse. Diese Einstufung kann z. B. einen sofortigen Aktualisierungsbedarf der Gesundheitsinformation auslösen, das Festhalten an der ursprünglichen Zeitplanung bestätigen oder zu einer Verschiebung des vorgesehenen Aktualisierungstermins führen.

C) Anstoß zu einer Aktualisierung

Als Anstoß zu Aktualisierungen können folgende Faktoren dienen:

- Leserbriefe oder Leserkommentare
- Interne oder externe Kritik
- Bedeutende neue Informationen aus dem Evidenzscanningsystem
- Erreichen des festgelegten Aktualisierungstermins

D) Geplante Aktualisierung

Sechs Monate vor Erreichen des festgelegten Aktualisierungstermins werden die Informationsprodukte hinsichtlich des zu erwartenden Überarbeitungsaufwands geprüft und die Aktualisierung wird geplant. Dabei werden berücksichtigt:

- die Zusammenfassung aus der Redaktionskonferenz zur ursprünglichen Veröffentlichung,
- Ergebnisse des Evidenzscanningsystems,
- geplante Aktualisierungen der zugrunde liegenden systematischen Übersichten durch die Autorinnen und Autoren (Autorenanfrage),
- zwischenzeitlich eingeführte Änderungen an den Methoden oder den redaktionellen Standards,

- Online-Bewertungen durch Nutzerinnen und Nutzer,
- Besucherzahlen,
- Rang der Information in Google-Suchen und
- die Notwendigkeit einer Nutzertestung, externen Begutachtung oder eines erneuten Stellungnahmeverfahrens.

Des Weiteren können zusätzliche Recherchen nach systematischen Übersichten und gegebenenfalls nach Studien für erforderlich gehalten werden. In der Regel werden im Rahmen einer Redaktionskonferenz die Entscheidungen über die vorgesehene Überarbeitung und weitere Recherchen bis hin zum Zurückziehen der Gesundheitsinformation getroffen.

Nach jeder Aktualisierung einer Gesundheitsinformation wird die Notwendigkeit einer Anpassung des festgelegten Aktualisierungstermins geprüft.

E) Qualitätssicherung der Aktualisierung

Wie weitreichend die einzelnen Qualitätssicherungsmaßnahmen sein sollen, hängt vom Überarbeitungsbedarf der jeweiligen Gesundheitsinformation ab. Die Aktualisierung eines Produkts mit geringen Änderungen aufgrund eines aktualisierten Cochrane Reviews wird beispielsweise nur eine interne Qualitätssicherung haben. Generell reichen die möglichen Maßnahmen von kleineren Korrekturen, die in der Verantwortung der Chefredaktion liegen, bis hin zu einer erneuten Begutachtung. Nach Veröffentlichung der aktualisierten Gesundheitsinformation werden alle Abonnenten des Newsletters informiert. Rückmeldungen von Leserinnen und Lesern können zu einer erneuten Aktualisierung führen.

F) Einstufung der Aktualisierung

Der Umfang jeder Aktualisierung eines Informationsberichts, Merkblatts oder einer Kurzantwort wird einer von drei Kategorien zugeordnet. Diese Zuordnung wird mit der Veröffentlichung der Aktualisierung auf der Website in der Historie angegeben: gering, mittel und groß (siehe Tabelle 10).

Tabelle 10: Kategorisierung der Aktualisierungen

Gering	Mittel	Groß
Kleinere Korrekturen oder sprachliche Verbesserungen des Textes, die die Botschaft der Gesundheitsinformation nicht betreffen.	Überarbeitung einer zugrunde liegenden systematischen Übersicht, die entweder zu keinen oder nur in geringem Umfang zu neuen Daten geführt hat. Die Botschaft der Gesundheitsinformation hat sich gar nicht oder nur geringfügig geändert. Austausch einer zugrunde liegenden systematischen Übersicht ohne eine wesentliche Änderung der Botschaft der Gesundheitsinformation.	Aktualisierung einer systematischen Übersicht oder der Austausch der zugrunde liegenden Übersicht mit wesentlichen Änderungen in der Botschaft der Gesundheitsinformation (z. B. die Änderung der Richtung eines Effekts). Das Zurückziehen einer zugrunde liegenden systematischen Übersicht oder der Gesundheitsinformation.

5.4 Informationsprodukte

5.4.1 Informationsberichte, Merkblätter und Kurzantworten

Die Gesundheitsinformationen für Bürgerinnen 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 Informationsprodukten des Instituts gehören:

- Informationsberichte: umfassende Beiträge, die die Grundlage für ein erweitertes Produktmodul zu einem relevanten gesundheitlichen Thema darstellen;
- Merkblätter: kurze, leicht verständliche Artikel; und
- Kurzantworten: Zusammenfassungen einzelner systematischer Reviews oder größerer Studien, darunter auch der von anderen Ressorts des Instituts erstellten.

Zusammengenommen bilden diese Produkte sowie die zusätzlichen Elemente ein evidenzbasiertes „Nachschlagewerk zur Gesundheit“. Studien zeigen, dass Informationssuchende sich mehrheitlich vor allem für Informationen über Behandlungen interessieren und für das, was sie selbst aktiv dazu beitragen können [173,298,373,439,478]. Außerdem fühlen sich deutsche Patientinnen und Patienten vor allem in Bezug auf diagnostische Untersuchungen und deren Ergebnisse nicht hinreichend informiert [523]. Darin spiegelt sich wider, dass ein Hauptschwerpunkt auf evidenzbasierten Informationen zu

Behandlungseffekten, diagnostischen Untersuchungen und Selbstmanagementstrategien liegen sollte. 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 Patientinformationen, 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 der stärker interessierten und motivierten Leserinnen und Lesern erkennen, die sich durch einen allzu einfachen Sprachstil bevormundet sehen könnten [129].

Die Inhalte der einzelnen Informationsberichte können je nach Thema variieren. Zu jedem Thema werden Informationen aus folgenden Bereichen gesichtet und auf Relevanz für den spezifischen Bericht überprüft:

- 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 Patientinnen und Patienten mit der betreffenden Erkrankung sowie von anderen Betroffenen, darunter Betreuungspersonen, Familienangehörige und Freunde

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 Internetnutzerinnen und -nutzer interessieren sich am meisten für Informationen zu den neuesten Forschungsergebnissen, und diese interessieren sie eventuell noch mehr als die Meinung von Expertinnen und Experten [416]. 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 2.1.5 beschrieben. Informationen zur Gesundheitsforschung sollten ein ähnlich hohes Qualitätssicherungsniveau gewährleisten wie der Forschungsbericht selbst [432]. Die Sicherung der Qualität und Relevanz von Patienteninformationen könnte auch im Rahmen von Patientenreviews erfolgen [393]. Informationsberichte und Merkblätter werden deshalb extern begutachtet, u. a. auch von Patientinnen und Patienten. Die Qualitätssicherung für alle Informationsberichte, Merkblätter und Kurzantworten umfasst auch, dass in der Regel einzelne Autorinnen und Autoren berücksichtigter systematischer Übersichten die Gelegenheit erhalten, sich zu diesen Patienteninformationen zu äußern. Die Entwürfe zu den Informationsberichten, Merkblättern und Kurzantworten werden im Rahmen des begrenzten Stellungnahmeverfahrens u. a. dem Kuratorium des Instituts vorgelegt, das den Patientenvertreterinnen und -vertretern in diesem Gremium ebenfalls die Gelegenheit gibt, diese Entwürfe zu kommentieren. Zusätzlich wird eine externe Nutzertestung aller Informationsberichte, Merkblätter und Kurzantworten und gelegentlich auch von zusätzlichen Elementen durchgeführt. Überdies sind auch die Patientinnen und Patienten, die über ihre Krankengeschichte und ihre Genesung berichten, eingeladen, sich zu den Entwürfen der Patienteninformationen, die zu ihren jeweiligen Geschichten gehören, zu äußern (siehe Abschnitt 5.4.4).

5.4.2 Zusätzliche Elemente

Die Informationsberichte, Merkblätter und Kurzantworten 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 5.2.4). Die Animationen stellen die am leichtesten verständlichen Informationsprodukte des Instituts dar.

Möglicherweise bevorzugen und vertrauen die Nutzerinnen und Nutzer auch eher solchen Internetseiten, die attraktiver gestaltet sind und Multimediaelemente enthalten [96,298,478]. Tatsächlich werden hochwertige Inhalte zuweilen allein aufgrund ihrer mangelhaften Präsentation abgelehnt [478].

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
- Erfahrungsberichte

Diese zusätzlichen Elemente sollen

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

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

5.4.3 Pressemitteilungen

Das Institut gibt zeitgleich mit der Veröffentlichung ausgewählter Gesundheitsinformationen Pressemitteilungen heraus. Die Pressemitteilungen sind ein wichtiges Mittel, um die Website www.gesundheitsinformation.de als Quelle für verlässliche patientenzentrierte Gesundheitsinformationen zu etablieren und die in Abschnitt 5.1 formulierten Institutsziele zu erreichen [213]. Die auf die Pressemitteilungen folgende Berichterstattung transportiert die Kernaussagen der Gesundheitsinformationen, weckt das Interesse am Informationsangebot des Instituts und erhöht die Bekanntheit der Website www.gesundheitsinformation.de. Über die Berichterstattung in verschiedenen Medien werden zudem Menschen angesprochen, die durch die Website allein nicht erreicht werden. Die Inhalte der Pressemitteilungen werden nach den gleichen Kommunikationsstandards erarbeitet wie die Gesundheitsinformationen des Instituts. Die resultierende Berichterstattung wird vom Ressort Gesundheitsinformation sowohl quantitativ als auch inhaltsanalytisch evaluiert. Zu welchen Gesundheitsinformationen Pressemitteilungen herausgegeben werden, hängt unter anderem davon ab,

- welche Krankheiten, diagnostischen Verfahren oder Therapien viele Menschen betreffen,
- welche Themen in den Medien und der allgemeinen Öffentlichkeit aktuell diskutiert werden,
- was Bürgerinnen und Bürger möglicherweise wissen möchten,
- ob die jeweiligen Forschungsergebnisse in der Öffentlichkeit bereits bekannt sind,
- ob bei einem bestimmten Thema die öffentliche Wahrnehmung den Ergebnissen der aktuellen Forschung widerspricht und / oder
- welche Themen besonders dafür geeignet sind, die allgemeine Gesundheitskompetenz und Wissenschaftskenntnisse zu fördern.

5.4.4 Erfahrungsberichte

Patientinnen und Patienten haben möglicherweise mehr Vertrauen in Gesundheitswebsites, wenn diese die Erfahrungen von Betroffenen beinhalten [478].

Viele Patientinnen und Patienten möchten von den Erfahrungen von anderen Betroffenen mit derselben Erkrankung hören oder darüber lesen [243,503]. 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 [205]. Die Bedeutung von Erfahrungsberichten in der medizinischen Praxis und im Gesundheitswesen wird zunehmend gewürdigt [210,494,556].

Einige Funktionen von Erfahrungsberichten sind Folgende [503]:

- Sie bieten die Möglichkeit, eigene Erfahrungen mit den Erfahrungen von anderen zu vergleichen.

- Das Lesen über die Gefühle anderer kann eine „Erlaubnis“ für das Zulassen ähnlicher Emotionen sein.
- Sie können Betroffenen zeigen, dass sie mit ihren Erfahrungen nicht allein sind.

Mit der Erhebung von Erfahrungsberichten von Patientinnen und Patienten sowie deren 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 nicht zu den evidenzbasierten Gesundheitsinformationen im Widerspruch stehen.

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 [128] im Internet kostenfrei nutzbar ist [243,244,556]. 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 Interviewpartnerinnen und -partner, insbesondere über Selbsthilfeorganisationen
- 2) Einholen der Einverständniserklärung 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

Besonderer Wert wird 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 Interviewerin oder des Interviewers sowie auf ein anonymisiertes Vorgehen gelegt. Zu jedem Informationsbericht sollen, wenn möglich, mindestens zwei Interviews geführt werden.

5.4.5 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 Websitestandards im Hinblick auf:

- Benutzerfreundlichkeit und Barrierefreiheit [264,318,392] (siehe Abschnitt 5.2.5)
- Datenschutz [269]
- Transparenz
- Sichtbarkeit für Suchmaschinen [509]
- Attraktivität für Nutzerinnen und 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 einzelner Informationsprodukte sowie die
- Hilfefunktionen und Funktionen, die eine Tour durch die Website ermöglichen.

Die Website bietet auch einen kostenlosen elektronischen Newsletter, den man wahlweise zweiwöchentlich 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 Nutzerin oder 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 [35,97,298].

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 [240]. Diese Stiftung mit Sitz in der Schweiz hat ein Programm zur Überprüfung und Zertifizierung der Standards von medizinischen Websites hinsichtlich des Inhalts, der Transparenz und des Datenschutzes 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, einschließlich eines Vergleichs mit ähnlichen Websites,
- Nutzerbewertungen und Nutzerfeedback, einschließlich der Reaktionen auf die laufenden Nutzerumfragen des Instituts,
- den Rang der Informationen im Rahmen von Google-Suchen und den Google PageRank der Website,
- 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 Zahl der Leserinnen und Leser oder Reichweite der Website [264,391]. Das Institut unterscheidet mehrere Hauptkategorien von Webmetriken:

- Messung des Websiteverkehrs („Traffic“), also der Anzahl von Personen, die die Website gezielt besuchen oder zufällig darauf stoßen
- Bestimmung der „Herkunft“ der Besucherinnen und Besucher (Suchmaschinen und Links von anderen Websites)
- Messung der Anzahl der besuchten Informationsseiten
- Reichweite der 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 [264,265]:

- die Anzahl der von Nutzerinnen und 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 Nutzerinnen und Nutzer nicht identifizieren kann. Die Datenschutzrichtlinien des Instituts sind auf der Website detailliert niedergelegt [269].

Der Website-Traffic stellt die Zahl der Personen dar, die insgesamt die Website überhaupt erreicht haben, nicht die Zahl der tatsächlichen Leserinnen und Leser [264]. 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 Glossar Begriffen 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
- Abgabe von Bewertungen zu den Gesundheitsinformationen

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 Leserinnen und Leser auch nach der Publikation auf seiner Website ständig zum Einreichen von Verbesserungsvorschlägen auf. 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 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 Informationsbericht, jedem Merkblatt und

jeder Kurzantwort wird online ein Dokumentationsverlauf erstellt, der über das Datum und den Grad der Änderungen Auskunft gibt, und jede Version wird archiviert.

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 Nutzerinnen und Nutzern
- Beratung mit Nutzerinnen, Nutzern und Patientenvertreterinnen und -vertretern
- unabhängige Evaluation

Bei einigen interaktiven Produkten ist das Testen einzelner Bestandteile von Informationen durch die Nutzerinnen und Nutzer Bestandteil der Arbeitsabläufe vor der Veröffentlichung. Diese Tests werden von den Mitarbeiterinnen und Mitarbeitern des Ressorts „Gesundheitsinformation“ durchgeführt. Offiziellere Nutzertests der Website werden außerhalb des Ressorts durchgeführt.

Mit der Freischaltung der Website (2006) 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 veröffentlicht [460]. Dadurch sollte ein Leistungsvergleich („Benchmarking“) zwischen der Website www.gesundheitsinformation.de und den offiziell evaluierten Modellprojekten (15 Websites) ermöglicht werden. Die Auswertung von 2561 ausgefüllten Fragebögen auf der Website 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 [97,99,313]. Bisher gibt es kein Instrument, das sich als zuverlässiger Indikator für die Qualität von Gesundheitsinformationen oder -websites erwiesen hat [166,193,286,378]. Die gebräuchlichen Instrumente weisen größere Lücken auf: DISCERN beispielsweise befasst sich nicht mit der inhaltlichen Qualität [97,313].

Befragungen von Patientinnen und Patienten, auch aus Deutschland [167,523], lassen darauf schließen, dass einige Aspekte, die bei der Evaluation von Gesundheitsinformationen manchmal als wichtig erachtet werden, den meisten Patientinnen und 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 Nutzerinnen und Nutzer als tatsächlich wichtig ansieht [416]. 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 Websiteanbieter werden leisten können.

Die externe Evaluation und vor allem die qualitative Evaluation durch potenzielle Nutzerinnen und Nutzer nimmt hinsichtlich der ständigen Weiterentwicklung der Informationen und der Website des Instituts einen hohen Stellenwert ein [106]. Das Institut beauftragt externe Sachverständige mit der inhaltlichen Evaluation einzelner Informationsprodukte und Informationspakete durch potenzielle Nutzerinnen und Nutzer [18,252,405].

Die externe Nutzertestung ist ein kontinuierliches Schlüsselement für die Qualitätssicherung der Gesundheitsinformationen vor ihrer Publikation; in der Regel erfolgt sie zeitgleich mit dem Stellungnahmeverfahren. Bei der Nutzertestung kommentieren mögliche Nutzerinnen und Nutzer, darunter auch Betroffene, die Textentwürfe im Hinblick auf ihren Informationsgehalt und ihre Verständlichkeit, sodass hier bereits zahlreiche Verbesserungen in die Endprodukte einfließen können. Wenn möglich, sollen auch alle Gesprächspartnerinnen und -partner bei der Erhebung von Erfahrungsberichten in die Evaluation der Informationsentwürfe einbezogen werden, an denen sie mitgearbeitet haben.

6 Informationsbeschaffung

Die Informationen, die die Grundlage der Institutsberichte darstellen, sind vielfältiger Art (z. B. Ergebnisse wissenschaftlicher Studien, Leitlinien, Registerdaten und andere Datensammlungen, Unterlagen von Zulassungsbehörden, Dossiers von pharmazeutischen Unternehmen). Im vorliegenden Kapitel wird der Ablauf einer themenbezogenen Recherche nach wissenschaftlicher Literatur beschrieben.

Im nachfolgenden Abschnitt 6.1 wird das Vorgehen des Instituts beschrieben, wenn die Informationsbeschaffung selbst durchgeführt wird. In Abschnitt 6.2 wird die Überprüfung einer Informationsbeschaffung, die durch andere durchgeführt wurde, beschrieben.

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

6.1 Eigene Informationsbeschaffung durch das Institut

Eine systematische Literaturrecherche hat zum Ziel, alle für die jeweilige Fragestellung **relevanten**, d. h. zum Erkenntnisgewinn beitragenden, Publikationen zu identifizieren. Die Literaturrecherche nach Primärliteratur orientiert sich dabei i. d. R. am Ziel einer hohen Sensitivität.

Basiert die Nutzenbewertung auf systematischen Übersichten, wird Vollständigkeit im Sinne einer vollständigen Berücksichtigung aller verfügbaren Primärstudien nicht angestrebt. Sollten sich Hinweise auf eine Unvollständigkeit der darin verwendeten Primärstudien ergeben, wird eine Nutzenbewertung auf Grundlage von Primärstudien erstellt, wenn die Robustheit der Ergebnisse nicht mehr gewährleistet ist. Unter Robustheit ist dabei eine ausreichende Sicherheit zu verstehen, dass sich das Ergebnis durch das Einbeziehen weiterer Informationen bzw. Studien nicht wesentlich verändern würde.

Folgende Aspekte müssen bei der systematischen Literaturrecherche vorab definiert werden:

- die Einschlusskriterien des Berichtsplans bzw. der Projektskizze, hinsichtlich
 - der medizinischen Kriterien (zum Beispiel Zielpopulation, Intervention),
 - des Studien- bzw. Leitlinientyps sowie der
 - formalen Merkmale der Publikation (zum Beispiel Abstractpublikation, Sprache etc.)
- die einzubeziehenden Datenquellen (z. B. bibliografische Datenbanken, Leitliniendatenbanken, Handsuche in ausgewählten Fachzeitschriften, Kontaktaufnahme mit Fachexpertinnen und Fachexperten, Industrie, Patientenorganisationen etc.)

Beispiele und Untersuchungen hierzu sind in einer Reihe von Publikationen zu finden [152,174,255,256,292,353,370,412,434,449,484]. 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.

6.1.1 Ablauf einer Recherche

Die Recherche in bibliografischen Datenbanken, Studienregistern sowie in Leitlinien-datenbanken und bei Leitlinienanbietern beinhaltet folgende Schritte:

- 1) ggf. Konkretisierung der Fragestellung
- 2) Umwandlung der Fragestellung in eine recherchierbare Fragestellung
- 3) Formulierung eines Recherchekonzepts
- 4) Auswahl der Datenbanken
- 5) Identifizierung von Suchwörtern
- 6) Formulierung der Suchstrategien
- 7) Qualitätssicherung (bei bibliografischer Recherche)
- 8) Durchführung der Suche
- 9) Speicherung der Ergebnisse in Textdateien und Import in ein Literaturverwaltungsprogramm (wenn ein standardisierter Export möglich ist)
- 10) Dokumentation der Recherche

Für die Identifizierung der Suchwörter und die Formulierung der Suchstrategie in bibliografischen Datenbanken 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 den Rückgriff auf Vorarbeiten anderer Arbeitsgruppen (systematische Übersichten zum Thema) generiert. Zusätzlich dazu findet eine formale interne Qualitätssicherung unter Berücksichtigung der Übersichtsarbeit von Sampson statt [450].

6.1.2 Bibliografische Datenbanken

A) Suche nach Primärliteratur

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 und EMBASE werden dabei immer ausgewählt (mindestens zwei große biomedizinische Datenbanken). Für die Erstellung von Gesundheitsinformationen erfolgt die Recherche nach qualitativen Studien zusätzlich in CINAHL und PsycInfo.

B) Suche nach systematischen Übersichten

Bei der Suche nach systematischen Übersichten sind teilweise andere Quellen zu berücksichtigen als bei der Suche nach Primärliteratur. Es wird regelhaft in Datenbanken, die ausschließlich oder überwiegend systematische Übersichten enthalten, recherchiert. Darüber hinaus wird in einer Auswahl biomedizinischer Datenbanken, die auch – aber nicht primär – systematische Übersichten enthalten (z. B. MEDLINE und EMBASE), recherchiert.

Welche Datenbanken oder sonstigen Quellen (z. B. Websites einzelner HTA-Institutionen) zusätzlich relevant und zu durchsuchen sind, ist themenabhängig zu entscheiden. Kostenpflichtige HTAs werden in Ausnahmefällen berücksichtigt, wenn z. B. anzunehmen ist, dass zusätzliche relevante Informationen daraus gewonnen werden können, oder ansonsten keine Informationen verfügbar sind.

6.1.3 Suche in Studienregistern

Die systematische Recherche soll nicht nur publizierte, sondern auch unpublizierte Studien identifizieren. „Unpubliziert“ bedeutet in diesem Zusammenhang, dass die Studien (bzw. einzelne Daten) nicht oder nur teilweise in Fachzeitschriften veröffentlicht wurden. Publikationen zu Studien werden i. d. R. über eine Recherche in bibliografischen Datenbanken wie MEDLINE oder EMBASE identifiziert. Bei der Suche nach unpublizierten Studien bzw. Daten kann auf Studienregister zurückgegriffen werden [332].

Die Nutzenbewertung des Instituts umfasst regelhaft die Suche in großen, allgemeinen Studienregistern sowie Metaregistern solcher Studienregister. Dazu gehören insbesondere das Studienregister Clinicaltrials.gov des US-amerikanischen National Institute of Health (NIH) sowie das Metaregister International Clinical Trials Registry Platform Search Portal (ICTRP) der WHO. Bei der Nutzenbewertung von Arzneimitteln werden zusätzlich Studienregister der pharmazeutischen Industrie (Studienregister einzelner Firmen und Metaregister) durchsucht. In krankheitsspezifischen Studienregistern wird nur im Ausnahmefall recherchiert. Das bei der Europäischen Zulassungsbehörde EMA angesiedelte europäische Studienregister (clinicaltrialsregister.eu) stellt wegen fehlender Funktionalitäten [548] und aufgrund des unvollständigen Datenbestands in der derzeitigen Form keine Quelle dar, die regelhaft bei der Nutzenbewertung des IQWiG herangezogen wird.

Neben der Information über die Existenz der Studie enthalten einige Register zunehmend auch Studienergebnisse. Dies gilt z. B. für das Register Clinicaltrials.gov und Studienregister der pharmazeutischen Industrie. Ergebnisse aus Studienregistern können, die grundsätzliche Relevanz der jeweiligen Studie vorausgesetzt, berücksichtigt werden.

6.1.4 Leitliniendatenbanken und -anbieter

Hat eine Recherche zum Ziel, Leitlinien zu identifizieren, erfolgt die Suche primär in Leitliniendatenbanken (z. B. des Guidelines International Network [G-I-N], der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften [AWMF] oder des

National Guideline Clearinghouse [NGC]), gegebenenfalls gefolgt von fachspezifischen bzw. fachübergreifenden Leitlinienanbietern. Ob ergänzend dazu eine Recherche nach Leitlinien in bibliografischen Datenbanken erfolgt, ist abhängig von der Art des zu erarbeitenden Berichts.

Bei der Suche in Leitliniendatenbanken und bei Leitlinienanbietern richtet sich die anzuwendende Suchstrategie nach dem Aufbau und den Möglichkeiten der jeweiligen Internetseiten. Nur wenige Seiten ermöglichen eine Suche mit Schlagwörtern, sodass in der Regel jeweils die gesamte Liste von veröffentlichten Leitlinien durchgesehen wird. Außerdem ist bei der Suche in Leitliniendatenbanken und bei Leitlinienanbietern häufig kein standardisierter Export möglich. Aus diesem Grund wird die Dokumentation der Recherche und der Anzahl der Treffer in einem standardisierten Rechercheprotokoll vorgenommen. Die Erfassung der potenziell relevanten Treffer erfolgt in einem Literaturverwaltungsprogramm. Ansonsten erfolgt der Ablauf wie in Abschnitt 6.1.1 dargestellt.

Innerhalb der Nutzenbewertung werden Leitlinien als Informationsquelle nicht grundsätzlich ausgeschlossen. Es erfolgt jedoch i. d. R. keine systematische Recherche nach Leitlinien.

