

Methodenhandbuch für Health Technology Assessment Version 1.2012

Wissenschaftlicher Ergebnisbericht

Methodenhandbuch für Health Technology Assessment Version 1.2012

Eine Kooperation von Donau-Universität Krems (DUK), EBM Review Center Graz,
Gesundheit Österreich GmbH (GÖG), Ludwig Boltzmann Institut für HTA (LBI-HTA)
und Privater Universität für Gesundheitswissenschaften, Medizinische Informatik und Technik (UMIT)

Barbara Fröschl / GÖG (Projektkoordination)

Bernhard Bornschein / UMIT

Sophie Brunner-Ziegler / GÖG

Annette Conrads-Frank / UMIT

Alexander Eisenmann / GÖG

Gerald Gartlehner / DUK

Ludwig Grillich / DUK

Angela Kaminski / DUK

Nikolai Mühlberger / UMIT

Daniela Pertl / GÖG

Petra Schnell-Inderst / UMIT

Gaby Sroczynski / UMIT

Uwe Siebert / UMIT

Kylie Thaler / DUK

Claudia Wild / LBI-HTA

Johannes Wurm / UMIT

Ingrid Zechmeister / LBI-HTA

Johannes Zsifkovits / GÖG

Begutachtung:

Karl Horvath / EBM Review Center Graz

Ursula Püringer / EBM Review Center Graz

Ulrich Grouven

Tobias Kurth

Dagmar Lühmann / Institut für Sozialmedizin, Lübeck

Enno Swart / Institut für Sozialmedizin und Gesundheitsökonomie (ISMHE)

Anja Neumann / Universität Duisburg-Essen

Projektassistenz:

Romana Landauer

Wien, im März 2012

Im Auftrag des Bundesministeriums für Gesundheit

Die GÖG bedankt sich bei allen Kooperationspartnern für die gute Zusammenarbeit!



ISBN 978-3-85159-155-2

Eigentümer, Herausgeber und Verleger: Gesundheit Österreich GmbH, Stubenring 6,
1010 Wien, Tel. +43 1 515 61, Fax 513 84 72, Homepage: www.goeg.at

Inhalt

Tabellen	VII
Abbildungen.....	IX
Boxen	XI
Abkürzungsverzeichnis	XII
Einleitung.....	1
1 Aufbau des Methodenhandbuchs	2
2 HTA-Produkte	5
2.1 HTA-Bericht.....	5
2.2 Rapid Assessment.....	5
2.3 Adaption bestehender Assessments	6
2.4 Budgetauswirkungsanalyse	7
3 Themenfindung und Priorisierung.....	8
3.1 Themenfindung	8
3.2 Priorisierung.....	9
4 Methodik der Durchführung eines Health Technology Assessment	10
4.1 Entwickeln der Fragestellung.....	10
4.1.1 Vorabrecherche von Hintergrundinformationen (Scoping)	10
4.1.2 Operationalisieren der Fragestellung	10
4.2 Definition der Selektionskriterien	15
4.3 Erstellen des Hintergrunds	16
4.3.1 Erstellen des gesundheitspolitischen Hintergrunds	16
4.3.2 Erstellen des wissenschaftlichen Hintergrunds.....	17
4.4 Erstellen eines Berichtsplans	18
4.5 Literatursuche.....	19
4.5.1 HTA-Bericht / Systematischer Review	20
4.5.2 Rapid Assessment.....	22
4.5.3 Adaption bestehender HTA-Berichte und Rapid Assessments.....	24
4.6 Literaturselektion.....	25
4.6.1 HTA-Bericht / Systematischer Review	26
4.6.2 Rapid Assessment.....	27
4.6.3 Adaption bestehender Assessments	27
4.7 Exkurs: Klassifikation von Studien	28
4.8 Exkurs: Hierarchie der Studien	34
5 Zusätzliche Datengenerierung und Datenanalyse.....	35
5.1 Grundlagen.....	35
5.1.1 Grundprinzipien sozialwissenschaftlicher Methoden	35
5.1.2 Interviews	37

5.1.3	Fragebogenerhebung.....	41
5.1.4	Zusammenfassende Darstellung der Methoden	43
5.2	Einsatz von Routinedaten im Rahmen von Health Technology Assessments	44
5.2.1	Einleitung: Routinedaten und HTA	44
5.2.2	Methode.....	45
5.2.3	Definitionen und Abgrenzung des Themas.....	46
5.2.4	Einsatz von Routinedaten im HTA-Prozess.....	49
5.2.5	Methoden zur Analyse von Routinedaten im Rahmen von HTA.....	60
5.2.6	Routinedaten in Österreich	70
5.2.7	Fazit.....	74
5.2.8	Danksagung	75
6	Beurteilung der medizinischen Studien und Extraktion der Daten.....	76
6.1	Beurteilung des Bias-Risikos (interne Validität)	76
6.1.1	Therapie.....	78
6.1.2	Qualitative Studien	92
6.1.3	Diagnostische Studien	95
6.1.4	Präventionsstudien	99
6.2	Beurteilung der externen Validität	101
6.2.1	Pragmatische Studien	101
6.3	Datenextraktion	104
7	Synthese der medizinischen Evidenz.....	107
7.1	Qualitative Zusammenfassung.....	107
7.2	Effektgrößen	109
7.2.1	Dichotome Effektmaße	110
7.2.2	Stetige Effektmaße.....	113
7.2.3	Inzidenzmaße.....	114
7.2.4	Überlebenszeitanalysen	114
7.3	Quantitative Zusammenfassung.....	115
7.3.1	Metaanalysen	115
7.3.2	Entscheidungsanalytische Nutzenmodellierung	146
7.4	Stärke der Evidenz.....	152
8	Ökonomische Bewertung.....	155
8.1	Ökonomische Übersichtsarbeit	155
8.1.1	Hintergrund	155
8.1.2	Exkurs: Grundlagen gesundheitsökonomischer Evaluation	156
8.1.3	Übersicht zu vorhandenen Methodenhandbüchern	158
8.1.4	Durchführung gesundheitsökonomischer Übersichtsarbeiten	159
8.1.5	Übertragbarkeit und Adaption von Ergebnissen vom Studien- auf den Entscheidungskontext.....	185
8.1.6	Resümee	195
8.2	Budgetauswirkungsanalyse.....	196
8.2.1	Definition	196
8.2.2	Abgrenzung der Budgetauswirkungsanalyse zu anderen gesundheitsökonomischen Studien.....	197
8.2.3	Erstellung und Durchführung einer Budgetauswirkungsanalyse.....	199
8.3	Gesundheitsökonomische Modellierung.....	202
8.3.1	Rahmenbedingungen des Modells.....	203

8.3.2	Vorgehen bei der Modellentwicklung	203
8.3.3	Modelltypen	209
8.3.4	Basisfall- und Sensitivitätsanalyse	219
8.3.5	Qualitätskriterien für die Modellierung	225
8.4	Ökonomische Empfehlung	227
8.4.1	Einleitung	227
8.4.2	Ökonomische Empfehlung im Detail	229
9	Psychosoziale, ethische, rechtliche und organisatorische Aspekte	241
9.1	Psychosoziale Aspekte	243
9.2	Ethische Aspekte	245
9.3	Rechtliche Aspekte.....	248
9.4	Organisatorische Aspekte	249
10	Qualitätssicherung.....	253
11	Implementierung und Impact	254
11.1	Implementierungsvorschläge	254
11.2	Impact	254
12	Ausblick	256
13	Glossar	257
14	Literaturverzeichnis	263
14.1	Literatur zu Kapitel 2 HTA-Produkte	263
14.2	Literatur zu Kapitel 3 Themenfindung und Priorisierung	263
14.3	Literatur zu Kapitel 4 Methodik der Durchführung eines Health Technology Assessment.....	263
14.4	Literatur zu Kapitel 5 Zusätzliche Datengenerierung und Datenanalyse	267
14.4.1	Literatur zu Kapitel 5.1 Grundlagen.....	267
14.4.2	Literatur zu Kapitel 5.2. Einsatz von Routinedaten im Rahmen von Health Technology Assessments	268
14.5	Literatur zu Kapitel 6 Beurteilung der medizinischen Studien und Extraktion der Daten und Kapitel 7 Synthese der medizinischen Evidenz.....	274
14.6	Literatur zu Kapitel 8 Ökonomische Bewertung.....	287
14.6.1	Literatur zu Kapitel 8.1. Ökonomische Übersichtsarbeit.....	287
14.6.2	Literatur zu Kapitel 8.2 Bugdetauswirkungsanalyse.....	303
14.6.3	Literatur zu Kapitel 8.3 Gesundheitsökonomische Modellierung.....	305
14.7	Literatur zu Kapitel 9 Soziale, ethische, rechtliche und organisatorische Aspekte	312
14.8	Literatur zu Kapitel 10 Qualitätssicherung.....	314
14.9	Literatur zu Kapitel 11 Implementierung und Impact	314
15	Anhang A	315
15.1	Methodenarbeitsgruppe.....	315
15.1.1	Department für Evidenzbasierte Medizin und Klinische Epidemiologie an der Donau-Universität Krems	315
15.1.2	EBM Review Center Graz	316

15.1.3	Gesundheit Österreich GmbH / Geschäftsbereich BIQG.....	316
15.1.4	Ludwig Boltzmann Institut Health Technology Assessment.....	317
15.1.5	UMIT	318
15.2	Nationale und internationale Methodenhandbücher	319
15.3	Quellen für Datenbanken.....	320
15.3.1	Quellen für die systematische Literatursuche	320
15.3.2	Quellen für die ergänzende Suche.....	321
15.4	Tabellarische Übersicht zu Datenbeständen	321
15.5	Dateneigenschaften ausgewählter Datenbestände.....	337
15.5.1	Administrative Daten	337
15.5.2	Register.....	348
15.5.3	Surveys	371
15.5.4	Amtliche Statistiken.....	373
15.5.5	Daten zu Planungszwecken.....	373
15.5.6	Daten zu Versorgungsstrukturen	373
15.5.7	Sonstige	374
15.5.8	Kompasse	376
15.5.9	Verlinkung unterschiedlicher administrativer Daten am Beispiel des GAP-DRG-Projekts	379
15.6	Checklisten zur Qualitätsbewertung der Studien	379
15.6.1	Checkliste zur Beurteilung von randomisierten kontrollierten Studien (RCT).....	380
15.6.2	Checkliste zur Beurteilung von Kohortenstudien	381
15.6.3	Checkliste zur Beurteilung von Fall-Kontroll-Studien	382
15.6.4	Checkliste zur Beurteilung von systematischen Reviews und Metaanalysen.....	383
15.6.5	Checkliste zur Beurteilung von diagnostischen Studien	384
15.7	Checklisten und Tabellen zum Kapitel ökonomische Übersichtsarbeit	385
15.8	Beispiel für eine Budgetauswirkungsanalyse.....	403

Tabellen

Tabelle 1.1:	HTA-Erstellungsprozess	2
Tabelle 4.1:	PICO-Prinzip zur Wirksamkeit von antimikrobiellen Präparaten zur Wundheilung (Beispiel).....	12
Tabelle 4.2:	PICO-Prinzip zur Diagnosegenauigkeit der automatischen Blutdruckmessung am Handgelenk (Beispiel).....	13
Tabelle 4.3:	PICO-Prinzip zum COPD-Screening (Beispiel)	13
Tabelle 4.4:	PICO-Prinzip zur Kosteneffektivität Langzeit-Analoginsulin (Beispiel).....	13
Tabelle 4.5:	PICO-Prinzip zur Sensitivität und Spezifität von Assessmentinstrumenten (Beispiel)	14
Tabelle 4.6:	PICO-Prinzip zu organisatorischen Modellen in Intensivstationen (Beispiel).....	14
Tabelle 4.7:	Überblick über Vor- und Nachteile unterschiedlicher Studiendesigns	32
Tabelle 4.8:	Vereinfachte Darstellung der Hierarchie der Evidenz unterschiedlicher Studientypen in absteigender Reihenfolge.....	34
Tabelle 5.1:	Übersicht über die dargestellten Methoden	43
Tabelle 5.2:	Möglichkeiten für den Einsatz von Routinedaten im HTA-Prozess	55
Tabelle 5.3:	Übersicht zur Datenbeschreibung	71
Tabelle 6.1:	Klassifizierung des Bias-Risikos	77
Tabelle 6.2:	Unterschiede zwischen explanatorischen und pragmatischen Studiendesigns	102
Tabelle 7.1:	Beispiel eines Evidenzprofils	108
Tabelle 7.2:	Vierfeldertafel zur Berechnung von Effektmaßen für dichotome Zielvariablen	110
Tabelle 7.3:	Beispiel für einen Forest Plot	141
Tabelle 7.4:	Beispiel für eine kumulative Metaanalyse.....	142
Tabelle 7.5:	Übersicht verfügbarer Software zur Durchführung von Metaanalysen.....	145
Tabelle 7.6:	Faktoren, die die Stärke der Evidenz beeinflussen.....	153
Tabelle 8.1:	Klinischer Ereignispfad.....	161

Tabelle 8.2:	Beispielfragen	162
Tabelle 8.3:	Beispiel für Ein- und Ausschlusskriterien	164
Tabelle 8.4:	Beispiel für Übersichtstabelle; Übersichtsarbeit aus dem Bereich der Kinder- und Jugendpsychiatrie	174
Tabelle 8.5:	Permutationsmatrix	176
Tabelle 8.6:	Hierarchische Matrix	177
Tabelle 8.7:	Beispiel für ein gesundheitsökonomisches Evidenzprofil auf Basis der GRADE-Methodologie bei Drogenersatztherapiestudien	180
Tabelle 8.8:	Beispiel für NICE-Evidenzsynthese nach GRADE-Methodologie für gesundheitsökonomische Studien zu Omega-3-Säurenethylester- Ergänzungen nach Herzinfarkt	183
Tabelle 8.9:	Übertragbarkeitsfaktoren von (Welte, Feenstra et al. 2004).....	186
Tabelle 8.10:	Beispiel für systematische Überprüfung der Übertragbarkeit: Transferability of the Belgium Netherlands Stent Study II to Germany	190
Tabelle 8.11:	Unterschiede zwischen Budgetauswirkungsanalysen und gesundheitsökonomischen Studien.....	198
Tabelle 8.12:	Vereinfachtes Beispiel zur Berechnung der Budgetauswirkung (Angabe zur Berechnung)	201
Tabelle 8.13:	Vereinfachtes Beispiel zur Berechnung der Budgetauswirkung (Berechnung).....	201
Tabelle 9.1:	Framework zur systematischen Analyse soziokultureller Implikationen medizinischer Technologien (beispielhaft).....	243
Tabelle 15.1:	Übersicht zu Datenbeständen	322
Tabelle 15.2:	Beispiel für Datenextraktionsformular für gesundheitsökonomische Evaluationen.....	385
Tabelle 15.3:	BMJ-Checkliste.....	386
Tabelle 15.4:	CHEC-Checkliste	387
Tabelle 15.5:	Checkliste zur Beurteilung der methodischen Qualität gesundheitsökonomischer Studien (entwickelt im Konsensusverfahren von den gesundheitsökonomischen Projektgruppen München, Hannover, Ulm)	388
Tabelle 15.6:	Checkliste (Philips, Ginnelly 2004)	391
Tabelle 15.7:	Evidenzsynthese zu Effektivität und Kosteneffektivität von HIV- Präventionsmaßnahmen bei unterschiedlichen Subgruppen.....	395

Tabelle 15.8:	GDP-PPP (1980–2008), Nationale Währung in US \$ (Teil 1)	397
Tabelle 15.9:	GDP-PPP (1980–2008), Nationale Währung in US \$ (Teil 2)	398
Tabelle 15.10:	GDP-PPP (1980–2008), Nationale Währung in US \$ (Teil 3)	399
Tabelle 15.11:	Bruttoinlandsprodukt – Preis Index (2000 = 100); Teil 1	400
Tabelle 15.12:	Bruttoinlandsprodukt – Preis Index (2000 = 100); Teil 2	401
Tabelle 15.13:	Bruttoinlandsprodukt – Preis Index (2000 = 100); Teil 3	402
Tabelle 15.14:	Beispiel zur Berechnung der Budgetauswirkung (Angabe)	404
Tabelle 15.15:	Voraussichtliche jährliche Fälle im Jahr 2009 und Inzidenzraten virusbedingter Erkrankungen bei Kindern unter 5 Jahren (Berechnung)	406
Tabelle 15.16:	Voraussichtliche Kosten für die Gesellschaft im Jahr 2009 für die Behandlung virusbedingter Erkrankungen bei Kindern unter 5 Jahren ohne Einführung eines Impfprogramms zur Virusimmunisierung (Berechnung)	407
Tabelle 15.17:	Voraussichtliche Fälle in fünf Jahren bei Einführung eines Impfprogramms zur Virusimmunisierung im Vergleich zur Nichteinführung* (Berechnung)	407
Tabelle 15.18:	Voraussichtliche Budgetauswirkung in fünf Jahren bei Einführung eines Impfprogramms zur Virusimmunisierung* (Berechnung)	408

