Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Institute for Quality and Efficiency in Health Care

Allgemeine Methoden

Version 4.2 vom 22.04.2015

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 978-3-9815265-1-6

© 2015 Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, Köln
www.iqwig.de

Anschrift des Herausgebers:
Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
Im Mediapark 8 (Köln Turm)
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. Somit leistet das Methodenpapier 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 im neuen Anhang,
- Verwendung von Prädiktionsintervallen für Meta-Analysen mit zufälligen Effekten in Abschnitt 8.3.8 (vorher Abschnitt 7.3.8).

In der vorliegenden Version 4.2 werden im Vergleich zur Version 4.1 der „Allgemeinen Methoden“ des Instituts vom 28.11.2013 kleinere Fehler beseitigt, redaktionelle Änderungen durchgeführt sowie aktuelle Literaturzitate ergänzt. Es haben folgende inhaltliche Änderungen stattgefunden:

- Ergänzung eines Abschnitts zu den Standards der Gesundheitsökonomie in Kapitel 1 (neuer Abschnitt 1.3),
- Ergänzung des neuen Produkts Potenzialbewertung gemäß § 137e SGB V sowie der entsprechenden Methoden in den Abschnitten 1.1, 2.1, 2.2, 3.8, 7.2 (vorher Abschnitt 6.2), 8.1.3 (vorher Abschnitt 7.1.3) und 8.1.4 (vorher Abschnitt 7.1.4),
- Überarbeitung der Ausführungen zu Gesundheitsinformationen nach Änderung der Formate infolge des Relaunches der Website www.gesundheitsinformation.de vom 13.02.2014 in Abschnitt 2.1.7 und Kapitel 6 (vorher Kapitel 5),
- Neufassung von Abschnitt 3.5 zu diagnostischen Verfahren mit Integration des alten Abschnitts 3.8 zu Prognosestudien,
- Integration der Methoden für Kosten-Nutzen-Bewertungen als neues Kapitel 4 und damit verbundene Anpassungen in Abschnitt 3.1.5,
- Ergänzung zum Umgang mit unaufgefordert übermittelten Daten in Kapitel 7 (vorher Kapitel 6),
- Ergänzungen zur Evidenzhierarchie nicht randomisierter Studien in Abschnitt 8.1.3 (vorher Abschnitt 7.1.3),
- Ergänzung zu patientenrelevanten Endpunkten in Abschnitt 8.3.3 (vorher Abschnitt 7.3.3).

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Inhaltsverzeichnis

Präambel ..... **ii**
Was ist neu? ..... **iii**
Inhaltsverzeichnis ..... **iv**
Tabellenverzeichnis ..... **x**
Abbildungsverzeichnis ..... **xi**
Abkürzungsverzeichnis ..... **xii**
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 Die Strategien der EbM 6
1.2.4 Die Bedeutung der Ergebnissicherheit 6
1.2.5 Die Verbindung von Ergebnissicherheit und Alltagsnähe 8
1.2.6 Nutzen im Einzelfall 10
1.3 Gesundheitsökonomie **10**
1.3.1 Bedeutung von Gesundheitsökonomie für das Institut 11
1.3.2 Die internationalen Standards der Gesundheitsökonomie 12
1.3.3 Methodische Standards in der Gesundheitsökonomie 13
1.4 Evidenzbasierte Entscheidungsfindung im Gesundheitswesen **14**
2 Produkte des Instituts **15**
2.1 Produktspezifische Verfahrensabläufe **15**
2.1.1 Bericht 17
2.1.2 Rapid Report 22
2.1.3 Dossierbewertung 23
2.1.4 Kosten-Nutzen-Bewertung nach § 35b SGB V 25
2.1.5 Potenzialbewertung 28
2.1.6 Addendum 29
2.1.7 Gesundheitsinformationen 30
2.1.8 Arbeitspapier 32
2.2 Generelle Aspekte bei der Produkterstellung **34**
2.2.1 Auswahl externer Sachverständiger 34
2.2.2 Gewährleistung der fachlichen Unabhängigkeit 35

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

2.2.3 Review der Produkte des Instituts 36
2.2.4 Veröffentlichung der Produkte des Instituts 37
3 Nutzenbewertung medizinischer Interventionen 38
3.1 Patientenrelevanter medizinischer Nutzen und Schaden 38
3.1.1 Definition des patientenrelevanten medizinischen Nutzens bzw. Schadens 38
3.1.2 Surrogate patientenrelevanter Endpunkte 40
3.1.3 Ermittlung des Schadens medizinischer Interventionen 42
3.1.4 Endpunktbezogene Bewertung 44
3.1.5 Zusammenfassende Bewertung 49
3.2 Spezielle Aspekte der Nutzenbewertung 49
3.2.1 Auswirkung nicht publizierter Studienergebnisse auf das Fazit 49
3.2.2 Dramatischer Effekt 51
3.2.3 Studiendauer 52
3.2.4 Patientenberichtete Endpunkte 53
3.2.5 Nutzen und Schaden in kleinen Populationen 53
3.3 Nutzenbewertung von Arzneimitteln 54
3.3.1 Stellenwert des Zulassungsstatus 54
3.3.2 Studien zur Nutzenbewertung von Arzneimitteln 56
3.3.3 Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V 57
3.4 Nichtmedikamentöse therapeutische Interventionen 62
3.5 Diagnostische Verfahren 63
3.6 Früherkennung und Screening 67
3.7 Prävention 68
3.8 Potenzialbewertung 69
4 Kosten-Nutzen-Bewertung medizinischer Interventionen 72
4.1 Einführung 72
4.1.1 Rechtliche Grundlagen für die Kosten-Nutzen-Bewertung nach SGB V 72
4.1.2 Perspektive 72
4.1.3 Zeithorizont 72
4.1.4 Auswahl der Komparatoren 73
4.1.5 Versorgungspfad 73
4.1.6 Modell 74
4.1.7 Datengrundlage 74
4.1.8 Unsicherheit 75
4.1.9 Interpretation der Ergebnisse 75
4.2 Modellierung 78
4.2.1 Grundlagen 78

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

4.2.2 Grundsätzliche Aspekte der Modellentwicklung ... 78
4.2.3 Einflussdiagramm und Modellkonzept ... 79
4.2.4 Auswahl der Modellierungstechnik ... 79
4.2.5 Modelldokumentation und Modellvalidierung ... 80
4.3 Nutzen ... 83
4.3.1 Übertragung und Darstellung des Nutzens ... 83
4.3.2 Endpunkte ... 83
4.3.3 Maß des Gesamtnutzens ... 83
4.3.4 Unsicherheit und Verteilung von Nutzendaten ... 87
4.4 Kosten ... 88
4.4.1 Perspektive und zu berücksichtigende Kosten ... 88
4.4.2 Abgrenzung der Kosten ... 89
4.4.3 Schritte zur Kostenbestimmung ... 91
4.4.4 Datengrundlage ... 94
4.4.5 Unsicherheit und Verteilung von Kostendaten ... 95
4.4.6 Inflationsbereinigung und Diskontierung ... 95
4.5 Epidemiologische Daten ... 95
4.5.1 Daten ... 95
4.5.2 Datengrundlage ... 96
4.5.3 Unsicherheit und Verteilung von epidemiologischen Daten ... 96
4.6 Ergebnisdarstellung als Effizienzgrenze ... 96
4.6.1 Definition ... 96
4.6.2 Verfahrensverlauf ... 97
4.6.3 Konstruktion der Effizienzgrenze ... 100
4.6.4 Sonderkonstellationen ... 102
4.7 Unsicherheit (Sensitivitätsanalysen) ... 103
4.7.1 Quantifizierung von Unsicherheit ... 103
4.7.2 Sensitivitätsanalysen ... 103
4.7.3 Darstellung von Unsicherheit mittels des Net Health Benefit ... 104
4.8 Ausgaben-Einfluss-Analyse (Budget-Impact-Analyse) ... 104
4.8.1 Perspektive in der Ausgaben-Einfluss-Analyse ... 105
4.8.2 Zeithorizont in der Ausgaben-Einfluss-Analyse ... 105
4.8.3 Szenarien in der Ausgaben-Einfluss-Analyse ... 105
4.8.4 Population in der Ausgaben-Einfluss-Analyse ... 105
4.8.5 Einzuschließende Kosten in der Ausgaben-Einfluss-Analyse ... 105
4.8.6 Ergebnisdarstellung in der Ausgaben-Einfluss-Analyse ... 106
4.9 Besonderheiten der Kosten-Nutzen-Bewertung nach § 35b SGB V ... 106

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

4.9.1 Gesetzliche Anforderungen und Verfahrensablauf ... 106
4.9.2 Der Net Health Benefit zur Berechnung zusatznutzenbasierter Erstattungspreise ... 107
4.9.3 Sensitivitätsanalysen zur Berechnung zusatznutzenbasierter Erstattungspreise 108
4.9.4 Interquartilsregion als Maß der Streuung für Preisverhandlungen ... 108
5 Leitlinien und Versorgungsanalyse ... 110
5.1 Hintergrund ... 110
5.2 Identifikation von Versorgungsstandards über Leitlinien ... 110
5.2.1 Versorgungsstandards in Leitlinien ... 110
5.2.2 Methodische Bewertung von Leitlinien ... 111
5.2.3 Strukturierte Aufbereitung und Auswertung von Empfehlungen ... 112
5.3 Empfehlungen zu Disease-Management-Programmen ... 114
5.4 Versorgungsanalyse ... 114
5.4.1 Hintergrund ... 114
5.4.2 Inhaltliche Aspekte einer Versorgungsanalyse ... 115
5.4.3 Ziele einer Versorgungsanalyse ... 116
5.4.4 Fragestellung der Versorgungsanalyse ... 116
5.4.5 Mögliche Versorgungsparameter ... 117
5.4.6 Ablauf einer Versorgungsanalyse ... 118
5.4.7 Ebenen der Versorgungsanalyse ... 120
5.4.8 Methodische Besonderheiten einer Versorgungsanalyse ... 120
5.4.9 Informationsbeschaffung ... 121
5.4.10 Bewertung der identifizierten Daten ... 123
5.4.11 Informationssynthese und -analyse ... 123
5.4.12 Darstellung und Bewertung der Versorgungsqualität ... 124
5.5 Validität von Leitlinienempfehlungen ... 124
5.5.1 Hintergrund ... 124
5.5.2 Validität von Leitlinienempfehlungen ... 125
5.5.3 Ziel der Analyse und Bewertung von Leitlinienempfehlungen ... 125
5.5.4 Vorgehen bei der Analyse und Bewertung der internen Validität ... 126
5.5.5 Mögliche Fragestellungen ... 126
6 Evidenzbasierte Gesundheitsinformation für Bürgerinnen und Bürger ... 127
6.1 Hintergrund und Ziele ... 127
6.2 Patientenzentrierte Kommunikation ... 128
6.2.1 Kommunikationsstandards ... 128
6.2.2 Methode der Patientenwege ... 130
6.2.3 Beteiligung von Bürgerinnen und Bürgern ... 133
6.2.4 Visuelle Kommunikation und Multimedia ... 133

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

6.2.5 Informationszugang und Barrierefreiheit 134
6.3 Themenwahl, Informationsbeschaffung und wissenschaftliche Bewertung 135
6.3.1 Themenwahl 135
6.3.2 Informationsbeschaffung für Gesundheitsinformationen 137
6.3.3 Evidenzbewertung 138
6.3.4 Aktualisierung 139
6.4 Informationsprodukte 140
6.4.1 Hauptformate 140
6.4.2 Ergänzende Formate 141
6.4.3 Erfahrungsberichte 142
6.4.4 Website 143
7 Informationsbeschaffung 145
7.1 Eigene Informationsbeschaffung durch das Institut 145
7.1.1 Ablauf einer Recherche 146
7.1.2 Bibliografische Datenbanken 146
7.1.3 Suche in Studienregistern 147
7.1.4 Leitliniendatenbanken und -anbieter 148
7.1.5 Herstelleranfragen 148
7.1.6 Sonstige Datenquellen für die Recherche 149
7.1.7 Auswahl relevanter Publikationen 149
7.1.8 Dokumentation der Recherche 150
7.1.9 Ergänzungsrecherchen bei systematischen Übersichten 151
7.2 Überprüfung der Informationsbeschaffung 151
8 Informationsbewertung 152
8.1 Qualitätsbewertung von Einzelstudien 152
8.1.1 Kriterien zum Einschluss von Studien 152
8.1.2 Zusammenhang zwischen Studientyp/-art und Fragestellung 152
8.1.3 Rangordnung verschiedener Studienarten / Evidenzgrade 153
8.1.4 Aspekte der Bewertung des Verzerrungspotenzials 154
8.1.5 Interpretation von kombinierten Endpunkten 157
8.1.6 Interpretation von Subgruppenanalysen 158
8.1.7 Bewertung der Konsistenz von Daten 160
8.2 Berücksichtigung systematischer Übersichten 161
8.2.1 Einordnung systematischer Übersichten 161
8.2.2 Nutzenbewertung auf Basis systematischer Übersichten 162
8.2.3 Berücksichtigung publizierter Meta-Analysen 164
8.3 Spezielle biometrische Aspekte 165

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

8.3.1 Darstellung von Effekten und Risiken ... 165
8.3.2 Beurteilung statistischer Signifikanz ... 166
8.3.3 Beurteilung klinischer Relevanz ... 167
8.3.4 Bewertung subjektiver Endpunkte bei offenen Studiendesigns ... 169
8.3.5 Nachweis der Verschiedenheit ... 170
8.3.6 Nachweis der Gleichheit ... 171
8.3.7 Adjustierung und multifaktorielle Verfahren ... 172
8.3.8 Meta-Analysen ... 173
8.3.9 Indirekte Vergleiche ... 178
8.3.10 Umgang mit nicht oder nicht vollständig publizierten Daten ... 179
8.3.11 Darstellung von Verzerrungsarten ... 180
8.4 Qualitative Methoden ... 182
8.4.1 Qualitative Studien ... 182
8.4.2 Konsultationstechniken ... 184
Anhang A – Rationale der Methodik zur Feststellung des Ausmaßes des Zusatznutzens ... 186
Literaturverzeichnis ... 198

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Tabellenverzeichnis

Seite

Tabelle 1: Übersicht über die Produkte des Instituts 17
Tabelle 2: Regelhaft abgeleitete Aussagesicherheiten für verschiedene Evidenzsituationen beim Vorliegen von Studien derselben qualitativen Ergebnissicherheit 48
Tabelle 3: Szenarien der Datenübermittlung und Konsequenzen für das Fazit einer Nutzenbewertung 50
Tabelle 4: Schwellenwerte zur Feststellung des Ausmaßes eines Effekts 60
Tabelle 5: Konzepte von Unsicherheit in der gesundheitsökonomischen Entscheidungsanalyse 75
Tabelle 6: Perspektive und relevante zu berücksichtigende Kosten 88
Tabelle 7: Beispiele möglicher Versorgungsparameter 117
Tabelle 8: Informationsquellen zur Identifizierung von deutschen Versorgungsstandards 121
Tabelle 9: Informationsquellen zur Identifizierung von Versorgungsdaten 122
Tabelle 10: Verschiedene Dimensionen eines Patientenweges 132
Tabelle 11: Mögliche Aspekte zur Priorisierung der Themen 137
Tabelle 12: Feststellung des Ausmaßes des Zusatznutzens – Kriterien gemäß AM-NutzenV 187
Tabelle 13: Feststellung des Ausmaßes des Zusatznutzens – Kriterien gemäß AM-NutzenV mit Ergänzungen* 189
Tabelle 14: Feststellung des Ausmaßes des Zusatznutzens – hierarchisierte Kriterien gemäß AM-NutzenV mit Ergänzungen* 190
Tabelle 15: Inferenzstatistische Schwellenwerte (Hypothesengrenzen) für relative Effektmaße 194
Tabelle 16: Tatsächliche Effekte für das relative Risiko 197

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Abbildungsverzeichnis

Seite

Abbildung 1: Ablauf der Berichtserstellung 19
Abbildung 2: Ablauf der Erstellung eines Rapid Reports 22
Abbildung 3: Ablauf der Erstellung einer Dossierbewertung 24
Abbildung 4: Ablauf der Kosten-Nutzen-Bewertung nach § 35b SGB V 26
Abbildung 5: Ablauf der Erstellung einer Potenzialbewertung 28
Abbildung 6: Ablauf der Erstellung eines Addendums 29
Abbildung 7: Ablauf der Gesundheitsinformationserstellung 31
Abbildung 8: Ablauf der Erstellung eines Arbeitspapiers 33
Abbildung 9: Darstellung der entscheidungsrelevanten Bereiche 97
Abbildung 10: Interpretation der Steigung der theoretischen Effizienzgrenze 98
Abbildung 11: Absolute versus erweiterte Dominanz 99
Abbildung 12: Darstellung der Effizienzgrenze 102
Abbildung 13: Darstellung eines NHB > 0 108
Abbildung 14: Interquartilsregion möglicher zusatznutzenbasierter Erstattungspreise (basierend auf PSA) als Maß der Streuung für Preisverhandlungen 109
Abbildung 15: Exemplarischer Ablauf einer Versorgungsanalyse 119
Abbildung 16: Tatsächliche Effekte in Abhängigkeit des Basisrisikos 196

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Abkürzungsverzeichnis

Abkürzung	Definition
AEA	Ausgaben-Einfluss-Analyse
AGREE	Appraisal of Guidelines Research and Evaluation in Europe
AHP-Verfahren	Analytic-Hierarchy-Process-Verfahren
AMG	Arzneimittelgesetz
AMIS	Arzneimittel-Informationssystem
AMNOG	Arzneimittelmarktneuordnungsgesetz
AM-NutzenV	Arzneimittel-Nutzenbewertungsverordnung
AMSTAR-Instrument	A Measurement Tool to Assess Systematic Reviews
ANP	Analytic Network Process (analytischer Netzwerkprozess)
ANP-Verfahren	Analytic-Network-Process-Verfahren
AOK	Allgemeine Ortskrankenkasse
AQUA	Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen
AQUIK	ambulante Qualitätsindikatoren und Kennzahlen
AWMF	Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e. V.
ÄZQ	Ärztliches Zentrum für Qualität in der Medizin
BÄK	Bundesärztekammer
BITV	barrierefreie Informationstechnik-Verordnung
BMG	Bundesministerium für Gesundheit
BMJV	Bundesministerium für Justiz und Verbraucherschutz
BQS	Institut für Qualität und Patientensicherheit
CA	Conjoint-Analyse
CONSORT	Consolidated Standards of Reporting Trials
DALY	Disability-Adjusted Life Year (um Behinderungen adjustiertes Lebensjahr)
DARE	Database of Abstracts of Reviews of Effects
DCE	Discrete-Choice-Experiment
DELBI-Instrument	Deutsches Leitlinien-Bewertungsinstrument
DMP	Disease-Management-Programm
DRG	Diagnosis Related Groups
EbM	evidenzbasierte Medizin
EBM	einheitlicher Bewertungsmaßstab
EMA	European Medicines Agency (Europäische Zulassungsbehörde)

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Abkürzung	Definition
FDA	Food and Drug Administration (behördliche Lebensmittelüberwachungs- und Arzneimittelzulassungsbehörde der USA)
G-BA	Gemeinsamer Bundesausschuss
GCP	Good Clinical Practice (gute klinische Praxis)
GEP	Gute Epidemiologische Praxis
G-I-N	Guidelines International Network
GKV	gesetzliche Krankenversicherung
GKV-Spitzenverband	Spitzenverband Bund der Krankenkassen
GMG	GKV-Modernisierungsgesetz
GoR	Grade of Recommendation
GRADE	The Grading of Recommendations Assessment, Development and Evaluation (short GRADE) Working Group
HERG	Health Experiences Research Group
HON	Health on the Net Foundation
HONcode	HON Code of Conduct
HTA	Health Technology Assessment
HVPI	harmonisierter Verbraucherpreisindex
HYE	Healthy Years Equivalent
ICD	International Statistical Classification of Diseases and Related Health Problems (Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme)
ICF	International Classification of Functioning, Disability and Health (Internationale Klassifikation der Funktionsfähigkeit, Behinderung und Gesundheit)
ICH	International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use
ICTRP	International Clinical Trials Registry Platform Search Portal
IKNV	inkrementelles Kosten-Nutzen-Verhältnis
IPD	Individual Patient Data (individuelle Patientendaten)
IPDAS	International Patient Decision Aid Standards
IQR	Interquartilsregion
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ISOQOL	International Society of Quality of Life Research
ISPOR	International Society For Pharmacoeconomics and Outcomes Research
KBV	Kassenärztliche Bundesvereinigung

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Abkürzung	Definition
KNB	Kosten-Nutzen-Bewertung
LoE	Level of Evidence
MAUI	Multi-Attribute-Utility-Instrument
MDK	Medizinischer Dienst der Krankenversicherung
MID	Minimal Important Difference
MOOSE	Meta-analysis of Observational Studies in Epidemiology
MORE	McMaster Online Rating of Evidence
MPG	Medizinproduktegesetz
MTC	Mixed Treatment Comparison
MTM	Multiple Treatment Meta-Analyse
NGC	National Guideline Clearinghouse
NHB	Net Health Benefit
NHC	National Health Committee
NIH	National Institute of Health
NUB	neue Untersuchungs- und Behandlungsmethoden
OECD	Organisation for Economic Co-operation and Development (Organisation für wirtschaftliche Zusammenarbeit und Entwicklung)
OR	Odds Ratio
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRO	Patient Reported Outcomes (patientenberichtete Endpunkte)
PSA	probabilistische Sensitivitätsanalyse
QALY	Quality-Adjusted Life Year (qualitätsadjustiertes Lebensjahr)
RCT	Randomized Controlled Trial (randomisierte kontrollierte Studie)
RD	Risikodifferenz
RKI	Robert Koch-Institut
ROC-Kurve	Receiver-Operating-Characteristic-Kurve
RR	Relatives Risiko
SGB V	Sozialgesetzbuch – Fünftes Buch – Gesetzliche Krankenversicherung
SMD	standardisierte Mittelwertdifferenzen
SMDM	Society for Medical Decision Making
STARD	Standards for Reporting of Diagnostic Accuracy
STE	Surrogate-Threshold-Effekt

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Abkürzung	Definition
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
SW	Schwellenwert
TREND	Transparent Reporting of Evaluations with Non-randomized Designs
UK NSC	United Kingdom National Screening Committee
US PSTF	United States Preventive Services Task Force
WHO	World Health Organization (Weltgesundheitsorganisation)
WIdO	Wissenschaftliches Institut der AOK
ZE	Zusatzentgelte

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 [135] als Einrichtung der Stiftung für Qualität und Wirtschaftlichkeit im Gesundheitswesen gegründet. Die gesetzlichen Grundlagen und Aufgaben des Instituts sind im Sozialgesetzbuch – Fünftes Buch – Gesetzliche Krankenversicherung (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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 [136]. 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 [80] und die Verfahrensordnung des G-BA [21].

Verbunden mit einer Nutzenbewertung kann der G-BA das Institut auch mit einer Kosten-Nutzen-Bewertung beauftragen. Die Rahmenbedingungen dieser Kosten-Nutzen-Bewertungen geben § 35b SGB V und § 139a SGB V vor. Dabei werden Kosten-Nutzen-Verhältnisse medizinischer Technologien gegenübergestellt mit dem Ziel, Informationen bereitzustellen, auf deren Grundlage die Angemessenheit und die Zumutbarkeit einer Kostenübernahme durch die Versichertengemeinschaft berücksichtigt werden können.

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).

Im Rahmen des Versorgungsstrukturgesetzes erfolgten 2012 Änderungen des § 137c SGB V und die Hinzufügung des § 137e SGB V. Hiermit erhielt der G-BA die Möglichkeit, klinische Studien zu neuen Untersuchungs- oder Behandlungsmethoden zu initiieren (Erprobung),

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

sofern der Nutzen einer Methode noch nicht hinreichend belegt ist, die Methode jedoch das Potenzial einer erforderlichen Behandlungsalternative erkennen lässt. Auch externe Antragsteller, beispielsweise Medizinproduktehersteller, können eine Erprobung beantragen, indem sie dem G-BA aussagekräftige Unterlagen zum Potenzial der Methode vorlegen. Die Feststellung des Potenzials einer Methode obliegt dem G-BA, der hierfür Kriterien festgelegt hat [211]. Der G-BA beauftragt in der Regel das Institut damit, Erprobungsanträge gemäß § 137e Abs. 7 SGB V dahin gehend zu prüfen, ob sich aus den Antragsunterlagen ein Potenzial ergibt.

Das Institut ist gemäß § 139a Absatz 4 Satz 1 SGB V gesetzlich verpflichtet, die „Bewertung des medizinischen Nutzens nach den international anerkannten Standards der evidenzbasierten Medizin und die ökonomische Bewertung nach den hierfür maßgeblichen international anerkannten Standards, insbesondere der Gesundheitsökonomie“ zu gewährleisten. Das Institut bestimmt auftragsbezogen die Methoden und Kriterien für die Erarbeitung von Bewertungen auf der Grundlage der in den jeweiligen Fachkreisen anerkannten internationalen 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. In Abschnitt 1.3 werden der Begriff der Gesundheitsökonomie 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 Informationen 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 Eindeutschung des englischen Begriffs „Evidence-based Medicine“. Gemeint ist damit eine medizinische Betreuung von Patientinnen und Patienten, die sich nicht allein auf Meinungen und

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 [17,178,231,236].

Instrumente, die einer subjektiv geprägten und damit häufig verzerrten Bewertung vorbeugen sollen (siehe u. a. Kapitel 8), 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 [383]. 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 [267]. 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 [33]. Etwa 20 Jahre später bemühten sich klinische Epidemiologen, die Methodenlehre auch in der klinischen Praxis zu etablieren [183]. 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) [31].

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. Die folgenden Abschnitte erläutern, 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 [474] 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 [255]. 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 [328]. 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 wurden und diese Überprüfung gezeigt hat, dass vorschneller Optimismus gefährliche Folgen für Patientinnen und Patienten haben kann [157,457]. 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 [211].

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

1.2.3 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 2 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 Nutzennachweis 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.

1.2.4 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 2 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 2 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 2 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 unberechtigterweise 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 [222]. In einer wissenschaftlich unsicheren Situation gewinnen diese Aspekte an Gewicht. Zudem hat der G-BA die Möglichkeit, selbst Studien zu fordern oder zu initiieren, um die identifizierten Evidenzlücken zu schließen.

1.2.5 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 allein kein Garant für eine hohe Ergebnissicherheit. Dazu sind z. B.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 Produkts 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Diese Ausführungen zeigen, dass Ergebnissicherheit und Alltagsnähe (oder interne und externe Validität) nicht fest miteinander verknüpft sind: Hohe Ergebnissicherheit und Alltagsnähe schließen einander 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 [253].

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 [199,201,218,381,561]. 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 [596], was letztlich über adäquate Subgruppenanalysen zumindest ansatzweise zu beantworten wäre.

1.2.6 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 [232,238,315,492]. Diese sind jedoch für die Nutzenbewertung einer Behandlungsmethode für zukünftige Patientinnen und Patienten in aller Regel nicht geeignet.

1.3 Gesundheitsökonomie

Mit dem Begriff Gesundheitsökonomie können 2 Sachverhalte zum Ausdruck gebracht werden.

Im weiten Sinne geht es um „die Analyse der wirtschaftlichen Aspekte des Gesundheitswesens unter Verwendung von Konzepten der ökonomischen Theorie“ [495]. Dazu werden u. a. Konzepte aus den Bereichen der mikroökonomischen Verhaltenstheorie, der Wettbewerbstheorie, der ökonomischen Theorie der Politik und der Managementtheorie herangezogen

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

[495]. Gegenstand einer solchen Untersuchung könnte z. B. sein, wie Akteure im Gesundheitswesen durch Anreizsetzung, z. B. die Praxisgebühr, ihr Verhalten ändern, oder ob die Verhandlungslösung nach AMNOG tatsächlich vor überhöhten Preisen neuer Arzneimittel schützt. Inwieweit solche Untersuchungen zur Steuerung des Gesundheitswesens genutzt werden können und sollen, kann sowohl methodisch als auch ethisch diskutiert werden, ist aber nicht weiter Gegenstand dieser kurzen Darstellung.

Im engen Sinn versteht man unter Gesundheitsökonomie die Kosten-Nutzen-Bewertung (KNB) als vergleichende sowie auch nicht vergleichende Untersuchungen, z. B. Krankheitskostenstudien oder Ausgaben-Einfluss-Analysen. Diese Analysen dienen der Information der Entscheidungsträger zu Kosten-Nutzen-Verhältnissen von Interventionen und stellen damit neben der Nutzenbewertung einen Bereich in sog. Health Technology Assessments dar.

1.3.1 Bedeutung von Gesundheitsökonomie für das Institut

Mit Gründung des Instituts 2004 stand es dem G-BA und dem BMG offen, eine KNB zu beauftragen. Eine KNB von Arzneimitteln war bis zur Gesetzesänderung 2007 nicht vorgesehen. Mit dem GKV-Wettbewerbsstärkungsgesetz wurde in § 35b SGB V die Kosten-Nutzen-Bewertung von Arzneimitteln verankert, um Informationen für die Empfehlung eines sogenannten Höchstbetrags zu gewinnen. Bis zu diesem Höchstbetrag sollten neue Arzneimittel erstattet werden, da dieser Höchstbetrag die angemessenen Kosten für den Zusatznutzen eines neuen Arzneimittels im Vergleich zu anderen Arzneimitteln und Behandlungsformen in einer Indikation darstellen sollte. Bedingung für die Beauftragung einer KNB musste also ein Nachweis des Zusatznutzens für ein neues Arzneimittel sein, der in einer Nutzenbewertung des IQWiG vorliegen musste. Die Entwicklung der aus dieser gesundheitsökonomischen Fragestellung resultierenden Methoden ist umfassend dokumentiert [285,287,288,290-292,294,295].

Mit dem Arzneimittelmarktneuordnungsgesetz (AMNOG), das zum 01.01.2011 in Kraft trat, verlagerte sich die Bedeutung der KNB im Verfahren der frühen Nutzenbewertung von Arzneimitteln; sie ist vornehmlich für den Fall des Scheiterns der Verhandlungen um einen Preis zwischen GKV-Spitzenverband und pharmazeutischen Unternehmen und der Nichteinigung im anschließenden Schiedsverfahren vorgesehen. Die Fragestellung ist allerdings geblieben: Nach § 35b Absatz 1 Satz 4 SGB V in Verbindung mit 5. Kapitel § 32 Absatz 3 der Verfahrensordnung des G-BA [211] ist die Angemessenheit und Zumutbarkeit einer Kostenübernahme durch die Versichertengemeinschaft zu berücksichtigen. Damit der G-BA diese angemessen berücksichtigen kann, müssen ihm entsprechende Informationen vorgelegt werden. Diese Informationen liefern die KNB (Angemessenheit) und die Ausgaben-Einfluss-Analyse (Zumutbarkeit). Die Bewertung der Angemessenheit und der Zumutbarkeit einer Kostenübernahme erfolgt im Hinblick darauf, ob unter Beachtung des Grundsatzes der Verhältnismäßigkeit eine begründbare Relation zwischen den Kosten und dem Nutzen des Arzneimittels besteht. Dabei soll das IQWiG nach 5. Kapitel § 32 Absatz 2 und 3 der Verfahrensordnung des G-BA eine Empfehlung vorlegen, auf deren Basis der G-BA einen

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Beschluss fällen soll [211]. Die Darstellung einer begründbaren Relation zwischen den Kosten und dem Nutzen muss also aus der KNB hervorgehen.

Auch wenn nirgendwo im Gesetz oder in den nachgeordneten Bestimmungen auf die Frage eingegangen wird, wie die Gesundheitsökonomie (siehe Abschnitt 1.3) zu verstehen ist, ergibt es sich aus der Anwendung, dass es sich um KNB und damit um die Gesundheitsökonomie im engen Sinne handelt.

1.3.2 Die internationalen Standards der Gesundheitsökonomie

Wie in jeder Wissenschaft gibt es auch in der Gesundheitsökonomie internationale Standards. Dazu gehören die Einteilung der KNB in die Studienformen Kosten-Effektivitäts- / Wirksamkeits-Analyse (Cost-Effectiveness-Analysis), Kosten-Nutzwert-Analyse (Cost-Utility-Analysis) und Kosten-Nutzen-Analyse (im engeren Sinne) (Cost-Benefit-Analysis). Manchmal werden dazu noch die Kosten-Kosten-, die Kosten-Konsequenzen- und die Kosten-Minimierungs-Analysen als eigene Formen benannt, die jedoch selten eingesetzt werden. Bei Letzterer wird auch diskutiert, ob es sich um eine eigenständige Form handelt [159].

Auch zum Vorgehen bei KNB gibt es internationale Standards. Auf seiten der Nutzenbewertung folgt das Institut mit seiner Orientierung an der evidenzbasierten Medizin und den daraus sich ergebenden Konkretionen im Bereich den internationalen Standards. Bevor man im Bereich der Gesundheitsökonomie von internationalen Standards spricht, muss man unterscheiden zwischen klar methodischen Fragen und Fragen, die auf Wertentscheidungen, Einschätzungen oder Umfragen basieren. Dies kann am Beispiel der Diskontierungsrate verdeutlicht werden. Mit einer Diskontierungsrate werden Nutzen und Kosten, die in verschiedenen Perioden anfallen, auf eine Periode abgezinst, um damit für eine Entscheidung jetzt vergleichbar zu sein. Die reine Durchführung der Diskontierung ist mathematisch klar geregelt, somit eine methodische Frage. Die Wahl der Diskontierungsrate und insbesondere die Entscheidung, ob Kosten und Nutzen mit derselben Rate oder ggf. sogar einer nicht konstanten Rate abgezinst werden sollen, unterliegt unter anderem Fragen der Einschätzung der zukünftigen wirtschaftlichen Entwicklung und der Generationengerechtigkeit [103,250,409,421,423,433,446] und ist damit eine Wertentscheidung.

Wie international anerkannte Instrumente der Bewertung von gesundheitsökonomischen Analysen zeigen [101,158,280,440], gibt es eine Reihe von Schritten und Aspekten, für die methodische Vorgaben bestehen und die transparent und nachvollziehbar abgearbeitet werden müssen. Dazu gehören:

- Definition der zu bewertenden Interventionen und ihrer Komparatoren. Eine Auswahl muss begründet werden, um Fehlentscheidungen auf Grundlage einer aus Interessen gesteuerten Komparatorenauswahl vorzubeugen.
- Perspektive der KNB

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

- Zeithorizont der KNB
- Art der KNB (siehe oben) und möglichst Begründung des Studientyps
- Kosten mit Darstellung von Ressourcenverbrauch und Ressourcenbewertung
- Inflationsbereinigung und Währungsumrechnung (falls notwendig)
- Entwicklung und Erläuterung des Modells und möglichst auch Begründung der Modellwahl, z. B. Entscheidungsbaum, Markov-Modell
- Diskontierungsrate
- Darstellung der Ergebnisse, z. B. in aggregierter und disaggregierter Form
- Untersuchung der Unsicherheit der Ergebnisse mittels deterministischer und probabilistischer Sensitivitätsanalysen
- Darstellung der Unsicherheit, z. B. mit sog. Kosten-Effektivitäts-Akzeptanz-Kurven oder dem sog. Net Benefit

Hierzu gibt es in Lehrbüchern und auch z. B. in den Leitlinien der ISPOR zu einzelnen dieser Themen oder Teilthemen Vorgaben für eine gute methodische Praxis.

1.3.3 Methodische Standards in der Gesundheitsökonomie

Im Weiteren werden auch andere Aspekte einer KNB als internationaler Standard verstanden. So wird oft eine an einem Schwellenwert je gewonnenes qualitätsadjustiertes Lebensjahr (QALY) orientierte Entscheidung im Gesundheitssystem als internationaler Standard der Gesundheitsökonomie ausgegeben. Dies ist kritisch zu sehen. Zum einen hat die ganz große Mehrheit der Länder, in denen KNB zur Entscheidungsfindung beitragen, keinen (festen) Schwellenwert. Zum anderen wäre dies eine Wertentscheidung und fiele somit nicht in die internationalen methodischen Standards, um die es sich nach § 139 a Absatz 4 Satz 1 SGB V nur handeln kann, wenn es um Bewertungen des Instituts geht.

Darüber hinaus stellt sich die Frage nach dem Maß des Gesamtnutzens nicht nur als methodische Frage, sondern immer auch unter dem Aspekt einer Wertentscheidung. Dabei haben Fragestellung und Ziel einer gesundheitsökonomischen Analyse einen Einfluss darauf, welches Instrument man als Maß des Gesamtnutzens einsetzen soll. D. h., auch die Frage, ob das QALY herangezogen werden soll, muss unter ethischen, rechtlichen und kulturellen Aspekten beleuchtet werden. Wissenschaftlich bzw. methodisch wiederum kann diskutiert werden, welche Annahmen in das QALY-Konzept einfließen, z. B. die Annahme, dass die Bewertung eines Zustands unabhängig von seiner Dauer ist (= constant proportional (time) trade off), und ob diese Annahmen tragfähig sind. Ebenfalls kann methodisch untersucht werden, ob die verschiedenen Erhebungsmethoden, z. B. indirekt vs. direkt bzw. verschiedene indirekte und direkte Verfahren in unterschiedlicher Koppelung, zu unterschiedlichen Ergebnissen führen und was das bedeuten kann. Eine Frage der Wertentscheidung auf Basis rechtlicher Vorgaben, z. B. des SGB V, ist wieder, bei wem die Nutzwerte zur Generierung

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

von QALYs erhoben werden sollen, bei den tatsächlich von einer Krankheit Betroffenen oder bei der Allgemeinbevölkerung.

Letztlich ist die Frage, auf Grundlage welcher Rechte, Ansprüche oder Bedürfnisse Ressourcen im Gesundheitswesen mit welchem Ziel und mit welcher Wirkung auf die Allokation und Distribution von Gütern bzw. Leistungen verteilt werden soll, nur als Wertentscheidung aufzufassen, und sie bestimmt wiederum, welche wissenschaftlichen Standards und Methoden eingesetzt werden sollen.

1.4 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 und der Gesundheitsökonomie.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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,
- Kosten-Nutzen-Bewertung nach § 35b SGB V,
- Potenzialbewertung,
- 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. Dies erfolgt im weiteren Verfahren durch den G-BA.

Weiterhin kann das Institut vom G-BA nach § 35b SGB V mit Kosten-Nutzen-Bewertungen von Arzneimitteln beauftragt werden. Für Kosten-Nutzen-Bewertungen nach § 35b SGB V ist

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

ein Stellungnahmeverfahren (Anhörung) durch das Institut vorgesehen. Beim G-BA wird ein weiteres Stellungnahmeverfahren durchgeführt.

Potenzialbewertungen erfolgen im Auftrag des G-BA und beziehen sich auf Erprobungsanträge gemäß § 137e SGB V. Ein Anhörungsverfahren beim Institut erfolgt nicht. Sofern es zu einer Erprobung kommt, führt der G-BA ein Stellungnahmeverfahren zur Erprobungsrichtlinie durch.

Addenda können vom G-BA oder BMG beauftragt werden, wenn sich nach Erstellung eines Produkts 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 2.1.8) 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 Instituts in einer Übersicht dargestellt. Die produktspezifischen Verfahrensabläufe sind in den sich daran anschließenden Abschnitten 2.1.1 bis 2.1.8 beschrieben.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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
Kosten-Nutzen-Bewertung nach § 35b SGB V	Bewertung des Kosten-Nutzen-Verhältnisses von Arzneimitteln gemäß § 35b SGB V	beschrieben in Abschnitt 2.1.4	G-BA
Potenzial-bewertung	Bewertung des Potenzials neuer Untersuchungs- und Behandlungsmethoden gemäß § 137e SGB V	beschrieben in Abschnitt 2.1.5	G-BA
Addendum	zeitnahe Stellungnahme zu Fragestellungen, die sich im Zuge der Beratung abgeschlossener Institutsprodukte ergeben haben	beschrieben in Abschnitt 2.1.6	G-BA, BMG
Gesundheits-information	allgemeinverständliche Informationen für Bürgerinnen und Bürger und Patientinnen und Patienten, breit gefächertes Themenspektrum	beschrieben in Abschnitt 2.1.7	G-BA, BMG; auch Initiierung durch das Institut
Arbeitspapier	Information zu versorgungsrelevanten Entwicklungen in der Medizin oder zu methodischen Aspekten	beschrieben in Abschnitt 2.1.8	Initiierung durch das Institut

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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

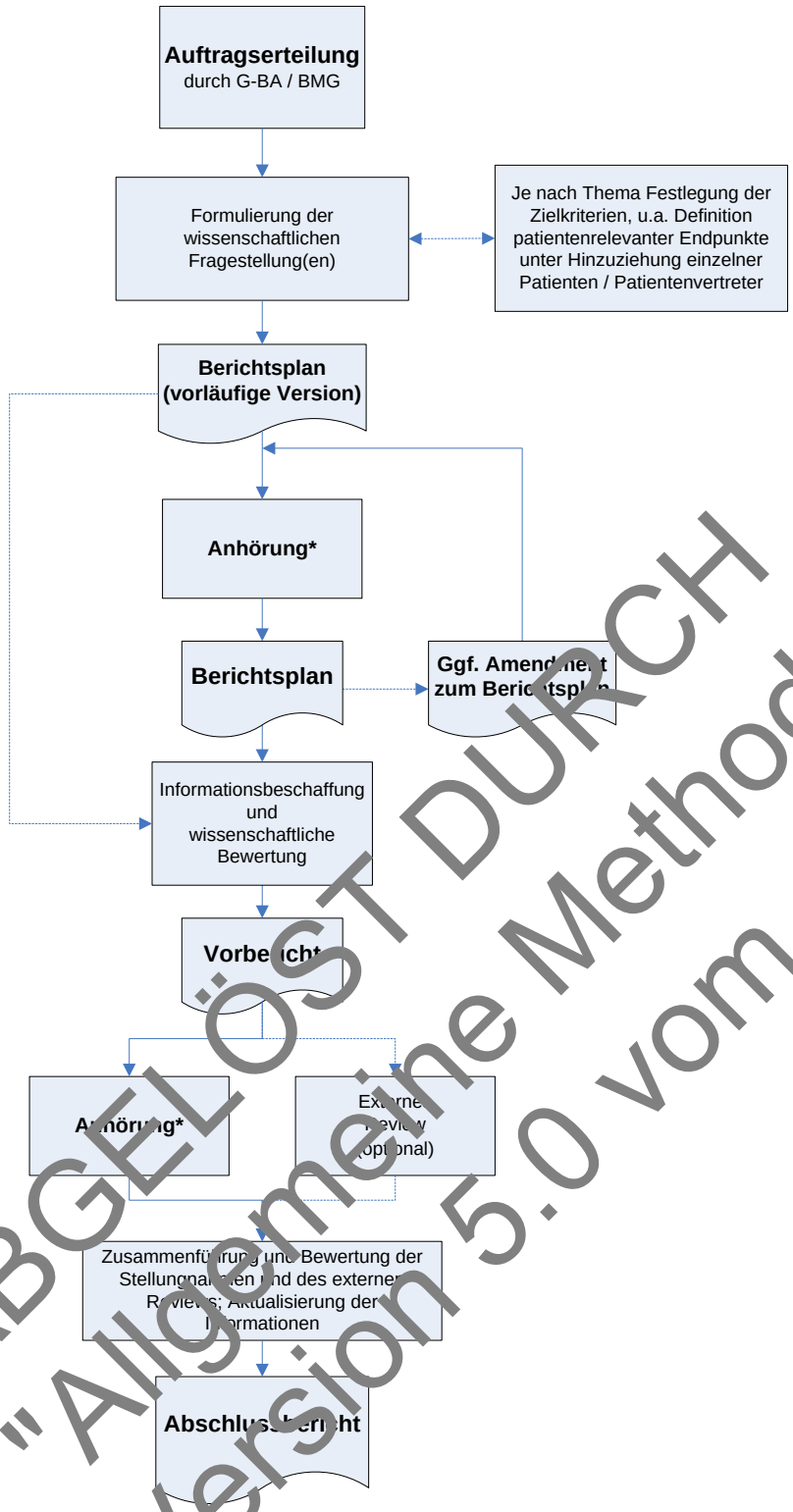
Beirat des Instituts einbezogen. Der interne Qualitätssicherungsprozess ist in diesem Flussdiagramm nicht dargestellt.

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. 5 Arbeitstage später.

Für eine Frist von mindestens 4 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"



ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

* 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. 5 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. 5 Arbeitstage nach Versendung an den Auftraggeber.