6.1.5 Herstelleranfragen

Bei der Nutzenbewertung des Instituts werden in der Regel die Hersteller der zu bewertenden Technologie(n) bezüglich bislang nicht veröffentlichter Informationen angefragt. Ziel dieser Anfrage ist die Identifizierung aller für die Nutzenbewertung relevanten Studien bzw. Informationen unabhängig vom Publikationsstatus. Im Rahmen der Arzneimittelbewertung wird diese Anfrage in der Regel in einem zweistufigen Verfahren durchgeführt. In der ersten Stufe fragt das Institut beim Hersteller eine vollständige Übersicht 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 [270]). 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 die selektive Bereitstellung von Informationen vermieden werden.

6.1.6 Sonstige Datenquellen für die Recherche

A) Kongressbände und ausgewählte Fachzeitschriften

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

B) Öffentlich zugängliche Dokumente von Zulassungsbehörden

Im Falle einer Arzneimittelbewertung, aber auch für die Bewertung bestimmter (nicht-medikamentöser) Medizinprodukte sind öffentlich zugängliche Zulassungsdatenbanken bzw. -korrespondenzen weitere mögliche Quellen für die Informationsgewinnung.

C) Informationen von Autorinnen und Autoren einzelner Publikationen

Im Rahmen der Leitlinien- bzw. Nutzenbewertung kann es im Einzelfall sinnvoll sein, Kontakt zu Autorinnen und Autoren bzw. Leitlinienerstellerinnen und Leitlinienerstellern aufzunehmen. Die Anfragen können sich z. B. auf spezielle Detailinformationen zu einzelnen Leitlinien oder auf nicht publizierte Teilaspekte zu Publikationen beziehen.

D) Durch den G-BA oder das BMG übermittelte Unterlagen

Sofern vom jeweiligen Auftraggeber (G-BA, 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.

6.1.7 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. in 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.

- Bei der Suche in Leitliniendatenbanken und bei Leitlinienanbietern erfolgt je nach Fragestellung ein ergänzender dritter Schritt, in dem geprüft wird, ob bei der Generierung und Formulierung der Leitlinie eine methodische Systematik zur Anwendung kam. In der Regel ist hiermit die Evidenzbasierung der Leitlinie gemeint (siehe Abschnitt 4.2). Das Institut legt im Rahmen der Erstellung des Berichtsplans vorab fest, ob aufgrund der Fragestellung nur evidenzbasierte Leitlinien in dem jeweiligen Bericht berücksichtigt werden sollen.

Alle Selektionsschritte werden 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.

Bei der Recherche nach Leitlinien werden die Schritte vom Volltextscreening an (ab dem zweiten Screening) von zwei Personen unabhängig voneinander ausgeführt. Die Qualitätssicherung des ersten Screeningschrittes findet systembedingt mithilfe des standardisierten Rechercheprotokolls statt.

6.1.8 Dokumentation der Recherche

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 als 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 Autorinnen und Autoren, Anfragen an Herstellerinnen und Hersteller etc.).

6.1.9 Ergänzungsrecherchen bei systematischen Übersichten

In den meisten Fällen wird eine ergänzende Literaturrecherche nach aktuellen Primärstudien erforderlich sein (Ergänzungsrecherche), die den Zeitraum zwischen den in der systematischen Übersicht durchgeführten Recherchen und der Erstellung des IQWiG-Berichts abdeckt. Auf eine ergänzende aktualisierende Primärliteraturrecherche kann bei der Nutzenbewertung anhand von systematischen Übersichten nur in begründeten Ausnahmefällen verzichtet werden. Dies gilt z. B., wenn eine ausreichende Sicherheit besteht, dass sich das Ergebnis durch das Einbeziehen weiterer Informationen bzw. Studien nicht wesentlich verändern würde (Robustheit).

Zudem kann es für spezifische, nicht in der systematischen Übersicht behandelte Fragestellungen notwendig sein, ergänzende Recherchen nach Primärliteratur durchzuführen.

6.2 Überprüfung der Informationsbeschaffung für Dossiers

Bei der Dossierbewertung erfolgt primär keine eigene Informationsbeschaffung; stattdessen wird eine Überprüfung der Informationsbeschaffung in den Dossiers durchgeführt.

Für die Erstellung eines Dossiers ist eine Suche in bibliografischen Datenbanken nicht immer erforderlich. Eine Suche in öffentlich zugänglichen Studienregistern ist hingegen grundsätzlich durch den pharmazeutischen Unternehmer durchzuführen; Näheres beschreibt die Verfahrensordnung des G-BA [198].

Bei der Dossierbewertung findet zum einen eine formale und inhaltliche Prüfung der Informationsbeschaffung statt. Diese bezieht sich auf die Suche in bibliografischen Datenbanken sowie die Suche in Studienregistern und stützt sich auf das in Abschnitt 6.1.1 beschriebene Vorgehen zur Qualitätssicherung sowie die Dokumentvorlagen zur Erstellung von Dossiers in den Vorgaben der Verfahrensordnung des G-BA [198].

Je nachdem, welche Ergebnisse die formale und inhaltliche Prüfung ergibt, schließt sich zum anderen eine eigene Recherche und / oder Studienselektion des Instituts zur Überprüfung der Vollständigkeit der Informationsbeschaffung an. Hierzu stehen verschiedene Strategien zur Verfügung, wie z. B. die stichprobenartige Überprüfung der im Dossier ausgeschlossenen Literaturzitate, die Anwendung spezieller Funktionen von Literaturdatenbanken (z. B. „Related Articles“-Feature in PubMed [452]) sowie die Durchführung einer vollständigen Literaturrecherche. Das Ergebnis der Überprüfung der Informationsbeschaffung und die Beschreibung des diesbezüglichen Vorgehens sind Teil der Dossierbewertung.

7 Informationsbewertung

In der Forschung bezeichnet der Begriff „Bias“ (Verzerrung) eine systematische Abweichung zwischen den Forschungsergebnissen und der „Wahrheit“ [444]. 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 [89]. Alle diese Methoden bilden auch die gesetzliche Grundlage für die Arbeit des Instituts.

7.1 Qualitätsbewertung von Einzelstudien

7.1.1 Kriterien zum Einschluss von Studien

Häufig ergibt sich das Problem, dass für eine Nutzenbewertung relevante Studien die Einschlusskriterien für die in der systematischen Übersicht definierte Patientenpopulation bzw. die definierten Prüf- und Vergleichsinterventionen nicht vollständig erfüllen. Das Institut verfährt hier in der Regel nach folgenden Kriterien:

Für das Einschlusskriterium bezüglich Population reicht es aus, wenn bei mindestens 80 % der in der Studie eingeschlossenen Patientinnen und Patienten dieses Kriterium erfüllt ist. Liegen für solche Studien entsprechende Subgruppenanalysen vor, wird auf diese Analysen zurückgegriffen. Studien, bei denen das Einschlusskriterium bezüglich Population bei weniger als 80 % der in der Studie eingeschlossenen Patientinnen und Patienten erfüllt ist, werden nur dann eingeschlossen, wenn entsprechende Subgruppenanalysen vorliegen oder wenn hinreichend sicher plausibel oder nachgewiesen ist, dass die in dieser Studie gewonnenen Ergebnisse auf die Zielpopulation der systematischen Übersicht anwendbar sind (zur Anwendbarkeit siehe Abschnitt 3.3.1).

Ebenfalls eingeschlossen werden Studien, die zu mindestens 80 % das Einschlusskriterium bezüglich Prüfintervention (Interventionsgruppe der Studie) erfüllen und zu mindestens 80 % das Einschlusskriterium bezüglich Vergleichsintervention (Vergleichsgruppe der Studie). Ist in einer Studie eines der beiden Kriterien verletzt, wird die Studie aus der Nutzenbewertung ausgeschlossen.

7.1.2 Zusammenhang 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 3.2.1). Gütekriterien, die die Aussagekraft kontrollierter Studien erhöhen, sind Randomisierung und Verblindung. Häufige Designs bei klinischen Studien bilden parallele Gruppen [414], Crossover-Studien [289] und clusterrandomisierte Studien [138]. Sind in einer Studie Zwischenauswertungen geplant, so muss dies mithilfe eines adäquaten sequenziellen Designs berücksichtigt werden [544].

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 für die Schätzung von Krankheitsprävalenzen. Weitere grundlegende klassische Studientypen der Epidemiologie sind Fall-Kontroll-Studien [57] zur Untersuchung des Zusammenhangs zwischen Expositionen und seltenen Erkrankungen und Kohortenstudien [58] 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 [294].

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).

7.1.3 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 [223,228]. Es gibt jedoch kein System der Evidenzbewertung, das allgemein akzeptiert und universell in allen systematischen Übersichten einsetzbar ist [295]. 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, weitgehend mit der Evidenzklassifizierung des G-BA übereinstimmt [198] und Eingang in die Verordnung über die Nutzenbewertung von Arzneimitteln nach § 35a SGB V [70] gefunden hat. 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 Autorinnen und Autoren wiederum kritisiert wird [553]. 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 [228].

7.1.4 Aspekte der Bewertung des Verzerrungspotenzials

Ein wesentlicher Gesichtspunkt bei der Interpretation von Studienergebnissen ist die Einschätzung von Verzerrungsaspekten (vgl. qualitative Ergebnisunsicherheit, Abschnitt 3.1.4). 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 möglicher Verzerrungen 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) [275,310] und bei epidemiologischen Studien die Leitlinien und Empfehlungen zur Sicherung „Guter Epidemiologischer Praxis“ (GEP) [116]. 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 kontrollierte Versuche [465] und das dazugehörige Erläuterungsdokument [368],
- ein Vorschlag zur Erweiterung des CONSORT-Statements für randomisierte Studien bei nichtmedikamentösen Verfahren [54] und das zugehörige Erläuterungsdokument [53],

- das CONSORT-Statement für clusterrandomisierte Studien [81],
- das CONSORT-Statement zur Dokumentation unerwünschter Ereignisse [277],
- das CONSORT-Statement für Nichtunterlegenheits- und Äquivalenzstudien [413],
- das CONSORT-Statement für pragmatische Versuche [559],
- das PRISMA(Preferred Reporting Items for Systematic Reviews and Meta-Analyses)-Statement für Meta-Analysen randomisierter Studien [369] und das zugehörige Erläuterungsdokument [337],
- das TREND(Transparent Reporting of Evaluations with Non-randomized Designs)-Statement für nicht randomisierte Interventionsstudien [113],
- das STROBE(Strengthening the Reporting of Observational Studies in Epidemiology)-Statement für Beobachtungsstudien in der Epidemiologie [535] und das dazugehörige Erläuterungsdokument [527],
- die MOOSE(Meta-analysis of Observational Studies in Epidemiology)-Checkliste für Meta-Analysen von Beobachtungsstudien in der Epidemiologie [497] und
- das STARD(Standards for Reporting of Diagnostic Accuracy)-Statement für Diagnosestudien [51] und das dazugehörige Erläuterungsdokument [52].

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 Einschätzung möglicher Verzerrungen beschreiben, sind die Standardwerke [150,222,248].

Zentrale Aspekte zur Bewertung des Verzerrungspotenzials der Ergebnisse randomisierter kontrollierter Studien (RCTs) durch das Institut stellen ein adäquates Concealment, d. h. die Unvorhersehbarkeit und Verheimlichung der Gruppenzuteilung (z. B. durch externe Randomisierung bei nicht verblindbaren Studien), die verblindete Zielgrößenerhebung bei Studien, in denen Ärztin oder Arzt und Patientin oder Patient nicht verblindbar sind, und die adäquate Anwendung des Intention-to-Treat-Prinzips dar.

Die Interpretation der Ergebnisse von unverblindeten Studien beziehungsweise Studien, in denen es (möglicherweise) zu einer Entblindung kam, muss im Vergleich zu der verblindeter 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 von den (unverblindeten) behandelnden Personen ist bzw. unabhängig von den behandelnden Personen verblindet 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.

Im Rahmen der Berichterstellung werden i. d. R. standardisierte Bewertungsbögen zur Bewertung der Verzerrungsaspekte von Studienergebnissen verwendet. Für kontrollierte Studien zur Nutzenbewertung von Interventionen werden insbesondere folgende endpunktübergreifende und endpunktspezifische Punkte regelhaft herangezogen:

Endpunktübergreifende Punkte:

- adäquate Erzeugung der Randomisierungssequenz (bei randomisierten Studien)
- Verdeckung der Gruppenzuteilung (Allocation Concealment; bei randomisierten Studien)
- zeitliche Parallelität der Interventionsgruppen (bei nicht randomisierten Studien)
- Vergleichbarkeit der Interventionsgruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren (bei nicht randomisierten Studien)
- Verblindung der Patienten und behandelnden Personen bzw. weiterbehandelten Personen
- ergebnisunabhängige Berichterstattung aller relevanten Endpunkte

Endpunktspezifische Punkte:

- Verblindung der Endpunkterheber
- adäquate Umsetzung des Intention-to-Treat-Prinzips
- ergebnisunabhängige Berichterstattung einzelner Endpunkte

Für randomisierte Studien wird anhand dieser Aspekte das Verzerrungspotenzial zusammenfassend als „niedrig“ oder „hoch“ eingestuft. Ein niedriges Verzerrungspotenzial liegt dann vor, wenn mit großer Wahrscheinlichkeit ausgeschlossen werden kann, dass die Ergebnisse relevant verzerrt sind. Unter einer relevanten Verzerrung ist zu verstehen, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

Für die Bewertung eines Endpunkts wird zunächst das Verzerrungspotenzial endpunktübergreifend als „niedrig“ oder „hoch“ eingestuft. Falls diese Einstufung als „hoch“ erfolgt, wird das Verzerrungspotenzial für den Endpunkt i. d. R. auch als „hoch“ bewertet. Ansonsten finden die endpunktspezifischen Punkte Berücksichtigung.

Eine Einstufung des Verzerrungspotenzials des Ergebnisses für einen Endpunkt als „hoch“ führt nicht zum Ausschluss aus der Nutzenbewertung. Die Klassifizierung dient vielmehr der Diskussion heterogener Studienergebnisse und beeinflusst die Sicherheit der Aussage.

Für nicht randomisierte vergleichende Studien wird in der Regel keine zusammenfassende Bewertung der Verzerrungsaspekte durchgeführt, da die Ergebnisse dieser Studien aufgrund der fehlenden Randomisierung generell ein hohes Verzerrungspotenzial besitzen.

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.

7.1.5 Interpretation von kombinierten Endpunkten

Ein „kombinierter Endpunkt“ umfasst eine von den Untersuchern definierte Gruppe von Ereignissen (z. B. Herzinfarkte, Schlaganfälle, kardiovaskuläre Todesfälle). Dabei unterscheiden sich die einzelnen Ereignisse dieser Gruppe nicht selten in ihrem Schweregrad und in ihrer Bedeutung für Patientinnen und Patienten und Ärztinnen und Ärzte (z. B. Krankenhausaufnahme und kardiovaskuläre Todesfälle). Daher muss man sich bei der Interpretation kombinierter Endpunkte der mit dieser Tatsache verbundenen Konsequenzen bewusst sein [94,175,188]. Die folgenden Ausführungen beschreiben die bei der Ergebnisinterpretation zu berücksichtigenden Aspekte. Sie beziehen sich aber ausdrücklich nicht auf die (ggf. abschließende) Nutzen- / Schadenbewertung mittels kombinierter Endpunkte, wenn beispielsweise der mögliche Schaden einer Intervention (z. B. Häufung schwerer Blutungen) mit dem Nutzen (Verminderung von Herzinfarkten) in einem Endpunkt erfasst wird.

Eine Voraussetzung für die Berücksichtigung eines kombinierten Endpunkts ist, dass die einzelnen Komponenten des kombinierten Endpunkts sämtliche im Berichtsplan definierte patientenrelevante Endpunkte darstellen. Hierbei können nur dann Surrogatendpunkte eingeschlossen werden, wenn sie vom Institut dezidiert als valide akzeptiert sind (siehe Abschnitt 3.1.2). Das Resultat für jedes im kombinierten Endpunkt zusammengefasste Einzelereignis soll auch isoliert berichtet werden. Die Komponenten sollen von ähnlicher „Schwere“ sein. D. h. nicht, dass sie von identischer Bedeutung sein müssen. Beispielsweise kann Mortalität mit symptomatischem Herzinfarkt oder Schlaganfall kombiniert werden, jedoch nicht mit „stummen“ Herzinfarkt oder „stationärer Aufnahme“.

Erfüllt ein kombinierter Endpunkt die oben genannten Voraussetzungen, dann sind bei der Interpretation für Aussagen zum Nutzen und Schaden folgende Aspekte zu beachten:

- Ist der Effekt der Intervention auf die einzelnen Komponenten des kombinierten Endpunkts in der Regel gleichgerichtet?
- Wurde ein relevanter, eigentlich in die Kombination passender Endpunkt ohne nachvollziehbare und akzeptable Begründung nicht eingeschlossen oder ausgeschlossen?
- Wurde der kombinierte Endpunkt vorab definiert oder post hoc eingeführt?

Sofern es die vorliegenden Daten und Datenstrukturen erlauben, können Sensitivitätsanalysen unter Ausschluss versus Hinzunahme einzelner Komponenten durchgeführt werden.

Bei entsprechenden Voraussetzungen können im Rahmen der Nutzenbewertung einzelne Endpunkte aus einem kombinierten Endpunkt ermittelt und berechnet werden.

7.1.6 Interpretation von Subgruppenanalysen

Subgruppenanalysen werden in der methodischen Literatur sehr kritisch diskutiert [22,401]. Die Interpretation der Ergebnisse von Subgruppenanalysen auf Studienebene 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 bezüglich mehr oder weniger arbiträrer subgruppenbildender Merkmale post hoc stattfinden, können ihre Ergebnisse nicht als methodisch korrekte Prüfung einer Hypothese betrachtet werden.
- **Multipltes Testen:** Bei der Analyse zahlreicher Subgruppen besteht eine manchmal recht hohe Wahrscheinlichkeit, dass die 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 [64].

Die Ergebnisse von Subgruppenanalysen sollten unter Beachtung der drei oben angesprochenen Aspekte in Bewertungen einfließen und nicht das Ergebnis der primären Analyse dominieren; dies gilt umso mehr, falls das primäre Studienziel nicht erreicht wurde. Eine Ausnahme von dieser Regel kann dann vorliegen, wenn sozialrechtliche Implikationen (s. u.) solche Analysen erforderlich machen. Darüber hinaus sind Subgruppenanalysen nicht oder nur sehr eingeschränkt interpretierbar, wenn das subgruppenbildende Merkmal nach Therapiebeginn (nach Randomisierung) erhoben wurde, z. B. bei sogenannten Respon-

deranalysen. Diese Aspekte spielen auch bei der Durchführung und Interpretation von Subgruppenanalysen im Rahmen von Meta-Analysen eine Rolle (siehe Abschnitt 7.3.8).

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.

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 [98].

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 [186], 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 [430]. Sun et al. [500] identifizieren Kriterien zur Bewertung der Verlässlichkeit von Subgruppenanalysen.

Eine wichtige Indikation für Subgruppenanalysen stellt eine mögliche Heterogenität des Effekts in verschiedenen, klar voneinander abgrenzbaren Patientenpopulationen dar [314,430]. 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 [314]. 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 [216]. Im Rahmen von systematischen Übersichten ist die Analyse der Heterogenität zwischen den einzelnen Studien und damit ggf. Subgruppenanalysen eine wissenschaftliche (siehe Abschnitt 7.3.8), aber auch sozialrechtliche Notwendigkeit, da das Institut nach § 139a Abs. 2 SGB V gehalten ist, alters-, geschlechts- und lebenslagenspezifische Besonderheiten zu berücksichtigen. Darüber hinaus soll das Institut gemäß der amtlichen Begründung für das GKV-Modernisierungsgesetz (GMG) auch erarbeiten, für welche Patientengruppen ein neues Arzneimittel eine maßgebliche Verbesserung des Behandlungserfolgs erwarten lässt mit dem Ziel, dass diese Patientinnen und Patienten das neue Arzneimittel erhalten sollen [118]. Eine entsprechende Zielsetzung findet sich auch in § 35a SGB V zur Bewertung des Nutzens von Arzneimitteln mit neuen Wirkstoffen [120]. Bei dieser Bewertung sollen Patientengruppen abgegrenzt werden, für die ein therapeutisch bedeutsamer Zusatznutzen besteht. Eine weitere sozialrechtliche Notwendigkeit für Subgruppenanalysen kann sich durch den Zulassungsstatus von Arzneimitteln

ergeben. Zum einen kann es Folge der Abwägungsentscheidung von Wirksamkeit und Risiken durch die Zulassungsbehörden sein, dass das Arzneimittel nur für einen Teil der in den Zulassungsstudien untersuchten Patientenpopulationen zugelassen wird. Den Abwägungen können dabei auch post hoc durchgeführte Subgruppenanalysen zugrunde liegen. Zum anderen können auch Studien, die nach der Zulassung durchgeführt wurden, Patientengruppen einschließen, für die das Arzneimittel in Deutschland nicht zugelassen ist. Dies gilt umso mehr, je stärker sich die Zulassungen international voneinander unterscheiden. In solchen Fällen muss ggf. auf Subgruppenanalysen zurückgegriffen werden, die den Zulassungsstatus des Arzneimittels abbilden, und zwar unabhängig davon, ob diese Analysen a priori geplant waren oder post hoc durchgeführt wurden.

7.1.7 Bewertung der Konsistenz von Daten

Zur Einschätzung der Aussagekraft der Studienergebnisse 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 [9]. Inkonsistenzen können innerhalb einer Publikation bestehen, aber auch zwischen verschiedenen Publikationen zu derselben Studie.

Ein Problem vieler Publikationen sind unvollständige Angaben im Methodik- und Ergebnisteil. Insbesondere die Darstellung von Lost-to-Follow-up-Patienten, Therapieabbrechern 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 Patientinnen und 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 Patientinnen und 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 [9].

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 Autorinnen und 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.

7.2 Berücksichtigung systematischer Übersichten

Systematische Übersichten sind Arbeiten, die auf systematische, reproduzierbare und transparente Weise Ergebnisse von Primärstudien zusammenfassend darstellen und bewerten. Dies sind auch HTA-Berichte, in denen normalerweise versucht wird, eine klinische und / oder eine für Patientinnen und Patienten relevante Frage zu beantworten. HTA-Berichte versuchen oftmals auch, Antworten auf zusätzliche Fragen zu geben, die für die Auftraggeber / Entscheidungsträger von Interesse sind [139,333,408]. Da im Folgenden nicht zwischen systematischen Übersichten und HTA-Berichten unterschieden werden muss, bezieht der Begriff „systematische Übersicht“ HTA-Berichte mit ein.

7.2.1 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 [1]. 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 [139,150,204,408].

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 [1,150,248].

Im Falle von systematischen Übersichten über die Wirkungen medizinischer Interventionen 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 [204]. Systematische Übersichten über Diagnose- und Screeningtests weisen im Vergleich zu Übersichten über Behandlungsmaßnahmen einige methodische Unterschiede auf [107].

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 7.2.2). Die vom Institut für Patientinnen und Patienten und Bürgerinnen 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 [284,400,402] oder anhand des AMSTAR-Instruments [473-475] nur geringe methodische Mängel aufweisen. 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 [47,103].

7.2.2 Nutzenbewertung auf Basis systematischer Übersichten

Eine Nutzenbewertung auf Basis systematischer Übersichten kann – sofern bestimmte Voraussetzungen erfüllt sind – eine ressourcensparende und verlässliche Evidenzbasis für Empfehlungen an den G-BA bzw. das BMG liefern [95,330]. Für die Verwendung systematischer Übersichten zur Nutzenbewertung ist es notwendig, dass die systematischen Übersichten von ausreichend hoher Qualität sind, d. h.

- nur minimales Verzerrungspotenzial aufweisen sowie
- die Evidenzlage vollständig, transparent und nachvollziehbar darstellen

und so erlauben, eine klare Schlussfolgerung zu ziehen [23,400,547]. Außerdem ist es eine notwendige Voraussetzung, dass die jeweilige Suche der Methodik des Instituts nicht widerspricht und die Übertragbarkeit der Ergebnisse auf die jeweilige Fragestellung, insbesondere unter Berücksichtigung der genannten Ein- und Ausschlusskriterien, möglich ist.

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 systematischer Übersichten. 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 beruhen müssen. Je nach Fragestellung ist es möglich, dass bestimmte Teile eines Auftrags auf Basis von systematischen Übersichten beantwortet werden, während für andere Teile auf die Primärliteratur zurückgegriffen wird.

B) Mindestzahl relevanter systematischer Übersichten

Alle qualitativ ausreichenden und thematisch relevanten systematischen Übersichten 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 erforderlich.

C) Qualitätsbewertung der Publikationen einschließlich Mindestanforderungen

Die Bewertung der allgemeinen Qualität systematischer Übersichten erfolgt anhand des validierten Qualitätsindex von Oxman und Guyatt [399,400,402] oder anhand des AMSTAR-Instruments [473-475]. Systematische Übersichten gelten nach dem Oxman-Guyatt-Index dann als qualitativ ausreichend, wenn diese von zwei voneinander unabhängigen Gutachtern in der Gesamtbewertung mit mindestens fünf von sieben möglichen Punkten bewertet wurden. Für das AMSTAR-Instrument ist kein derartiger Grenzwert festgelegt, ein solcher muss also ggf. vorab definiert werden. Darüber hinaus werden regelhaft die Sponsoren der systematischen Übersichten sowie die Interessenkonflikte der Autorinnen und Autoren dokumentiert und diskutiert. Die jeweiligen Indexkriterien können je nach Anforderung des Projekts um weitere Items ergänzt werden (beispielsweise Vollständigkeit der Recherche, Recherchen nach unpublizierten Studien z. B. in Registern oder zusätzliche Aspekte bei systematischen Übersichten diagnostischer Studien).