Abbildungen

Abbildung 4.1:	Darstellung zur Einteilung von Studientypen	30
Abbildung 7.1:	Anzahl der Publikationen im Bereich Metaanalyse in MEDLINE/PubMed	115
Abbildung 7.2:	Rehospitalisierungen bei Disease-Management-Programmen im Vergleich zur Standardbehandlung	117
Abbildung 7.3:	Funnel Plot und gepoolter Punktschätzer vor und nach Anwendung des „Trim & Fill“-Ansatzes	128
Abbildung 7.4:	Beispiel für einen Funnel Plot	143
Abbildung 8.1:	Vier-Felder-Matrix	178
Abbildung 8.2:	Zu berücksichtigende Aspekte in einer Budgetauswirkungsanalyse	200

Abbildung 8.3:	Entscheidungsbaum zur Anwendung eines diagnostischen Tests vor Behandlung.....	210
Abbildung 8.4:	Zustandsübergangsdiagramm für das Markov-Modell einer chronischen Krankheit mit Übergangswahrscheinlichkeiten	212
Abbildung 8.5:	Darstellung des Ablaufs eines Markov-Modells für eine chronische Erkrankung	213
Abbildung 8.6:	Diskrete Ereignissimulation, Flussdiagramm für ein Organtransplantationsmodell.....	215
Abbildung 8.7:	Diskrete Ereignissimulation, Ermittlung der Überlebenszeit für Patienten in der Simulation	216
Abbildung 8.8:	Zervixkrebs-Screeningstrategien, Effizienzgrenze	220
Abbildung 8.9:	Tornado-Diagramm von Einweg-Sensitivitätsanalysen zum inkrementellen Kosten-Effektivitäts-Verhältnis einer Behandlung mit pegyliertem Interferon und Ribavirin im Vergleich zur Behandlung mit Interferon und Ribavirin	223
Abbildung 8.10:	Kosten-Nutzwert-Akzeptanzkurven zu drei diagnostischen Strategien für die Langzeitbehandlung von Patienten mit Sodbrennen	225

Boxen

Box 5.1:	Einsatz von Routinedaten für Priorisierung: Projektbeispiel „Hämocomplettan® und Fibrogammin® bei erworbenen Fibrinogenmangelzuständen“	56
Box 5.2:	Einsatz von Routinedaten für die Statusbestimmung einer Technologie: Projektbeispiel „Halten Statine, was sie versprechen? Prognostizierte und tatsächliche Auswirkungen auf kardiovaskuläre Erkrankungen in Österreich“	57
Box 5.3:	Einsatz von Routinedaten für die Beschreibung der Krankheitslast: Projektbeispiel „Eltern–Kind–Vorsorge neu. Teil I: Epidemiologie – Häufigkeiten von Risikofaktoren und Erkrankungen in Schwangerschaft und früher Kindheit“	57
Box 5.4:	Einsatz von Routinedaten für die Evaluierung der Effektivität einer Technologie: Projektbeispiel „Ambulante kardiologische Rehabilitation“ ...	58
Box 5.5:	Einsatz von Routinedaten für die Kostenberechnung: Projektbeispiel „Ökonomische Evaluation der HPV–Impfung“	59
Box 5.6:	Einsatz von Routinedaten für Impact–Evaluierung: Projektbeispiel „Auswirkung der HTA–Forschung auf das Gesundheitswesen in Österreich“	60
Box 5.7:	Umgang mit Confounding am Projektbeispiel „Ambulante kardiologische Rehabilitation“	67
Box 8.1:	Studientypen von gesundheitsökonomischen Evaluationen.....	157
Box 8.2:	Mögliche Ziele gesundheitsökonomischer Übersichtsarbeiten	160
Box 8.3:	Beispiel für die qualitative Beschreibung der Übertragbarkeit (Faktor „Produktivität“)	189
Box 8.4:	Inflationsbereinigung mit Preisindizes	192
Box 8.5:	Beispiel für Währungskonversion	193

Abkürzungsverzeichnis

BGK	Bundesgesundheitskommission
BIQG	Bundesinstitut für Qualität im Gesundheitswesen (Geschäftsbereich der Gesundheit Österreich GmbH)
BMG	Bundesministerium für Gesundheit
CDC	Centers for Disease Control and Prevention
CINAHL	The Cumulative Index to Nursing and Allied Health Literature / Index zur pflegerelevanten Literatursuche
DACEHTA	Danish Centre for Health Technology Assessment
DIMDI	Deutsches Institut für Medizinische Dokumentation und Information
DLD	Diagnosen- und Leistungsdokumentation
EBM	Evidence-based medicine
EMA	European Medicines Agency
EMS	Epidemiologisches Meldesystem
EU	Europäische Union
FDA	Food and Drug Administration
GeroLit	Gerontologische Literatur
GÖG/BIQG	Gesundheit Österreich GmbH / Geschäftsbereich BIQG
GRADE	The Grading of Recommendations Assessment, Development and Evaluation
HTA	Health Technology Assessment
HTA-AG	HTA-Arbeitsgruppe
HVB	Österreichischer Hauptverband der Sozialversicherungsträger
IET	Institut für klinische Epidemiologie der TILAK
ITT	Intention to treat
KAL	Katalog ambulanter Leistungen
LBI	Ludwig Boltzmann Institut
LKF-System	Leistungsorientierte Krankenanstaltenfinanzierung
LSD	Landessanitätsdirektion
MBDS	Minimal basic data set
MeSH	Medical Subject Headings
NHH	Number needed to harm
NICE	National Institute for Health and Clinical Excellence
NNT	Number needed to treat
OR	Odds Ratio
PICO	Population, Intervention, Control, Outcome
QUADAS	Quality Assessment of Diagnostic Accuracy Studies
RA	Rapid Assessment
RCT	Randomized controlled trials / randomisierte kontrollierte Studien

RR	Relatives Risiko
RRR	Relative Risikoreduktion
TILAK	Tiroler Landeskrankenanstalten GmbH
UMIT	Private Universität für Gesundheitswissenschaften, Medizinische Informatik und Technik

Einleitung

Die systematische Analyse und Bewertung medizinischer Verfahren und Technologien (nicht zuletzt Pflegeinterventionen) auf Basis von Health Technology Assessments (HTA) gewinnt in vielen europäischen Ländern seit Mitte der 1990er Jahre für versorgungs- und gesundheitspolitische Entscheidungen zunehmend an Bedeutung. Dabei wird in den meisten Fällen sowohl die medizinische und pflegerische Effektivität als auch die ökonomische Effizienz von Maßnahmen beurteilt; außerdem sollen soziale, juristische, organisatorische und ethische Fragestellungen in einen HTA-Bericht Eingang finden. In den letzten Jahren sind vermehrt HTA-Berichte publiziert worden, die Qualität dieser Berichte ist jedoch sehr unterschiedlich. Neben einem standardisierten und nachvollziehbaren Prozedere (vgl. z. B. HTA-Prozesshandbuch, GÖG/BIQG 2010) müssen bei der Erstellung von HTA-Berichten auch fachlich-methodische Anforderungen eingehalten werden.

Vor diesem Hintergrund beauftragte das Bundesministerium für Gesundheit (BMG) die Gesundheit Österreich GmbH / Geschäftsbereich BIQG mit der Erstellung eines Methodenhandbuchs für Health Technology Assessment. Dieses wird in Kooperation mit anderen HTA-Instituten (Ludwig Boltzmann Institut für HTA, Donau-Universität Krems, UMIT, Grazer EBM-Center) erstellt, regelmäßig aktualisiert und erweitert. Zusätzlich werden die Inhalte des Handbuchs von externen Begutachterinnen/Begutachtern bewertet sowie einem öffentlichen Stellungnahmeverfahren unterzogen.

Vorliegender Bericht (Version 1.2012) ist die zweite, stark erweiterte Version des Methodenhandbuchs für Health Technology Assessment. Die in dieser Version neu hinzugekommenen Teile wurden einem Stellungnahmeverfahren unterzogen.

Maßgebliche internationale Methodenhandbücher, die für die Erstellung herangezogen werden und sich zum ergänzenden Nachlesen anbieten, sind im Anhang aufgelistet (unter 15.2).

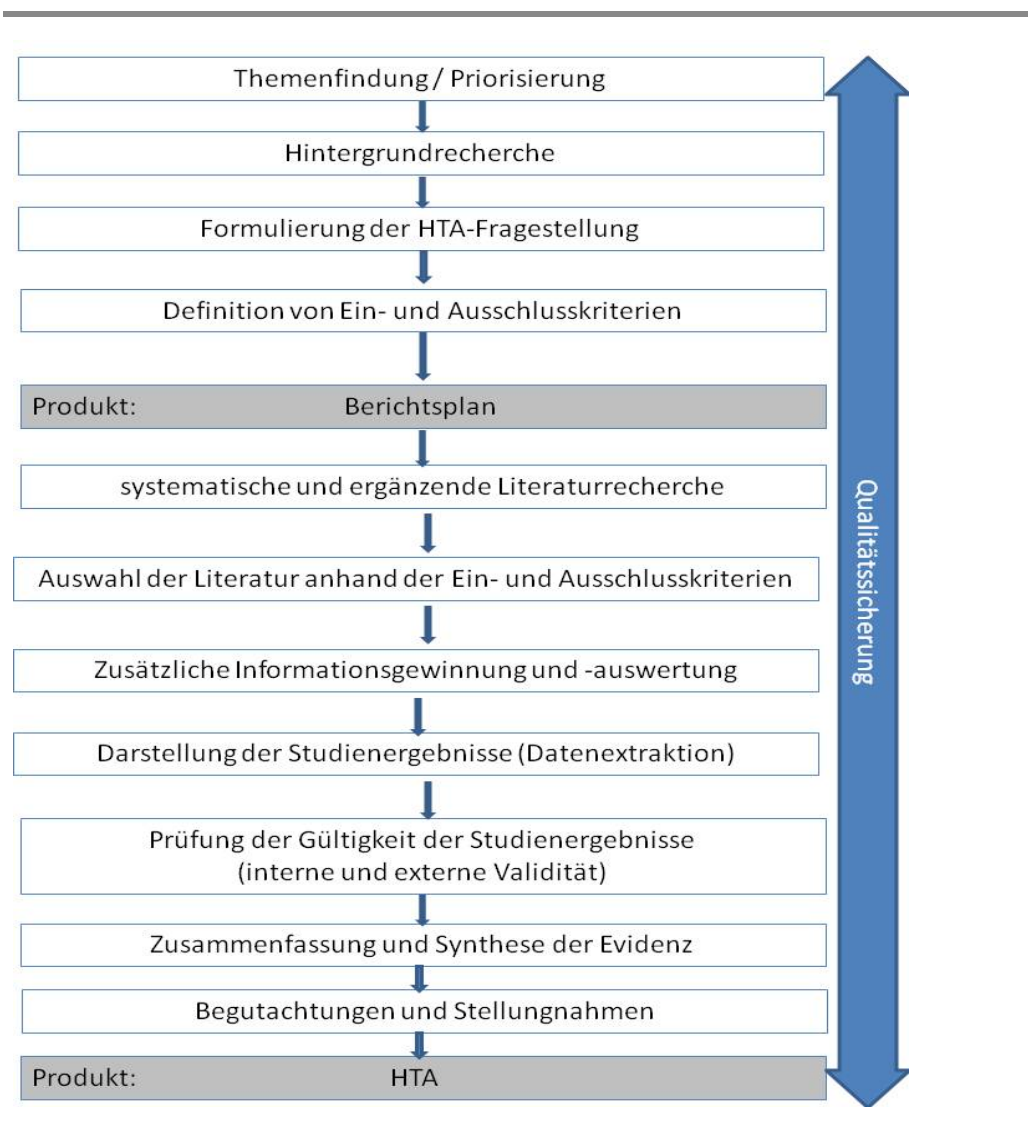
Das Handbuch dient dazu, die methodische Vorgehensweise nachvollziehbar zu machen und zu standardisieren. Ein Anspruch daran ist, dass es praxisbezogen und für eine interessierte Öffentlichkeit verständlich und nachvollziehbar sein soll.

Die im Handbuch beschriebenen Methoden sollen bei allen vom BMG und von der Bundesgesundheitsagentur beauftragten HTA-Projekten angewandt werden. Im Rahmen der nationalen HTA-Strategie wird das mittelfristige Ziel verfolgt, dass auch andere (öffentliche) Einrichtungen bzw. Auftraggeber/innen sowie HTA-Anbieter/innen dieses Handbuch ihren Projekten zugrunde legen. Damit soll es gelingen, in Österreich einheitliche methodische Standards zu etablieren.

1 Aufbau des Methodenhandbuchs

Der Aufbau des vorliegenden Methodenhandbuchs ist am Prozess der Durchführung von HTA-Berichten orientiert. In nachfolgender Abbildung ist dieser Prozess grafisch dargestellt:

Tabelle 1.1:
HTA-Erstellungsprozess



Quelle und Darstellung: GÖG/BIQG 2009

In Kapitel zwei werden zur besseren Verständlichkeit des Methodenhandbuchs die unterschiedlichen HTA-Produkte (HTA-Bericht, Rapid Assessment, Adaption bestehender Assessments) dargestellt. Kapitel drei „Themenfindung und Priorisierung“ beschreibt überblicksartig, wie HTA-Themen festgelegt werden können.

In Kapitel vier werden methodische Aspekte der Durchführung von Health Technology Assessments erläutert. Diese umfassen methodische Vorgaben und Vorschläge bezüglich der Durchführung von Hintergrundrecherchen, der Formulierung der Fragestellung nach dem PICO-Schema, der Definition von Ein- und Ausschlusskriterien, der Erstellung des Berichtsplans, der Literatursuche (Entwicklung einer Suchstrategie, Auswahl der Datenbanken usw.) und -selektion. Außerdem wird als Exkurs auf die Klassifikation von Studiendesigns und deren Hierarchie bei HTA eingegangen.

In Kapitel fünf werden Tools der zusätzlichen Datengenerierung und -analyse dargestellt. Eine zusätzliche Datengenerierung stellt eher einen Randbereich von HTA dar, da sie nur dann notwendig ist, wenn nicht genügend publizierte Evidenz vorhanden ist. Deswegen werden in diesem Kapitel Grundbegriffe der Datengenerierung erläutert. Außerdem wird auf die Möglichkeit der systematischen Auswertung von Daten (z. B. Registerdaten) im Rahmen von HTA eingegangen.

In Kapitel sechs wird dargestellt, wie die externe und die interne Validität von medizinischen und pflegewissenschaftlichen Studien (unterteilt nach Studientypen) zu beurteilen ist, außerdem wird gezeigt, wie die Datenextraktion erfolgen soll.

Kapitel sieben beschäftigt sich mit der Synthese der Evidenz. Es beschreibt Möglichkeiten, Evidenz zusammenzufassen, Effektgrößen darzustellen und die Stärke der vorhandenen Evidenz nach dem GRADE-Schema zu beurteilen. Weiters wird auf Möglichkeiten der quantitativen Zusammenfassung von Studien und auf Modellierungen eingegangen.

In Kapitel acht wird auf ökonomische Bewertungen eingegangen. Es wird erläutert, wie ökonomische Übersichtsarbeiten zu erstellen sind. Außerdem behandelt werden die Methodik von Budgetauswirkungsanalysen und gesundheitsökonomische Modellierungen. Am Ende des Kapitels findet sich eine ökonomische Empfehlung für Österreich.

Obwohl Konsens darüber besteht, dass soziokulturelle Aspekte in HTA-Berichten von großer Bedeutung sind, liegt keine einheitliche Methodik zur Durchführung vor. Kapitel neun widmet sich darum den möglichen Fragestellungen und einer Systematik der methodischen Herangehensweise in der Bewertung sozialer, organisatorischer, ethischer oder rechtlicher Aspekte und Wirkungen von medizinischen Technologien und Pflegeinterventionen.

Kapitel zehn beschäftigt sich mit Qualitätssicherung. Qualitätssicherungsmaßnahmen fließen allerdings, wie in Abbildung 1.1 dargestellt, in alle Bereiche eines Health Technology Assessment ein; insofern werden Aspekte der Qualitätssicherung (wie z. B. die Prüfung der internen und externen Validität) auch in anderen Kapiteln behandelt. Für das BIQG bindende Qualitätssicherungsmaßnahmen wie Durchführung eines Lektorats, internes und externes Begutachtungsverfahren und schriftliches Stimmungsverfahren sind dem GÖG/BIGQ-Prozesshandbuch Version 1.2010 zu entnehmen.

In einem eigenen Kapitel (Kapitel elf) wird auf die Implementierungsmöglichkeiten von Health Technology Assessment im Gesundheitssystem eingegangen. Bei diesem Kapitel handelt es sich wie beim Kapitel zur Qualitätssicherung um Vorschläge bzw. Möglichkeiten und nicht um verbindliche Arbeitsschritte.