Für eine Frist von mindestens 4 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. 4 Wochen später) dem Kuratorium der Stiftung zugeleitet. In der Regel weitere 4 Wochen später erfolgt die Veröffentlichung dieser Dokumente auf der Website des Instituts. Sollten Stellungnahmen zu Abschlussberichten eingehen, die substanzielle nicht berücksichtigte Evidenz enthalten, oder erlangt das Institut auf andere Weise Kenntnis von solcher Evidenz, wird dem Auftraggeber begründet mitgeteilt, ob eine Neubeauftragung zu dem Thema (ggf. Aktualisierung des Berichts) aus Sicht des Instituts erforderlich erscheint. Der Auftraggeber entscheidet über die

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 der Trägerorganisationen 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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.

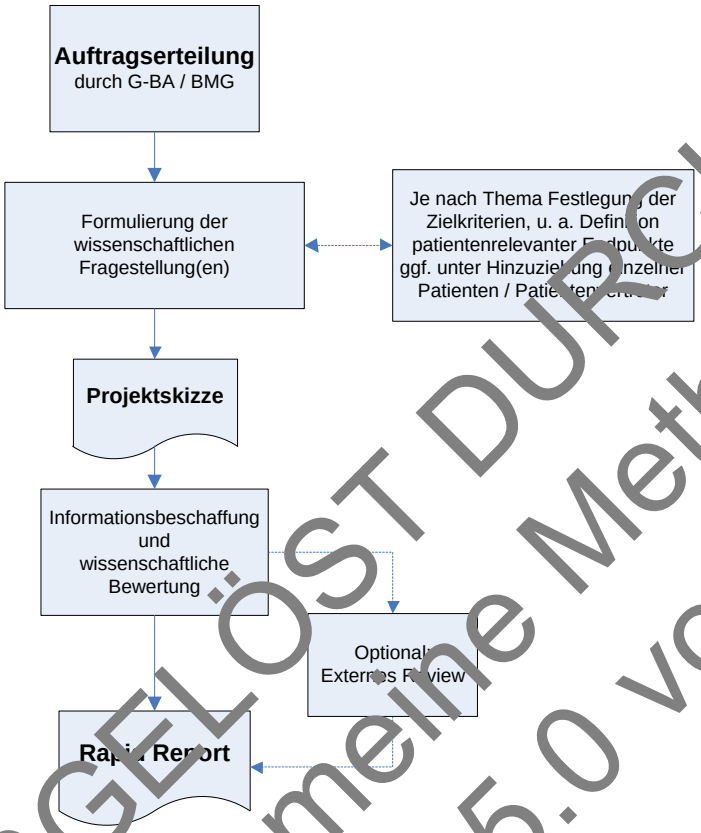


Abbildung 2: Ablauf der Erstellung eines Rapid Reports

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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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. 4 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

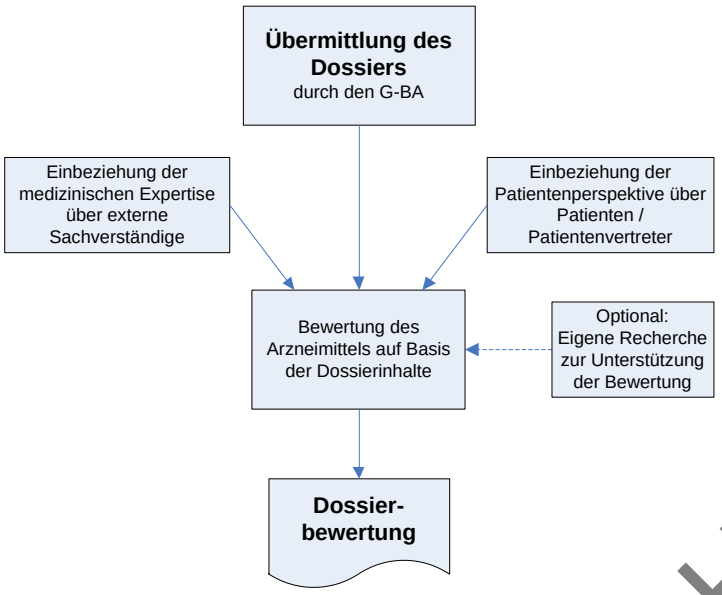


Abbildung 3: Ablauf der Erstellung einer Dossierbewertung

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 3 Monate nach dem für die Einreichung des Dossiers

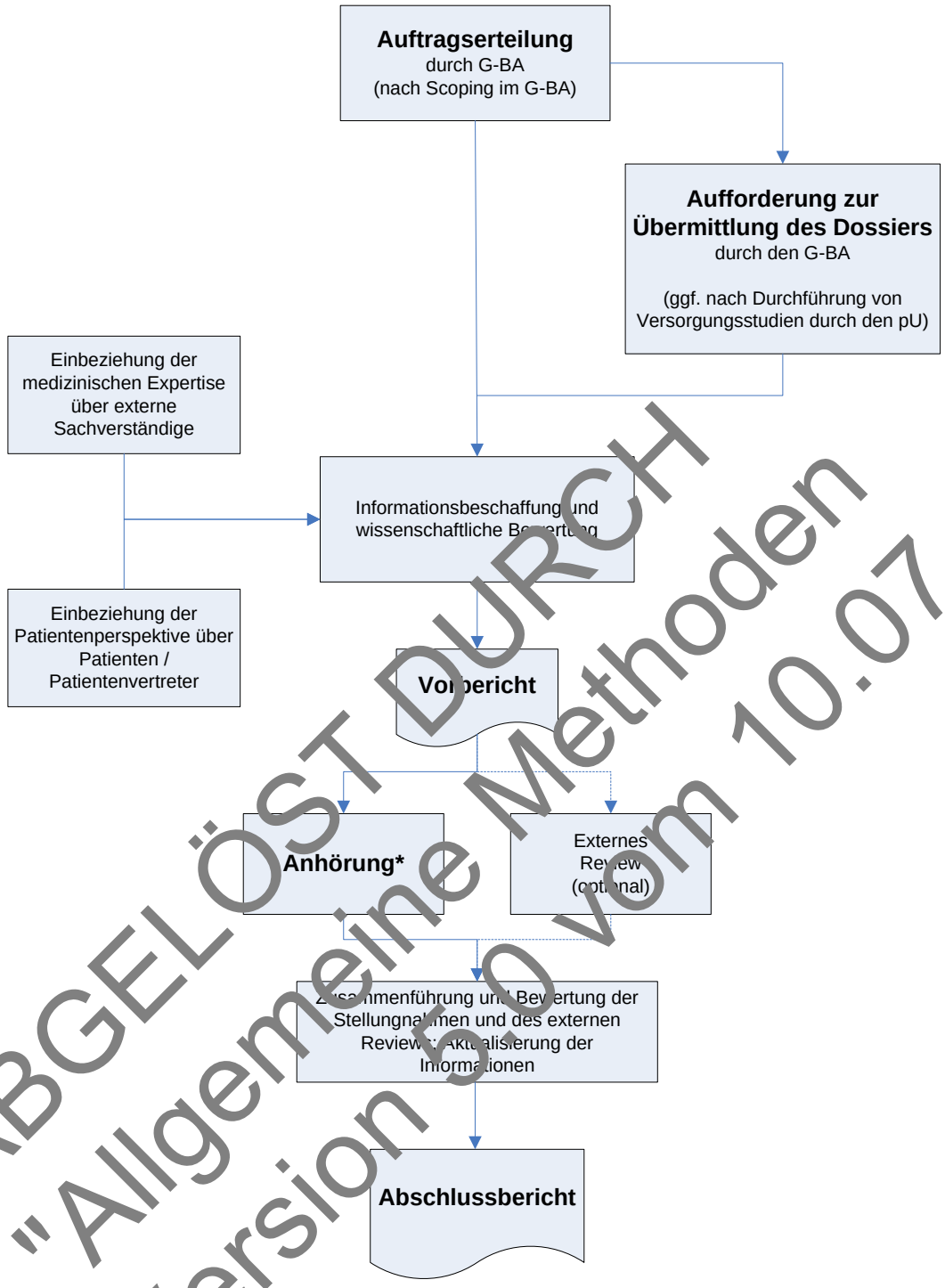
ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 Kosten-Nutzen-Bewertung nach § 35b SGB V

Der Ablauf der Kosten-Nutzen-Bewertung nach § 35b SGB V ist in Abbildung 4 schematisch dargestellt. Alle Arbeitsschritte werden in Verantwortung des Instituts getätigt. Dabei werden regelhaft externe Sachverständige beteiligt. Bei Bedarf wird auch der wissenschaftliche Beirat des Instituts einbezogen. Der interne Qualitätssicherungsprozess ist in diesem Flussdiagramm nicht dargestellt.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"



ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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

Abbildung 4: Ablauf der Kosten-Nutzen-Bewertung nach § 35b SGB V

Vor der **Auftragserteilung** durch den G-BA bereitet der G-BA wesentliche Auftragsinhalte vor (im Zuge des „Scopings“; siehe Abschnitt 4.9.1) und gibt den Stellungnahmeberechtigten die Gelegenheit zur Stellungnahme. Zeitgleich zur Auftragserteilung teilt der G-BA in seinem Beschluss mit, ob Versorgungsstudien zu berücksichtigen sind, die der G-BA mit dem pharmazeutischen Unternehmer vereinbart hat.

Parallel dazu fordert der G-BA den pharmazeutischen Unternehmer zur **Einreichung des Dossiers** auf. Das Dossier des pharmazeutischen Unternehmers fließt in diese Bewertung ein.

Im **Vorbericht** werden die Ergebnisse der Informationsbeschaffung und der wissenschaftlichen Bewertung dargestellt. Bei der inhaltlichen Bewertung werden regelhaft die medizinische Expertise über externe Sachverständige und die Patientenperspektive über Patientinnen und Patienten bzw. Patientenorganisationen einbezogen.

Die **medizinische Expertise** wird primär auf Basis eines Fragenkatalogs eingeholt, der zu Beginn der Bewertung an die externen Sachverständigen verschickt wird. Bei der Bewertung werden die Rückmeldungen der externen Sachverständigen berücksichtigt. 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 **Patientenperspektive** wird auf Basis eines Fragebogens erhoben, der zu Beginn der Bewertung an Patientinnen und Patienten bzw. Patientenorganisationen verschickt wird. Bei der Bewertung werden die in diesem Fragebogen getroffenen Angaben z. B. zu relevanten Endpunkten und zu wichtigen Subgruppen berücksichtigt. Patientinnen und Patienten bzw. Patientenorganisationen werden über die in § 140f SGB V benannten maßgeblichen Organisationen identifiziert.

Der **Vorbericht** enthält die vorläufige Empfehlung an den G-BA. Er wird nach Fertigstellung zunächst dem G-BA, dem Vorstand der Stiftung, dem Stiftungsrat und dem Kuratorium zugeleitet. Er wird auf der Webseite des Instituts zeitnah nach Versendung an den G-BA veröffentlicht. Für eine Frist von 3 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 und ist dem G-BA innerhalb von 3 Monaten nach Einleitung des Stellungnahmeverfahrens zum Vor-

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

bericht zu übermitteln (siehe Verfahrensordnung des G-BA 5. Kapitel § 31 [211]). Der Abschlussbericht und die Dokumentation der Anhörung zum Vorbericht werden zunächst dem G-BA, dem Vorstand der Stiftung sowie dem Stiftungsrat und anschließend dem Kuratorium der Stiftung zugeleitet. Die Dokumente werden danach auf der Website des Instituts veröffentlicht. Sollten Stellungnahmen zu Abschlussberichten eingehen, die substanzielle nicht berücksichtigte Evidenz enthalten, oder erlangt das Institut auf andere Weise Kenntnis von solcher Evidenz, wird dem G-BA begründet mitgeteilt, ob eine Neubeauftragung zu dem Thema (ggf. Aktualisierung des Berichts) erforderlich erscheint. Der G-BA entscheidet über die Beauftragung des Instituts. Eine solche Aktualisierung unterliegt den allgemeinen methodischen und verfahrenstechnischen Anforderungen an Produkte des Instituts.

2.1.5 Potenzialbewertung

Der Ablauf der Erstellung einer Potenzialbewertung ist in Abbildung 5 schematisch dargestellt. Alle Arbeitsschritte werden in Verantwortung des Instituts getätigt. Dabei können externe Sachverständige beteiligt werden (siehe Abschnitt 2.2.1). Der interne Qualitätssicherungsprozess ist in diesem Flussdiagramm nicht dargestellt.

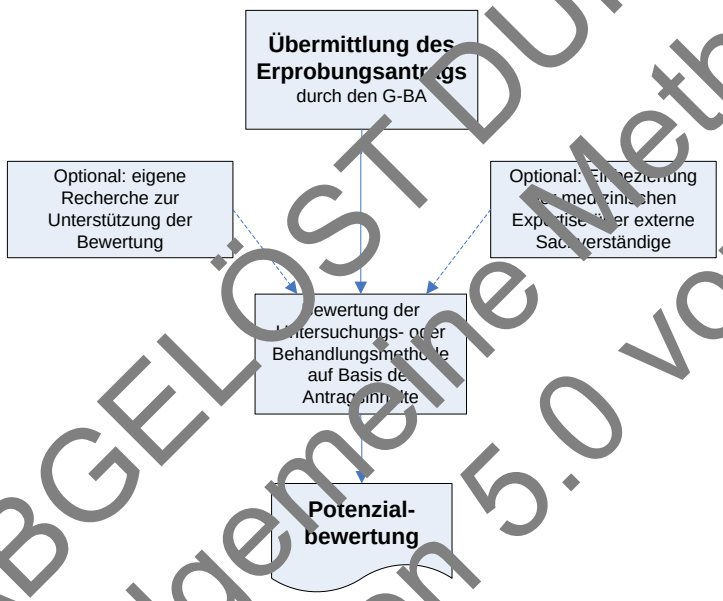


Abbildung 5: Ablauf der Erstellung einer Potenzialbewertung

Nach **Übermittlung des Erprobungsantrags** durch den G-BA erfolgt die inhaltliche Bewertung der Antragsinhalte in Verantwortung des Instituts. Hierbei kann externe medizinische Expertise miteinbezogen werden. Dies erfolgt in gleicher Weise wie bei Dossierbewertungen allerdings unter Berücksichtigung der spezifischen Anforderungen an die Wahrung der hohen Vertraulichkeit im Rahmen von Potenzialbewertungen.

Grundlage für die Bewertung ist der vom Antragsteller an den G-BA und dann weiter an das Institut übermittelte Antrag. Optional kann zur Unterstützung der Bewertung eine **eigene**

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Recherche des Instituts durchgeführt werden. Da die Eckpunkte einer Erprobungsstudie ein optionaler Antragsinhalt sind, konkretisiert das Institut diese Eckpunkte, wenn der Antragsteller hierzu keine Angaben macht.

Den Abschluss des Prozesses bildet die Erstellung der **Potenzialbewertung**. Nach § 137e SGB V muss der G-BA innerhalb von 3 Monaten eine Entscheidung zum Potenzial der beantragten Untersuchungs- oder Behandlungsmethode treffen. Daher werden Potenzialbewertungen durch das Institut regelhaft binnen 6 Wochen fertiggestellt. Nach Fertigstellung wird die Potenzialbewertung dem G-BA zugeleitet. Eine Publikation der Potenzialbewertung erfolgt nicht, weil das Antragsverfahren gemäß § 137e SGB V der hohen Vertraulichkeit unterliegt. Erst wenn im weiteren Verlauf der G-BA eine Erprobungsrichtlinie erlässt, wird auch die Potenzialbewertung veröffentlicht.

2.1.6 Addendum

Der Ablauf der Erstellung eines Addendums ist in Abbildung 6 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.

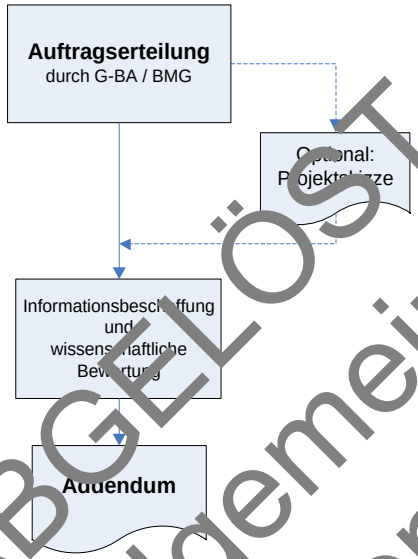


Abbildung 6: Ablauf der Erstellung eines Addendums

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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

liegenden Institutsprodukts beteiligt waren. Das Verfahren zur Veröffentlichung eines Addendums orientiert sich an dem des ursprünglichen Institutsprodukts. So wird beispielsweise ein Addendum zu Berichten zunächst an den Auftraggeber sowie an Stiftungsrat und Vorstand versandt. In der Regel eine Woche später wird es dem Kuratorium der Stiftung übermittelt und weitere 3 Wochen später auf der Website des Instituts veröffentlicht.

2.1.7 Gesundheitsinformationen

Das Institut erstellt **Gesundheitsinformationen** für Bürgerinnen und Bürger in verschiedenen Formaten, die in Abschnitt 6.4 ausführlicher vorgestellt werden.

Diese Informationen werden der Öffentlichkeit primär über die Website www.gesundheitsinformation.de zur Verfügung gestellt (auf Englisch unter informedhealthonline.org). Den Kern der Website bilden gesundheits- oder krankheitsbezogene Themen. Je nach Breite und Tiefe können innerhalb eines Themas unterschiedliche Textformate kombiniert werden.

Der Ablauf der Erstellung von Gesundheitsinformationen ist in Abbildung 7 schematisch dargestellt. In die Erstellung der Gesundheitsinformationen werden bei mehreren Schritten externe Personen einbezogen. Deren Aufgaben werden in Kapitel 6 genauer dargestellt.

Die vom Institut erstellten Gesundheitsinformationen gehen hervor aus

- Direktaufträgen des G-BA oder des BMG zur Erstellung von Patienteninformationen;
- allgemein verständlichen Zusammenfassungen (Begleitinformationen) der anderen Produkte des Instituts;
- 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 2.1.8) 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 der Themenwahl ist in Abschnitt 6.3.1 beschrieben. Nach Festlegung der innerhalb eines Themas zu behandelnden Aspekte erfolgt die **Informationsbeschaffung**, gefolgt von der **wissenschaftlichen Bewertung** der identifizierten Publikationen. Die Methodik der Informationsbeschaffung für Gesundheitsinformationen, der wissenschaftlichen Bewertung und darüber hinaus der Einbeziehung von Patientinnen und Patienten wird in Kapitel 6 beschrieben.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

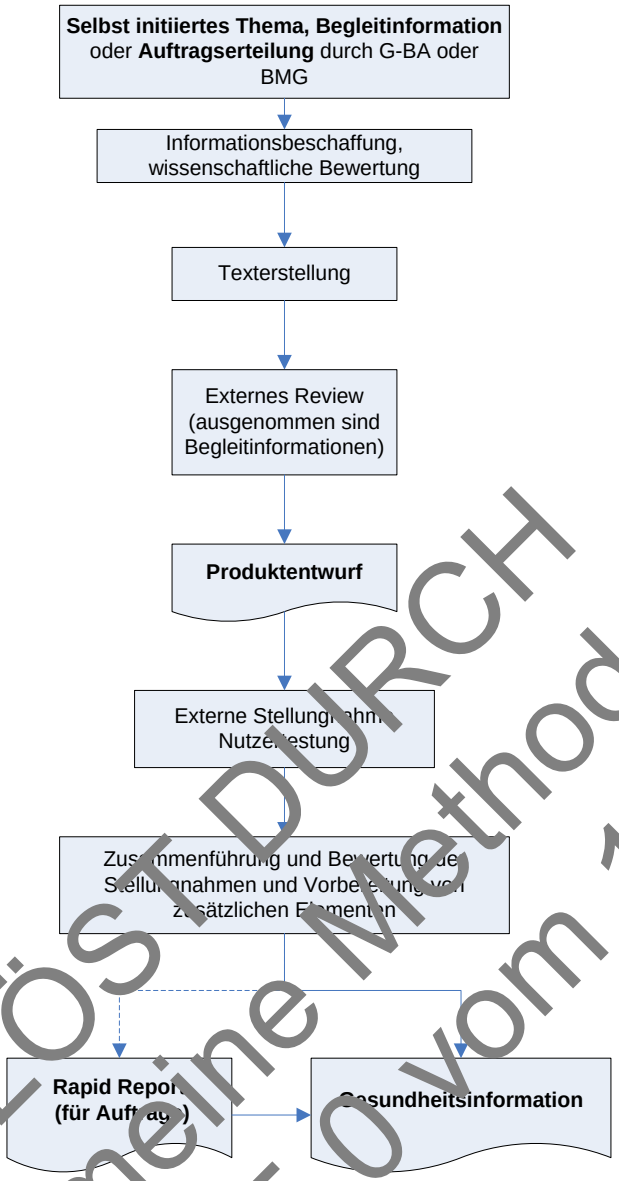


Abbildung 7: Ablauf der Gesundheitsinformationserstellung

Evidenzbasis einer Begleitinformation ist der zugrunde liegende IQWiG-Bericht. Eine darüber hinausgehende Informationsbeschaffung und wissenschaftliche Bewertung ist in diesem Fall optional, beispielsweise im Hinblick auf ergänzende Hintergrundinformationen oder Aspekte der qualitativen Forschung.

Im Anschluss an die **Texterstellung** und die ressortinterne Qualitätssicherung werden die Entwürfe zum **externen Review** verschickt. Begleitinformationen werden intern von der Projektleitung des jeweiligen Ressorts begutachtet.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Der aus Sicht des Instituts fertige Entwurf einer Gesundheitsinformation wird im Rahmen einer einmonatigen Beratungsperiode zur begrenzten **Stellungnahme** an den Auftraggeber, das Kuratorium und die weiteren Gremien des Instituts verschickt. Im Kuratorium sind Organisationen der Leistungserbringer und der Sozialpartner und Selbstverwaltungsorgane der Trägerorganisationen 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 an. Eine Gesundheitsinformation durchläuft außerdem vor der Veröffentlichung – in der Regel zeitgleich mit dem Stellungnahmeverfahren – eine externe Nutzertestung. Bei der Nutzertestung kommentiert eine Gruppe von Patientinnen und Patienten oder möglichen Nutzerinnen und Nutzern die Texte im Hinblick auf ihren Informationsgehalt und ihre Verständlichkeit.

Die während der Beratungsperiode eingegangenen Stellungnahmen und das Ergebnis der Nutzertestung werden geprüft, kommentiert und zusammengefasst. Sie können Anlass für eine Überarbeitung der vorgelegten Gesundheitsinformation sein.

Im Falle der Direktbeauftragung durch den G-BA oder das BMG wird die Gesundheitsinformation in Form eines Rapid Reports erarbeitet. Die Erstellung und Publikation unterliegt dem im IQWiG üblichen, in Abschnitt 2.1.2 dargestellten Verfahren. Der Rapid Report wird zunächst dem Auftraggeber, dem Vorstand der Stiftung sowie dem Stiftungsrat und anschließend (i. d. R. 4 Wochen später) dem Kuratorium der Stiftung zugeleitet. In der Regel weitere 4 Wochen später erfolgt die Veröffentlichung des Rapid Reports auf der Website (www.iqwig.de) des Instituts. Anschließend wird in der Regel die betreffende Gesundheitsinformation selbst auf www.gesundheitsinformation.de publiziert. Die Texte für allgemein verständliche Erläuterungen von G-BA-Richtlinien werden erst nach Veröffentlichung der Richtlinien selbst unter www.gesundheitsinformation.de veröffentlicht.

Korrekturen, Verbesserungen und Aktualisierungen der veröffentlichten Gesundheitsinformationen werden primär intern vorgenommen. Es können bei umfassenden oder grundlegenden inhaltlichen Änderungen externe Gutachterinnen und Gutachter eingebunden werden. Eine ausführlichere Beschreibung der Aktualisierungsmechanismen findet sich in Kapitel 6.

2.1.8 Arbeitspapier

Der Ablauf der Erstellung eines **Arbeitspapiers** ist in Abbildung 8 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

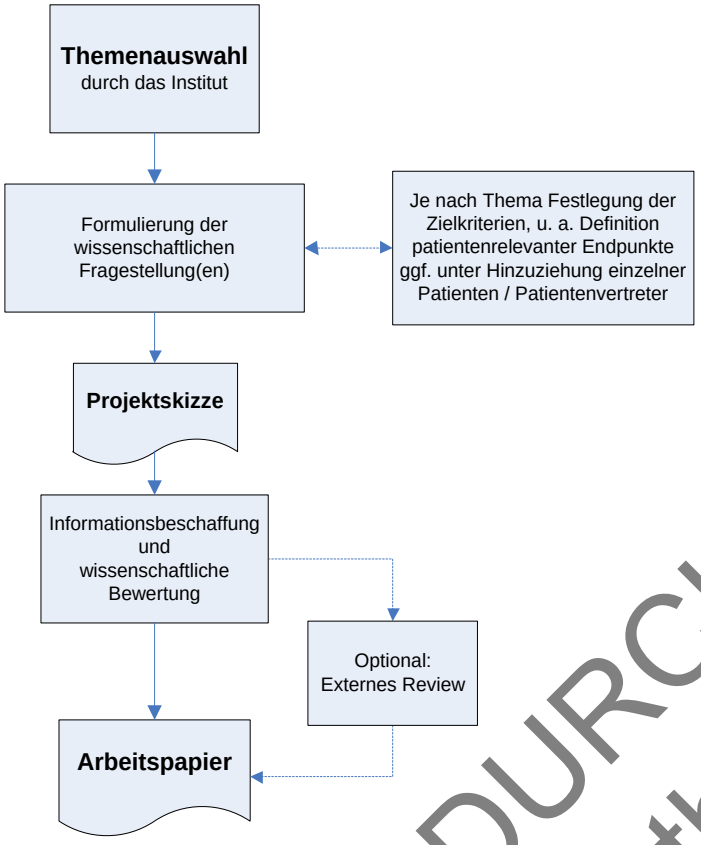


Abbildung 8: Ablauf der Erstellung eines Arbeitspapiers

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 Absatz 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 3 Wochen später auf der Website des Instituts veröffentlicht. Sollten Stellungnahmen zu Arbeitspapieren eingehen, die substanzielle nicht berücksichtigte Evidenz enthalten, oder erlangt das Institut auf andere Weise Kenntnis von solcher Evidenz, wird geprüft, ob eine Aktualisierung des Arbeitspapiers aus Sicht des Instituts erforderlich erscheint. 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 nach den allgemeinen Grundsätzen des Beschaffungswesens in einem transparenten und nicht diskriminierenden Wettbewerb.

Auftragsbekanntmachungen für Forschungsaufträge nach § 139b Absatz 3 SGB V sind auf der Website des Instituts veröffentlicht. Ausnahmen sind möglich, z. B. bei einem besonders eilbedürftigen Auftrag. Aufträge mit einem Auftragsvolumen oberhalb des aktuell geltenden Schwellenwerts der Vergabebestimmungen der Europäischen Union (EU) werden EU-weit

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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, Kosten-Nutzen-Bewertungen nach § 35b SGB V, Potenzialbewertung gemäß § 137e SGB V und Erstellung von Gesundheitsinformationen 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 Instituts 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 Instituts 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 [339,366] 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 6 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.

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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 Instituts 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 Absatz 4 SGB V).

Zur Wahrung der Unabhängigkeit des Instituts 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 (mit Ausnahme von Potenzialbewertungen, siehe hierzu § 19 der Verfahrensordnung des G-BA [211]), 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 [595]. 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

patientenrelevanter Zielgrößen zur Feststellung krankheits- und behandlungsbedingter Veränderungen erlauben:

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

Diese Zielgrößen sind auch im SGB V als vorrangig zu berücksichtigende Zielgrößen genannt, z. B. in § 35 Abs. 1b SGB V. Ergänzend kann der interventions- und erkrankungsbezogene Aufwand mit der Behandlung berücksichtigt werden. Auch die Patientenzufriedenheit kann ergänzend einbezogen werden, sofern hierbei gesundheitsbezogene Aspekte abgebildet werden. Ein Nutzen oder Zusatznutzen kann sich allein auf Basis dieser beiden 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 [81]. 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 Absatz 1 Satz 4 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 sollen nur Instrumente eingesetzt werden, die für den Einsatz in klinischen Studien geeignet und entsprechend evaluiert sind [174]. Darüber hinaus können für die Nutzenbewertung valide Surrogatendpunkte berücksichtigt werden (siehe Abschnitt 3.1.2).

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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.

3.1.3 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 [15,194,444]. Die meisten Surrogatendpunkte sind jedoch in dieser Hinsicht nicht verlässlich und können bei der Nutzenbewertung irreführend sein [102,210,277]. Daher werden Surrogatendpunkte im Rahmen der Nutzenbewertung des Instituts 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 [28,586]. Die Notwendigkeit

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 Surrogatendpunkts gibt es weder ein Standardverfahren noch eine allgemein beste Schätzmethode noch ein allgemein akzeptiertes Kriterium, dessen Erfüllung den Nachweis der Validität bedeuten würde [380]. Allerdings widmet sich die aktuelle methodische Literatur sehr häufig korrelationsbasierten Verfahren zur Surrogatvalidierung mit Schätzung von Korrelationsmaßen auf Studien- und individueller Ebene [286]. In der Nutzenbewertung des Instituts 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 [86,400]. Alternative Methoden [586] werden nur in begründeten Ausnahmefällen in Betracht gezogen.

Bei korrelationsbasierten Verfahren wird zum 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 [86,88]. Da in der Nutzenbewertung des Instituts Aussagen bezogen auf Gruppen von Patientinnen und Patienten gemacht werden, stützt sich die Bewertung der Validität eines Surrogatendpunkts hier primär auf den Grad der Korrelation auf der Ebene der Behandlungseffekte. Zusätzlich zur Höhe der Korrelation wird für die Bewertung der Validität eines Surrogatendpunkts die Aussagesicherheit der Ergebnisse zur Validierung berücksichtigt. Dazu werden verschiedene Kriterien herangezogen [286]. 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 [193,219,227,380]. 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) [85,286]. Hierzu wird der Effekt auf das Surrogat, der sich aus den Studien zur Nutzenbewertung ergibt, dem STE gegenübergestellt [88,400].

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Die Schlussfolgerungen für eine Nutzenbewertung des Instituts 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 Instituts 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 [109].

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 [602]. 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 [109]. 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 [41,304]. 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 [48,164,303]. 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 [302]. 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 [127].

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 2 oder mehr ansonsten gleichwertigen Therapieoptionen erheblich unterscheiden könnten,
- insbesondere bei derjenigen 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 3 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 8.3.5).

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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Fehler (Verzerrung, siehe Abschnitt 8.3.11), 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 8.1.3), 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 8.1.4).

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 3 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 8.3.8) sinnvollerweise 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Falls das Prädiktionsintervall zur Darstellung der Heterogenität in einer Meta-Analyse mit zufälligen Effekten (siehe Abschnitt 8.3.8) 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 2 oder mehr Studien zeigen in eine Richtung. Für diese „gerichteten" Studien gelten alle folgenden Bedingungen:

- Das Gesamtgewicht dieser Studien ist ≥ 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 8.3.8). 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 8.3.8), 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:
  - Alle Studien weisen statistisch signifikante Ergebnisse derselben Effektrichtung auf: Die gleichgerichteten Effekte sind deutlich gleichgerichtet.
  - 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 signifikanten 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 8.3.9). 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).

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 Ergebnissicherheit	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 %).

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

- 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 Nutzenaussage 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 bezüglich 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 lebenslagenspezifische 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 wie in der frühen Nutzenbewertung nach § 35a SGB V) 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 oder über ihre Gewichtung zu einer Gesamtaussage zu kommen. Die gleichzeitige Würdigung von Nutzen und Schaden wird themenspezifisch konkretisiert (siehe auch Abschnitt 4.3.3).

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 [179,295] (siehe auch Abschnitt 8.3.11).

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Die Verzerrung der publizierten Evidenz durch Publikationsbias und Outcome Reporting Bias wurde umfangreich beschrieben [160,390,522]. 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 7.1.5).

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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

War die Datenübermittlung unvollständig, hängt die Konsequenz für das Fazit davon ab, ob durch andere Rechercheschritte 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 Rechercheschritte 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. [214] 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 [214]. Diese Größenordnung dient dem Institut als Orientierung und bedeutet keine starre Grenze. Glasziou et al. [214] haben ihrer Empfehlung

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 [342].

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. [176]). 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 [282]. 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 können auch andere Nutzendimensionen mittels PRO erfasst werden, wie z. B. Symptome der Erkrankung. Wie auch für die Erfassung der Lebensqualität sind hierfür Instrumente zu fordern, die für den Einsatz in klinischen Studien geeignet sind [174]. 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 [198]. D. h., dass auch im Falle von PROs einschließlich gesundheitsbezogener Lebensqualität, Symptomen und der 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 [600]. Dies ist bei der Interpretation solcher Studien zu beachten (siehe auch Abschnitte 8.1.4 und 8.3.4). 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 [198].

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 [598]. Davon unberührt, haben auch Patientinnen und Patienten mit seltenen Erkrankungen einen Anspruch auf eine möglichst sichere Information über Behandlungsoptionen [171]. 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 [173], 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 [175].

Für den Fall extrem seltener Erkrankungen oder extrem spezifischer Krankheitskonstellationen kann die Forderung nach (parallel) vergleichenden Studien unangemessen sein [598]. 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 [82]. 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 3 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 [226]. 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 [422].

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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 [358] (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 [332]).

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 8.1.4), 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 8.1.

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) 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.

3.2.3 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,

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 Dossiervorlagen beschrieben, die Bestandteil der Verfahrensordnung des G-BA sind [211]. 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) [80].

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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 wieder um 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 (Überlebenszeitenanalysen),
- 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 3 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 unterscheidenden Schwellenwerte für jede der 3 Kategorien der Qualität der Zielgrößen wieder.

Tabelle 4: Schwellenwerte zur Feststellung des Ausmaßes eines Effekts

		Zielgrößenkategorie		
		Gesamtmortalität	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 ≥ 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 2 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 [553]. 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 [116], um die entsprechenden Schwellenwerte zur Ausmaßfeststellung anzulegen. Andernfalls 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 [211] 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 aktuellen DIN EN ISO-Norm 14155 (Abschnitt A.6.1 [138]) 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 [389]. 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 [389]). 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 für 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 [211], 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 potenzielle Nutzen einer

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Methode insbesondere gegen die Risiken der Anwendung beim Patienten abzuwägen, die mit einem Wirksamkeitsnachweis geringerer Aussagekraft einhergehen“ [211]. 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

Diagnostische Verfahren zeichnen sich dadurch aus, dass ihr gesundheitsbezogener Nutzen (oder Schaden) im Wesentlichen erst dadurch zustande kommt, dass sich ihnen therapeutische oder präventive Verfahren anschließen. Die alleinige Gewinnung diagnostischer Informationen (ohne medizinische Konsequenzen) hat regelhaft keinen sozialrechtlich relevanten Nutzen.

Dies gilt in gleicher Weise sowohl für diagnostische Informationen, die sich auf den augenblicklichen Gesundheitszustand beziehen, wie auch für prognostische Informationen (oder Marker), die sich auf einen zukünftigen Gesundheitszustand beziehen. Im Folgenden werden daher Verfahren zur Bestimmung diagnostischer oder prognostischer Informationen gemeinsam als diagnostische Verfahren betrachtet.

Allgemein kann der Prozess zur Evaluierung diagnostischer Verfahren in Analogie zur Arzneimittelprüfung in verschiedene, hierarchisch angeordnete Phasen oder Stufen eingeteilt werden [204,329]. Prospektiv geplante, kontrollierte Diagnosestudien der Phase 4 nach Köbberling et al. [329] bzw. Studien der Stufe 5 nach Fryback und Thornbury [204] 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 [211]). 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 [197], und Beispiele zeigen, dass sie auch mit vergleichsweise moderatem Aufwand durchführbar sind [16,568].

Das Institut folgt dieser Logik und führt Nutzenbewertungen zu diagnostischen Verfahren in erster Linie auf der Grundlage solcher Studien mit patientenrelevanten Endpunkten durch. Die

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 Erhebung diagnostischer oder prognostischer Informationen 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.

Aussagen zum Nutzen diagnostischer Verfahren stützen sich im Idealfall auf randomisierte Studien, die in verschiedenster Form durchgeführt werden können [50,51,188,360,378,484]. In einer Studie mit Strategiedesign kommen bei 2 (oder mehr) Patientengruppen jeweils unterschiedliche Strategien zur Anwendung, die sich jeweils aus diagnostischer Maßnahme und therapeutischer Konsequenz zusammensetzen. Ebenfalls von hoher Aussagekraft sind randomisierte Studien, in denen bei allen Patienten zunächst die konventionelle und die zu prüfende diagnostische Maßnahme durchgeführt werden und anschließend lediglich diejenigen Patienten randomisiert werden, bei denen der zu prüfende Test ein anderes Ergebnis und damit eine andere therapeutische Konsequenz als der konventionelle Test ergeben hat (Diskordanzdesign). Ebenfalls von hoher Evidenz und priorität für die Nutzenbewertung diagnostischer Verfahren heranzuziehen sind solche Studien, bei denen die Wechselwirkung zwischen diagnostischer oder prognostischer Information und therapeutischem Nutzen geprüft wird (Interaktionsdesign [484,541]). Viele diagnostische oder prognostische Merkmale – insbesondere genetische Marker – können in prospektiv vergleichenden Studien auch retrospektiv bestimmt und auf eine mögliche Wechselwirkung hin geprüft werden (sogenanntes „prospektiv-retrospektives“ Design [516]). Die Validität solcher „prospektiv-retrospektiven“ Designs hängt insbesondere davon ab, dass eine prospektive Planung der Analysen (insbesondere auch der Grenzwertfestlegung) vorhanden ist. Bei allen Studien im Interaktionsdesign ist es darüber hinaus wichtig, dass die eingesetzten Therapien dem aktuellen Standard entsprechen, dass die Informationen (z. B. Gewebeproben) zum interessierenden Merkmal für alle Studienteilnehmer, oder zumindest für eine repräsentative Stichprobe vollständig vorliegen und dass bei Analyse mehrerer Merkmale das Problem einer multiplen Signifikanztestung adäquat berücksichtigt wird (siehe hierzu auch Abschnitt 8.3.2) [485].

Insgesamt entscheidend ist weniger, inwieweit eine diagnostische oder prognostische Information einen aktuellen oder zukünftigen Gesundheitszustand feststellen kann, sondern dass diese Information auch prädiktive Bedeutung hat, also den höheren (oder geringeren) Nutzen einer Folgebehandlung vorhersagen kann [188,517]. Wichtig bei dieser notwendigerweise verknüpften Betrachtungsweise von diagnostischer und therapeutischer Intervention ist, dass ein Nutzen insgesamt sich in der Regel nur dann ergeben kann, wenn beide Interventionen ihr

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Ziel erfüllen: Sowohl bei unzureichender prädiktiver Diskriminationsfähigkeit der diagnostischen Intervention als auch bei fehlender Effektivität der therapeutischen Intervention wird eine Studie keinen Nutzen der diagnostischen Intervention zeigen können.

Neben Strategie- und Interaktionsdesign existiert mit dem Anreicherungsdesign eine dritte wesentliche Form von RCTs zu diagnostischen Fragestellungen [379,541]. Bei diesem Design erfolgt allein auf der Basis des zu prüfenden diagnostischen Tests eine Randomisierung (und damit ein Einschluss) lediglich eines Teils der Patienten, beispielsweise der Testpositiven, welche dann 1 von 2 Therapieformen erhalten. Einem solchen Design fehlt im Vergleich zum Interaktionsdesign die Prüfung eines möglichen Therapieeffektes beim Rest der Patienten, also beispielsweise den Testnegativen. Daher lassen sich aus solchen Designs nur dann tragfähige Schlüsse ziehen, wenn aufgrund anderer Informationen ausgeschlossen werden kann, dass ein in der randomisierten Patientengruppe beobachteter Effekt nicht auch in der nicht randomisierten Gruppe vorhanden gewesen wäre.

Die bisherigen Ausführungen beziehen sich primär auf diagnostische Verfahren, die durch eine Erhöhung der Testgüte (also Sensitivität, Spezifität oder beide) mehr Patienten einer bestimmten therapeutischen Konsequenz zuführen. In diesen Fällen ist es regelhaft notwendig, durch eine Erfassung der gesamten diagnostisch-therapeutischen Behandlungskette den Einfluss des diagnostischen Verfahrens auf patientenrelevante Endpunkte zu untersuchen. Es kann jedoch vorkommen, dass das zu prüfende diagnostische Verfahren ein anderes bereits etabliertes diagnostisches Verfahren lediglich ersetzen soll, ohne dass der neue Test zusätzliche Patienten identifiziert oder ausschließt. Wenn der neue Test direkte patientenrelevante Vorteile aufweist, wie zum Beispiel geringere Invasivität oder Strahlungsfreiheit, wird es nicht immer notwendig sein, erneut die gesamte diagnostisch-therapeutische Kette zu untersuchen, weil die therapeutischen Konsequenzen, die sich aus dem neuen Test ergeben, sich nicht von denen des bisherigen Tests unterscheiden [42,51,394]. Ausreichend für einen Nutzennachweis können in diesen Fällen Testgütestudien sein, in denen gezeigt wird, dass das Testergebnis des bisherigen Tests (= Referenzstandard) und das Testergebnis des zu prüfenden Tests (= Indextest) in einem hinreichend hohen Anteil der Patienten identisch ist (einseitige Äquivalenzfragestellung).

Für einen Vergleich zweier oder mehrerer diagnostischer Tests hinsichtlich bestimmter Testgüteeigenschaften 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.

Wenn eine Studie aussagekräftige Daten zum Nutzen, zur diagnostischen Güte oder zur prognostischen Wertigkeit liefern soll, ist es wesentlich, dass hierbei ein Vergleich zur bisherigen diagnostischen Vorgehensweise erfolgt [542]. Nur auf diese Weise lässt sich der Mehrwert der diagnostischen oder prognostischen Information sicher bestimmen. Bei Studien

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

zur Testgüte bedeutet dies, dass neben Sensitivität und Spezifität der neuen und der bisherigen Methode von besonderem Interesse ist, inwieweit die diagnostischen Maßnahmen pro Patient unterschiedliche Befunde ergeben. Bei Studien zu prognostischen Markern kommt dagegen häufig multifaktoriellen Regressionsmodellen eine zentrale Rolle zu, sodass Abschnitt 8.3.7 zu beachten ist. Bei der Auswahl nicht randomisierter Studiendesigns zu diagnostischen Methoden findet die in Abschnitt 8.1.3 dargestellte Rangordnung verschiedener Studiendesigns regelhaft Anwendung.

Bei der Bewertung der Ergebnissicherheit von Studien zur diagnostischen Güte orientiert sich das Institut in erster Linie an den QUADAS-2-Kriterien [592,593], die allerdings ggf. projektspezifisch angepasst werden. Die Kriterien des STARD-Statements [52,53] werden herangezogen, um im Einzelfall über den Ein- oder Ausschluss nicht im Volltext publizierter Studien zu entscheiden (siehe hierzu auch die Abschnitte 8.1.4 und 8.3.11). Für die methodische Bewertung von Prognosestudien gibt es trotz einzelner guter Vorschläge keine allgemein akzeptierten Qualitätskriterien [11,251,252,515]. Publikationsstandards für Studien zu prognostischen Markern existieren nur in allgemeiner Form [579], es gibt jedoch Publikationsstandards zu prognostischen Markern in der Onkologie [14,393].

Studien der Stufe 3 und 4 nach Fryback und Thornbury [204] 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 [204,239]. Der Stellenwert derartiger, letztlich unkontrollierter Studien im Rahmen von Nutzenbewertungen diagnostischer (oder auch prognostischer) Verfahren muss als weitgehend unklar angesehen werden. Informationen zu Managementänderungen allein können deshalb nicht für einen Nutzennachweis herangezogen werden, solange keine Informationen über die patientenrelevanten Auswirkungen solcher Änderungen vorliegen.

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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 – wie oben bereits ausgeführt – vergleichende Studien zu den diagnostischen Strategien mit und ohne neues Verfahren erforderlich [197,367].

Mehrere diagnostische Einzeltests / -informationen werden zum Teil über Algorithmen, Scores oder Ähnliches zu einem Gesamttest zusammengefasst. Bei der Bewertung solcher kombinierten Tests sind die gleichen Prinzipien wie bei Einzeltests anzuwenden. Insbesondere müssen die Validierung und die klinische Evaluation eines jeden neuen Tests unabhängig erfolgen von der Testentwicklung (beispielsweise Grenzwertfestlegung, Score-Gewichtung oder Auswertungsalgorithmus) [531].