D) Ergebnisse

Die Ergebnisse einer Nutzenbewertung anhand von systematischen Übersichten werden zu jeder Fragestellung, wenn möglich, in tabellarischer Form zusammengefasst. Beim Vorliegen inkonsistenter Ergebnisse aus mehreren Arbeiten zu einem Endpunkt werden mögliche Erklärungen für diese Heterogenität beschrieben [285].

Liefert die Zusammenstellung der systematischen Übersichten 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 systematischer Übersichten fassen die Ergebnisse der zugrunde liegenden systematischen Übersichten zusammen, ergänzen sie ggf. um die Zusammenfassung der aktuelleren Primärstudien oder Primärstudien zu Fragestellungen, die nicht durch die systematischen Übersichten abgedeckt werden, und ziehen daraus ein eigenständiges Fazit.

Die Empfehlungen auf Basis systematischer Übersichten beruhen nicht auf der Zusammenfassung der Empfehlungen bzw. Schlussfolgerungen der zugrunde liegenden systematischen Übersichten. 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.

7.2.3 Berücksichtigung publizierter Meta-Analysen

Den Bewertungen des Instituts 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 7.3.8). 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 [480]. 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 [158]. 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.

7.3 Spezielle biometrische Aspekte

7.3.1 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* [109]. Besondere Aspekte, die beim Vorliegen ordinaler Daten berücksichtigt werden müssen, beschreibt Agresti [6,7].

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. [13].

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 [8].

7.3.2 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 [517]. 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 [172,305].

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 [274]. Das Institut folgt in der Regel diesem Vorgehen. Das Institut folgt weiterhin 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 [40]. Das Problem der Multiplizität ist in systematischen Übersichten nicht komplett lösbar, sollte aber dennoch zumindest bei der Interpretation der Ergebnisse beachtet werden [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) [12]. Zum Nachweis einer Gleichheit wendet das Institut adäquate Methoden für Äquivalenzhypothesen an.

Als prinzipielle Alternative zu statistischen Signifikanztests können Bayes'sche Methoden angesehen werden [488,489]. Das Institut wird in Abhängigkeit von der Fragestellung bei Bedarf auch Bayes'sche Methoden anwenden (z. B. bei indirekten Vergleichen, siehe Abschnitt 7.3.9).

7.3.3 Beurteilung klinischer Relevanz

Mit dem Begriff „klinische Relevanz“ werden in der Literatur unterschiedliche Konzepte bezeichnet. Zum einen kann es auf der Gruppenebene um die Frage gehen, ob ein Unterschied zwischen zwei Therapiealternativen bei einem patientenrelevanten Endpunkt (z. B. schwerwiegende Nebenwirkungen) groß genug ist, um eine allgemeine Anwendung der besseren Alternative zu empfehlen. Zum anderen wird unter klinischer Relevanz die Frage verstanden, ob eine Veränderung (z. B. der beobachtete Unterschied von 1 Punkt auf einer Symptomskala) für einzelne Patientinnen und Patienten relevant ist. Soweit das zweite Konzept im Sinne einer Responderdefinition und entsprechenden Responderanalysen zur Betrachtung von Gruppenunterschieden führt, sind beide Konzepte für die Bewertungen des Instituts von Bedeutung.

Allgemein spielt im Rahmen von systematischen Übersichten und Meta-Analysen die Bewertung der klinischen Relevanz von Gruppenunterschieden eine besondere Rolle, da hier oftmals eine Power erreicht wird, die das „statistische Aufdecken“ kleinster Effekte ermöglicht [526]. Dabei ist die klinische Relevanz eines Effekts oder Risikos 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 [433]. In einer kleinen Studie ist ein sehr kleiner p-Wert nur bei einem deutlichen Effekt zu erwarten, während in einer großen Studie auch bei einem verschwindend kleinen Effekt eine deutliche Signifikanz nicht ungewöhnlich ist [171,261]. 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, unabhängig davon, welches der o. g. Konzepte adressiert wird. So finden sich beispielsweise nur in wenigen Leitlinien Hinweise auf die Definition von relevanten beziehungsweise irrelevanten Unterschieden zwischen Gruppen [324,505]. Auch Methodenhandbücher zur Erstellung von systematischen Übersichten, wie z. B. das Cochrane Handbook [248], enthalten i. d. R. entweder keine oder keine eindeutige Anleitung zur Bewertung klinischer Relevanz auf System- oder Individualebene. Es gibt aber verschiedene Ansätze zur Beurteilung der klinischen Relevanz von Studienergebnissen. So kann z. B. der beobachtete Unterschied (Effektschätzer und das dazugehörige Konfidenz-

intervall) ohne prädefinierte Schwellen allein auf Basis medizinischer Sachkenntnis bewertet werden. Als formales Relevanzkriterium kann alternativ gefordert werden, dass das Konfidenzintervall vollständig oberhalb einer bestimmten „Irrelevanzschwelle“ liegen muss, um hinreichend sicher einen klar irrelevanten Effekt ausschließen zu können. Dies entspricht dann der Anwendung eines statistischen Tests mit Verschiebung der Nullhypothese zum statistischen Nachweis relevanter klinischer Effekte [551]. Ein weiterer Vorschlag sieht vor, dass die Relevanz allein auf Basis des Effektschätzers (im Vergleich zu einer „Relevanzschwelle“) beurteilt wird, vorausgesetzt, es liegt ein statistisch signifikanter Unterschied zwischen den Interventionsgruppen vor [301]. Im Gegensatz zur Anwendung eines statistischen Tests mit Verschiebung der Nullhypothese kann die Wahrscheinlichkeit für den Fehler 1. Art durch die Relevanzbewertung anhand des Effektschätzers nicht kontrolliert werden. Darüber hinaus ist dieser Ansatz u. U. weniger effizient. Schließlich liegt eine weitere Möglichkeit zur Relevanzbewertung darin, ein Relevanzkriterium auf individueller Ebene zu formulieren, z. B. im Sinne einer Responderdefinition [302]. Dabei gibt es auch Ansätze, bei denen sich das Responsekriterium innerhalb einer Studie zwischen den untersuchten Probanden unterscheidet, indem vorab individuelle Therapieziele formuliert werden [426].

Bei der Betrachtung patientenrelevanter Endpunkte, die mithilfe von (komplexen) Skalen operationalisiert werden, ist es in besonderer Weise notwendig, neben der statistischen Signifikanz der Effekte die Relevanz der beobachteten Wirkungen der untersuchten Interventionen zu bewerten, da die Komplexität der Skalen häufig eine sinnvolle Interpretation geringfügiger Unterschiede erschwert. Hier geht es also um die Frage, ob der beobachtete Unterschied zwischen zwei Gruppen überhaupt für die Patientinnen und Patienten spürbar ist. Diese Relevanzbewertung kann auf Basis von Mittelwertdifferenzen und Responderanalysen vorgenommen werden [466]. Ein wesentliches Problem bei der Relevanzbewertung ist die Tatsache, dass Relevanzkriterien skalenspezifisch nicht definiert sind oder aber adäquate Auswertungen auf Basis skalenspezifischer Relevanzkriterien (z. B. Responderanalysen) fehlen [375]. Welches Verfahren bei den Bewertungen des Instituts gewählt werden kann, hängt daher von der Verfügbarkeit der Daten aus den Primärstudien ab.

Um skalen- und indikationsspezifischen Besonderheiten gerecht zu werden, wendet das Institut für die Relevanzbewertung regelhaft folgende Hierarchie an, deren Stufen durch das Vorliegen verschiedener Relevanzkriterien determiniert sind.

- 1) Falls für die jeweilige Skala eine begründete Irrelevanzschwelle für den Gruppenunterschied (Mittelwertdifferenz) vorliegt bzw. ableitbar ist, wird diese für die Relevanzbewertung herangezogen. Liegt das zum beobachteten Effekt korrespondierende Konfidenzintervall vollständig oberhalb dieser Irrelevanzschwelle, ist statistisch abgesichert, dass die Effektstärke nicht in einem sicher irrelevanten Bereich liegt. Das Institut wertet dies als ausreichend für den Nachweis eines relevanten Effekts, da sich in diesem Fall die beobachteten Effekte in der Regel deutlich oberhalb der Irrelevanzschwelle (und zumindest nahe an einer Relevanzschwelle) realisieren. Für dieses

Kriterium ist zum einen eine validierte bzw. etablierte Irrelevanzschwelle geeignet. Zum anderen lässt sich aus einer validierten, etablierten oder anderweitig gut begründeten Relevanzschwelle (z. B. aus Fallzahlplanungen) eine Irrelevanzschwelle ableiten. Eine Möglichkeit besteht dann darin, als Irrelevanzschwelle die untere Grenze des Konfidenzintervalls festzulegen, welche sich bei einer zur klassischen Nullhypothese ausreichend gepowerten Studie ergibt, falls der geschätzte Effekt genau der Relevanzschwelle entspricht.

- 2) Falls skalenspezifische begründete Irrelevanzschwellen nicht vorliegen oder ableitbar sind, kommen Responderanalysen in Betracht. Hier ist es erforderlich, dass bei diesen Analysen ein validiertes bzw. etabliertes Responsekriterium angewendet wurde (z. B. im Sinne einer individuellen Minimal Important Difference [MID]) [423]. Liegt bei einer solchen Auswertung ein statistisch signifikanter Unterschied bezüglich der Anteile an Respondern zwischen den Gruppen vor, wird dies, sofern keine spezifischen Gründe dagegen sprechen, als Nachweis eines relevanten Effekts angesehen, da die Responsedefinition bereits eine Schwelle der Relevanz beinhaltet.
- 3) Falls weder skalenspezifische Irrelevanzschwellen noch Responderanalysen vorliegen, wird für die Relevanzbewertung auf ein allgemeines statistisches Maß in Form von standardisierten Mittelwertdifferenzen (SMD in Form von Hedges' g) zurückgegriffen. Als Irrelevanzschwelle wird dann 0,2 verwendet: Liegt das zum Effektschätzer korrespondierende Konfidenzintervall vollständig oberhalb dieser Irrelevanzschwelle, wird davon ausgegangen, dass die Effektstärke nicht in einem sicher irrelevanten Bereich liegt. Dies soll gewährleisten, dass der Effekt hinreichend sicher mindestens als „klein“ angesehen werden kann [169].

7.3.4 Bewertung subjektiver Endpunkte bei offenen Studiendesigns

In verschiedenen empirischen Arbeiten wurde gezeigt, dass in randomisierten kontrollierten Studien mit fehlender Verblindung bei subjektiven Endpunkten im Mittel eine Verzerrung der Effekte zugunsten der untersuchten Intervention vorliegt. Zu den subjektiven Endpunkten gehören beispielsweise patientenberichtete Endpunkte sowie Endpunkte, deren Erhebung und Einschätzung stark von den behandelnden bzw. endpunkterhebenden Personen abhängen. Eine Zusammenfassung dieser Arbeiten liefern Wood et al. [555]. Demnach sind solche Ergebnisse potenziell hochgradig verzerrt. Ein allgemein akzeptierter Umgang mit diesem Problem im Rahmen von systematischen Übersichten existiert nicht. I. d. R. wird das Institut in dieser Situation keinen Beleg für einen Nutzen oder Schaden aus statistisch signifikanten Ergebnissen ableiten.

Eine Möglichkeit, dem hohen Verzerrungspotenzial bei subjektiven Endpunkten in offenen Studien Rechnung zu tragen, besteht in der Formulierung einer adjustierten Entscheidungsgrenze. Nur dann, wenn das Konfidenzintervall des interessierenden Gruppenunterschieds einen bestimmten Abstand zum Nulleffekt aufweist, wird der Interventionseffekt als so groß angesehen, dass er nicht allein durch Verzerrung zu erklären ist. Das Verfahren der

Anwendung einer adjustierten Entscheidungsgrenze an sich stellt das Testen einer verschobenen Nullhypothese dar, wie es seit Jahrzehnten in der Anwendung ist, u. a. notwendigerweise bei der Prüfung von Äquivalenz- und Nichtunterlegenheitshypothesen [159]. Notwendig für die Anwendung adjustierter Entscheidungsgrenzen ist die prospektive Festlegung des konkreten Grenzwerts. Das Institut wird die Wahl des Grenzwerts bei Anwendung projektspezifisch durch empirische Daten, wie sie z. B. Wood et al. [555] liefern, begründen.

7.3.5 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. Dafür 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 [46,114].

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.

7.3.6 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 [12]. Für den Nachweis einer Gleichheit ist die Anwendung von Methoden zur Untersuchung von Äquivalenzhypothesen erforderlich [288].

Wichtig ist hierbei zunächst das Verständnis, dass eine exakte Gleichheit – also z. B., 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 bzw. höchstens irrelevanten Unterlegenheit erfolgen. In letzterem Fall spricht man von einer „Nichtunterlegenheitshypothese“ [100,274,428].

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 Analyseverfahren ab [427].

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 % [288]. In der Regel verwendet das Institut jedoch nach internationalem Vorgehen 95 %-Konfidenzintervalle.

Im Vergleich zu Überlegenheitsstudien besitzen Äquivalenzstudien spezielle methodische Probleme. Zum einen ist es häufig schwierig, Äquivalenzbereiche sinnvoll zu definieren [324]. Zum anderen schützen zum Beispiel die üblichen Designkriterien Randomisierung und Verblindung nicht mehr hinreichend sicher vor Verzerrungen [470]. 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 [288]. Somit ist bei der Bewertung von Äquivalenzstudien besondere Vorsicht geboten.

7.3.7 Adjustierung und multifaktorielle Verfahren

Vor allem in nicht randomisierten Studien spielen multifaktorielle Verfahren, die es ermöglichen, den Effekt von Confoundern auszugleichen, eine zentrale Rolle [296]. Ein weiteres wichtiges Anwendungsgebiet für multifaktorielle Methoden sind Studien mit

mehreren Interventionen [360]. Die Darstellung von Ergebnissen multifaktorieller Verfahren ist jedoch in der medizinischen Fachliteratur leider oft unzureichend [38,380]. Um die Qualität einer solchen Analyse bewerten zu können, ist eine Darstellung der wesentlichen Aspekte im Rahmen der statistischen Modellbildung notwendig [231,435] sowie Angaben zur Güte des gewählten Modells (Goodness-of-Fit) [257]. 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 [26,39,296].

In eigenen regressionsanalytischen Berechnungen setzt das Institut moderne Verfahren ein [230]. 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 [230,499]. Wichtig ist hierbei eine sorgfältige Vorauswahl der potenziellen erklärenden Variablen unter sachwissenschaftlicher Kenntnis [111]. Bei Bedarf werden auch moderne Verfahren wie die Lasso-Methode eingesetzt [511]. Bei der Modellierung stetiger Einflussgrößen greift das Institut im Bedarfsfall auf flexible Modellansätze wie zum Beispiel das Verfahren der Fractional Polynomials zurück [436,457], um eine adäquate Beschreibung nicht monotoner Zusammenhänge zu ermöglichen.

7.3.8 Meta-Analysen

A) Allgemeines

In der Literatur verwendete Begriffe wie „Literaturübersicht“, „systematische Übersicht“, „Meta-Analyse“, „gepoolte Analyse“ oder „Forschungssynthese“ sind häufig unterschiedlich definiert und nicht klar voneinander abgegrenzt [150]. 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 der Konfidenzintervalle grafisch aufgetragen werden [335]. Es kommen zum einen Modelle mit festem Effekt 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 [154,471,531]. Liegen Informationen darüber vor, dass die Effekte der einzelnen Studien homogen sind, ist eine Meta-Analyse unter der Annahme eines festen Effekts ausreichend. Solche Informationen werden jedoch häufig nicht vorliegen, sodass bei der Evaluierung der Studien in ihrer Gesamtheit die Annahme zufälliger Effekte hilfreich ist [472]. 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 substanziiell kleinere Überdeckungswahrscheinlichkeit aufweisen können [61]. Das Institut verwendet daher vorrangig Modelle mit zufälligen Effekten und weicht nur in begründeten Ausnahmefällen auf Modelle mit festem Effekt 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.

B) Heterogenität

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 zusammenfassenden Studien trotz Vergleichbarkeit häufig heterogene Effekte zeigen [250]. In dieser Situation ist es erforderlich, die Heterogenität der Studien bezüglich der Ergebnisse zu untersuchen [203]. 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 [282,308]. Daneben gilt es auch, das Ausmaß der Heterogenität zu quantifizieren. Zu diesem Zweck gibt es spezielle statistische Methoden wie z. B. das I^2 -Maß [249]. Für dieses Maß existieren Untersuchungen, die eine grobe Einschätzung der Heterogenität zulassen (z. B. die Kategorien wahrscheinlich unbedeutend (0 bis 40 %), mittelmäßig (30 bis 60 %), substanziiell (50 bis 90 %) und erheblich (75 bis 100 %) [109]). Ist die Heterogenität der Studien zu groß, so ist eine statistische Zusammenfassung der Studienergebnisse unter Umständen nicht sinnvoll [109]. Die Spezifizierung, wann eine „zu große“ Heterogenität vorliegt, ist kontextabhängig. In der Regel wird von einer statistischen Zusammenfassung abgesehen, falls der Heterogenitätstest einen p-Wert unter 0,2 liefert. Es spielt auch die Lage der Effekte eine Rolle. Zeigen die einzelnen Studien einen deutlichen und gleichgerichteten Effekt, dann kann auch die Zusammenfassung heterogener Ergebnisse mittels eines Modells mit zufälligen Effekten zu einer Nutzensaussage führen. In dieser Situation kann ggf. aber auch ohne quantitative Zusammenfassung eine positive Nutzensaussage getroffen werden (siehe Abschnitt 3.1.4). In den übrigen Situationen führt das Institut keine Meta-Analyse durch. In diese Entscheidung sollten jedoch neben statistischen Maßzahlen 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 [192]. In solchen Fällen sollte die Datenanalyse über ein relatives Effektmaß erfolgen; für die

deskriptive Darstellung können dann unter Umständen absolute Maße für spezifische Basisrisiken hieraus abgeleitet werden.

Bei einer großen 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 [506,524]. 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 [506]. 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 [211]. 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 [480,506] (siehe Abschnitt 7.2.3).

Zur Darstellung der Heterogenität im Rahmen einer Meta-Analyse mit zufälligen Effekten verwendet das Institut Prädiktionsintervalle [217,246,425]. Im Gegensatz zu einem Konfidenzintervall, das die Präzision eines geschätzten Effekts quantifiziert, überdeckt ein 95 %-Prädiktionsintervall den wahren Effekt einer einzelnen (neuen) Studie mit Wahrscheinlichkeit 95 %. Dabei ist es wichtig zu beachten, dass ein Prädiktionsintervall nicht zur Beurteilung der statistischen Signifikanz eines Effekts herangezogen wird. Das Institut folgt dem Vorschlag von Guddat et al. [217], das Prädiktionsintervall deutlich unterscheidbar von einem Konfidenzintervall in Form eines Rechtecks in einen Forest Plot einzufügen. Die Anwendung von Meta-Analysen mit zufälligen Effekten und zugehörigen Prädiktionsintervallen im Fall von sehr wenigen Studien (z. B. weniger als 5) wird in der Literatur kritisch gesehen, da eine mögliche Heterogenität nur sehr unpräzise geschätzt werden kann [246]. Das Institut stellt Prädiktionsintervalle in Forest Plots von Meta-Analysen mit zufälligen Effekten in der Regel dar, wenn mindestens 4 Studien vorhanden sind und wenn die grafische Darstellung der Heterogenität wichtig ist. Dies ist insbesondere dann der Fall, wenn aufgrund von zu großer Heterogenität gar kein gepoolter Effekt präsentiert wird.

Prädiktionsintervalle werden daher insbesondere in Forest Plots verwendet, wenn aufgrund zu starker Heterogenität kein Gesamteffekt geschätzt und dargestellt wird. In diesen heterogenen Situationen ist das Prädiktionsintervall eine wertvolle Hilfe bei der Beurteilung, ob die Studieneffekte gleichgerichtet sind oder nicht und ob es sich im ersten Fall um deutlich gleichgerichtete oder mäßig gleichgerichtete Effekte handelt (siehe Abschnitt 3.1.4).

C) Subgruppenanalysen im Rahmen von Meta-Analysen

Neben den allgemeinen Aspekten, die bei der Interpretation von Subgruppenanalysen beachtet werden müssen (siehe Abschnitt 7.1.6), gibt es besondere Aspekte, die bei Subgruppenanalysen im Rahmen von Meta-Analysen eine Rolle spielen. Während im Allgemeinen post hoc durchgeführte Subgruppenanalysen auf Studienebene kritisch zu interpretieren sind, ist man in einer systematischen Übersicht dennoch auf die Verwendung der Ergebnisse solcher Analysen auf Studienebene angewiesen, wenn im Rahmen der systematischen Übersicht genau diese Subgruppen untersucht werden sollen. Analog zum Vorgehen, Studien mit zu großer Heterogenität nicht mithilfe von Meta-Analysen zusammenzufassen, sollten auch Ergebnisse von Subgruppen nicht zu einem gemeinsamen Effektschätzer zusammengefasst werden, wenn sich die Subgruppen zu stark voneinander unterscheiden. Das Institut interpretiert im Rahmen von Meta-Analysen die Ergebnisse eines Heterogenitäts- oder Interaktionstests bezüglich wichtiger Subgruppen in der Regel wie folgt. Ein zum Niveau $\alpha = 0,05$ signifikantes Ergebnis wird als Beleg unterschiedlicher Effekte, ein zum Niveau $\alpha = 0,20$ signifikantes Ergebnis wird als Hinweis auf unterschiedliche Effekte in den Gruppen gewertet. Liegt mindestens ein Hinweis auf unterschiedliche Effekte in den Subgruppen vor, so werden neben dem Gesamteffekt auch die einzelnen Subgruppen-ergebnisse berichtet. Liegt ein Beleg für unterschiedliche Effekte in den Subgruppen vor, so werden die Ergebnisse aller Subgruppen nicht zu einem gemeinsamen Effektschätzer gepoolt. Bei mehr als zwei Subgruppen werden – wenn möglich – die paarweisen statistischen Tests auf das Vorliegen von Subgruppeneffekten durchgeführt und Paare, die zum Niveau $\alpha = 0,20$ nicht statistisch signifikant sind, zu einer Gruppe zusammengefasst. Die Ergebnisse der verbleibenden Gruppen werden getrennt berichtet und es werden getrennte Nutzensaussagen für diese Gruppen abgeleitet [482].

D) Geringe Zahl von Ereignissen

Ein häufiges Problem in Meta-Analysen bei binären Daten ist das Vorhandensein von sogenannten Nullzellen, also die Beobachtung von keinem einzigen Ereignis in einer Interventionsgruppe einer Studie. Das Institut folgt dem üblichen Vorgehen, beim Auftreten von Nullzellen den Korrekturwert von 0,5 zu jeder Zelhäufigkeit der entsprechenden Vierfeldertafel zu addieren [109]. Dieses Vorgehen ist adäquat, wenn nicht zu viele Nullzellen vorkommen. Im Fall einer insgesamt geringen Zahl von Ereignissen ist es unter Umständen notwendig, auf andere Methoden zurückzugreifen. Bei sehr seltenen Ereignissen kann die sogenannte Peto-Odds-Ratio-Methode verwendet werden, die keinen Korrekturterm beim Vorliegen von Nullzellen erfordert [56,109].

Kommen sogar Studien vor, in denen in beiden Studienarmen kein Ereignis beobachtet wird (sogenannte Doppelnulstudien), so werden diese Studien in der Praxis häufig aus der meta-analytischen Berechnung ausgeschlossen. Dieses Verfahren sollte vermieden werden, wenn zu viele Doppelnulstudien auftreten. Es gibt mehrere Methoden, um den Ausschluss von Doppelnulstudien zu vermeiden. Unter Umständen kann als Effektmaß die absolute Risikodifferenz verwendet werden, die gerade bei sehr seltenen Ereignissen häufig nicht zu

den sonst üblichen Heterogenitäten führt. Ein in der Praxis bislang selten angewendetes Verfahren stellt die logistische Regression mit zufälligen Effekten dar [519]. Neuere Verfahren wie exakte Methoden [510] oder die Anwendung der Arcus-Sinus-Differenz [438] stellen interessante Alternativen dar, sind aber noch nicht ausreichend untersucht. Das Institut wird in Abhängigkeit der jeweiligen Datensituation ein geeignetes Verfahren auswählen und ggf. mithilfe von Sensitivitätsanalysen die Robustheit der Ergebnisse untersuchen.

E) Meta-Analysen diagnostischer Studien

Auch die Ergebnisse von Studien zur diagnostischen Güte können mithilfe meta-analytischer Techniken statistisch zusammengefasst werden [124,281]. Wie in Abschnitt 3.5 ausgeführt, sind Studien, die allein die diagnostische Güte untersuchen, jedoch meist von nachrangiger Bedeutung in der Bewertung diagnostischer Verfahren, sodass auch Meta-Analysen von Studien zur diagnostischen Güte einen in gleicher Weise eingeschränkten Stellenwert haben.

Für eine Meta-Analyse von Studien zur diagnostischen Güte gelten die gleichen grundlegenden Prinzipien wie für Meta-Analysen von Therapiestudien [124,421]. Dies beinhaltet insbesondere die Notwendigkeit einer systematischen Literaturübersicht, die Bewertung der methodischen Qualität der Primärstudien, die Durchführung von Sensitivitätsanalysen und die Untersuchung des möglichen Einflusses von Publikationsbias.