Quellenangaben z. B. zu Datenbanken und Kostendaten sowie Tabellen zur Qualitätsbewertung finden sich im Anhang.

2 HTA-Produkte

HTA ist ein systematischer Prozess zur Analyse, Evaluierung und Bewertung wissenschaftlicher Informationen über Auswirkungen bereits eingeführter und neuer gesundheitsrelevanter Verfahren, Leistungen, Technologien und Prozeduren auf die Gesundheit bzw. auf die Gesundheitsversorgung der Bevölkerung. Es gibt HTA-Produkte unterschiedlichen Umfangs. Im Folgenden werden drei Produkte beschrieben (HTA-Bericht, Rapid Assessment, Adaption bestehender Assessments).

2.1 HTA-Bericht

Die untersuchten inhaltlichen Dimensionen sind Wirksamkeit, Sicherheit und Wirtschaftlichkeit sowie soziale, ethische, rechtliche und organisatorische Aspekte. Je nach Relevanz für die vorliegende Fragestellung erfolgt eine Gewichtung.

Bei einem umfassenden HTA-Bericht erfolgen Recherche und Aufarbeitung des aktuellen, international verfügbaren Wissensstandes zur festgelegten Fragestellung mithilfe unterschiedlicher Tools wie einer systematischen Übersichtsarbeit (in deren Rahmen eine Metaanalyse durchgeführt werden kann) oder verschiedener Erhebungsmethoden. Außerdem wird im Rahmen der Berichtserstellung entschieden, welches Vorgehen bei der ökonomischen Bewertung zu wählen ist (beispielsweise eine ökonomische Modellrechnung). Auf den Berichtsergebnissen aufbauend werden Empfehlungen abgegeben, die sich auf den gesundheitspolitischen Entscheidungszusammenhang der jeweiligen Fragestellung beziehen.

Für die einzelnen Prozessschritte der Erstellung des Berichts siehe die jeweils aktuelle Version des Prozesshandbuchs der GÖG/BIQG.

2.2 Rapid Assessment

Unter einem Rapid Assessment (RA) ist eine kurzfristige Verfahrensbewertung zu verstehen, die eingesetzt wird, wenn dringliche Entscheidungen getroffen werden müssen bzw. wenn eingeschränkte finanzielle und personelle Ressourcen zur Verfügung stehen. Innerhalb eines knapp bemessenen Zeitrahmens (ca. drei Monaten) wird eine Zusammenfassung der Evidenz erstellt. Dabei wird in der Regel eine sehr eingeschränkte Fragestellung bearbeitet.

Die Erstellung von RA erfolgt weitgehend unter Ausführung derselben Arbeitsschritte wie bei umfassenden HTA-Berichten. Durch den stark eingeschränkten Zeitrahmen und

die eingeschränkten Ressourcen werden jedoch manche Arbeitsschritte massiv verkürzt, was zu einer geringeren Detailtiefe von RA führt. Aufgrund der vereinfachten Methodik besteht außerdem ein erhöhtes Risiko für Verzerrungen. Weiters erfolgt im Rahmen eines RA keine ökonomische Bewertung, außer es wird eine rein ökonomische Bewertung (ohne vorangegangene Bewertung der Effektivität) durchgeführt. Bei alleiniger Bewertung der Effektivität werden keine Aussagen zu entstehenden Kosten, ökonomischen Auswirkungen bzw. entsprechenden Finanzierungsfragen in jeweils vorliegenden RA getroffen. Wesentlich ist aber auch beim RA, dass das methodische Vorgehen präzise und nachvollziehbar beschrieben wird.

RA dürfen daher nicht als kostengünstige Minimalvariante eines ausführlichen HTA-Berichtes (miss-)verstanden werden. Die Aussagekraft von Ergebnis, Diskussion und Empfehlungen eines RA sind unter diesem Aspekt zu relativieren.

Für die einzelnen Prozessschritte der Erstellung eines RA siehe die jeweils aktuelle Version des Prozesshandbuchs der GÖG/BIQG.

2.3 Adaption bestehender Assessments

Bei der Adaption bestehender Assessments ist zwischen einem umfassenden **Update** und einer **Abänderung** einzelner Berichtsteile zu unterscheiden. Ein Update durchläuft einen Prozess zur Identifizierung neuer Evidenz, die in einen bereits vorhandenen und vollständigen Bericht eingearbeitet wird. Demnach entspricht die Erweiterung einer Literatursuche auf neue Quellen oder auch eine ausgiebige, jedoch erfolglose Suche nach neuer Evidenz einem Update. Eine Abänderung umfasst hingegen einzelne Korrekturmaßnahmen oder Analysen im bereits publizierten Bericht (beispielsweise die Korrektur von Rechtschreibfehlern, notwendige Änderungen in der Methodik oder im Hintergrund, i. d. R. ohne Suche nach neuer Evidenz) (Higgins 2011, Moher 2008).

Die Adaption bestehender Berichte kann je nach Umfang der notwendigen Arbeitsschritte einen großen bzw. kleinen Ressourcenaufwand bedeuten. So kann beispielsweise ein Update eines bestehenden Berichts ohne Veränderung der Fragestellung erstellt werden. Jedoch ist auch ein Update eines Berichts möglich, bei dem Fragestellung und (oder) Selektionskriterien überarbeitet werden und das Update als Erweiterung bzw. Einschränkung der ursprünglichen Fragestellung dient (siehe Kapitel 4.5.3). Updates können einerseits von den jeweiligen Autorinnen und Autoren bzw. Institutionen des entsprechenden HTA-Berichts / der entsprechenden systematischen Übersichtsarbeit bzw. des entsprechenden Rapid Assessments durchgeführt werden. Andererseits kann für bereits bestehende Berichte auch von anderen Autorinnen und Autoren bzw. Institutionen, die beim Originalbericht nicht beteiligt waren, ein Update erstellt werden, um eine (teilweise) neue bzw. im Umfang abgeänderte Fragestellung

zu beantworten. Dabei ist die Übertragbarkeit der Ergebnisse aus dem Originalbericht auf das Update aufgrund möglicher Unterschiede der zugrunde liegenden Gesundheitssysteme genau zu prüfen. Abänderungen einzelner Berichtsteile werden zumeist von den jeweiligen Autorinnen und Autoren bzw. der jeweiligen Institution des Originalberichts durchgeführt (Higgins 2011, Moher 2008).

Die Gründe für die Adaption bestehender Assessments, vor allem für die Erstellung eines Updates zu einem vorhandenen Bericht, liegen in der schnellen Weiterentwicklung vorhandener Evidenz und der (möglicherweise) geringeren Aussagekraft der vorliegenden Berichtsergebnisse und der damit verbundenen Schlussfolgerungen und Empfehlungen. Weiters kann die Notwendigkeit eines Updates u. a. begründet sein in der Verfügbarkeit besser wirksamer Alternativtherapien bzw. in der klinischen Ausweitung oder geplanten Änderung der Therapie, im Vorhandensein neuer Anwendungsmöglichkeiten, im Auftreten neuer Nebenwirkungen oder Schäden für die Patientin / den Patienten, in der Publikation entgegengesetzter Ergebnisse, der Änderung der Beschreibung der Effektivität oder in statistischen Änderungen, die die Entscheidungsfindung auf Grundlage des vorliegenden Bericht beeinflussen würde (Agency for Healthcare Research and Quality 2007, Garritty 2010, Higgins 2011, Moher 2008).

2.4 Budgetauswirkungsanalyse

Zur gesundheitsökonomischen Bewertung einer medizinischen Technologie oder Strukturveränderung kann neben gesundheitsökonomischen Studien wie der Kosten-Wirksamkeits-Analyse auch eine Budgetauswirkungsanalyse (engl.: budget impact analysis, BIA) durchgeführt werden.

Im Rahmen einer Budgetauswirkungsanalyse werden die direkten finanziellen Auswirkungen (Folgekosten bzw. -erlöse) bewertet, die entstehen, wenn eine neue Gesundheitstechnologie eingeführt bzw. erstattet wird bzw. wenn die Nutzung einer bestehenden Gesundheitstechnologie in einem Gesundheitssystem geändert wird (Garattini 2011, Mauskopf 2007, Nuijten 2011, Patented Medicine Prices Review Board 2007, Trueman 2001). Darunter fallen beispielsweise Finanzierungsflüsse, die für den potenziellen Anteil der Patientinnen und Patienten, die die neue bzw. veränderte Therapie erhalten werden, anfallen, Finanzierungsflüsse, die der Verbreitung der Therapie im Gesundheitssystem dienen, sowie Finanzierungsflüsse, die aufgrund der Anwendung der Therapie bei zuvor unbehandelten Patientinnen und Patienten anfallen werden (Mauskopf 2007). Detaillierte Angaben zur Erstellung einer Budgetauswirkungsanalyse finden sich im Kapitel 8.2.

Autorin: Daniela Pertl (GÖG/BIQG)

3 Themenfindung und Priorisierung

3.1 Themenfindung

Die Themenfindung steht am Beginn der Erstellung eines Health Technology Assessment (HTA) und hat jene Forschungslücken zu identifizieren, die aktuell oder in näherer Zukunft durch Erstellung eines HTA unterschiedlichen Umfangs (HTA-Berichte, Rapid Assessments, Adaption bestehender Assessments) zu bearbeiten sind (GÖG/BIQG 2009).

International werden unterschiedliche Methoden zur Themenfindung diskutiert und angewendet. Häufig bieten HTA-Einrichtungen eine öffentlich zugängliche Themeneingabe über ihre Website an, führen persönliche Befragungen von Stakeholdern und Organisationen im Gesundheitswesen durch oder identifizieren neue Technologien durch Frühwarnsysteme und Horizon Scanning. Das konkrete Vorgehen bei der Themenfindung ist eng mit dem Aufgabenprofil und der Funktion der jeweiligen HTA-Einrichtung im Gesundheitswesen verknüpft. Beispielsweise können HTA-Einrichtungen auch ohne priorisiertes Themenfindungsverfahren mit der Bearbeitung konkreter Fragestellungen von nationalen oder internationalen Institutionen beauftragt werden. Meist werden öffentlich geführte HTA-Institutionen von staatlichen Einrichtungen mit Themen von besonderer Dringlichkeit oder von öffentlichem Interesse direkt beauftragt (GÖG/BIQG 2008).

Von zahlreichen HTA-Einrichtungen wird der prinzipielle Anspruch formuliert, dass die Themenfindung auf möglichst breiter Basis (Stakeholder und Öffentlichkeit) sowie anhand eines strukturierten Prozesses durchgeführt werden sollte. Dadurch soll das Spektrum möglicher Fragestellungen erweitert und die Relevanz der Themen erhöht werden. Ein strukturierter, transparenter Prozess dient der Nachvollziehbarkeit und Akzeptanz der Themenfindung (GÖG/BIQG 2008).

3.2 Priorisierung

Unter Priorisierung ist die vorrangige Behandlung von Fragestellungen zu verstehen, die zuvor durch einheitliche und transparent dargelegte Kriterien identifiziert wurden. Durch Priorisierung wird das Ziel verfolgt, besonders relevante Fragestellungen herauszufiltern und für jede (priorisierte) Fragestellung eine sinnvolle methodische Vorgehensweise zu bestimmen (HTA-Bericht, Rapid Assessment, Adaption bestehender Assessments) (GÖG/BIQG 2009).

Eine international gültige Methode zur Priorisierung existiert nicht, da diese wesentlich durch den Charakter der erstellenden HTA-Einrichtung bestimmt wird. Für HTA-Einrichtungen, die mit der Bearbeitung konkreter Fragestellungen beauftragt werden, ist eine formale Priorisierung in der Regel hinfällig, da die Fragestellung vom Auftraggeber vorgegeben wird. Für zumeist öffentlich geführte HTA-Einrichtungen, die HTA-Berichte zu wesentlichen Fragen im Gesundheitswesen im Rahmen eines HTA-Programms erstellen, ergibt sich die Priorisierung der Themen nicht automatisch aus ihrem Auftrag. Hierfür müssen gesonderte Verfahren gewählt werden, um, unter Einbeziehung diverser Stakeholder, relevante Themen zu identifizieren, z. B. Reihung besonders relevanter Themenvorschläge nach zuvor festgelegten Kriterien (vgl. Anhang) durch eine Priorisierungsgruppe (Perleth et al. 2008).

Autorin: Daniela Pertl (GÖG/BIQG)

4 Methodik der Durchführung eines Health Technology Assessment

In folgendem Kapitel wird auf die Methodik für die Erstellung von HTA-Produkten eingegangen. Nicht immer ist es sinnvoll, alle unten genannten Schritte – der Struktur des Methodenhandbuchs folgend – nacheinander durchzuführen, vielfach kommt es zu Überschneidungen in den Arbeitsabläufen.

4.1 Entwickeln der Fragestellung

4.1.1 Vorabrecherche von Hintergrundinformationen (Scoping)

Um eine genaue Fragestellung definieren zu können, ist eine (unsystematische) Vorabrecherche notwendig, sofern diese nicht schon durch den Prozess der Themenfindung und Prioritätensetzung erfolgt ist. Es bietet sich z. B. an, Leitlinien und bereits existierende HTA-Berichte und systematische Übersichtsarbeiten zu suchen und zu beschaffen und eine mögliche Adaption bestehender Berichte zur Beantwortung der Fragestellung zu überprüfen (siehe Kapitel 2.3). Auch Gespräche mit der Auftraggeberin / dem Auftraggeber sowie mit Fachexpertinnen/Fachexperten sind sinnvoll.

4.1.2 Operationalisieren der Fragestellung

Unter Operationalisierung wird die „Übersetzung“ einer vorläufig formulierten HTA-Frage in eine prüfbare HTA-Fragestellung verstanden (Counsell 1997). Eine Operationalisierung der Fragestellung ist in der Regel erforderlich, da die Fragestellung von der Auftraggeberin / vom Auftraggeber bzw. im Rahmen der öffentlichen Themeneingabe meistens zu ungenau formuliert wird. Die Beteiligung von externen Fachexperten und Stakeholdern bei der Operationalisierung der Fragestellung sowie im gesamten Erstellungsprozess eines HTA-Berichts bzw. eines Updates zu einem bestehenden Assessment wird ausdrücklich empfohlen. Das konkrete Ausmaß dieser Beteiligung ist zwischen der Auftraggeberin / dem Auftraggeber und dem HTA-Berichtsersteller abzustimmen.

4.1.2.1 PICO-Prinzip

Zur Operationalisierung einer HTA-Fragestellung wird die Anwendung des sogenannten PICO-Prinzips bereits im Rahmen der Erstellung des Berichtsplans empfohlen. PICO steht als Akronym für *Population, Intervention, Control, Outcome* und beschreibt damit vier grundlegende Komponenten einer HTA-Frage (Counsell 1997). Vor allem auf Fragen zu medizinischen Interventionen, zu pflegerischen Interventionen, auf diagnostische Fragen und solche zu Screening-Programmen ist das PICO-Prinzip anwendbar. Für ökonomische und organisatorische Fragen ist in der Regel eine Anpassung einzelner PICO-Komponenten erforderlich. Für methodische Fragen ist das PICO-Prinzip meistens nicht sinnvoll anwendbar. Korrekt formulierte PICO-Fragen bilden die Grundlage einer effizienten systematischen Literatursuche (Gartlehner 2009).

4.1.2.2 PICO-Komponenten

Bei Fragestellungen zu medizinischen und pflegerischen Interventionen beschreibt die PICO-Komponente *Population* eine Gruppe von Patientinnen und Patienten oder Studienprobanden, die anhand unterschiedlicher Merkmale (z. B. Erkrankungen, Zusatzerkrankungen, Geschlecht, Alter) charakterisiert sind. Bei organisatorischen Fragen ist die betroffene Organisationseinheit oder die Versorgungsebene (z. B. ambulant, stationär) zu definieren.

Unter *Intervention* werden die in Bezug auf die jeweiligen Fragestellungen relevanten Therapien, Maßnahmen, diagnostischen Tests, Expositionen (z. B. Umweltschadstoffbelastung), Risikofaktoren oder Versorgungsabläufe definiert. Bei organisatorischen Fragestellungen können z. B. angewendete Leitlinien dieser PICO-Komponente zugeordnet werden.

Die PICO-Komponente *Control* (Kontrollintervention, Kontrollgruppe) beschreibt bei Fragestellungen zu Interventionen die Vergleichsintervention, das heißt die Goldstandard-Intervention, eine Scheinintervention (Placebo) oder alternative Interventionen. Bei Vorhandensein eines etablierten Therapieverfahrens (Standardverfahren) sollte ein Vergleich der zu evaluierenden (neuen) Intervention mit dem Standardverfahren erfolgen. Bei Fragestellungen zum Screening ist die Kontrollintervention meistens „kein Screening“ oder das bisher angewandte Screeningverfahren. Bei Fragestellungen zur Wirkung einer Exposition oder eines Risikofaktors ist in der Regel in der Kontrollgruppe die Exposition beziehungsweise der Risikofaktor nicht vorhanden (Machin et al. 2005). Bei diagnostischen Fragestellungen ist die Goldstandard-Diagnostik in der Regel die adäquate Kontrollintervention.