Auch Biomarker, die im Rahmen einer „personalisierten“ oder besser stratifizierten Medizin eingesetzt werden, sind mit den hier charakterisierten Methoden zu evaluieren [268,541]. Dies gilt sowohl für Biomarker, die vor der Entscheidung über den Beginn einer (oder einer anderen) Therapie bestimmt werden, als auch für solche, die während einer Therapie bestimmt werden, um über Fortführung, Abbruch, Wechsel oder Anpassung der Therapie zu entscheiden [520,567]. Auch hier ist die Unterscheidung zwischen der prognostischen und der prädiktiven Wertigkeit eines Merkmals wesentlich. Prognostische Marker geben Informationen über einen zukünftigen Gesundheitszustand und beziehen sich hierbei in aller Regel auf den Verlauf unter Behandlung und nicht auf den natürlichen Verlauf der Erkrankung ohne Behandlung. Dass ein Biomarker prognostische Bedeutung hat, bedeutet nicht, dass er auch prädiktive Bedeutung hat. Gleiches gilt auch umgekehrt nicht.

Schließlich kann es auch bei der Bewertung von Diagnostika 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.

2.6 Früherkennung und Screening

Screeningprogramme setzen sich aus unterschiedlichen „Bausteinen“ zusammen, die entweder als Ganzes oder in Teilen untersucht werden (können) [120,513]. Bei der Bewertung erfolgt eine Orientierung an international akzeptierten Standards bzw. Kriterien, wie zum Beispiel denjenigen des UK National Screening Committee (UK NSC [564]), der US Preventive Services Task Force (US PSTF [247,437,490]) oder des National Health Committee (NHC) aus Neuseeland [406].

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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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. [329]), 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 [581]. 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 [254]. 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 [554].

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

In Einzelfällen ist zu prüfen, inwieweit die Berücksichtigung weiterer Studiendesigns sinnvoll sein kann [308]. So werden beispielsweise massenmediale Kampagnen häufig im Rahmen eines sogenannten Interrupted Time Series Designs evaluiert (z. B. bei [572]) 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 [106].

Für den Nutzen auf Populationsebene ist neben der Effektivität eines solchen Programms auch die Höhe der Teilnahmerate entscheidend. Außerdem ist die Frage relevant, welche Personen erreicht werden; so gibt es Hinweise darauf, dass Präventionsprogramme gerade von Bevölkerungsgruppen mit erhöhtem Krankheitsrisiko weniger in Anspruch genommen werden [343]. Beide Aspekte stehen deshalb im besonderen Fokus von Bewertungen des Instituts.

3.8 Potenzialbewertung

Potenzialbewertungen zielen im Gegensatz zu Nutzenbewertungen darauf ab zu prüfen, ob neue Untersuchungs- oder Behandlungsmethoden möglicherweise einen Nutzen aufweisen. Potenzial bedeutet hierbei, dass erstens die bisher vorliegenden Erkenntnisse einen möglichen Nutzen erkennen lassen und zweitens auf der Basis dieser Erkenntnisse eine Studie geplant werden kann, die eine Bewertung des Nutzens der Methode auf einem ausreichend sicheren Erkenntnisniveau erlaubt (siehe § 14 Abs. 3 und 4 der Verfahrensordnung des G-BA [211]).

Eine Potenzialbewertung gemäß § 137e Abs. 7 SGB V beruht auf einem Antrag, für den der G-BA Form und notwendige Inhalte definiert hat. Antragsberechtigt sind die Hersteller eines Medizinprodukts, auf dessen Einsatz die technische Anwendung einer neuen Untersuchungs- oder Behandlungsmethode maßgeblich beruht, und Unternehmen, die in sonstiger Weise als Anbieter einer neuen Methode ein wirtschaftliches Interesse an einer Erbringung zulasten der Krankenkassen haben. Im Antrag sind aussagekräftige Unterlagen vorzulegen, die sich insbesondere auf den aktuellen Erkenntnisstand und den zu erwartenden Nutzen der neuen Untersuchungs- oder Behandlungsmethode beziehen (siehe § 20 Abs. 2 Nr. 5 der Verfahrensordnung des G-BA [211]). Optional kann auch ein Vorschlag zu den Eckpunkten einer Erprobungsstudie eingereicht werden. Ein Antrag zu einer Methode kann sich auf eine oder mehrere Indikationen beziehen.

Im Rahmen der Potenzialbewertung wird vom Institut die Plausibilität der Angaben des Antragstellers geprüft. Diese Prüfung bezieht sich insbesondere auf die Sinnhaftigkeit der im Antrag dargestellten medizinischen Fragestellung(en), die Güte der vom Antragsteller durchgeführten Literatursuchen (siehe Abschnitt 7.2), die Einschätzung der Ergebnissicherheit der relevanten Studien und die Korrektheit der im Antrag dargestellten Ergebnisse. Die Bewertung mündet in einer Aussage zum Potenzial der beantragten Untersuchungs- oder Behandlungsmethode. Wenn aus Sicht des Instituts ein Potenzial feststellbar ist, wird die vom Antragsteller vorgeschlagene Erprobungsstudie geprüft; sofern der Antrag keinen solchen

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

oder einen ungeeigneten Vorschlag enthält, konkretisiert das Institut die Eckpunkte einer möglichen Erprobungsstudie.

Aufgrund der besonderen Zielsetzung sind bei der Potenzialbewertung im Vergleich zur Nutzenbewertung deutlich niedrigere Anforderungen an die Evidenz zu stellen. Letztendlich besteht erst das Ziel der Erprobung darin, eine adäquate Datengrundlage für eine zukünftige Nutzenbewertung herzustellen. Dementsprechend lässt sich ein Potenzial insbesondere auch auf der Basis nicht randomisierter Studien begründen. Darüber hinaus finden auch weitere methodische Grundsätze der Nutzenbewertung bei der Potenzialbewertung keine oder nur eingeschränkte Anwendung, wie im Folgenden dargestellt wird.

Im Gegensatz zur Nutzenbewertung wird im Rahmen von Potenzialbewertungen aufgrund der niedrigeren Anforderungen an die Evidenz eine erweiterte Bewertung der qualitativen Ergebnissicherheit nicht randomisierter Studien vorgenommen. Hierbei werden neben den in Abschnitt 3.1.4 für randomisierte Studien genannten Stufen (hohe oder mäßige Ergebnissicherheit) folgende Grade verwendet:

- **geringe qualitative Ergebnissicherheit**: Ergebnis einer höherwertigen nicht randomisiert vergleichenden Studie (beispielsweise quasirandomisierte kontrollierte Studien, nicht randomisierte kontrollierte Studien mit aktiver Zuteilung der Intervention nach vorab geplanter Regel, prospektive vergleichende Kohortenstudien mit passiver Zuteilung der Intervention) mit adäquater Kontrolle für Confounder,
- **sehr geringe qualitative Ergebnissicherheit**: Ergebnis einer höherwertigen nicht randomisiert vergleichenden Studie (siehe vorheriger Punkt) jedoch ohne adäquate Kontrolle für Confounder oder Ergebnis einer sonstigen nicht randomisiert vergleichenden Studie (beispielsweise retrospektive vergleichende Kohortenstudien, historisch kontrollierte Studien, Fall-Kontroll-Studien),
- **minimale qualitative Ergebnissicherheit**: Ergebnis einer nicht vergleichenden Studie (beispielsweise einarmige Kohortenstudien, Verlaufsbeobachtungen oder Fallserien, Querschnittsstudien oder sonstige nicht vergleichende Studien).

Ein wesentlicher Aspekt der Ergebnissicherheit ist die Kontrolle für Confounder, die insbesondere durch multifaktorielle statistische Methoden – wie in Abschnitt 8.3.7 beschrieben – angestrebt werden kann. Auch weitere Faktoren werden bei der Einschätzung der Ergebnissicherheit berücksichtigt (siehe Abschnitt 8.1.4).

Da auch hochwertige, nicht randomisierte Studien ein erhebliches Verzerrungspotenzial bergen, muss bei der Ableitung eines Potenzials aus solchen Studien geprüft werden, ob die vorliegenden Studien bezüglich der interessierenden Intervention Unterschiede in einer Größenordnung zeigen, die erwarten lassen, dass durch geeignete zukünftige Studien ein Nutzen nachgewiesen werden kann, und nicht mehr allein durch einen durchschnittlich erwartbaren Einfluss von Verzerrung (Bias) erklärt werden können. Daher ergibt sich ein

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Potenzial insbesondere dann, wenn Studien geringer Ergebnissicherheit mindestens kleine Effekte zeigen, wenn Studien sehr geringer Ergebnissicherheit mindestens mittlere Effekte zeigen oder wenn Studien minimaler Ergebnissicherheit mindestens große Effekte zeigen. Als ungefähre Grenzen zwischen kleinen, mittleren und großen Effekten können für das relative Risiko Werte von 0,8 und 0,5 dienen [150,434]. Abweichend zum Vorgehen bei Nutzenbewertungen (siehe Abschnitt 3.1.2) werden bei der Potenzialbewertung auch Surrogatendpunkte berücksichtigt, für die noch keine hinreichende Validität gezeigt werden konnte. Surrogatendpunkte sollen jedoch etabliert und plausibel sein, um hierüber ein Potenzial begründen zu können.

Sofern das Potenzial diagnostischer Methoden zu bewerten ist, finden auch Daten zur Testgüte Berücksichtigung. Hierbei ist die Ergebnissicherheit der zugrunde liegenden Studien zu prüfen (siehe Abschnitte 3.5 und 8.3.11). Im zweiten Schritt erfolgt eine Plausibilitätsprüfung der diagnostischen Methode hinsichtlich der vom Antragsteller postulierten Effekte bezüglich patientenrelevanter Endpunkte, d. h. möglicher direkter Effekte der Methode sowie therapeutischer Konsequenzen, über die die diagnostische Methode patientenrelevante Zielgrößen beeinflussen könnte.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

4 Kosten-Nutzen-Bewertung medizinischer Interventionen

4.1 Einführung

Gemäß SGB V bestimmt das Institut auftragsbezogen über die Methoden und Kriterien für die Erarbeitung von Kosten-Nutzen-Bewertungen (KNB) auf der Grundlage der in den jeweiligen Fachkreisen anerkannten internationalen Standards der evidenzbasierten Medizin und der Gesundheitsökonomie. Für jede KNB müssen Entscheidungen zur Perspektive, zum Zeithorizont, zur Auswahl der Komparatoren, zum zugrunde liegenden Versorgungspfad, zum Modell, zur Datengrundlage und zur Darstellung der Unsicherheit gefällt werden. Vor dem Hintergrund der Beauftragung durch den G-BA werden diese grundlegenden Kriterien einer KNB kurz erläutert. Alle Abweichungen von den hier vorgelegten Methoden sind im Einzelfall zu begründen.

4.1.1 Rechtliche Grundlagen für die Kosten-Nutzen-Bewertung nach SGB V

Nach § 139a Absatz 3 Nr. 2 SGB V kann das Institut zu Fragen der Qualität und Wirtschaftlichkeit der im Rahmen der gesetzlichen Krankenversicherungen erbrachten Leistungen beauftragt werden. Kosten-Nutzen-Bewertungen (KNB) von Arzneimitteln können vom G-BA auch nach § 35b SGB V beauftragt werden. Weiterhin kann eine KNB nach § 139b Absatz 2 SGB V vom BMG beauftragt werden.

[79] Im Folgenden werden zunächst methodische Aspekte berührt, die für KNB generell gelten. In Abschnitt 4.9 werden dann Abweichungen erläutert, die sich bei KNB von Arzneimitteln nach § 35b SGB V ergeben.

4.1.2 Perspektive

Auftragsabhängig können die (reine) GKV-Perspektive, die Perspektive der GKV-Versichertengemeinschaft, die Sozialversicherungsperspektive bzw. die Perspektive einzelner Sozialversicherungsträger sowie die gesellschaftliche Perspektive berücksichtigt werden. Im Unterschied zur reinen GKV-Perspektive werden bei der GKV-Versichertenperspektive auch Kosten berücksichtigt, die den Versicherten entstehen, z. B. aus Zuzahlungen (siehe Abschnitt 4.4.1). Je nach Auftrag für eine KNB kann es erforderlich sein, die Perspektive einzelner Sozialversicherungsträger zusätzlich zur Perspektive der GKV-Versichertengemeinschaft einzunehmen. Die Entscheidung, ob weitere Perspektiven in eine KNB aufgenommen werden sollen, hängt allein von der Frage ab, ob dies für die Entscheidungsträger relevant ist. Die Ergebnisse der Bewertung aus einer erweiterten Perspektive werden dem Entscheidungsträger gesondert ausgewiesen zur Verfügung gestellt.

4.1.3 Zeithorizont

Der Zeithorizont muss mindestens die durchschnittliche Studiendauer abbilden und somit für die Erstattungsentscheidung relevante Kosten- und Nutzenunterschiede zwischen den Interventionen einer KNB einbeziehen. Bei chronischen Erkrankungen sollte möglichst ein längerer Zeithorizont gewählt werden, insbesondere wenn Lebenszeitgewinne erwartet

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

werden [68,159,377,555]. Kosten und Nutzen sollen immer über denselben Zeithorizont modelliert werden.

Der angemessene Zeithorizont ist häufig länger als der Zeitraum, den die verfügbaren Primärdaten aus prospektiven Studien abdecken. Dann sollte unter Abwägung der Dauer der Studien ein der Krankheit angemessener Zeithorizont gewählt werden [259,555].

4.1.4 Auswahl der Komparatoren

Für die Herleitung einer Effizienzgrenze, der vom Institut gewählten Darstellungsform der Ergebnisse einer KNB (siehe Abschnitt 4.6), sind alle in einem Indikationsgebiet versorgungsrelevanten Interventionen in eine KNB einzubeziehen. Substanzen z. B. können auch zu Wirkstoffklassen zusammengefasst werden, wenn dies medizinisch sinnvoll erscheint und ausreichende Homogenität vorliegt (siehe Abschnitt 8.3.8).

4.1.5 Versorgungspfad

Für jede KNB sollte(n) zunächst (ein) Versorgungspfad(e) für das Indikationsgebiet entwickelt werden. Ein Versorgungspfad beschreibt Behandlungsabläufe für Patienten mit einer/mehreren spezifischen Indikation(en) in einer zeitlichen Abfolge und strukturiert sie nach Sektoren, beteiligten Berufsgruppen, Stadien und ggf. weiteren Aspekten. Dieser Versorgungspfad dient als Basis, um das entscheidungsanalytische Modell (siehe Abschnitte 4.1.6 und 4.2) zu erstellen. Weiterhin basieren auf dem Versorgungspfad auch die Recherchen nach Daten zu Kosten und weiteren notwendigen Daten für das Modell.

Zunächst sind auftragsbezogen der Krankheitsverlauf und die Versorgung in Deutschland kurz für die relevante Indikation mit Angabe der Quellen zu beschreiben. Die relevanten Interventionen und Behandlungsschritte in verschiedenen Leistungsbereichen inklusive der Leistungserbringer müssen innerhalb der Grenzen der Zulassung und des Wirtschaftlichkeitsgebots erbracht werden. Darüber hinaus muss die Anwendung innerhalb der Vorgaben der Richtlinien und Therapiehinweise bewertet werden, die im System der gesetzlichen Krankenversicherung gelten. Weiterhin sind aktuelle Therapieempfehlungen für Deutschland darzustellen, wobei gültige Leitlinien herangezogen werden sollten. Aus dem beschriebenen Versorgungskontext sollen die für die KNB relevanten Komponenten abgegrenzt werden, sodass ein modellrelevanter Versorgungspfad beschrieben werden kann. Wenn einzelne Komponenten dezidiert nicht in den Versorgungspfad aufgenommen werden, sollte diese Entscheidung begründet werden.

Selbst wenn eine Piggy-back-Studie vorliegt, das ist eine klinische Studie, in der neben Nutzen und Schaden einer Technologie zugleich auch Kosten erhoben werden, sollte begleitend ebenfalls ein Versorgungspfad abgebildet werden, damit die Kosten und weitere Daten, die in der Piggy-back-Studie erhoben wurden, anhand eines beigefügten Versorgungspfads nachvollzogen werden können.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

4.1.6 Modell

Piggy-back-Studien liegen sehr selten vor. Darüber hinaus werden ökonomische Daten in klinischen Studien meist nicht erhoben. Häufig sind die Daten gar nicht ausreichend für eine umfassende Aufbereitung der Kosten einer Intervention. Denn zum einen liefern klinische Studien selten Informationen zu den langfristigen ökonomischen Konsequenzen, die mit der Einführung einer neuen Intervention einhergehen. Zum anderen greifen sie nicht immer adäquat und vollständig die für die Kostenseite in Deutschland relevanten Versorgungsaspekte auf. Darüber hinaus kann auch ein protokollinduzierter Ressourcenverbrauch im Rahmen klinischer Studien Fehleinschätzungen auf der Kostenseite induzieren. Aus diesen Gründen ist die Modellierung der Kosten einer Intervention eine wesentliche Komponente der KNB (siehe Abschnitt 4.4). Ebenso kann in einer KNB der Nutzen modelliert werden, wenn einer KNB ein längerer als der studienbelegte Zeithorizont zugrunde gelegt werden soll (siehe Abschnitt 4.3).

4.1.7 Datengrundlage

Grundlagen und Bewertung von Daten, die auf der Nutzenseite in eine KNB einfließen, werden in Abschnitt 3.3 erläutert. Für das Maß des Gesamtnutzens werden einzelne Studien oder Datenerhebungen (siehe auch Abschnitt 4.3.3) verwendet.

Daten, die für die Abbildung der Versorgung, der Epidemiologie und der Kosten in die KNB eingehen, können auf verschiedenen Wegen erhoben werden und aus unterschiedlichen Quellen stammen. Dazu gehören Sekundärdaten, Leitlinien, Expertenbefragungen sowie Preiskataloge bzw. -verzeichnisse (siehe auch Abschnitte 4.4.4 und 4.5.2).

Auswertungen von Sekundärdaten sollten sich an den Leitlinien und Empfehlungen zur guten Praxis Sekundärdatenanalyse ausrichten [19]. Insbesondere sollen die Auswahl der Datenbasis, die Größe und relevante Merkmale der Stichprobe und der Studienpopulation (inkl. Ein- und Ausschlusskriterien), die statistischen Methoden und die Kontrolle von Störgrößen transparent beschrieben und begründet werden. Die Generalisierbarkeit und die Repräsentativität der Ergebnisse sollten erläutert werden. Die einzelnen Auswertungsschritte müssen nachvollziehbar sein; Plausibilitätskontrollen sind sicherzustellen. Werden Leitlinien verwendet, sollten diese aus dem deutschen Gesundheitssystem stammen und möglichst evidenzbasiert sein. Darunter werden Leitlinien verstanden, deren Empfehlungen auf einer systematischen Literaturrecherche beruhen, grundsätzlich mit einer Evidenz- und / oder Empfehlungseinstufung (Level of Evidence [LoE] und / oder Grade of Recommendation [GoR]) versehen sind und mit den Referenzen der ihnen zugrunde liegenden Primär- und / oder Sekundärliteratur verknüpft sind (modifiziert nach AGREE) [4].

Befragungen von Experten folgen den allgemein anerkannten Methoden und Vorgehensweisen der quantitativen Sozialforschung. Das heißt, dass bei Befragung von Experten explizit Angaben zu Rekrutierung, Anzahl und Expertise der Experten, Fragestellung, individuellen Antworten (nicht nur Mittelwerte), Art der Konsensfindung sowie Darstellung

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

und Umgang mit den Ergebnissen zu machen sind. Preiskataloge bzw. -verzeichnisse müssen aktuell sein und die für die GKV relevanten Preise abbilden.

4.1.8 Unsicherheit

Nach international üblichem Vorgehen werden folgende Arten der Unsicherheit in der gesundheitsökonomischen Entscheidungsanalyse unterschieden [63]:

Tabelle 5: Konzepte von Unsicherheit in der gesundheitsökonomischen Entscheidungsanalyse

Bezeichnung	**Konzept**	**Andere Bezeichnungen in der Literatur**	**Vergleichbares Konzept bei Regressions-modellen**
stochastische Unsicherheit	zufällige Variabilität im Ergebnis bei gleichen Probanden	Variabilität, First-Order Uncertainty	Fehlerterm
Parameterunsicherheit	Unsicherheit in der Schätzung des interessierenden Parameters	Second-Order Uncertainty	Standardfehler des Schätzers
strukturelle Unsicherheit	zugrunde liegende Annahmen im entscheidungsanalytischen Modell	Modellunsicherheit	Art des Regressions-modells, z. B. linear, loglinear

(adaptiert und übersetzt aus Briggs et al. [63])

Aufgrund seiner Vielschichtigkeit muss die Untersuchung der Unsicherheit in allen Bereichen einer KNB bedacht werden. Dazu orientiert sich das Institut an der Einteilung der Unsicherheit (siehe Tabelle 5).

Dazu werden in den Abschnitten 4.2, 4.3, 4.4 und 4.5 schon grundsätzliche Bemerkungen zur Unsicherheit und den Verteilungsannahmen gemacht. In Abschnitt 4.7 werden dann die Durchführung und die Darstellung der Untersuchung von Unsicherheit dargestellt.

4.1.9 Interpretation der Ergebnisse

Die Ergebnisse werden tabellarisch und grafisch in Form einer Effizienzgrenze dargeboten.

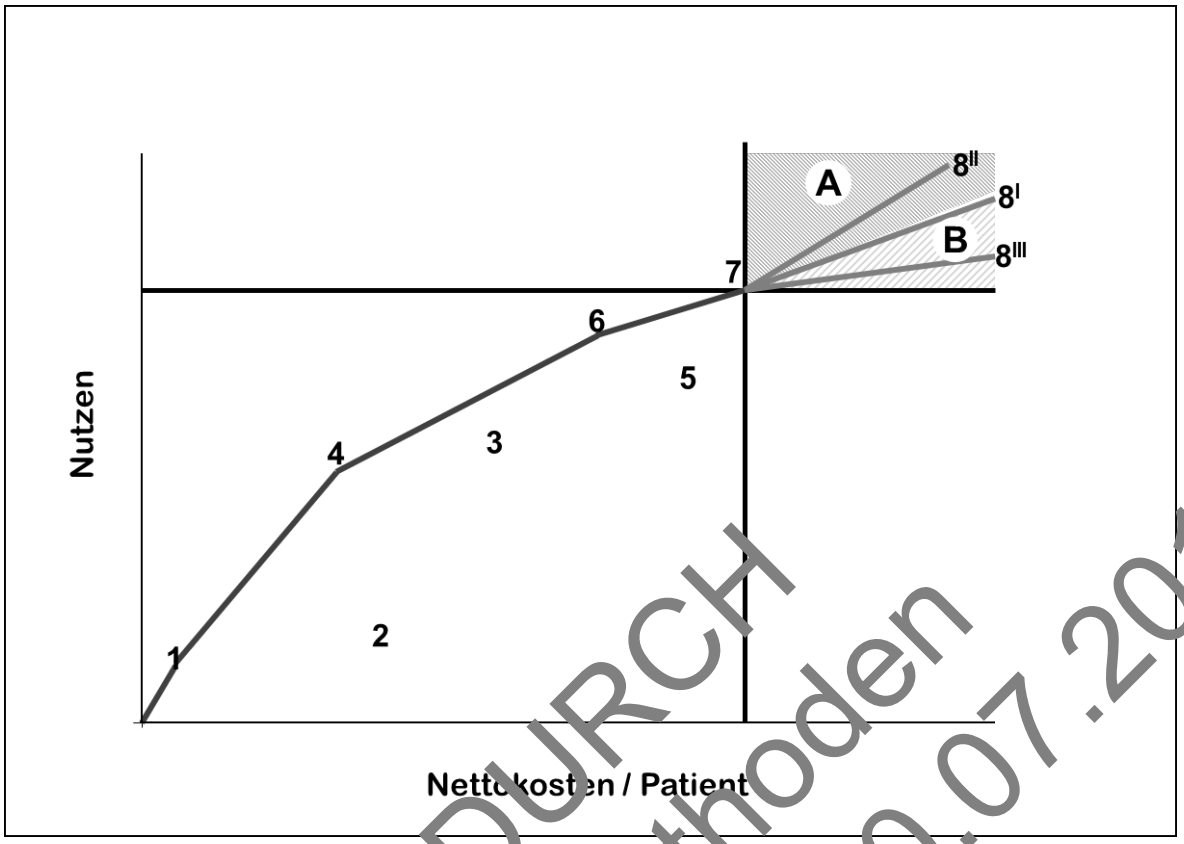
Eine Effizienzgrenze ist der Abbildung 9 zu entnehmen. Die Interventionen 1 bis 7 werden als Komparatoren mit ihren Kosten-Nutzen-Verhältnissen abgetragen. Die Interventionen 1, 4, 6 und 7 spannen eine Effizienzgrenze auf. Das letzte Segment der Effizienzgrenze kann unter der Annahme, dass es den Kehrwert der gegenwärtigen Zahlungsbereitschaft darstellt, linear extrapoliert werden. Dann gilt: Interventionen, die sich endpunktbezogen auf der Fläche A befinden (vgl. Intervention 8^{II}), haben nach ihrem Kosten-Nutzen-Verhältnis gegenüber dem extrapolierten letzten Segment der Effizienzgrenze ein besseres Kosten-Nutzen-Verhältnis

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

und könnten somit zum angegebenen Preis erstattet werden. Interventionen auf der Fläche B (vgl. Intervention 8^{III}) haben nach ihrem Kosten-Nutzen-Verhältnis gegenüber dem extrapolierten letzten Segment der Effizienzgrenze ein ungünstigeres Kosten-Nutzen-Verhältnis, sodass ihr Preis bezogen auf die Effizienzgrenze als nicht angemessen angesehen werden kann, da sich die vorhandene Effizienz verschlechtern würde. Der Entscheidungsträger kann unter Berücksichtigung des Kriteriums der Angemessenheit der Kosten von Interventionen auf der Fläche B einen Erstattungsbetrag verhandeln. Interventionen mit gleichbleibendem Nutzen-Kosten-Verhältnis (vgl. Intervention 8^{I}) erfüllen ebenfalls das Kriterium, dass ihr Preis im Vergleich zum extrapolierten letzten Segment der Effizienzgrenze angemessen wäre.

Die Effizienzgrenze zeigt auf, welche Interventionen den verhältnismäßig größten Nutzen im Verhältnis zu den entstehenden Kosten aufweisen. Ineffiziente Interventionen sind z. B. sowohl teurer als auch in Bezug auf einen Endpunkt von geringerem Nutzen als andere Interventionen. Wenn sowohl die Kosten als auch der durch die neue Intervention generierte Nutzen höher sind als diejenigen, die bereits in der Effizienzgrenze abgebildet sind, ist aus der Effizienzgrenze selbst nicht direkt ableitbar, welche Kosten dieser Intervention angemessen sind. Somit müssen weitere Kriterien herangezogen werden, um zu beurteilen, ob die Anwendung einer Zusatznutzen stiftenden, aber kostenintensiveren neuen Therapie angemessen ist. Das Institut geht davon aus, dass eine Verschlechterung der Effizienz in einer Indikation durch die Aufnahme neuer Interventionen unangemessen ist. Diese Effizienz wird durch die lineare Extrapolation der Steigung des letzten Segments der Effizienzgrenze umgesetzt. Daher werden bei einem gegebenen Nutzen einer zu beurteilenden Intervention solche Kosten-Nutzen-Verhältnisse als angemessen angesehen, die gemessen an der Effizienzgrenze nicht zu einer Verschlechterung der Effizienz in einem gegebenen Indikationsgebiet führen (siehe Abbildung 9).

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"



Nettokosten: Nettokosten sind die Kosten, die sich aus Versorgung ergeben, bereinigt um Kosteneinsparungen, sogenannte Cost-Offsets.

Abbildung 9: Darstellung der entscheidungsrelevanten Bereiche

Wenn ein Maß des Gesamtnutzens festgelegt wird (siehe Abschnitt 4.3.3), ist dieses als primäres Ergebnis anzusehen. Ist die Bestimmung mehrerer Effizienzgrenzen für die Bewertung einer Intervention erforderlich, so bleibt es dem Entscheidungsträger unter Beachtung der Relevanz der patientenrelevanten Endpunkte vorbehalten, eine Gewichtung vorzunehmen. Eine ähnliche Vorgehensweise kann auch bei Zulassung einer Intervention in mehreren Indikationsgebieten erwogen werden.

Die Zumutbarkeit einer Kostenübernahme durch die GKV-Versichertengemeinschaft hängt einerseits von der Angemessenheit des Preises einer Intervention ab, andererseits aber auch von den damit verbundenen künftigen Gesamtausgaben in Abhängigkeit der finanziellen Leistungsfähigkeit und Zahlungsbereitschaft der GKV-Versichertengemeinschaft. Da weder die finanzielle Leistungsfähigkeit noch die Zahlungsbereitschaft der GKV-Versichertengemeinschaft bewertet werden, wird zur Zumutbarkeit der Kostenübernahme keine konkrete Empfehlung abgegeben. Um zukünftige finanzielle Auswirkungen einer Kostenübernahme darzustellen, soll eine Ausgaben-Einfluss-Analyse (Budget-Impact-Analysis) durchgeführt werden, die dem Entscheidungsträger als Informationsgrundlage zur Entscheidung über die Zumutbarkeit dienen kann.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

4.2 Modellierung

4.2.1 Grundlagen

In einem gesundheitsökonomischen Entscheidungsmodell („Modell“) als Kernstück einer KNB werden Daten zu Nutzen und Kosten aus verschiedenen Quellen zusammengeführt, um Kosten-Nutzen-Verhältnisse von Interventionen im Krankheitsverlauf zu berechnen. Eine Zusammenführung von Daten aus verschiedenen Quellen mithilfe eines Modells ist oft aus einer Vielzahl von Gründen notwendig. Zumeist werden in einer Studie nicht alle für die Entscheidung relevanten Variablen erhoben. Auch wird ein gesundheitsökonomisches Modell explizit dazu verwendet, Nutzen und Kosten über den studienbelegten Zeitraum hinaus zu extrapolieren.

Gesundheitsökonomische Modelle sind somit, wie mathematisch-formalisierte Modelle, auch eine vereinfachte Abbildung der Realität. Durch eine bewusste Reduktion der Komplexität auf die für das Entscheidungsproblem relevanten Entscheidungsfaktoren und -variablen wird zudem analytische Klarheit geschaffen.

Für gesundheitsökonomische Modelle ist eine gründliche Dokumentation von zentraler Bedeutung. Diese Dokumentation sollte grundsätzlich zweigeteilt sein: zum einen eine allgemeine anschauliche Dokumentation des Vorgehens mit Darstellung und Begründung der getroffenen Entscheidungen und der Auswahl der Daten(-quellen); zum anderen eine technische Dokumentation, welche die funktionalen / mathematischen Beziehungen der Modellkomponenten darstellt, sodass ein sachverständiger Dritter die Ergebnisse des Modells unabhängig von einer bestimmten Software replizieren kann.

Der Grad der Komplexität bzw. das Maß der Reduktion eines Modells hängen immer von der Fragestellung ab und lassen sich a priori nicht festlegen. Daher ist neben der internen Validität eines Modells die Anwendbarkeit zu beschreiben und zu belegen. Die Modellstruktur (z. B. Gesundheitszustände), die sich innerhalb der Zulassung und im Rahmen der Leistungserbringung nach SGB V bewegen muss, wird zur Sicherung der externen Validität regelhaft mit externen klinischen Sachverständigen abgestimmt.

4.2.2 Grundsätzliche Aspekte der Modellentwicklung

Die Modelle müssen im Ergebnis detailliert den Nutzen und die Kosten abbilden, die sich in Deutschland für die zu betrachtende Intervention ergeben. Dazu müssen folgende Informationen in das Modell eingehen:

- Ergebnisse zu den Effekten (Nutzen und Schaden) der Interventionen,
- vollständige Erfassung der Krankheitskosten und
- alle Krankheits- und Behandlungsaspekte, die relevanten Einfluss auf die Nutzen- oder Kostenkomponenten des Modells haben können, z. B. in den Bereichen Demografie, Epidemiologie, Versorgungspfad(e).

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Da oft Daten zu einzelnen Aspekten fehlen, ist es besonders wichtig, den Einfluss dann getroffener Annahmen und Modellinputs auf die Ergebnisse durch Sensitivitätsanalysen zu explorieren.

Um die Validität und formale / inhaltliche Nachvollziehbarkeit von Modellierungen sicherzustellen, müssen folgende Bedingungen erfüllt sein:

- vollständige Transparenz mit klar beschriebenen und begründeten Modellinputs und Annahmen,
- hinreichende Tiefe zur ausreichenden Abbildung der modellierten Erkrankung, der mit ihr assoziierten Kosten und der betreffenden Versorgungspfade,
- ausreichende Flexibilität zur Berechnung multipler Szenarien bei variierenden Annahmen und Settings,
- Möglichkeit der Bestimmung von Unsicherheit in den vorhergesagten Kosten und Nutzenkomponenten.

4.2.3 Einflussdiagramm und Modellkonzept

Auf Basis der Überlegungen und Informationen, die zur Erstellung des Versorgungspfades geführt haben, werden die Grundlagen für das Modell in einem Einflussdiagramm und einem Modellkonzept dargelegt.

Ein Einflussdiagramm bildet grafisch die wesentlichen Zusammenhänge von Krankheitsverlauf, Patientencharakteristika, Pathophysiologie und Behandlung des Modells ab. Es zeigt die Faktoren auf, die Einfluss auf eine zu modellierende Fragestellung bzw. Fragestellungen haben oder haben können. Trotz seines Namens ist das Einflussdiagramm nicht per se eine kausale Darstellung.

Das Modellkonzept ergibt sich aus dem Einflussdiagramm und stellt das beabsichtigte Design in deutlich größerer Tiefe dar. Da auch die differenziertesten Modelle Vereinfachungen der Realität mit erforderlichen Annahmen und Einschränkungen in Bezug auf eingeschlossene Inhalte darstellen [3,159], kann das Modell nur dann richtig verstanden werden, wenn das Modellkonzept präzisiert und nachvollziehbar dokumentiert wird.

4.2.4 Auswahl der Modellierungstechnik

Die Wahl der geeigneten Modellierungstechnik hängt von der Fragestellung, den Charakteristika der zu bewertenden Intervention, der entsprechenden Erkrankung und den Rahmenbedingungen ab. Bei der Auswahl der Modellierungstechnik ist für das Institut der Grundgedanke leitend, dass das ökonomische Modell so differenziert und komplex sein sollte, wie es für die adäquate Beantwortung der gestellten Fragestellung(en) erforderlich ist. Die Datenlage an sich sollte nicht die Wahl der Modellierungstechnik bestimmen. Sollte durch die Wahl der Modellierungstechnik eine Modifikation des Modellkonzepts notwendig sein, sollte die Wahl der Modellierungstechnik nochmals überprüft werden [87].

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Die gewählte Modellierungstechnik muss auch mit bereits durchgeführten / veröffentlichten Modellierungen zu gleichen bzw. eng verwandten Entscheidungsproblemen verglichen werden. Wenn von diesen schon vorliegenden Modellen abgewichen wird, sollte dies diskutiert und begründet werden. Da die angemessene Modellierungstechnik jedoch immer von der zugrunde liegenden Fragestellung abhängt, sind feststehende A-priori-Vorgaben nicht sinnvoll, zumal sich der internationale Standard der Gesundheitsökonomie kontinuierlich weiterentwickelt [300]. Die folgenden zentralen Problemkreise sind grundsätzlich zu berücksichtigen:

- Zeitliche Dimension: Für welchen Zeithorizont werden Aussagen getroffen, wird extrapoliert und wie wird Zeit innerhalb des Modells strukturiert (z. B. kontinuierlich / diskret, Zykluslänge)?
- Analyseeinheit: Welche Analyse- bzw. Beobachtungseinheit wird abgebildet (z. B. Individuen, Kohorten) und welche Merkmale werden berücksichtigt (Alter, Geschlecht etc.)?
- Interaktion: Welche Interaktion zwischen den Analyseeinheiten selbst, d. h. Patienten, oder anderen Elementen des Modells werden abgebildet?

Da für eine Modellierung oft Daten aus unterschiedlichen Quellen zusammengeführt werden, kann es notwendig sein, diese Daten in das gleiche Format zu transformieren, z. B. denselben Zeitbezug.

4.2.5 Modelldokumentation und Modellvalidierung

A) Modellvalidierung und strukturelle Unsicherheit

Ein Simulationsmodell, das für eine Fragestellung valide ist, kann für eine andere nicht valide sein [350]. Der externe Validierungsprozess muss daher jede intendierte Verwendung des Modells abdecken und bei Einsatz für andere Fragestellungen muss die Validierung erneut erfolgen. Über das Vorgehen zur Validierung eines Modells herrscht Uneinigkeit, aber es gibt einige grundlegende Schritte, die eingehalten werden müssen [486].

Ein Schlüsselelement der Validierung befasst sich mit der Frage, ob das Modell inhaltlich die Realität von Krankheitsverlauf und Behandlung hinreichend abbildet. Die Plausibilitätsprüfung (Augenscheinvalidität) bezieht sich auf das Einflussdiagramm, das Modellkonzept, die Datenbeschaffung, die Erarbeitung der funktionalen Beziehungen und die Auswahl der Modellierungstechnik.

Ein weiteres Schlüsselelement der Validierung ist die korrekte technische Umsetzung des Modells (interne bzw. technische Validierung). Dieser Aspekt bezieht sich auf die Frage, ob die technische Implementation tatsächlich das Modellkonzept korrekt umsetzt, z. B. sind die Ergebnisse numerisch korrekt und stabil.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Ein drittes Element der Validität ist die Vorhersagevalidität. Inwieweit sagt das Modell die Zukunft voraus, d. h. finden die vorhergesagten Ergebnisse eine Entsprechung in der Realität. Diese ist sicherlich die wünschenswerteste Form der Validität, ist jedoch die am schwierigsten belegbare, wenn dies überhaupt möglich ist [571]. Jedoch ist ein Abgleich von Modellergebnissen mit vergangenen, vergleichbaren Beobachtungen sinnvoll und Unterschiede sollten erklärbar sein. Dies gilt auch im Vergleich zu anderen gesundheitsökonomischen Modellen (Cross Validity).

Eine spezielle Form der Unsicherheit bei der Modellerstellung bzw. -validierung ist die sogenannte strukturelle Unsicherheit. Bei der strukturellen Unsicherheit wird hinterfragt, inwieweit die funktionalen Zusammenhänge, die dem Modell zugrunde liegen, tatsächlich valide sind und ob andere funktionale Formen nicht zutreffender wären. Falls es bei der Planung und Erstellung eines Modells augenscheinlich wird, dass die strukturelle Unsicherheit bei der zugrunde liegenden Fragestellung von Relevanz ist, kann es notwendig sein, mehrere (alternative) Modelle zu erstellen, um die Folgen dieser Form der Unsicherheit auf das Ergebnis zu quantifizieren [540].

B) Allgemeine Dokumentation

Das Institut erstellt einen detaillierten technischen Bericht mit Beschreibung aller Modellierungsschritte von der Entwicklung des Einflussdiagramms bis zur finalen Validierung. Zusätzlich ist eine voll ausführbare Version des Modells einschließlich eines Anwendermanuals zur Verfügung zu stellen. In Anlehnung an andere vorgeschlagene Leitlinien [107,410,584] sollte die Dokumentation des Modells Folgendes beinhalten:

- das zur Anleitung der Modellentwicklung verwendete Einflussdiagramm
- Details zum Modellkonzept
  - Beschreibung der Zielpopulation(en) im Rahmen der Evaluation einschließlich Subgruppen
  - Beschreibung der evaluierten Interventionen
  - Auswahl und Begründung der Modellsettings (Simulationsgröße, Zeithorizont, Diskontierungsraten etc.)
  - Überblick zu aktuellen gesundheitsökonomischen Evaluationen im untersuchten Indikationsgebiet
- Beschreibung aller Datenquellen. Eine Begründung der Auswahl von Datenquellen muss gegeben werden.
- Details zu allen im Modell eingesetzten funktionalen Beziehungen. Wenn diese speziell für das Modell entwickelt wurden, müssen detaillierte Angaben zu den angewendeten Methoden gegeben werden.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

- Auflistung aller Annahmen hinsichtlich der Datenquellen und der Modellstruktur. Besonders wichtig ist eine detaillierte Darstellung jeglicher Annahme und verwendeten Technik zur Projektion über den Zeitraum hinaus, für welchen die Daten gelten.
- Begründung der verwendeten Modellierungstechnik
  - Beschreibung, in welcher Art die Technik den erforderlichen Eigenschaften entspricht
- Überblick über die eingesetzten Validierungstechniken und ihre Ergebnisse
- Detaillierte Ergebnisdarstellung einschließlich einer Bewertung des Einflusses von:
  - Anwendung der Intervention in relevanten Subgruppen
  - Unsicherheit der Inputdaten (siehe Abschnitt 4.7 zu Sensitivitätsanalysen)
- Interpretation der Ergebnisse einschließlich einer Beschreibung der Limitationen des angewendeten Ansatzes

C) Technische Dokumentation und elektronische Version des Modells

Die technische Dokumentation ist von zentraler Bedeutung für das Verständnis und die Bewertung des zugrunde liegenden gesundheitsökonomischen Modells. Alle verwendeten Variablen sind zu benennen und zu definieren. Die funktionalen / mathematischen Beziehungen der Modellkomponenten sind darzustellen und gegebenenfalls zu begründen. Die formal-mathematischen Beziehungen sollen alle in das Modell einfließenden Inputvariablen (z. B. Gesundheitszustände) mit den entsprechenden Operatoren (z. B. altersspezifische Übergangswahrscheinlichkeiten) verbinden. Auch muss weiterhin die Herleitung von Zwischen- bzw. Endwerten dargestellt werden.

Alle kalkulatorischen Schritte innerhalb der Software sind nachvollziehbar zu dokumentieren. Dies kann in aller Regel durch die Dokumentation des Programmcodes erfolgen, mit der die elektronische Version des Modells implementiert ist. Bei Tabellenkalkulationsprogrammen (z. B. Excel) ist die Reihenfolge der Kalkulationsschritte nicht direkt der elektronischen Version zu entnehmen. Diese müssen dann ggf. schriftlich so dokumentiert werden, dass die Reihenfolge der Berechnungsschritte ersichtlich ist.

Eine elektronische Version des Modells muss zur Verfügung gestellt werden mit dem Einvernehmen, dass das Modell öffentlich zugänglich gemacht wird und ggf. für zukünftige Evaluationen adaptiert werden kann. Die elektronische Modellversion muss vollständig zugänglich sein und den Reviewern sowie der Öffentlichkeit ermöglichen, alle in der Analyse eingesetzten Formeln und Beziehungen einzusehen und das Modell mit unterschiedlichen Inputdaten laufen zu lassen. Um das Review des Modells zu erleichtern, sollte der elektronischen Version eine Anwenderanleitung beigefügt werden, die beschreibt, welche Software und Hardware benötigt werden, wie Modellinputs verändert werden können, wo diese Inputs im Modell gefunden werden können, wie das Modell ausgeführt werden kann und wie Ergebnisse extrahiert werden können.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

4.3 Nutzen

Die Methoden, mit denen der Nutzen im Rahmen der Nutzenbewertung bestimmt wird, werden in Kapitel 3 beschrieben. In einer KNB kann der studienbelegte Nutzen vom modellierten Nutzen unterschieden werden, wenn einer KNB ein längerer als der studienbelegte Zeithorizont zugrunde gelegt wird.

4.3.1 Übertragung und Darstellung des Nutzens

Damit der Nutzen in die KNB mithilfe der Effizienzgrenze einfließen kann, muss er approximativ kardinalskaliert sein. In der KNB kann der approximativ kardinalskalierte Nutzen, der ggf. direkt aus den Studienergebnissen ableitbar ist, oder ein transformierter approximativ kardinalskalierter Nutzen auf der vertikalen Achse abgetragen werden. Die Einschränkung der Bedingung, dass ein Nutzen „nur“ approximativ kardinalskaliert sein muss, ergibt sich aus folgender Überlegung: Eine in der Nutzenerhebung verwendete Skala muss nicht über ihre gesamte Spannbreite kardinalskaliert sein. Es genügt, wenn diese über den Bereich, der für die Definition des patientenrelevanten Zusatznutzens von Bedeutung ist, das Kriterium erfüllt, kardinalskaliert zu sein. Oft zeigen beispielsweise unterschiedliche Erhebungsinstrumente sogenannte Boden-(Floor-) oder Decken-(Ceiling-) Effekte an den Rändern ihrer Wertebereiche, sind jedoch im restlichen Wertebereich kardinalskaliert [57,182,452].