Bei Meta-Analysen diagnostischer Studien ist in der Praxis in den meisten Fällen mit Heterogenität zu rechnen, daher empfiehlt sich hier in der Regel die Verwendung von Modellen mit zufälligen Effekten [124]. Eine solche meta-analytische Zusammenfassung von Studien zur diagnostischen Güte kann durch getrennte Modelle für Sensitivität und Spezifität erfolgen. Bei Interesse an einer summarischen Receiver-Operating-Characteristic(ROC)-Kurve und / oder einem zweidimensionalen Schätzer für Sensitivität und Spezifität haben jedoch neuere bivariate Meta-Analysen mit zufälligen Effekten Vorteile [227,422]. Diese Verfahren ermöglichen auch die Berücksichtigung erklärender Variablen [226]. Die grafische Darstellung der Ergebnisse erfolgt entweder über die separate Darstellung der Sensitivitäten und Spezifitäten in Form modifizierter Forest Plots oder eine zweidimensionale Abbildung der Schätzer für Sensitivität und Spezifität. Analog zu den Konfidenz- und Prädiktionsintervallen in Meta-Analysen von Therapiestudien können bei bivariaten Meta-Analysen von diagnostischen Studien Konfidenz- und Prädiktionsregionen im ROC-Raum dargestellt werden.

F) Kumulative Meta-Analysen

Es wird seit einiger Zeit verstärkt diskutiert, ob man bei wiederholten Aktualisierungen systematischer Übersichten die darin enthaltenen Meta-Analysen als kumulative Meta-Analysen mit Korrektur für multiples Testen berechnen und darstellen sollte [49,62,63,395,507,543]. Das Institut verwendet standardmäßig die übliche Form von Meta-Analysen und greift in der Regel nicht auf Methoden für kumulative Meta-Analysen zurück.

Für den denkbaren Fall, dass das Institut mit der regelmäßigen Aktualisierung einer systematischen Übersicht beauftragt wird, die so lange aktualisiert wird, bis eine

Entscheidung auf der Basis eines statistisch signifikanten Resultats vorgenommen werden kann, wird das Institut jedoch die Anwendung von Methoden für kumulative Meta-Analysen mit Korrektur für multiples Testen in Erwägung ziehen.

7.3.9 Indirekte Vergleiche

Zurzeit sind Methoden in der Entwicklung, die es ermöglichen, Evidenz aus direkten und indirekten Vergleichen zu kombinieren. Diese Verfahren werden Mixed Treatment Comparison (MTC) Meta-Analyse [347-349], Multiple Treatment Meta-Analyse (MTM) [79] oder auch Netzwerk-Meta-Analyse [350,447] genannt. Diese Methoden stellen eine wichtige Weiterentwicklung der üblichen meta-analytischen Verfahren dar. Allerdings gibt es noch zahlreiche ungelöste methodische Probleme, sodass gegenwärtig von einer routinemäßigen Anwendung dieser Methoden im Rahmen der Nutzenbewertung abzuraten ist [196,448,485,501]. Aus diesem Grund verwendet das Institut zur Nutzenbewertung von Interventionen primär direkt vergleichende Studien (placebokontrollierte Studien sowie Head-to-Head-Vergleiche), d. h., Aussagen für die Nutzenbewertung werden i. d. R. nur aus Ergebnissen direkt vergleichender Studien abgeleitet.

In bestimmten Situationen wie z. B. bei Bewertungen des Nutzens von Arzneimitteln mit neuen Wirkstoffen [120] sowie bei Kosten-Nutzen-Bewertungen (siehe unten) kann es jedoch erforderlich sein, indirekte Vergleiche einzubeziehen und daraus Aussagen für die Nutzenbewertung unter Berücksichtigung einer geringeren Ergebnissicherheit abzuleiten.

Bei der Kosten-Nutzen-Bewertung von Interventionen sind i. d. R. gemeinsame quantitative Vergleiche multipler (d. h. mehr als zwei) Interventionen notwendig. Die Einschränkung auf direkte Head-to-Head-Vergleiche würde bedeuten, dass sich die Kosten-Nutzen-Bewertung auf einen einzigen paarweisen Vergleich beschränken würde oder sogar ganz unmöglich wäre. Um eine Kosten-Nutzen-Bewertung multipler Interventionen zu ermöglichen, kann das Institut unter Inkaufnahme einer – im Vergleich zum Ansatz der reinen Nutzenbewertung – geringeren Ergebnissicherheit auch indirekte Vergleiche zur Bewertung von Kosten-Nutzen-Verhältnissen heranziehen [267].

Allerdings müssen adäquate Methoden für indirekte Vergleiche verwendet werden. Die Anwendung nicht adjustierter indirekter Vergleiche (d. h. die naive Verwendung einzelner Studienarme) wird vom Institut abgelehnt; es werden ausschließlich adjustierte indirekte Vergleiche akzeptiert. Hierzu zählen insbesondere das Verfahren von Bucher et al. [67] sowie die oben erwähnten Methoden der MTC Meta-Analysen. Neben den Annahmen paarweiser Meta-Analysen, die auch hier erfüllt sein müssen, muss bei indirekten Vergleichen zusätzlich eine ausreichende Konsistenz der in den einzelnen Studien geschätzten Effekte gegeben sein. Letzteres ist ein kritischer Punkt, da indirekte Vergleiche nur valide Ergebnisse liefern, wenn die Konsistenzannahmen erfüllt sind. Obwohl Techniken zur Untersuchung von Inkonsistenzen in der Entwicklung sind [126,348], gibt es in diesem Bereich noch viele offene methodische Fragen. Daher ist eine vollständige Beschreibung des verwendeten Modells zusammen mit verbleibenden Unklarheiten notwendig [501]. Darüber hinaus ist eine

wesentliche Voraussetzung für die Berücksichtigung eines indirekten Vergleichs, dass dieser auf die interessierende Fragestellung insgesamt ausgerichtet ist und nicht nur selektiv auf z. B. einzelne Endpunkte.

7.3.10 Umgang mit nicht oder nicht vollständig publizierten Daten

In der Praxis ergibt sich häufig das Problem, dass für die 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 Autorinnen und Autoren von Publikationen oder den Sponsoren von Studien (siehe Abschnitte 3.2.1 und 6.1.5). 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 [125,259,404]. 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.

7.3.11 Darstellung von Verzerrungsarten

Unter „Verzerrung“ (Bias) versteht man eine systematische Abweichung der Schätzung eines Effekts aus Studiendaten vom wahren Effekt. Es gibt vielfältige mögliche Ursachen, die eine Verzerrung hervorrufen können [86]. Nur die wichtigsten Verzerrungsarten werden im Folgenden beschrieben; eine ausführliche Übersicht über verschiedene Verzerrungsarten in unterschiedlichsten Situationen gibt Feinstein [170].

Ein „Selection Bias“ entsteht durch eine Verletzung der Zufallsprinzipien bei Stichprobenziehungen, d. h. bei der Zuteilung der Patienten zu den Interventionsgruppen. Speziell bei Gruppenvergleichen 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 [247], 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.

Neben der Vergleichbarkeit der Gruppen bezüglich potenzieller prognostischer Faktoren spielen die Behandlungsgleichheit und die Beobachtungsgleichheit aller Probanden eine entscheidende Rolle. Eine Verzerrung durch unterschiedliche Behandlungen (mit Ausnahme der zu untersuchenden Intervention) wird als „Performance Bias“ bezeichnet. Eine Verletzung der Beobachtungsgleichheit kann zu einem „Detection Bias“ führen. Die Verblindung ist ein wirksamer Schutz vor beiden Biasarten [291], die in der Epidemiologie als „Information Bias“ zusammengefasst werden.

Protokollverletzungen und Studienaustritte können bei Nichtberücksichtigung in der Auswertung das Studienergebnis systematisch verzerren, was als „Attrition Bias“ bezeichnet wird. Zur Verminderung von Attrition Bias kann in Studien, die eine Überlegenheit zeigen wollen, 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 [291,317].

Ein ähnliches Problem stellen fehlende Werte auch anderer Ursache dar. Fehlende Werte, die nicht durch einen Zufallsmechanismus zustande kommen, können ebenso eine Verzerrung im Ergebnis verursachen [344]. Die möglichen Ursachen und Auswirkungen fehlender Werte sind daher im Einzelfall zu diskutieren. Bei Bedarf sind statistische Verfahren erforderlich, die eine Verzerrung berücksichtigen bzw. ausgleichen. Ersetzungsverfahren (Imputationsverfahren) für fehlende Werte sind dabei nur eine Klasse von verschiedenen Verfahren, von denen keines als allgemein akzeptiert gilt. Beispielsweise empfiehlt die EMA, in Sensitivitätsanalysen unterschiedliche Verfahren zum Umgang mit fehlenden Werten gegenüberzustellen [163].

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 [181].

Ein „Reporting Bias“ wird verursacht durch die selektive Berichterstattung lediglich einer Teilmenge aller relevanten Daten und kann zur Überschätzung des Nutzens einer Intervention in systematischen Übersichten führen. Werden innerhalb einer Publikation in Abhängigkeit der Ergebnisse Teilmengen der Analysen oder Zielgrößen nicht bzw. weniger detailliert oder in einer von der Planung abweichenden Weise berichtet, dann liegt ein „Bias durch ergebnisgesteuerte Berichterstattung“ („Selective“ bzw. „Outcome Reporting Bias“) vor [84,142,247]. „Publication Bias“ dagegen beschreibt die Tatsache, dass Studien, die einen statistisch signifikanten negativen oder keinen statistisch signifikanten Unterschied zwischen der Interventions- und der Kontrollgruppe finden, gar nicht oder ggf. später publiziert werden als Studien mit positiven und statistisch signifikanten Resultaten [496]. Bei der Zusammenfassung der publizierten Ergebnisse kann es deshalb zu einer systematischen Verzerrung des gemeinsamen Effektschätzers kommen. Für die Entdeckung und Berücksichtigung von Publication Bias können grafische Methoden wie der Funnel-Plot [151] und statistische Methoden wie die Meta-Regression eingesetzt werden. Diese können das Vorliegen eines Publication Bias weder sicher nachweisen noch ausschließen, was die Bedeutung der Suche auch nach unpublizierten Daten unterstreicht. Beispielsweise können anhand von Studienregistern oder durch Anfragen bei Herstellern unpublizierte Informationen identifiziert bzw. erschlossen werden [327,351,409,495,496].

In Studien zur Ermittlung der Güte einer diagnostischen Strategie (Indextest) kann es zu verzerrten Ergebnissen kommen, wenn der Referenztest die Probanden nicht korrekt in gesund und krank unterscheidet („Misclassification Bias“). Wenn nur eine nicht zufällige Stichprobe von Probanden, die den Indextest erhalten haben, dem Referenztest unterzogen wird („Partial Verification Bias“) oder wenn der verwendete Referenztest vom Ergebnis des Indextests abhängt („Differential Verification Bias“), besteht die Gefahr, verzerrte Schätzungen der diagnostischen Güte zu erhalten. In Fällen, in denen der Indextest selbst Bestandteil des Referenztests ist, kann es zu Überschätzungen der diagnostischen Güte kommen („Incorporation Bias“) [331].

Ein weiterer in der internationalen Literatur aufgeführter Bias ist der sogenannte „Spectrum Bias“. Dieser spielt eine Rolle in Studien, in denen die Stichprobe zur Validierung eines diagnostischen Tests aus bereits als erkrankt bekannten Personen und gesunden Freiwilligen als Kontrollgruppe zusammengesetzt wird [341]. Die Validierung eines Testes in solchen Studien führt häufig zu Schätzungen für die Sensitivität und Spezifität, die höher sind, als sie es in einer klinischen Situation wären, bei der Patientinnen und Patienten mit Verdacht auf eine Erkrankung untersucht werden [545]. Der Begriff „Bias“ (im Sinne einer systematischen Beeinträchtigung der internen Validität) ist allerdings in diesem Zusammenhang nicht

günstig, da die Ergebnisse solcher Studien bei sachgerechter Durchführung durchaus intern valide sind [545]. Trotzdem können Studien des oben beschriebenen Designs Merkmale (insbesondere bezüglich der Stichprobenszusammensetzung) aufweisen, aufgrund derer sie für klinische Fragestellungen im Sinne der externen Validität nicht aussagekräftig sind.

Wie auch bei Interventionsstudien ist bei diagnostischen Studien die vollständige Berücksichtigung aller Teilnehmer der Studie (auch derjenigen mit nicht eindeutigen Testergebnissen) notwendig, um eine systematische Verzerrung des Ergebnisses zu vermeiden [331]. Während es zur Bedeutung des sowie zum Umgang mit Publication Bias im Zusammenhang mit Interventionsstudien bereits zahlreiche Untersuchungen gibt, ist dieses Problem für Studien zur diagnostischen Genauigkeit noch deutlich weniger erforscht [331].

Ein allgemeines Problem bei der Schätzung von Effekten ist eine Verzerrung durch Messfehler in den erhobenen Studiendaten [82,87]. In der Praxis sind Messfehler kaum zu vermeiden und es ist bekannt, dass auch nicht differenzielle Messfehler zu einer Verzerrung 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.

7.4 Qualitative Methoden

7.4.1 Qualitative Studien

Qualitative Forschungsmethoden werden eingesetzt, um subjektives Erleben, subjektive Erfahrungen und individuelles Handeln sowie die soziale Realität zu erforschen und zu verstehen [130,229,355,382]. Sie können Zugang zu den Sichtweisen und Erfahrungen von Patientinnen und Patienten, Angehörigen und des medizinischen Personals mit einem bestimmten Krankheitsbild oder einer Intervention schaffen.

Instrumente der qualitativen Forschung sind z. B. Fokusgruppen etwa mit Teilnehmenden einer randomisierten kontrollierten Studie. Qualitative Daten können auch über Interviews, Beobachtungen und schriftliche Dokumente wie Tagebücher erhoben werden.

An die Erhebung schließt sich eine Analyse an, die zumeist das Ziel hat, übergreifende Themen und Konzepte in den 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 [132].

Eine systematische Synthese verschiedener qualitativer Studien zu einer Forschungsfrage ist auch möglich [25,316,367,508]. Es gibt jedoch derzeit keine allgemein akzeptierte Vorgehensweise zur Synthese qualitativer Studien und zur Kombination qualitativer und quantitativer Daten [132,133].

A) Qualitative Studien bei der Erstellung von Gesundheitsinformationen

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

Es erfolgen eine Auswahl relevanter Publikationen anhand vorher festgelegter Ein- und Ausschlusskriterien sowie eine Abschätzung der Studienqualität anhand vorher definierter Kriterien. 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.

In den letzten Jahren wurden verschiedene Instrumente zur Qualitätsbewertung qualitativer Studien entwickelt [102]. 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 [130,133,229,382].

B) Qualitative Studien bei der Erstellung von Berichten

Verschiedene Informationsquellen können die Erarbeitung systematischer Übersichten unterstützen [131,336,504]. Eine dieser Quellen können Forschungsergebnisse aus qualitativen Studien sein [229,336,384,504]. Qualitative Studien scheinen sich in systematischen Übersichten zur Nutzenbewertung medizinischer Leistungen zu etablieren [130,131,384].

Qualitative Forschung kann Informationen zur Akzeptanz und Eignung von Interventionen in der Praxis liefern [25,130]. Die Ergebnisse qualitativer Forschung können bei der Interpretation einer systematischen Übersicht hilfreich sein [504] und im Rahmen von Primärstudien oder systematischen Übersichten zur Ermittlung patientenrelevanter Endpunkte eingesetzt werden [130,132,316,382,384].

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

7.4.2 Konsultationstechniken

Die Bearbeitung der an das Institut herangetragenen Fragen und Aufgaben erfordert oft die Konsultation von Patientinnen und Patienten, Patientenvertreterinnen und -vertretern sowie nationalen und internationalen Expertinnen und 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 Expertinnen und Experten sowie im Rahmen der Stellungnahmeverfahren. Im Rahmen der Entwicklung von Gesundheitsinformationen dienen Konsultationstechniken dazu, Patientinnen und Patienten und Patientenvertreterinnen und -vertreter 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) [522], z. B. Interviews mit Patientenvertreterinnen und -vertretern zur Ermittlung patientenrelevanter Endpunkte
- Gruppenmeetings und Konsultationen [385,388,389], z. B. im Rahmen von Erörterungen
- Gruppeninterviews und Fokusgruppen [130,522] z. B. mit Patientinnen und Patienten bezüglich der Evaluation von Gesundheitsinformationen
- Erfassung von Meinungen und Surveys (einschließlich Online-Umfragen und Feedback-mechanismen), z. B. bei der Ermittlung der Informationsbedürfnisse der Leserinnen und 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 Expertinnen und Experten sind mit einem Mehraufwand verbunden. Die Einbeziehung von Patientinnen und Patienten in wissenschaftliche Prozesse ermöglicht jedoch die Berücksichtigung von und die Ausrichtung des Forschungsprozesses auf Fragen und Bedürfnisse von Patientinnen und Patienten [398].

Anhang A - Rationale der Methodik zur Feststellung des Ausmaßes des Zusatznutzens

In diesem Anhang wird die Rationale für das methodische Vorgehen bei der Feststellung des Ausmaßes des Zusatznutzens gemäß der Arzneimittel-Nutzenbewertungsverordnung (AM-NutzenV) beschrieben.

Gemäß § 5 Abs. 4 Satz 1 der Arzneimittel-Nutzenbewertungsverordnung (AM-NutzenV) ist im Dossier darzulegen und folgerichtig auch zu bewerten, „in welchem Ausmaß ein Zusatznutzen vorliegt“. In § 5 Abs. 7 der AM-NutzenV findet sich dazu eine Einteilung in 6 Kategorien: (1) erheblicher Zusatznutzen, (2) beträchtlicher Zusatznutzen, (3) geringer Zusatznutzen, (4) nicht quantifizierbarer Zusatznutzen, (5) kein Zusatznutzen belegt, (6) geringerer Nutzen. Weiterhin liefert § 5 Abs. 7 der AM-NutzenV für die Kategorien 1 bis 3 eine Definition sowie beispielhafte, besonders zu berücksichtigende Kriterien als Orientierung für die Darlegung und Bewertung. Die dort formulierten Kriterien beschreiben sowohl qualitative Momente (Art der Zielgrößen) als auch explizit quantitative Momente (z. B. „erhebliche Verlängerung der Überlebensdauer“ vs. „moderate Verlängerung der Überlebensdauer“). Darüber hinaus ist offensichtlich eine Hierarchisierung dieser Zielgrößen intendiert, da teilweise dieselben Attribute (z. B. das Attribut „relevant“) bei unterschiedlichen Zielgrößen zu einem unterschiedlichen Ausmaß führen. In Tabelle 11 sind die diesbezüglichen Angaben für die primär relevanten Ausmaßkategorien erheblicher, beträchtlicher und geringer Zusatznutzen aufgeführt. Es ergibt sich für die Nutzenbewertung die Aufgabe, auf der Basis dieser Vorgaben das Ausmaß des Zusatznutzens zu operationalisieren.

Die in § 5 Abs. 7 der AM-NutzenV gelieferten Kriterien für das Ausmaß des Zusatznutzens benennen (Rechts-)Begriffe, die zum Teil eindeutig bestimmt (z. B. „Überlebensdauer“, „schwerwiegende Nebenwirkungen“), teilweise weniger eindeutig bestimmt sind (z. B. „Abschwächung schwerwiegender Symptome“). Darüber hinaus sind die Kategorien nicht für alle aufgeführten Kriterien erschöpfend besetzt, z. B. werden für die „Überlebensdauer“ nur Beispiele für die Kategorien „erheblicher“ und „beträchtlicher“ Zusatznutzen genannt.

Durch die Formulierung „insbesondere“ in § 5 Abs. 7 zu den Kategorien 1-3 macht der Verordnungsgeber deutlich, dass die den Kategorien zugeordneten Kriterien nicht abschließend zu verstehen sind. Es ist nicht davon auszugehen, dass der Verordnungsgeber einer weniger als „moderaten Verlängerung der Überlebensdauer“ nicht zumindest einen „geringen Zusatznutzen“ anerkennen wollte. Weiterhin erscheint die Zielgröße (gesundheitsbezogene) Lebensqualität, die in § 2 Abs. 3 der AM-NutzenV explizit als Nutzenkriterium formuliert wird, überhaupt nicht in der Kriterienliste für das Ausmaß des Zusatznutzens.

Tabelle 11: Feststellung des Ausmaßes des Zusatznutzens – Kriterien gemäß AM-NutzenV

Ausmaßskategorie	Erheblich nachhaltige und gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte große Verbesserung des therapielevanten Nutzens	Heilung	Erhebliche Verlängerung der Überlebensdauer	Langfristige Freiheit von schwerwiegenden Symptomen	Weitgehende Vermeidung schwerwiegender Nebenwirkungen
	Beträchtlich gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte deutliche Verbesserung des therapielevanten Nutzens	Spürbare Linderung der Erkrankung	Moderate Verlängerung der Überlebensdauer	Abschwächung schwerwiegender Symptome	Relevante Vermeidung schwerwiegender Nebenwirkungen Bedeutsame Vermeidung anderer Nebenwirkungen
	Gering gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte moderate und nicht nur geringfügige Verbesserung des therapielevanten Nutzens			Verringerung von nicht schwerwiegenden Symptomen	Relevante Vermeidung von Nebenwirkungen

In einem ersten Schritt ist es also sinnvoll, die Kriterienliste anzupassen und durch qualitativ und quantitativ gleichwertige Kriterien zu ergänzen. In Tabelle 12 sind die Ergänzungen der Vorgaben der AM-NutzenV aufgeführt. Dabei wurden die Aspekte „Heilung“ und „spürbare Linderung der Erkrankung“ nicht explizit berücksichtigt. Der Begriff der „Heilung“ bedarf grundsätzlich einer Operationalisierung, die sich regelhaft auf Kriterien stützen wird, die sich auch in den Endpunkten Mortalität und Morbidität abbilden lassen (z. B. Überleben über mindestens einen definierten Zeitraum in der Onkologie). Da der Begriff „Heilung“ in der AM-NutzenV ausschließlich mit einem erheblichen Zusatznutzen verknüpft wird, ist die jeweilige konkrete Operationalisierung anhand der verwendeten Endpunkte daraufhin zu prüfen, ob sie einer relevanten Verbesserung der Mortalität bzw. schwerwiegender Ereignisse gleichkommt. Die Verkürzung der Symptombdauer, z. B. bei banalen Infektionskrankheiten, ist in diesem Sinne nicht als Heilung anzusehen.

Ausgehend von diesen Ergänzungen ist eine Umstrukturierung der Zielgrößenkategorien angezeigt, um die in der AM-NutzenV intendierte Hierarchisierung der Zielgrößen abzubilden und gemäß § 5 Abs. 7 der AM-NutzenV den Schweregrad der Erkrankung zu berücksichtigen. Dazu werden die Zielgrößen gemäß ihrer Bedeutung wie folgt gruppiert (siehe Tabelle 13):

1. Gesamtmortalität
2.
 - schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen)
 - schwerwiegende (bzw. schwere) Nebenwirkungen
 - gesundheitsbezogene Lebensqualität
3.
 - nicht schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen)
 - nicht schwerwiegende (bzw. schwere) Nebenwirkungen

Der gesundheitsbezogenen Lebensqualität wird die gleiche Bedeutung beigemessen wie schwerwiegenden (bzw. schweren) Symptomen, Folgekomplikationen und Nebenwirkungen. Die möglichen Ausmaßkategorien für die nicht schwerwiegenden Zielgrößen bleiben auf „beträchtlich“ und „gering“ beschränkt.

Die Vorgaben der AM-NutzenV machen deutlich, dass zur Feststellung des Ausmaßes des Zusatznutzens zunächst auf Endpunktebene eine Effektstärkenbeschreibung zu erfolgen hat. Für jede Zielgröße wird separat die Effektstärke – unabhängig von ihrer Richtung – in die 3 Ausmaßkategorien (gering, beträchtlich, erheblich) eingestuft. Im Rahmen einer Gesamtabwägung sind diese einzelnen Ausmaße anschließend zu einer globalen Aussage zum Ausmaß des Zusatznutzens zusammenzufassen. Das schrittweise Vorgehen ist in Abschnitt 3.3.3 beschrieben.

Tabelle 12: Feststellung des Ausmaßes des Zusatznutzens – Kriterien gemäß AM-NutzenV mit Ergänzungen*

		Zielgrößenkategorie			
		<i>Gesamtmortalität</i>	<i>Symptome (Morbidität)</i>	<i>Gesundheitsbezogene Lebensqualität</i>	<i>Nebenwirkungen</i>
Ausmaßkategorie	Erheblich nachhaltige und gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte große Verbesserung des therapielevanten Nutzens	Erhebliche Verlängerung der Überlebensdauer	Langfristige Freiheit von schwerwiegenden (bzw. schweren) Symptomen (bzw. Folgekomplikationen)	<i>Erhebliche Verbesserung der Lebensqualität</i>	Weitgehende Vermeidung schwerwiegender (bzw. schwerer) Nebenwirkungen
	Beträchtlich gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte deutliche Verbesserung des therapielevanten Nutzens	Moderate Verlängerung der Überlebensdauer	Abschwächung schwerwiegender (bzw. schwerer) Symptome (bzw. Folgekomplikationen) <i>Bedeutsame Verringerung von nicht schwerwiegenden (bzw. schweren) Symptomen (bzw. Folgekomplikationen)</i>	<i>Bedeutsame Verbesserung der Lebensqualität</i>	Relevante Vermeidung schwerwiegender (bzw. schwerer) Nebenwirkungen Bedeutsame Vermeidung anderer (nicht schwerwiegender bzw. schwerer) Nebenwirkungen
	Gering gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte moderate und nicht nur geringfügige Verbesserung des therapielevanten Nutzens	<i>Jegliche Verlängerung der Überlebensdauer</i>	<i>Jegliche Verringerung schwerwiegender (bzw. schwerer) Symptome (bzw. Folgekomplikationen)</i> Verringerung von nicht schwerwiegenden (bzw. schweren) Symptomen (bzw. Folgekomplikationen)	<i>Relevante Verbesserung der Lebensqualität</i>	<i>Jegliche (statistisch signifikante) Verringerung schwerwiegender (bzw. schwerer) Nebenwirkungen</i> Relevante Vermeidung von (anderen, nicht schwerwiegenden bzw. schweren) Nebenwirkungen

*Ergänzungen gegenüber AM-NutzenV *kursiv* gesetzt

Tabelle 13: Feststellung des Ausmaßes des Zusatznutzens – hierarchisierte Kriterien gemäß AM-NutzenV mit Ergänzungen*

		Zielgrößenkategorie			
		Gesamtmortalität	Schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen	Gesundheitsbezogene Lebensqualität	Nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen
Ausmaßkategorie	Erheblich nachhaltige und gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte große Verbesserung des therapielevanten Nutzens	Erhebliche Verlängerung der Überlebensdauer	Langfristige Freiheit bzw. weitgehende Vermeidung	<i>Erhebliche Verbesserung</i>	<i>Nicht besetzt</i>
	Beträchtlich gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte deutliche Verbesserung des therapielevanten Nutzens	Moderate Verlängerung der Überlebensdauer	Abschwächung bzw. relevante Vermeidung	<i>Bedeutsame Verbesserung</i>	Bedeutsame Vermeidung
	Gering gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte moderate und nicht nur geringfügige Verbesserung des therapielevanten Nutzens	<i>Jegliche Verlängerung der Überlebensdauer</i>	<i>Jegliche Verringerung</i>	<i>Relevante Verbesserung</i>	Relevante Vermeidung

*Ergänzungen gegenüber AM-NutzenV *kursiv* gesetzt

Entsprechend § 2 Abs. 3 der AM-NutzenV ist der Begriff „Nutzen“ als Effekt definiert und in § 2 Abs. 4 der AM-NutzenV der Begriff „Zusatznutzen“ als ein solcher Effekt im Vergleich zu der zweckmäßigen Vergleichstherapie. Daraus kann abgeleitet werden, dass die Feststellung des Ausmaßes des Zusatznutzens neben der hierarchischen Betrachtung von Zielgrößen auch auf der Basis von Effektstärken zu erfolgen hat.