Mit der PICO-Komponente *Outcome* werden Studienendpunkte (Zielvariablen) festgelegt. Studienendpunkte unterscheiden sich unter anderem hinsichtlich ihrer klinischen Relevanz für die Patientin / den Patienten, wobei patientenrelevante Studienendpunkte von sogenannten Surrogatparametern abgegrenzt werden können. Patientenrelevante Studienendpunkte sind Endpunkte mit unmittelbarer Bedeutung für die Patientin / den Patienten und beziehen sich insbesondere auf die Sterblichkeit (Mortalität), auf die Einschränkung der Funktionsfähigkeit durch nicht-tödliche Krankheitsereignisse (Morbidität) sowie auf die gesundheitsbezogene Lebensqualität (Biomarkers Definitions Working Group 2001). Demgegenüber stellen Surrogatparameter Endpunkte dar, die keine unmittelbare Relevanz für die Patientin / den Patienten haben, jedoch mit patientenrelevanten Endpunkten assoziiert sind (IQWiG 2008a). In der Regel sind Surrogatparameter physiologische oder biochemische Messwerte, die vergleichsweise einfach bestimmt werden können und als Ersatzmessgrößen stellvertretend für patientenrelevante Endpunkte erhoben werden (z. B. Senkung von LDL-Cholesterin als Surrogatparameter zur Reduktion der kardiovaskulären Mortalität; Biomarkers Definitions Working Group 2001). Eine weitere Klassifikation von Studienendpunkten (Zielvariablen) betrifft deren Festlegung als primäre oder sekundäre Endpunkte im Rahmen der Studienkonzeption. Als primärer Studienendpunkt ist diejenige Zielvariable definiert, anhand der das Ergebnis einer Studie bewertet wird (Counsell 1997) und für die in der Regel die Fallzahlplanung durchgeführt wird. Sekundäre Endpunkte sind nachrangige Zielvariablen, die zusätzliche Ergebnisse zum Studienhauptergebnis liefern.

Nachfolgend sind sechs Beispiele für konkrete PICO-Fragestellungen dargestellt.

Beispiel-Fragestellung zu klinischen, therapeutischen, medizinischen oder pflegerischen Interventionen: *Können antimikrobielle Präparate bei Patienten/Patientinnen mit chronischen Wunden die Wundheilung fördern?*

Tabelle 4.1:

PICO-Prinzip zur Wirksamkeit von antimikrobiellen Präparaten zur Wundheilung (Beispiel)

PICO-Komponenten		Beispiel
P	Population	Patientinnen/Patienten mit verschiedenen Formen chronischer Wunden (z. B. diabetischer Ulcus, venöser Ulcus, Druckulcus)
I	Intervention	Antimikrobielle Präparate
C	Control	Placebo bzw. Standardtherapie
O	Outcome	Wundheilung (komplette Heilung, verbleibende Wundfläche, Wundheilungsscores)

Quelle: Kunz et al. 2009; GÖG/BIQG-eigene Darstellung

Beispiel-Fragestellung diagnostische Tests: *Ist die vollautomatische Blutdruckmessung am Handgelenk im Vergleich zur mechanischen, auskultatorischen Messung mittels*

Druckmanschette am Oberarm eine adäquate Methode zur Bestimmung des arteriellen Blutdrucks?

Tabelle 4.2:

PICO-Prinzip zur Diagnosegenauigkeit der automatischen Blutdruckmessung am Handgelenk (Beispiel)

PICO-Komponenten		Beispiel
P	Population	Allgemeinbevölkerung mit normalem Blutdruck, Patientinnen/Patienten mit arterieller Hyper- und Hypotonie
I	Intervention	Vollautomatische Blutdruckmessung am Handgelenk
C	Control	Auskultatorische Messung mittels Druckmanschette am Oberarm (Goldstandard)
O	Outcome	Messgenauigkeit (in mmHg), Zuverlässigkeit der Blutdruckmessung (Reliabilität)

Quelle und Darstellung: GÖG/BIQG

Beispiel-Fragestellung zu Screening: *Kann ein Screening auf COPD (chronisch obstruktive Lungenerkrankung) mittels Spirometrie bei asymptomatischen Erwachsenen den weiteren Verlauf einer erkannten COPD günstig beeinflussen?*

Tabelle 4.3:

PICO-Prinzip zum COPD-Screening (Beispiel)

PICO-Komponenten		Beispiel
P	Population	Erwachsene, allgemeine Bevölkerung, keine COPD-Symptome
I	Intervention	Früherkennung der COPD mittels Spirometrie
C	Control	Keine Früherkennung mittels Spirometrie
O	Outcome	Verlauf der COPD (Lungenfunktion, Exazerbationsrate, Mortalität)

Quelle und Darstellung: GÖG/BIQG

Beispiel für eine gesundheitsökonomische Fragestellung: *Wie ist die Kosteneffektivität von Langzeit-Analoginsulin (Insulin glargin) gegenüber NPH-Insulin bei Patientinnen und Patienten mit Diabetes mellitus Typ I einzuschätzen?*

Tabelle 4.4:

PICO-Prinzip zur Kosteneffektivität Langzeit-Analoginsulin (Beispiel)

PICO-Komponenten		Beispiel
P	Population	Patientinnen/Patienten mit Diabetes mellitus Typ I
I	Intervention	Therapie mit Langzeit-Analoginsulin (Insulin glargin)
C	Control	Therapie mit NPH-Insulin
O	Outcome	Kosten pro qualitätsbereinigtem Lebensjahr (QALY)

Quelle und Darstellung: GÖG/BIQG

Beispiel für eine Fragestellung im Bereich Pflege: *Wie ist die Sensitivität und Spezifität von Instrumenten zum Assessment von Dekubitus zu bewerten?*

Tabelle 4.5:
PICO-Prinzip zur Sensitivität und Spezifität von Assessmentinstrumenten (Beispiel)

PICO-Komponenten	Beispiel
P Population	Zu pflegende Patientinnen/Patienten
I Intervention	Verschiedene Assessmentinstrumente (z. B. Skalen) zur Erfassung von Dekubitus
C Control	Andere Assessmentinstrumente bzw. kein Instrument
O Outcome	Sensitivität, Spezifität

Beispiel für eine organisatorische Fragestellung: *Welche organisatorischen Modelle und Ablaufstrukturen führen zu einer Verbesserung der klinischen und ökonomischen Ergebnisse in Intensivstationen?*

Tabelle 4.6:
PICO-Prinzip zu organisatorischen Modellen in Intensivstationen (Beispiel)

PICO-Komponenten	Beispiel
P Population (adaptiert: Organisationseinheit)	Intensivstationen
I Intervention	Organisatorische Modelle, Ablaufstrukturen
C Control	Keine Kontrollgruppe
O Outcome	Verbesserung der klinischen und ökonomischen Ergebnisse (z. B. Mortalität, iatrogene Morbidität, Wiedereinweisung, direkter Ressourceneinsatz, Liegedauer)

Quelle: LBI-HTA 2007a; GÖG/BIQG-eigene Darstellung

4.1.2.3 Änderung der Fragestellung während der Erstellung eines HTA-Produkts

Grundsätzlich ist die HTA-Fragestellung a priori zu erstellen, d. h. vor Beginn der systematischen Literatursuche, um eine Anpassung der HTA-Fragestellung an die identifizierte Evidenz zu verhindern. Es kann jedoch vorkommen, dass bei der Erstselektion der Evidenz im Zuge der Literaturrecherche offensichtlich wird, dass die ursprünglich formulierte HTA-Fragestellung zu unpräzise ist. In diesem Ausnahmefall ist eine Modifikation der Fragestellung möglich, jedoch sind eine Abstimmung mit der Auftraggeberin / dem Auftraggeber und eine transparente Dokumentation im Berichtsplan verpflichtend (siehe auch Prozesshandbuch der GÖG/BIQG, Version 1.2010). Zu einem späteren Zeitpunkt, das heißt nach Durchsicht der Volltexte (Zweitselektion), ist eine Änderung der Fragestellung jedenfalls nicht mehr zulässig (Kunz 2009).

4.1.2.4 Nutzen und Schaden von Interventionen

Im Kontext von HTA bezeichnet der Begriff Nutzen kausal begründete positive Auswirkungen, der Begriff Schaden kausal begründete negative Auswirkungen von Interventionen (z. B. Therapien, medizinischen Maßnahmen, diagnostischen Tests) im Vergleich zu Kontrollinterventionen (z. B. zu Goldstandard-Intervention, alternativer Intervention, Scheinintervention, keiner Intervention), bezogen auf patientenrelevante Endpunkte (IQWiG 2008a). In HTA-Produkten sollten (unerwünschte) Nebenwirkungen von Interventionen ausdrücklich berücksichtigt werden. Studientypen, die zur Erfassung von Nebenwirkungen geeignet sind, sollen daher im Rahmen des Selektionsprozesses (siehe Kapitel zu Ein- und Ausschlusskriterien) mit einbezogen werden. Unter der Voraussetzung einer geeigneten Fragestellung sollte in HTA-Produkten eine explizite Abwägung von Nutzen und Schaden einer Intervention erfolgen.

4.2 Definition der Selektionskriterien

Grundsätzlich legen die Selektionskriterien (Ein- und Ausschlusskriterien) fest, welche Studien im Bericht zur Bewertung der Fragestellung herangezogen werden und welche nicht. Die Selektionskriterien sind eine Grundlage für die Erstellung der Suchstrategie und werden für die Erst- und Zweitselektion verwendet. Im Lauf der Erstellung eines HTA-Berichts, eines Rapid Assessment oder eines Updates eines bestehenden Assessments kann es notwendig sein, die Kriterien zu adaptieren. Wesentlich ist, dass solche Adaptionen im Bericht dokumentiert werden. Die Selektionskriterien können wie im Folgenden dargestellt untergliedert werden:

1. Formale Kriterien

- » Sprache
- » Suchzeitraum
- » Art der Publikation (z. B. peer reviewed journal, nicht publizierte [graue] Literatur)

2. Kriterien zur inhaltlichen Relevanz

- » Patientencharakteristika
- » Intervention
- » Outcome-Parameter

3. Kriterien in Bezug auf Qualität/Validität

- » Studiendesign (z. B. Metaanalyse, HTA, RCT)
- » Spezifische Aspekte (wie Länge der Beobachtung, Größe der Studiengruppe)

Der festgelegte Suchzeitraum sollte begründbar sein, z. B. ab Einführung (Zulassung) einer der untersuchten Technologien oder ab dem Zeitpunkt der zuletzt durchgeführten Literatursuche bei Updates bestehender Berichte. Die Kriterien zur inhaltlichen Relevanz sollten sich aus der operationalisierten Fragestellung ergeben. Bei den Kriterien zu Qualität/Validität sollte überlegt werden, welche Studiendesigns am geeignetsten sind, um einen Effekt nachzuweisen. Es empfiehlt sich aber, wenn eine mangelnde Verfügbarkeit von für die Fragestellung relevanter Literatur zu erwarten ist, zu Beginn der Selektion weitere Studiendesigns einzuschließen. Ist die Verfügbarkeit von Literatur doch besser als erwartet, können die Selektionskriterien im Lauf des Selektionsprozesses stärker eingeschränkt werden (CRD 2001, DCAHTA 2008). Wenn ein HTA-Bericht (oder RA oder ein Update eines bestehenden Berichts) in mehrere Unterfragestellungen (z. B. Unterfragestellung zu den Nebenwirkungen eines Verfahrens) unterteilt ist, werden in der Regel für die einzelnen Unterfragestellungen jeweils eigene Selektionskriterien definiert. So werden beispielsweise zur Bewertung der Nebenwirkungen eines Verfahrens andere Studiendesigns herangezogen als zum Nachweis seiner Effektivität. Zur Hierarchie von Studientypen siehe 4.7.

4.3 Erstellen des Hintergrunds

Der gesundheitspolitische und wissenschaftliche Hintergrund soll für die Leserin / den Leser des HTA-Berichts bzw. RA bzw. des Updates eines bestehenden Berichts einen kurzen Überblick über die Thematik geben. Die eigentliche Fragestellung wird erst im Ergebnisteil des Berichts behandelt und bei der Datensynthese beantwortet. Endergebnisse im Bezug auf die Fragestellung, die zum Zeitpunkt der Erstellung des jeweiligen Hintergrunds schon vorliegen bzw. vorliegen könnten, werden an dieser Stelle nicht berücksichtigt.

4.3.1 Erstellen des gesundheitspolitischen Hintergrunds

Der gesundheitspolitische Hintergrund hat (soweit relevant) folgende Punkte zu beinhalten:

- » Beschreibung der Ausgangslage zur Fragestellung:
 - » Hintergrund der Beauftragung (wer ist/sind der/die Auftraggeber/innen, stammt das Thema aus der priorisierten Themenliste? etc.);
 - » Erklärung, ob ein gänzlich neuer Bericht oder ein Update eines bestehenden Berichts erstellt wird (die Durchführung eines Updates muss begründet werden);

- » Begründung für die Erstellung des Berichts in Bezug auf den derzeitigen Wissensstand.
- » Beleuchtung der Fragestellung aus verschiedenen Perspektiven (z. B. aus Public-Health-Sicht, Patienten-Sicht, Sicht der Gesundheitspolitik etc.):
 - » Erläuterung des Gesundheitsproblems, z. B. epidemiologische Aspekte zur Technologie/Intervention (Zielgruppe, Prävalenz, Inzidenz etc.); Krankheitslast für Individuen und die Gesellschaft;
 - » derzeitiger Einsatz der Technologie/Intervention, z. B. Verfügbarkeit, betroffene Sektoren, Form der Verwendung, Lebenszyklus, derzeitige Erstattungssituation der betroffenen Technologie, betroffene Finanzierungsträger, derzeitige Verbreitung, Nennung von Alternativen zur Technologie, regionale/zeitliche/andere Variationen in der Verwendungshäufigkeit.

Die Informationen für den gesundheitspolitischen Hintergrund werden u. a. durch die Vorabrecherche von Hintergrundinformationen identifiziert. Weiters bietet es sich je nach Fragestellung an, österreichrelevante Quellen heranzuziehen (EUnetHTA 2008).

4.3.2 Erstellen des wissenschaftlichen Hintergrunds

Der wissenschaftliche Hintergrund hat (soweit relevant) folgende Punkte zu beinhalten:

- » Beschreibung des derzeitigen Wissensstands und der Limitationen
 - a) Beschreibung und fachliche/technische Merkmale der Technologie/ Intervention sowie Vergleichsintervention
z. B. Definition der Technologie/Intervention, Zeitpunkt der Entwicklung, Verwendungszweck, Anwender, Art und Weise der Anwendung, benötigte Materialien, Erfordernisse wie Räumlichkeiten, technische Ausstattung, Personal, Ausbildung bzw. Information, Kosten der Technologie;
 - b) ggf. kurze Nennung weiterer/alternativ infrage kommender Interventionen und Einordnung in Hinblick auf die Fragestellung (inkl. Begründung ihres Ausschlusses);
 - c) Zielgrößen (die als sinnvoll/aussagekräftig erachtet werden – Patientenrelevanz);
 - d) ökonomischer Hintergrund (ggf. auch als eigene Überschrift)
- » Zielsetzung des Berichts (was soll durch den Bericht an Wissen ergänzt werden?).

Die Informationen für den wissenschaftlichen Hintergrund werden u. a. durch die Vorabrecherche von Hintergrundinformationen bzw. ggf. durch eingeschlossene Volltexte der systematischen Literatursuche identifiziert. Weiters bietet es sich je nach Fragestellung an, in medizinischen und pflegewissenschaftlichen Fachjournalen sowie

nach international und national anerkannten Klassifikationen oder Klassifikationsgruppen (z. B. World Health Organization Family of International Classifications [WHO-FIC]) zu suchen bzw. österreichrelevante Quellen heranzuziehen (EUnetHTA 2008).