Hier wird keine spezielle Vorgehensweise zur Wertbestimmung des Nutzens auf einer Kardinalskala empfohlen, da jedes Indikationsgebiet unterschiedliche Möglichkeiten zur Bewertung des Nutzens bieten kann, die der Anforderung nach Kardinalskalierung genügen.

4.3.2 Endpunkte

Der Nutzen kann auf der vertikalen Achse der Effizienzgrenze durch einzelne oder aggregierte patientenrelevante Endpunkte dargestellt werden (zur Definition des patientenrelevanten medizinischen Nutzens bzw. Schadens siehe Abschnitt 3.1.1). Wenn mehrere patientenrelevante Endpunkte nebeneinander dargestellt werden, wird für jeden patientenrelevanten Endpunkt eine eigene Effizienzgrenze erstellt. Andernfalls wird der Nutzen aggregiert zu einem einzigen Maß des Gesamtnutzens, das anschließend in einer Effizienzgrenze abgetragen wird. Ein Maß des Gesamtnutzens ist in einer ganz allgemeinen Definition eine Aggregierung der Bewertung von Nutzen und Schaden in einer Größe, wobei unterschiedliche patientenrelevante Endpunkte zu einem einzigen Maß zusammengefasst werden. Es kann sowohl in der Nutzenbewertung als auch in der Kosten-Nutzen-Bewertung berücksichtigt werden. Die in diesem Kapitel dargestellten Anforderungen an die Erhebung eines Maßes des Gesamtnutzens gelten auch, wenn es im Rahmen der Nutzenbewertung eingesetzt wird.

4.3.3 Maß des Gesamtnutzens

International gibt es unterschiedliche Maße, den Gesamtnutzen auszudrücken bzw. zu erfassen. Dazu gehören das qualitätsadjustierte Lebensjahr (QALY = Quality-Adjusted Life Year)

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

und das behinderungsbereinigte Lebensjahr (DALY = Disability-Adjusted Life Year). Maße wie das „Saved Young Life Equivalent [416]“ oder das „Healthy Years Equivalent“ (HYE) [206] sind mit dem Anspruch angetreten, Schwächen des QALY zu korrigieren, das am weitesten verbreitet ist.

In diesem Kontext werden je nach methodischem Ansatz oder ökonomischer Theorie in der wissenschaftlichen Literatur die Begriffe Präferenzen (Preferences), Nutzwerte bzw. Teilnutzenwerte (Utilities) oder Werte (Values) verwendet [159]. Auf die weitere Debatte der Begriffe und des Stellenwerts der Erhebungsmethoden im Zusammenhang der Frage eines „Welfarist“- versus „Extra-Welfarist“-Rahmens sei hier nur hingewiesen [69]. In Anlehnung an das SGB V wird im Weiteren von Gewichten gesprochen, mit deren Hilfe einzelne patientenrelevante Endpunkte in ein Maß des Gesamtnutzens überführt werden können.

Wenn der G-BA für eine KNB nach § 35b Absatz 1 Satz 2 SGB V das Maß des Gesamtnutzens festlegt (siehe Abschnitt 4.9), werden ein entsprechendes Instrument und ggf. die dafür festgelegten Erhebungsmethoden oder eine schon festgelegte Gewichtung von Endpunkten nach den Vorgaben im Auftrag eingesetzt. Die Ergebnisse sollten dem Entscheidungsträger zusammen mit der Endpunktgewichtung zur Verfügung gestellt werden. Hierdurch ergibt sich für den Entscheidungsträger die Option, einen aus mehreren zusatznutzenbasierten Erstattungspreisen gewichteten Erstattungsbetrag zu verhandeln.

A) Das QALY als Maß des Gesamtnutzens

Zur Berechnung des QALYs werden Gewichte für Gesundheitszustände erhoben. Befragte wägen dabei ab, wie sie diese Gesundheitszustände empfinden bzw. einschätzen. Das Ergebnis ist dann eine Indexzahl für jeden Gesundheitszustand. Unter Integration der Dauer der entsprechenden Gesundheitszustände können diese zumeist als Nutzwerte (Utilities) bezeichneten Gewichte in QALYs überführt werden. Die Erhebung und Berechnung von Nutzwerten wird z. B. dargestellt in Puhan et al. [445], Lipscomb et al. [363] und Tierney et al. [553].

Das Institut schließt nicht aus, in Kosten-Nutzen-Bewertungen auf QALYs als Maß für den Gesamtnutzen zurückzugreifen. QALYs sollten nur verwendet werden, wenn die einfließenden Werte zu den Gesundheitszuständen zunächst bei den Betroffenen erhoben worden sind, die aktuell oder in der Vergangenheit diese Gesundheitszustände erlebt haben. Die Daten sollten bei den an klinischen Studien Teilnehmenden erhoben worden sein. Wenn generische Indexinstrumente eingesetzt wurden, muss ein in Deutschland validierter Tarif bei der Ermittlung des Nutzwertes angewendet werden. Die Verwendung von QALYs sowie ihre Erhebung und Umrechnung in einen deutschen Tarif müssen in jedem Fall nachvollziehbar dargestellt und begründet werden. Im Übrigen gelten alle üblichen Standards, die an entsprechende Verfahren und Instrumente angelegt werden: D. h., Nachweise der Objektivität, Reliabilität, Validität und Reagibilität / Änderungssensitivität müssen vorliegen. Parallel zum Einsatz eines generischen Instruments sollten krankheitsspezifische Instrumente zur

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Erhebung der Lebensqualität in den klinischen Studien eingesetzt werden. Daher wird auch vom Mapping krankheitsspezifischer auf generische Instrumente abgeraten.

Angesichts der anhaltenden Diskussion um die Vor- und Nachteile verschiedener Instrumente, insbesondere der Multi-Attribute-Utility-Instrumente (MAUI), mit denen Lebensqualität, subjektives Wohlbefinden oder Nutzwerte erhoben bzw. abgebildet werden sollen/können, muss man sagen, dass keine allgemeine Empfehlung abgegeben werden kann. Die Wahl eines Instruments hängt davon ab, welches dieser 3 Konzepte im Vordergrund stehen soll und welche Dimensionen von Lebensqualität vorzugsweise erhoben werden sollen [450].

Die wissenschaftliche Debatte um die ethischen und methodischen Probleme des QALY-Konzepts an sich und deren Lösung oder einer daran geknüpften Zahlungsbereitschafts-schwelle in einer KNB sowie des Einsatzes des QALY zur reinen Abwägung von Nutzen und Schaden soll hier nicht erneut aufgerollt werden. In diesem Zusammenhang sei auf eine Reihe von Publikationen verwiesen [137,153,154,246,363,377,391,417,573].

B) Erhebung von Präferenzen zur Erstellung eines Maßes des Gesamtnutzens

Neben den oben genannten krankheitsübergreifenden Maßen können auch Verfahren der multikriteriellen Entscheidungsfindung oder der Präferenzerhebung eingesetzt werden, wenn ein Maß des Gesamtnutzens zum Vergleich von Interventionen erhoben werden soll. Für Endpunkte, die mittels dieser Verfahren gewichtet werden, gelten alle Vorgaben nach SGB V und AMNutzenV. Surrogate können nur bei nachgewiesener Validität herangezogen werden. Im Bereich der Gesundheitsversorgung setzen sich als Methoden der multikriteriellen Entscheidungsfindung oder der Präferenzerhebung weitgehend das Analytic-Hierarchy-Process(AHP)-Verfahren und die Conjoint-Analyse (CA) durch [62,121,277,382,466]. Das Institut kann daher indikationsbezogen auf diese Verfahren zur Generierung eines Maßes des Gesamtnutzens zurückgreifen. Allerdings gibt es noch ungelöste methodische Probleme beim Einsatz dieser Verfahren, sodass gegenwärtig eine routinemäßige Anwendung dieser Methoden nicht vorgesehen ist.

Für das AHP-Verfahren [151,152] wird ein Entscheidungsproblem in sogenannte Kriterien zerlegt. Diese werden dann in eine Hierarchie gebracht. So kann z. B. ein Arzneimittel nach den Kriterien Mortalität, Morbidität und Lebensqualität beurteilt werden. Die Kriterien können in weitere Subkriterien zerlegt werden, die Endpunkten entsprechen können [276]. Die am AHP-Verfahren 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 [468,470,471] können über den sog. rechten Eigenvektor die Gewichte für die Kriterien bzw. Subkriterien ermittelt werden, die sich zu 1 aufsummieren müssen. Eine methodische Weiter-entwicklung, die auch zulässt, dass Kriterien voneinander abhängig sind, ist das Analytic-Network-Process(ANP)-Verfahren [467,469].

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Die CA gehört zur Gruppe der Stated-Preference-Techniken [62]. Eine Entscheidung wird in sogenannte Attribute zerlegt, die Endpunkten entsprechen können. Für jedes Attribut werden Ausprägungen (Level) angegeben. Die Wahlalternativen (Stimuli) werden für ein Discrete-Choice-Experiment (DCE = wahlbasierte CA) aus den Attributen mit unterschiedlichen Ausprägungen zusammengesetzt. Die Befragten werden dann mit einem Set von (theoretischen) Szenarien konfrontiert (Wahlszenario = Choice Set), die aus mindestens 2 Wahlalternativen (Stimuli) bestehen. Aus der Wahl der Szenarien werden dann in einem Regressionsmodell Koeffizienten für die Ausprägungen der Attribute ermittelt. Der Einfluss der Attribute auf die Entscheidung kann dargestellt werden, indem im Anschluss daran Gewichte für die Attribute gebildet werden. Diese Gewichte können wiederum auf 1 normiert werden.

Ausgehend von ihrer Entwicklung wurde das AHP-Verfahren ausgerichtet auf die Entscheidungsfindung bei widerstrebenden Zielen in Gremien, z. B. dem Management eines Unternehmens, und die CA auf die Erhebung von Präferenzen, um Kaufentscheidungen voraussehen zu können und Produkte anpassen zu können. Mittlerweile spielen beide Verfahren eine Rolle bei der Identifikation und Priorisierung patientenrelevanter Endpunkte, z. B. vor Planung einer Studie, und der Bestimmung des Nettonutzens (Maß des Gesamtnutzens) von Interventionen [118,402].

Eine klare Zuordnung, wann welches Verfahren bevorzugt werden soll, lässt sich somit kaum ableiten. Ein AHP-Verfahren wäre wohl eher angebracht, wenn in einer geschlossenen Gruppe eine Entscheidung gefällt werden soll [276,278], wohingegen man eine CA durchführen würde, wenn man auch die Kompensationsleistung für entgangenen Nutzen berücksichtigen wollte, wenn eine Intervention nicht erstattet wird. Nebenbei bemerkt, es gibt auch die Möglichkeit, mittels CA QALYs zu berechnen [196,229]. Für die Auswahl eines der beiden Verfahren sind aber folgende Kriterien heranzuziehen: Für die CA gilt, dass maximal 6 bis 7 Attribute einbezogen werden können. Beim AHP-Verfahren gibt es diese Grenze nicht. Weiterhin scheint das AHP-Verfahren die Befragten vor geringere kognitive Anforderungen zu stellen, was man je nach Indikation berücksichtigen könnte. Diese Einschätzungen können sich derzeit nur teilweise auf empirische Daten stützen, sodass eine evidenzgesteuerte Auswahl eines der beiden Verfahren derzeit nicht möglich ist. Außerdem besteht zu einigen Fragen wie der Reliabilität beider Verfahren Forschungsbedarf.

Die Stärken und Schwächen beider Methoden können hier nicht im Detail beschrieben werden [413]. Nachvollziehbarkeit in Planung, Durchführung, Auswertung und Bewertung einer jeden Umsetzung ist daher zentral. Für die CA gibt es eine grundlegende Liste von Kriterien, um hohe Qualität, Transparenz und Verlässlichkeit des Ergebnisses einer CA zu gewährleisten [61], viele der Anforderungen gelten auch für die Durchführung eines AHP-Verfahrens.

Im Detail sollten bei Planung, Durchführung, Auswertung und Bewertung der Ergebnisse von Erhebungen mithilfe eines der beiden Verfahren folgende Anforderungen erfüllt werden:

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

- Vollständigkeit der Kriterien oder Attribute,
- umfassende Dokumentation des Vorgehens der Auswahl der zu Befragenden und Darstellung, inwieweit sie anhand soziodemografischer und krankheitsspezifischer Faktoren repräsentativ für das Kollektiv an Betroffenen sind.

Nicht nur wer befragt wird, sondern auch die Rekrutierungswege müssen berichtet werden. Weiterhin gilt es, eine Fallzahl zu planen. Für die CA gibt es Faustformeln für eine Fallzahlschätzung [312]. Für das AHP-Verfahren gibt es derzeit keine Methode, um eine Fallzahl zu schätzen. Für ein AHP-Verfahren können aber zumindest Kriterien der Repräsentativität herangezogen werden, die auch anderen Befragungen zugrunde gelegt werden (Größe der Stichprobe, Art der Ziehung etc.):

- Untersuchung der befragten Population auf Homogenität
- umfassende Dokumentation der Auswertung samt Übergabe der Rohdaten inklusive der Fragen im originalen Wortlaut
- Sprache, Wahl und Steuerung der Umsetzung inklusive einer Einschätzung der Verzerrung durch die Art des Designs

  Eine den Befragten angemessene Sprache sollte gewählt werden.
- Untersuchung der Konsistenz und der Unsicherheit der Ergebnisse und Durchführung von entsprechenden Analysen (z. B. Sensitivitätsanalysen)

4.3.4 Unsicherheit und Verteilung von Nutzendaten

Für geschätzte Effekte im Rahmen einer Nutzenbewertung können im Allgemeinen Konfidenzintervalle bzw. Credible Intervals (falls Bayes'sche Methoden gewählt werden, siehe Abschnitte 8.3.2 und 8.3.9) berechnet werden, die die Präzision bzw. Unsicherheit der Punktschätzer anzeigen. Für die weitere Untersuchung der Unsicherheit sind geeignete Annahmen zu treffen, da viele Effekte nicht normalverteilt sind.

Schätzer aus indirekten Vergleichen (siehe Abschnitt 8.3.9) sind mit höherer Unsicherheit behaftet als Schätzer aus direkten Vergleichen; darauf wird bei der Bewertung der Unsicherheit hingewiesen. Für Schätzer aus indirekten Vergleichen, die aufgrund von unterschiedlichen Annahmen zu A-priori-Verteilungen z. B. voneinander abweichen, sind ggf. Szenarioanalysen durchzuführen.

Insbesondere auch für das Gesamtmaß des Nutzens gilt es, die in Abschnitt 4.7 geforderten Untersuchungen der Unsicherheit (Sensitivitätsanalysen) durchzuführen.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

4.4 Kosten

4.4.1 Perspektive und zu berücksichtigende Kosten

Auftragsabhängig können die (reine) GKV-Perspektive, die Perspektive der GKV-Versichertengemeinschaft, die Sozialversicherungsperspektive bzw. die Perspektive einzelner Sozialversicherungsträger sowie die gesellschaftliche Perspektive berücksichtigt werden. Im Folgenden werden die relevanten zu berücksichtigenden Kosten nach Perspektiven abgegrenzt.

Aus der (reinen) GKV-Perspektive werden alle direkten erstattungsfähigen Kosten und Transferleistungen (z. B. Krankengeld) berücksichtigt. Weiterhin können, soweit für die KNB relevant, die Anteile der Beitragszahlungen an Rentenversicherung, Pflegeversicherung und Arbeitslosenversicherung, die die GKV bei Erkrankung nach 6 Wochen Arbeitsunfähigkeit übernehmen muss, und Beitragsausfälle (während der Krankengeldzahlung) berücksichtigt werden.

In der Perspektive der GKV-Versichertengemeinschaft sind zusätzlich zu den direkten erstattungsfähigen Kosten eigene Aufwendungen der Versicherten zu berücksichtigen, die nicht erstattungsfähig sind (siehe Abschnitt 4.4.2). Krankengeld wird hingegen nicht berechnet, da das Geld lediglich von der GKV zu den Versicherten umverteilt wird, sodass keine Mehrkosten für die Versichertengemeinschaft entstehen [465]. Beitragsausfälle zur gesetzlichen Krankenversicherung durch Krankheit werden damit ebenso nicht berücksichtigt.

Tabelle 6: Perspektive und relevante zu berücksichtigende Kosten[1]

Kostenkategorie / Perspektive	Direkte medizinische Kosten		Direkte nichtmedizinische Kosten		Indirekte Kosten	Transferleistungen
	erstattungsfähig	nicht erstattungsfähig	erstattungsfähig	nicht erstattungsfähig	-	-
Gesellschaft	ja	ja	ja	ja	ja	nein
Sozialversicherung	ja	nein	Ja	nein	nein	ja
GKV-Versichertengemeinschaft	ja	ja	ja	ja	nein	nein
GKV	ja	nein	ja	nein	nein	ja
GKV: gesetzliche Krankenversicherung						

[1] Die Inhalte der jeweiligen Kostenkategorie können je nach eingenommener Perspektive unterschiedlich sein. In einer engeren Auslegung der GKV-Versichertengemeinschaft z. B. werden ggf. Zuzahlungen berücksichtigt, aber keine weiteren Ausgaben der Versicherten. Dies wird in den Aufträgen des G-BA konkretisiert.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Anders als bei der Perspektive der GKV-Versichertengemeinschaft werden bei der Sozialversicherungsperspektive bzw. der Perspektive einzelner Sozialversicherungsträger keine Zuzahlungen der Versicherten berechnet. Berücksichtigt werden erkrankungsbedingte erstattungsfähige Ausgaben inklusive Transferleistungen.

In der gesellschaftlichen Perspektive werden Kostenkomponenten unabhängig davon berücksichtigt, wer sie trägt und wer von den Effekten einer Intervention betroffen ist. Generell sind die Kosten zu berücksichtigen, die bei allen Sozialversicherungsträgern und sonstigen Betroffenen anfallen (siehe Tabelle 6). Zeitaufwand bei Patienten und / oder ggf. bei Angehörigen, der einen Arbeitsausfall darstellt, wird nicht nochmals als Zeitaufwand berücksichtigt. Dies würde mit der Berücksichtigung von Produktivitätsverlusten zu Doppelzählungen führen. Ebenso werden Transferzahlungen und durch die GKV finanzierte Beitragszahlungen an die Sozialversicherungen nicht berücksichtigt, weil sie nur umverteilt werden und aus volkswirtschaftlicher Sicht keine Mehrkosten entstehen [465].

Generell ist bei der Bestimmung der Kosten in jeder Perspektive zu prüfen, ob diese Kosten und ggf. Einsparungen für die betrachteten Interventionen bzw. Indikationsgebiete und Patientengruppen relevant sind.

4.4.2 Abgrenzung der Kosten

A) Direkte Kosten

Direkte Kosten beziehen sich auf den Ressourcenkonsum bei der aktuellen und zukünftigen Erstellung von Gesundheitsleistungen. Sie werden weiter differenziert in direkte medizinische und direkte nicht medizinische Kosten. Unter direkten medizinischen Kosten ist der Ressourcenverbrauch zu verstehen, der im Gesundheitssektor bei der Gesundheitsversorgung entsteht. Sie schließen Kosten z. B. für Krankenhausaufenthalte, ambulante Arztkontakte, Arzneimittel und Heil- und Hilfsmittel ein. Direkte nicht medizinische Kosten umfassen die Ressourcen, die die Erstellung medizinischer Leistungen im Gesundheitssektor unterstützen, z. B. Fahrtkosten zu medizinischen Interventionen oder den bewerteten Zeitaufwand von Betroffenen und ihren pflegenden Angehörigen, der durch die Erkrankungen entsteht.

Erstattungsfähige Kosten umfassen Ausgaben für Gesundheitsleistungen, die von der GKV bzw. von anderen Sozialversicherungsträgern finanziert werden. Nicht erstattungsfähige medizinische Kosten sind Leistungen, die von den Versicherten direkt getragen werden, wie Zuzahlungen für Arzneimittel, Heil- und Hilfsmittel und ambulante Arztkontakte. Nicht erstattungsfähige nicht medizinische Kosten sind z. B. krankheitsbedingte Nettoeinkommens-

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

verluste[2] (z. B. finanzielle Einbußen durch den Empfang von Krankengeld, das unterhalb des Nettoeinkommens liegt) oder der Zeitaufwand von Betroffenen und Angehörigen.

Die meisten empirischen Studien berücksichtigen die Auswirkungen auf die Freizeit von Betroffenen und Angehörigen nicht. Insofern bezieht das Institut nicht regelhaft den Zeitaufwand für Betroffene und ihre Angehörigen in die gesellschaftliche Perspektive ein. Für den Fall, dass dennoch repräsentative und valide Informationsquellen zum Zeitaufwand vorliegen, kann dieser Zeitaufwand in Sensitivitätsanalysen zur Perspektive der Gesellschaft berücksichtigt werden. Die Lebensqualität von Angehörigen wird im Allgemeinen auf der Nutzenseite nicht berücksichtigt. Sollten deren Freizeitverluste betrachtet werden, sollten sie auch auf der Kostenseite bewertet werden [68,311,425,585].

B) Indirekte Kosten

Indirekte Kosten bezeichnen den Produktivitätsausfall bei Arbeitsunfähigkeit, Erwerbsunfähigkeit (bei langfristiger Erkrankung oder Behinderung) und vorzeitigem Tod.

Das Institut berücksichtigt Produktivitätsausfälle primär auf der Kostenseite. So empfiehlt es auch überwiegend die Literatur [70,71,94,159,311,500,501]. Produktivitätsverluste wegen vorzeitigen Todes (Mortalitätskosten) sind nicht auf der Kostenseite auszuweisen, wenn Mortalität bereits auf der Nutzenseite berücksichtigt wird, um Doppelzählungen zu vermeiden. Nur für die Fälle, bei denen der betrachtete Endpunkt nicht die Mortalität oder Lebenszeit betrifft, werden die Mortalitätskosten auf der Kostenseite abgebildet. Kosten für die Gesellschaft (Ausfall von Steuern und Sozialversicherungsbeiträgen) werden immer auf der Kostenseite abgebildet [311,500,501].

International wird diskutiert, auch unbezahlte Arbeit (z. B. Hausarbeit) in einer KNB zu berücksichtigen. Das Institut sieht dies im Regelfall vorerst nicht vor.

C) Transferleistungen

Transferleistungen können, soweit für die KNB relevant, berücksichtigt werden. Transferleistungen sind generell nicht zu berücksichtigen, wenn Zahlungen lediglich umverteilt werden und damit für die gewählte Perspektive keine Mehrkosten entstehen.

D) Intangible Kosten

Mit intangiblen Kosten werden nicht direkt als Ressourcenverbrauch berechenbare bzw. in Geldeinheiten bewertbare Erfahrungen wie Schmerz oder Angst aufseiten der Behandelten

[2] Streng genommen bezeichnen krankheitsbedingte Nettoeinkommensverluste die Differenz aus dem Nettoeinkommen von Gesunden und dem Nettoeinkommen von Erkrankten unter Berücksichtigung von Zuzahlungen für Gesundheitsleistungen zur Behandlung der Erkrankung. Im Rahmen der Perspektive der GKV-Versichertengemeinschaft werden Zuzahlungen jedoch als nicht erstattungsfähige Kosten berücksichtigt, sodass die Nettoeinkommensverluste aus der Differenz zwischen dem gezahlten Krankengeld und Nettoeinkommen eines Gesunden ermittelt werden können.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

bezeichnet. Nach internationalen Standards der Gesundheitsökonomie sollten sie auf der Nutzenseite berichtet werden, soweit Daten für diese Angaben vorhanden sind.

E) Zukünftige Kosten

Weiterhin wird in der gesundheitsökonomischen Literatur häufig eine Unterscheidung vorgeschlagen zwischen interventionsassoziierten und nicht interventionsassoziierten (zukünftigen) Kosten. Interventionsassoziierte Kosten sind z. B. die nach einem Herzinfarkt anfallenden Arzneimittel und Nachkontrollen, wohingegen nicht interventionsassoziierte Kosten z. B. die Behandlungskosten eines Jahre später auftretenden Karzinoms wären, dessen Behandlung mit der des Herzinfarkts nichts zu tun hat.

Die Berücksichtigung von nicht interventionsassoziierten Kosten wird kontrovers diskutiert [68,159,207,371]. Interventions- und nicht interventionsassoziierte Kosten werden auftragsspezifisch gegeneinander abgegrenzt. Ist die Verlängerung des Lebens für die KNB relevant, werden im Basisfall die interventionsassoziierten zukünftigen Kosten berücksichtigt (sowohl bei gegebener Lebenserwartung als auch bei gewonnenen Lebensjahren). Nicht interventionsassoziierte zukünftige Kosten können in separaten Sensitivitätsanalysen erfasst werden (nicht bei gegebener Lebenserwartung, da identisch für alle Strategien, wohl aber bei gewonnenen Lebensjahren).

F) Investitions- und Implementierungskosten

Wenn explizit für die GKV oder die GKV-Versichertengemeinschaft einmalige Kosten zur Finanzierung der Bereitstellung oder Umsetzung entstehen, sind die Investitions- und Implementierungskosten angemessen zu berücksichtigen. Dies sollte über Sensitivitätsanalysen untersucht werden.

4.4.3 Schritte zur Kostenbestimmung

Grundsätzlich sind die Kosten so präzise wie möglich zu bestimmen. Methoden, verwendete Quellen und Ergebnisse sind für die einzelnen Schritte der Kostenbestimmung zu beschreiben. Die Bestimmung der in das Modell eingehenden Kosten folgt üblicherweise einem vierstufigen Prozess:

- Identifikation der Ressourcen,
- Mengenerfassung der Ressourcen,
- Bewertung der Ressourcen und
- Berechnung der in das Modell einfließenden Kosten nach Gesundheitszuständen und ggf. Zyklen.

A) Identifikation der Ressourcen

Im Rahmen der Identifizierung der Ressourcen müssen die Gesundheitsleistungen bestimmt werden, die zur Behandlung der Erkrankung angewendet werden (siehe Abschnitt 4.1.5). Die

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Angaben sollten möglichst aktuell sein und können aus den in Abschnitt 4.4.4 beschriebenen Quellen gezogen werden.

B) Mengenerfassung der Ressourcen

Die Anwendungshäufigkeit, der Anteil der relevanten Patientenpopulation, die die jeweilige Leistung in Anspruch genommen hat, und die Dauer der Inanspruchnahme müssen bestimmt werden. Kosten für Leistungen, die sehr selten erbracht werden und / oder nur einen geringen Einfluss auf die Ergebnisse ausüben, sollten zwar beschrieben werden, werden aber nicht zwangsläufig in die Berechnung einbezogen [159].

Um den Ressourcenkonsum zu quantifizieren, können sowohl der Mikro- als auch der Makro(Gross)-Costing-Ansatz [543,544] verwendet und kombiniert werden. Der Präzisionsgrad der Mengenerfassung wird also unter anderem durch das Vergütungssystem und den entsprechenden Aggregationsgrad der Leistungen bestimmt.

Beide Ansätze können im Bottom-up-Ansatz oder als Top-down-Ansatz [494,543,544] angewendet werden, wenn entweder ausgehend von den einzelnen Patienten die verbrauchten Ressourcen gemessen werden oder ausgehend von hoch aggregierten Daten (Ausgaben für eine Krankheit) eine (durchschnittliche) Aufteilung auf Patienten vorgenommen wird.

C) Bewertung der Ressourcen

GKV-Versichertenperspektive

Regulierte und verhandelte Preise (also Preise, die sich nicht ausschließlich über Marktmechanismen entwickelt haben) bestimmen im Allgemeinen die Ausgaben und repräsentieren die Opportunitätskosten der GKV-Versichertengemeinschaft. Wie zuvor beschrieben, bestimmt das Vergütungssystem den maximalen Präzisionsgrad bei der Ausgabenbestimmung der erstattungsfähigen Kosten. So repräsentieren aus der Perspektive der GKV-Versichertengemeinschaft DRGs die bestmögliche Bewertung für den stationären Sektor und der einheitliche Bewertungsmaßstab (EBM) die bestmögliche Bewertung für den ambulanten Sektor.

Bei der Kostenbestimmung für Arzneimittel ist zwischen dem stationären und ambulanten Bereich zu unterscheiden. Im Bereich der stationären Versorgung sind die Arzneimittel in der Regel Teil der entsprechenden pauschalen Vergütung. Wenn für relevante Arzneimittel Zusatzentgelte (ZE) vereinbart wurden oder diese über neue Untersuchungs- und Behandlungsmethoden (NUB) abgerechnet werden können, sind diese Kosten entsprechend zu ermitteln und in der KNB anzusetzen. Im ambulanten Bereich werden zunächst die Apothekenabgabepreise zur Grundlage genommen. Sofern Festbeträge vorhanden sind, müssen diese angegeben werden. Diese werden um Apotheken- und Herstellerrabatt reduziert. Kassenspezifische Rabatte werden nicht abgebildet. Grundsätzlich werden dem Prinzip der Effizienz folgend die günstigsten Vertreter eines Wirkstoffs bzw. einer Wirkstoffklasse ausgewählt. Relevante Preisänderungen im Zeitverlauf müssen berücksichtigt werden.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Nicht erstattungsfähige Kosten sind teilweise reguliert, sodass hier auf entsprechende Standardisierung bei der Ressourcenbewertung zurückgegriffen werden kann (z. B. Zuzahlungsregelungen im stationären Bereich und für Arzneimittel). Diese Kosten werden in der GKV-Versichertenperspektive gesondert ausgewiesen.

Besonderheiten in weiteren Perspektiven

In der Sozialversicherungsperspektive sind je nach Versicherungszweig unter Umständen nur aggregierte Daten verfügbar. In diesem Fall sollten die Ressourcen mittels Top-down-Ansatz aus den jeweiligen Statistiken bewertet werden.

Bei der Berechnung der Kosten aus der gesellschaftlichen Perspektive wäre theoretisch zu beachten, dass die gesellschaftlichen Opportunitätskosten sich in der Regel von den administrativen Preisen unterscheiden, weil die administrativen Preise nur die Perspektive des Bezahlenden darstellen. So sind z. B. in den Fallpauschalen nicht die Kosten für den Bau von Krankenhäusern enthalten, die man jeder Fallpauschale aus gesellschaftlicher Perspektive zuschlagen müsste. Das Institut ist sich dieser theoretischen Diskussion bewusst, richtet sich aber nach den internationalen Standards anderer Health-Technology-Assessment (HTA)-Organisationen, die auch in der gesellschaftlichen Perspektive administrative Preise nutzen, da ein anderes Vorgehen – aufgrund fehlender Daten, z. B. zu den tatsächlichen Kosten, die für den Bau von Krankenhäusern den Fallpauschalen zugeschlagen werden müssten – mit hoher Unsicherheit behaftet wäre. International üblich in gesundheitsökonomischen Evaluationen ist zumeist nur die zusätzliche Betrachtung der indirekten Kosten. Wird der Zeitaufwand von Betroffenen oder Angehörigen in der Kostenbestimmung berücksichtigt, wird dieser mit dem Nettolohn bewertet.

Bewertung von indirekten Kosten

Für Produktivitätsverluste berücksichtigt das Institut im Basisfall den Friktionskostenansatz [225,334], da der Humankapitalansatz auf einigen unrealistischen Annahmen (insbesondere Vollbeschäftigung am Arbeitsmarkt) beruht. In Sensitivitätsanalysen kann diese Schätzung dem Humankapitalansatz gegenübergestellt werden.

In der KNB orientiert sich die Bewertung von indirekten Kosten an den individuellen Arbeitskosten (d. h. Bruttolohnsatz und Lohnnebenkosten – in Deutschland Arbeitgeberbeiträge zur Sozialversicherung) oder den durchschnittlichen Arbeitskosten. Die Berechnung der durchschnittlichen Arbeitskosten pro Arbeitstag basiert auf den gewichteten durchschnittlichen Arbeitskosten von in Vollzeit und Teilzeit erwerbstätigen Personen in Deutschland. Näherungsweise kann hierzu das „Arbeitnehmerentgelt in Deutschland pro Jahr“ dividiert durch die „Anzahl der Arbeitnehmer mal 365“ herangezogen werden (wobei dann Sonn- und Feiertage bei den Arbeitsunfähigkeitstagen berücksichtigt werden müssen). Die Anwendung auf Selbstständige ist zu diskutieren [220]. Die Friktionskosten werden mit 80 % der Lohnkosten angenommen (analog zu den Niederlanden [334]). Die Friktionsperiode wird, sofern keine aktuellen Daten verfügbar sind, in Anlehnung an die durchschnittliche tatsächliche Besetzungsdauer in Deutschland für das Jahr 2012, mit 82 Tagen angesetzt [58].

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Sollte der Humankapitalansatz in einer Sensitivitätsanalyse untersucht werden, werden die zukünftigen Produktivitätsverluste auf Basis des durchschnittlichen Alters der Patienten bis zum Erreichen der Regelaltersgrenze berechnet.

D) Darstellung der in das Modell einfließenden Kosten nach Zuständen oder Zyklen

Bevor die Kosten in das Modell eingespeist werden können, müssen sie als durchschnittliche Kosten je Patient nach Gesundheitszuständen und je nach Modell auch nach Zyklen vorliegen.

Je nach Indikation, Intervention, Endpunkten und Modell liegt unter Umständen keine direkte Information zu den Kosten der jeweiligen Gesundheitszustände im Modell vor. Dann können die durchschnittlichen Kosten einer Intervention je Patient und Kostenkategorie (Leistungsbereiche und indirekte Kosten) für den Betrachtungszeitraum durch Annahmen aus weiteren Quellen (siehe Abschnitt 4.4.4) auf die verschiedenen Gesundheitszustände und Zyklen des Modells verteilt werden.

Für absorbierende Zustände in einem Markov-Modell kann es notwendig sein, Übergangskosten zu berechnen, die nur einmalig beim Übergang in diesen Gesundheitszustand anfallen. Dies ist dann zu empfehlen, wenn anzunehmen ist, dass die Kosten in diesem Zustand im ersten Zyklus wesentlich höher ausfallen als in den Folgezyklen.

4.4.4 Datengrundlage

Kosten, die in das Modell einzuspeisen sind, müssen, wie oben beschrieben, für die verschiedenen Gesundheitszustände und ggf. Zyklen eines Modells berechnet werden. Das Vorgehen bei der Datenerhebung und -auswertung, alle Berechnungen und die Ergebnisse sind transparent darzustellen.

Zur Identifikation und zur Mengenerfassung der Ressourcen können Informationen aus 3 Arten von Quellen gewonnen werden: Sekundärdaten (vor allem aus GKV-Routinedaten), Leitlinien und Expertenmeinungen. Zur Preisbestimmung stützt sich das Institut auf die jeweils relevanten regulierten oder verhandelten Preise, z. B. aus der Lauer-Taxe, dem EBM, dem DRG-Katalog oder Statistiken aus der Rentenversicherung oder des Statistischen Bundesamts.

Sekundärdaten in Form von ausgewerteten GKV-Routinedaten anhand einer repräsentativen Stichprobe sind hierbei die Datenquelle erster Wahl. Wenn aktuelle Analysen nicht der Literatur entnommen werden können, sind bevorzugt eigene Auswertungen durchzuführen.

Ergänzend können Leitlinien oder Ergebnisse aus Expertenbefragungen hinzugezogen werden, wenn Routinedaten nicht hinreichend die Versorgung in allen Zuständen des Modells darstellen. Bevorzugt sollten evidenzbasierte Leitlinien aus dem deutschen Gesundheitssystem verwendet werden (siehe Abschnitt 4.1.7). Sind diese in dem zu untersuchenden Indikationsgebiet nicht vorhanden, ist abzuwägen und transparent darzulegen, ob andere deutsche Leitlinien verwendet werden können oder ob auf Expertenbefragungen zurück-

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

gegriffen werden soll. Expertenbefragungen sind nur dann eine Option, wenn die Daten nicht repräsentativeren Quellen entnommen werden können bzw. wenn Letztere den in den Gesundheitszuständen erforderlichen Detailgrad nicht vollständig abdecken (siehe auch Abschnitt 4.1.7).

Eine Übertragbarkeit von Versorgungspfaden und Kostendaten aus anderen Gesundheitssystemen ist aufgrund von Systemunterschieden selten gegeben und nur unter sehr strengen Voraussetzungen möglich [354,511]. Die Übertragbarkeit von Kostendaten aus folgenden Ländern wird nicht grundsätzlich ausgeschlossen, da sie ein dem deutschen ähnliches System von ambulanter und stationärer Versorgung haben: Österreich, Schweiz, Niederlande, Belgien, Frankreich. Eine Verwendung muss aber jeweils begründet und diskutiert werden. Kostendaten aus weiteren Ländern dürfen in einer KNB nicht verwendet werden.

4.4.5 Unsicherheit und Verteilung von Kostendaten

Die Unsicherheit in Kostendaten sollte angemessen adressiert werden. Kostendaten sind inhärent stetig, positiv, ohne Obergrenze und im Allgemeinen nicht normalverteilt, sondern für gewöhnlich rechtsschief verteilt [159].

4.4.6 Inflationsbereinigung und Diskontierung

A) Inflationsbereinigung

Wenn Kostendaten aus unterschiedlichen Zeitperioden stammen, muss inflationsbereinigt werden. Als Quelle für die jährliche Inflation soll der Harmonisierte Verbraucherpreisindex (HVPI) des Statistischen Bundesamts verwendet werden [525]. Im Rahmen einer Sensitivitätsanalyse können weitere Preissteigerungsraten für einzelne Bereiche der Versorgung, z. B. Arzneimittel, aus anderen Quellen einbezogen werden.

B) Diskontierung

Wenn Kosten und Nutzen in Perioden anfallen, die länger als ein Jahr dauern, werden sie nach dem ersten Jahr im Basisfall mit einer identischen konstanten Rate von 3 % auf die aktuelle Periode diskontiert [32,89,129,159,304]. In Sensitivitätsanalysen sollten ebenfalls identische konstante Raten von 0 und 5 % verwendet werden. Abweichungen hiervon müssen begründet werden.

4.5 Epidemiologische Daten

4.5.1 Daten

Für eine KNB sind aktuelle epidemiologische Daten unerlässlich. Daten zu Prävalenz und Inzidenz in Deutschland dienen neben der Abschätzung der Krankheitslast auch dazu, im Rahmen der Ausgaben-Einfluss-Analyse die Veränderungen im Budget der GKV zu quantifizieren. Daher sind Aussagen nötig, ob in den nächsten 5 Jahren Veränderungen in der Inzidenz, Prävalenz oder Mortalität zu erwarten sind. Weiterhin sind Daten zur Mortalität von

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Bedeutung, um krankheitsbedingtes Versterben und auch die sog. Hintergrundmortalität abzubilden.

Eine besondere Rolle für die Modellierung spielen die Basiswahrscheinlichkeiten für Ereignisse. In einem Modell sind für jeden Endpunkt Angaben zu den endpunktbezogenen Ereignishäufigkeiten bzw. -wahrscheinlichkeiten nötig, die als Ausgangswerte in das entscheidungsanalytische Modell eingehen.

4.5.2 Datengrundlage

Epidemiologische Daten können aus Sekundärdaten wie öffentlichen Datensammlungen und GKV-Routinedaten (siehe Abschnitt 4.4.4) sowie Registerdaten und ggf. wissenschaftlichen Publikationen kommen (siehe Abschnitt 4.1.7). Öffentliche Datensammlungen, beispielsweise vom Robert Koch-Institut, sind aufgrund der hohen methodischen Konsistenz primär heranzuziehen, wenn vorhanden und in geeigneter Form (beispielsweise passende Altersklassen) erhältlich. Eine Sonderstellung nehmen Registerdaten ein. Unabhängig von der Einschätzung der Qualität eines Registers haben diese Daten oft nur einen regionalen Bezug. Daher muss die Übertragbarkeit geprüft werden. Sind wissenschaftliche Publikationen vorhanden, in denen epidemiologische Kennziffern bestimmt wurden, können diese unter Umständen direkt verwendet werden. Die Verwendbarkeit muss im Einzelfall geklärt werden, da die Studien oft methodisch verschiedene Ansätze verwenden. Kohortenstudien bzw. hinreichend große und repräsentative Stichproben sind zu bevorzugen. Die methodische Qualität der zugrunde liegenden Studie kann u.a. anhand der Vorgaben der „gute[n] epidemiologischen Praxis“ bewertet werden.

4.5.3 Unsicherheit und Verteilung von epidemiologischen Daten

Die Unsicherheit in epidemiologischen Daten sollte angemessen adressiert werden. Insbesondere die Unsicherheit von Daten zum Basisrisiko und zur Mortalität muss in den Sensitivitätsanalysen ausreichend und den Verteilungen angemessen berücksichtigt werden.

4.6 Ergebnisdarstellung als Effizienzgrenze

Als Basis der ökonomischen Bewertung von Interventionen innerhalb eines Indikationsgebiets wird eine Effizienzgrenze gezeichnet. Sie wird aus den effizientesten Interventionen der verfügbaren Komparatoren erzeugt und kann der Ableitung von Handlungsempfehlungen für die zu prüfende(n) Intervention(en) dienen. Sie kann Informationen zur Verhandlung von Erstattungsbeträgen liefern, ohne auf einen in Deutschland zurzeit nicht konsentierten Schwellenwert für die Zahlungsbereitschaft zu rekurrieren.

4.6.1 Definition

Die Effizienzgrenze stellt den endpunktbezogenen Nutzen der verfügbaren Interventionen innerhalb eines Indikationsgebiets den Nettokosten dieser Interventionen grafisch gegenüber. Dabei wird, falls erforderlich, der Nutzen in ein approximativ kardinalskaliertes Maß

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

überführt[3]. Diejenigen Interventionen, die nach Nutzen und Kosten die effizientesten sind, bilden die Effizienzgrenze.

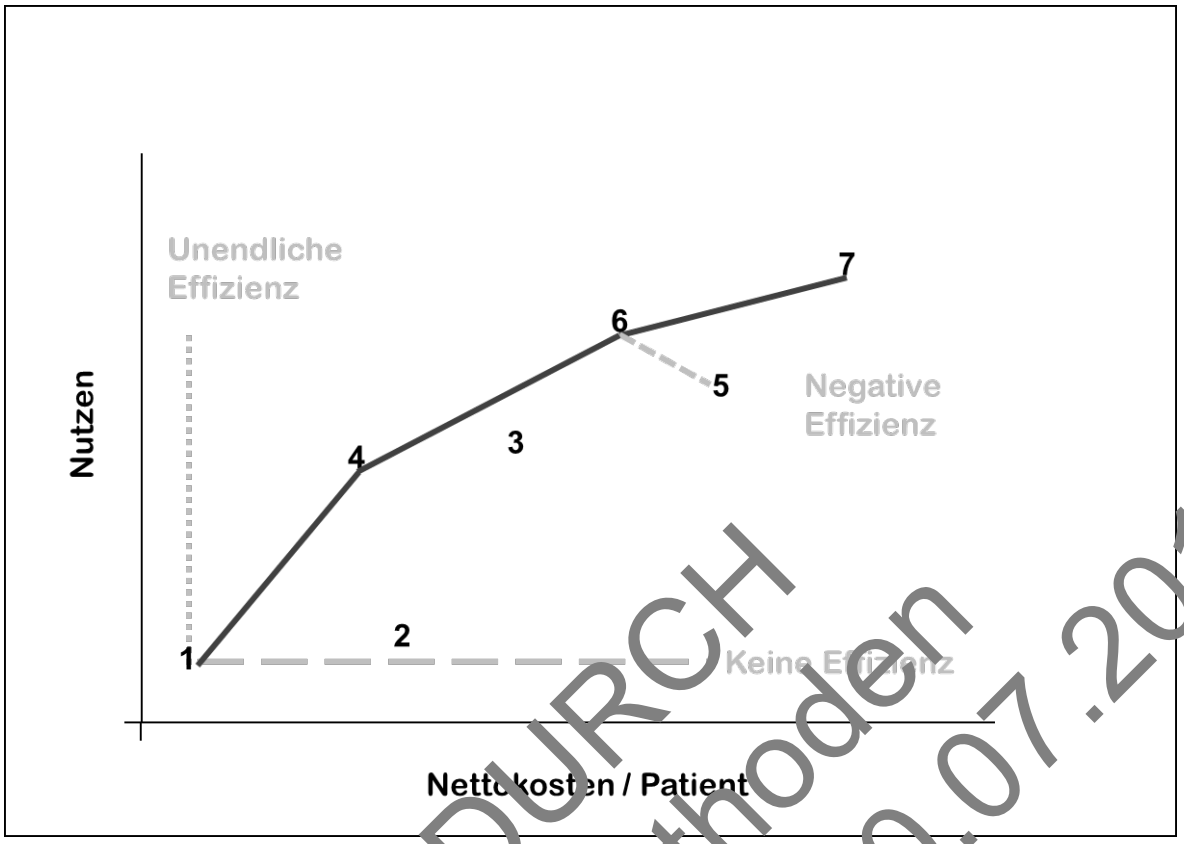
4.6.2 Verfahrensverlauf

Im Verfahren muss unterschieden werden zwischen der/den zu bewertenden Intervention(en) und den Interventionen, die die Effizienzgrenze bilden (Komparatoren). Letztere sind Interventionen, die derzeit in Deutschland in dem zur Bewertung anstehenden Indikationsgebiet angewendet und erstattet werden. Deren Kosten und Nutzen werden ermittelt und grafisch dargestellt. Erstere sind die zu prüfenden Interventionen.