Zu den Fragen, welche Effektstärken für die einzelnen Zielgrößen zu welcher Ausmaßkategorie führen und welche Effektmaße für diese Bewertung zu wählen sind, finden sich in der AM-NutzenV keine Angaben. Diese Fragen können prinzipiell nur bedingt methodisch beantwortet werden. Dennoch besteht die Notwendigkeit, das in den Dossiers dargelegte Ausmaß des Zusatznutzens zu bewerten (§ 7 Abs. 2, AM-NutzenV) und selbst Aussagen zum Ausmaß zu machen. Um hierbei zunächst die im weiteren Abwägungsprozess notwendigerweise zu treffenden Werturteile möglichst gering zu halten und diese explizit zu machen, bedarf es einer

- expliziten Operationalisierung, um ein transparentes und nachvollziehbares Verfahren sicherzustellen, sowie einer
- abstrakten Operationalisierung, um größtmögliche Konsistenz zwischen den Nutzenbewertungen zu erzielen.

Vor diesem Hintergrund ist zunächst die Wahl eines geeigneten Effektmaßes zu treffen. Es sei zunächst die Situation binärer Daten (Analyse von Vierfeldertafeln) im Fokus. Relative Effektmaße – hierunter fallen im Wesentlichen das relative Risiko (RR) und das Odds Ratio (OR) – haben in diesem Zusammenhang gegenüber absoluten Maßen wie der Risikodifferenz (RD) folgende Vorteile:

- Die Risikodifferenz beschreibt nicht die Effektivität einer Therapie als solche, da sie stark vom Basisrisiko in der Kontrollgruppe abhängt. Dieses variiert jedoch zwischen Regionen, Populationen und im Zeitverlauf sowie insbesondere auch zwischen verschiedenen Vergleichstherapien. Eine Risikodifferenz muss daher als beschreibendes Maß einer konkreten Studie, nicht als fixe Maßzahl eines Therapieverfahrens aufgefasst werden, ein Problem auch und vor allem für Meta-Analysen [483]. Diese hohe Sensitivität für Rahmenbedingungen stellt die Übertragbarkeit von absoluten Effektmaßen aus Studien in die Versorgung infrage. Daher ist es übliche Praxis, Effekte in klinischen Studien vorzugsweise als relatives Risiko, Odds Ratio bzw. Hazard (oder auch Incidence) Ratio auszudrücken [108].
- Die Höhe der Risikodifferenz wird von der Höhe des Basisrisikos (absolutes Risiko in der Kontrollgruppe) begrenzt. Liegt dieses bei 1 %, dann kann die Risikodifferenz niemals über 0,01 liegen, beträgt es 10 %, dann nicht über 0,1 usw. Die Risikodifferenz könnte nur dann ihr Optimum 1 erreichen, wenn das Basisrisiko bei 100 % läge. Würde nun beispielsweise eine mindestens 20 %ige absolute Risikoreduktion als wesentliche therapeutische Verbesserung definiert, so wäre (für diese beispielhafte Forderung) bei

Erkrankungen mit (langfristigen) Überlebensraten $> 80\%$ grundsätzlich kein erheblicher Zusatznutzen (für den entsprechenden Endpunkt) mehr darstellbar.

- Ein weiterer Nachteil der Verwendung von absoluten Risikoreduktionen als Effektmaß zur Operationalisierung zur Feststellung des Ausmaßes des Zusatznutzens ist zudem, dass ein genauer Zeitpunkt definiert werden muss, bei dem diese absolute Risikoreduktion bestimmt wird (z. B. nach 1, 2, 5 oder 10 Jahren), sofern es dazu keine allgemein akzeptierten Festlegungen gibt (z. B. 30-Tage-Mortalität bei Myokardinfarkt).

Zusammenfassend mögen absolute Risikoreduktionen in einer individuellen Entscheidungssituation eher handlungsleitend sein, für allgemeine Aussagen im Sinne einer Bewertung des Zusatznutzens eines Arzneimittels sind dagegen relative Effektmaße besser geeignet.

Den relativen Maßen ist gemeinsam, dass der Nulleffekt (kein Gruppenunterschied) bei 1 liegt. Im Folgenden wird auf Effekte unterhalb von 1 eingegangen. Hierauf können Effekte oberhalb von 1 durch Kehrwertbildung zurückgeführt werden. Das Konzept sieht vor, dass ein 95 %-Konfidenzintervall im Sinne einer verschobenen Hypothesengrenze einen jeweiligen Schwellenwert unterschreitet, damit das Ergebnis als erheblicher, beträchtlicher oder geringer Zusatznutzen eingestuft wird. Ein solches inferenzstatistisches Vorgehen hat gegenüber der Betrachtung von Punktschätzern zwei wesentliche Vorteile: (i) Die Präzision der Schätzung fließt in die Bewertung ein; (ii) die statistischen Irrtumsmöglichkeiten lassen sich damit einhergehend auf übliche kleine Werte (z. B. 5 %) beschränken.

Die Schwellenwerte variieren bezüglich der in Tabelle 13 abgebildeten 2 Dimensionen Zielgrößenkategorie und Ausmaßkategorie des Effekts. Die Schwellenwerte sollten umso näher an 1 liegen (unterhalb von 1), je mehr Bedeutung einer Zielgröße zugemessen wird. Dadurch wird der Anforderung der AM-NutzenV nach einer Berücksichtigung der Krankheitsschwere Rechnung getragen. Demgegenüber sollten die Schwellenwerte umso weiter weg von 1 liegen (unterhalb von 1), je stärker das Ausmaß des Effekts attestiert wird.

Der oben beschriebenen expliziten und abstrakten Operationalisierung folgend ist eine Rasterung der Schwellenwerte von 0,05 vorgesehen [272]. Im Folgenden wird kurz die weitere Entwicklung der Methodik erläutert, die zu diesen Schwellenwerten geführt hat. Die weiteren Ausführungen werden aufzeigen, dass diese Wahl von 0,05 in der Praxis anwendbar ist und zu vernünftigen Aussagen führt.

Den Ausgangspunkt bildete die Frage, welche Größenordnung die tatsächlichen Effekte haben sollten, um z. B. das Ausmaß „erheblich“ innezuhaben. Dazu wurde ursprünglich ein relatives Risiko von 0,50 – von Djulbegovic et al. [134] als Anforderung für einen „Durchbruch“ postuliert – als Effekt erheblichen Ausmaßes für die Zielgröße Gesamtmortalität verankert [272].

Es stellte sich für diesen tatsächlichen Effekt (0,5) die Frage, wie der Schwellenwert gewählt werden muss, um mit einer adäquaten Power die Ausmaßkategorie „erheblich“ auch erreichen

zu können. Die entsprechenden Überlegungen dazu können im Detail der ersten durch das Institut durchgeführten Dossierbewertung entnommen werden [272], werden aber auch am Ende dieses Anhangs noch einmal aufgegriffen. Sie führten dazu, dass für einen Schwellenwert von 0,85 die gleichzeitige Anforderung nach Realisierbarkeit und Stringenz als erfüllt angesehen werden kann.

Im nächsten Schritt mussten dann für die Ausmaßmatrix die übrigen tatsächlichen Effekte festgelegt und die dazugehörigen Schwellenwerte ermittelt werden. Dabei war zu beachten, dass die Anforderungen von der Zielgrößenkategorie „Mortalität“ ausgehend für weniger schwerwiegende Zielgrößen zunehmen und von der Ausmaßkategorie „erheblich“ ausgehend für niedrigere Ausmaßkategorien abnehmen sollten. Eine Rasterung von 1/6 für die tatsächlichen Effekte erwies sich dabei als pragmatische Lösung. Nachfolgend werden die Schwellenwerte für die jeweiligen Ausmaßkategorien beschrieben.

1. Gesamtmortalität

Jegliche zum üblichen Irrtumsniveau 5 % statistisch signifikante Verlängerung der Überlebensdauer wird zumindest als „geringer Zusatznutzen“ eingestuft, da für die Gesamtmortalität die Anforderung „mehr als geringfügig“ bereits durch den Endpunkt selbst als erfüllt angesehen wird. Demnach beträgt der auf das 95 %-Konfidenzintervall bezogene Schwellenwert hier 1. Als „beträchtlicher“ Effekt wird eine Verlängerung der Überlebensdauer bezeichnet, wenn ein Schwellenwert von 0,95 unterschritten wird. Als „erheblich“ wird eine Verlängerung der Überlebensdauer bewertet, wenn der Schwellenwert von 0,85 durch die obere Grenze des 95 %-Konfidenzintervalls unterschritten wird.

2. • schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen)

- **schwerwiegende (bzw. schwere) Nebenwirkungen**
- **gesundheitsbezogene Lebensqualität**

Auch für schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und schwerwiegende (bzw. schwere) Nebenwirkungen entspricht jegliche statistisch signifikante Verminderung zumindest einem „geringen“ Effekt, weil die Anforderung „mehr als geringfügig“ bereits durch die Qualität des Endpunkts selbst erfüllt ist. Ein „beträchtlicher“ Effekt erfordert – in Abgrenzung zu gewünschten Effekten für die Gesamtmortalität – die Unterschreitung eines Schwellenwerts von 0,90. Voraussetzung für die Einstufung eines Effekts für diese Zielgrößen als „erheblich“ ist ebenfalls in Abgrenzung zu Effekten bezüglich der Gesamtmortalität die Unterschreitung eines Schwellenwerts von 0,75. Um aus diesen Zielgrößen einen erheblichen Effekt ableiten zu können, soll außerdem das Risiko für das untersuchte Ereignis in mindestens einer der zu vergleichenden Gruppen 5 % oder höher sein. Dieses zusätzliche Kriterium stützt die Relevanz des Ereignisses auf Populationsebene und trägt den besonderen Anforderungen an diese Kategorie des Zusatznutzens Rechnung.

Voraussetzung zur Feststellung des Ausmaßes des Effekts für Endpunkte zur gesundheitsbezogenen Lebensqualität (wie für alle patientenberichtete Endpunkte) ist, dass sowohl die eingesetzten Instrumente als auch die Responsekriterien validiert oder zumindest

unzweifelhaft etabliert sind. Liegen solche Ergebnisse dichotom im Sinne von Respondern / Non-Respondern vor, gelten dieselben im vorherigen Absatz genannten Kriterien (Risiko für die Kategorie „erheblich“ soll mindestens 5 % betragen) wie für schwerwiegende Symptome.

3. • nicht schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen)
• nicht schwerwiegende (bzw. schwere) Nebenwirkungen

Die Festlegung der Schwellenwerte für die nicht schwerwiegenden (bzw. schweren) Symptome (bzw. Folgekomplikationen) und die nicht schwerwiegenden (bzw. schweren) Nebenwirkungen berücksichtigt den im Vergleich zu den Kategorien 1 und 2 verminderten Schweregrad. Die Einstufung eines Effekts für diese Zielgrößen als „erheblich“ ist grundsätzlich nicht angezeigt. Voraussetzung für die Einstufung eines Effekts als „beträchtlich“ ist die Unterschreitung eines Schwellenwerts von 0,80. Ein „geringer Zusatznutzen“ erfordert die Unterschreitung eines Schwellenwerts von 0,90. Dies ist in der in § 5 Abs. 7 der AM-NutzenV formulierten Anforderung an einen geringen Zusatznutzen, dass es sich um eine moderate und nicht nur geringfügige Verbesserung handeln muss, begründet. Dem Verfahren ist somit implizit, dass (auch statistisch signifikante) Effekte, die aber nur als geringfügig bewertet werden, zur Einstufung in die Kategorie „kein Zusatznutzen“ führen.

In der folgenden Tabelle 14 sind die jeweiligen Schwellenwerte für alle Ausmaßkategorien und Zielgrößenkategorien abgebildet.

Tabelle 14: Inferenzstatistische Schwellenwerte (Hypothesengrenzen) für relative Effektmaße

		Zielgrößenkategorie		
		Gesamt-mortalität	Schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen und gesundheitsbezogene Lebensqualität ^a	Nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen
Ausmaßkategorie	Erheblich	0,85	0,75 und Risiko $\geq 5\%$ ^b	Nicht besetzt
	Beträchtlich	0,95	0,90	0,80
	Gering	1,00	1,00	0,90

a: Voraussetzung ist wie für alle patientenberichtete Endpunkte die Verwendung eines validierten bzw. etablierten Instruments sowie eines validierten bzw. etablierten Responsekriteriums. Werte gelten für Non-Response.

b: Risiko muss für mindestens eine der beiden zu vergleichenden Gruppen mindestens 5 % betragen.

Detaillierte methodische Rationale für die Festlegung der Schwellenwerte

Den Ausgangspunkt bildete die (fiktive) Planung einer Studie zur Testung der üblichen Hypothesen

$$H_0: RR \geq RR_0 \text{ vs. } H_1: RR < RR_0$$

anhand des relativen Risikos mit $RR_0 = 1$. Durch die Festlegung des Signifikanzniveaus, der Power, des Risikos in der Kontrollgruppe und des tatsächlichen Effekts (RR_1) ergibt sich die benötigte Fallzahl.

Eine solche Studie hätte für alle gegenüber 1 verschobenen Hypothesengrenzen ($RR_0 < 1$) eine geringere Power. Um für eine interessierende verschobene Hypothesengrenze (die oben genannten Schwellenwerte) dieselbe Power zu erhalten, die zur Testung der üblichen (nicht verschobenen) Hypothesen festgelegt wurde, muss die Fallzahl erhöht werden, und zwar entweder innerhalb der Studie oder durch Kombination mehrerer Studien. Vom Regelfall des Vorliegens von zwei (z. B. pivotalen) Studien ausgehend, wurde eine Verdoppelung der Fallzahl angenommen. Die Hypothesengrenze für die verschobenen Hypothesen wurde dann gerade so gewählt, dass die Power aus zwei Einzelstudien zu den üblichen Hypothesen der Power der gemeinsamen (gepoolten) Analyse zu den verschobenen Hypothesen entspricht. Diese Hypothesengrenze diente als Schwellenwert für die obere Grenze des zweiseitigen 95 %-Konfidenzintervalls für das relative Risiko. Bei Vorgabe eines Signifikanzniveaus von 5 % (zweiseitig) und einer Power von 90 % (sowohl für die übliche als auch für die verschobene Hypothesengrenze), einer Verdopplung der Fallzahl für die verschobene Hypothesengrenze ergab sich z. B. für den für die Zielgröße „Mortalität“ und die Ausmaßkategorie „erheblich“ postulierten tatsächlichen Effekt von 0,5 ein Schwellenwert von (gerundet) 0,85.

Die im Anhang A der Nutzenbewertung zu Ticagrelor [272] aufgeführte Formel für den Zusammenhang des tatsächlichen Effekts und des Schwellenwerts ist unabhängig von den sonstigen Vorgaben und beruht auf dem Algorithmus, der in der Prozedur „Power“ der Software SAS verwendet wird. In der entsprechenden Dokumentation für diesen Algorithmus [456] wird auf die Arbeit von Fleiss et al. [178] verwiesen. Ein Austausch mit Herrn Röhmel (damals Sprecher der Arbeitsgruppe Pharmazeutische Forschung der Deutschen Region der Internationalen Biometrischen Gesellschaft) sowie direkt mit dem Technical Support von SAS ergab, dass die Gültigkeit dieses Algorithmus offensichtlich nicht publiziert ist. Es stellte sich die Frage, welche tatsächlichen Effekte bei genauerer Berechnung notwendig sind, um mit einer hohen Wahrscheinlichkeit die jeweilige Ausmaßkategorie zu erreichen.

Die tatsächlichen Effekte wurden daher per Monte-Carlo-Simulationen folgendermaßen ermittelt.

- 1) Das Signifikanzniveau für die oben genannte Hypothese beträgt 2,5 %, die Power beträgt 90 %. Der Parameter RR_1 durchläuft alle Werte zwischen 0,2 und 0,95 mit Schrittweite

0,01. Das Risiko in der Kontrollgruppe p_C durchläuft alle Werte zwischen 0,05 und 0,95 mit Schrittweite 0,05. Für jedes dieser Tupel (RR_1, p_C) wird die benötigte Fallzahl n nach der Formel von Farrington und Manning [168] mit $RR_0 = 1$ berechnet und anschließend verdoppelt ($m := 2n$).

- 2) Für jedes Tripel (RR_1, p_C, m) durchläuft ein Schwellenwert SW absteigend alle Werte zwischen 1 und 0 mit Schrittweite -0,005. Für jedes SW wird die Power für die oben genannte Hypothese mit $RR_0 = SW$ approximiert. Das Signifikanzniveau beträgt 2,5 %. Dazu werden per Zufallsgenerator 50 000 Vierfeldertafeln simuliert, die obere Konfidenzintervallgrenze für das relative Risiko mittels Normalverteilungsapproximation und Deltamethode zur Varianzschätzung berechnet und anschließend der Anteil an Simulationsdurchläufen bestimmt, für die die obere Konfidenzintervallgrenze kleiner als SW ist. Der Durchlauf von SW wird gestoppt, sobald eine approximierte Power kleiner als 90 % ist. Das dazugehörige Tripel (RR_1, p_C, SW) wird in eine Liste geschrieben.
- 3) Nach Durchlauf aller Parameter in Schritt 1 und 2 werden alle Tripel der Liste selektiert, für die der Schwellenwert SW um weniger als 0,01 von einem der Werte 0,75, 0,8, 0,85, 0,9 oder 0,95 abweicht.

In Abbildung 8 sind die resultierenden (genaueren) tatsächlichen Effekte in Abhängigkeit des Risikos in der Kontrollgruppe für alle oben festgelegten Schwellenwerte aufgetragen (Punkte durch geglättete Kurven approximiert).

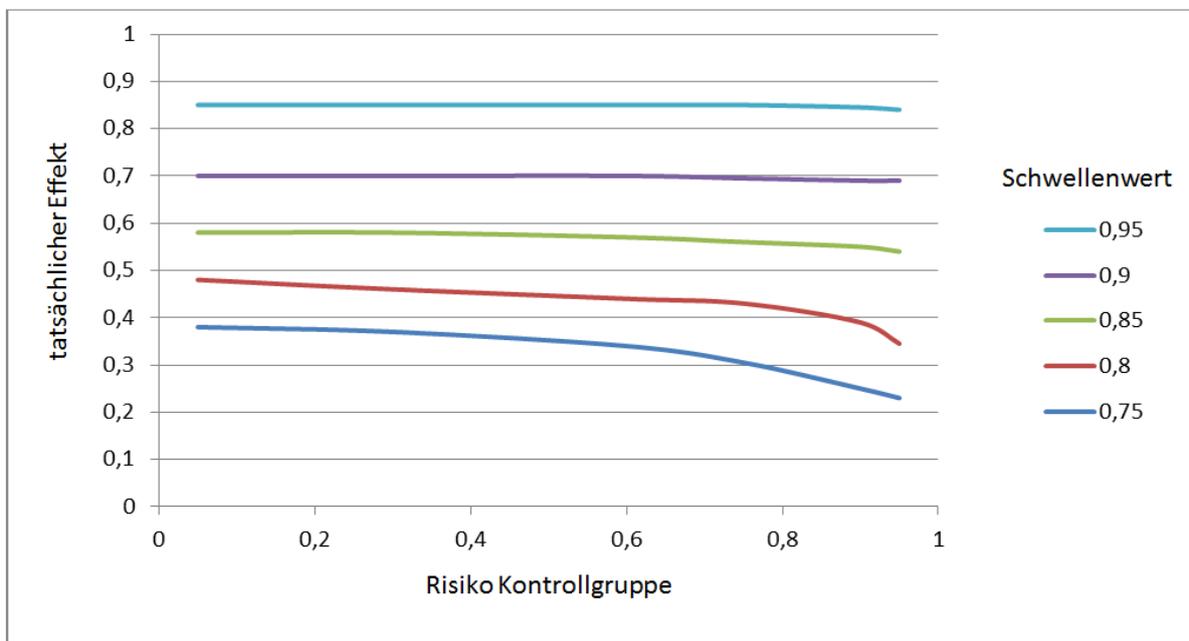


Abbildung 8: Tatsächliche Effekte in Abhängigkeit des Basisrisikos

In Tabelle 15 sind die Bereiche, in denen sich die tatsächlichen Effekte (in Abhängigkeit des Risikos der Kontrollgruppe) realisieren, pro Zielgrößen- und Ausmaßkategorie noch einmal eingetragen.

Tabelle 15: Tatsächliche Effekte für das relative Risiko

		Zielgrößenkategorie		
		Gesamt-mortalität	Schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplika-tionen) und Nebenwirkungen und gesundheitsbezogene Lebensqualität	Nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplika-tionen) und Nebenwirkungen
Ausmaßkategorie	Erheblich	0,53–0,58	0,24–0,38	Entfällt
	Beträchtlich	0,84–0,85	0,69–0,71	0,34–0,48
	Gering	Entfällt	Entfällt	0,69–0,71

Bezogen auf die Gesamtmortalität sind für ein erhebliches Ausmaß tatsächliche relative Risiken im Bereich 0,55 – also weiterhin etwa einer Halbierung des Risikos entsprechend – zu veranschlagen. Für das Ausmaß „beträchtlich“ muss der tatsächliche Effekt bei etwa 0,85 liegen. Für schwerwiegende Symptome und gleichwertige Zielgrößen bedarf es für ein erhebliches Ausmaß tatsächlich einer Risikoreduktion auf etwa ein Viertel bis ein Drittel. Verglichen mit den ursprünglich veranschlagten tatsächlichen Effekten [272] ergibt sich für die nahe an 1 liegenden Schwellenwerte eine gute Übereinstimmung. Bei den weiter von 1 entfernten Schwellenwerten zeigen die Simulationsergebnisse etwas moderatere Anforderungen an die Stärke der tatsächlichen Effekte. Die in Tabelle 14 veranschlagte Rasterung der Schwellenwerte erscheint vernünftig und praktikabel.

8 Literaturverzeichnis

1. Editorial commentary: avoiding biased comparisons [online]. In: James Lind Library. 2007 [Zugriff: 19.04.2013]. URL: <http://www.jameslindlibrary.org/essays/bias/avoiding-biased-comparisons.html>.
2. SGB V Handbuch Sozialgesetzbuch V: Krankenversicherung. Altötting: KKF-Verlag; 2011.
3. AGREE Collaboration. Appraisal of guidelines for research & evaluation: AGREE instrument; training manual [online]. 01.2003 [Zugriff: 08.08.2011]. URL: <http://www.agreecollaboration.org/pdf/aitraining.pdf>.
4. AGREE Collaboration. Appraisal of guidelines for research & evaluation (AGREE) instrument. London: AGREE Research Trust; 2006. URL: <http://www.agreetrust.org/?o=1085>.
5. AGREE Next Steps Consortium. Appraisal of guidelines for research and evaluation II: AGREE II instrument [online]. 05.2009 [Zugriff: 23.04.2013]. URL: <http://www.agreetrust.org/index.aspx?o=1397>.
6. Agresti A. Modelling ordered categorical data: recent advances and future challenges. Stat Med 1999; 18(18): 2191-2207.
7. Agresti A (Ed). Categorical data analysis. Hoboken: Wiley; 2002.
8. 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.
9. 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.
10. Altenhofen L, Blumenstock G, Diel F, Döbler K, Geraedts M, Jäckel WH et al. Qualitätsindikatoren: Manual für Autoren. Neukirchen: Make a Book; 2009. (ÄZQ-Schriftenreihe; Band 36). URL: <http://www.aeqz.de/mdb/edocs/pdf/schriftenreihe/schriftenreihe36.pdf>.
11. 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. S. 228-247.
12. Altman DG, Bland JM. Statistic notes: absence of evidence is not evidence of absence. BMJ 1995; 311(7003): 485.
13. Altman DG, Machin D, Bryant TM, Gardner MJ. Statistics with confidence: confidence intervals and statistical guidelines. London: BMJ Publishing Group; 2000.
14. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. BMJ 2009; 338(7708): 1432-1435.