4.4 Erstellen eines Berichtsplans

Für jeden HTA-Bericht, für jedes Rapid Assessment und für jedes Update zu einem bestehenden Bericht wird ein Berichtsplan erstellt, der fester Bestandteil auch eines Endberichts ist. Es kann im Lauf des Prozesses der Berichtserstellung zu Abweichungen vom Berichtsplan kommen, diese müssen laufend transparent und nachvollziehbar im Berichtsplan bzw. in weiterer Folge im Endbericht beschrieben werden. Die Elemente, die ein Berichtsplan enthalten muss, sind (Liberati 2009, LBI-HTA 2007):

- » die genaue Fragestellung (PICO) und der gesundheitspolitische Hintergrund zur Fragestellung;
- » Intervention und Vergleichsintervention, Zielgruppe, Zielgrößen, Ausgaben;
- » Methodik:
Schlagwortgruppen (ohne exakte Ausformulierung der Schlagworte) für die Suchstrategie,
Ein- und Ausschlusskriterien für die Literaturselektion (Studiendesign);
- » (Art der) Qualitätssicherung bei Literaturselektion, Kriterien zur Bewertung der Studienqualität (Qualitätskriterien);
- » Darstellung der Verfahren zur Analyse der Ergebnisse (z. B. qualitative Synthese in Evidenztabelle oder bei quantitativer Synthese z. B. verwendete Methoden bei Metaanalysen oder Modellierungen);
- » Beschreibung Reviewverfahren;
- » Darlegung von Interessenkonflikten;
- » Einschätzung zu und Festlegen von notwendigen (grundlegenden) Abweichungen von der unten skizzierten Vorgehensweise – diese können z. B. inhaltlich durch die Fragestellung bedingt sein oder durch mangelnde Literaturverfügbarkeit.

Im Unterschied zu Rapid Assessments umfasst der Berichtsplan von HTA-Berichten zusätzlich:

- » die Einschätzung (wenn zu diesem Zeitpunkt möglich), ob eine Metaanalyse sinnvoll erscheint oder ob die Arbeit ausschließlich als systematische Übersichtsarbeit erstellt wird, sowie die Frage:

- » Sind soziale, ethische und rechtliche bzw. organisatorische Aspekte zu behandeln bzw. welche dieser Aspekte sind zu behandeln und mit welcher Methodik (vgl. auch Kapitel 9)?

Gegebenenfalls kann ein Zeitplan in den Berichtsplan inkludiert werden, der einen Überblick über die Dauer der einzelnen Arbeitsschritte und Meilensteine bietet.

4.5 Literatursuche

Die Literatursuche für HTA-Berichte kann aus verschiedenen Elementen bestehen. Immer enthält die Literatursuche eine systematische Suche mittels Suchstrategie in verschiedenen Datenbanken. Andere Suchverfahren und Datenquellen, die herangezogen werden können, sind: Handsuche, Suche nach nicht publizierter (grauer) Literatur und Identifikation von Literatur durch Expertinnen und Experten (siehe dazu 4.5.1.3.1 bis 4.5.1.3.3). Bei HTA-Berichten und Updates zu bestehenden Berichten ist die Literatursuche immer umfangreicher als bei Rapid Assessments (siehe dazu 4.5.1 und 4.5.2 und 4.5.3).

Die Herausforderung an eine systematische Literatursuche besteht darin, zum einen möglichst alle relevanten Publikationen zur jeweiligen Fragestellung zu finden und zum anderen die Zahl der irrelevanten Treffer möglichst gering zu halten. Diese Herausforderung wird durch die Kennzahlen Sensitivität und Spezifität beschrieben. Sensitivität, im Englischen oft als „recall“ bezeichnet, gibt mit dem Anteil der bei der Suche gefundenen relevanten Publikationen die Vollständigkeit des Suchergebnisses an. Die Spezifität, im Englischen als „precision“ bezeichnet, beschreibt hingegen mit dem Anteil der relevanten Publikationen an der Gesamtergebnismenge die Genauigkeit des Suchergebnisses. Der Anspruch an die systematische Literatursuche ist, dass (möglichst) alle relevanten Primärpublikationen identifiziert werden (sensitiver Ansatz) (Guba 2008). Anzumerken ist, dass das Fehlen von Literatur jedoch nicht die Unwirksamkeit eines Verfahrens belegt.

Die systematische Literatursuche und die Entwicklung der Suchstrategie kann durch eine/n Informationsspezialistin bzw. -spezialisten durchgeführt werden. Eine solche Vorgangsweise ist vorteilhaft für die Qualität der Literatursuche, da Informationsspezialisten die einzelnen Datenbanken gut kennen und über deren Besonderheiten bezüglich der Suchbegriffe Bescheid wissen. Wichtig ist dabei eine enge Zusammenarbeit (insbesondere hinsichtlich der Suchbegriffe) mit dem Projektteam bzw. dem/der Projektkoordinator/in.

Alle Elemente der Literatursuche sind sowohl bei HTA-Berichten als auch bei Rapid Assessments und bei Updates zu bestehenden Berichten präzise und nachvollziehbar zu dokumentieren; im PRISMA Statement findet sich hierfür ein gutes Schema (Liberati et al. 2009, DACEHTA 2008, The Cochrane Collaboration 2008).

4.5.1 HTA-Bericht / Systematischer Review

4.5.1.1 Auswahl der Datenbanken für die systematische Literatursuche

Die Datenbankenauswahl hängt von der Forschungsfrage ab. Grundsätzlich zu konsultieren sind die beiden großen biomedizinischen Datenbanken Medline und Embase (in erster Linie zur Suche nach Primärstudien), die Datenbanken des Center for Reviews and Dissemination, des NICE, die Cochrane-Datenbanken sowie die DAHTA-Datenbank des DIMDI (Sekundärstudien wie systematische Reviews, Metaanalysen und HTA-Berichte) (DACEHTA 2008, Liberati 2009).

Je nach Fragestellung können zusätzliche Datenbanken herangezogen werden. Quellenangaben zu den Datenbanken finden sich im Anhang unter Punkt 15.3.1.

4.5.1.2 Suchstrategie für die systematische Literatursuche

Für die Entwicklung der Suchstrategie werden Suchbegriffe gesammelt, die sich aus den PICO-Komponenten ergeben, außerdem werden die Selektionskriterien z. B. für die Einschränkung des Studiendesigns, der Sprache oder des Zeitraums herangezogen. In der Regel werden außerdem themenrelevante Publikationen, die durch die Vorabrecherche identifiziert wurden, verwendet. Zu berücksichtigen ist, dass für unterschiedliche Datenbanken zum Teil verschiedene Suchstrategien formuliert werden müssen. In der Regel verlangen Datenbanken für Primärstudien aufgrund der viel größeren Anzahl an verschlagworteten Publikationen, aber auch wegen der bei Sekundärstudien oft weiter gefassten Fragestellungen, nach einer detaillierteren Suchstrategie als Datenbanken für Sekundärstudien. Für die meisten Datenbanken sind englische Begriffe notwendig. Diese Begriffe werden miteinander zu einer Suchstrategie verknüpft. Dafür werden Operatoren verwendet; die relevantesten sind die sogenannten Boole'schen Operatoren. Entweder erweitern sie die Suche, oder sie engen sie ein. Grundsätzlich gibt es folgende drei Operatoren:

1. AND (einschränkend, da nur Literaturstellen im Suchergebnis enthalten sind, die alle mit AND verknüpften Begriffe enthalten),
2. OR (erweiternd, da alle Literaturstellen im Suchergebnis enthalten sind, die nur einen oder mehrere der mit OR verknüpften Begriffe enthalten),
3. NOT (einschränkend, schließt Datenbankeinträge aus, in denen ein Begriff vorkommt).

Die ersten beiden Operatoren kommen in beinahe jeder Suchstrategie zum Einsatz. Der Operator NOT ist wenn überhaupt mit Bedacht einzusetzen, da er die Gefahr in sich birgt, dass durch ihn relevante Publikationen aus der Trefferauswahl fallen.

Die Suche in den Datenbanken erfolgt einerseits durch eine Freitextsuche, d. h., es wird in den Feldern der Datenbank mit Begriffen gesucht, die durch die Autoren/ Autorinnen festgelegt werden. Andererseits kann in den meisten Datenbanken mit standardisierten Schlagwörtern gesucht werden (subject heading). Diese Schlagwörter sind Teil eines Ordnungssystems (Thesaurus), in dem Begriffe systematisch dargestellt und durch synonyme Bezeichnungen beschrieben werden. Die Schlagwörter fassen diese Synonyme zusammen, wodurch ein kontrolliertes Vokabular entsteht. In der Medline-Datenbank heißt der Thesaurus MeSH (Medical Subject Headings), die Verschlagwortung (Indexierung) der Dokumente erfolgt durch sogenannte MeSH Terms (Schlagworte). Der Vorteil einer Indexierung der Artikel liegt darin, dass Synonyme und verschiedene Schreibweisen automatisch berücksichtigt werden und eine Suche nach Ober- und Unterbegriffen möglich ist. Der Nachteil liegt in der teilweise uneinheitlichen Verschlagwortung von Dokumenten, die sich zwischen den einzelnen Datenbanken, aber auch innerhalb einer Datenbank unterscheiden kann. Weiters können aktuelle, noch nicht verschlagwortete Artikel in der Datenbank nicht gefunden werden. Um die Qualität der Suchstrategie (wie oben beschrieben entwickelt) zu überprüfen, sollte kontrolliert werden, ob durch die systematische Suche allgemein bekannte relevante Primärpublikationen aufgefunden werden können. Fehlen wesentliche Primärpublikationen, so müssen die Ursachen hierfür geklärt werden; gegebenenfalls ist die Suchstrategie zu überarbeiten und die Suche zu wiederholen (Guba 2008).

4.5.1.3 Ergänzende Suche

Auch wenn die systematische Literatursuche noch so genau und gewissenhaft durchgeführt wird, so kann mit ihr trotzdem (fast) nie alle potenziell relevante Literatur aufgefunden werden. Einer ergänzenden Suche kommt daher je nach Fragestellung/Themengebiet unterschiedlich große Bedeutung zu. Im Rahmen eines HTA-Berichts sind die Handsuche sowie die Suche nach „grauer“ Literatur obligater Bestandteil der Literatursuche. Zusätzlich kann es sinnvoll sein, Expertinnen und Experten zu befragen.

4.5.1.3.1 Handsuche

Bei der Handsuche werden Referenzlisten der neuesten und/oder relevantesten Sekundärstudien, die durch die systematische Literatursuche aufgefunden wurden, kontrolliert. Außerdem kann in der Zitationsdatenbank Scopus gesucht werden (Quellenangaben hierfür finden sich im Anhang unter 15.3.2). Gegebenenfalls kann so um publizierte und durch die systematische Literatursuche nicht identifizierte passende Literatur ergänzt werden (LBI-HTA 2007).

4.5.1.3.2 Suche nach nicht publizierter (grauer) Literatur

Als graue Literatur werden sowohl Artikel bezeichnet, die nicht in den bibliographischen Datenbanken indiziert sind (z. B. von Patientenorganisationen oder medizinischen Fachgesellschaften, die ihre Studien nicht in Datenbanken veröffentlichen), als auch Studien, die (noch) nicht publiziert wurden (z. B. laufende klinische Studien). Zur grauen Literatur zählen Studien, die gar nicht veröffentlicht werden. Häufig werden Studien, die keine signifikanten Ergebnisse liefern, nicht veröffentlicht, was in der Folge zu einem Publikationsbias in der systematischen Übersichtsarbeit führen kann (wenn solche Studien nicht durch eine zusätzliche Suche ergänzt werden). Nicht veröffentlichte Studien finden sich z. B. in Studienregistern oder bei Zulassungsbehörden wie z. B. der European Medicines Agency (EMA) oder der amerikanischen Food and Drug Administration (FDA); Quellenangaben hierfür finden sich im Anhang unter 15.3.2 (LBI-HTA 2007, DACEHTA 2008, The Cochrane Collaboration 2008).

4.5.1.3.3 Identifikation von Literatur durch Expertinnen und Experten

Eine Befragung z. B. von Fachexpertinnen bzw. Fachexperten, externen Gutachterinnen bzw. Gutachtern, Patientenvertretungen oder Ansprechpartnern aus der Industrie nach relevanter Literatur kann ebenfalls als ergänzende Suche eingesetzt werden (LBI-HTA 2007).

4.5.2 Rapid Assessment

Bei Rapid Assessments wird die Literatursuche in der Regel mit einer höheren Spezifität als bei HTA-Berichten durchgeführt. Bei Rapid Assessments ist die methodische Vorgehensweise aufgrund zeitlich und finanziell begrenzter Ressourcen stärker eingeschränkt als bei HTA-Berichten. Wesentlich ist eine transparente Dokumentation des methodischen Vorgehens. Wo thematisch und inhaltlich möglich und sinnvoll, sollte wie im Folgenden beschrieben vorgegangen werden (IQWiG 2008a).

4.5.2.1 Systematische Suche nach Sekundärstudien

Zunächst wird nach bereits publizierten thematisch passenden Sekundärstudien gesucht (HTA-Berichte, systematische Reviews, Metaanalysen). Auf jeden Fall sollte die Datenbank des Center for Reviews and Dissemination und die Cochrane-Datenbank konsultiert werden. Des Weiteren können z. B. noch die Datenbanken von DIMDI und NICE herangezogen werden. Werden zur Fragestellung oder zu Teilbereichen der Fragestellung passende Sekundärstudien gefunden, so wird zunächst deren methodologische Qualität überprüft (siehe dazu Punkt 15.3).

4.5.2.2 Systematische Suche nach Primärstudien

Ist die Qualität der aufgefundenen HTA-Berichte und/oder systematischen Reviews ausreichend, wird der Review durch eine systematische Literatursuche nach Primärstudien ergänzt. Die Entwicklung der Suchstrategie erfolgt dabei wie bei HTA-Berichten (siehe 4.5.1.1). Auf jeden Fall wird dafür in Medline und Embase ab dem Zeitpunkt, zu dem die systematische Literatursuche der publizierten Sekundärstudie/n endete, eine systematische Suche nach Primärliteratur durchgeführt. Ist nur ein Teilbereich der Forschungsfragestellung abgedeckt, müssen auch die nicht abgedeckten Bereiche durch eine entsprechende systematische Literatursuche ergänzt werden.

Ist keine thematisch und methodisch passende Sekundärstudie vorhanden, wird in Medline und Embase eine systematische Literatursuche über einen eingeschränkten Zeitraum durchgeführt (je nach Fragestellung und Literaturverfügbarkeit unterschiedlich). Weitere themenspezifische Literaturdatenbanken können selbstverständlich ergänzt werden (LBI-HTA 2007).

4.5.2.3 Ergänzende Suche

Eine Handsuche sollte stets auch bei einem RA durchgeführt werden. Die Identifikation von „grauer“ Literatur und von Literatur durch Expertinnen und Experten kann bei Rapid Assessments bei Bedarf mit der gleichen Methodik wie bei HTA-Berichten durchgeführt werden (siehe 4.5.1.3). In der Regel kommt der ergänzenden Suche aufgrund zeitlich und finanziell begrenzter Ressourcen aber eine geringere Bedeutung als bei HTA-Berichten zu.

4.5.3 Adaption bestehender HTA-Berichte und Rapid Assessments

Der Umfang der Literatursuche bei der Adaption bestehender Berichte hängt von der Art des HTA-Produkts, das adaptiert werden soll, ab (HTA-Bericht / systematische Übersichtsarbeit oder Rapid Assessment). Bei Erstellung eines Updates ist die Literatursuche (in unterschiedlichen Umfang) in jedem Fall durchzuführen (siehe Kapitel 4.5.3.1). Dabei ist zu berücksichtigen, ob die ursprüngliche Fragestellung und die Selektionskriterien verändert werden sollen oder nicht. Je nachdem variieren auch die zu setzenden methodischen und ablauforientierten Arbeitsschritte.

Generell ist bei der Adaption bestehender HTA-Produkte zu beachten, dass ggf. Textänderungen in den Abschnitten zu Hintergrund und Methodik, in den Ergebnissen bzw. Schlussfolgerungen und Empfehlungen notwendig sind und diese Berichtsteile dementsprechend überarbeitet werden müssen. Jegliche Abänderungen bzw. Updates des Originalberichts müssen dokumentiert und für die Leserin / den Leser ersichtlich werden.

4.5.3.1 Systematische Suche bei Erstellung eines Updates

Bei der systematischen Literatursuche für die Erstellung eines Updates sind zu Beginn die Suchstrategie und die verwendeten Datenbanken des Originalberichts zu überprüfen.

Grundsätzlich ist die systematische Suche für das Update genauso wie im Originalbericht durchzuführen, wenn die ursprüngliche Fragestellung und die Selektionskriterien nicht verändert werden sollen. Das heißt, dass dieselben Datenbanken und dieselbe Suchstrategie wie im Originalbericht auch für die systematische Suche des Updates heranzuziehen sind. Zu beachten ist, dass die Suchzeiträume für das Update entsprechend angepasst werden müssen. Wird dabei keine neue Evidenz für das Update identifiziert bzw. entsprechen keine Studien den Einschlusskriterien des Updates, wird dieses Suchergebnis in den betreffenden Berichtsteilen dokumentiert. Identifiziert die systematische Suche neue Evidenz, die den Einschlusskriterien des Updates entspricht, wird die weitere methodische Vorgehensweise wie im Originalbericht beschrieben angewandt.