In der Darstellung der Effizienzgrenze werden von links nach rechts die Interventionen mit jeweils höherer Effizienz aufgetragen. Die Steigung der theoretischen Verbindungslinie zwischen 2 Interventionen (das Liniensegment) gibt den zusätzlichen Nutzen pro zusätzlichen Kosten an (siehe Abbildung 10).

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

[3] Hat der aus der Nutzenbewertung ermittelte patientenrelevante Zusatznutzen bereits approximativ kardinalskalierte Eigenschaften, kann er direkt in die KNB überführt werden.

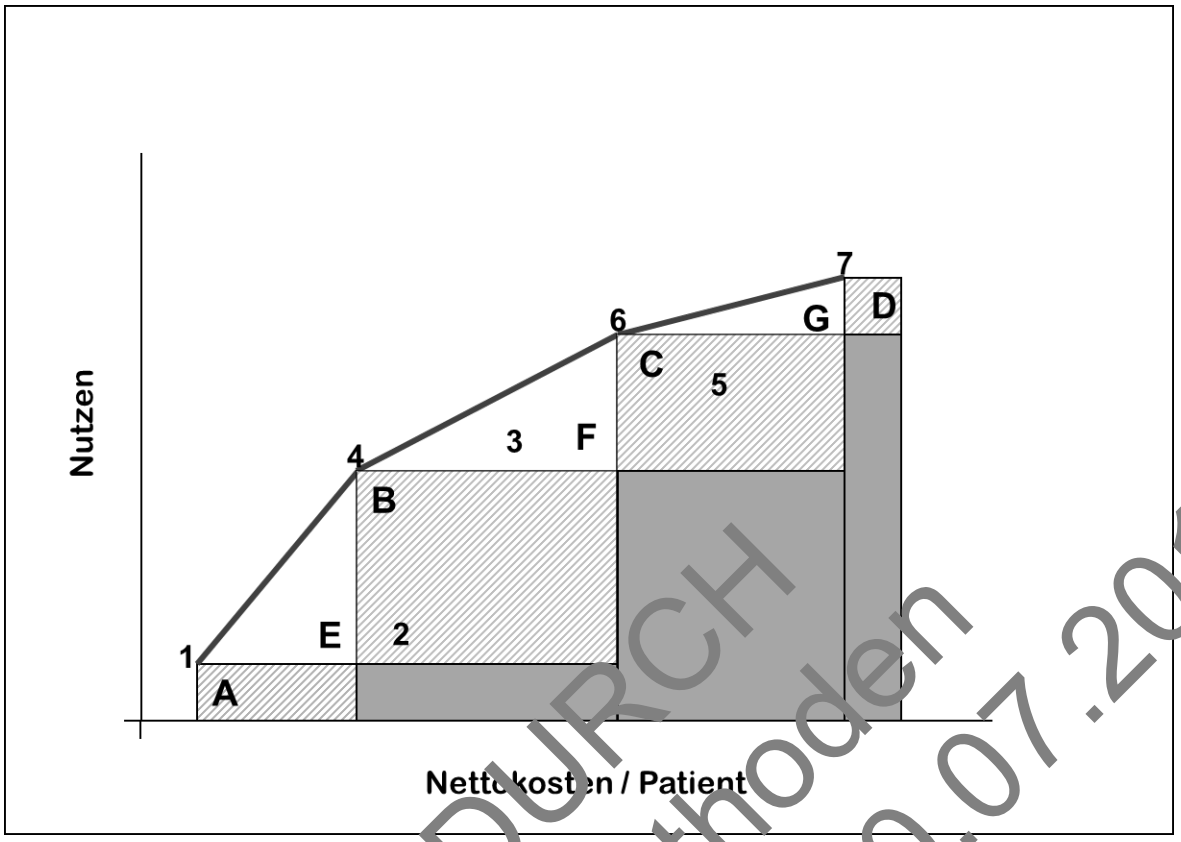


Eine Horizontale (Steigungswinkel = 0°) bedeutet keine Effizienz, während eine Vertikale (Steigungswinkel = 90°) für unendliche Effizienz steht. Eine in aufsteigender Reihenfolge positive Steigung (z. B. zwischen Intervention 6 und 7) stellt einen zusätzlichen Nutzen bei erhöhten Kosten dar, während eine negative Steigung (z. B. zwischen Intervention 6 und 5) einen geringeren Nutzen bei höheren Kosten bedeutet.

Abbildung 10: Interpretation der Steigung der theoretischen Effizienzgrenze

Die Positionen von Interventionen wie beispielsweise Intervention 3 in Abbildung 10 erfordern eine weitergehende Interpretation, weil sie keine negative Effizienz im Vergleich zu bereits eingeführten Interventionen (z. B. Intervention 4) haben. In Abbildung 11 ist die Fläche unter der theoretischen Effizienzgrenze durch eine Reihe von Rechtecken (A bis D) aufgeteilt. Jedes dieser Rechtecke enthält alle Interventionen, die eine negative Effizienz (höhere Kosten bei geringerem Nutzen) gegenüber mindestens einer bereits im Markt befindlichen Intervention auf der theoretischen Effizienzgrenze haben. Interventionen in diesen Teilbereichen (z. B. Intervention 2 bzw. Intervention 5 in Abbildung 11) sind eindeutig ineffizient. Es verbleiben die Dreiecke E, F und G, in welchen die Interventionen nicht eindeutig ineffizient sind. Üblicherweise sind Interventionen, die in diesen Dreiecken eingetragen werden, kein Bestandteil der Effizienzgrenze, weil die theoretische Kombination der beiden Interventionen, die die Hypotenuse des Dreiecks bilden, einen höheren Nutzen bei geringeren Kosten bietet (sogenannte erweiterte Dominanz).

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"



Die theoretische Effizienzgrenze (durchgezogene Linie) verbindet diejenigen Interventionen, die in Bezug zu jeder anderen Intervention bzw. anderen Kombinationen effizient sind. Interventionen in den Rechtecken A bis D (z. B. Intervention 2 bzw. Intervention 5) sind eindeutig ineffizient. Intervention 3 befindet sich in einer der verbleibenden dreieckigen Flächen (E bis G) und ist nicht eindeutig ineffizient. Theoretisch würde sich durch die Kombination von Intervention 4 und Intervention 6 eine erweiterte Dominanz ergeben, jedoch kann sich diese in der Praxis als nicht durchführbar erweisen.

Abbildung 11: Absolute versus erweiterte Dominanz

Eine solche Kombination ist in der Praxis nicht immer möglich. Dies würde nämlich voraussetzen, dass bei einem fixen Preis der Intervention 3 die Leistungsempfänger auf Intervention 4 und 6 umverteilt werden müssten, um eine höhere Effizienz zu erreichen. Dies kann klinisch unerwünscht sein und schwierig zu rechtfertigen, weil es zu einer Schlechterstellung derjenigen führen würde, die Intervention 4 erhalten. Die Alternative, einen Wechsel der Leistungsempfänger zwischen beiden Therapien über die Zeit hinweg zu ermöglichen, ist eindeutig bei den meisten chirurgischen und vermutlich auch bei vielen medikamentösen Interventionen nicht möglich. So kann es viele Situationen geben, in welchen Interventionen, die sich in den dreieckigen Flächen befinden, einen Teil der praktischen Effizienzgrenze ausmachen. Wird das Kriterium der erweiterten Dominanz nicht angewendet, resultiert eine stufenförmige absolute Effizienzgrenze, welche sich aus der Verbindung der oberen Segmente der schraffierten Rechtecke gegenüber den dreieckigen Flächen ergibt. Hierbei ist jedoch zu beachten, dass die absolute Effizienzgrenze keine Steigung mehr im Sinne eines Kehrwerts der Zahlungsbereitschaft wiedergibt und somit keine Schwellenwerte zu erheben wären.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

4.6.3 Konstruktion der Effizienzgrenze

Die Effizienzgrenze wird so konstruiert, dass sie die relevanten Interventionen in einem vorgegebenen Indikationsgebiet darstellt. Dazu gehört:

- Vollständige und detaillierte Festlegung des Indikationsgebiets, das von Interesse ist. Dies kann die genaue Erkrankung, die Behandlungsgegebenheiten (z. B. stationäre Versorgung), die Zielpopulation, die Therapiesequenz (erste, zweite Therapiewahl etc.) sowie die Angabe über eine Mono- oder Kombinationstherapie beinhalten.
- Positionierung der vorhandenen Therapien anhand ihres Nutzens und ihrer Kosten
- Eintragen der Interventionen in ein Koordinatensystem mit dem Nutzen auf der vertikalen (y-)Achse und den Kosten auf der horizontalen (x-)Achse.[4] Hierbei ist gemäß guter wissenschaftlicher Praxis auf eine (zumindest je Endpunkt) gleichbleibende Skalierung der Achsen zu achten.
- Auftragen der Effizienzgrenze

Bei der Evaluation neuer Interventionen werden anschließend zusätzlich deren gesundheitliche Effekte und Kosten in dem betreffenden Indikationsgebiet dargestellt.

A) Vertikale Achse

- Auf der vertikalen Achse wird der Nutzen und Schaden abgetragen. Hierbei ist auf einen positiven Wertebereich zu achten, sodass die Effizienzgrenze den vergrößerten Nutzen bzw. verringerten Schaden abbildet (ggf. ist z. B. eine Multiplikation mit „-1“ erforderlich oder die Umrechnung auf das Komplementärereignis „1 - Schaden“).
- Der Nutzen bzw. Schaden wird anhand von patientenrelevanten Endpunkten dargestellt, die geeignet operationalisiert werden müssen (z. B. Lebensqualitätsscores).
- Der Nutzen bzw. Schaden wird auf die vertikale Achse übertragen. Diese Übertragung kann unter Einbeziehung von Modellierungen erfolgen.

B) Horizontale Achse

- Auf der horizontalen Achse werden die gesamten Nettokosten pro Patient eingetragen.
- Die Kosten werden regelhaft aus der Perspektive der GKV-Versichertengemeinschaft berechnet und können auftragsabhängig zusätzliche Kosten aus ggf. erweiterten Perspektiven (z. B. Sozialversicherungsperspektive, gesellschaftliche Perspektive) beinhalten.
- Als Kosten werden die gegenwärtig zu erwartenden Kosten verwendet.

[4] Eine tabellarische Darstellung ist auch möglich. Die Zusammenhänge werden aber dadurch nicht so anschaulich.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Um die Kosten für jede Intervention abzuschätzen und im Koordinatensystem zur Effizienzgrenze einzutragen, müssen mehrere Voraussetzungen eingehalten werden. Die Kosten sollten demjenigen Betrag entsprechen, der in der Praxis anfallen würde. Auf der Effizienzgrenze müssen die Gesamtnettokosten pro Patient abgetragen werden.

Um das Kosten-Nutzen-Verhältnis von (neuen) Interventionen mit mehr Nutzen und mehr Kosten als die Komparatoren zu bestimmen, wird das letzte Segment der Effizienzgrenze verlängert (siehe Abschnitt 4.1.9 sowie Abbildung 9 und Abbildung 12).

In Abhängigkeit von der Anzahl der Endpunkte, die aus der vorab durchgeführten Nutzenbewertung übernommen werden, können mehrere Effizienzgrenzen hergeleitet und präsentiert werden.[5] Sofern eine Endpunktgewichtung vorgenommen wurde, wird diese ebenfalls vorgelegt. Wenn ein Maß des Gesamtnutzens festgelegt wurde, ist dieses als primäres Ergebnis anzusehen.

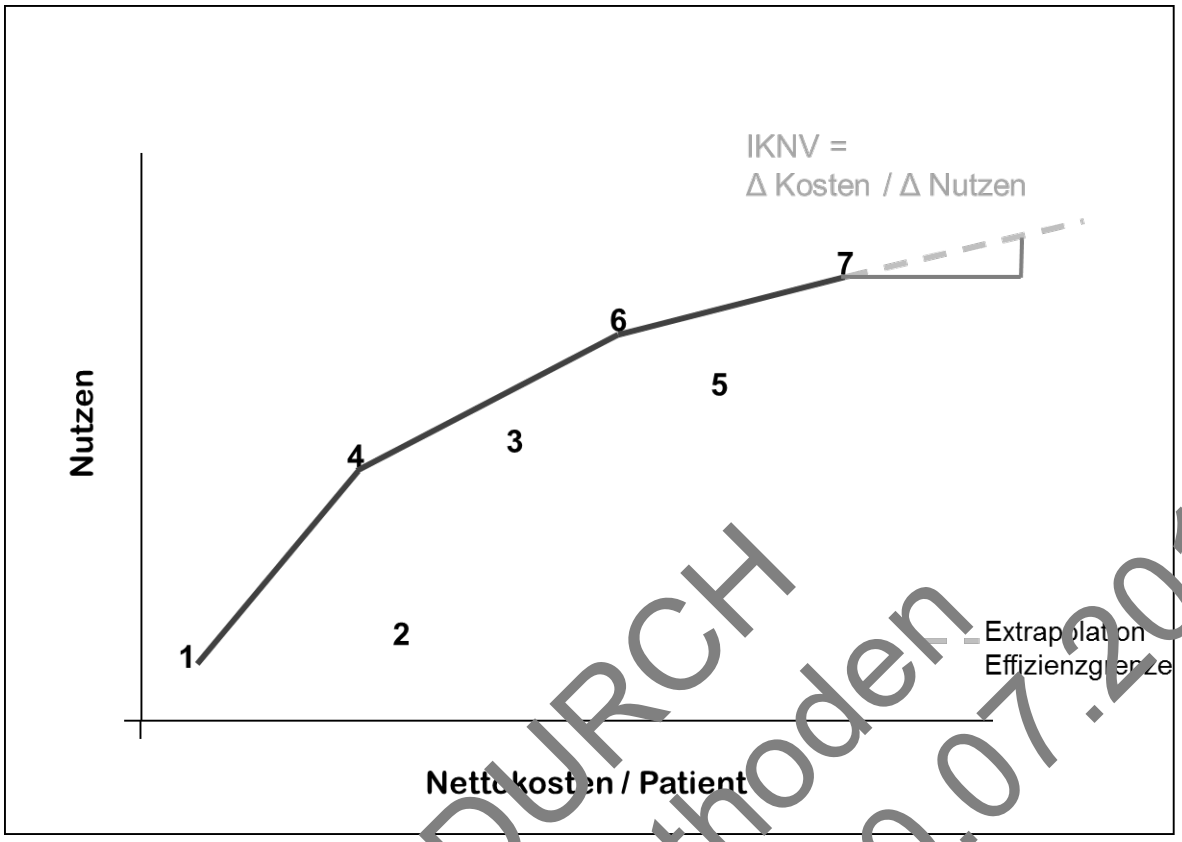
C) Definition des Ursprungs des Koordinatensystems

Der Punkt „keine Intervention“ (d. h. der natürliche Verlauf) erfordert ebenfalls eine Bewertung. Obwohl er möglicherweise als Koordinatenursprung angesehen werden könnte (null Nutzen, null Kosten), ist dies selten angemessen, da auch eine nicht durchgeführte Intervention noch Kosten und Gesundheitseffekte verursachen kann, beispielsweise infolge der nicht behandelten Erkrankung, der Überwachung usw. Daher sollten auch Daten für den natürlichen Verlauf erhoben werden. Eine gängige Annahme ist hierbei, dass Placebo am ehesten dem natürlichen Verlauf entspricht. Dies ist auftragsbezogen zu prüfen.

Wenn der Ursprung der Effizienzgrenze nicht dem Nullpunkt entspricht, müssen die Effizienzgrenzen (zumindest je Endpunkt) in gleich skalierte Koordinatensysteme eingetragen werden. Die am weitesten unten und links liegende Intervention wird im Allgemeinen zum Ursprung der Effizienzgrenze (siehe Abbildung 12). Eine Verschiebung des Nullpunkts (des Koordinatensystems) ist aus Gründen der Vergleichbarkeit der Darstellung verschiedener Effizienzgrenzen abzulehnen.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

[5] Dies bezieht sich auch auf die gesonderte Darstellung von divergierenden Schadenaspekten in Abgrenzung zum patientenrelevanten Zusatznutzen.



Die Effizienzgrenze beginnt in einem vom Nullpunkt des Koordinatensystems verschiedenen Ursprung. Die Verlängerung zeigt das inkrementelle Kosten-Nutzen-Verhältnis (IKNV), an dem eine (neue) Intervention mit mehr Nutzen und mehr Kosten als die Komparatoren gemessen wird.

Abbildung 12: Darstellung der Effizienzgrenze

4.6.4 Sonderkonstellationen

Es gibt 2 Sonderkonstellationen, in denen sich trotz vollständiger Information nicht unmittelbar eine Empfehlung anhand der Effizienzgrenze für eine neue Intervention ableiten lässt:

1) Die letzte Intervention auf der Effizienzgrenze dominiert alle anderen Interventionen und verursacht dieselben Kosten wie das Referenzszenario. Die Steigung wäre damit unendlich (siehe Darstellung in Abbildung 10).
2) Die letzte Intervention auf der Effizienzgrenze vor der Einführung der Innovation ist kostengünstiger und hat mehr Nutzen als alle Komparatoren einschließlich des Ursprungs.

In beiden Fällen würde sich ein neuer Ursprung ergeben, auf dem die jeweils letzte Intervention vor Einführung der innovativen Intervention liegen würde.

Hier kann die Ausgaben-Einfluss-Analyse weitere Daten liefern, indem die Auswirkungen auf das Budget dargestellt werden (siehe Abschnitt 4.1.9 und Abschnitt 4.8).

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

4.7 Unsicherheit (Sensitivitätsanalysen)

Die Arten der Unsicherheit wurden oben (siehe Abschnitt 4.1.8) dargestellt. Die Unsicherheit vieler Modellparameter ergibt sich daraus, dass ihr Wert aus Stichproben geschätzt wird. Diese Art von Unsicherheit wird oft durch Konfidenzintervalle oder andere statistische Ansätze zur Beschreibung von Variabilität erfasst.

4.7.1 Quantifizierung von Unsicherheit

Bei den Kosten kann Unsicherheit über Annahmen zum Ressourcenverbrauch bestehen, z. B. Dosierung eines Arzneimittels über die Zeit. Auch kann das Modell stochastisch ausgelegt sein (es verwendet Zufallszahlen bei den Monte-Carlo-Ziehungen). Um diese Art von Unsicherheit einzugrenzen, können verschiedene Techniken angewendet werden [349,456,508].

Auch aus der in Abschnitt 4.2 beschriebenen Form von möglicher Variabilität in der Modellstruktur resultiert Unsicherheit, die bei der Untersuchung berücksichtigt werden muss. Zuletzt können selbst zuvor festgesetzte Inputparameter wie die Diskontierungsrate variiert werden, um die Unsicherheit abzubilden, die sich aus unterschiedlichen Diskontierungsraten ergibt (siehe Abschnitt 4.4.6).

4.7.2 Sensitivitätsanalysen

Parameterunsicherheit sowie andere Arten der Unsicherheit, die nicht reduzierbar sind, werden quantifiziert. Das Institut berücksichtigt sowohl univariate und multivariate deterministische als auch probabilistische Sensitivitätsanalysen und legt bei der Bearbeitung die Empfehlungen der gemeinsamen Modeling Good Research Practices Task Force Working Group der ISPOR and SMDM zugrunde [63].

Alle dazu unternommenen Analysen sollten mit minimalen und maximalen Werten für die eingesetzten Parameterwerte und zugrunde liegenden Annahmen vollständig dokumentiert werden. Für probabilistische Sensitivitätsanalysen (PSAs) müssen folgende Aspekte präzisiert werden: eingesetzte Wahrscheinlichkeitsverteilungen und ihre Quellen, Korrelationen zwischen Inputparametern und jeglichen strukturellen Varianten.

Strukturelle Sensitivitätsanalysen werden durchgeführt, um den Einfluss einer Variierung der Annahmen in der Modellstruktur, z. B. Anzahl oder Art der Modellzustände, zu eruieren.

Darstellung der Ergebnisse der Sensitivitätsanalysen

Für die deterministische Sensitivitätsanalyse sind extreme Ausprägungen der Inputparameter anzugeben, für die die neue Intervention möglicherweise kostensparend ist bzw. oberhalb oder unterhalb der Effizienzgrenze liegt. Für univariate und multivariate Analysen müssen die Ergebnisse tabellarisch und in einem Tornado-Diagramm dargestellt werden, in dem für die entsprechenden Intervalle der Inputparameter die Ausprägungen der Ergebnisse als Intervall dargestellt werden.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Für die PSAs wird der Anteil der Simulationen in Prozent angegeben, für die sich eine Kostenersparnis bzw. eine Lage oberhalb oder unterhalb der Effizienzgrenze ergibt. Im Falle von PSAs werden die Ergebnisse als kumulative Kostenverteilungen präsentiert.

4.7.3 Darstellung von Unsicherheit mittels des Net Health Benefit

Bei der Darstellung der Ergebnisse der Sensitivitätsanalysen ist zu beachten, dass die Berücksichtigung der Parameterunsicherheit zum einen die Lage mehrerer oder aller Interventionen verändern kann, die die Effizienzgrenze bilden. Zum anderen kann sich auch die Lage der zu prüfenden Intervention verändern, die dieser Effizienzgrenze gegenübergestellt wird.

Der Net Health Benefit (NHB) ist ein etabliertes Verfahren zur Darstellung von Ergebnissen aus PSAs [532]. Durch die NHB-Berechnung wird diesem Problem Rechnung getragen, da der NHB eine Funktion sowohl von Zusatznutzen und -kosten als auch der Effizienzgrenze ist und die Position der zu bewertenden Intervention als Abstand zu der sich verlagernden Effizienzgrenze bzw. zu dem sich verlagernden letzten Segment der Effizienzgrenze abbildet. Daher sollten sowohl die Basisfallanalysen als auch die deterministischen und die probabilistischen Sensitivitätsanalysen basierend auf dem Konzept der NHB-Berechnung durchgeführt werden.

4.8 Ausgaben-Einfluss-Analyse (Budget-Impact-Analyse)

Eine Ausgaben-Einfluss-Analyse (AEA) ist eine Bewertung der direkten finanziellen Konsequenzen, die mit der Erstattung einer Intervention in einem Gesundheitssystem in Zusammenhang stehen [558]. In einem Berechnungsmodell für eine AEA werden der Anteil der Patienten, die die neue Intervention möglicherweise erhalten werden, sowie die Verbreitung der Intervention im Gesundheitssystem einschließlich ihrer Anwendung bei zuvor unbehandelten Patienten berücksichtigt. Eine AEA prognostiziert insbesondere, wie eine Veränderung im Mix der Interventionen für eine bestimmte Krankheit die Ausgaben für ein Indikationsgebiet zukünftig beeinflussen kann [386].

Zweck einer AEA ist es weniger, die finanziellen Konsequenzen einer Intervention exakt abzuschätzen. Vielmehr soll ein zuverlässiger Berechnungsrahmen zur Verfügung gestellt werden, der dem Entscheidungsträger ermöglicht, die möglichen Ausgabeneffekte einer neuen Intervention (oder der geänderten Nutzung von bereits vorhandenen Interventionen) zu verstehen [386]. Ein derartiges Modell ist erforderlich, weil viele der Ausgangsgrößen je nach Konstellation variieren können und darüber hinaus mit Unsicherheit behaftet sind. Daher ist das Ergebnis der AEA nicht ein einzelner Wert für die Ausgabenabschätzung, sondern vielmehr ein Wertebereich, der aus dem Modell resultiert.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

4.8.1 Perspektive in der Ausgaben-Einfluss-Analyse

Die AEA sollte aus der Perspektive der GKV oder eines anderen relevanten Ausgabenträgers durchgeführt werden (siehe auch Abschnitt 4.4.1). Jegliche außerhalb dieser Perspektive anfallenden Ausgaben oder erzielten Einsparungen werden nicht einbezogen.

4.8.2 Zeithorizont in der Ausgaben-Einfluss-Analyse

Die AEA sollte den Zeithorizont umfassen, der für Kostenträger die höchste Relevanz für ihre Ausgaben hat [386]. Da sich der Einfluss auf das Ausgabengeschehen nach der Einführung der neuen Intervention im Laufe der Zeit wahrscheinlich verändert – sowohl aufgrund der Marktanpassung als auch aufgrund längerfristiger Auswirkungen auf die betreffende Krankheit –, sollte dieser für eine Periode von 1 und 3 Jahren abgeschätzt und dargeboten werden [385]. Das Ergebnis muss als Ausgaben und Einsparungen pro Jahr dargestellt werden anstatt in Form eines einzigen „gegenwärtigen Nettowerts“ [386]. Daher darf in diesem Fall keine Diskontierung der Finanzflüsse vorgenommen werden. Sollte das Ergebnis als Gesamtkostenbetrag für 3 Jahre dargestellt werden, können die Kosten entsprechend diskontiert werden (siehe Abschnitt 4.4.6).

4.8.3 Szenarien in der Ausgaben-Einfluss-Analyse

Eine AEA vergleicht Versorgungsszenarien – jedes definiert durch eine Zusammenstellung von Interventionen – anstatt spezifische einzelne Interventionen [386]. Es müssen mindestens 2 Szenarien berücksichtigt werden: zum einen das Referenzszenario, definiert durch die aktuelle Kombination aus Interventionen, zum anderen die prognostizierte neue Kombination aus Interventionen.

4.8.4 Population in der Ausgaben-Einfluss-Analyse

Die Anzahl der Versicherten, die für die neue Intervention infrage kommen, ist einer der Schlüsselfaktoren zur Bestimmung der zu erwartenden Ausgaben für die neue Intervention. Die voraussichtliche Anzahl der Empfänger ergibt sich aus der prognostizierten Inanspruchnahme der Intervention innerhalb der Zielpopulation. Jede erwartete Off-Label-Anwendung der neuen Intervention sollte nicht in der primären AEA, sondern kann im Rahmen von Sensitivitätsanalysen berücksichtigt werden [426]. Bei der Vorhersage der Zahl der Anwender müssen sowohl die Substitution vorhandener Interventionen als auch der induzierte Bedarf berücksichtigt werden.

4.8.5 Einzuschließende Kosten in der Ausgaben-Einfluss-Analyse

Die Kosten (Nettokosten, d. h. um Kosteneinsparungen, sogenannte Cost-Offsets, bereinigt) sollten entsprechend den in Abschnitt 4.4 beschriebenen Methoden abgeschätzt werden.

Für die AEA werden Investitions- und Implementierungskosten – soweit möglich und von der GKV getragen – identifiziert und quantifiziert. Sie sollten separat und nach Kostenkategorien

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

geordnet dargestellt werden, wobei eine vollständige Erläuterung der Methode und der zur Kosteneinschätzung verwendeten Quellen enthalten sein muss.

4.8.6 Ergebnisdarstellung in der Ausgaben-Einfluss-Analyse

Die Ergebnisse (in €) sollten als Wertebereich präsentiert werden und nicht als einzelner Punktschätzer. Weiterhin sollten sowohl ein Gesamtbetrag als auch ein Anteil an den Jahresausgaben ausgewiesen werden.

4.9 Besonderheiten der Kosten-Nutzen-Bewertung nach § 35b SGB V

4.9.1 Gesetzliche Anforderungen und Verfahrensablauf

Für die KNB nach § 35b SGB V gelten einige besondere Anforderungen. Standardmäßig gibt es im Rahmen der Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V 2 Konstellationen, die zu einer KNB führen können:

1) Ist ein pharmazeutischer Unternehmer mit dem Beschluss des G-BA nicht einverstanden, dass das zu bewertende Arzneimittel keinen Zusatznutzen hat oder keine therapeutische Verbesserung darstellt, kann der pharmazeutische Unternehmer nach § 35a Absatz 5a SGB V verlangen, dass der G-BA eine KNB nach § 35b SGB V oder nach § 139a Absatz 3 Nummer 5 SGB V in Auftrag gibt.
2) Nach einem Schiedsspruch der Schiedsstelle können gemäß § 130b Absatz 8 SGB V sowohl der pharmazeutische Unternehmer als auch der GKV-Spitzenverband eine KNB nach § 35b SGB V beantragen.

Sollte ein pharmazeutischer Unternehmer und / oder der GKV-Spitzenverband beim G-BA eine KNB gemäß § 35b SGB V beantragen, ergeben sich weitere Besonderheiten im Ablauf des Verfahrens, die in Abschnitt 2.1.4 beschrieben werden.

Nach § 130b Absatz 8 Satz 3 SGB V dient eine KNB von Arzneimitteln nach § 35b SGB V dem Zweck der Verhandlung eines Erstattungsbetrags, der im Vergleich zu (einer) zweckmäßigen Vergleichstherapie(n) verhandelt werden soll. Nach § 35b SGB V legt der G-BA in seinem Auftrag einer KNB fest:

- zweckmäßige Vergleichstherapie und andere Arzneimittel und Behandlungsformen, mit denen das zu bewertende Arzneimittel verglichen werden soll,
- Patientengruppen,
- Zeitraum,
- Art von Nutzen und Kosten und
- Maß des Gesamtnutzens.

Basis für die KNB sind 1) die Ergebnisse klinischer Studien, 2) die Ergebnisse der Versorgungsstudien, die mit dem G-BA vereinbart wurden oder die der G-BA auf Antrag des

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

pharmazeutischen Unternehmers anerkennt, und 3) die Nachweise des pharmazeutischen Unternehmers (vgl. § 35b Absatz 1 Satz 3 SGB V). Aufgrund der Gesetzeslage in Deutschland (§ 35b Absatz 1 SGB V) wird zudem regelhaft die Perspektive der GKV-Versichertengemeinschaft eingenommen. Näheres beschreibt die Verfahrensordnung des G-BA [211].

4.9.2 Der Net Health Benefit zur Berechnung zusatznutzenbasierter Erstattungspreise

Wie in Abschnitt 4.7.3 erläutert, kann der NHB zur Darstellung der Unsicherheit herangezogen werden. Auf Basis des Erwartungswerts des NHB der zu bewertenden Intervention kann über die weitere Berechnung des kostenbereinigten (Zusatz-)Nutzens der zu bewertenden Intervention auch ein zusatznutzenbasierter Erstattungspreis abgeleitet werden [533].

Der inkrementelle NHB wird mithilfe der Effektschätzer für den Nutzen und der Kosten der jeweiligen Interventionen sowie eines Schwellenwerts berechnet. In dieser Anwendung entspricht der Schwellenwert dem Kehrwert der Steigung des letzten (und gegebenenfalls extrapolierten) Segments der Effizienzgrenze für kosteneffektive Interventionen (siehe Abbildung 13). Wäre der NHB in etwa gleich 0, so läge Intervention 8 auf der Effizienzgrenze, die durch die Steigung (1/λ) des letzten Segments der Effizienzgrenze bestimmt wird, und kann im Vergleich zu den die Effizienzgrenze bildenden (per Definition kosteneffektiven) Interventionen ebenfalls als kosteneffektiv bewertet werden. Entsprechend wird ein zusatznutzenbasierter Erstattungspreis mittels des NHB durch Umformung und Berechnung der maximalen Interventionskosten ermittelt, die notwendig sind, um zu gewährleisten, dass der NHB mindestens gleich 0 ist. Praktisch kann der NHB mithilfe des Modells durch iterative Berechnungen geschätzt werden.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

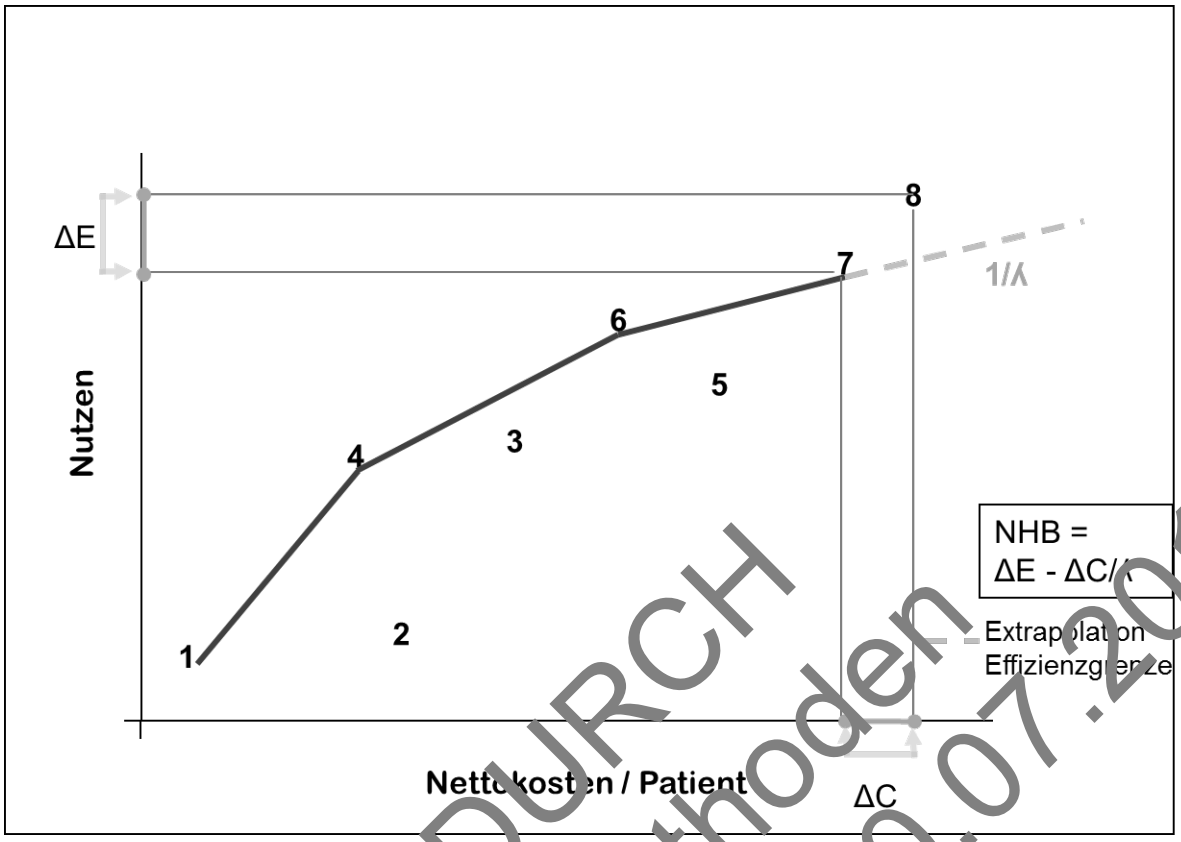


Abbildung 13: Darstellung eines NHB > 0

4.9.3 Sensitivitätsanalysen zur Berechnung zusatznutzenbasierter Erstattungspreise

Für den zusatznutzenbasierten Erstattungspreis können je Effizienzgrenze Preis-Akzeptanzkurven [187] und / oder NHB-Werte dargestellt werden (siehe Abschnitt 4.9.2).

Bei Verwendung des NHB sollten die Ergebnisse der PSAs über die Berechnung und Mittelung der jeweiligen NHB-Erwartungswerte für die zu prüfende Intervention für eine hinreichend große Zahl von Durchläufen dargestellt werden. Bei jedem Durchlauf können sich sowohl die Effizienzgrenze als auch die Lage der zu prüfenden Intervention relativ zu der Effizienzgrenze und somit der jeweilige NHB-Wert verändern. Hieraus lassen sich der gemittelte NHB-Wert der zu prüfenden Intervention sowie eine Interquartilsregion berechnen (siehe Abschnitt 4.9.4). Der NHB-Erwartungswert in Kombination mit der Interquartilsregion gibt an, wie groß erwartungsgemäß der kostenbereinigte (Zusatz-)Nutzen unter Berücksichtigung der Modellunsicherheit beim aktuellen zusatznutzenbasierten Erstattungspreis ist.

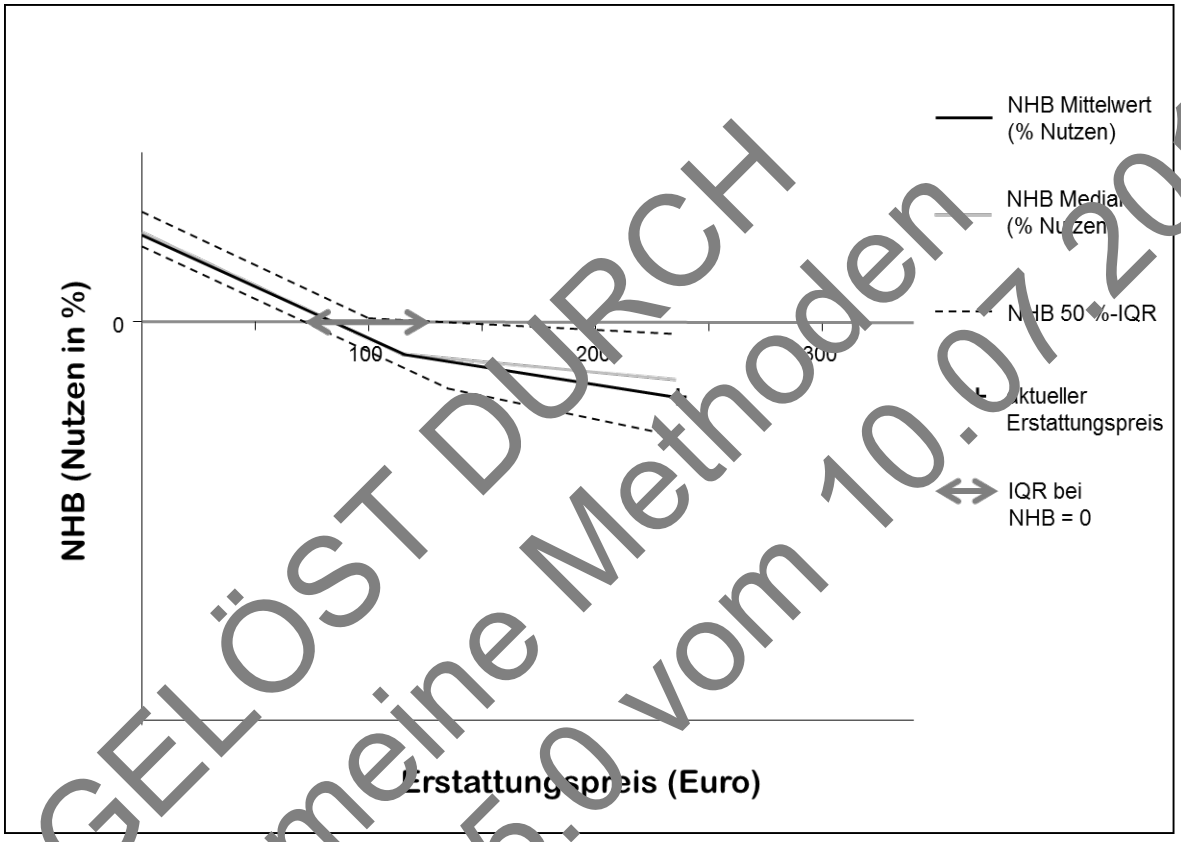
4.9.4 Interquartilsregion als Maß der Streuung für Preisverhandlungen

Um dem Spitzenverband Bund der Krankenkassen (kurz: GKV-Spitzenverband) und dem pharmazeutischen Unternehmer auf Basis der Ergebnisse der Sensitivitätsanalysen (siehe Abschnitt 4.7) ein Maß der Streuung für die Verhandlungen zu geben, wird eine Interquartilsregion (IQR) angegeben. Die IQR umschließt alle Werte des NHB aus den Simulationen,

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

die vom unteren und vom oberen Quartil begrenzt werden (siehe Abschnitt 4.9.3). Das heißt, dass die Interquartilsregion die 50 % der Simulationen in den PSAs abdeckt, die oberhalb der 25 % niedrigsten Ergebnisse und unterhalb der 25 % höchsten Ergebnisse liegen (siehe Abbildung 14). Grundsätzlich kann auch die Angabe von anderen Regionen mit anderen Maßen sinnvoll sein.

Die IQR erlaubt es, unter Berücksichtigung der gesamten Unsicherheit (umgesetzt durch PSAs) einen Spielraum für mögliche Verhandlungen um Erstattungsbeträge zu eröffnen, in dessen Grenzen auch die Unsicherheit der Effektschätzer und der Kosten berücksichtigt wird.



Die durchgezogene Linie gibt für jeden möglichen Erstattungspreis (x-Achse) den im Mittel zu erwartenden NHB an. An der Stelle, an der die durchgezogene Linie die x-Achse kreuzt, kann ein zusatznutzenbasierter Erstattungspreis abgelesen werden, bei dem der im Mittel zu erwartende NHB 0 ist, d. h. weder positiv noch negativ.

Abbildung 14: Interquartilsregion möglicher zusatznutzenbasierter Erstattungspreise (basierend auf PSA) als Maß der Streuung für Preisverhandlungen

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

5 Leitlinien und Versorgungsanalyse

5.1 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. [191,221] 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 5.3). Ebenso können durch den Abgleich dieser Standards mit konkreten Versorgungsstrukturen, -prozessen und -ergebnissen Versorgungslücken und Verbesserungspotenziale aufgedeckt werden (siehe Abschnitt 5.4). 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.

5.2 Identifikation von Versorgungsstandards über Leitlinien

5.2.1 Versorgungsstandards in Leitlinien

Ärztliches Handeln, das nach medizinisch-wissenschaftlicher Erkenntnis und / oder ärztlicher Erfahrung in der Profession akzeptiert ist, stellt den medizinischen Standard dar [248]. Eine Form, einen medizinischen Standard wissenschaftlich und institutionell festzusetzen, ist die Leitlinie.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 [4,296]). 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 [557].

5.2.2 Methodische Bewertung von Leitlinien

Die Informationsbeschaffung erfolgt entsprechend den in Kapitel 7 beschriebenen Vorgehensweisen.

International werden verschiedene Instrumente zur methodischen Leitlinienbewertung eingesetzt [577]. Das von einem Netzwerk von Forschern und Gesundheitspolitikern entwickelte und validierte AGREE-Instrument (Appraisal of Guidelines Research and Evaluation in Europe) [4,374] und dessen Weiterentwicklung in Form des AGREE-II-Instruments [5,72-74] 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-II-Instrument [5] enthält 23 Einzelfragen (Items), die in 6 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 2 Reviewer unabhängig voneinander durchgeführt.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 [5].

5.2.3 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 [24,161,235,341,499]. 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

entsprechenden Risiken einer Empfehlung beinhalten [24,235,499]. 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 konsentierte 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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). 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.

5.3 Empfehlungen zu Disease-Management Programmen

Die über das in Abschnitt 5.2 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.

5.4 Versorgungsanalyse

5.4.1 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 [248]. Dieser medizinische Standard wird als Versorgungsstandard bezeichnet. Er kann durch Gesetze, Verordnungen und Richtlinien vorgegeben sein oder in Leitlinien identifiziert werden (siehe

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

auch Abschnitt 5.2). Auch die Referenzwerte von Qualitätsindikatoren können als Versorgungsstandards interpretiert werden [190].

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 [249,297]. Durch den Abgleich von Soll- und Istzustand 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.

5.4.2 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 2 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 [439]. Zur Bewertung der Versorgungsqualität wird die Versorgungssituation mit einem normativen Standard, dem Versorgungsstandard, abgeglichen, sofern ein solcher vorhanden ist.

5.4.3 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 [472] 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.

5.4.4 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),

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

- 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.

5.4.5 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 7).