15. American Society of Clinical Oncology. Outcomes of cancer treatment for technology assessment and cancer treatment guidelines. *J Clin Oncol* 1996; 14(2): 671-679.
16. Amir E, Seruga B, Martinez-Lopez J, Kwong R, Pandiella A, Tannock IF et al. Oncogenic targets, magnitude of benefit, and market pricing of antineoplastic drugs. *J Clin Oncol* 2011; 29(18): 2543-2549.
17. 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.
18. Arbeits- und Koordinierungsstelle Gesundheitsversorgungsforschung. Evaluation der Gesundheitsinformationsseite "Spezial: Wechseljahre": Interviews mit Nutzerinnen zum Informationspaket „Wechseljahre“ des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen“; Abschlussbericht [online]. 15.08.2008 [Zugriff: 22.04.2013]. URL: <http://www.akg.uni-bremen.de/pages/download.php?ID=27&SPRACHE=de&TABLE=AP&TYPE=PDF>.
19. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften, Ärztliche Zentralstelle Qualitätssicherung. Das Leitlinien-Manual von AWMF und ÄZQ. *Z Arztl Fortbild Qualitätssich* 2001; 95(Suppl 1): 5-84.
20. Arbeitskreis Versorgungsforschung beim Wissenschaftlichen Beirat. Definition und Abgrenzung der Versorgungsforschung [online]. 08.09.2004 [Zugriff: 23.04.2013]. URL: <http://www.bundesaerztekammer.de/downloads/Definition.pdf>.
21. Arnold M. Gesundheitssystemforschung. In: Hurrelmann K, Laaser U (Ed). *Gesundheitswissenschaften: Handbuch für Lehre, Forschung und Praxis*. Weinheim: Beltz; 1993. S. 423-437.
22. 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.
23. 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.
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. *BMC Health Serv Res* 2004; 4: 38.
25. Atkins S, Lewin S, Smith H, Engel M, Fretheim A, Volmink J. Conducting a meta-ethnography of qualitative literature: lessons learnt. *BMC Med Res Methodol* 2008; 8: 21.
26. 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.
27. Baker SG. Surrogate endpoints: wishful thinking or reality? *J Natl Cancer Inst* 2006; 98(8): 502-503.

28. Baltussen R, Niessen L. Priority setting of health interventions: the need for multi-criteria decision analysis. *Cost Eff Resour Alloc* 2006; 4: 14.
29. 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.
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. Health information on the internet. In: Heggenhougen HK, Quah SR (Ed). *International encyclopedia of public health: volume 3*; G-I. Amsterdam: Academic Press; 2008. S. 678-682.
33. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med* 2010; 7(9): e1000326.
34. Bastian H, Kaiser T, Matschewsky S. Förderung allgemeiner Gesundheits- und Wissenschaftskennntnisse mittels Bürger- und Patienteninformationen: die Rolle des IQWiG. *Z Arztl Fortbild Qualitatssich* 2005; 99(6): 379-385.
35. 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.
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(9): 857-865.
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. S. 515-538. (Handbook of Statistics; Band 8).
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, Zimbrunn B, Heininger U. Reporting of vaccine safety data in publications: systematic review. *Pharmacoepidemiol Drug Saf* 2005; 14(2): 101-106.
49. Borm GF, Donders AR. Updating meta-analyses leads to larger type I errors than publication bias. *J Clin Epidemiol* 2009; 62(8): 825-830, 830.e1-830.e10.
50. Bossuyt PM, Irwig LM, Craig J, Glasziou PP. Comparative accuracy: assessing new tests against existing diagnostic pathways. *BMJ* 2006; 332(7549): 1089-1092.
51. 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.
52. 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.
53. 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.
54. 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.
55. Boynton J, Glanville J, McDaid D, Lefebvre C. Identifying systematic reviews in MEDLINE: developing an objective approach to search strategy design. *Journal of Information Science* 1998; 24(3): 137-155.
56. Bradburn MJ, Deeks JJ, Berlin JA, Russell Localio A. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Stat Med* 2007; 26(1): 53-77.

57. Breslow NE, Day NE. Statistical methods in cancer research; volume I: the analysis of case-control studies. Lyon: International Agency for Research on Cancer; 1980. (IARC Scientific Publications; Band 32). URL: <http://www.iarc.fr/en/publications/pdfs-online/stat/sp32/SP32.pdf>.
58. Breslow NE, Day NE. Statistical methods in cancer research; volume II: the design and analysis of cohort studies. Lyon: International Agency for Research on Cancer; 1987. (IARC Scientific Publications; Band 82). URL: <http://www.iarc.fr/en/publications/pdfs-online/stat/sp82/SP82.pdf>.
59. Bridges JF, Kinter ET, Kidane L, Heinzen RR, McCormick C. Things are looking up since we started listening to patients: trends in the application of conjoint analysis in health 1982-2007. *Patient* 2008; 1(4): 273-282.
60. 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): 1-29.
61. Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. *Stat Med* 2001; 20(6): 825-840.
62. Brok J, Thorlund K, Gluud C, Wetterslev J. Trial sequential analysis reveals insufficient information size and potentially false positive results in many meta-analyses. *J Clin Epidemiol* 2008; 61(8): 763-769.
63. Brok J, Thorlund K, Wetterslev J, Gluud C. Apparently conclusive meta-analyses may be inconclusive: trial sequential analysis adjustment of random error risk due to repetitive testing of accumulating data in apparently conclusive neonatal meta-analyses. *Int J Epidemiol* 2009; 38(1): 287-298.
64. Brookes ST, Whitely E, Egger M, Davey Smith G, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *J Clin Epidemiol* 2004; 57(3): 229-236.
65. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G et al. Development of the AGREE II; part 1: performance, usefulness and areas for improvement. *CMAJ* 2010; 182(10): 1045-1052.
66. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G et al. Development of the AGREE II; part 2: assessment of validity of items and tools to support application. *CMAJ* 2010; 182(10): E472-E478.
67. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 1997; 50(6): 683-691.
68. Bundesministerium der Justiz. Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz - BStatG) [online]. 07.09.2007 [Zugriff: 06.05.2013]. URL: http://www.gesetze-im-internet.de/bundesrecht/bstatg_1987/gesamt.pdf.

69. Bundesministerium der Justiz. Verordnung zur Schaffung barrierefreier Informationstechnik nach dem Behindertengleichstellungsgesetz (Barrierefreie-Informationstechnik-Verordnung - BITV 2.0) [online]. 12.09.2011 [Zugriff: 06.05.2013]. URL: http://www.gesetze-im-internet.de/bitv_2_0/BJNR184300011.html.
70. Bundesministerium für Gesundheit. Verordnung über die Nutzenbewertung von Arzneimitteln nach § 35a Absatz 1 SGB V für Erstattungsvereinbarungen nach § 130b SGB V (Arzneimittel-Nutzenbewertungsverordnung – AM-NutzenV). Bundesgesetzblatt Teil 1 2010; (68): 2324-2328.
71. Bundessozialgericht. Urteil: Aktenzeichen B 6 A 1/08 R [online]. 06.05.2009 [Zugriff: 19.04.2013]. URL: <http://juris.bundessozialgericht.de/cgi-bin/rechtsprechung/document.py?Gericht=bsg&Art=en&sid=965bc60820d25990f7f287c0fa2b4c2c&nr=11110&pos=0&anz=1>.
72. Bundesverfassungsgericht. Leitsatz zum Beschluss des Ersten Senats: Aktenzeichen 1 BvR 347/98 [online]. 06.12.2005 [Zugriff: 23.04.2013]. URL: http://www.bverfg.de/entscheidungen/rs20051206_1bvr034798.html.
73. Bundesversicherungsamt. Festlegungen [online]. [Zugriff: 23.04.2013]. URL: http://www.bundesversicherungsamt.de/cln_115/nn_1046668/DE/Risikostrukturausgleich/Festlegungen/festlegungen_node.html?nnn=true.
74. Burgers JS. Guideline quality and guideline content: are they related? Clin Chem 2006; 52(1): 3-4.
75. Burgers JS, Bailey JV, Klazinga NS, Van der Bij AK, Grol R, Feder G. Inside guidelines: comparative analysis of recommendations and evidence in diabetes guidelines from 13 countries. Diabetes Care 2002; 25(11): 1933-1939.
76. Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. Pharm Stat 2006; 5(3): 173-186.
77. Burzykowski T, Molenberghs G, Buyse M (Ed). The evaluation of surrogate endpoints. New York: Springer; 2005.
78. Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. Biostatistics 2000; 1(1): 49-67.
79. Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ 2005; 331(7521): 897-900.
80. Campbell F, Dickinson HO, Cook JV, Beyer FR, Eccles M, Mason JM. Methods underpinning national clinical guidelines for hypertension: describing the evidence shortfall. BMC Health Serv Res 2006; 6: 47.
81. Campbell MK, Elbourne DR, Altman DG. CONSORT statement: extension to cluster randomised trials. BMJ 2004; 328(7441): 702-708.

82. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. Measurement error in nonlinear models: a modern perspective. London: Chapman & Hall; 2006.
83. Centre for Evidence-based Medicine. Levels of evidence (March 2009) [online]. 03.2009 [Zugriff: 19.04.2013]. URL: <http://www.cebm.net/index.aspx?o=1025>.
84. Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA 2004; 291(20): 2457-2465.
85. Charles C, Gafni A, Whelan T, O'Brien MA. Treatment decision aids: conceptual issues and future directions. Health Expect 2005; 8(2): 114-125.
86. Chavalarias D, Ioannidis JP. Science mapping analysis characterizes 235 biases in biomedical research. J Clin Epidemiol 2010; 63(11): 1205-1215.
87. Cheng CL, Van Ness JW. Statistical regression with measurement error. London: Arnold; 1999.
88. Clement S, Ibrahim S, Crichton N, Wolf M, Rowlands G. Complex interventions to improve the health of people with limited literacy: a systematic review. Patient Educ Couns 2009; 75(3): 340-351.
89. Cochrane Collaboration. Our principles [online]. 22.06.2012 [Zugriff: 23.04.2013]. URL: <http://www.cochrane.org/about-us/our-principles>.
90. Cochrane Effective Practice and Organisation of Care Review Group. The data collection checklist [online]. 06.2002 [Zugriff: 23.04.2013]. URL: <http://epoc.cochrane.org/sites/epoc.cochrane.org/files/uploads/datacollectionchecklist.pdf>.
91. Commission of the European Communities. eEurope 2002: quality criteria for health related websites [online]. 29.11.2002 [Zugriff: 04.03.2011]. URL: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2002:0667:FIN:EN:PDF>.
92. Committee of Experts on Management of Safety and Quality in Health Care, Expert Group on Safe Medication Practices. Glossary of terms related to patient and medication safety [online]. 20.10.2005 [Zugriff: 19.04.2013]. URL: http://www.who.int/patientsafety/highlights/COE_patient_and_medication_safety_gl.pdf.
93. Corbin JM, Strauss AL. Weiterleben lernen: Verlauf und Bewältigung chronischer Krankheit. Bern: Huber; 2003.
94. Cordoba G, Schwartz L, Woloshin S, Bae H, Gøtzsche PC. Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review. BMJ 2010; 341: c3920.
95. Cornell JE, Laine C. The science and art of deduction: complex systematic overviews. Ann Intern Med 2008; 148(10): 786-788.
96. 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.

97. 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. URL: http://www.pickereurope.org/assets/content/pdf/Project_Reports/Health-information-quality-web-version-FINAL.pdf.
98. Cui L, Hung HMJ, Wang SJ, Tsong Y. Issues related to subgroup analysis in clinical trials. *J Biopharm Stat* 2002; 12(3): 347-358.
99. Cullen R. Health information on the internet: a study of providers, quality, and users. Westport: Praeger; 2006.
100. 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.
101. Da Costa BR, Rutjes AWS, Johnston BC, Reichenbach S, Nuesch E, Tonia T et al. Methods to convert continuous outcomes into odds ratios of treatment response and numbers needed to treat: meta-epidemiological study. *Int J Epidemiol* 2012; 41(5): 1445-1459.
102. Daly J, Willis K, Small R, Green J, Welch N, Kealy M et al. A hierarchy of evidence for assessing qualitative health research. *J Clin Epidemiol* 2007; 60(1): 43-49.
103. Dans AL, Dans LF, Guyatt GH, Richardson S. 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.
104. Dans LF, Silvestre MA, Dans AL. Trade-off between benefit and harm is crucial in health screening recommendations; part I: general principles. *J Clin Epidemiol* 2011; 64(3): 231-239.
105. 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.
106. De Jonchere K, Gartlehner G, Goology L, Mustajoki P, Permanand G. Gesundheitsinformationen für Patienten und die Öffentlichkeit: zusammengestellt vom Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen; Gutachten der Weltgesundheitsorganisation 2008/2009 [online]. 2010 [Zugriff: 23.04.2013]. URL: http://www.euro.who.int/_data/assets/pdf_file/0004/94990/E93735g.pdf.
107. Deeks JJ. Systematic reviews in health care: systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001; 323(7305): 157-162.
108. Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat Med* 2002; 21(11): 1575-1600.
109. Deeks JJ, Higgins JPT, Altman DG. Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S (Ed). *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley; 2008. S. 243-296.

110. Delnoij DM, Groenewegen PP. Health services and systems research in Europe: overview of the literature 1995-2005. *Eur J Public Health* 2007; 17(Suppl 1): 10-13.
111. Derksen S, Keselman HJ. Backward, forward, and stepwise automated subset selection algorithms: frequency of obtaining authentic and noise variables. *Br J Math Stat Psychol* 1992; 45(2): 265-282.
112. 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.
113. Des Jarlais DC, Lyles C, Crepaz N. 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.
114. Desu MM, Raghavarao D. *Sample size methodology*. Boston: Academic Press; 1990.
115. Detmer DE. Building the national health information infrastructure for personal health, health care services, public health, and research. *BMC Med Inform Decis Mak* 2003; 3: 1.
116. Deutsche Gesellschaft für Epidemiologie. Leitlinien und Empfehlungen zur Sicherung von guter epidemiologischer Praxis (GEP): Langversion [online]. 03.2008 [Zugriff: 23.04.2013]. URL: http://www.gmds.de/pdf/publikationen/stellungnahmen/stell_gep_ergaenzung.pdf.
117. Deutsche Rentenversicherung Bund (Ed). *Rentenversicherung in Zeitreihen: Ausgabe 2012*. Berlin: DRV; 2008. (DRV-Schriften; Band 22). URL: http://www.deutsche-rentenversicherung.de/cae/servlet/contentblob/238700/publicationFile/50912/rv_in_zeitreihen.pdf.
118. Deutscher Bundestag. Gesetzentwurf der Fraktionen SPD, CDU/CSU und BÜNDNIS 90/DIE GRÜNEN: Entwurf eines Gesetzes zur Modernisierung der gesetzlichen Krankenversicherung (GKV-Modernisierungsgesetz - GMG); BT Drucksache 15/1525 [online]. 08.09.2003 [Zugriff: 16.01.2013]. URL: <http://dipbt.bundestag.de/doc/btd/15/015/1501525.pdf>.
119. Deutscher Bundestag. Gesetz zur Modernisierung der gesetzlichen Krankenversicherung (GKV-Modernisierungsgesetz - GMG). *Bundesgesetzblatt Teil 1* 2003; (55): 2190-2258.
120. Deutscher Bundestag. Gesetz zur Neuordnung des Arzneimittelmarktes in der gesetzlichen Krankenversicherung (Arzneimittelmarktneuordnungsgesetz – AMNOG) vom 22. Dezember 2010. *Bundesgesetzblatt Teil 1* 2010; (67): 2262-2277.
121. Deutscher Ethikrat (Ed). *Nutzen und Kosten im Gesundheitswesen: zur normativen Funktion ihrer Bewertung; Stellungnahme*. Berlin: Deutscher Ethikrat; 2011. URL: <http://www.ethikrat.org/dateien/pdf/stellungnahme-nutzen-und-kosten-im-gesundheitswesen.pdf>.

122. Deutsches Institut für Normung. Klinische Prüfung von Medizinprodukten an Menschen: gute klinische Praxis (ISO 14155:2011 + Cor. 1:2011); deutsche Fassung EN ISO 14155:2011 + AC:2011. Berlin: Beuth; 2012.
123. Deutsches Netzwerk Evidenzbasierte Medizin. Die "Gute Praxis Gesundheitsinformation". Z Evid Fortbild Qual Gesundheitswes 2010; 104(1): 66-68.
124. Devillé WL, Buntinx F, Bouter LM, Montori VM, De Vet HCW, Van der Windt DAWM et al. Conducting systematic reviews of diagnostic studies: didactic guidelines. BMC Med Res Methodol 2002; 2: 9.
125. Di Pietrantonj C. Four-fold table cell frequencies imputation in meta analysis. Stat Med 2006; 25(13): 2299-2322.
126. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. Stat Med 2010; 29(7-8): 932-944.
127. Dierks ML, Seidel G, Horch K, Schwartz FW. Bürger- und Patientenorientierung im Gesundheitswesen. Berlin: Robert Koch-Institut; 2006. (Gesundheitsberichterstattung des Bundes; Band 32). URL: <http://www.rki.de/DE/Content/Gesundheitsmonitoring/Gesundheitsberichterstattung/GBEDo wnloadsT/buergerorientierung.pdf? blob=publicationFile>.
128. DIPEX. Healthtalkonline [online]. [Zugriff: 19.04.2013]. URL: <http://www.healthtalkonline.org>.
129. 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.
130. Dixon-Woods M, Agarwal S, Young B, Jones D, Sutton A. Integrative approaches to qualitative and quantitative evidence. London: Health Development Agency; 2004. URL: http://www.nice.org.uk/niceMedia/pdf/Integrative_approaches_evidence.pdf.
131. Dixon-Woods M, Fitzpatrick R. Qualitative research in systematic reviews: has established a place for itself. BMJ 2001; 323(7316): 765-766.
132. 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.
133. 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.
134. Djulbegovic B, Kumar A, Soares HP, Hozo I, Bepler G, Clarke M et al. Treatment success in cancer: new cancer treatment successes identified in phase 3 randomized controlled trials conducted by the National Cancer Institute-sponsored cooperative oncology groups, 1955 to 2006. Arch Intern Med 2008; 168(6): 632-642.
135. Dolan JG. Shared decision-making: transferring research into practice; the Analytic Hierarchy Process (AHP). Patient Educ Couns 2008; 73(3): 418-425.

136. Dolan JG, Isselhardt BJ Jr, Cappuccio JD. The Analytic Hierarchy Process in medical decision making: a tutorial. *Med Decis Making* 1989; 9(1): 40-50.
137. Dolan P, Shaw R, Tsuchiya A, Williams A. QALY maximisation and people's preferences: a methodological review of the literature. *Health Econ* 2005; 14(2): 197-208.
138. Donner A, Klar J. Design and analysis of cluster randomization trials in health research. London: Arnold; 2000.
139. 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.
140. Drazen JM. COX-2 inhibitors: a lesson in unexpected problems. *N Engl J Med* 2005; 352(11): 1131-1132.
141. Drummond MF, Sculpher MJ, Torrance GW, O'Brian BJ, Stoddart GL. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press; 2005.
142. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS One* 2008; 3(8): e3081.
143. Ebell MH, Siwek J, Weiss BD, Woolf SH, Susman J, Ewigman B et al. Strength of Recommendation Taxonomy (SORT): a patient-centered approach to grading evidence in the medical literature. *Am Fam Physician* 2004; 69(3): 548-556.
144. Eccles MP, Rousseau N, Freemantle N. Updating evidence-based clinical guidelines. *J Health Serv Res Policy* 2002; 7(2): 98-103.
145. Edwards A, Elwyn G (Ed). Shared decision-making in health care: achieving evidence-based patient choice. Oxford: Oxford University Press; 2009.
146. Edwards AGK, Bastian H. Risk communication: making evidence part of patient choices? In: Edwards AGK, Elwyn GJ (Ed). Evidence-based patient choice: inevitable or impossible? Oxford: Oxford University Press; 2001. S. 144-160.
147. Edwards AGK, Elwyn GJ, Mulley A. Explaining risks: turning numerical data into meaningful pictures. *BMJ* 2002; 324(7341): 827-830.
148. Edwards AGK, 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.
149. 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.
150. Egger M, Davey Smith G, Altman DG (Ed). Systematic reviews in health care: meta-analysis in context. London: BMJ Publishing Group; 2001.

151. 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.
152. 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.
153. Elwyn GJ, O'Connor A, Stacey D, Volk R, Edwards AGK, Coulter A et al. Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. *BMJ* 2006; 333(7565): 417-424.
154. Engels EA, Schmid CH, Terrin N, Olkin I, Lau J. Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. *Stat Med* 2000; 19(13-35): 1707-1728.
155. Entwistle VA, Sheldon TA, Sowden A, Watt IS. 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.
156. Epstein RM, Alper BS, Quill TE. Communicating evidence for participatory decision making. *JAMA* 2004; 291(19): 2359-2366.
157. 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. *Amtsblatt der Europäischen Gemeinschaften* 2000; 43(L18): 1-5.
158. European Medicines Agency. Points to consider on application with: 1. meta-analyses; 2. one pivotal study [online]. 31.05.2001 [Zugriff: 22.09.2010]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003657.pdf.
159. European Medicines Agency. Guideline on the choice of the non-inferiority margin [online]. 27.07.2005 [Zugriff: 23.04.2013]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003636.pdf.
160. European Medicines Agency. Reflection paper on the regulatory guidance for the use of Health Related Quality of Life (HRQL) measures in the evaluation of medicinal products [online]. 27.07.2005 [Zugriff: 23.04.2013]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003637.pdf.
161. European Medicines Agency. Guideline on clinical trials in small populations [online]. 27.07.2006 [Zugriff: 23.04.2013]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003615.pdf.

162. European Medicines Agency. Guideline on clinical investigation of medicinal products in the treatment of diabetes mellitus: draft [online]. 20.01.2010 [Zugriff: 23.04.2013]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/02/WC500073570.pdf.
163. European Medicines Agency. Guideline on missing data in confirmatory clinical trials [online]. 02.07.2010 [Zugriff: 23.04.2013]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500096793.pdf.
164. Evidence-Based Medicine Working Group. Evidence-based medicine: a new approach to teaching the practice of medicine. *JAMA* 1992; 268(17): 2420-2425.
165. Eyding D, Lelgemann M, Grouven U, Harter M, Kromp M, Kaiser T et al. Reboxetine for acute treatment of major depression: systematic review and meta-analysis of published and unpublished placebo and selective serotonin reuptake inhibitor controlled trials. *BMJ* 2010; 341: c4737.
166. Eysenbach G. Recent advances: consumer health informatics. *BMJ* 2000; 320(7251): 1713-1716.
167. Eysenbach G, Köhler 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.
168. Farrington CP, Manning G. Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk. *Stat Med* 1990; 9(12): 1447-1454.
169. Fayers P, Machin D. Quality of life: the assessment, analysis and interpretation of patient-reported outcomes. Chichester: Wiley; 2007.
170. Feinstein AR. Clinical epidemiology: the architecture of clinical research. Philadelphia: Saunders; 1985.
171. Feinstein AR. Invidious comparisons and unmet clinical challenges. *Am J Med* 1992; 92(2): 117-120.
172. Feise RJ. Do multiple outcome measures require p-value adjustment? *BMC Med Res Methodol* 2002; 2: 8.
173. 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.
174. 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.

175. Ferreira-González I, Busse JW, Heels-Ansdell D, Montori VM, Alk EA, Byrant DM et al. Problems with use of composite end points in cardiocascular trials: systematic review of randomized controlled trials. *BMJ* 2007; 334(7597): 786-792.
176. Fessler J, Fischer J, Franzen D, Geraedts M, Graf HJ, Kroegel C et al. Leitlinien-Clearingbericht "COPD": Leitlinien-Clearingverfahren von Bundesärztekammer und Kassenärztlicher Bundesvereinigung in Kooperation mit Deutscher Krankenhausgesellschaft, Spitzenverbänden der Krankenkassen und Gesetzlicher Rentenversicherung. Niebüll: Videel; 2003. (ÄZQ-Schriftenreihe; Band 14). URL: <http://www.leitlinien.de/mdb/edocs/pdf/schriftenreihe/schriftenreihe14.pdf>.
177. Field MJ, Lohr KN (Ed). *Clinical practice guidelines: directions for a new program*. Washington: National Academy Press; 1990.
178. Fleiss JL, Tytun A, Ury HK. A simple approximation for calculating sample sizes for comparing independent proportions. *Biometrics* 1980; 36(2): 343-346.
179. Fleming TR. Surrogate endpoints and FDA's accelerated approval process. *Health Aff (Millwood)* 2005; 24(1): 67-78.
180. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med* 1996; 125(7): 605-613.
181. Fletcher RH, Fletcher SW. *Klinische Epidemiologie: Grundlagen und Anwendung*. Bern: Huber; 2007.
182. Food and Drug Administration. *Guidance for industry: developing medical imaging drug and biological products; part 2: clinical indications* [online]. 06.2004 [Zugriff: 23.04.2013]. URL: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071603.pdf>.
183. Food and Drug Administration. *Guidance for industry: patient-reported outcome measures; use in medical product development to support labeling claims* [online]. 12.2009 [Zugriff: 23.04.2013]. URL: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf>.
184. Førde OH. Is imposing risk awareness cultural imperialism? *Soc Sci Med* 1998; 47(9): 1155-1159.
185. 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.
186. 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.