Bei einem Berichts-Update kann es jedoch auch notwendig sein, Fragestellung und (oder) Suchstrategie und (oder) Selektionskriterien des Originalberichts zu überarbeiten, beispielsweise bei Ein- oder Ausschluss einer Kontrollgruppe, einer Patienten-

population, eines Endpunkts, bei Änderungen der Indexierung, bei Änderungen der Syntax für die Suche bzw. bei Kritik an der letzten Suchstrategie etc. Das Update dient somit als (teilweise) Erweiterung bzw. Einschränkung der Fragestellung aus dem Originalbericht. Hierfür sollte die systematische Suche einerseits mit der ursprünglichen Suchstrategie und in den Datenbanken des Originalberichts durchgeführt werden. Andererseits sollte die systematische Literatursuche auch mit der abgeänderten Suchstrategie des Updates durchgeführt werden. Dabei können je nach Änderung der Fragestellung bzw. Suchstrategie zusätzliche Datenbanken herangezogen werden. Für die Suchen ist dabei der Zeitraum zwischen dem Endpunkt der Originalsuche und dem Suchzeitpunkt des Updates zu wählen. Gegebenenfalls kann es notwendig sein, die abgeänderte Suchstrategie des Updates auch auf den Zeitraum der systematischen Suche des Originalberichts auszuweiten, um alle relevanten Studien zu identifizieren.

Bei einer Änderung der Selektionskriterien sollte die systematische Suche des Originalberichts mit den veränderten, neuen Selektionskriterien auf relevante Studien überprüft werden. Bei Einschluss neuer Kontrollgruppen bzw. Outcomes sollten die bereits eingeschlossenen Studien des Originalberichts zusätzlich auf mögliche Informationen, die für die überarbeitete Fragestellung im Update relevant sind, überprüft werden.

4.5.3.2 Ergänzende Suche

Eine Handsuche und eine Suche nach „grauer“ Literatur sollten stets auch bei einem Berichts-Update durchgeführt werden. Zusätzlich kann es sinnvoll sein, Expertinnen und Experten zu befragen. Die Identifikation von Literatur kann dabei mit der gleichen Methodik wie bei HTA-Berichten durchgeführt werden (siehe 4.5.1.3), sollte sich jedoch auch am Originalbericht orientieren. Quellen, die beispielsweise im Originalbericht per Handsuche identifiziert wurden, sollten auf aktuellere Versionen für das Update überprüft werden.

4.6 Literaturselektion

Die systematische Literatursuche führt aufgrund des in erster Linie sensitiven Vorgehens (vgl. 4.5, „sensitiver Ansatz“) zu einer großen Anzahl an Treffern, die für die Bewertung in einem HTA-Bericht, einem Rapid Assessment bzw. eines Updates zu einem bestehenden Bericht nicht relevant sind. Die Auswahl der relevanten Publikationen erfolgt in einem mehrstufigen Selektionsprozess.

4.6.1 HTA-Bericht / Systematischer Review

Erstselektion (Selektion der Abstracts) und Zweitselektion (Selektion der Volltexte) erfolgen in der Regel durch zwei Projektmitarbeiter/innen unabhängig voneinander. Bei der Erstselektion kann alternativ zuerst eine Projektmitarbeiterin / ein Projektmitarbeiter alle Abstracts selektieren und ein zweiter im Anschluss nur die zunächst ausgeschlossenen. Bei abweichenden Bewertungen wird über Diskussion ein Konsens erzielt. Bei der Erstselektion kann – mit dem Ziel hoher Sensitivität – alternativ auch vereinbart werden, dass jede positive Bewertung zum Einschluss führt.

4.6.1.1 Erstselektion (Abstracts)

Anhand des Abstracts (bzw. wenn keines vorliegt: anhand des Titels) werden im Rahmen der Erstselektion alle Publikationen mit den vorab definierten Ein- und Ausschlusskriterien selektiert (siehe 4.2). Die Publikationen, die einem Ausschlusskriterium entsprechen, werden ausgeschlossen. Ist ein Abstract nicht eindeutig ein- oder auszuschließen, wird der Volltext bestellt.

Nach der Erstselektion wird im Berichtsplan (vgl. Kapitel 4.3) eine Literaturliste mit den eingeschlossenen Abstracts nach Erstselektion ergänzt.

4.6.1.2 Zweitselektion (Volltexte)

Für alle verbleibenden potenziell relevanten Publikationen wird der Volltext bestellt. Die Volltexte werden anhand der Selektionskriterien abermals selektiert, wobei es sinnvoll sein kann, die Kriterien vor diesem Schritt zu adaptieren. Beispielsweise kann die Mindeststudienlaufzeit (oder die Mindeststudienteilnehmerzahl), wenn ausreichend Publikationen vorhanden sind, für den Einschluss erhöht werden. Wenn eine Fragestellung nicht durch RCTs beantwortet werden kann, sollten auch Studien mit anderen Studiendesigns eingeschlossen werden, wenn die Studienqualität im Rahmen der Fragestellung vertretbar erscheint.

Das heißt, dass der endgültige Einschluss von Studien erst nach Prüfung der internen Validität geschehen kann bzw. wenn ein guter Überblick über die insgesamt verfügbare Evidenz gewonnen werden konnte.

Auf Volltextebene werden für jede ausgeschlossene Studie die Ausschlussgründe (laut Selektionskriterien) im Bericht angeführt (DACEHTA 2008, The Cochrane Collaboration 2008).

4.6.2 Rapid Assessment

4.6.2.1 Erstselektion (Abstracts)

Die Selektion der Abstracts aus den Primärdatenbanken erfolgt wie für einen HTA-Bericht / systematischen Review. Die Abstracts werden jedoch nur von jeweils einer erfahrenen Projektmitarbeiterin / einem erfahrenen Projektmitarbeiter selektiert. Das bedeutet, dass jedes Projektteammitglied einen Teil der vorliegenden Abstracts selektiert, nur unklare Abstracts werden noch einmal im Gesamtteam eingeschätzt. Ist ein Abstract nicht eindeutig ein- oder auszuschließen, wird im Zweifelsfall der Volltext bestellt.

Nach der Erstselektion wird im Berichtsplan (vgl. Kapitel 4.3) eine Literaturliste mit den eingeschlossenen Abstracts nach Erstselektion ergänzt (vgl. Kapitel 10; Prozesshandbuch GÖG/BIGQ Version 1.2010).

4.6.2.2 Zweitselektion (Volltexte)

Die Selektion der Volltexte erfolgt wie für einen HTA-Bericht / systematischen Review. Die Volltexte werden jedoch nur von jeweils einer Projektmitarbeiterin / einem Projektmitarbeiter selektiert. Das heißt, dass jedes Projektteammitglied einen Teil der vorliegenden Volltexte selektiert. Jene Volltexte, die nicht klar zuteilbar sind, werden im Gesamtteam noch einmal gemeinsam durchgeschaut (vgl. Prozesshandbuch GÖG/BIGQ Version 1.2010).

4.6.3 Adaption bestehender Assessments

Die durch die Literatursuche neu identifizierten Studien eines Updates werden, wie bei HTA-Berichten und Rapid Assessment, ebenfalls einer Erstselektion auf Abstractebene und einer Zweitselektion auf Volltextebene unterzogen (siehe Kapitel 4.6.1 bzw. 4.6.2).

Bei gleichbleibender Fragestellung und gleichbleibenden Selektionskriterien werden die neu identifizierten Studien im Update grundsätzlich wie im Originalbericht bewertet, d. h., es werden die gleichen Ein- und Ausschlusskriterien angewandt.

Bei Identifizierung potenzieller neuer Studien für die Erstellung eines Updates mit adaptierter Fragestellung und/oder adaptierten Selektionskriterien sind die veränderten Selektionskriterien anzuwenden. Jegliche Adaption der Selektionskriterien muss im Update dokumentiert werden.

4.7 Exkurs: Klassifikation von Studien

Die Wahl des geeigneten Studientyps ist ein wichtiger Aspekt des Designs medizinischer Studien. Der Studientyp bestimmt wesentlich wissenschaftliche Qualität und Aussagekraft einer Untersuchung und wird durch die Fragestellung bzw. die vorhandenen Ressourcen determiniert.

Bislang konnte sich international kein einheitliches Klassifikationsschema der Studientypen durchsetzen, u. a. aus dem Grund, dass sich Studien nach unterschiedlichen Aspekten systematisieren lassen. Die folgende Begriffserklärung und die Abbildung 4.1 dienen dazu, eine grobe Übersicht über unterschiedliche Einteilungsmöglichkeiten zu gewinnen.

Prinzipiell können Studien in der klinischen und epidemiologischen Forschung nach den übergeordneten Merkmalen **1) analytisch versus deskriptiv**, **2) interventionell versus rein beobachtend** und **3) prospektiv versus retrospektiv** unterteilt werden, wobei fließende Übergänge zwischen den jeweiligen Studientypen gängig sind und die Anforderungen an Planung, Durchführung und statistische Auswertung bei den einzelnen Typen gleichermaßen hoch sein sollten.

Analytisch: Grundsätzlich werden in analytischen Studien die Determinanten von Erkrankungen und die Wirksamkeit von Interventionen untersucht. Analytische Studien dienen der Überprüfung von Hypothesen und versuchen kausale Zusammenhänge zwischen einer Intervention oder Exposition und einem definierten Endpunkt herzustellen.

Deskriptiv: Deskriptive Studien dienen der Beschreibung der Verteilung von Erkrankungen in einer definierten Population und der Generierung neuer Hypothesen, erlauben jedoch keine Beurteilung von kausalen Zusammenhängen. Deskriptive Studien sind für HTA-Berichte in der Regel dann relevant, wenn keine für die Beurteilung eines kausalen Zusammenhangs geeignete Evidenz gefunden werden konnte.

Interventionell: Interventionsstudien haben immer analytischen Charakter, sind prospektiv (siehe unten) und beobachten die Studienpopulation, nachdem eine Intervention (deren Effektivität und Sicherheit geprüft werden soll) gesetzt wurde.

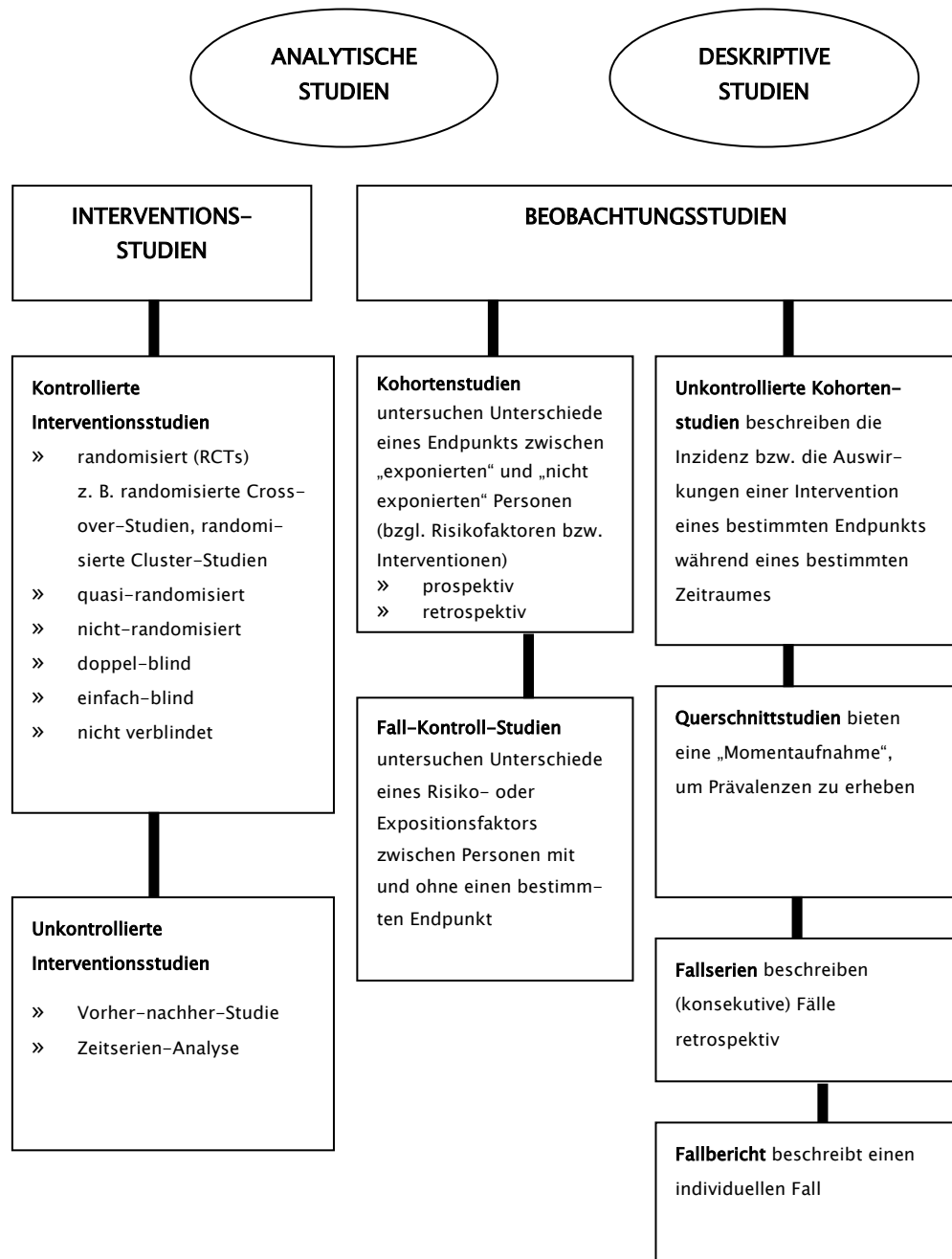
Beobachtend: In Beobachtungsstudien werden Gruppen mit und ohne Intervention bzw. Gruppen mit unterschiedlichen Interventionen beobachtet. Es werden dabei keine Experimente oder zusätzlichen Interventionen durchgeführt. Beobachtungsstudien können pro- oder retrospektiv und analytisch oder deskriptiv sein.

Prospektiv: Prospektive Studien sind vorausschauend, wobei die Daten zur Exposition vor der Erfassung des Ereignisses erhoben werden. Vorteilhaft bei prospektiven Studien ist, dass im Vergleich zu retrospektiven Studien eine höhere Datenqualität erzielt werden kann durch Ermittlung des optimalen Stichprobenumfangs; außerdem ermöglichen sie, Daten entsprechend der genauen Fragestellung zu erheben. Auch können bestimmte Formen des Informationsbias (z. B. Recallbias) bei prospektiven Studien verhindert werden. Dem gegenüber steht ein hoher zeitlicher, organisatorischer und finanzieller Aufwand.

Retrospektiv: Retrospektive Studien sind rückblickend; die Ereignisse, die von Interesse sind, sind in der Vergangenheit eingetreten. Dem Vorteil des zeitlich und finanziell geringen Aufwands steht häufig eine mangelnde Datenqualität aufgrund unzureichender Dokumentation der zu untersuchenden Ereignisse gegenüber.

Die folgende Abbildung beruht auf modifizierten Schemata zur Einteilung von Studientypen der Task Force on Community Preventive Services der CDC (Centers for Disease Control 2006) und des NICE (NICE 2006).

Abbildung 4.1:
Darstellung zur Einteilung von Studientypen



Quelle: modifiziert nach CDC 2006 und NICE 2006; GÖG/BIQG-eigene Darstellung

Genauere Beschreibungen zu den einzelnen Studientypen finden sich in Kapitel 5.2. Die im Folgenden dargestellte Tabelle 4.7 zeigt einen Überblick über die wichtigsten Studientypen.

Tabelle 4.7:
Überblick über Vor- und Nachteile unterschiedlicher Studiendesigns

Studiendesign	Beschreibung	Vorteil	Nachteil
RCT	<p>Probanden werden rekrutiert. Die Intervention wird nach dem Zufallsprinzip zugeteilt. Beobachtung über die Zeit und Erfassung des Endpunktes. Es gibt unterschiedliche Typen von RCTs, zwei wesentliche sind: Randomisierte Cross-over-Studien: Alle Teilnehmer/innen erhalten alle Interventionen, aber in unterschiedlicher Reihenfolge. Die Reihenfolge wird randomisiert. Randomisierte Cluster-Studien: Ein Cluster von Personen, z. B. eine ganze Klinik und nicht Einzelpersonen, wird für die Studie herangezogen.</p>	<p>Klarer kausaler Zusammenhang zwischen Intervention und Endpunkt Erfassung der Inzidenz Hohe interne Validität</p>	<p>Finanziell aufwendig (Häufig) geringe externe Validität¹</p>
Kohortenstudie	<p>Probanden werden bei prospektiven Kohortenstudien rekrutiert, bei retrospektiven Kohortenstudien werden die Gruppen im Nachhinein gebildet. Messung des Risikofaktors Beobachtung über die Zeit und Erfassung des Endpunktes zu einem festgelegten Zeitpunkt</p>	<p>Klarer kausaler Zusammenhang zwischen Intervention und Endpunkt Erfassung der Inzidenz Geeignet zum Studium seltener Risikofaktoren und mehrerer Endpunkte</p>	<p>Zeitlich, organisatorisch und finanziell aufwendig Kein Hypothesenbeweis Nicht geeignet zum Studium seltener Endpunkte</p>

Fortsetzung nächste Seite

¹
Siehe Kapitel 6.2.