Tabelle 7: 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. Säuglingssterblichkeit / 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

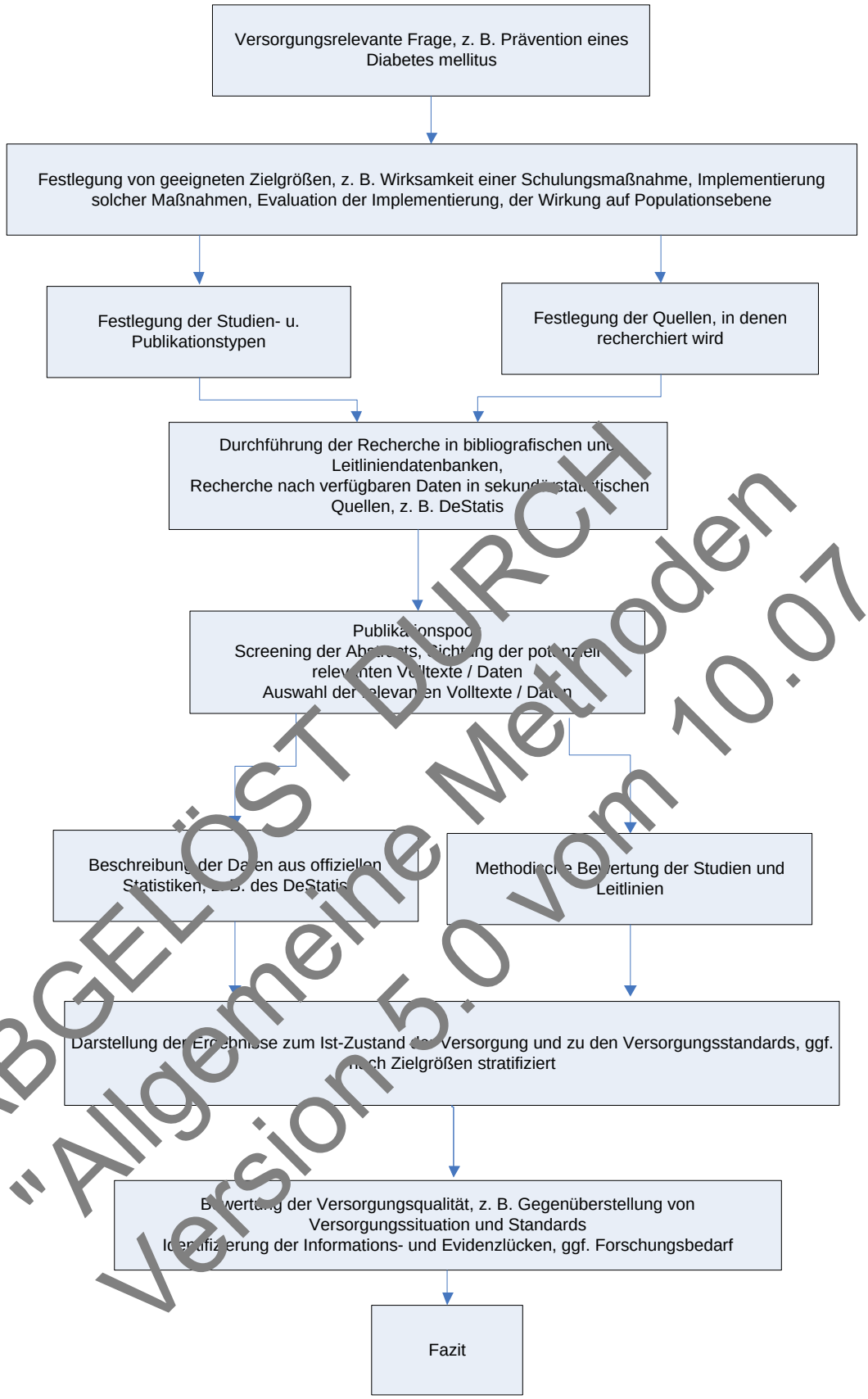
Sie geben Auskunft über die Häufigkeit der Erkrankung [346]. Anhand der Case-Fatality-Rate kann der Schweregrad einer Krankheit eingeschätzt werden [256]. Die Folgen einer Erkrankung können anhand der Daten gemäß der International Classification of Functioning (ICF) und der Daten der Rentenversicherung, z. B. zu Erwerbsminderungsrenten, beurteilt werden [133,526]. 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 [320,325,587].

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 5.4.5 beschrieben – identifiziert.

5.4.6 Ablauf einer Versorgungsanalyse

Der Ablauf einer Versorgungsanalyse wird in Abbildung 15 beispielhaft dargestellt.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"



ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Abbildung 15: Exemplarischer Ablauf einer Versorgungsanalyse

5.4.7 Ebenen der Versorgungsanalyse

Die Beschreibung der Versorgung anhand der o. g. Parameter kann bezogen auf 3 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) [260,346]. Zielgrößen auf dieser Ebene sind ebenfalls patientenrelevante Endpunkte wie Mortalität und Morbidität allerdings mit dem Bezugspunkt allgemeine Bevölkerung [346]. 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 [21,125]. 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 3 Ebenen (Individualmedizin, Bevölkerungsmedizin und Gesundheitssystem) beziehen. Auch können auf allen Ebenen zeitliche Entwicklungen und regionale Variationen (Disparitäten) untersucht werden [131]; dazu können u. a. auch geografische Informationssysteme zum Einsatz kommen.

5.4.8 Methodische Besonderheiten der Versorgungsanalyse

Mit Bezug auf die Komplexität des Gesundheitssystems und die o. g. Ebenen (siehe Abschnitte 5.4.5 und 5.4.7) 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 5.4.9) und mit verschiedenen Methoden bearbeitet. Die methodische Bewertung erfolgt – soweit möglich – mit geeigneten Instrumenten (siehe Abschnitt 5.4.9).

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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

5.4.9 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 7.1).

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 5.2 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 8 gibt einen Überblick über mögliche Quellen zur Identifizierung von Versorgungsstandards.

Tabelle 8: 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	BMJV, 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 7.1) 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) im Berichtsplan verbindlich festgelegt. Im Folgenden

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

werden mögliche Datenquellen zur Identifizierung von Versorgungsdaten benannt (siehe Tabelle 9).

Tabelle 9: Informationsquellen zur Identifizierung von Versorgungsdaten

Informationen zu	**Beispiele für Quellen**
Morbidität und Mortalität, z. B. Inzidenzen und Prävalenzen (Populationsebene)	▪ 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)	▪ Versorgungsstudien
Inanspruchnahme- und Verordnungsverhalten	▪ 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	▪ Schiedsstellen der Landesärztekammern ▪ Qualitätsindikatoren der OECD ▪ weitere Publikationen der GKV
Messen der Versorgungsqualität mit Indikatoren ▪ Qualität der Versorgung auf Systemebene ▪ Qualität in der ambulanten ärztlichen Versorgung ▪ Qualität in der stationären Versorgung ▪ Qualität in der Pflege	 ▪ 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	▪ Evaluationsberichte von DMPs
Gesundheitssystem / Gesundheitssystem-vergleich	▪ z. B. Publikationen der WHO (z. B. World Health Report)

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 7.1). 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.

5.4.10 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 [213], 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 [77,431]. 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 [431,563].
- Abschließend 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 5.2.3 beschrieben.

5.4.11 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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

5.4.12 Darstellung und Bewertung der Versorgungsqualität

Die Bewertung der Versorgungsqualität besteht aus 3 Schritten:

Schritt 1: Beschreibung des Istzustands

Die Beschreibung des Istzustands 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 Sollzustands

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 Sollzustand

Danach wird der Istzustand 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 Sollzustand? 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.

5.5 Validität von Leitlinienempfehlungen

5.5.1 Hintergrund

Obwohl die Methodik zur Leitlinienentwicklung zunehmend weiterentwickelt wird [18,233], weisen Leitlinien immer noch erhebliche Unterschiede in ihrem Erstellungsprozess, ihrer Darstellung, ihrer methodischen Qualität und nicht zuletzt in ihren Inhalten auf

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

[83,84,92,269,281,384,388,392]. Eine hohe methodische Qualität von Leitlinien korreliert zudem nicht notwendigerweise mit der inhaltlichen Qualität der darin enthaltenen Empfehlungen [582]. Für viele Fragestellungen reicht die Überprüfung der methodischen Qualität einer Leitlinie daher nicht aus, um die Wertigkeit einzelner Empfehlungen einzuschätzen [233]. Hieraus ergibt sich die Notwendigkeit, dass Inhalte von Leitlinien, insbesondere in Bezug auf die Validität einzelner Empfehlungen, analysiert und überprüft werden müssen.

5.5.2 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).

5.5.3 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

5.5.4 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.

5.5.5 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 lang wirksamen 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

6 Evidenzbasierte Gesundheitsinformation für Bürgerinnen und Bürger

6.1 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 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 [35,104,143,322,336]. Die 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 [475] und ist damit auch ein wesentlicher Bestandteil evidenzbasierter Gesundheitsinformationen. Evidenzbasierte Gesundheitsinformationen müssen darüber hinaus neben Informationen über den Nutzen und Schaden auch Informationen über die Unsicherheiten und Ungewissheiten enthalten, die bezüglich der beschriebenen medizinischen Interventionen bestehen [113,139,169,482,527]. Weitere Anforderungen sind, dass

- der Inhalt auf klaren wissenschaftlichen Belegen (strenger Evidenz aus systematischen Übersichten) basiert,
- die 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 medizinischer Maßnahmen verständlich dargestellt werden,

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

- die 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.

6.2 Patientenzentrierte Kommunikation

6.2.1 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. Zudem sollen die Gesundheitsinformationen des IQWiG eine heterogene Zielgruppe verständlich ansprechen, deren Mitglieder sich unter anderem durch Gesundheitskompetenz, Erkrankungen, Bildungsniveau, persönlichen Hintergrund, Alter und Geschlecht unterscheiden.

Das primäre Mittel einer Qualitätsbewertung im Hinblick auf die Verständlichkeit von Texten für die Zielgruppe ist die Begutachtung von Entwürfen durch Testleserinnen und Testleser sowie Leserbewertungen, die möglichst diese Unterschiede abbilden (siehe unten).

Die Evidenz zu erklären und bei der Vermittlung gesundheitsbezogener Informationen objektiv zu bleiben, stellt eine zusätzliche Herausforderung dar [170,322,498,556]. Zur Gewährleistung eines objektiven und nicht direktiven Vorgehens wird in den Gesundheitsinformationen ausgewogen dargestellt, was als wissenschaftlich bekannt gilt. Es werden keine Empfehlungen ausgesprochen. Diesem Anspruch wird durch eine nicht direktive Informationsaufbereitung und eine neutrale Sprache Rechnung getragen.

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 weitere Formate 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,
- ein Risiko als absolutes Risiko, ggf. ergänzt um weitere Informationen wie das relative Risiko, anzugeben, wenn es verlässliche und entscheidungsrelevante Daten, und

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

- verzerrte Informationen zugunsten oder zuungunsten der Produkte eines bestimmten Unternehmens zu vermeiden, indem Handelsnamen nur so weit verwendet werden, wie es für das Verständnis unerlässlich ist.

Evidenz aus Screeningverfahren zeigt, dass die Angabe von absoluten Risikoschätzern für Patientinnen und Patienten zur persönlichen Orientierung bei ihren Entscheidungen hilfreich sein kann [163]. Wenn es verlässliche Zahlen gibt, die Menschen helfen können, ihr Risiko abzuschätzen, werden diese Informationen dargestellt. 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 [98,167] – eine Ungewissheit bestehen bleibt [420], 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. Bei der Entwicklung von Entscheidungshilfen erfolgt eine Orientierung an den International Patient Decision Aid Standards (IPDAS) [167,270].

Menschen zu informieren ist nicht der einzige Zweck von Gesundheitskommunikation. Ein besonderes Anliegen ist es auch, den Nutzerinnen und Nutzern emotionale Unterstützung zu bieten [186]; ferner können Gesundheitsinformationen 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 [419]. Diese Fähigkeiten bilden die Voraussetzung für autonomes Denken und Handeln auf seiten 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 [145,221,575].

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,
- 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

durchgehende Verwendung maskuliner Personenbezeichnungen (generisches Maskulinum) führt zu einer gedanklichen Unterrepräsentation von Frauen, worin eine sprachliche Benachteiligung von Frauen zu sehen ist [305]. In den Texten der Gesundheitsinformationen wird daher ein weitestgehend geschlechterneutraler Sprachstil verwendet, 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 wurde für die Gesundheitsinformationen ein Leitfaden zur Textgestaltung sowie zu Sensibilisierungs- und anderen Maßnahmen für die redaktionellen Mitarbeiterinnen und Mitarbeiter entwickelt [408]. Dieser wird, basierend auf der Evaluation der Produkte sowie neuer Evidenz aus dem Bereich der evidenzbasierten Kommunikation, kontinuierlich weiterentwickelt.

6.2.2 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. Hierzu wird sich einer Methode bedient, die die möglichen Wege von Patientinnen und Patienten mit einer bestimmten Erkrankung nachzeichnet. Sie wird im Folgenden als „Methode der Patientenwege“ bezeichnet.

Patientenwege dienen der kurzen Zusammenfassung und grafischen Aufbereitung der verschiedenen sozialen, emotionalen, kognitiven und klinischen Dimensionen, die mit einer Krankheit einhergehen können. Die Methode lehnt sich an die medizinsoziologischen Modelle der „Illness Trajectory“ [110] und der „Patientenkarriere“ [212,340] sowie verschiedene Modelle der „Patient Journey“ [345] 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 [212,340]. Ein weiterer Ansatz ist die Krankheitsverlaufskurve oder „Illness Trajectory“ nach Corbin und Strauss [110].

Aus diesen verschiedenen Ansätzen wurde die Methode der Patientenwege abgeleitet. Auch wenn Patientinnen und Patienten zu Expertinnen und Experten des Lebens mit einer

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

bestimmten Erkrankung werden und damit in gewisser Weise eine „Karriere“ beschreiten können, wird der Begriff „Patientenwege“ bevorzugt. Dieser Begriff erfasst möglichst viele der potenziellen Verläufe von Patientinnen und Patienten. Bei der Erstellung 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 werden folgende Fragen gestellt:

- 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 10 angegebenen Dimensionen unterstützt.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Tabelle 10: 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
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
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 erstellt werden. So wird versucht, systematisch zu erfassen, 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 healthtalk.org-Datenbank [144], 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 [403,459].

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

6.2.3 Beteiligung von Bürgerinnen und Bürgern

Es gibt Hinweise darauf, dass die Beteiligung von Betroffenen an der Erstellung von Gesundheitsinformationen deren Relevanz steigern kann [415]. Die Orientierung an der Sichtweise und an den Informationsbedürfnissen der Bürgerinnen und Bürger zählt zu den Anforderungen an evidenzbasierte Gesundheitsinformationen [139] und ist ein zentraler Bestandteil bei der Erstellung von Gesundheitsinformationen [603]. Dabei kommen verschiedene Maßnahmen zum Einsatz. Im Rahmen der Themenpriorisierung und -fokussierung 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 6.3.1). Außerdem werden von Patientinnen und Patienten sowie deren Angehörigen Erfahrungsberichte erhoben, um diese individuellen Erfahrungen und Erlebnisse mit Krankheit und Pflegebedürftigkeit in ihren verschiedenen Dimensionen anderen Betroffenen beziehungsweise Interessierten als Ergänzung zu den Gesundheitsinformationen zur Verfügung zu stellen (siehe Abschnitt 6.4.3). Im Rahmen der Qualitätssicherung können Gutachten von Patientinnen und Patienten oder Patientenvertreterinnen und -vertretern zu Textentwürfen eingeholt werden. Eine weitere Maßnahme zur Einbindung der Öffentlichkeit ist das begrenzte Stellungnahmeverfahren. Dabei wird u. a. dem Kuratorium die Möglichkeit zur Kommentierung der Textentwürfe gegeben. Dem Kuratorium gehören auch Vertreterinnen und Vertreter relevanter Institutionen für die Wahrnehmung von Patienteninteressen und von Selbsthilfe-Organisationen chronisch kranker und behinderter Menschen an. In der Regel zeitgleich mit dem Stellungnahmeverfahren durchlaufen die Gesundheitsinformationen 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 die Möglichkeit, dem Herausgeber eine Rückmeldung zu geben. Die einzelnen Gesundheitsinformationen können online kommentiert werden. Weiterhin erfolgt eine kontinuierliche Nutzerbefragung auf der Website (siehe Abschnitt 6.2.4).

6.2.4 Visuelle Kommunikation und Multimedia

Ein Text allein ist möglicherweise weniger verständlich und einprägsam als Informationen, in denen die Erklärungen durch Bilder ergänzt werden [170,274,333,362,556]. Auch gesprochener Text kann zum besseren Verständnis beitragen [274,483]. Die verständlichste Kommunikationsform – vor allem für Menschen mit geringerem Bildungsstand – könnten Erläuterungen darstellen, in denen Text, Bild und Ton miteinander kombiniert sind [274]. Wo es sinnvoll erscheint, werden Texte entsprechend aufbereitet, 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

zu verstehen [162,362,498]. 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 [145,321,575]. 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 Kommentarfunktion zu den einzelnen Texten,
- Themenvorschläge und ein allgemeines Online-Kontaktformular,
- ständige Umfragen zur Benutzerfreundlichkeit der Website und
- gelegentliche Online-Umfragen zu spezifischen Gesundheitsthemen [326].

6.2.5 Informationszugang und Barrierefreiheit

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).

Es wird gewährleistet, dass die Website gesundheitsinformation.de die international anerkannten Zugänglichkeitsrichtlinien für Personen mit Behinderungen [583] sowie die deutsche BITV (barrierefreie Informationstechnik-Verordnung) erfüllt [78]. Außerdem wird sich um die entsprechende Evaluation und Optimierung der Nutzbarkeit der Website bemüht.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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.

Die Gesundheitsinformationen werden sowohl in deutscher als auch in englischer Sprache veröffentlicht. Die Sicherung der bestmöglichen Informationsqualität erfordert eine breite internationale Beteiligung. Die Herausgabe der Gesundheitsinformationen auch in englischer Sprache ermöglicht es, vom Feedback internationaler Wissenschaftlerinnen und Wissenschaftler, insbesondere auch von der Begutachtung durch die Autorinnen und Autoren systematischer Reviews, zu profitieren. Die Verfügbarkeit einer englischsprachigen Version erweitert zudem die internationalen Möglichkeiten zur Übersetzung der Texte in andere Sprachen.

6.3 Themenwahl, Informationsbeschaffung und wissenschaftliche Bewertung

6.3.1 Themenwahl

Die vom Institut erstellten Gesundheitsinformationen gehen hervor

- aus Direktaufträgen des G-BA oder des BMG zur Erstellung von Patienteninformationen,
- aus Zusammenfassungen der anderen Produkte des Instituts und Begleitinformationen zu diesen Produkten und
- aus der gesetzlichen Aufgabe des Instituts zur Bereitstellung von Gesundheitsinformationen für 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 [29,30].

Gemäß § 139a Abs. 3 Nr. 6 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 [210].

Eine breit akzeptierte Definition bzw. eine klar umrissene Zusammenstellung der „Krankheiten mit erheblicher epidemiologischer Bedeutung“ kann nicht ausgemacht werden. Eine praxisnahe Festlegung der epidemiologischen Bedeutung von Krankheiten muss sich auf

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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.

Es wird angestrebt, zu den Diagnose- oder Krankheitsgruppen Informationen zu verfassen, von denen mindestens ein Prozent der Bevölkerung dauerhaft (Prävalenz) oder bezogen auf den Zeitraum eines Jahres (Inzidenz) betroffen sind. Dieser Themenkatalog kann in begründeten Fällen auch erweitert werden, zum Beispiel in Abhängigkeit von Auftragsthemen des IQWiG. Zur Erstellung des Themenkatalogs wird primär der Versorgungsreport des Wissenschaftlichen Instituts der AOK (WIdO) herangezogen, der regelmäßig auf www.wido.de/vsreport.html veröffentlicht wird. Der jährlich aktualisierte Report beinhaltet auf der Basis von etwa 24 Millionen AOK-Versicherten Angaben zu Prävalenzen und Hospitalisierungsraten für die 1500 häufigsten Erkrankungen (gruppiert nach ICD10). Der Themenkatalog für die Gesundheitsinformationen wird regelmäßig überprüft und wenn notwendig angepasst. Der aktuelle Stand der Umsetzung ist auf der Website einsehbar.

Ergänzend können bei der Priorisierung der Themen noch weitere Quellen berücksichtigt werden, 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 Call-Center 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 [493],
- 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Zusätzlich werden bei der Priorisierung weitere wissenschaftliche, redaktionelle und nutzerbezogene Aspekte berücksichtigt. Eine Übersicht ist in Tabelle 11 dargestellt.

Tabelle 11: Mögliche Aspekte zur Priorisierung der Themen

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

6.3.2 Informationsbeschaffung für Gesundheitsinformationen

Die Gesundheitsinformationen stützen sich vorwiegend auf systematische Übersichten und qualitative Studien. Wenn ein Thema im Detail recherchiert werden soll, wird in der Regel nach folgenden Informationen gesucht, 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 8.4),
- 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Die Ergebnisse dieser ersten Auswertung von Patienten- und Informationsbedürfnissen vermitteln ein Bild davon, 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 können themenspezifisch auch die Patientenvertreterinnen und -vertreter befragt werden, um weitere Aspekte zu identifizieren und um zu erörtern, welche Relevanz den Rechercheergebnissen in Deutschland zukommt.

Die wissenschaftliche Grundlage einer Gesundheitsinformation bildet in der Regel die themenspezifische systematische Recherche nach systematischen Übersichten zu Nutzen und Schaden medizinischer Interventionen, wie beispielsweise diagnostische Maßnahmen, inklusive Früherkennung oder medikamentöse und nichtmedikamentöse Behandlungsoptionen. Die Recherchen schließen die Database of Abstracts of Reviews of Effects (Other Reviews) (DARE), die Cochrane Database of Systematic Reviews (Cochrane Reviews), die Health Technology Assessment Database (Technology Assessments) sowie MEDLINE ein, sind aber nicht darauf begrenzt. Berücksichtigt werden in erster Linie systematische Übersichten mit Suchen, die in den letzten 3 Jahren durchgeführt wurden [509,510]. Themenspezifisch kann dieser Zeitraum auch ausgeweitet werden.

Weiterhin werden gegebenenfalls zusätzliche Recherchen zu Ursachen, Verlauf, Prognose und Epidemiologie durchgeführt. 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.

In Ausnahmefällen und vorwiegend im Rahmen des Aktualisierungsprozesses wird eine Recherche nach Primärstudien in Betracht gezogen. Wenn es beispielsweise zu einem wichtigen Thema keine neuere Übersicht gibt, wird eine Aktualisierungsrecherche nach Primärstudien erwogen. 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 [254]. Ergänzend können weitere Datenbanken hinzugezogen werden.

6.2.3 Evidenzbewertung

Die Gesundheitsinformationen stützen sich zum großen Teil auf systematische Übersichten (siehe Abschnitt 8.2). Damit eine systematische Übersicht über die Wirkung einer Intervention für Gesundheitsinformationen verwendet werden kann, muss diese gewisse Mindestvoraussetzungen erfüllen, d. h., sie darf nach dem Oxman-Guyatt-Index nur geringe methodische Mängel aufweisen [309,428,430]. Um sich als Basis für eine Aussage zum Nutzen und Schaden einer medizinischen Intervention zu eignen, sollte eine systematische Übersicht mindestens 2 Studien umfassen, deren Qualität von den Autorinnen und Autoren der Übersicht als adäquat erachtet wird und die Daten zu mindestens einem patientenrelevanten Endpunkt enthalten. Berücksichtigt werden auch die Relevanz und die

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Übertragbarkeit der Evidenz, insbesondere in Bezug auf Geschlecht und Alter (siehe Abschnitt 8.4).

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,
- Qualität, Ausführlichkeit und Aktualität der Suche,
- Durchführung von Sensitivitätsanalysen und Umgang mit Heterogenität,
- Adressierung und Umgang mit ggf. vorhandenem Verzerrungspotenzial.

Als Quelle für die in den Gesundheitsinformationen 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 [310].

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 [216]. Bei der Bewertung von solchen systematischen Übersichten werden zur Orientierung die Kriterien des Oxford Centre for Evidence-Based Medicine sowie das Evidenzbewertungssystem der McMaster-Universität angewendet [96,254]. Die Methoden zur Bewertung qualitativer Studien werden in Abschnitt 8.4 beschrieben.

6.3.4 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 [108] und die das deutsche Grundsatzpapier „Gute Praxis Gesundheitsinformation“ [139] vorsieht. Die Evidenz nimmt exponentiell zu. Das gilt sowohl für Studien [34,539] als auch für systematische Übersichten [34,399]. Neue Evidenz kann dazu führen, dass existierende Übersichten als überholt bzw. nicht mehr aktuell gelten [203,480,510,578], oft kommt es allerdings vor, dass die ursprünglichen Schlussfolgerungen durch die neue Evidenz nicht grundlegend verändert, sondern eher noch bekräftigt werden [301,443,538].

Eine Studie über Leitlinienempfehlungen kam zu dem Schluss, dass nach 3 Jahren mehr als 90 %, nach 6 Jahren jedoch nur noch etwa 50 % der in Leitlinien abgegebenen Empfehlungen

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

aktuell sein dürften [509]. Bei manchen Themen, zu denen beispielsweise sehr starke Evidenz vorliegt, kann die Halbwertszeit der Evidenz deutlich über, in anderen Bereichen dagegen wiederum unter diesen 3 Jahren liegen [510]. 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 3 Jahre als Zeitintervall für eine Aktualitätsprüfung an. Auf Basis dieses Intervalls wird bei der Veröffentlichung eines Themenpaketes das Datum festgelegt, zu dem die Information spätestens aktualisiert sein soll. Zudem werden im Rahmen eines Evidenzscannings regelmäßig folgende Quellen gesichtet: Cochrane Database of Systematic Reviews (Cochrane Reviews), McMaster Online Rating of Evidence (MORE) und PubMed. Ferner werden deutsche, europäische und US-amerikanische Zulassungsbehörden 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. Die mögliche Auswirkung einer durch das Evidenzscanning identifizierten relevanten systematischen Übersicht, Studie oder Meldung auf den Aktualisierungsbedarf einer Gesundheitsinformation wird erhoben und bewertet. Diese Bewertung kann die ursprüngliche Zeitplanung bestätigen oder zu einer Änderung des vorgesehenen Aktualisierungstermins führen.

6.4 Informationsprodukte

6.4.1 Hauptformate

Der Kern der Website gesundheitsinformation.de ist auf die Darstellung gesundheits- oder krankheitsbezogener Themen ausgerichtet. Ein Thema kann unterschiedliche Textformate beinhalten. Die einzelnen Textformate sollen die wesentlichen Aspekte des jeweiligen Themas abdecken und zentrale Fragen der Nutzerinnen und Nutzer beantworten. Diese einzelnen Formate sollen zudem unterschiedlichen Informationsbedürfnissen verschiedener Adressatengruppen gerecht werden.

Zu den Hauptformaten gehören:

- **Überblick:** Der Überblick führt in das Thema ein und bietet die Grundlage und Anknüpfung für die in Folge aufgeführten vertiefenden Informationsformate. Der Überblick ist nach einer festen Struktur gegliedert.
- **Mehr Wissen:** Dieses Format informiert ausführlicher über bestimmte Aspekte eines Themas, wie zum Beispiel über medikamentöse und nichtmedikamentöse Behandlungsoptionen einer Erkrankung oder bestimmte diagnostische Maßnahmen. Ein „Mehr Wissen" beschreibt wenn möglich auch Vor- und Nachteile einzelner Behandlungsmöglichkeiten oder bei mangelnder Evidenz auch die resultierenden Unsicherheiten. Ein weiterer Inhalt eines „Mehr Wissen" kann die Beschreibung des Lebens mit einer Erkrankung sein. Dabei wird versucht, sowohl die Perspektive der unmittelbar Betroffenen als auch der Angehörigen zu berücksichtigen.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

- **Was Studien sagen:** Diese Texte sind sachliche, in der Regel auf den Ergebnissen qualitativ hochwertiger, systematischer Evidenzsynthesen beruhende Zusammenfassungen des aktuellen Wissensstandes zu einer im Titel formulierten Frage. Sie beschreiben die Studien ausführlicher und erklären, wie die Antwort auf die Forschungsfrage gefunden wurde.

Zusammengenommen bilden diese Formate sowie die im nachfolgenden Abschnitt 6.4.2 beschriebenen ergänzenden Formate ein evidenzbasiertes „Nachschlagewerk zur Gesundheit“.

Der Ablauf der Erstellung von Gesundheitsinformationen ist in Abschnitt 2.1.7 beschrieben. Informationen zur Gesundheitsforschung sollten ein ähnlich hohes Qualitätssicherungsniveau gewährleisten wie der Forschungsbericht selbst [460]. Die inhaltliche Qualitätssicherung der Texte erfolgt daher durch eine externe Begutachtung durch medizinische Fachexperten, themenspezifisch ggf. auch durch Patientinnen und Patienten. Die Entwürfe 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 zeitlich parallel eine externe Nutzertestung durchgeführt. Überdies sind auch die Patientinnen und Patienten, mit deren Hilfe wir unsere Erfahrungsberichte erstellen, eingeladen, sich zu den Entwürfen der Patienteninformationen, die zu ihren jeweiligen Geschichten gehören, zu äußern (siehe Abschnitt 6.4.3).

6.4.2 Ergänzende Formate

Die Hauptformate können um verschiedene ergänzende Formate erweitert werden, um die Verständlichkeit zu verbessern und das Interesse daran zu steigern. Beispielsweise kann die Integration von Bildern, Ton und Animation dazu beitragen, die Verständlichkeit der Website zu erhöhen, vor allem für Menschen mit eingeschränkter Lesekompetenz (siehe Abschnitt 6.2.4).

Als ergänzende Formate sind folgende zu verstehen:

- Erfahrungsberichte von Betroffenen, Näheres findet sich in Abschnitt 6.4.3,
- Grafiken, Fotos und anderes Bildmaterial,
- Animationen mit Ton und Bild,
- Quiz,
- Glossar zu medizinisch-wissenschaftlichen Begriffen,
- „Kurz erklärt“ – Texte, die themenübergreifend Anatomie, Körperfunktionen, Behandlungs- und Untersuchungsverfahren sowie Prinzipien und Methoden der evidenzbasierten Medizin erklären,
- (Um-)Rechner.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Diese ergänzenden Formate 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 Elemente werden in der Regel auch einer externen Testung durch Nutzerinnen und Nutzer unterzogen. Eine besondere Bedeutung kommt dabei der Barrierefreiheit zu.

6.4.3 Erfahrungsberichte

Patientinnen und Patienten haben möglicherweise mehr Vertrauen in Gesundheitswebsites, wenn diese die Erfahrungen von Betroffenen beinhalten [512].

Viele Patientinnen und Patienten möchten von den Erfahrungen von anderen Betroffenen mit derselben Erkrankung hören oder darüber lesen [257,539]. 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 [217]. Die Bedeutung von Erfahrungsberichten in der medizinischen Praxis und im Gesundheitswesen wird zunehmend gewürdigt [223,528,601].

Einige Funktionen von Erfahrungsberichten sind folgende [539]:

- 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 Multimediawebsite healthtalk.org [144], die im Internet kostenfrei nutzbar ist [257,258,601]. Die Inhalte der Seite stammen von der Database of Personal Experience of Health and Illness (DIPEx) und der Health Experiences Research Group (HERG) an der Universität Oxford.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Die Methodik zur Erhebung, Bearbeitung und Veröffentlichung von Erfahrungsberichten orientiert sich an den etablierten Vorgehensweisen von DIPEx.

Erfahrungsberichte werden 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 2 Interviews geführt werden.

6.4.4 Website

Die Verbreitung der Gesundheitsinformationen erfolgt in erster Linie über die zweisprachige Website www.gesundheitsinformation.de bzw. www.patientenmedhealthonline.org. Ziel ist die Gewährleistung eines hohen Websitestandards im Hinblick auf:

- Benutzerfreundlichkeit und Barrierefreiheit [283,339,414] (siehe Abschnitt 6.2.5),
- Datenschutz [289],
- Transparenz,
- Sichtbarkeit für Suchmaschinen [550],
- Attraktivität für Nutzerinnen und Nutzer,
- Nutzerinteraktivität.

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. Kerninformationsdaten werden auch im RSS-Format bereitgestellt, damit man als Nutzerin oder Nutzer eines RSS-Readers die Informationen der Website verfolgen kann. Darüber hinaus können diese Inhalte automatisch in externe Websites eingebunden werden.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Die Website gesundheitsinformation.de ist von der Health on the Net Foundation (HON) zertifiziert und erfüllt die 8 Prinzipien des HON Code of Conduct (HONcode) für medizinische und Gesundheitswebsites und folgt den Anforderungen der Guten Praxis Gesundheitsinformation.

ABGELÖST DURCH
"Allgemeine Methoden
Version 5.0 vom 10.07.2017"

7 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 7.1 wird das Vorgehen des Instituts beschrieben, wenn die Informationsbeschaffung selbst durchgeführt wird. In Abschnitt 7.2 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.

Ebenso werden unaufgefordert übermittelte Daten, d. h. außerhalb von Stellungnahmeverfahren oder außerhalb sonstiger bestehender Regelungen (z. B. Herstelleranfragen), inhaltlich nicht berücksichtigt. Die unaufgeforderte Übermittlung von Studiendaten birgt das Risiko, dass sie selektiv geschieht und es dadurch zu einer Verzerrung des Ergebnisses der Nutzenbewertung kommt.

7.1 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),

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

- die einzubeziehenden Datenquellen (z. B. bibliografische Datenbanken, Leitliniendatenbanken, Handsuche in ausgewählten Fachzeitschriften, Kontaktaufnahme mit Fachexpertinnen und Fachexperten, Industrie, Patientenorganisationen).

Beispiele und Untersuchungen hierzu sind in einer Reihe von Publikationen zu finden [271,272,398,491,522]. 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.

7.1.1 Ablauf einer Recherche

Die Recherche in bibliografischen Datenbanken, Studienregistern sowie in Leitliniendatenbanken 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 selbstentwickelte 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 [478,479].

7.1.2 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Die großen biomedizinischen Datenbanken wie beispielsweise MEDLINE und EMBASE werden dabei immer ausgewählt (mindestens 2 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.

7.1.3 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 [352].

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), das Metaregister International Clinical Trials Registry Platform Search Portal (ICTRP) der WHO sowie das EU Clinical Trials Register der Europäischen Zulassungsbehörde EMA. Bei der Nutzenbewertung von Arzneimitteln werden zusätzlich Pharmnet.BUND Klinische Prüfung, Arzneimittel-Informationssystem (AMIS) sowie Studienregister der pharmazeutischen Industrie (Studienregister einzelner Firmen und Metaregister) durchsucht. In krankheitsspezifischen Studienregistern wird nur im Ausnahmefall recherchiert.

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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

7.1.4 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 7.1.1 dargestellt.

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

7.1.5 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 Instituts 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 [293]). 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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.

7.1.6 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.

7.1.7 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- / Ausschluss-kriterien genügenden Publikationen in 2 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

- 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 5.2). 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 2 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 2 Personen unabhängig voneinander ausgeführt. Die Qualitätssicherung des ersten Screeningschritts findet systembedingt mithilfe des standardisierten Rechercheprotokolls statt.

7.1.8 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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

7.1.9 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.

7.2 Überprüfung der Informationsbeschaffung

Bei der Dossierbewertung und der Potenzialbewertung erfolgt primär keine eigene Informationsbeschaffung; stattdessen wird eine Überprüfung der Informationsbeschaffung in den Dossiers bzw. in den Anträgen zur Erprobung durchgeführt.

Für die Erstellung eines Dossiers bzw. Antrags sind eine Suche in bibliografischen Datenbanken und eine Suche in öffentlich zugänglichen Studienregistern grundsätzlich durch den Antragsteller durchzuführen; die genauen Vorgaben finden sich in der Verfahrensordnung des G-BA [209,211].

Bei der Dossierbewertung und der Potenzialbewertung finden 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 7.1.1 beschriebene Vorgehen zur Qualitätssicherung sowie die Dokumentvorlagen zur Erstellung von Dossiers und Anträgen zur Erprobung in den Vorgaben der Verfahrensordnung des G-BA [209,211].

Je nachdem, welche Ergebnisse sich aus der formalen und inhaltlichen Prüfung der Dossiers ergeben, schließen sich in den Dossiers 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 [481,580]) sowie ggf. die Durchführung einer vollständigen Literaturrecherche. Das Ergebnis der Überprüfung der Informationsbeschaffung und die Beschreibung des diesbezüglichen Vorgehens sind Teil der Dossierbewertung.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

8 Informationsbewertung

In der Forschung bezeichnet der Begriff „Bias“ (Verzerrung) eine systematische Abweichung zwischen den Forschungsergebnissen und der „Wahrheit“ [473]. 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 [105]. Alle diese Methoden bilden auch die gesetzliche Grundlage für die Arbeit des Instituts.

8.1 Qualitätsbewertung von Einzelstudien

8.1.1 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.

8.1.2 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 [442], Cross-over-Studien [314] und clusterrandomisierte Studien [155]. Sind in einer Studie Zwischenauswertungen geplant, so muss dies mithilfe eines adäquaten sequenziellen Designs berücksichtigt werden [590].

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 [59] zur Untersuchung des Zusammenhangs zwischen Expositionen und seltenen Erkrankungen und Kohortenstudien [60] 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 [317].

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).

3.1.3 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 [237,242]. Es gibt jedoch kein System der Evidenzbewertung, das allgemein akzeptiert und universell in allen systematischen Übersichten einsetzbar ist [318,588]. Aufgrund der Komplexität der Studienbewertung lässt sich anhand einer Evidenzhierarchie auch kein abschließendes Qualitätsurteil bilden [24,599]. Das Institut orientiert sich im Allgemeinen an einer groben Hierarchie von Studientypen, die weitgehend akzeptiert wird, weitgehend mit der Evidenzklassifizierung des G-BA übereinstimmt [211] und Eingang in die Verordnung über die Nutzenbewertung von Arzneimitteln nach § 35a SGB V [80] gefunden hat. Zumindest für die Prüfung von Interventionseffekten wird der höchste Evidenzgrad RCTs und

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

systematischen Übersichten von RCTs zugeordnet. Einzelne RCTs werden in einigen Einteilungen noch in solche hoher oder weniger hoher Qualität eingestuft (siehe Abschnitt 3.1.4).

Spätestens bei der Einteilung nicht randomisierter Studien bezüglich ihres Verzerrungspotenzials jedoch kann das Studiendesign allein keine hinreichende Orientierung mehr bieten [234,261,576], auch wenn eine grundsätzliche Unterscheidung zwischen vergleichenden und nicht vergleichenden Studien sinnvoll erscheint. Wie in Abschnitt 3.8 dargestellt, wird das Institut bei der Einteilung nicht randomisierter vergleichender Studien neben anderen Designaspekten vor allem die Kontrolle möglicher Confounder prüfen. Diese Graduierung bezieht sich jedoch auf das Verzerrungspotenzial (siehe Abschnitt 8.1.4) und nicht auf den Evidenzgrad einer Studie.

8.1.4 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) [299,331] und bei epidemiologischen Studien die Leitlinien und Empfehlungen zur Sicherung „Guter Epidemiologischer Praxis“ (GEP) [132]. 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 [496] und das dazugehörige Erläuterungsdokument [396],

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

- ein Vorschlag zur Erweiterung des CONSORT-Statements für randomisierte Studien bei nichtmedikamentösen Verfahren [55] und das zugehörige Erläuterungsdokument [54],
- das CONSORT-Statement für clusterrandomisierte Studien [93],
- das CONSORT-Statement zur Dokumentation unerwünschter Ereignisse [302],
- das CONSORT-Statement für Nichtunterlegenheits- und Äquivalenzstudien [441],
- das CONSORT-Statement für pragmatische Versuche [604],
- die CONSORT-PRO-Erweiterung für patientenberichtete Endpunkte [91],
- das PRISMA(Preferred Reporting Items for Systematic Reviews and Meta-Analyses)-Statement für Meta-Analysen randomisierter Studien [397] und das zugehörige Erläuterungsdokument [357],
- das TREND(Transparent Reporting of Evaluations with Non-randomized Designs)-Statement für nicht randomisierte Interventionsstudien [128],
- das STROBE(Strengthening the Reporting of Observational Studies in Epidemiology)-Statement für Beobachtungsstudien in der Epidemiologie [570] und das dazugehörige Erläuterungsdokument [570],
- die MOOSE(Meta-analysis of Observational Studies in Epidemiology)-Checkliste für Meta-Analysen von Beobachtungsstudien in der Epidemiologie [534],
- das STARD(Standards for Reporting of Diagnostic Accuracy)-Statement für Diagnosestudien [52] und das dazugehörige Erläuterungsdokument [53] und
- die ISOQOL (International Society of Quality of Life Research) Reporting Standards für patientenberichtete Endpunkte [75].

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 [165,236,264].

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 den 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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. Spezifisch von diesem Vorgehen abgewichen wird im Rahmen der Potenzialbewertung (siehe Abschnitt 3.8).

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.

8.1.5 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 [111,189,202]. 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

„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 „stummem“ 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.

8.1.6 Interpretation von Subgruppenanalysen

Subgruppenanalysen werden in der methodischen Literatur sehr kritisch diskutiert [22,429]. Die Interpretation der Ergebnisse von Subgruppenanalysen auf Studienebene wird im Wesentlichen durch 3 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.
- Multiples 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 [67].

Die Ergebnisse von Subgruppenanalysen sollten unter Beachtung der 3 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 Responderanalysen. Diese Aspekte spielen auch bei der Durchführung und Interpretation von Subgruppenanalysen im Rahmen von Meta-Analysen eine Rolle (siehe Abschnitt 8.3.8).

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 [114].

Ungeachtet der oben getroffenen Einschränkungen können für manche Fragestellungen Subgruppenanalysen die auf absehbare Zeit beste wissenschaftliche Evidenz für die Bewertung von Effekten in ebendiesen Subgruppen darstellen [200], 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 [458]. Sun et al. [536] 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 [335,458]. 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 [335]. 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 [228]. Im Rahmen von systematischen Übersichten ist die Analyse der Heterogenität zwischen den einzelnen Studien und damit ggf. Subgruppenanalysen eine wissenschaftliche (siehe Abschnitt 8.3.8), 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 [134]. Eine entsprechende Zielsetzung findet

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

sich auch in § 35a SGB V zur Bewertung des Nutzens von Arzneimitteln mit neuen Wirkstoffen [136]. 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.

8.1.7 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 Instituts 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,

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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.

8.2 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 [156,353,435]. Da im Folgenden nicht zwischen systematischen Übersichten und HTA-Berichten unterschieden werden muss, bezieht der Begriff „systematische Übersicht" HTA-Berichte mit ein.

8.2.1 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 [156,165,216,435].

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 [1,165,264].

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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

aus anderen Arten von Primärstudien [216]. Systematische Übersichten über Diagnose- und Screeningtests weisen im Vergleich zu Übersichten über Behandlungsmaßnahmen einige methodische Unterschiede auf [122].

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 8.2.2). 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 6.3.3).

Damit das Institut eine systematische Übersicht über Behandlungseffekte verwenden kann, muss diese gewisse Mindestvoraussetzungen erfüllen, d. h., sie darf nach dem Oxman-Guyatt-Index [309,428,430] oder anhand des AMSTAR-Instruments [505-507] 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 [47,119].

8.2.2 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 [112,348]. 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 [23,428,594]. 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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).

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 2 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 [427,428,430] oder anhand des AMSTAR-Instruments [505-507]. Systematische Übersichten gelten nach dem Oxman-Guyatt-Index dann als qualitativ ausreichend, wenn diese von 2 voneinander unabhängigen Gutachtern in der Gesamtbewertung mit mindestens 5 von 7 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 [310].

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.

8.2.3 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 8.3.8). 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

aggregierten Daten beruhen, im Allgemeinen eine höhere Ergebnissicherheit, d. h. präzisere und nicht durch ökologischen Bias verzerrte Ergebnisse [514]. 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 [172]. 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.

8.3 Spezielle biometrische Aspekte

8.3.1 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* [124]. Besondere Aspekte, die beim Vorliegen ordinaler Daten berücksichtigt werden müssen, beschreibt Agresti [67].

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].

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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].

8.3.2 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 beachten. 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 [560]. 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 [185,527].

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 [298]. 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 [523,524]. Das Institut wird in Abhängigkeit von der Fragestellung bei Bedarf auch Bayes'sche Methoden anwenden (z. B. bei indirekten Vergleichen, siehe Abschnitt 8.3.9).