187. 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.
188. Freemantle N, Calvert M. Weighing the pros and cons for composite outcomes in clinical trials. *J Clin Epidemiol* 2007; 60(7): 658-659.
189. French SD, McDonald S, McKenzie JE, Green SE. Investing in updating: how do conclusions change when Cochrane systematic reviews are updated? *BMC Med Res Methodol* 2005; 5: 33.
190. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991; 11(2): 88-94.
191. Fujimori M, Uchitomi Y. Preferences of cancer patients regarding communication of bad news: a systematic literature review. *Jpn J Clin Oncol* 2009; 39(4): 201-216.
192. 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.
193. 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.
194. Garcia-Castillo D, Feters MD. Quality in medical translations: a review. *J Health Care Poor Underserved* 2007; 18(1): 74-84.
195. Garrison LP Jr, Towse A, Bresnahan BW. Assessing a structured, quantitative health outcomes approach to drug risk-benefit analysis. *Health Aff (Millwood)* 2007; 26(3): 684-695.
196. Gartlehner G, Moore CG. Direct versus indirect comparisons: a summary of the evidence. *Int J Technol Assess Health Care* 2008; 24(2): 170-177.
197. Gemeinsamer Bundesausschuss. Beschluss des Gemeinsamen Bundesausschusses über die Anpassung der Beauftragung des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen zur Erstellung von Patienteninformationen [online]. 13.03.2008 [Zugriff: 23.04.2013]. URL: <http://www.g-ba.de/downloads/39-261-650/2008-03-13-IQWiG-Anpassung-Generalauftrag.pdf>.
198. Gemeinsamer Bundesausschuss. Verfahrensordnung des Gemeinsamen Bundesausschusses [online]. 06.12.2012 [Zugriff: 15.04.2013]. URL: http://www.g-ba.de/downloads/62-492-667/VerfO_2012-12-06.pdf.
199. Gemoets D, Rosemblat G, Tse T, Logan R. Assessing readability of consumer health information: an exploratory study. *Medinfo* 2004; 11(Pt 2): 869-873.
200. Gerhardt U. *Patientenkarrieren*. Frankfurt am Main: Suhrkamp; 1986.
201. Gesellschaft für Evaluation. *Standards für Evaluation*. Mainz: DeGEval; 2008. URL: http://www.degeval.de/images/stories/Publikationen/DeGEval_-_Standards.pdf.

202. Glasziou PP, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. *BMJ* 2007; 334(7589): 349-351.
203. Glasziou PP, Sanders SL. Investigating causes of heterogeneity in systematic reviews. *Stat Med* 2002; 21(11): 1503-1511.
204. Glasziou PP, Vandenbroucke JP, Chalmers I. Assessing the quality of research. *BMJ* 2004; 328(7430): 39-41.
205. 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.
206. 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.
207. Gøtzsche PC, Liberati A, Torri V, Rossetti L. Beware of surrogate outcome measures. *Int J Technol Assess Health Care* 1996; 12(2): 238-246.
208. Graham RM, Mancher M, Miller-Wolman D, Greenfield S, Steinberg E. Clinical practice guidelines we can trust. Washington: National Academies Press; 2011. URL: http://www.awmf.org/fileadmin/user_upload/Leitlinien/International/IOM_CPG_lang_2011.pdf.
209. Gray JAM. How to get better value healthcare. Oxford: Oxford Press; 2007.
210. Greenhalgh T, Hurwitz B. Narrative based medicine: why study narrative? *BMJ* 1999; 318(7175): 48-50.
211. Greenland S, Morgenstern H. Ecological bias, confounding, and effect modification. *Int J Epidemiol* 1989; 18(1): 269-274.
212. 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-e36.
213. Grilli R, Ramsay C, Minozzi S. Mass media interventions: effects on health services utilisation. *Cochrane Database Syst Rev* 2002; (1): CD000389.
214. Grimes DA, Schulz K. An overview of clinical research: the lay of the land. *Lancet* 2002; 359(9300): 57-61.
215. Grimes DA, Schulz KF. Surrogate end points in clinical research: hazardous to your health. *Obstet Gynecol* 2005; 105(5 Pt 1): 1114-1118.
216. Grouin JM, Coste M, Lewis J. Subgroup analyses in randomized clinical trials: statistical and regulatory issues. *J Biopharm Stat* 2005; 15(5): 869-882.
217. Guddat C, Grouven U, Bender R, Skipka G. A note on the graphical presentation of prediction intervals in random-effects meta-analyses. *Systematic Reviews* 2012; 1: 34.
218. Guyatt GH. Evidence-based medicine. *ACP J Club* 1991; 14(Suppl 2): A16.

219. Guyatt GH, Jaeschke R, Roberts R. N-of-1 randomized clinical trials in pharmacoepidemiology. In: Strom BL (Ed). *Pharmacoepidemiology*. Chichester: Wiley; 2005. S. 665-680.
220. Guyatt GH, Oxman AD, Kunz R, Falck-Ytter Y, Vist GE, Liberati A et al. Going from evidence to recommendations. *BMJ* 2008; 336(7652): 1049-1051.
221. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008; 336(7650): 924-926.
222. Guyatt GH, Rennie D. *Users' guides to the medical literature: a manual for evidence-based clinical practice*. Chicago: American Medical Association; 2002.
223. 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.
224. 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.
225. 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.
226. Hamza TH, Van Houwelingen HC, Heijnenbrok-Kal MH, Stijnen T. Associating explanatory variables with summary receiver operating characteristic curves in diagnostic meta-analysis. *J Clin Epidemiol* 2009; 62(12): 1284-1291.
227. Harbord RM, Whiting P, Sterne JA, Egger M, Deeks JJ, Shang A et al. An empirical comparison of methods for meta-analysis of diagnostic accuracy showed hierarchical models are necessary. *J Clin Epidemiol* 2008; 61(11): 1095-1103.
228. Harbour RT, Miller J. A new system for grading recommendations in evidence based guidelines. *BMJ* 2001; 323(7308): 334-336.
229. 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.
230. Harrell FE. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. New York: Springer; 2001.
231. 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.
232. Harris J. QALYfying the value of life. *J Med Ethics* 1987; 13(3): 117-123.

233. 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.
234. Hart D (Ed). *Ärztliche Leitlinien im Medizin- und Gesundheitsrecht: Recht und Empirie professioneller Normbildung*. Baden-Baden: Nomos; 2005. (Gesundheitsrecht und Gesundheitswissenschaften; Band 9).
235. Harteloh P. The meaning of quality in health care: a conceptual analysis. *Health Care Anal* 2003; 11(3): 259-267.
236. Hayden JA, Cote P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. *Ann Intern Med* 2006; 144(6): 427-437.
237. Haynes RB. Forming research questions. *J Clin Epidemiol* 2006; 59(9): 881-886.
238. 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.
239. Haynes RB, Devereaux PJ, Guyatt GH. Clinical expertise in the era of evidence-based medicine and patient choice. *Evid Based Med* 2002; 7(2): 36-38.
240. Health on the Net Foundation. Der HONcode in Kürze [online]. 04.1997 [Zugriff: 19.04.2013]. URL: <http://www.hon.ch/HONcode/German>.
241. Hemingway H, Riley RD, Altman DG. Ten steps towards improving prognosis research. *BMJ* 2009; 339: b4184.
242. Hennekens CH, Buring JE. *Epidemiology in medicine*. Boston: Little, Brown and Company; 1987.
243. Herxheimer A, McPherson A, Miller R, Shepperd S, Yaphe J, Ziebland S. Database of Patients' Experiences (DIPEX): a multi-media approach to sharing experiences and information. *Lancet* 2000; 355(9214): 1540-1543.
244. Herxheimer A, Ziebland S. DIPEX: fresh insights for medical practice. *J R Soc Med* 2003; 96(5): 209-210.
245. Hicks NJ. Evidence-based health care. *Bandolier* 1997; 4(5): 8.
246. Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc A* 2009; 172(1): 137-159.
247. Higgins JPT, Altman DG. Assessing risk of bias in included studies. In: Higgins JPT, Green S (Ed). *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley; 2008. S. 187-242.
248. Higgins JPT, Green S (Ed). *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley; 2008.

249. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; 21(11): 1539-1558.
250. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327(7414): 557-560.
251. Hill AB (Ed). *Controlled clinical trials*. Oxford: Blackwell; 1960.
252. Hirschberg I. *Bewertung und Wirkung von evidenzbasierten Gesundheitsinformationen: die Perspektive der Nutzer*; Public Health; Magisterarbeit. München: GRIN Verlag; 2010. (Patientenorientierung und Gesundheitskompetenz; Band 1).
253. Hirsh J, Guyatt G. Clinical experts or methodologists to write clinical guidelines? *Lancet* 2009; 374(9686): 273-275.
254. 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.
255. Hopewell S, Clarke M, Lefebvre C, Scherer R. Handsearching versus electronic searching to identify reports of randomized trials. *Cochrane Database Syst Rev* 2007; (2): MR000001.
256. Hopewell S, McDonald S, Clarke M, Egger M. Grey literature in meta-analyses of randomized trials of health care interventions. *Cochrane Database Syst Rev* 2007; (2): MR000010.
257. 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.
258. 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.
259. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol* 2005; 5: 13.
260. Hummel JM, IJzerman MJ. *The use of the Analytic Hierarchy Process in health care decision making*. Enschede: University of Twente; 2009.
261. Hung HM, O'Neill RT, Bauer P, Köhne K. The behavior of the P-value when the alternative hypothesis is true. *Biometrics* 1997; 53(1): 11-22.
262. Hussain T, Michel G, Shiffman RN. The Yale Guideline Recommendation Corpus: a representative sample of the knowledge content of guidelines. *Int J Med Inf* 2009; 78(5): 354-363.

263. ICH Expert Working Group. 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; E1; current step 4 version [online]. 27.10.1994 [Zugriff: 19.04.2013]. URL: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E1/Step4/E1_Guideline.pdf.
264. Inan H. Measuring the success of your website: a customer-centric approach to website management. Frenchs Forest: Pearson Education Australia; 2002.
265. Informationsgemeinschaft zur Feststellung der Verbreitung von Werbeträgern. Definitionen und technische Erläuterungen: Version 1.9; Anlage 1 zu den IVW-Richtlinien für Online-Angebote [online]. 14.11.2005 [Zugriff: 19.04.2013]. URL: http://daten.ivw.eu/download/pdf/Online_RichtlinienV1_9_Anlage1.pdf.
266. Initi@tive D²¹. (N)Onliner Atlas 2009: eine Topographie des digitalen Grabens durch Deutschland; Nutzung und Nichtnutzung des Internets, Strukturen und regionale Verteilung [online]. 06.2009 [Zugriff: 23.04.2013]. URL: <http://www.initiated21.de/wp-content/uploads/2009/06/NONLINER2009.pdf>.
267. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden zur Bewertung von Verhältnissen zwischen Nutzen und Kosten [online]. 12.10.2009 [Zugriff: 04.11.2013]. URL: https://www.iqwig.de/download/Methodik_fuer_die_Bewertung_von_Verhaeltnissen_zwischen_Kosten_und_Nutzen.pdf.
268. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Aussagekraft von Surrogatendpunkten in der Onkologie: Rapid Report; Auftrag A10-05; Version 1.1 [online]. 21.11.2011 [Zugriff: 23.04.2013]. (IQWiG-Berichte; Band 80). URL: https://www.iqwig.de/download/A10-05_Rapid_Report_Version_1-1_Surrogatendpunkte_in_der_Onkologie.pdf.
269. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Nutzungsbedingungen [online]. 13.09.2012 [Zugriff: 06.05.2013]. URL: <http://www.gesundheitsinformation.de/nutzungsbedingungen.51.de.html>.
270. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Vereinbarung über die vertrauliche Behandlung von Unterlagen [online]. 19.08.2005 [Zugriff: 19.04.2013]. URL: <http://www.iqwig.de/download/IQWiG-VFA-Mustervertrag.pdf>.
271. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Bupropion, Mirtazapin und Reboxetin bei der Behandlung von Depression: Abschlussbericht; Auftrag A05-20C [online]. 09.11.2009 [Zugriff: 23.04.2013]. (IQWiG-Berichte; Band 68). URL: https://www.iqwig.de/download/A05-20C_Abschlussbericht_Bupropion_Mirtazapin_und_Reboxetin_bei_Depressionen.pdf.

272. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Ticagrelor: Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag A11-02 [online]. 29.09.2011 [Zugriff: 04.11.2013]. (IQWiG-Berichte; Band 96). URL: https://www.iqwig.de/de/projekte_ergebnisse/projekte/arzneimittelbewertung/a11_02_ticagrelor_nutzenbewertung_gemass_35a_sgb_v_dossierbewertung.1425.html.
273. Institute of Medicine. Crossing the quality chasm: a new health system for the 21st century. Washington: National Academy Press; 2001. URL: http://books.nap.edu/openbook.php?record_id=10027.
274. International Conference on Harmonisation Expert Working Group. ICH harmonised tripartite guideline: statistical principles for clinical trials; an introductory note on an international guideline. *Stat Med* 1999; 18(15): 1905-1942.
275. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. Webauftritt [online]. [Zugriff: 07.05.2013]. URL: <http://www.ich.org>.
276. Ioannidis JPA. Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 2005; 294(2): 218-228.
277. 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.
278. Ioannidis JPA, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. *JAMA* 2001; 285(4): 437-443.
279. Ioannidis JPA, Mulrow CD, Goodman SN. Adverse events: the more you search, the more you find. *Ann Intern Med* 2006; 144(4): 298-300.
280. Irmen L, Linner U. Die Repräsentation generisch maskuliner Personenbezeichnungen: eine theoretische Integration bisheriger Befunde. *Z Psychol* 2005; 213(3): 167-175.
281. Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC et al. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med* 1994; 120(8): 667-676.
282. Jackson D. The power of the standard test for the presence of heterogeneity in meta-analysis. *Stat Med* 2006; 25(15): 2688-2699.
283. Jackson N, Waters E. Criteria for the systematic review of health promotion and public health interventions. *Health Promot Int* 2005; 20(4): 367-374.
284. Jadad AR. Randomised controlled trials: a user's guide. London: BMJ Books; 1998.
285. Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. *CMAJ* 1997; 156(10): 1411-1416.
286. Jadad AR, Gagliardi A. Rating health information on the internet: navigating to knowledge or to Babel? *JAMA* 1998; 279(8): 611-614.

287. Johnston ME, Brouwers MC, Browman GP. Keeping cancer guidelines current: results of a comprehensive prospective literature monitoring strategy for twenty clinical practice guidelines. *Int J Technol Assess Health Care* 2003; 19(4): 646-655.
288. Jones B, Jarvis P, Lewis J, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. *BMJ* 1996; 313(7048): 36-39.
289. Jones B, Kenward MG. Design and analysis of cross-over trials. London: Chapman and Hall; 1989. (Monographs on Statistics and Applied Probability; Band 34).
290. Jull A, Bennett D. Do n-of-1 trials really tailor treatment? *Lancet* 2005; 365(9476): 1992-1994.
291. Jüni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ* 2001; 323(7303): 42-46.
292. 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.
293. Kaltenthaler E, Brazier J, De Nigris E, Tumur 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): 1-168.
294. Kass PH, Gold EB. Modern epidemiologic study designs. In: Ahrens W, Pigeot I (Ed). *Handbook of epidemiology*. Berlin: Springer; 2005. S. 321-344.
295. 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.
296. Katz MH. Multivariable analysis: a primer for readers of medical research. *Ann Intern Med* 2003; 138(8): 644-650.
297. Kelley E, Hurst J. Health Care Quality Indicators Project: conceptual framework paper [online]. 09.03.2006 [Zugriff: 19.04.2013]. (OECD Health Working Papers; Band 23). URL: <http://www.oecd.org/dataoecd/1/36/36262363.pdf>.
298. 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-e27.
299. Kettunen T, Liimatainen L, Villberg J, Perko U. Developing empowering health counseling measurement: preliminary results. *Patient Educ Couns* 2006; 64(1-3): 159-166.
300. Kickbusch IS. Health literacy: addressing the health and education divide. *Health Promot Int* 2001; 16(3): 289-297.
301. Kieser M. Assessment of clinical relevance by considering point estimates and associated confidence intervals. *Pharm Stat* 2005; 4(2): 101-107.

302. Kieser M, Röhmel 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.
303. Klusen N, Meusch M (Ed). Wettbewerb und Solidarität im europäischen Gesundheitsmarkt. Baden-Baden: Nomos Verlagsgesellschaft; 2006. (Beiträge zum Gesundheitsmanagement; Band 16).
304. Knelangen M, Zschorlich B, Büchter R, Fechtelpeter D, Rhodes T, Bastian H. Online-Umfragen auf Gesundheitsinformation.de: Ermittlung potenzieller Informationsbedürfnisse für evidenzbasierte Gesundheitsinformationen. *Z Evid Fortbild Qual Gesundhwes* 2010; 104(8-9): 667-673.
305. Knottnerus JA, Bouter LM. The ethics of sample size: two-sided testing and one-sided thinking. *J Clin Epidemiol* 2001; 54(2): 109-110.
306. 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 evidenzbasierte Medizin in Klinik und Praxis*. Köln: Deutscher Ärzteverlag; 2007. S. 3-14.
307. Köbberling J, Trampisch HJ, Windeler J. Memorandum for the evaluation of diagnostic measures. *J Clin Chem Clin Biochem* 1990; 28(12): 873-879.
308. Koch A, Ziegler S. Metaanalyse als Werkzeug zum Erkenntnisgewinn. *Med Klin* 2000; 95(2): 109-116.
309. Koch G. No improvement: still less than half of the Cochrane reviews are up to date. In: XIV Cochrane Colloquium: programme and abstract book; 23.-26.10.2006; Dublin, Irland. 2006. S. 104.
310. Kolman J, Meng P, Scott G. Good clinical practice: standard operating procedures for clinical researchers. Chichester: Wiley; 1998.
311. 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. *Amtsblatt der Europäischen Gemeinschaften* 2003; 46(L159): 46-94.
312. Kools M, Van de Wiel MW, Ruiters 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.
313. Köpke S, Berger B, Steckelberg A, Meyer G. In Deutschland gebräuchliche Bewertungsinstrumente für Patienteninformationen: eine kritische Analyse. *Z Arztl Fortbild Qualitätssich* 2005; 99(6): 353-357.
314. Kraemer HC, Frank E, Kupfer DJ. Moderators of treatment outcomes: clinical, research, and policy importance. *JAMA* 2006; 296(10): 1286-1289.
315. Kranich C. Patientenkompetenz: was müssen Patienten wissen und können? *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 2004; 47(10): 950-956.

316. Kristensen FB, Sigmund H. Health technology assessment handbook. Kopenhagen: Danish Centre for Health Technology Assessment; 2007. URL: http://www.sst.dk/~media/Planlaegning%20og%20kvalitet/MTV%20metode/HTA_Handbook_net_final.ashx.
317. Kristman V, Manno M, Cote P. Loss to follow-up in cohort studies: how much is too much? *Eur J Epidemiol* 2004; 19(8): 751-760.
318. Krug S. Don't make me think! Web Usability; das intuitive Web. Heidelberg: mitp; 2006.
319. Kulbe A. Grundwissen Psychologie, Soziologie und Pädagogik: Lehrbuch für Pflegeberufe. Stuttgart: Kohlhammer; 2009.
320. Kunz R, Djulbegovic B, Schünemann HJ, Stanulla M, Muti P, Guyatt G. Misconceptions, challenges, uncertainty, and progress in guideline recommendations. *Semin Hematol* 2008; 45(3): 167-175.
321. Kunz R, Lelgemann M, Guyatt GH, Antes G, Falck-Ytter Y, Schünemann H. Von der Evidenz zur Empfehlung. In: Kunz R, Ollenschläger G, Raspe H, Jonitz G, Donner-Banzhoff N (Ed). *Lehrbuch evidenzbasierte Medizin in Klinik und Praxis*. Köln: Deutscher-Ärzte-Verlag; 2007. S. 231-247.
322. Kunz R, Vist GE, Oxman AD. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database Syst Rev* 2007; (2): MR000012.
323. Laaser U, Hurrelmann K. Gesundheitsförderung und Krankheitsprävention. In: Hurrelmann K, Laaser U (Ed). *Handbuch Gesundheitswissenschaften*. Weinheim: Juventa Verlag; 1998. S. 395-424.
324. Lange S, Freitag G. Choice of delta: requirements and reality; results of a systematic review. *Biom J* 2005; 47(1): 12-27.
325. Lapsley P. The patient's journey: travelling through life with a chronic illness. *BMJ* 2004; 329(7466): 582-583.
326. Last JM, Spasoff RA, Harris SS, Thuriaux MC (Ed). *A dictionary of epidemiology*. Oxford: Oxford University Press; 2001.
327. Lau J, Ioannidis JPA, Terrin N, Schmid CH, Olkin I. The case of the misleading funnel plot. *BMJ* 2006; 333(7568): 597-600.
328. Laupacis A, Wells G, Richardson WS, Tugwell P. Users' guides to the medical literature; V: how to use an article about prognosis. *JAMA* 1994; 272(3): 234-237.
329. Lausen B, Potapov S, Prokosch HU. Gesundheitsbezogene Internetnutzung in Deutschland 2007. *GMS Med Inform Biom Epidemiol* 2007; 4(2): Doc06.
330. Lavis JN. How can we support the use of systematic reviews in policymaking? *PLoS Med* 2009; 6(11): e1000141.

331. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Systematic reviews of diagnostic test accuracy. *Ann Intern Med* 2008; 149(12): 889-897.
332. Lefebvre C, Manheimer E, Glanville J. Searching for studies. In: Higgins JPT, Green S (Ed). *Cochrane handbook for systematic reviews of interventions*. New York: Wiley; 2008. S. 95-150.
333. 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.
334. Lewis D. Computer-based approaches to patient education: a review of the literature. *J Am Med Inform Assoc* 1999; 6(4): 272-282.
335. Lewis S, Clarke M. Forest plots: trying to see the wood and the trees. *BMJ* 2001; 322(7300): 1479-1480.
336. Leys M. Health care policy: qualitative evidence and health technology assessment. *Health Policy* 2003; 65(3): 217-226.
337. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009; 339: b2700.
338. 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.
339. Lieb K, Klemperer D, Koch K, Baethge C, Ollenschläger G, Ludwig WD. Interessenskonflikt in der Medizin: mit Transparenz Vertrauen stärken. *Dtsch Arztebl* 2011; 108(6): A256-A260.
340. Lijmer JG, Bossuyt PM. Various randomized designs can be used to evaluate medical tests. *J Clin Epidemiol* 2009; 62(4): 364-373.
341. Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, Van der Meulen JH et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999; 282(11): 1061-1066.
342. Lipkus IM, Hollands JG. The visual communication of risk. *J Natl Cancer Inst Monogr* 1999; (25): 149-163.
343. Lipscomb J, Drummond M, Fryback D, Gold M, Revicki D. Retaining, and enhancing, the QALY. *Value Health* 2009; 12(Suppl 1): S18-S26.
344. Little RJA, Rubin DB. *Statistical analysis with missing data*. Hoboken: Wiley; 2002.
345. Lo B, Field MJ (Ed). *Conflict of interest in medical research, education, and practice*. Washington: National Academies Press; 2009.

346. 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.
347. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004; 23(20): 3105-3124.
348. Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. *J Am Stat Assoc* 2006; 101(474): 447-459.
349. Lu G, Ades AE, Sutton AJ, Cooper NJ, Briggs AH, Caldwell DM. Meta-analysis of mixed treatment comparisons at multiple follow-up times. *Stat Med* 2007; 26(20): 3681-3699.
350. Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med* 2002; 21(16): 2313-2324.
351. Macaskill P, Walter SD, Irwig LM. A comparison of methods to detect publication bias in meta-analysis. *Stat Med* 2001; 20(4): 641-654.
352. MacDermid JC, Brooks D, Solway S, Switzer-McIntyre S, Brosseau L, Graham ID. Reliability and validity of the AGREE instrument used by physical therapists in assessment of clinical practice guidelines. *BMC Health Serv Res* 2005; 5: 18.
353. MacLean CH, Morton SC, Ofman JJ, Roth EA, Shekelle PG. How useful are unpublished data from the Food and Drug Administration in meta-analysis? *J Clin Epidemiol* 2003; 56(1): 44-51.
354. Maetzel A. Der Gebrauch von Nutzwerten im gesundheitsökonomischen Vergleich von Interventionen bei verschiedenen Krankheitsbildern: eine Einführung. *Z Rheumatol* 2004; 63(5): 380-384.
355. Malterud K. The art and science of clinical knowledge: evidence beyond measures and numbers. *Lancet* 2001; 358(9279): 397-400.
356. Mangiapane S, Velasco Garrido M. Surrogatendpunkte als Parameter der Nutzenbewertung [online]. 2009 [Zugriff: 24.04.2013]. (Schriftenreihe Health Technology Assessment; Band 91). URL: http://portal.dimdi.de/de/hta/hta_berichte/hta250_bericht_de.pdf.
357. 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.
358. Martini P. Methodenlehre der therapeutischen Untersuchung. Berlin: Springer; 1932.
359. Matthys J, De Meyere M, Van Driel ML, De Sutter A. Differences among international pharyngitis guidelines: not just academic. *Ann Fam Med* 2007; 5(5): 436-443.
360. McAlister FA, Straus SE, Sackett DL, Altman DG. Analysis and reporting of factorial trials: a systematic review. *JAMA* 2003; 289(19): 2545-2553.

361. McAlister FA, Van Diepen S, Padwal RS, Johnson JA, Majumdar SR. How evidence-based are the recommendations in evidence-based guidelines? *PLoS Med* 2007; 4(8): e250.
362. McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. *BMJ* 2002; 324(7351): 1448-1451.
363. McGauran N, Wieseler B, Kreis J, Schüler YB, Kölsch H, Kaiser T. Reporting bias in medical research: a narrative review. *Trials* 2010; 11(1): 37.
364. McGregor M, Caro JJ. QALYs: are they helpful to decision makers? *Pharmacoeconomics* 2006; 24(10): 947-952.
365. McMurray J, Swedberg K. Treatment of chronic heart failure: a comparison between the major guidelines. *Eur Heart J* 2006; 27(15): 1773-1777.
366. Medical Services Advisory Committee. Guidelines for the assessment of diagnostic technologies. Canberra: MSAC; 2005. URL: [http://www.health.gov.au/internet/msac/publishing.nsf/Content/D81BE529B98B3DB6CA2575AD0082FD1B/\\$File/Guidelines%20for%20the%20assessment%20of%20diagnostic%20technologies%20Sept%202005.pdf](http://www.health.gov.au/internet/msac/publishing.nsf/Content/D81BE529B98B3DB6CA2575AD0082FD1B/$File/Guidelines%20for%20the%20assessment%20of%20diagnostic%20technologies%20Sept%202005.pdf).
367. Mills E, Jadad AR, Ross C, Wilson K. Systematic review of qualitative studies exploring parental beliefs and attitudes toward childhood vaccination identifies common barriers to vaccination. *J Clin Epidemiol* 2005; 58(11): 1081-1088.
368. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; 340: c869.
369. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009; 339: b2535.
370. Moher D, Pham B, Klassen TP, Schulz KF, 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.
371. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med* 2007; 4(3): e78.
372. Moher D, Tsertsvadze A. Systematic reviews: when is an update an update? *Lancet* 2006; 367(9514): 881-883.
373. 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.
374. Molenberghs G, Burzykowski T, Alonso A, Assam P, Tilahun A, Buyse M. A unified framework for the evaluation of surrogate endpoints in mental-health clinical trials. *Stat Methods Med Res* 2010; 19(3): 205-236.