Fortsetzung Tabelle 4.7 – Seite 2 von 2

Studiendesign	Beschreibung	Vorteil	Nachteil
Fall-Kontroll-Studie	Fälle werden gesammelt Kontrollen werden ausgewählt Messung des Risikofaktors retrospektiv	Kosten- und zeitgünstig Mehrere Risikofaktoren möglich Geeignet zum Studium mehrerer Risikofaktoren und seltener Endpunkte Benötigt kleinere Fallzahlen als Kohortenstudie	Anfällig für Fehlerquellen Kein Hypothesenbeweis Oft unklarer zeitlicher Zusammenhang zwischen Risikofaktor und Endpunkt Nicht geeignet zum Studium seltener Risiko- faktoren Erfassung der Inzidenz meist nicht möglich
Fallbericht/Fallserie	Bericht über individuellen Krankheitsfall / Reihe von aufeinanderfolgenden Krankheitsfällen	Verwendung in der Arbeitsmedizin und bei der Beobachtung von Nebenwirkungen Geeignet für Lehr- und Fortbildungszwecke bei neuen und/oder seltenen Erkrankungen	Geringe externe Validität Kein Hypothesenbeweis
Querschnittstudien	Messung von Risikofaktor und Endpunkt zum selben Zeitpunkt in einer definierten Population	Kostengünstig	Kein Hypothesenbeweis Erfassung der Inzidenz nicht möglich

Quelle und Darstellung: GÖG/BIQG

4.8 Exkurs: Hierarchie der Studien

Auch bezüglich der Rangordnung von Studien für die Erstellung von HTA-Berichten / systematischen Übersichtsarbeiten besteht international kein Konsens. Unumstritten ist, dass unterschiedliche Studiendesigns unterschiedliche Limitierungen (z. B.: Studiengröße und Länge der Studie) und unterschiedliche Anfälligkeit für das Auftreten von Bias und Confoundern aufweisen. Am breitesten anerkannt und akzeptiert ist die folgende hierarchische Grobeinteilung von Studientypen (Tabelle 4.8):

Tabelle 4.8:
Vereinfachte Darstellung der Hierarchie der Evidenz
unterschiedlicher Studientypen in absteigender Reihenfolge

Systematische Übersichtsarbeiten und Metaanalysen von RCTs (wobei die einzelnen RCTs zusätzlich je nach Qualität unterteilt werden)
RCTs (randomisierte Cross-over-Studien, randomisierte Cluster-Studien)
Quasi experimentelle Studien: nicht randomisierte Interventionsstudien (CT), Vorher-nachher-Studien, Zeitserien-Analyse
Prospektive und retrospektive kontrollierte Beobachtungsstudien: Kohortenstudien, Fall-Kontroll-Studien, Fallserien

Quelle: adaptiert nach CRD 2009; GÖG/BIQG-eigene Darstellung

Studien, die weiter unten in der Hierarchie rangieren, sind meistens anfälliger für das Auftreten von Fehlerquellen als jene, die sich im oberen Bereich befinden. Es ist jedoch zu bedenken, dass Daten aus RCTs durch ausgewählte Populationsgruppen oder Vergleichsinterventionen und kurze Beobachtungsdauern limitiert sind. Daher ist bei der Verwendung von RCTs die Ergänzung durch hochqualitative Beobachtungsstudien hilfreich. Unabhängig von der Rangordnung wird der Evidenzgrad jeder einzelnen Studie durch ihre Qualität und aktuelle Relevanz bestimmt, weshalb die präsentierte Hierarchie der Studienevidenz nicht als dogmatisch zu betrachten ist.

Bei Gesundheitsförderung, aber auch in manchen Bereichen der Primärprävention ist die Orientierung an RCTs oft nicht oder nur indirekt möglich, da bei RCTs über einen längeren Zeitraum meist kein kausaler Zusammenhang nachweisbar ist (der für primärpräventive und gesundheitsfördernde Maßnahmen aber relevant ist). Außerdem ist eine Randomisierung zwar möglich, allerdings sind in einem Randomized controlled trial (RCT) eher passive Studienteilnehmer/innen (in Hinblick auf mögliche Interaktionen) erwünscht, die sich streng an das vorgegebene Protokoll halten. Dieser Ansatz stellt eine Antithese zu komplexen sozialen Interventionen z. B. im Rahmen von Gesundheitsförderungsmaßnahmen dar, da hier aktive Teilnahme und Eigeninitiative zentral für den Erfolg sind (Schug et al. 2006). In einem solchen komplexen Fall sind andere Studiendesigns verstärkt heranzuziehen.

Autorinnen und Autoren: Barbara Fröschl, Sophie Brunner-Ziegler, Alexander Eisenmann, Daniela Pertl (GÖG/BIQG)

5 Zusätzliche Datengenerierung und Datenanalyse

5.1 Grundlagen

Dieses Kapitel stellt Methoden zur Generierung, Analyse und Verwendung von Daten dar. An dieser Stelle muss darauf hingewiesen werden, dass eine eigene Erhebung von Primär- oder Sekundärdaten insbesondere dann sinnvoll ist, wenn für die Beantwortung der Frage nicht ausreichend publizierte (und den Selektionskriterien entsprechende) Literatur vorhanden ist (siehe Kapitel 4.5) bzw. wenn sich diese als inadäquat herausstellt. Relevant sind eigene Datenerhebungen beispielsweise auch, um etwaige publizierte Literatur an den lokalen Kontext anzupassen. Eine eigene Erhebung kann nur im Rahmen eines ausführlichen HTA-Berichts durchgeführt werden. Die zeitlichen und finanziellen Ressourcen von Rapid Assessments reichen nicht dafür aus.

In jedem Anwendungsbereich der empirischen Sozialforschung und also auch in den Gesundheitswissenschaften muss seitens der Forscherinnen und Forscher besonders auf die Qualitätssicherung im Forschungsprozess geachtet werden. Die folgenden Ausführungen sollen die zugrunde liegenden Qualitätskriterien bzw. die Anwendbarkeit der verschiedenen Methoden verständlicher machen und den Autorinnen und Autoren somit die Durchführung einer Studie erleichtern. Hier wird auf jene Methoden fokussiert, die für die Gesundheitsforschung von besonderer Relevanz sind. Im qualitativen Bereich sind dies vor allem unterschiedliche Varianten von Interviews, im quantitativen Bereich sind Fragebogenerhebungen sowie registerbasierte Studien zu nennen. Für ausführlichere Darstellungen der gesamten Methodenpalette der empirischen Sozialforschung siehe Schnell, Hill & Esser 1995; Bortz/Döring 2002 sowie Diekmann 2003 und Diekmann 2004.

Autoren und Autorinnen, die sich für eine eigene Datenerhebung entscheiden, sollten sich mit dem Spektrum an empirischen Forschungsmethoden auseinandersetzen.

5.1.1 Grundprinzipien sozialwissenschaftlicher Methoden

Innerhalb der sozialwissenschaftlichen Forschung kann unterschieden werden zwischen qualitativer und quantitativer Forschung. Je nach Fragestellung, Erkenntnisinteresse, theoretischer Perspektive und/oder vorhandenen Ressourcen können unterschiedliche Methoden angewendet werden. Als wichtigste Methoden der qualitativen

Forschung gelten Interviews, Beobachtung und andere Formen der Feldforschung. Fragebogenerhebung kann sowohl als qualitatives wie als quantitatives Messinstrument verwendet werden, gilt jedoch primär als quantitatives Verfahren.

Maßgeblich bei einer empirischen Forschung ist, was erforscht werden soll. Konkret bedeutet dies, dass sich die Wahl des Erhebungsinstruments von der Forschungsfrage (der HTA-Fragestellung) ableitet. Zunächst bedarf es deshalb einer grundlegenden Klärung des Erkenntnisinteresses, der Fragestellung und darauf aufbauend der Vorgehensweise.

Die Beantwortung folgender Fragen ist hierbei hilfreich.

- » Welche unterschiedlichen (qualitativen sowie quantitativen) Methoden existieren?
- » Welche Methode eignet sich am besten zur Erforschung der spezifischen Fragestellung?
- » Welche Methode eignet sich am besten zur Analyse der zu generierenden Daten?
- » Auf welchen theoretischen Konzepten beruht die Methode?
- » Welche Ressourcen stehen mir zur Verfügung?

Die drei wichtigsten Grundprinzipien der Qualitätssicherung in der empirischen Forschung sind die (interne sowie externe) Validität, die Reliabilität des Messverfahrens sowie die Objektivität des Messinstruments. Diese Kriterien müssen erfüllt sein, um eine erfolgreiche und effiziente Forschung durchführen zu können. Zu den Gütekriterien empirischer Forschung im Detail siehe Diekmann 2003, 216–227.

Befragungsmethoden zählen zu den am häufigsten verwendeten Methoden innerhalb der empirischen Forschung. Hier ist zu unterscheiden zwischen schriftlichen und mündlichen Befragungen (Interviews). Welche Methode im konkreten Fall vorzuziehen ist, hängt primär von dem Forschungsvorhaben und der zu erforschenden Fragestellung ab. Aber auch die finanziellen und zeitlichen Ressourcen, die zu befragende Personengruppe und die beabsichtigte Nutzung der Ergebnisse spielen bei der Entscheidung eine Rolle.

Qualitative Methoden eignen sich besonders für Fragestellungen mit Patientinnen- und Patienten-Bezug in Verbindung mit Organisationsstrukturen oder für die Erforschung von Zusammenhängen und Verbindungen zwischen einzelnen HTA-Elementen. Quantitative Verfahren sind primär im Bereich der Epidemiologie und Gesundheitssystemforschung von Nutzen (DACEHTA 2008). Bei einzelnen Forschungsvorhaben im HTA-Bereich kann jedoch auch eine Kombination unterschiedlicher Verfahren von Vorteil sein. So könnte beispielsweise zur Konkretisierung der Forschungsfrage zuerst ein qualitatives Verfahren durchgeführt werden, um in einem weiteren Schritt eine größer angelegte Fragebogenerhebung zu realisieren.

Generell werden qualitative Verfahren vor allem zur ersten Orientierung bzw. zu explorativen Zwecken eingesetzt. Wenn noch relativ wenig über das Forschungsgebiet bekannt ist, bieten sich qualitative Methoden zur detaillierten Informationsbeschaffung eher an als quantitative Verfahren. Besonders auch zur Erforschung komplexerer Zusammenhänge und subjektiver Meinungen werden qualitative Verfahren bevorzugt verwendet. Im Gegensatz dazu zeichnen sich quantitative Verfahren vor allem durch ihre höhere Standardisierung und die Möglichkeit zur numerischen Messung aus. Das quantitativ gewonnene Datenmaterial ermöglicht somit eine „objektivere“ Darstellung. Angewendet werden quantitative Verfahren vor allem zur Befragung größerer Populationen und spezifischer Fragestellungen.

Ein nicht zu vernachlässigendes Element während der Forschungsplanung und -durchführung stellt die Komponente der Datenauswertung dar. Die Frage, wie die generierten Daten ausgewertet werden und vor allem, welche zeitlichen Ressourcen für die Auswertung benötigt werden, sollte stets mit berücksichtigt werden. Der Aufwand der Datenauswertung wird oftmals unterschätzt. Wie es unterschiedliche Methoden der Datengenerierung gibt, existieren auch unterschiedliche Formen zur Auswertung der Daten. Die Transkription von Interviews und die inhaltliche Interpretation der Daten (im qualitativen Bereich) bzw. statistische Auswertungsverfahren (im quantitativen Bereich) – all dies benötigt Zeit und sollte im veranschlagten Zeitplan für das HTA-Projekt berücksichtigt werden. Die Auswertung der Erhebungsergebnisse ist nicht weniger zentraler Bestandteil einer empirischen Forschung als die gründliche Forschungsplanung und die Durchführung.

5.1.2 Interviews

Das Interview ist das am weitesten verbreitete Messverfahren innerhalb der qualitativen Sozialforschung, was ihm auch den Beinamen „Königsweg der Sozialforschung“ (König 1972) eintrug. Das Instrumentarium Interview existiert jedoch nicht, Interviews variieren in Form und Methodik. Je nach Anzahl der befragten Personen kann zunächst unterschieden werden zwischen Einzelinterviews („Face-to-face“-Befragung) und Gruppeninterviews („Fokusgruppendifkussion“ und „Delphi-Verfahren“).

5.1.2.1 Einzelinterviews

Die Durchführung bzw. die Form des Interviews hängt stark mit seiner Strukturierung und dem Grad der Standardisierung zusammen. Zwischen den beiden Polen „völlig strukturiert“ und „völlig offen“ gibt es unterschiedliche Variationen und Mischformen.

Das standardisierte Interview zeichnet sich durch vorgegebene Fragen (meistens auch mit vorgegebenen Antwortkategorien) aus, welche alle in derselben Reihenfolge gestellt werden. Das Ziel dieser Methode ist eine möglichst genaue, vergleichbare und „objektive“ Erhebung. Dieses Vorgehen eignet sich jedoch nur für Themenbereiche, über die bereits genug Vorkenntnisse bestehen.

Im Gegensatz dazu charakterisiert sich das nicht-standardisierte Interview durch seine offene Gesprächsführung und minimale Strukturierung. Es bedarf nur eines theoretischen Rahmens und weniger Vorkenntnisse vonseiten des/der Interviewenden, weshalb es sich vor allem für explorative Studien eignet. Auch für schwierige Themenbereiche, die einen sensiblen Umgang der Interviewerin oder des Interviewers erfordern, wird diese Form des Interviews verstärkt eingesetzt.

In der Mitte dieser beiden Extreme anzusiedeln wären Mischformen, die strukturierte Fragebogenelemente mit offenen Fragen und einer offenen Interviewführung vereinen. Exemplarisch für diese halbstandardisierten Interviews ist das Leitfadeninterview, bei dem Themenbereiche und Inhalte vorgegeben sind, die Interviewerin bzw. der Interviewer jedoch Spielraum in der Gestaltung des Gesprächs hat.

Generell eignet sich das Einzelinterview vor allem zur Erforschung von Meinungen, Erfahrungen, Wünschen, Motiven, Kenntnissen und Verhaltensweisen diverser Personen (Patientinnen und Patienten, Kundinnen und Kunden, Bürgerinnen und Bürger etc.) zu der zu beurteilenden Maßnahme oder Technologie. Eine andere Anwendungsvariante des Einzelinterviews kann die Befragung von Expertinnen und Experten zu spezifischen Themengebieten und Fragestellungen in Bezug auf das Gesundheitssystem sein. Besonders bei mangelnder Literatur und bei schwammigen Forschungsfragen können Expertinnen und Experten zur Informationsbeschaffung beitragen bzw. bei der Konkretisierung der Fragestellung behilflich sein.

Nachteile des Einzelinterviews ergeben sich vor allem durch die relativ aufwendige Form der Datengenerierung. Bei größeren Forschungsprojekten mit einem umfangreichen Befragungskollektiv ist diese Variante der qualitativen Forschung kaum einsetzbar bzw. mit hohen Kosten und Ressourcen verbunden. Auch die Auswertung der Daten kann (je nach Methode oder Analyseverfahren) relativ aufwendig und zeitintensiv sein. Die Transkription und die inhaltliche Analyse der Interviews erfordert Zeit – dies gilt es in der Konzeption und der Planung zu bedenken. Literaturhinweise zu qualitativen Methoden bzw. Interviews finden sich bei Froschauer/Lueger 2003, Rubin 2004 und Flick 2006¹.

1

Im Internet können kostenlos Artikel der Zeitschrift Forum Qualitative Sozialforschung eingesehen werden, teilweise auch mit Gesundheitsbezug (<http://qualitative-research.net>).

5.1.2.2 Fokusgruppeninterviews

Bei Fokusgruppeninterviews handelt es sich um Diskussionen mehrerer Personen (in der Regel bis zu zehn Beteiligte) zu einem spezifischen Thema (Morgan 1988). In Bezug auf den Gesundheitsbereich könnte ein mögliches Thema beispielsweise die Verbesserung der Notfallversorgung von Unfallverletzten sein, wobei hier Rettungsbedienstete bzw. Angestellte von Notfallambulanzen und Spitälern die „Expertinnen und Experten“ darstellen (Geyer/Siegrist 1998). Die Zusammensetzung der Fokusgruppen sollte vorerst möglichst heterogen sein, um so die gesamte Vielfalt und Bandbreite eines Themengebietes abzudecken. In einer zweiten Etappe empfiehlt es sich, Interviews mit homogeneren Fokusgruppen durchzuführen. Die Interviewerin oder der Interviewer gibt das Thema vor und strukturiert die Diskussion.