8.3.3 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 2 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 [569]. 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 [461]. 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 [184,279]. 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Definition von relevanten beziehungsweise irrelevanten Unterschieden zwischen Gruppen [344,546]. Auch Methodenhandbücher zur Erstellung von systematischen Übersichten, wie z. B. das Cochrane Handbook [264], 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 Konfidenzintervall) 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 [597]. 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 [323]. 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 [324]. Dabei gibt es auch Ansätze, bei denen sich das Responsekriterium innerhalb einer Studie zwischen den untersuchten Probanden unterscheidet, indem vorab individuelle Therapieziele formuliert werden [453].

Patientenrelevante Endpunkte können auch mithilfe von (komplexen) Skalen erhoben werden. Eine Voraussetzung für die Berücksichtigung solcher Endpunkte ist die Verwendung von validierten bzw. etablierten Instrumenten. 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 2 Gruppen überhaupt für die Patientinnen und Patienten spürbar ist. Diese Relevanzbewertung kann auf Basis von Mittelwertdifferenzen und Responderanalysen vorgenommen werden [497]. 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 [401]. Welches Verfahren bei den Bewertungen des Instituts gewählt werden kann, hängt daher von der Verfügbarkeit der Daten aus den Primärstudien ab.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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]) [449]. 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 [181].

8.3.4 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Einschätzung stark von den behandelnden bzw. endpunkterhebenden Personen abhängen. Eine Zusammenfassung dieser Arbeiten liefern Wood et al. [600]. Demnach sind solche Ergebnisse potenziell hochgradig verzerrt. Ein allgemein akzeptierter Umgang mit diesem Problem im Rahmen von systematischen Übersichten existiert nicht. In der Regel 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 [173]. 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 aus meta-epidemiologischer Forschung begründen [489,600].

8.3.5 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 [46,130].

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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.

8.3.6 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 [313]. 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 2 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“ [115,298,455].

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 Analysemethode ab [454].

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 % [313]. 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 [344]. Zum anderen schützen zum Beispiel die üblichen Designkriterien Randomisierung und Verblindung nicht mehr hinreichend sicher vor Verzerrungen [502]. Auch ohne Kenntnis der

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 [313]. Somit ist bei der Bewertung von Äquivalenzstudien besondere Vorsicht geboten.

8.3.7 Adjustierung und multifaktorielle Verfahren

Vor allem in nicht randomisierten Studien spielen multifaktorielle Verfahren, die es ermöglichen, den Effekt von Confoundern auszugleichen, eine zentrale Rolle [319]. Ein weiteres wichtiges Anwendungsgebiet für multifaktorielle Methoden sind Studien mit mehreren Interventionen [387]. Die Darstellung von Ergebnissen multifaktorieller Verfahren ist jedoch in der medizinischen Fachliteratur leider oft unzureichend [38,404]. Um die Qualität einer solchen Analyse bewerten zu können, sind eine Darstellung der wesentlichen Aspekte im Rahmen der statistischen Modellbildung notwendig [245,462] sowie Angaben zur Güte des gewählten Modells (Goodness-of-Fit) [273]. 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 [27,39,319].

In eigenen regressionsanalytischen Berechnungen setzt das Institut moderne Verfahren ein [244]. Sehr vorsichtig zu interpretieren sind hierbei Ergebnisse multifaktorieller Modelle, die aus einem Variablenselektionsprozess hervorgegangen sind. Falls bei der Modellwahl solche

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Variablenselektionsverfahren unumgänglich sind, wird eine Form der Backward-Elimination eingesetzt, da diese dem Verfahren der Forward-Selektion vorzuziehen ist [244,535]. Wichtig ist hierbei eine sorgfältige Vorauswahl der potenziellen erklärenden Variablen unter sachwissenschaftlicher Kenntnis [126]. Bei Bedarf werden auch moderne Verfahren wie die Lasso-Methode eingesetzt [552]. Bei der Modellierung stetiger Einflussgrößen greift das Institut im Bedarfsfall auf flexible Modellansätze wie zum Beispiel das Verfahren der Fractional Polynomials zurück [463,488], um eine adäquate Beschreibung nicht monotoner Zusammenhänge zu ermöglichen.

8.3.8 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 [165]. 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 [355]. Es kommen zum einen Modelle mit

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 [168,503,574]. 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 [504]. Des Weiteren ist zu beachten, dass die aus einem Modell mit festen Effekten berechneten Konfidenzintervalle für den erwarteten Gesamteffekt selbst bei Vorhandensein einer geringen Heterogenität im Vergleich zu Konfidenzintervallen aus einem Modell mit zufälligen Effekten eine substanziell kleinere Überdeckungswahrscheinlichkeit aufweisen können [64]. 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-analysen 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 zusammenzufassenden Studien trotz Vergleichbarkeit häufig heterogene Effekte zeigen [266]. In dieser Situation ist es erforderlich, die Heterogenität der Studien bezüglich der Ergebnisse zu untersuchen [215]. 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 [307,330]. 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 [265]. 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 %), substanziell (50 bis 90 %) und erheblich (75 bis 100 %) [124]). Ist die Heterogenität der Studien zu groß, so ist eine statistische Zusammenfassung der Studienergebnisse unter Umständen nicht sinnvoll [124]. 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 Nutzenaussage führen. In dieser Situation kann ggf. aber auch ohne quantitative

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Zusammenfassung eine positive Nutzenaussage 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 [205]. 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 [547,566]. 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 [547]. 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 [224]. 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 [438,514,547] (siehe Abschnitt 4.2.3).

Zur Darstellung der Heterogenität im Rahmen einer Meta-Analyse mit zufälligen Effekten verwendet das Institut Prädiktionsintervalle [230,262,451]. 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. [230], 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 [262]. Das Institut stellt Prädiktionsintervalle in Forest Plots von Meta-Analysen mit zufälligen Effekten

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 8.1.6), 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 Subgruppenergebnisse 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 2 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 Nutzenaussagen für diese Gruppen abgeleitet [518].

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 Zellhäufigkeit der entsprechenden Vierfeldertafel zu addieren [124]. Dieses Vorgehen ist adäquat, wenn nicht zu viele Nullzellen vorkommen. Im Fall einer insgesamt geringen Zahl von Ereignissen ist es unter

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 [56,124].

Kommen sogar Studien vor, in denen in beiden Studienarmen kein Ereignis beobachtet wird (sogenannte Doppelnullstudien), so werden diese Studien in der Praxis häufig aus der meta-analytischen Berechnung ausgeschlossen. Dieses Verfahren sollte vermieden werden, wenn zu viele Doppelnullstudien auftreten. Es gibt mehrere Methoden, um den Ausschluss von Doppelnullstudien 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 [562]. Neuere Verfahren wie exakte Methoden [551] oder die Anwendung der Arcus-Sinus-Differenz [404] 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 [140,306]. 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 [140,447]. 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 [140]. 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 [241,448]. Diese Verfahren ermöglichen auch die Berücksichtigung erklärender Variablen [240]. 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 Prädiktions-intervallen in Meta-Analysen von Therapiestudien können bei bivariaten Meta-Analysen von diagnostischen Studien Konfidenz- und Prädiktionsregionen im ROC-Raum dargestellt werden.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 [49,65,66,418,548,589]. 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.

8.3.9 Indirekte Vergleiche

Unter „Methoden für indirekte Vergleiche“ werden sowohl Verfahren für einen einfachen indirekten Vergleich von 2 Interventionen verstanden als auch Verfahren, in denen direkte und indirekte Evidenz kombiniert wird. Letztere werden Mixed-Treatment-Comparison (MTC)-Meta-Analyse [368-370], Multiple-Treatment-Meta-Analyse (MTM) [90] oder auch Netzwerk-Meta-Analyse [372,476] 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 [26,208,477,521,537]. 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 [136] 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 2) 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 [284].

Allerdings müssen adäquate Methoden für indirekte Vergleiche verwendet werden. Die Anwendung nicht adjustierter indirekter Vergleiche (d. h. die naive Verwendung einzelner

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Studienarme) wird vom Institut abgelehnt; es werden ausschließlich adjustierte indirekte Vergleiche akzeptiert. Hierzu zählen insbesondere das Verfahren von Bucher et al. [76] sowie die oben erwähnten Methoden der MTC-Meta-Analysen. Neben den Annahmen paarweiser Meta-Analysen, die auch hier erfüllt sein müssen, muss bei MTC-Meta-Analysen zusätzlich eine ausreichende Konsistenz der geschätzten Effekte aus direkter und indirekter Evidenz gegeben sein. Letzteres ist ein kritischer Punkt, da MTC-Meta-Analysen nur valide Ergebnisse liefern, wenn die Konsistenzannahme erfüllt ist. Obwohl Techniken zur Untersuchung von Inkonsistenzen in der Entwicklung sind [142,369], gibt es in diesem Bereich noch viele offene methodische Fragen. Daher ist eine vollständige Beschreibung des verwendeten Modells zusammen mit verbleibenden Unklarheiten notwendig [537]. 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.

8.3.10 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 7.1.5). 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 [141,275,432]. 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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.

8.3.11 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 [99]. Nur die wichtigsten Verzerrungsarten werden im Folgenden beschrieben; eine ausführliche Übersicht über verschiedene Verzerrungsarten in unterschiedlichsten Situationen gibt Feinstein [183].

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 [263], 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 [316], 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 [316,338].

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 [365]. Die möglichen Ursachen und Auswirkungen fehlender Werte

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 [177].

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 [195].

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 abweichender Weise berichtet, dann liegt ein „Bias durch ergebnisgesteuerte Berichterstattung" („Selective" bzw. „Outcome Reporting Bias") vor [97,160,263]. „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 [530]. 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 [166] 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 [347,373,436,529,530].

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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Bestandteil des Referenztests ist, kann es zu Überschätzungen der diagnostischen Güte kommen („Incorporation Bias“) [351].

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 [361]. 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 [591]. 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 [591]. Trotzdem können Studien des oben beschriebenen Designs Merkmale (insbesondere bezüglich der Stichprobenzusammensetzung) 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 [351]. 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 [351].

Ein allgemeines Problem bei der Schätzung von Effekten ist eine Verzerrung durch Messfehler in den erhobenen Studiendaten [95,100]. 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.

8.4 Qualitative Methoden

8.4.1 Qualitative Studien

Qualitative Forschungsmethoden werden eingesetzt, um subjektives Erleben, subjektive Erfahrungen und individuelles Handeln sowie die soziale Realität zu erforschen und zu verstehen [146,243,376,405]. 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 [148].

Eine systematische Synthese verschiedener qualitativer Studien zu einer Forschungsfrage ist auch möglich [25,337,395,549]. Es gibt jedoch derzeit keine allgemein akzeptierte Vorgehensweise zur Synthese qualitativer Studien und zur Kombination qualitativer und quantitativer Daten [148,149].

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 [117]. Die Hauptaufgabe des Instituts 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 [146,149,243,405].

B) Qualitative Studien bei der Erstellung von Berichten

Verschiedene Informationsquellen können die Erarbeitung systematischer Übersichten unterstützen [147,356,545]. Eine dieser Quellen können Forschungsergebnisse aus qualitativen Studien sein [243,356,406,545]. Qualitative Studien scheinen sich in systematischen Übersichten zur Nutzenbewertung medizinischer Leistungen zu etablieren [146,147,406].

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Qualitative Forschung kann Informationen zur Akzeptanz und Eignung von Interventionen in der Praxis liefern [25,146]. Die Ergebnisse qualitativer Forschung können bei der Interpretation einer systematischen Übersicht hilfreich sein [545] und im Rahmen von Primärstudien oder systematischen Übersichten zur Ermittlung patientenrelevanter Endpunkte eingesetzt werden [146,148,337,405,406].

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.

8.4.2 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 Eruierung von Informationsbedürfnissen, bei der Evaluation von Gesundheitsinformationen und im Rahmen des Stellungnahmeprozesses einzubeziehen.

Folgende Konsultationstechniken werden vom Institut eingesetzt:

- Interviews mit Key Informants (Schlüsselinformanten) [565], z. B. Interviews mit Patientenvertreterinnen und -vertretern zur Ermittlung patientenrelevanter Endpunkte,
- Gruppenmeetings und Konsultationen [407,411,412], z. B. im Rahmen von Erörterungen,
- Gruppeninterviews und Fokusgruppen [146,565] z. B. mit Patientinnen und Patienten bezüglich der Evaluation von Gesundheitsinformationen,
- Erfassung von Meinungen und Surveys (einschließlich Online-Umfragen und Feedbackmechanismen), z. B. bei der Ermittlung der Informationsbedürfnisse der Leserinnen und Leser von www.gesundheitsinformation.de.

Wenn ein tiefgehendes 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 wissen-

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

schaftliche Prozesse ermöglicht jedoch die Berücksichtigung von und die Ausrichtung des Forschungsprozesses auf Fragen und Bedürfnisse von Patientinnen und Patienten [424].

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 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 12 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 gelisteten 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 bis 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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

Ausmaßkategorie	**Erheblich** **nachhaltige** und gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte **große Verbesserung** des therapierelevanten 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 therapierelevanten 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 therapierelevanten Nutzens			Verringerung von nicht schwerwiegenden Symptomen	relevante Vermeidung von Nebenwirkungen

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

In einem ersten Schritt ist es also sinnvoll, die Kriterienliste anzupassen und durch qualitativ und quantitativ gleichwertige Kriterien zu ergänzen. In Tabelle 13 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 Symptomdauer, 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 14):

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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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

		Zielgrößenkategorie			
		Gesamtmortalität	***Symptome (Morbidität)***	***Gesundheitsbezogene Lebensqualität***	***Nebenwirkungen***
Ausmaßkategorie	**Erheblich** **nachhaltige** und gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte **große Verbesserung** des therapierelevanten Nutzens	erhebliche Verlängerung der Überlebensdauer	Langfristige Freiheit von schwerwiegenden *(bzw. schweren)* Symptomen *(bzw. Folgekomplikationen)*	*erhebliche Verbesserung der Lebensqualität*	weitgehende Vermeidung schwerwiegender *(bzw. schwerer)* Nebenwirkungen
	Beträchtlich gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte **deutliche Verbesserung** des therapierelevanten Nutzens	moderate Verlängerung der Überlebensdauer	Abschwächung schwerwiegender *(bzw. schwerer)* Symptome *(bzw. Folgekomplikationen)* *Bedeutsame Verringerung von nicht schwerwiegenden (bzw. schweren) Symptomen (bzw. Folgekomplikationen)*	*bedeutsame Verbesserung der Lebensqualität*	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 therapierelevanten Nutzens	*jegliche Verlängerung der Überlebensdauer*	*jegliche Verringerung schwerwiegender (bzw. schwerer) Symptome (bzw. Folgekomplikationen)* Verringerung von nicht schwerwiegenden *(bzw. schweren)* Symptomen *(bzw. Folgekomplikationen)*	*relevante Verbesserung der Lebensqualität*	*jegliche (statistisch signifikante) Verringerung schwerwiegender (bzw. schwerer) Nebenwirkungen* relevante Vermeidung von *(anderen, nicht schwerwiegenden bzw. schweren)* Nebenwirkungen

*Ergänzungen gegenüber AM-NutzenV *kursiv* gesetzt

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Tabelle 14: Feststellung des Ausmaßes des Zusatznutzens – hierarchisierte Kriterien gemäß AM-NutzenV mit Ergänzungen*

		Zielgrößenkategorie			
		Gesamtmortalität	**Schwerwiegende *(bzw. schwere)* Symptome *(bzw. Folgekomplikationen)* und Nebenwirkungen**	***Gesundheits-bezogene 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 therapierelevanten Nutzens	erhebliche Verlängerung der Überlebensdauer	langfristige Freiheit bzw. weitgehende Vermeidung	*erhebliche Verbesserung*	*nicht besetzt*
	Beträchtlich gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte **deutliche Verbesserung** des therapierelevanten Nutzens	moderate Verlängerung der Überlebensdauer	Abschwächung bzw. relevante Vermeidung	*bedeutsame Verbesserung*	bedeutsame Vermeidung
	Gering gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte **moderate und nicht nur geringfügige Verbesserung** des therapierelevanten Nutzens	*jegliche Verlängerung der Überlebensdauer*	*jegliche Verringerung*	*relevante Verbesserung*	relevante Vermeidung

*Ergänzungen gegenüber AM-Nutzen V *kursiv* gesetzt

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 [519]. 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 [123].
- 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 2 wesentliche Vorteile: (i) Die Präzision der Schätzung fließt in die Bewertung ein; (ii) die statistischen Irrtumswahrscheinlichkeiten lassen sich damit einhergehend auf übliche kleine Werte (z. B. 5 %) beschränken.

Die Schwellenwerte variieren bezüglich der in Tabelle 14 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 [296]. 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. [150] als Anforderung für einen „Durchbruch“ postuliert – als Effekt erheblichen Ausmaßes für die Zielgröße Gesamtmortalität verankert [296].

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

ABGELÖST DURCH: "Allgemeine Methoden Version 5.0 vom 10.07.2017"

zu können. Die entsprechenden Überlegungen dazu können im Detail der ersten durch das Institut durchgeführten Dossierbewertung entnommen werden [296], 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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 15 sind die jeweiligen Schwellenwerte für alle Ausmaßkategorien und Zielgrößenkategorien abgebildet.

Tabelle 15: Inferenzstatistische Schwellenwerte (Hypothesengrenzen) für relative Effektmaße

		Zielgrößenkategorie		
		Gesamtmortalität	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 ≥ 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Detaillierte methodische Rationale für die Festlegung der Schwellenwerte

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

$H_0: RR \geq RR_0 \;\; vs. \;\; H_1: RR < RR_0$

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 2 (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 2 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 Verdoppelung 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 [296] 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 [187] wird auf die Arbeit von Fleiss et al. [192] 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

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

mit Schrittweite 0,05. Für jedes dieser Tupel (RR_1, p_C) wird die benötigte Fallzahl n nach der Formel von Farrington und Manning [180] 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 16 sind die resultierenden (genaueren) tatsächlichen Effekte in Abhängigkeit des Risikos in der Kontrollgruppe für alle oben festgelegten Schwellenwerte aufgetragen (Punkte durch geglättete Kurven approximiert).

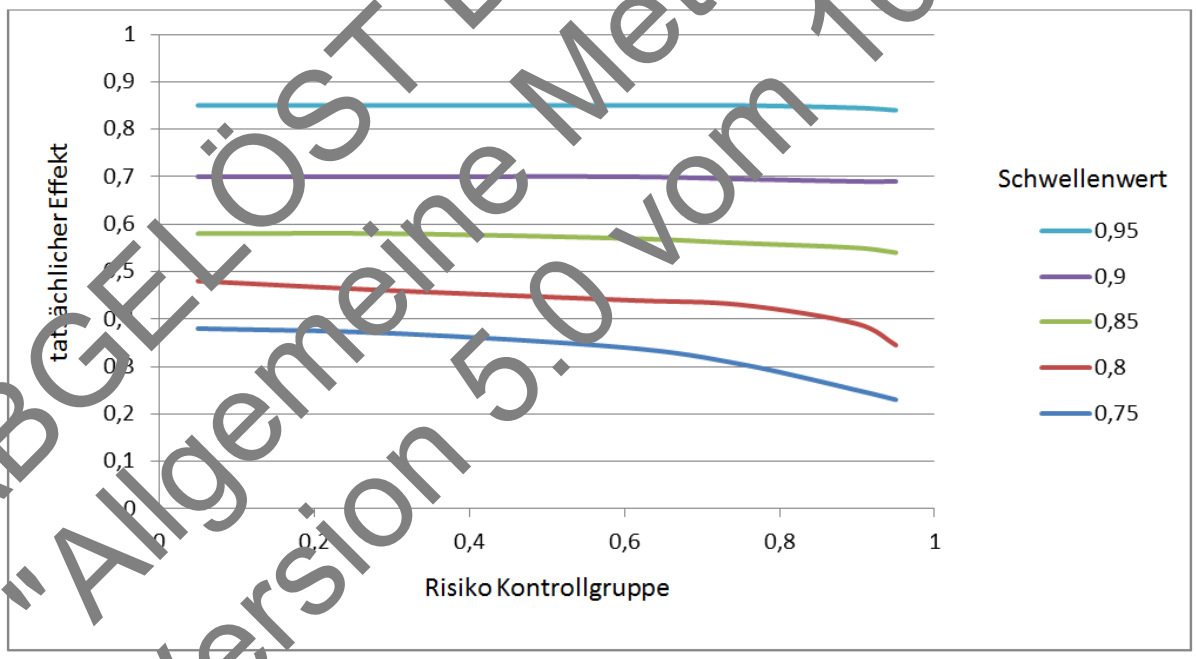


Abbildung 16: Tatsächliche Effekte in Abhängigkeit des Basisrisikos

In Tabelle 16 sind die Bereiche, in denen sich die tatsächlichen Effekte (in Abhängigkeit des Risikos der Kontrollgruppe) realisieren, pro Zielgrößen- und Ausmaßkategorie noch einmal eingetragen.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

Tabelle 16: Tatsächliche Effekte für das relative Risiko

		Zielgrößenkategorie		
		Gesamt-mortalität	Schwerwiegende (bzw. schwere) Symptome (bzw. Folgekompli-kationen) und Nebenwirkungen und gesundheitsbezogene Lebensqualität	Nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) 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 [296] 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 15 veranschlagte Rasterung der Schwellenwerte erscheint vernünftig und praktikabel.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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. Ades AE, Claxton K, Sculpher MJ. Evidence synthesis, parameter correlation and probabilistic sensitivity analysis. Health Econ 2006; 15(4): 373-381.

4. AGREE Collaboration. Appraisal of guidelines for research & evaluation: AGREE instrument. London: St. George's Hospital Medical School; 2001. URL: http://apps.who.int/rhl/agreeinstrumentfinal.pdf.

5. AGREE Next Steps Consortium. Appraisal of guidelines for research & 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.aezq.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, McShane LM, Sauerbrei W, Taube SE. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. PLoS Med 2012; 9(5): e1001216.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften, Ärztliche Zentralstelle Qualitätssicherung. Das Leitlinien-Manual von AWMF und ÄZQ. Z Arztl Fortbild Qualitatssich 2001; 95(Suppl 1): 5-84.

19. Arbeitsgruppe Erhebung und Nutzung von Sekundärdaten der Deutschen Gesellschaft für Sozialmedizin und Prävention und der Deutschen Gesellschaft für Epidemiologie. Gute Praxis Sekundärdatenanalyse (GPS): Leitlinien und Empfehlungen; 3. Fassung 2012 [online] 2012 [Zugriff: 18.03.2015]. URL: http://dgepi.de/fileadmin/pdf/leitlinien_GPS_fassung3.pdf.

20. Arbeitskreis Versorgungsforschung beim Wissenschaftlichen Beirat. Definition und Abgrenzung der Versorgungsforschung [online]. 08.09.2004 [Zugriff: 18.03.2015]. URL: http://www.bundesaerztekammer.de/downloads/Definition.pdf.

21. Arnold M. Gesundheitssystemforschung. In: Herrmann K, Haaser 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 M, 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. Bafeta A, Trinquart L, Seror R, Ravaud P. Reporting of results from network meta-analyses: methodological systematic review. BMJ 2014; 348: g1741.

27. 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.

28. Baker SG. Surrogate endpoints: wishful thinking or reality? J Natl Cancer Inst 2006; 98(8): 502-503.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

29. Baltussen R, Niessen L. Priority setting of health interventions: the need for multi-criteria decision analysis. Cost Eff Resour Alloc 2006; 4: 14.

30. 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.

31. Banta D. The development of health technology assessment. Health Policy 2003; 63(2): 121-132.

32. Barro RJ, Sala-i-Martin X. World real interest rates. In: Blanchard OJ, Fischer S (Ed). NBER Macroeconomics Annual 1990. Cambridge: MIT Press; 1990. S. 15-61.

33. 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.

34. 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.

35. Bastian H, Kaiser T, Matschewsky S. Förderung allgemeiner Gesundheits- und Wissenschaftskenntnisse mittels Bürger- und Patienteninformationen: die Rolle des IQWiG. Z Arztl Fortbild Qualitatssich 2005; 99(6): 379-385.

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. Biesheuvel CJ, Grobbee DE, Moons KG. Distraction from randomization in diagnostic research. Ann Epidemiol 2006; 16(7): 540-544.

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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

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, Zumbrunn 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 P. Comparative accuracy: assessing new tests against existing diagnostic pathways. BMJ 2006; 332(7549): 1089-1092.

51. Bossuyt PM, Lijmer JG, Mol BW. Randomised comparisons of medical tests: sometimes invalid, not always efficient. Lancet 2000; 356(9244): 1844-1847.

52. 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.

53. 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.

54. 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.

55. 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.

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. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. Health Econ 2004; 13(9): 873-884.

58. Brenzel H, Kettner A, Kubis A, Moczall A, Müller A, Rebien M et al. Neueinstellungen im Jahr 2012: Strukturwandel und Demografie prägten die Personalsuche [online]. 08.2013 [Zugriff: 18.03.2015]. (IAB Kurzbericht; Band 17/2013). URL: http://doku.iab.de/kurzber/2013/kb1713.pdf.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

59. 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.

60. 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.

61. Bridges JF, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA et al. Conjoint analysis applications in health: a checklist; a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value Health 2011; 14(4): 403-413.

62. 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.

63. Briggs AH, Weinstein MC, Fenwick EA, Karnon J, Sculpher MJ, Paltiel AD. Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-6. Med Decis Making 2012; 32(5): 722-732.

64. Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. Stat Med 2001; 20(6): 825-840.

65. 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.

66. 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.

67. 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.

68. Brouwer W, Rutten F, Koopmanschap M. Costing in economic evaluations. In: Drummond M, McGuire A (Ed). Economic evaluation in health care: merging theory with practice. Oxford: Oxford University Press; 2001. S. 68-93.

69. Brouwer WB, Culyer AJ, Van Exel NJ, Rutten FF. Welfarism vs. extra-welfarism. J Health Econ 2008; 27(2): 325-338.

70. Brouwer WBF, Koopmanschap MA, Rutten FFH. Productivity costs in cost-effectiveness analysis: numerator or denominator; a further discussion. Health Econ 1997; 6(5): 511-514.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

71. Brouwer WBF, Koopmanschap MA, Rutten FFH. Productivity costs measurement through quality of life: a response to the recommendation of the Washington Panel. Health Econ 1997; 6(3): 253-259.

72. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G et al. AGREE II: advancing guideline development, reporting and evaluation in health care. CMAJ 2010; 182(18): E839-E842.

73. 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.

74. 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.

75. Brundage M, Blazeby J, Revicki D, Bass B, De Vet H, Duffy H et al. Patient-reported outcomes in randomized clinical trials: development of ISOQOL reporting standards. Qual Life Res 2013; 22(6): 1161-1175.

76. 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.

77. Bundesministerium der Justiz. Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz - BStatG) [online]. 25.06.2013 [Zugriff: 18.03.2015]. URL: http://www.gesetze-im-internet.de/bundesrecht/bstatg_1987/gesamt.pdf.

78. Bundesministerium der Justiz. Verordnung zur Schaffung barrierefreier Informationstechnik nach dem Behindertengleichstellungsgesetz (Barrierefreie-Informationstechnik-Verordnung - BITV 2.0) [online]. 12.09.2011 [Zugriff: 18.03.2015]. URL: http://www.gesetze-im-internet.de/bitv_2_0/BJNR184300011.html.

79. Bundesministerium für Gesundheit. Stellungnahme zur Methodik der Kosten-Nutzen-Bewertung von Arzneimitteln [online]. 06.08.2008 [Zugriff: 09.10.2009]. URL: http://www.bmg.bund.de/cln_117/nn_1168258/SharedDocs/Standardartikel/DE/AZ/K/Glossar-Kosten-Nutzen-Bewertung/Stellungnahme.html.

80. 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) [online]. 27.03.2014 [Zugriff: 18.03.2015]. URL: http://www.gesetze-im-internet.de/bundesrecht/am-nutzenv/gesamt.pdf.

81. 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

82. Bundesverfassungsgericht. Leitsatz zum Beschluss des Ersten Senats: Aktenzeichen 1 BvR 347/98 [online]. 06.12.2005 [Zugriff: 18.03.2015]. URL: http://www.bverfg.de/entscheidungen/rs20051206_1bvr034798.html.

83. Burgers JS. Guideline quality and guideline content: are they related? Clin Chem 2006; 52(1): 3-4.

84. 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.

85. Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. Pharm Stat 2006; 5(3): 173-186.

86. Burzykowski T, Molenberghs G, Buyse M (Ed). The evaluation of surrogate endpoints. New York: Springer; 2005.

87. Buxton MJ, Drummond MF, Van Hout BA, Prince RL, Sheldon TA, Szucs T et al. Modelling in economic evaluation: an unavoidable fact of life. Health Econ 1997; 6(3): 217-227.

88. 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.

89. Cairns J. Discounting in economic evaluation. In: Drummond MF, McGuire A (Ed). Economic evaluation in health care: merging theory with practice. Oxford: Oxford University Press; 2001. S. 236-255.

90. Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ 2005; 331(7521): 897-900.

91. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. JAMA 2013; 309(8): 814-822.

92. 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.

93. Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: extension to cluster randomised trials. BMJ 2012; 345: e5661.

94. Canadian Agency for Drugs and Technologies in Health. Guidelines for the economic evaluation of health technologies: Canada. Ottawa: CADTH; 2006. URL: http://www.cadth.ca/media/pdf/186_EconomicGuidelines_e.pdf.

95. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. Measurement error in nonlinear models: a modern perspective. London: Chapman & Hall; 2006.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

96. Centre for Evidence-based Medicine. Levels of evidence (March 2009) [online]. 03.2009 [Zugriff: 18.03.2015]. URL: http://www.cebm.net/index.aspx?o=1025.

97. 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.

98. Charles C, Gafni A, Whelan T, O'Brien MA. Treatment decision aids: conceptual issues and future directions. Health Expect 2005; 8(2): 114-125.

99. Chavalarias D, Ioannidis JP. Science mapping analysis characterizes 235 biases in biomedical research. J Clin Epidemiol 2010; 63(11): 1205-1215.

100. Cheng CL, Van Ness JW. Statistical regression with measurement error. London: Arnold; 1999.

101. Chiou CF, Hay JW, Wallace JF, Bloom BS, Neumann PJ, Sullivan SD et al. Development and validation of a grading system for the quality of cost-effectiveness studies. Med Care 2003; 41(1): 32-44.

102. Ciani O, Buyse M, Garside R, Pavey T, Stein K, Sterne JA et al. Comparison of treatment effect sizes associated with surrogate and final patient relevant outcomes in randomised controlled trials: meta-epidemiological study. BMJ 2013; 346: f457.

103. Claxton K, Paulden M, Gravelle H, Brouwer W, Culyer AJ. Discounting and decision making in the economic evaluation of health-care technologies. Health Econ 2011; 20(1): 2-15.

104. 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.

105. Cochrane Collaboration. Our principles [online]. 16.01.2014 [Zugriff: 18.03.2015]. URL: http://www.cochrane.org/about-us/our-principles.

106. Cochrane Effective Practice and Organisation of Care Review Group. The data collection checklist [online]. 06.2002 [Zugriff: 18.03.2015]. URL: http://epoc.cochrane.org/sites/epoc.cochrane.org/files/uploads/datacollectionchecklist.pdf.

107. Collège des Économistes de la Santé. French guidelines for the economic evaluation of health care technologies [online]. 09.2004 [Zugriff: 18.03.2015]. URL: http://www.ces-asso.org/docs/France_Guidelines_HE_Evaluation.PDF.

108. Commission of the European Communities. eEurope 2002: quality criteria for health related websites [online]. 29.11.2002 [Zugriff: 18.03.2015]. URL: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2002:0667:FIN:EN:PDF.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

109. 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: 18.03.2015]. URL: http://www.who.int/patientsafety/highlights/COE_patient_and_medication_safety_gl.pdf.

110. Corbin JM, Strauss AL. Weiterleben lernen: Verlauf und Bewältigung chronischer Krankheit. Bern: Huber; 2003.

111. 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.

112. Cornell JE, Laine C. The science and art of deduction: complex systematic overviews. Ann Intern Med 2008; 148(10): 786-788.

113. 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.

114. Cui L, Hung HMJ, Wang SJ, Tsong Y. Issues related to subgroup analysis in clinical trials. J Biopharm Stat 2002; 12(3): 347-358.

115. 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.

116. Da Costa BR, Rutjes AWS, Johnston BC, Reichenbach S, Nüesch 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.

117. 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.

118. Danner M, Hummel JM, Volz F, Van Manen JG, Wiegard B, Dintsios CM et al. Integrating patients' views into health technology assessment: Analytic Hierarchy Process (AHP) as a method to elicit patient preferences. Int J Technol Assess Health Care 2011; 27(4): 369-375.

119. 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.

120. 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.

121. De Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Econ 2012; 21(2): 145-172.

122. Deeks JJ. Systematic reviews in health care: systematic reviews of evaluations of diagnostic and screening tests. BMJ 2001; 323(7305): 157-162.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

123. 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.

124. 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.

125. 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.

126. 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.

127. 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.

128. 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.

129. Desroches B, Francis M. World real interest rates: a global savings and investment perspective [online]. 03.2007 [Zugriff: 18.03.2015]. (Bank of Canada Working Papers; Band 2007-16). URL: http://www.bankofcanada.ca/wp-content/uploads/2010/03/wp07-16.pdf.

130. Desu MM, Raghavarao D. Sample size methodology. Boston: Academic Press; 1990.

131. 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.

132. Deutsche Gesellschaft für Epidemiologie. Leitlinien und Empfehlungen zur Sicherung von guter epidemiologischer Praxis (GEP): Langversion [online]. 03.2008 [Zugriff: 18.03.2015]. URL: http://www.gmds.de/pdf/publikationen/stellungnahmen/stell_gep_ergaenzung.pdf.

133. 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.

134. 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: 18.03.2015]. URL: http://dipbt.bundestag.de/doc/btd/15/015/1501525.pdf.

135. Deutscher Bundestag. Gesetz zur Modernisierung der gesetzlichen Krankenversicherung (GKV-Modernisierungsgesetz - GMG). Bundesgesetzblatt Teil 1 2003; (55): 2190-2258.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

136. Deutscher Bundestag. Gesetz zur Neuordnung des Arzneimittelmarktes in der gesetzlichen Krankenversicherung (Arzneimittelmarktneuordnungsgesetz – AMNOG) vom 22. Dezember 2010. Bundesgesetzblatt Teil 1 2010; (67): 2262-2277.

137. 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.

138. 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.

139. Deutsches Netzwerk Evidenzbasierte Medizin. Die "Gute Praxis Gesundheitsinformation". Z Evid Fortbild Qual Gesundhwes 2010; 104(1): 66-68.

140. 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.

141. Di Pietrantonj C. Four-fold table cell frequencies imputation in meta-analysis. Stat Med 2006; 25(13): 2299-2322.

142. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. Stat Med 2010; 29(7-8): 932-944.

143. 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/GBEDownloadsT/buergerorientierung.pdf?__blob=publicationFile.

144. DIPEx. Healthtalk.org [online]. [Zugriff: 18.03.2015]. URL: http://www.healthtalk.org/.

145. 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.

146. 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.

147. Dixon-Woods M, Fitzpatrick R. Qualitative research in systematic reviews: has established a place for itself. BMJ 2001; 323(7316): 765-766.

148. 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.

149. 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

150. 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.

151. Dolan JG. Shared decision-making: transferring research into practice; the Analytic Hierarchy Process (AHP). Patient Educ Couns 2008; 73(3): 418-425.

152. 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.

153. Dolan P, Edlin R, Tsuchiya A. The relative societal value of health gains to different beneficiaries: final report [online]. 31.01.2008 [Zugriff: 11.07.2011]. URL: http://www.haps.bham.ac.uk/publichealth/methodology/docs/publications/JH11_Social_Value_QALY_Final_Report_Paul_Dolan_et_al_2008.pdf.

154. 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.

155. Donner A, Klar J. Design and analysis of cluster randomization trials in health research. London: Arnold; 2000.

156. 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.

157. Drazen JM. COX-2 inhibitors: a lesson in unexpected problems. N Engl J Med 2005; 352(11): 1131-1132.

158. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. BMJ 1996; 313(7052): 275-283.

159. 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.

160. 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.

161. 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.

162. Edwards AGK, Elwyn GJ, Mulley A. Explaining risks: turning numerical data into meaningful pictures. BMJ 2002; 324(7341): 827-830.

163. 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

164. 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.

165. Egger M, Davey Smith G, Altman DG (Ed). Systematic reviews in health care: meta-analysis in context. London: BMJ Publishing Group; 2001.

166. 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.

167. 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.

168. 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.

169. 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.

170. Epstein RM, Alper BS, Quill TE. Communicating evidence for participatory decision making. JAMA 2004; 291(19): 2359-2366.

171. 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.

172. European Medicines Agency. Points to consider on application with: 1. meta-analyses; 2. one pivotal study [online]. 31.05.2001 [Zugriff: 18.03.2015]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003637.pdf.

173. European Medicines Agency. Guideline on the choice of the non-inferiority margin [online]. 27.07.2005 [Zugriff: 18.03.2015]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003636.pdf.

174. 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: 18.03.2015]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003637.pdf.

175. European Medicines Agency. Guideline on clinical trials in small populations [online]. 27.07.2006 [Zugriff: 18.03.2015]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003615.pdf.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

176. European Medicines Agency. Guideline on clinical investigation of medicinal products in the treatment of diabetes mellitus: draft [online]. 20.01.2010 [Zugriff: 18.03.2015]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/02/WC500073570.pdf.

177. European Medicines Agency. Guideline on missing data in confirmatory clinical trials [online]. 02.07.2010 [Zugriff: 18.03.2015]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500096793.pdf.

178. Evidence-Based Medicine Working Group. Evidence-based medicine: a new approach to teaching the practice of medicine. JAMA 1992; 268(17): 2420-2425.

179. 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.

180. 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.

181. Fayers P, Machin D. Quality of life: the assessment, analysis and interpretation of patient-reported outcomes. Chichester: Wiley; 2007.

182. Feeny D. As good as it gets but good enough for which applications? Med Decis Making 2006; 26(4): 307-309.

183. Feinstein AR. Clinical epidemiology: the architecture of clinical research. Philadelphia: Saunders; 1985.

184. Feinstein AR. Invidious comparisons and unmet clinical challenges. Am J Med 1992; 92(2): 117-120.

185. Feise RJ. Do multiple outcome measures require p-value adjustment? BMC Med Res Methodol 2002; 2: 8.

186. 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.

187. Fenwick E, O'Brien BJ, Briggs A. Cost-effectiveness acceptability curves-facts, fallacies and frequently asked questions. Health Econ 2004; 13(5): 405-415.

188. Ferrante di Ruffano L, Hyde CJ, McCaffery KJ, Bossuyt PM, Deeks JJ. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. BMJ 2012; 344: e686.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

189. Ferreira-Gonzáles 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.

190. 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.

191. Field MJ, Lohr KN (Ed). Clinical practice guidelines: directions for a new program. Washington: National Academy Press; 1990.

192. Fleiss JL, Tytun A, Ury HK. A simple approximation for calculating sample sizes for comparing independent proportions. Biometrics 1980; 36(2): 343-346.

193. Fleming TR. Surrogate endpoints and FDA's accelerated approval process. Health Aff (Millwood) 2005; 24(1): 67-78.

194. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? Ann Intern Med 1996; 125(7): 605-613.

195. Fletcher RH, Fletcher SW. Klinische Epidemiologie: Grundlagen und Anwendung. Bern: Huber; 2007.

196. Flynn TN. Using conjoint analysis and choice experiments to estimate QALY values: issues to consider. Pharmacoeconomics 2010; 28(9): 711-722.

197. Food and Drug Administration. Guidance for industry: developing medical imaging drug and biological products; part 2: clinical indications [online]. 06.2004 [Zugriff: 18.03.2015]. URL: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071663.pdf.

198. Food and Drug Administration. Guidance for industry: patient-reported outcome measures; use in medical product development to support labeling claims [online]. 12.2009 [Zugriff: 18.03.2015]. URL: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf.

199. 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 DIAMONDS-study. BMC Med Res Methodol 2007; 7: 16.

200. 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

201. 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.

202. Freemantle N, Calvert M. Weighing the pros and cons for composite outcomes in clinical trials. J Clin Epidemiol 2007; 60(7): 658-659.

203. 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.

204. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. Med Decis Making 1991; 11(2): 88-94.

205. 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.

206. Gafni A, Birch S, Mehrez A. Economics, health and health economics: HYEs versus QALYs. J Health Econ 1993; 12(3): 325-339.

207. Garber AM, Weinstein MC, Torrance GW, Kamlet MS. Theoretical foundations of cost-effectiveness analysis. In: Gold MR, Siegel JE, Russell LB, Weinstein MC (Ed). Cost-effectiveness in health and medicine. New York: Oxford University Press; 1996. S. 25-53.

208. Gartlehner G, Moore CG. Direct versus indirect comparisons: a summary of the evidence. Int J Technol Assess Health Care 2008; 24(2): 170-177.

209. Gemeinsamer Bundesausschuss. Anlage I zum 2. Kapitel der Verfahrensordnung: Antrag zur Erprobung von Untersuchungs- und Behandlungsmethoden nach § 137e des Fünften Buches Sozialgesetzbuch (SGB V) [online]. [Zugriff: 18.03.2015]. URL: http://www.g-ba.de/downloads/17-98-3627/Anlage%20I_2-Kapitel-VerfO_Erprobungsantrag_Formular.pdf.

210. 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: 18.03.2015]. URL: http://www.g-ba.de/downloads/39-261-650/2008-03-13-IQWiG-Anpassung-Generalauftrag.pdf.

211. Gemeinsamer Bundesausschuss. Verfahrensordnung des Gemeinsamen Bundesausschusses [online]. 19.11.2014 [Zugriff: 19.03.2015]. URL: https://www.g-ba.de/downloads/62-492-938/VerfO_2014-06-19_iK-2014-11-19.pdf.

212. Gerhardt U. Patientenkarrieren. Frankfurt am Main: Suhrkamp; 1986.

213. Gesellschaft für Evaluation. Standards für Evaluation. Mainz: DeGEval; 2008. URL: http://www.degeval.de/fileadmin/user_upload/Sonstiges/STANDARDS_2008-12.pdf.

214. Glasziou PP, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. BMJ 2007; 334(7589): 349-351.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

215. Glasziou PP, Sanders SL. Investigating causes of heterogeneity in systematic reviews. Stat Med 2002; 21(11): 1503-1511.

216. Glasziou PP, Vandenbroucke JP, Chalmers I. Assessing the quality of research. BMJ 2004; 328(7430): 39-41.

217. 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.

218. 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.

219. Gøtzsche PC, Liberati A, Torri V, Rossetti L. Beware of surrogate outcome measures. Int J Technol Assess Health Care 1996; 12(2): 238-246.

220. Graf von der Schulenburg JM, Greiner W, Jost F, Klusen N, Kubin M, Leidl R et al. Deutsche Empfehlungen zur gesundheitsökonomischen Evaluation: dritte und aktualisierte Fassung des Hannoveraner Konsens. Gesundheitsökonomie & Qualitätsmanagement 2007; 12(5): 285-290.

221. 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/ICM-CPG_lang_2011.pdf.

222. Gray JAM. How to get better value healthcare. Oxford: Offox Press; 2007.

223. Greenhalgh T, Hurwitz B. Narrative based medicine: why study narrative? BMJ 1999; 318(7175): 48-50.

224. Greenland S, Morgenstern H. Ecological bias, confounding, and effect modification. Int J Epidemiol 1989; 18(1): 269-274.

225. Greiner W, Damm O. Die Berechnung von Kosten und Nutzen. In: Schöffski O, Graf von der Schulenburg JM (Ed). Gesundheitsökonomische Evaluationen. Berlin: Springer; 2012. S. 23-42.

226. Grimes DA, Schulz K. An overview of clinical research: the lay of the land. Lancet 2002; 359(9300): 57-61.

227. Grimes DA, Schulz KF. Surrogate end points in clinical research: hazardous to your health. Obstet Gynecol 2005; 105(5 Pt 1): 1114-1118.

228. Grouin JM, Coste M, Lewis J. Subgroup analyses in randomized clinical trials: statistical and regulatory issues. J Biopharm Stat 2005; 15(5): 869-882.

229. Gu Y, Norman R, Viney R. Estimating health state utility values from discrete choice experiments: a QALY space model approach [online]. 2013 [Zugriff: 18.03.2015]. URL: http://www.icmconference.org.uk/index.php/icmc/ICMC2013/paper/viewFile/537/210.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

230. Guddat C, Grouven U, Bender R, Skipka G. A note on the graphical presentation of prediction intervals in random-effects meta-analyses. Syst Rev 2012; 1: 34.