375. Molnar FJ, Man-Son-Hing M, Fergusson D. Systematic review of measures of clinical significance employed in randomized controlled trials of drugs for dementia. *J Am Geriatr Soc* 2009; 57(3): 536-546.
376. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ* 2009; 338: b606.
377. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ* 2009; 338: b375.
378. 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.
379. Müller HP, Schmidt K, Conen D. Qualitätsmanagement: interne Leitlinien und Patientenpfade. *Med Klin* 2001; 96(11): 692-697.
380. 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.
381. Munday J. *Introducing translation studies: theories and applications*. London: Routledge; 2001.
382. 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): 1-274.
383. Murray E, Burns J, See TS, Lai R, Nazareth I. Interactive health communication applications for people with chronic disease. *Cochrane Database Syst Rev* 2005; (4): CD004274.
384. National Advisory Committee on Health and Disability. *Screening to improve health in New Zealand: criteria to assess screening*. Wellington: National Health Committee; 2003. URL: http://www.nsu.govt.nz/files/NSU/Screening_to_improve_health.pdf.
385. National Health and Medical Research Council. *Statement on consumer and community participation in health and medical research*. Canberra: Commonwealth of Australia; 2002. URL: http://www.nhmrc.gov.au/files_nhmrc/file/publications/synopses/r22.pdf.
386. National Health and Medical Research Council. *Cultural competency in health: a guide for policy, partnerships and participation*. Canberra: Commonwealth of Australia; 2006. URL: http://www.nhmrc.gov.au/files_nhmrc/file/publications/synopses/hp19.pdf.
387. National Institute of Clinical Studies. *The impact of the internet on consumers health behaviour* [online]. 02.2003 [Zugriff: 24.04.2013]. URL: http://www.nhmrc.gov.au/files_nhmrc/file/nics/material_resources/The%20Impact%20of%20the%20Internet%20on%20Consumers%20Health%20Behaviour.pdf.

388. National Resource Centre for Consumer Participation in Health. Methods and models of consumer participation [online]. 01.09.2008 [Zugriff: 06.05.2013]. (Information Series; Band 2). URL: <http://www.healthissuescentre.org.au/documents/items/2008/09/231154-upload-00001.pdf>.
389. National Resource Centre for Consumer Participation in Health. Feedback, participation and consumer diversity: a literature review. Canberra: Commonwealth of Australia; 2000. URL: <http://www.healthissuescentre.org.au/documents/items/2008/08/226293-upload-00001.pdf>.
390. NHS Centre for Reviews and Dissemination. The Database of Abstracts of Reviews of Effects (DARE). Effectiveness Matters 2002; 6(2): 1-4.
391. Nielsen J. Designing web usability. Indianapolis: New Riders Publishing; 2000.
392. Nielsen J, Loranger H. Web Usability. München: Addison-Wesley; 2008.
393. 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.
394. Nord E. Cost-value analysis in health care: making sense out of QALYs. Cambridge: Cambridge University Press; 1999.
395. Nuesch E, Jüni P. Commentary: which meta-analyses are conclusive? Int J Epidemiol 2009; 38(1): 298-303.
396. Nutbeam D. Health promotion glossary. Health Promot Int 1998; 13(4): 349-364.
397. O'Connor AM, Bennett CL, Stacey D, Barry M, Col NF, Eden KB et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2009; (3): CD001431.
398. 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): 1-148.
399. Oxman AD, Guyatt GH. Guidelines for reading literature reviews. Can Med Assoc J 1988; 138(8): 697-703.
400. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol 1991; 44(11): 1271-1278.
401. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Ann Intern Med 1992; 116(1): 78-84.
402. 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.

403. Parkin DM, Chen VW, Ferlay J, Galceran J, Storm HH (Ed). Comparability and quality control in cancer registration. Lyon: International Agency for Research on Cancer; 1994. (IARC Technical Reports; Band 19).
404. 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.
405. Patientenuniversität an der Medizinischen Hochschule Hannover. Nutzertestung von Gesundheitsinformationen des Instituts für Qualität und Wirtschaftlichkeit (IQWiG): Abschlussbericht [online]. 10.2010 [Zugriff: 22.04.2013]. URL: http://www.mh-hannover.de/fileadmin/institute/epidemiologie/public_health/downloads/KursbeschrMar2011/NutzertestungAbschlussberichtIQWiG_16_12.pdf.
406. Pearson SD, Rawlins MD. Quality, innovation, and value for money: NICE and the British National Health Service. *JAMA* 2005; 294(20): 2618-2622.
407. Pennekamp PH, Diedrich O, Schmitt O, Kraft CN. Prävalenz und Stellenwert der Internetnutzung orthopädischer Patienten. *Zeitschrift für Orthopädie* 2006; 144(5): 459-463.
408. 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.
409. 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.
410. Petitti DB, Teutsch SM, Barton MB, Sawaya GF, Ockene JK, DeWitt T. Update on the methods of the U.S. Preventive Services Task Force: insufficient evidence. *Ann Intern Med* 2009; 150(3): 199-205.
411. Pfaff H, Glaeske G, Neugebauer EA, Schrappe M. Memorandum III: Methoden für die Versorgungsforschung (Teil 1). *Gesundheitswesen* 2009; 71(8-9): 505-510.
412. 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.
413. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA* 2006; 295(10): 1152-1160.
414. Pocock SJ. *Clinical trials: a practical approach*. Chichester: Wiley; 1983.
415. 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.
416. Poensgen A, Larsson S. Patients, physicians and the internet: myth, reality, and implications [online]. 01.2001 [Zugriff: 23.04.2013]. URL: <http://www.bostonconsulting.com.au/documents/file13629.pdf>.

417. Pollock K. Concordance in medical consultations: a critical review. Oxford: Radcliffe Publishing; 2005.
418. 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.
419. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med* 1989; 8(4): 431-440.
420. Puhan MA, Singh S, Weiss CO, Varadhan R, Boyd CM. A framework for organizing and selecting quantitative approaches for benefit-harm assessment. *BMC Med Res Methodol* 2012; 12: 173.
421. Raum E, Perleth M. Methoden der Metaanalyse von diagnostischen Genauigkeitsstudien. Köln: Deutsches Institut für Medizinische Dokumentation und Information; 2003. (Schriftenreihe Health Technology Assessment; Band 2). URL: http://portal.dimdi.de/de/hta/hta_berichte/hta025_bericht_de.pdf.
422. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005; 58(10): 982-990.
423. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008; 61(2): 102-109.
424. Richter JG, Becker A, Specker C, Monser R, Schneider M. Krankheitsbezogene Internetnutzung bei Patienten mit entzündlich-rheumatischen Systemerkrankungen. *Z Rheumatol* 2004; 63(3): 216-222.
425. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. *Br Med J* 2011; 342: 964-967.
426. Rockwood K, Fay S, Song X, MacKnight C, Gorman M. Attainment of treatment goals by people with Alzheimer's disease receiving galantamine: a randomized controlled trial. *Can Med Assoc J* 2006; 174(8): 1099-1105.
427. 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.
428. Röhm J, Hauschke D, Koch A, Pigeot I. Biometrische Verfahren zum Wirksamkeitsnachweis im Zulassungsverfahren: Nicht-Unterlegenheit in klinischen Studien. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 2005; 48(5): 562-571.
429. Rossouw JE. Estrogens for prevention of coronary heart disease: putting the brakes on the bandwagon. *Circulation* 1996; 94(11): 2982-2985.

430. Rothwell PM. Treating individuals 2: subgroup analysis in randomised controlled trials; importance, indications, and interpretation. *Lancet* 2005; 365(9454): 176-186.
431. Rotter T, Kinsman L, James E, Machotta A, Gothe H, Willis J et al. Clinical pathways: effects on professional practice, patient outcomes, length of stay and hospital costs. *Cochrane Database Syst Rev* 2010; (3): CD006632.
432. Royal Society. Science and the public interest: communicating the results of new scientific research to the public [online]. 04.2006 [Zugriff: 23.04.2013]. URL: http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/publications/2006/8315.pdf.
433. Royall RM. The effect of sample size on the meaning of significance tests. *Am Stat* 1986; 40(4): 313-315.
434. 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): 1-51.
435. Royston P. A strategy for modelling the effect of a continuous covariate in medicine and epidemiology. *Stat Med* 2000; 19(14): 1831-1847.
436. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Appl Stat* 1994; 43(3): 429-467.
437. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. *BMJ* 2009; 338: b604.
438. Rücker G, Schwarzer G, Carpenter J, Olkin I. Why add anything to nothing? The arcsine difference as a measure of treatment effect in meta-analysis with zero cells. *Stat Med* 2009; 28(5): 721-738.
439. 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.
440. Saaty TL. A scaling method for priorities in hierarchical structures. *J Math Psychol* 1977; 15(3): 234-281.
441. Saaty TL. Decision making with the Analytic Hierarchy Process. *International Journal of Services Sciences* 2008; 1(1): 83-98.
442. Saaty TL, Vargas LG. The Analytic Hierarchy Process: wash criteria should not be ignored. *International Journal of Management and Decision Making* 2006; 7(2/3): 180-188.
443. Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen. Bedarfsgerechtigkeit und Wirtschaftlichkeit; Band III: Über- Unter- und Fehlversorgung; Gutachten 2000/2001; ausführliche Zusammenfassung [online]. 08.2001 [Zugriff: 06.05.2013]. URL: http://www.svr-gesundheit.de/fileadmin/user_upload/Gutachten/2000-2001/Kurzfa-de-01.pdf.

444. Sackett DL. Bias in analytic research. *J Chronic Dis* 1979; 32(1-2): 51-63.
445. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ* 1996; 312(7023): 71-72.
446. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine: how to practice and teach EBM. Edinburgh: Churchill Livingstone; 2000.
447. Salanti G, Higgins JPT, Ades AE, Ioannidis JPA. Evaluation of networks of randomized trials. *Stat Methods Med Res* 2008; 17(3): 279-301.
448. Salanti G, Marinho V, Higgins JPT. A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. *J Clin Epidemiol* 2009; 62(8): 857-864.
449. 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.
450. Sampson M, McGowan J, Cogo E, Grimshaw J, Moher D, Lefebvre C. An evidence-based practice guideline for the peer review of electronic search strategies. *J Clin Epidemiol* 2009; 62(9): 944-952.
451. Sampson M, Shojania KG, McGowan J, Daniel R, Rader T, Iansavichene AE et al. Surveillance search techniques identified the need to update systematic reviews. *J Clin Epidemiol* 2008; 61(8): 755-762.
452. Sampson MJ. Updating searches for systematic reviews [Dissertation]. Aberystwyth: Universität; 2009.
453. Sanger S, Lang B, Klemperer D, Thomeczek C, Dierks ML. Manual Patienteninformation: Empfehlungen zur Erstellung evidenzbasierter Patienteninformationen. Berlin: rztliches Zentrum fur Qualitat in der Medizin; 2006. (ZQ-Schriftenreihe; Band 25). URL: <http://www.aezq.de/mdb/edocs/pdf/schriftenreihe/schriftenreihe25.pdf>.
454. 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.
455. 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.
456. SAS Institute. SAS/STAT 9.2 user's guide: second edition [online]. 2009 [Zugriff: 04.11.2013]. URL: <http://support.sas.com/documentation/cdl/en/statug/63033/PDF/default/statug.pdf>.
457. 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.
458. Sawaya GF, Guirguis-Blake J, LeFevre M, Harris R, Petitti D. Update on the methods of the U.S. Preventive Services Task Force: estimating certainty and magnitude of net benefit. *Ann Intern Med* 2007; 147(12): 871-875.

459. Sawicki PT. Qualität der Gesundheitsversorgung in Deutschland: ein randomisierter Sechs-Länder-Vergleich aus Patientensicht. *Med Klin* 2005; 100(11): 755-768.
460. 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 für die Spitzenverbände der GKV [online]. 11.2004 [Zugriff: 06.05.2013]. URL: http://www.gkv-spitzenverband.de/media/dokumente/krankenversicherung_1/praevention_selbsthilfe_beratung/beratung/Abschlussbericht_UPD_1_Modellphase_Uni_Bielefeld.pdf.
461. Schluter PJ, Ware RS. Single patient (n-of-1) trials with binary treatment preference. *Stat Med* 2005; 24(17): 2625-2636.
462. Schmidt-Kaehler S. Patienteninformation Online: theoretische Grundlagen, Planung und Entwicklung eines Konzeptes für die Patientenschulung im Internet. Bern: Huber; 2004.
463. 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.
464. Schöffski O, Graf von der Schulenburg JM (Ed). Gesundheitsökonomische Evaluationen. Berlin: Springer; 2012.
465. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; 340: c332.
466. Schünemann HJ, Akl EA, Guyatt GH. Interpreting the results of patient reported outcome measures in clinical trials: the clinician's perspective. *Health Qual Life Outcomes* 2006; 4: 62.
467. Schünemann HJ, Best D, Vist GE, Oxman AD. Letters, numbers, symbols and words: how to communicate grades of evidence and recommendations. *Can Med Assoc J* 2003; 169(7): 677-680.
468. Schünemann HJ, Fretheim A, Oxman AD. Improving the use of research evidence in guideline development; 9: grading evidence and recommendations. *Health Res Policy Syst* 2006; 4: 21.
469. Scott JT, Harmsen M, Pictor MJ, Entwistle VA, Sowden AJ, Watt I. Recordings or summaries of consultations for people with cancer. *Cochrane Database Syst Rev* 2008; (3): CD001539.
470. Senn SJ. Inherent difficulties with active control equivalence studies. *Stat Med* 1993; 12(24): 2367-2375.
471. Senn SJ. The many modes of meta. *Drug Inf J* 2000; 34(2): 535-549.
472. Senn SJ. Trying to be precise about vagueness. *Stat Med* 2007; 26(7): 1417-1430.

473. Shea BJ, Bouter LM, Peterson J, Boers M, Andersson N, Ortiz Z et al. External validation of a measurement tool to assess systematic reviews (AMSTAR). *PLoS One* 2007; 2(12): e1350.
474. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007; 7: 10.
475. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol* 2009; 62(10): 1013-1020.
476. 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.
477. 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.
478. 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.
479. Silvestre MAA, Dans LF, Dans AL. Trade-off between benefit and harm is crucial in health screening recommendations; part II: evidence summaries. *J Clin Epidemiol* 2011; 64(3): 240-249.
480. Simmonds MC, Higgins JPT. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. *Stat Med* 2007; 26(15): 2982-2999.
481. Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. *Br J Cancer* 1994; 69(6): 979-985.
482. Skipka G, Bender R. Intervention effects in the case of heterogeneity between three subgroups: assessment within the framework of systematic reviews. *Methods Inf Med* 2010; 49(6): 613-617.
483. Smeeth L, Haines A, Ebrahim S. Numbers needed to treat derived from meta-analyses: sometimes informative, usually misleading. *Br Med J* 1999; 318(7197): 1548-1551.
484. Song F, Eastwood AJ, Gilbody S, Duley L, Sutton A. Publication and related biases. *Health Technol Assess* 2000; 4(10): 1-115.
485. Song F, Loke YK, Walsh T, Glenny AM, Eastwood AJ, Altman DG. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. *BMJ* 2009; 338: b1147.

486. Song F, Parekh S, Hooper L, Loke YK, Ryder J, Sutton AJ et al. Dissemination and publication of research findings: an updated review of related biases. *Health Technol Assess* 2010; 14(8): 1-193.
487. 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.
488. Spiegelhalter DJ, Freedman LS, Parmar MK. Bayesian approaches to randomized trials. *J R Stat Soc Ser A* 1994; 157(3): 357-416.
489. 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.
490. Statistisches Bundesamt. Statistik der schwerbehinderten Menschen 2007: Kurzbericht [online]. 01.2009 [Zugriff: 22.04.2013]. URL: https://www.destatis.de/DE/Publikationen/Thematisch/Gesundheit/BehinderteMenschen/SozialSchwerbehinderteKB5227101079004.pdf?__blob=publicationFile.
491. Statistisches Bundesamt. Leben in Deutschland: Haushalte, Familien und Gesundheit; Ergebnisse des Mikrozensus 2005 [online]. 06.2006 [Zugriff: 07.05.2013]. URL: https://www.destatis.de/DE/PresseService/Presse/Pressekonferenzen/2006/Mikrozensus/Presseebroschuere.pdf?__blob=publicationFile.
492. Steckelberg A, Berger B, Köpke S, Heesen C, Mühlhauser I. Kriterien für evidenzbasierte Patienteninformationen. *Z Arztl Fortbild Qualitätssich* 2005; 99(6): 343-351.
493. Steckelberg A, Kasper J, Redegeld M, Mühlhauser I. Risk information: barrier to informed choice? A focus group study. *Soz Praventivmed* 2004; 49(6): 375-380.
494. Steiner JF. The use of stories in clinical research and health policy. *JAMA* 2005; 294(22): 2901-2904.
495. 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.
496. Sterne JAC, Egger M, Moher D. Addressing reporting biases. In: Higgins JPT, Green S (Ed). *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley; 2008. S. 297-333.
497. 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.
498. 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.
499. Sun GW, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *J Clin Epidemiol* 1996; 49(8): 907-916.

500. Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ* 2010; 340: c117.
501. Sutton A, Ades AE, Cooper N, Abrams K. Use of indirect and mixed treatment comparisons for technology assessment. *Pharmacoeconomics* 2008; 26(9): 753-767.
502. Sutton AJ, Donegan S, Takwoingi Y, Garner P, Gamble C, Donald A. An encouraging assessment of methods to inform priorities for updating systematic reviews. *J Clin Epidemiol* 2009; 62(3): 241-251.
503. Swift TL, Dieppe PA. Using expert patients' narratives as an educational resource. *Patient Educ Couns* 2005; 57(1): 115-121.
504. 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.
505. Thomas S. Klinische Relevanz von Therapieeffekten: systematische Sichtung, Klassifizierung und Bewertung methodischer Konzepte [Dissertation]. Duisburg/Essen: Universität; 2009.
506. Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002; 21(11): 1559-1573.
507. Thorlund K, Devereaux PJ, Wetterslev J, Guyatt G, Ioannidis JPA, Thabane L et al. Can trial sequential monitoring boundaries reduce spurious inferences from meta-analyses? *Int J Epidemiol* 2009; 38(1): 276-286.
508. Thorne S. The role of qualitative research within an evidence-based context: can metasynthesis be the answer? *Int J Nurs Stud* 2009; 46(4): 569-575.
509. Thurow S. Search engine visibility. Indianapolis: New Riders; 2003.
510. Tian L, Cai T, Pfeffer MA, Piankov N, Cremieux PY, Wei LJ. Exact and efficient inference procedure for meta-analysis and its application to the analysis of independent 2×2 tables with all available data but without artificial continuity correction. *Biostatistics* 2009; 10(2): 275-281.
511. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Series B Stat Methodol* 1996; 58(1): 267-288.
512. Tierney JF, Stewart LA, Gherzi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials* 2007; 8: 16.
513. Torgerson DJ. Contamination in trials: is cluster randomisation the answer? *BMJ* 2001; 322(7282): 355-357.
514. 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.
515. Tricoci P, Allen JM, Kramer JM, Califf RM, Smith SC Jr. Scientific evidence underlying the ACC/AHA clinical practice guidelines. *JAMA* 2009; 301(8): 831-841.

516. Tsay MY, Yang YH. Bibliometric analysis of the literature of randomized controlled trials. *J Med Libr Assoc* 2005; 93(4): 450-458.
517. Tukey JW. Some thoughts on clinical trials, especially problems of multiplicity. *Science* 1977; 198(4318): 679-684.
518. 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.
519. Turner RM, Omar RZ, Yang M, Goldstein H, Thompson SG. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Stat Med* 2000; 19(24): 3417-3432.
520. Tyczynski JE, Demaret E, Parkin DM (Ed). Standards and guidelines for cancer registration in Europe: the ENCR recommendations vol.1. Lyon: IARC Press; 2003. (IARC Technical Publications; Band 40).
521. UK National Screening Committee. Programme appraisal criteria: criteria for appraising the viability, effectiveness and appropriateness of a screening programme [online]. [Zugriff: 18.04.2013]. URL: <http://www.screening.nhs.uk/criteria>.
522. USAID Center for Development Information and Evaluation. Conducting key informant interviews [online]. 1996 [Zugriff: 07.05.2013]. (Performance Monitoring and Evaluation TIPS; Band 2). URL: http://pdf.usaid.gov/pdf_docs/PNABS541.pdf.
523. 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.
524. Van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med* 2002; 21(4): 589-624.
525. 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.
526. Van Tulder M, Furlan A, Bombardier C, Bouter L. Updated method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group. *Spine (Phila Pa 1976)* 2003; 28(12): 1290-1299.
527. Vandembroucke 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.
528. Vergouwe Y, Royston P, Moons KG, Altman DG. Development and validation of a prediction model with missing predictor data: a practical approach. *J Clin Epidemiol* 2010; 63(2): 205-214.

529. Vidanapathirana J, Abramson MJ, Forbes A, Fairley C. Mass media interventions for promoting HIV testing. *Cochrane Database Syst Rev* 2005; (3): CD004775.
530. Vijan S. Should we abandon QALYs as a resource allocation tool? *Pharmacoeconomics* 2006; 24(10): 953-954.
531. 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.
532. Virtanen H, Leino-Kilpi H, Salantera S. Empowering discourse in patient education. *Patient Educ Couns* 2007; 66(2): 140-146.
533. Vlayen J, Aertgeerts B, Hannes K, Sermeus W, Ramaekers D. A systematic review of appraisal tools for clinical practice guidelines: multiple similarities and one common deficit. *Int J Qual Health Care* 2005; 17(3): 235-242.
534. Voisin CE, De la Varre C, Whitener L, Gartlehner G. Strategies in assessing the need for updating evidence-based guidelines for six clinical topics: an exploration of two search methodologies. *Health Info Libr J* 2008; 25(3): 198-207.
535. 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.
536. 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.
537. Walter U, Schwartz FW. Prävention. In: Schwartz FW, Badura B, Busse R, Leidl R, Raspe H, Siegrist J et al (Ed). *Das Public Health Buch: Gesundheit und Gesundheitswesen*. München: Urban und Fischer; 2003. S. 189-214.
538. 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.
539. Watine J, Friedberg B, Nagy E, Onody R, Oosterhuis W, Bunting PS et al. Conflict between guideline methodologic quality and recommendation validity: a potential problem for practitioners. *Clin Chem* 2006; 52(1): 65-72.
540. Web Accessibility Initiative. Webauftritt [online]. [Zugriff: 22.04.2013]. URL: <http://www.w3.org/WAI>.
541. Weir CJ, Walley RJ. Statistical evaluation of biomarkers as surrogate endpoints: a literature review. *Stat Med* 2006; 25(2): 183-203.
542. Wendt C. Gesundheitssysteme im internationalen Vergleich. *Gesundheitswesen* 2006; 68(10): 593-599.

543. Wetterslev J, Thorlund K, Brok J, Gluud C. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. *J Clin Epidemiol* 2008; 61(1): 64-75.
544. Whitehead J. The design and analysis of sequential clinical trials. Chichester: Horwood; 1983.
545. Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med* 2004; 140(3): 189-202.
546. Whiting P, Rutjes AWS, 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.
547. Whitlock EP, Lin JS, Chou R, Shekelle P, Robinson KA. Using existing systematic reviews in complex systematic reviews. *Ann Intern Med* 2008; 148(10): 776-782.
548. Wieseler B, McGauran N, Kaiser T. Still waiting for functional EU Clinical Trials Register. *BMJ* 2011; 342: d3834.
549. Windeler J. Bedeutung randomisierter klinischer Studien mit relevanten Endpunkten für die Nutzenbewertung [online]. In: Gesundheitsforschungsrat des Bundesministeriums für Bildung und Forschung (Ed). Diskussionsforum zur Nutzenbewertung im Gesundheitswesen: Begriffsdefinitionen und Einführung; Dokumentation des ersten gemeinsamen Workshops von GFR und IQWiG am 4. September 2007 in Berlin. 12.2007. S. 26-31 [Zugriff: 19.04.2013]. URL: http://www.gesundheitsforschung-bmbf.de/media/DLR_Nutzenbewert_07-11-22_Druckversion.pdf.
550. Windeler J. Externe Validität. *Z Evid Fortbild Qual Gesundhwes* 2008; 102(4): 253-259.
551. Windeler J, Conradt C. Wie können "Signifikanz" und "Relevanz" verbunden werden? *Med Klin* 1999; 94(11): 648-651.
552. Windeler J, Lange S. Nutzenbewertung in besonderen Situationen: seltene Erkrankungen. *Z Evid Fortbild Qual Gesundhwes* 2008; 102(1): 25-30.
553. Windeler J, Ziegler S. Evidenzklassifizierungen. *Z Arztl Fortbild Qualitatssich* 2003; 97(6): 513-514.
554. 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.
555. 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.
556. 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.

557. 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.

558. Zschorlich B, Knelangen M, Bastian H. Die Entwicklung von Gesundheitsinformationen unter Beteiligung von Bürgerinnen und Bürgern am Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Gesundheitswesen 2011; 73(7): 423-429.

559. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. BMJ 2008; 337: a2390.