231. Guyatt GH. Evidence-based medicine. ACP J Club 1991; 14(Suppl 2): A16.

232. 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.

233. 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.

234. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P et al. GRADE guidelines; 4: rating the quality of evidence; study limitations (risk of bias). J Clin Epidemiol 2011; 64(4): 407-415.

235. 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.

236. Guyatt GH, Rennie D. Users' guides to the medical literature: a manual for evidence-based clinical practice. Chicago: American Medical Association; 2002.

237. 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.

238. 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.

239. 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.

240. Hamza TH, Van Houwelingen HC, Heijenbrok-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.

241. 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.

242. Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. BMJ 2001; 323(7308): 334-336.

243. 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

244. Harrell FE. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. New York: Springer; 2001.

245. 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.

246. Harris J. QALYfying the value of life. J Med Ethics 1987; 13(3): 117-123.

247. 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.

248. Hart D (Ed). Ärztliche Leitlinien im Medizin- und Gesundheitsrecht: Recht und Empirie professioneller Normbildung. Baden-Baden: Nomos; 2005. (Gesundheitsrecht und Gesundheitswissenschaften; Band 9).

249. Harteloh P. The meaning of quality in health care: a conceptual analysis. Health Care Anal 2003; 11(3): 259-267.

250. Haute Autorité de Santé. Choices in methods for economic evaluation [online]. 10.2012 [Zugriff: 18.03.2015]. URL: http://www.has-sante.fr/portail/upload/docs/application/pdf/2012-10/choices_in_methods_for_economic_evaluation.pdf.

251. Hayden JA, Cote P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. Ann Intern Med 2006; 144(6): 427-437.

252. Hayden JA, Van der Windt DA, Cartwright JL, Cote P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med 2013; 158(4): 280-286.

253. Haynes RB. Forming research questions. J Clin Epidemiol 2006; 59(9): 881-886.

254. 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): 1801-1803.

255. 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.

256. Hennekens CH, Buring JE. Epidemiology in medicine. Boston: Little, Brown and Company; 1987.

257. 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.

258. Herxheimer A, Ziebland S. DIPEx: fresh insights for medical practice. J R Soc Med 2003; 96(5): 209-210.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

259. Hessel F, Kohlmann T, Krauth C, Nowy R, Seitz R, Siebert U et al. Gesundheitsökonomische Evaluation in der Rehabilitation; Teil 1: Prinzipien und Empfehlungen für die Leistungserfassung. In: Verband Deutscher Rentenversicherungsträger (Ed). Förderschwerpunkt "Rehabilitationswissenschaften": Empfehlungen der Arbeitsgruppen "Generische Methoden", "Routinedaten" und "Reha-Ökonomie". Frankfurt: VDR; 1999. S. 103-193. (DRV-Schriften; Band 16).

260. Hicks NJ. Evidence-based health care. Bandolier 1997; 4(5): 8.

261. Higgins JP, Ramsay C, Reeves BC, Deeks JJ, Shea B, Valentine JC et al. Issues relating to study design and risk of bias when including non-randomized studies in systematic reviews on the effects of interventions. Res Syn Meth 2013; 4(1): 12-25.

262. Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. J R Stat Soc A 2009; 172(1): 137-159.

263. 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.

264. Higgins JPT, Green S (Ed). Cochrane handbook for systematic reviews of interventions. Chichester: Wiley; 2008.

265. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002; 21(11): 1539-1558.

266. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003; 327(7414): 557-560.

267. Hill AB (Ed). Controlled clinical trials. Oxford: Blackwell; 1960.

268. Hingorani AD, Windt DA, Riley RD, Abrams K, Moons KG, Steyerberg EW et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. BMJ 2013; 346: e5793.

269. Hirsh J, Guyatt G. Clinical experts or methodologists to write clinical guidelines? Lancet 2009; 374(9686): 273-275.

270. 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.

271. 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.

272. Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. Cochrane Database Syst Rev 2009; (1): MR000006.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

273. 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.

274. 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.

275. 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.

276. Hummel JM, IJzerman MJ. The use oft the Analytic Hierarchy Process in health care decision making. Enschede: University of Twente; 2009.

277. Hummel M, IJzerman M. The past and future of the AHP in health care decision making [online]. In: Proceedings of the XI International Symposium on the Analytic Hierarchy Process (ISAHP); 15.-18.06.2011; Sorrent, Italien. [Zugriff: 18.03.2015]. URL: http://doc.utwente.nl/79775/1/past_and_future.pdf.

278. Hummel MJM, Steuten LMG, Groothuis-Oudshoorn KGM, Ijzerman MJ. How the Analytic Hierarchy Process may fill missing gaps in early decision modeling. ISPOR Connections 2011; 17(3): 10-11.

279. 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.

280. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. Pharmacoeconomics 2013; 31(5): 361-367.

281. 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.

282. 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: 18.03.2015]. URL: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E1/Step4/E1_Guideline.pdf.

283. Inan H. Measuring the success of your website: a customer-centric approach to website management. French's Forest: Pearson Education Australia; 2002.

284. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden zur Bewertung von Verhältnissen zwischen Nutzen und Kosten [online]. 12.10.2009 [Zugriff: 18.03.2015]. URL: https://www.iqwig.de/download/Methodik_fuer_die_Bewertung_von_Verhaeltnissen_zwischen_Kosten_und_Nutzen.pdf.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

285. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Anhang: Dokumention der Stellungnahmen zur „Methodik für die Bewertung von Verhältnissen zwischen Nutzen und Kosten im System der deutschen gesetzlichen Krankenversicherung Version 1.0“ [online]. 30.09.2008 [Zugriff: 18.03.2015]. URL: https://www.iqwig.de/download/08-10-14_Dokumentation_der_Stellungnahmen_KNB_Version_1_0.pdf.

286. 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: 18.03.2015]. (IQWiG-Berichte; Band 80). URL: https://www.iqwig.de/download/A10-05_Rapid_Report_Version_1-1_Surrogatendpunkte_in_der_Onkologie.pdf.

287. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Dokumentation und Würdigung der Stellungnahmen zur „Aktualisierung einiger Abschnitte der Allgemeinen Methoden Version 4.0 sowie neue Abschnitte zur Erstellung der Allgemeinen Methoden Version 4.1“ [online]. 28.11.2013 [Zugriff: 18.03.2015]. URL: https://www.iqwig.de/download/Dokumentation_und_Wuerdigung_der_Stellungnahmen_IQWiG_Methoden_4-1.pdf.

288. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Entwurf einer Methodik für die Bewertung von Verhältnissen zwischen Nutzen und Kosten im System der deutschen gesetzlichen Krankenversicherung: Version 2.0 [online]. 16.03.2009 [Zugriff: 18.03.2015]. URL: https://www.iqwig.de/download/09-03-18_Entwurf_Methoden_Kosten-Nutzen-Bewertung_Version_2_0.pdf.

289. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Gesundheitsinformationen.de: Hinweise zur Nutzung [online]. 23.01.2014 [Zugriff: 19.03.2015]. URL: http://www.gesundheitsinformation.de/hinweise-zur-nutzung.2010.de.html.

290. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Methodik für die Bewertung von Verhältnissen zwischen Nutzen und Kosten im System der deutschen gesetzlichen Krankenversicherung: Version 1.1 [online]. 09.10.2008 [Zugriff: 18.03.2015]. URL: https://www.iqwig.de/download/08-10-14_Entwurf_Methoden_Kosten-Nutzen-Bewertung_Version_1_1.pdf.

291. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Technischer Anhang: Modellierung [online]. 09.10.2008 [Zugriff: 18.03.2015]. URL: https://www.iqwig.de/download/TA_KNB_Modellierung_v_1_0.pdf.

292. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Technischer Anhang: Unsicherheit [online]. 09.10.2008 [Zugriff: 18.03.2015]. URL: https://www.iqwig.de/download/TA_KNB_Unsicherheit_v_1_0.pdf.

293. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Vereinbarung über die vertrauliche Behandlung von Unterlagen [online]. 19.08.2005 [Zugriff: 18.03.2015]. URL: http://www.iqwig.de/download/IQWiG-VFA-Mustervertrag.pdf.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

294. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Würdigung der Stellungnahmen zur „Methodik für die Bewertung von Verhältnissen zwischen Nutzen und Kosten im System der deutschen gesetzlichen Krankenversicherung Version 1.0“ [online]. 09.10.2008 [Zugriff: 18.03.2015]. URL: https://www.iqwig.de/download/08-10-14_Wuerdigung_der_Stellungnahmen_KNB_Version_1_0.pdf.

295. 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: 18.03.2015]. (IQWiG-Berichte; Band 68). URL: https://www.iqwig.de/download/A05-20C_Abschlussbericht_Bupropion_Mirtazapin_und_Reboxetin_bei_Depressionen.pdf.

296. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Ticagrelor: Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag A11-02 [online]. 29.09.2011 [Zugriff: 18.03.2015]. (IQWiG-Berichte; Band 96). URL: https://www.iqwig.de/download/A11-02_Ticagrelor_Nutzenbewertung_35a_SGB_V_.pdf.

297. 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.

298. 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.

299. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. Webauftritt [online]. [Zugriff: 18.03.2015]. URL: http://www.ich.org.

300. International Society for Pharmacoeconomics and Outcomes Research. ISPOR good practices for outcomes research index [online]. [Zugriff: 18.03.2015]. URL: http://www.ispor.org/workpaper/practices_index.asp.

301. Ioannidis JPA. Contradicted and initially stronger effects in highly cited clinical research. JAMA 2005; 294(2): 218-228.

302. Ioannidis JPA, Evans S, Gotzsche 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.

303. Ioannidis JPA, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. JAMA 2001; 285(4): 437-443.

304. Ioannidis JPA, Mulrow CD, Goodman SN. Adverse events: the more you search, the more you find. Ann Intern Med 2006; 144(4): 298-300.

305. Irmen L, Linner U. Die Repräsentation generisch maskuliner Personenbezeichnungen: eine theoretische Integration bisheriger Befunde. Z Psychol 2005; 213(3): 167-175.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

306. 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.

307. Jackson D. The power of the standard test for the presence of heterogeneity in meta-analysis. Stat Med 2006; 25(15): 2688-2699.

308. Jackson N, Waters E. Criteria for the systematic review of health promotion and public health interventions. Health Promot Int 2005; 20(4): 367-374.

309. Jadad AR. Randomised controlled trials: a user's guide. London: BMJ Books; 1998.

310. Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. CMAJ 1997; 156(10): 1411-1416.

311. Johannesson M. Avoiding double-counting in pharmacoeconomic studies. Pharmacoeconomics 1997; 11(5): 385-388.

312. Johnson RF. Sample size issues for conjoint analysis. In: Orme BK (Ed). Getting started with conjoint analysis: strategies for product design and pricing research. Madison: Research Publishers LLC; 2010. S. 57-66.

313. Jones B, Jarvis P, Lewis J, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. BMJ 1996; 313(7048): 36-39.

314. Jones B, Kenward MG. Design and analysis of cross-over trials. London: Chapman and Hall; 1989. (Monographs on Statistics and Applied Probability; Band 34).

315. Jull A, Bennett D. Do n-of-1 trials really tailor treatment? Lancet 2005; 365(9476): 1992-1994.

316. Jüni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. BMJ 2001; 323(7303): 42-46.

317. Kasl PH, Gold EB. Modern epidemiologic study designs. In: Ahrens W, Pigeot I (Ed). Handbook of epidemiology. Berlin: Springer; 2005. S. 321-344.

318. 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.

319. Katz MH. Multivariable analysis: a primer for readers of medical research. Ann Intern Med 2003; 138(8): 644-650.

320. Kelley E, Hurst J. Health Care Quality Indicators Project: conceptual framework paper [online]. 09.03.2006 [Zugriff: 18.03.2015]. (OECD Health Working Papers; Band 23). URL: http://www.oecd.org/dataoecd/1/36/36262363.pdf.

321. Kettunen T, Liimatainen L, Villberg J, Perko U. Developing empowering health counseling measurement: preliminary results. Patient Educ Couns 2006; 64(1-3): 159-166.

322. Kickbusch IS. Health literacy: addressing the health and education divide. Health Promot Int 2001; 16(3): 289-297.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

323. Kieser M. Assessment of clinical relevance by considering point estimates and associated confidence intervals. Pharm Stat 2005; 4(2): 101-107.

324. 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.

325. Klusen N, Meusch M (Ed). Wettbewerb und Solidarität im europäischen Gesundheitsmarkt. Baden-Baden: Nomos Verlagsgesellschaft; 2006. (Beiträge zum Gesundheitsmanagement; Band 16).

326. 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.

327. Knottnerus JA, Bouter LM. The ethics of sample size: two-sided testing and one-sided thinking. J Clin Epidemiol 2001; 54(2): 109-110.

328. 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.

329. Köbberling J, Trampisch HJ, Windeler J. Memorandum for the evaluation of diagnostic measures. J Clin Chem Clin Biochem 1990; 28(12): 873-879.

330. Koch A, Ziegler S. Metaanalyse als Werkzeug zum Erkenntnisgewinn. Med Klin 2000; 95(2): 109-116.

331. Kolman J, Meng P, Scott G. Good clinical practice: standard operating procedures for clinical researchers. Chichester: Wiley; 1998.

332. 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.

333. Kools M, Van de Wiel MW, Ruiter RA, Kok G. Pictures and text in instructions for medical devices: effects on recall and actual performance. Patient Educ Couns 2006; 64(1-3): 104-111.

334. Koopmanschap MA, Rutten FFH, Van Ineveld BM, Van Roijen L. The friction cost method for measuring indirect costs of disease. J Health Econ 1995; 14(2): 171-189.

335. Kraemer HC, Frank E, Kupfer DJ. Moderators of treatment outcomes: clinical, research, and policy importance. JAMA 2006; 296(10): 1286-1289.

336. Kranich C. Patientenkompetenz: was müssen Patienten wissen und können? Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2004; 47(10): 950-956.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

337. Kristensen FB, Sigmund H. Health technology assessment handbook. Kopenhagen: Danish Centre for Heath Technology Assessment; 2007. URL: http://sundhedsstyrelsen.dk/publ/Publ2008/MTV/Metode/HTA_Handbook_net_final.pdf.

338. 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.

339. Krug S. Don't make me think! Web Usability; das intuitive Web. Heidelberg: mitp; 2006.

340. Kulbe A. Grundwissen Psychologie, Soziologie und Pädagogik: Lehrbuch für Pflegeberufe. Stuttgart: Kohlhammer; 2009.

341. 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.

342. 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.

343. Laaser U, Hurrelmann K. Gesundheitsförderung und Krankheitsprävention. In: Hurrelmann K, Laaser U (Ed). Handbuch Gesundheitswissenschaften. Weinheim: Juventa Verlag; 1998. S. 395-424.

344. Lange S, Freitag G. Choice of delta: requirements and reality, results of a systematic review. Biom J 2005; 47(1): 12-27.

345. Lapsley P. The patient's journey: traveling through life with a chronic illness. BMJ 2004; 329(7466): 582-583.

346. Last JM, Spasoff RA, Harris SS, Thuriaux MC (Ed). A dictionary of epidemiology. Oxford: Oxford University Press; 2001.

347. Lau J, Ioannidis JPA, Terrin N, Schmid CH, Olkin I. The case of the misleading funnel plot. BMJ 2006; 333(7568): 597-600.

348. Lavis JN. How can we support the use of systematic reviews in policymaking? PLoS Med 2009; 6(11): e1000141.

349. Law AM, Kelton WD. Simulation modelling and analysis. Boston: McGraw Hill; 2000.

350. Law AM, McComas MG. How to build valid and credible simulation models. In: Peters BA, Smith JS, Medeiros DJ, Rohrer MW (Ed). Proceedings of the 2001 Winter Simulation Conference; 09.-12.09.2001; Arlington, USA. 2001. S. 22-29. URL: http://www.informs-sim.org/wsc01papers/004.PDF.

351. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Systematic reviews of diagnostic test accuracy. Ann Intern Med 2008; 149(12): 889-897.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

352. 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.

353. 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.

354. Leidl R, Graf von der Schulenburg JM, Wasem J (Ed). Ansätze und Methoden der ökonomischen Evaluation: eine internationale Perspektive. Baden-Baden: Nomos Verlagsgesellschaft; 1999. (Health Technology Assessments; Band 9).

355. Lewis S, Clarke M. Forest plots: trying to see the wood and the trees. BMJ 2001; 322(7300): 1479-1480.

356. Leys M. Health care policy: qualitative evidence and health technology assessment. Health Policy 2003; 65(3): 217-226.

357. 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.

358. 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.

359. 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.

360. Lijmer JG, Bossuyt PM, Various randomized designs can be used to evaluate medical tests. J Clin Epidemiol 2009; 62(4): 364-373.

361. Lijmer JG, Mol BW, Heisterkamp S, Bonsel 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.

362. Lipkus IM, Hollands JG. The visual communication of risk. J Natl Cancer Inst Monogr 1999; (25): 149-163.

363. Lipscomb J, Drummond M, Fryback D, Gold M, Revicki D. Retaining, and enhancing, the QALY. Value Health 2009; 12(Suppl 1): S18-S26.

364. Lipscomb J, Weinstein MC, Torrance GW. Time preference. In: Gold MR, Siegel JE, Russell LB, Weinstein MC (Ed). Cost-effectiveness in health and medicine. New York: Oxford University Press; 1996. S. 214-246.

365. Little RJA, Rubin DB. Statistical analysis with missing data. Hoboken: Wiley; 2002.

366. Lo B, Field MJ (Ed). Conflict of interest in medical research, education, and practice. Washington: National Academies Press; 2009.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

367. 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.

368. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med 2004; 23(20): 3105-3124.

369. Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. J Am Stat Assoc 2006; 101(474): 447-459.

370. 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.

371. Luce BR, Manning WG, Siegel JE, Lipscomb J. Estimating costs in cost-effectiveness analysis. In: Gold MR, Russell LB, Siegel JE, Weinstein MC (Ed). Cost-effectiveness in health and medicine. New York: Oxford University Press; 1996. S. 176-213.

372. Lumley T. Network meta-analysis for indirect treatment comparisons. Stat Med 2002; 21(16): 2313-2324.

373. Macaskill P, Walter SD, Irwig LM. A comparison of methods to detect publication bias in meta-analysis. Stat Med 2001; 20(4): 641-654.

374. 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.

375. Maetzel A. Der Gebrauch von Nutzwerten im gesundheitsökonomischen Vergleich von Interventionen bei verschiedenen Krankheitsbildern: eine Einführung. Z Rheumatol 2004; 63(5): 380-384.

376. Malterud K. The art and science of clinical knowledge: evidence beyond measures and numbers. Lancet 2001; 358(9279): 397-400.

377. Mandelblatt JS, Fryback DG, Weinstein MC, Russell LB, Gold MR, Hadorn DC. Assessing the effectiveness of health interventions. In: Gold MR, Siegel JE, Russell LB, Weinstein MC (Ed). Cost-effectiveness in health and medicine. New York: Oxford University Press; 1996. S. 135-175.

378. Mandrekar SJ, Sargent DJ. Clinical trial designs for predictive biomarker validation: theoretical considerations and practical challenges. J Clin Oncol 2009; 27(24): 4027-4034.

379. Mandrekar SJ, Sargent DJ. All-comers versus enrichment design strategy in phase II trials. J Thorac Oncol 2011; 6(4): 658-660.

380. Mangiapane S, Velasco Garrido M. Surrogatendpunkte als Parameter der Nutzenbewertung [online]. 2009 [Zugriff: 18.03.2015]. (Schriftenreihe Health Technology Assessment; Band 91). URL: http://portal.dimdi.de/de/hta/hta_berichte/hta250_bericht_de.pdf.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

381. 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.

382. Marsh K, Lanitis T, Neasham D, Orfanos P, Caro J. Assessing the value of healthcare interventions using multi-criteria decision analysis: a review of the literature. Pharmacoeconomics 2014; 32(4): 345-365.

383. Martini P. Methodenlehre der therapeutischen Untersuchung. Berlin: Springer; 1932.

384. 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.

385. Mauskopf JA, Earnshaw S, Mullins CD. Budget impact analysis: review of the state of the art. Expert Rev Pharmacoecon Outcomes Res 2005; 5(1): 65-79.

386. Mauskopf JA, Sullivan SD, Annemans L, Caro J, Mullins CD, Nuijten M et al. Principles of good practice for budget impact analysis: report of the ISPOR Task Force on Good Research Practices; budget impact analysis. Value Health 2007; 10(5): 336-347.

387. McAlister FA, Straus SE, Sackett DL, Altman DG. Analysis and reporting of factorial trials: a systematic review. JAMA 2003; 289(15): 2545-2553.

388. 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.

389. McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. BMJ 2002; 324(7351): 1448-1451.

390. 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.

391. McGregor M, Caro JJ. QALYs: are they helpful to decision makers? Pharmacoeconomics 2006; 24(10): 947-952.

392. McMurray J, Swedberg K. Treatment of chronic heart failure: a comparison between the major guidelines. Eur Heart J 2006; 27(15): 1773-1777.

393. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting recommendations for tumor marker prognostic studies (REMARK). J Natl Cancer Inst 2005; 97(16): 1180-1184.

394. Merlin T, Lehman S, Hiller JE, Ryan P. The "linked evidence approach" to assess medical tests: a critical analysis. Int J Technol Assess Health Care 2013; 29(3): 343-350.

395. 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.

396. 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

397. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009; 339: b2535.

398. Moher D, Pham B, Lawson ML, Klassen TP. The inclusion of reports of randomised trials published in languages other than English in systematic reviews. Health Technol Assess 2003; 7(41): 1-90.

399. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. PLoS Med 2007; 4(3): e78.

400. 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.

401. 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.

402. Mühlbacher AC, Bethge S, Tockhorn A. Präferenzmessung im Gesundheitswesen: Grundlage von Discrete-Choice-Experimenten. Gesundheitsökonomie & Qualitätsmanagement 2013; 18(4): 159-172.

403. Müller HP, Schmidt K, Conen D. Qualitätsmanagement: interne Leitlinien und Patientenpfade. Med Klin 2001; 96(11): 692-697.

404. 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.

405. 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.

406. National Advisory Committee on Health and Disability. Screening to improve health in New Zealand: criteria to assess screening. Wellington: National Health Committee; 2003. URL: https://www.nsu.govt.nz/system/files/resources/screening-to-improve-health.pdf.

407. 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.

408. 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.

409. National Institute for Health and Care Excellence. Guide to the processes of technology appraisal. London: NICE; 2014. URL: http://www.nice.org.uk/article/pmg19/resources/non-guidance-guide-to-the-processes-of-technology-appraisal-pdf.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

410. National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal. London: NICE; 2013. URL: http://www.nice.org.uk/article/pmg9/resources/non-guidance-guide-to-the-methods-of-technology-appraisal-2013-pdf.

411. 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.

412. 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.

413. Neidhardt K, Wasmuth T, Schmid A. Die Gewichtung multipler patientenrelevanter Endpunkte: ein methodischer Vergleich von Conjoint Analyse und Analytic Hierarchy Process unter Berücksichtigung des Effizienzgrenzenkonzepts des IQWiG; Diskussionspapier [online]. 02.2012 [Zugriff: 18.03.2015]. (Wirtschaftswissenschaftliche Diskussionspapiere; Band 02-12). URL: http://www.fiwi.uni-bayreuth.de/de/download/WP_02-12.pdf.

414. Nielsen J, Loranger H. Web Usability. München: Addison-Wesley; 2008.

415. 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.

416. Nord E. An alternative to QALYs: the saved young life equivalent (SAVE). BMJ 1992; 305(6858): 875-877.

417. Nord E. Cost-value analysis in health care: making sense out of QALYs. Cambridge: Cambridge University Press; 1999.

418. Nüesch E, Jüni P. Commentary: which meta-analyses are conclusive? Int J Epidemiol 2009; 38(1): 298-303.

419. Nutbeam D. Health promotion glossary. Health Promot Int 1998; 13(4): 349-364.

420. 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.

421. O'Mahony JF, Paulden M. NICE's selective application of differential discounting: ambiguous, inconsistent, and unjustified. Value Health 2014; 17(5): 493-496.

422. Odgaard-Jensen J, Vist GE, Timmer A, Kunz R, Akl EA, Schünemann H et al. Randomisation to protect against selection bias in healthcare trials. Cochrane Database Syst Rev 2011; (4): MR000012.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

423. Oliver A. A normative perspective on discounting health outcomes. J Health Serv Res Policy 2013; 18(3): 186-189.

424. 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.

425. Oostenbrink JB, Koopmanschap MA, Rutten FF. Standardisation of costs: the Dutch Manual for Costing in economic evaluations. Pharmacoeconomics 2002; 20(7): 443-454.

426. Orlewska E, Mierzejewski P. Proposal of Polish guidelines for conducting financial analysis and their comparison to existing guidance on budget impact in other countries. Value Health 2004; 7(1): 1-10.

427. Oxman AD, Guyatt GH. Guidelines for reading literature reviews. Can Med Assoc J 1988; 138(8): 697-703.

428. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol 1991; 44(11): 1271-1278.

429. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Ann Intern Med 1992; 116(1): 78-84.

430. 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.

431. 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).

432. 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.

433. Paulden M, Claxton K. Budget allocation and the revealed social rate of time preference for health. Health Econ 2012; 21(5): 612-618.

434. Pereira TV, Horwitz RI, Ioannidis JP. Empirical evaluation of very large treatment effects of medical interventions. JAMA 2012; 308(16): 1676-1684.

435. 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.

436. 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.

437. 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

438. Petkova E, Tarpey T, Huang L, Deng L. Interpreting meta-regression: application to recent controversies in antidepressants' efficacy. Stat Med 2013; 32(17): 2875-2892.

439. Pfaff H, Glaeske G, Neugebauer EA, Schrappe M. Memorandum III: Methoden für die Versorgungsforschung (Teil 1). Gesundheitswesen 2009; 71(8-9): 505-510.

440. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. Health Technol Assess 2004; 8(36): iii-iv, ix-xi, 1-158.

441. Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. JAMA 2012; 308(24): 2594-2604.

442. Pocock SJ. Clinical trials: a practical approach. Chichester: Wiley; 1983.

443. Poynard T, Munteanu M, Ratziu 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.

444. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Stat Med 1989; 8(4): 431-440.

445. 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.

446. Raftery JP. How should we value future health? Was NICE right to change? Value Health 2013; 16(5): 699-700.

447. 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.

448. 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.

449. 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.

450. Richardson J, Iezzi A, M.A. K, Maxwell A. Cross-national comparison of twelve quality of life instruments: MIC paper 2. Melbourne: Centre for Health Economics; 2012. (Research Papers; Band 78). URL: http://www.buseco.monash.edu.au/centres/che/pubs/researchpaper78.pdf.

451. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. BMJ 2011; 342: 964-967.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

452. Ringbaek T, Brøndum E, Martinez G, Lange P. EuroQoL in assessment of the effect of pulmonary rehabilitation COPD patients. Respir Med 2008; 102(11): 1563-1567.

453. 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.

454. 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.

455. Röhmel 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.

456. Ross SM. Simulation. Amsterdam: Elsevier Academic Press; 2006.

457. Rossouw JE. Estrogens for prevention of coronary heart disease: putting the brakes on the bandwagon. Circulation 1996; 94(11): 2982-2985.

458. Rothwell PM. Treating individuals 2: subgroup analysis in randomised controlled trials; importance, indications, and interpretation. Lancet 2005; 365(9454): 176-186.

459. 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.

460. Royal Society. Science and the public interest: communicating the results of new scientific research to the public [online]. 04.2006 [Zugriff: 18.03.2015]. URL: http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/publications/2006/8315.pdf.

461. Royall RM. The effect of sample size on the meaning of significance tests. Am Stat 1986; 40(4): 313-315.

462. Royston P. A strategy for modelling the effect of a continuous covariate in medicine and epidemiology. Stat Med 2000; 19(14): 1831-1847.

463. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. Appl Stat 1994; 43(3): 429-467.

464. 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.

465. Russell LB, Siegen JE, Daniels N, Gold MR, Luce BR, Mandelblatt JS. Cost-effectiveness analysis as a guide to resource allocation in health: roles and limitations. In: Gold MR, Siegel JE, Russell LB, Weinstein MC (Ed). Cost-effectiveness in health and medicine. New York: Oxford University Press; 1996. S. 3-24.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

466. Ryan M, Gerard K, Amaya-Amaya M (Ed). Using discrete choice experiments to value health and health care. Dordrecht: Springer; 2008. (The Economics of Non-Market Goods and Resources; Band 11).

467. Saaty T, Vargas LG. Decision making with the analytic network process: economic, political, social and technological applications with benefits, opportunities, costs and risks. New York: Springer Science and Business Media; 2006. (International Series in Operations Research Management Science; Band 95).

468. Saaty TL. A scaling method for priorities in hierarchical structures. J Math Psychol 1977; 15(3): 234-281.

469. Saaty TL. Theory and applications of the analytic network process: decision making with benefits, opportunities, costs and risks. Pittsburg: RWS Publications; 2005.

470. Saaty TL. Decision making with the Analytic Hierarchy Process. International Journal of Services Sciences 2008; 1(1): 83-98.

471. 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.

472. 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: 18.03.2015]. URL: http://www.svr-gesundheit.de/fileadmin/user_upload/Gutachten/2000-2001/Kurzf-de-01.pdf.

473. Sackett DL. Bias in analytic research. J Chronic Dis 1979; 32(1-2): 51-63.

474. 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.

475. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine: how to practice and teach EBM. Edinburgh: Churchill Livingstone; 2000.

476. Salanti G, Higgins JPT, Ades AE, Ioannidis JPA. Evaluation of networks of randomized trials. Stat Methods Med Res 2008; 17(3): 279-301.

477. 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.

478. 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.

479. Sampson M, McGowan J, Lefebvre C, Moher D, Grimshaw J. PRESS: Peer Review of Electronic Search Strategies. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2008. URL: http://www.cadth.ca/media/pdf/477_PRESS-Peer-Review-Electronic-Search-Strategies_tr_e.pdf.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

480. 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.

481. Sampson MJ. Updating searches for systematic reviews [Dissertation]. Aberystwyth: Universität; 2009.

482. Sänger S, Lang B, Klemperer D, Thomeczek C, Dierks ML. Manual Patienteninformation: Empfehlungen zur Erstellung evidenzbasierter Patienteninformationen. Berlin: Ärztliches Zentrum für Qualität in der Medizin; 2006. (ÄZQ-Schriftenreihe; Band 25). URL: http://www.aezq.de/mdb/edocs/pdf/schriftenreihe/schriftenreihe25.pdf.

483. 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.

484. 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.

485. Sargent DJ, Mandrekar SJ. Statistical issues in the validation of prognostic, predictive, and surrogate biomarkers. Clin Trials 2013; 10(6): 647-652.

486. Sargent RG. Validation and verification of simulation models. In: Ingalls RG, Rossetti MD, Smith JS, Peters BA (Ed). Proceedings of the 2004 Winter Simulation Conference; 05.-08.12.2004; Washington, USA. 2004. S. 17-28. URL: http://www.informs-sim.org/wsc04papers/004.pdf.

487. SAS Institute. SAS/STAT 9.2 user's guide. second edition [online]. 2009 [Zugriff: 18.03.2015]. URL: http://support.sas.com/documentation/cdl/en/statug/63033/PDF/default/statug.pdf.

488. 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.

489. Savović J, Jones HE, Altman DG, Harris RJ, Jűni P, Pildal J et al. Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. Health Technol Assess 2012; 16(35): 1-82.

490. 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.

491. Scherer RW, Langenberg P, Von Elm E. Full publication of results initially presented in abstracts. Cochrane Database Syst Rev 2007; (2): MR000005.

492. Schluter PJ, Ware RS. Single patient (n-of-1) trials with binary treatment preference. Stat Med 2005; 24(17): 2625-2636.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

493. 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.

494. Schöffski O. Grundformen gesundheitsökonomischer Evaluationen. In: Schöfski O, Graf von der Schulenburg JM (Ed). Gesundheitsökonomische Evaluationen. Berlin: Springer; 2012. S. 43-70.

495. Schöffski O, Graf von der Schulenburg JM (Ed). Gesundheitsökonomische Evaluationen. Berlin: Springer; 2012.

496. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010; 340: c332.

497. 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.

498. 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.

499. 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.

500. Sculpher M. The role and estimation of productivity costs in economic evaluation. In: Drummond MF, McGuire A (Ed). Economic evaluation in health care: merging theorey with practice. Oxford: Oxford University Press; 2001. S. 94-112.

501. Sculpher MJ, O'Brien BJ. Income effects of reduced health and health effects of reduced income: implications for health state valuation. Med Decis Making 2000; 20(2): 207-215.

502. Senn SJ. Inherent difficulties with active control equivalence studies. Stat Med 1993; 12(24): 2367-2375.

503. Senn SJ. The many modes of meta. Drug Inf J 2000; 34(2): 535-549.

504. Senn SJ. Trying to be precise about vagueness. Stat Med 2007; 26(7): 1417-1430.

505. 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.

506. 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.

507. 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

508. Shechter SM, Schaefer AJ, Braithwaite RS, Roberts MS. Increasing the efficiency of Monte Carlo cohort simulations with variance reduction techniques. Med Decis Making 2006; 26(5): 550-553.

509. 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.

510. 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.

511. Siebert U. Entscheidungsanalytische Modelle zur Sicherung der Übertragbarkeit internationaler Evidenz von HTA auf den Kontext des deutschen Gesundheitssystems: ein Methodenbeitrag zu HTA [online]. 2005 [Zugriff: 18.03.2015]. (Schriftenreihe Health Technology Assessment; Band 16). URL: http://portal.dimdi.de/de/hta/hta_berichte/hta099_bericht_de.pdf.

512. 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.

513. 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.

514. 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.

515. Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. Br J Cancer 1994; 69(6): 979-985.

516. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. J Natl Cancer Inst 2009; 101(21): 1446-1452.

517. Siontis KC, Siontis GCM, Contopoulos-Ioannidis DG, Ioannidis JPA. Diagnostic tests often fail to lead to changes in patient outcomes. J Clin Epidemiol 2014; 67(6): 612-621.

518. Skipka G, Bender R. Interpretation effects in the case of heterogeneity between three subgroups: assessment within the framework of systematic reviews. Methods Inf Med 2010; 49(6): 613-617.

519. Smeeth L, Haines A, Ebrahim S. Numbers needed to treat derived from meta-analyses: sometimes informative, usually misleading. BMJ 1999; 318(7197): 1548-1551.

520. Sölétormos G, Duffy MJ, Hayes DF, Sturgeon CM, Barak V, Bossuyt PM et al. Design of tumor biomarker-monitoring trials: a proposal by the European Group on Tumor Markers. Clin Chem 2013; 59(1): 52-59.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

521. 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.

522. 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.

523. Spiegelhalter DJ, Freedman LS, Parmar MK. Bayesian approaches to randomized trials. J R Stat Soc Ser A 1994; 157(3): 357-416.

524. 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.

525. Statistisches Bundesamt. Preise: Harmonisierter Verbraucherpreisindex; September 2013 [online]. 11.10.2013 [Zugriff: 17.10.2013]. URL: https://www.destatis.de/DE/Publikationen/Thematisch/Preise/Verbraucherpreise/HarmonisierteVerbraucherpreisindizesPDF_5611201.pdf?__blob=publicationFile.

526. Statistisches Bundesamt. Statistik der schwerbehinderten Menschen 2007: Kurzbericht [online]. 01.2009 [Zugriff: 18.03.2015]. URL: https://www.destatis.de/DE/Publikationen/Thematisch/Gesundheit/BehinderteMenschen/SozialSchwerbehinderteKB5227101079004.pdf?__blob=publicationFile.

527. Steckelberg A, Berger B, Köpke S, Heesen C, Mühlhauser I. Kriterien für evidenzbasierte Patienteninformationen. Z Ärztl Fortbild Qualitatssich 2005; 99(6): 343-351.

528. Steiner JF. The use of stories in clinical research and health policy. JAMA 2005; 294(22): 2901-2904.

529. 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.

530. Sterne JAC, Egger M, Moher D. Addressing reporting biases. In: Higgins JPT, Green S (Hrsg). Cochrane handbook for systematic reviews of interventions. Chichester: Wiley; 2008. S. 297-333.

531. Steyerberg EW, Moons KG, Van der Windt DA, Hayden JA, Perel P, Schroter S et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS Med 2013; 10(2): e1001381.

532. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Med Decis Making 1998; 18(2 Suppl): S68-S80.

533. Stollenwerk B, Lhachimi SK, Briggs A, Fenwick E, Caro JJ, Siebert U. Communicating the parameter uncertainty in the IQWiG efficiency frontier to decision-makers. Health Econ 04.03.2014 [Epub ahead of print].

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

534. 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.

535. 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.

536. 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.

537. Sutton A, Ades AE, Cooper N, Abrams K. Use of indirect and mixed treatment comparisons for technology assessment. Pharmacoeconomics 2008; 26(9): 753-767.

538. 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.

539. Swift TL, Dieppe PA. Using expert patients' narratives as an educational resource. Patient Educ Couns 2005; 57(1): 115-121.

540. Tainio M, Tuomisto JT, Hänninen O, Ruuskanen J, Jantunen MJ, Pekkanen J. Parameter and model uncertainty in a life-table model for fine particles (PM2.5): a statistical modeling study. Environ Health 2007; 6: 24.

541. Tajik P, Zwinderman AH, Mol BW, Bossuyt PM. Trial designs for personalizing cancer care: a systematic review and classification. Clin Cancer Res 2013; 19(17): 4578-4588.

542. Takwoingi Y, Leeflang MMG, Deeks JJ. Empirical evidence of the importance of comparative studies of diagnostic test accuracy. Ann Intern Med 2013; 158(7): 544-554.

543. Tan SS. Microcosting in economic evaluations: issues of accuracy, feasibility,consistency and generalisability [Dissertation]. Rotterdam: Erasmus Universität; 2009. URL: http://repub.eur.nl/res/pub/17254/091127_Tan,%20Siok%20Swan.pdf.

544. Tan SS, Bouwmans CAM, Rutten FFH, Hakkaart-van Roijen L. Update of the Dutch manual for costing in economic evaluations. Int J Technol Assess Health Care 2012; 28(2): 152-158.

545. 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.

546. Thomas S. Klinische Relevanz von Therapieeffekten: systematische Sichtung, Klassifizierung und Bewertung methodischer Konzepte [Dissertation]. Duisburg/Essen: Universität; 2009.

547. Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? Stat Med 2002; 21(11): 1559-1573.

548. 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

549. 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.

550. Thurow S. Search engine visibility. Indianapolis: New Riders; 2003.

551. 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.

552. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Series B Stat Methodol 1996; 58(1): 267-288.

553. Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. Trials 2007; 8: 16.

554. Torgerson DJ. Contamination in trials: is cluster randomisation the answer? BMJ 2001; 322(7282): 355-357.

555. Torrance GW, Siegel JE, Luce BR, Gold MR, Russell LB, Weinstein MC. Framing and designing the cost-effectiveness analysis. In: Gold MR, Siegel JE, Russell LB, Weinstein MC (Ed). Cost-effectiveness in health and medicine. New York: Oxford University Press; 1996. S. 54-81.

556. 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.

557. 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.

558. Trueman P, Drummond M, Hutton J. Developing guidance for budget impact analysis. Pharmacoeconomics 2001; 19(6): 609-621.

559. Tsay MY, Yang YH. Bibliometric analysis of the literature of randomized controlled trials. J Med Libr Assoc 2005; 93(4): 450-458.

560. Tukey JW. Some thoughts on clinical trials, especially problems of multiplicity. Science 1977; 198(4318): 679-684.

561. 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.

562. 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.

563. 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).

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

564. UK National Screening Committee. Programme appraisal criteria: criteria for appraising the viability, effectiveness and appropriateness of a screening programme [online]. [Zugriff: 18.03.2015]. URL: http://www.screening.nhs.uk/criteria.

565. USAID Center for Development Information and Evaluation. Conducting key informant interviews [online]. 1996 [Zugriff: 18.03.2015]. (Performance Monitoring and Evaluation TIPS; Band 2). URL: http://pdf.usaid.gov/pdf_docs/PNABS541.pdf.

566. Van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. Stat Med 2002; 21(4): 589-624.

567. Van Tinteren H, Hoekstra OS, Boers M. Do we need randomised trials to evaluate diagnostic procedures? Eur J Nucl Med Mol Imaging 2004; 31(1): 129-131.

568. 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.

569. 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.

570. Vandenbroucke JP, Von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ et al. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. Ann Intern Med 2007; 147(8): W163-W194.

571. Veerman JL, Mackenbach JP, Barendregt JJ. Validity of predictions in health impact assessment. J Epidemiol Community Health 2007; 61(4): 362-366.

572. Vidanapathirana J, Abramson MJ, Forbes A, Fairley C. Mass media interventions for promoting HIV testing. Cochrane Database Syst Rev 2005; (3): CD004775.

573. Vijan S. Should we abandon QALYs as a resource allocation tool? Pharmacoeconomics 2006; 24(10): 953-957.

574. 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.

575. Vincent H, Leino-Kilpi H, Salantera S. Empowering discourse in patient education. Patient Educ Couns 2007; 66(2): 140-146.

576. Viswanathan M, Ansari MT, Berkman ND, Chang S, Hartling L, McPheeters M et al. Assessing the risk of bias of individual studies in systematic reviews of health care interventions: Agency for Healthcare Research and quality methods guide for comparative effectiveness reviews [online]. 03.2012 [Zugriff: 18.03.2015]. URL: http://effectivehealthcare.ahrq.gov/ehc/products/322/998/MethodsGuideforCERs_Viswanathan_IndividualStudies.pdf.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

577. 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.

578. 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.

579. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ 2007; 335(7624): 806-808.

580. Waffenschmidt S, Janzen T, Hausner E, Kaiser T. Simple search techniques in PubMed are potentially suitable for evaluating the completeness of systematic reviews. J Clin Epidemiol 2013; 66(6): 660-665.

581. 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.

582. 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.

583. Web Accessibility Initiative. Webauftritt [online]. [Zugriff: 18.03.2015]. URL: http://www.w3.org/WAI.

584. Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C et al. Principles of good practice for decision analytic modeling in health care evaluation: report of the ISPOR Task Force on Good Research Practices: Modeling studies. Value Health 2003; 6(1): 9-17.

585. Weinstein MC, Siegel JE, Garber AM, Lipscomb J, Luce BR, Manning WG et al. Productivity costs, time costs and health-related quality of life: a response to the Erasmus Group. Health Econ 1997; 6(5): 505-510.

586. Weir CJ, Walley RJ. Statistical evaluation of biomarkers as surrogate endpoints: a literature review. Stat Med 2006; 25(2): 183-203.

587. Wendt C. Gesundheitssysteme im internationalen Vergleich. Gesundheitswesen 2006; 68(10): 593-599.

588. West S, King V, Carey TS, Lohr KN, McKoy N, Sutton SF et al. Systems to rate the strength of scientific evidence: AHRQ publication no. 02-E016 [online]. 03.2002 [Zugriff: 16.04.2014]. (Evidence Report/Technology Assessment (Summaries); Band 47). URL: http://archive.ahrq.gov/clinic/epcsums/strengthsum.pdf.

589. 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

590. Whitehead J. The design and analysis of sequential clinical trials. Chichester: Horwood; 1983.

591. 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.

592. Whiting PF, Rutjes AW, Westwood ME, Mallett S. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. J Clin Epidemiol 2013; 66(10): 1093-1104.

593. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011; 155(8): 529-536.

594. 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.

595. 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: 18.03.2015]. URL: http://www.gesundheitsforschung-bmbf.de/_media/DLR_Nutzenbewert_07-11-22_Druckversion.pdf.

596. Windeler J. Externe Validität. Z Evid Fortbild Qual Gesundhwes 2008; 102(4): 253-259.

597. Windeler J, Conradt C. Wie können "Signifikanz" und "Relevanz" verbunden werden? Med Klin 1999; 94(11): 648-651.

598. Windeler J, Lange S. Nutzenbewertung in besonderen Situationen: seltene Erkrankungen. Z Evid Fortbild Qual Gesundhwes 2008; 102(1): 25-30.

599. Windeler J, Ziegler S. Evidenzklassifizierungen. Z Arztl Fortbild Qualitatssich 2003; 97(6): 513-514.

600. Wood L, Egger M, Gluud LL, Schulz KF, Jüni 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.

601. 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.

602. 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"

603. 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.

604. 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.

ABGELÖST DURCH "Allgemeine Methoden Version 5.0 vom 10.07.2017"