Das Spezifische dieser Methode liegt in der Möglichkeit, durch Interaktion verschiedener Personen eine große Bandbreite an Daten zu generieren und Aspekte zur Fragestellung in den Vordergrund zu rücken, die in einem Einzelinterview eventuell untergegangen oder nicht ans Tageslicht getreten wären. Die Dynamik einer Gruppeninteraktion kann dazu beitragen, neue bzw. mehr Erkenntnisse zu produzieren. Bei der Erarbeitung von Lösungskonzepten im Gesundheitsbereich sind Gruppendiskussionen tendenziell produktiver als Einzelinterviews. Weitere Vorteile dieser Methode im Gegensatz zu Einzelinterviews liegen in der Tatsache, dass

- » durch die Kommunikation interessante Aspekte zur Fragestellung auftauchen,
- » eine förderliche, „natürlichere“ Atmosphäre geschaffen wird,
- » unterschiedliche Perspektiven vertreten werden,
- » das Abdriften in eine Richtung verhindert wird.

Besonders eignet sich die Methode der Fokusgruppeninterviews bei einer erstmaligen Orientierung im Forschungsfeld, als Unterstützung zur Entwicklung eines eigenen Fragebogens oder Interviewleitfadens sowie zur Diskussion über die Ergebnisse vorheriger Studien. Nachteile dieser Methode liegen vor allem in der mangelnden Vergleichbarkeit und Strukturierung der Diskussionen. Gleichzeitig bedeuten Fokusgruppendifkussionen einen erhöhten Aufwand bei Durchführung, Aufzeichnung und Interpretation (Flick 2006). Literatur zu Fokusgruppeninterviews siehe Morgan 1988, Loos/Schäffer 2001 und Bohnsack 2000.

5.1.2.3 Delphi-Verfahren

Das Delphi-Verfahren ist eine spezielle Variante der Gruppendiskussion. Hier wird – wie bei der Fokusgruppendifkussion – eine Gruppe von Expertinnen und Experten zu einem bestimmten Thema befragt, jedoch in schriftlicher Form. Ziel der Interaktion soll

es sein, für ein komplexes Problem gemeinsam ein Lösungskonzept zu erarbeiten. Anwendung findet das Delphi-Verfahren vor allem bei schwierigen Themenkomplexen, bei denen die einzelnen Lösungsbeiträge zu einem umfassenden Lösungsvorschlag zusammengefasst werden. Im Gesundheitsbereich könnten dies beispielsweise Maßnahmen zur Prävention von Geschlechtskrankheiten sein. Die Erfahrungen bzw. Kenntnisse von Expertinnen und Experten unterschiedlicher Fachrichtungen sollen den Prozess der Problemlösung unterstützen. Im HTA-Bereich stellen sich Delphi-Verfahren des Weiteren auch zur Abstimmung von Prioritäten als sinnvoll dar. Welchen Themen (bzw. Subfragestellungen) in einem spezifischen Projekt besondere Beachtung geschenkt werden sollten, kann durch die Delphi-Methode am besten ermittelt werden. Bei mangelnder Evidenz können Delphi-Verfahren auch zum Zwecke der Themenfindung oder zur Beantwortung der Forschungsfrage durchgeführt werden.

Konkret läuft das Delphi-Verfahren in mehreren Schritten ab:

1. Eine Gruppe von Expertinnen und Experten unterschiedlicher Disziplinen bzw. Fachrichtungen wird mittels eines standardisierten Fragebogens zum interessierenden Sachverhalt befragt.
2. Die Studienautorinnen und -autoren fertigen auf Basis der ausgewerteten Fragebögen einen überarbeiteten Fragebogen an. Dieser beinhaltet gleichzeitig die Lösungsvorschläge der Expertengruppe.
3. In einer zweiten Runde können die befragten Expertinnen und Experten nach Durchsicht der anderen Vorschläge den eigenen Beitrag jeweils korrigieren, überarbeiten und/oder adaptieren.
4. Diese Befragungsrunden werden so lange durchgeführt, bis keine weiteren Informationen zu erzielen sind (meistens zwei bis drei Runden).
5. Auf Basis dieser Informationen erstellen die Studienautorinnen und -autoren ein konsensuales Lösungskonzept.

Ein großer Vorteil der Delphi-Methode liegt in der Anonymität der Teilnehmerinnen und Teilnehmer. Im Unterschied zu anderen Gruppendiskussionsverfahren besteht in diesem Verfahren nicht die Gefahr, „das Gesicht zu verlieren“. Antworten können in einer nächsten Runde korrigiert oder geändert werden. Außerdem kann folgender Nachteil der Gruppendiskussion vermieden werden: die Beeinflussung der persönlichen Meinung einzelner Gruppenmitglieder durch dominierende Personen oder durch Druck zur Konformität. Eine sorgfältige Auswahl von Expertinnen und Experten aus verschiedenen Fachrichtungen stellt die Grundlage für ein Funktionieren dieser Methode dar. Ein Nachteil des Verfahrens ist der hohe Zeitaufwand (zwischen den unterschiedlichen Wellen sollten mindestens zwei Wochen Abstand sein). Weiterführende Literatur zum Delphi-Verfahren siehe Linstone/Turoff 1975, Häder 2002 und Schulz 2009.

5.1.3 Fragebogenerhebung

Die Fragebogenerhebung eignet sich primär für die Befragung größerer Gruppen, da sie ein verhältnismäßig kostengünstiges Verfahren darstellt. Vor allem in Verbindung mit der Erforschung individueller Verhaltensweisen und Charakteristika werden Fragebögen eingesetzt. Diese werden häufig im quantitativen Bereich eingesetzt. Im Unterschied zur qualitativen Datengenerierung mittels Interviews sollten diese quantitativen Methoden der empirischen Sozialforschung eine hohe Strukturiertheit aufweisen. In den Gesundheitswissenschaften eignet sich die Methode der standardisierten quantitativen Befragung besonders für Gesundheitssystemforschung und Epidemiologie.

Fragebögen gehören zu den komplexesten und arbeitstechnisch aufwendigsten Forschungsmethoden innerhalb der empirischen Sozialforschung. Sowohl die Konzeption der Fragebögen, die Durchführung der Erhebung als auch die Auswertung erfordern Zeit und Ressourcen. Diese Faktoren sollten bei der Entscheidung für eine Fragebogenerhebung mitbedacht werden. Nichtsdestotrotz gibt es Bereiche und Forschungsfelder, in denen die Anwendung von Fragebögen unerlässlich ist. Forschungen im Bereich Lebensqualität, Stress oder Mental vulnerability sind ohne Fragebögen schwer durchführbar. Im Bereich des Risikoverhaltens hat sich auch primär die Fragebogenerhebung als sinnvoll herausgestellt (DACEHTA 2008). In HTA-Projekten empfiehlt sich der Einsatz von Fragebögen vor allem für die Erhebung von Kostendaten (z. B. Medikamentenkosten).

Wie bei allen sozialwissenschaftlichen Methoden muss auch hier auf Validität, Reliabilität, Objektivität und Praktikabilität geachtet werden. Anders als bei qualitativen Verfahren können jedoch bei einem Fragebogen Schwächen oder Fehler (z. B. bei der Frageformulierung) im späteren Forschungsprozess schwer bis gar nicht korrigiert werden. Während der Vorbereitung des Fragebogens bis hin zur Auswertung sollten deshalb bestimmte Qualitätskriterien beachtet werden.

Kriterien:

- » Antwortkategorien sollten so präzise und vergleichbar wie möglich sein.
- » Verfahren und Ergebnisse sollten reproduzierbar sein.
- » Ergebnisse sollten mit anderen Studien vergleichbar sein.
- » Fragebogen muss ethischen Kriterien entsprechen (DACEHTA 2008).

Um die angestrebte Qualität des Fragebogens zu gewährleisten, sollte dieser vor der Verwendung getestet werden. Das häufigste Mittel zur Qualitätssicherung von Fragebögen besteht in der Durchführung einer Pilotphase (kognitives Interview und anschließende Pretests). Zum Zweck einer ersten Überprüfung wird der Fragebogen ausgewählten Personen der Zielgruppe zum Ausfüllen und Beurteilen vorgelegt.

Durch die Evaluierung dieses ersten Durchgangs sollen Fehlerquellen im Messinstrument vermieden und die Qualität des Fragebogens gesichert werden.

Insgesamt lassen sich folgende Vor- und Nachteile der Fragebogenerhebung festhalten:

Vorteile:

- » geringer Interviewer-Effekt
(Verzerrung durch Meinungen und Einstellungen des Interviewers)
- » geringere Kosten und geringerer Aufwand als bei anderen Verfahren
- » längere Bedenkzeit bei der Beantwortung der Fragen
- » Möglichkeit der „Anonymität“ der Befragten
- » höhere Standardisierung des Befragungsinstruments

Nachteile:

- » bei Problemen keine Hilfe durch Interviewerin oder Interviewer
- » unkontrollierte Erhebungssituation (Gefahr mangelnder Vergleichbarkeit)
- » Fragebogen muss selbsterklärend sein (großer Aufwand in der Erstellung)

Autorin: Maria Hörtner

5.1.4 Zusammenfassende Darstellung der Methoden

Tabelle 5.1:
Übersicht über die dargestellten Methoden

	Studientyp	Vorteile	Nachteile	Forschungsinteresse	Zeithorizont	Anwendungsgebiete
Einzel- interview	qualitativ	Flexible Gestaltung der Interviewsituation Subjektive Sicht des/der Interviewpartners/in kommt zur Geltung	Nur für kleine Befragungs- gruppe anwendbar Auswertungsprobleme Keine „objektiven“ Daten	Explorative Zwecke Erforschung „subjektiver“ Meinungen und Bedeutungen	Hoher Aufwand bei der Auswertung Transkription mit zu berücksichtigen	Expertenbefragung Literaturbeschaffung Konkretisieren der For- schungsfrage Beantwortung der Forschungsfrage
Fokus- gruppen	qualitativ	Gruppendynamik bringt neue Impulse Unterschiedliche Perspektiven	Mangelnde Vergleichbarkeit Schwer zu planender Ablauf	Erforschung von Meinungen und Einstellungen Erstmalige Orientierung im Forschungsfeld	Hoher Aufwand bei der Auswertung	Erarbeiten von Lösungs- konzepten
Delphi- Verfahren	qualitativ	Anonymität Kein Gruppendruck (z. B. Druck zur Konformität etc.)	Sorgfältige Auswahl von Expertinnen und Experten ist Grundlage für Funktionieren	Individuelles Urteil von Fachexpertenschaft Ideenfindung oder -bewertung	Hoher Zeitaufwand durch die wiederholten Befragungsrunden (mind. 2 Wochen Abstand dazwischen)	Erarbeiten von Lösungs- konzepten Themenfindung Themenpriorisierung Beantworten der Forschungs- frage (bei mangelnder Evidenz)
Fragebogen- erhebung	quantitativ	Kostengünstig Standardisiertes Befragungs- instrument Hohe Strukturierbarkeit Geringer Interviewer-Effekt	Unkontrollierte Erhebungs- situation	Erforschung individueller Verhaltensweisen und Charakteristika Erforschung von Fakten und „objektiven“ Daten	Zeitintensive Konzeption, Durchführung und Auswertung	Erheben von Kostendaten Gesundheitssystem- forschung

Quelle und Darstellung: GÖG/BIQG

5.2 Einsatz von Routinedaten im Rahmen von Health Technology Assessments

5.2.1 Einleitung: Routinedaten und HTA

Bei der Durchführung eines Health Technology Assessments (HTA) sind neben Daten aus klinischen Studien zur Wirksamkeit von medizinischen Interventionen oft Routinedaten (z. B. administrative Daten) und Daten aus anderen Quellen als wissenschaftlichen Studien notwendig. Zum einen ist der Einsatz solcher Daten erforderlich, weil sich Informationen aus klinischen Studien nur eingeschränkt zur Einschätzung des Nutzens einer Technologie unter Alltagsbedingungen eignen. Dies betrifft insbesondere die Bewertung von Risiken und Nebenwirkungen einer Intervention, die mit dem Design einer klinischen Studie oft nur ungenügend erfasst werden können, weil sie zum Beispiel selten oder erst nach längerem Zeitraum auftreten. Aber auch für andere Inhalte eines HTAs (z. B. für ökonomische Evaluationen) sind Daten abseits von klinischen Studien nötig (z. B. administrative Daten zum Mengenverbrauch einer Technologie). Schlussendlich befasst sich HTA zunehmend mit der Evaluation komplexer Leistungen. Beispiele dafür sind die Evaluierung von Disease-Management-Programmen, von Behandlungsabläufen oder von Vorsorgeprogrammen, die erstens nur eingeschränkt mit klassischen randomisiert-kontrollierten Studiendesigns zu evaluieren sind und zweitens aufgrund ihrer Kontextabhängigkeit nur mit kontextspezifischen Versorgungsdaten sinnvoll evaluiert werden können.

Da das Ziel von HTA die Unterstützung von Entscheidungen in Politik und Praxis ist, muss der Inhalt des Assessments auch die für den jeweiligen Entscheidungskontext relevanten Informationen beinhalten. Für manche Inhalte eines HTAs ist der Einsatz von Daten mit konkretem Bezug zum österreichischen Versorgungskontext daher sogar eine notwendige Voraussetzung, um brauchbares Wissen zu produzieren. Beispielsweise wird eine Abschätzung der ökonomischen Folgen einer Technologie nur dann für österreichische Entscheidungsträger relevant sein, wenn die Berechnung mit österreichischen Kostendaten erfolgt.

Die Identifikation und Aufbereitung solcher Daten, die – im Gegensatz zu publizierten Daten aus klinischen Studien – auch den direkten Zugang zu Daten inkludiert, gehören bisher nicht zum Standardinstrumentarium von HTA (Perleth, Antes et al. 2005) und nehmen demnach in den meisten HTA-Methodenhandbüchern nur einen geringen Stellenwert ein. Am ehesten finden sich Empfehlungen im Zusammenhang mit Wirtschaftlichkeitsanalysen (z. B. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen 2009).

Die beschriebenen Datenquellen und der Umgang mit ihnen werden in der Literatur hauptsächlich unter dem Thema Versorgungsforschung oder Qualitätssicherung behandelt (z. B. Glaeske, Augustin et al. 2009, Müller, Augustin et al. 2010, Perleth und Busse 2004, Van Emmerich und Metzinger 2010, Wissenschaftliches Institut der AOK 2011). An der Auseinandersetzung zum Umgang mit (Versorgungs-)Daten abseits von klinischen Studien im Rahmen von HTA wird daher sichtbar, dass HTA eng an die Versorgungsforschung anknüpft oder sogar als Teil der Versorgungsforschung betrachtet werden kann (Perleth und Busse 2004), insbesondere weil HTA zunehmend nicht nur zur Unterstützung für evidenzbasierte Entscheidungen zu einzelnen Technologien eingesetzt wird, sondern für die evidenzbasierte Planung und Steuerung des Gesundheitswesens auf allen Ebenen. Mit dem Ziel, eine explizit auf wissenschaftliche Evidenz gestützte Form der Entscheidungsfindung zu etablieren (Perleth und Busse 2004), sind unterschiedliche Datenquellen notwendig, die über das klassische Datenrepertoire von HTA – die klinischen Studien – hinausgehen.

Dieses Kapitel widmet sich daher dem Einsatz von österreichischen Routinedaten (z. B. administrativen Daten, Daten aus amtlichen Statistiken) im Rahmen von HTA (für eine Begriffsdefinition siehe Kapitel 3 und Glossar im Anhang unter 13). Es soll neben methodischen Aspekten zum Umgang mit solchen Daten vor allem einen Überblick über vorhandene Daten in Österreich geben und Informationen liefern zu deren Zugang, zu rechtlichen Rahmenbedingungen der Nutzung (z. B. Datenschutz) und zu den Eigenschaften der Daten (inkl. Rechtsgrundlage ihrer Entstehung). Wenngleich für manche Fragestellungen (z. B. zur Arzneimittelsicherheit) internationale Routinedaten genutzt werden können, beschränkt sich die Beschreibung der Daten auf österreichische Datenquellen oder auf solche internationale Datenquellen, die österreichische Daten enthalten (z. B. WHO Health Survey).

5.2.2 Methode

In die Beschreibung der Datenbestände (siehe Kapitel 5.2.6 und 15.5 im Anhang) wurden in einem ersten Schritt jene Datenquellen aufgenommen, die bereits in HTA-Berichten eines HTA-Instituts (des Ludwig Boltzmann Instituts für HTA) verwendet wurden. Mitarbeiter/innen der Koordinierungsstelle der nationalen HTA-Strategie an der Gesundheit Österreich GmbH / Geschäftsbereich BIQG wurden repräsentativ für weitere HTA-Expertinnen und -Experten um Ergänzung der Datenbestände gebeten (insbesondere um solche, die an der GÖG verwaltet werden).

Zur Beschreibung der Dateneigenschaften wurden fallweise „Dateneigner“ (Institutionen, die die Daten erheben, speichern und nützen sowie die rechtliche Verfügungsgewalt darüber besitzen) (Hoffmann und Glaeske 2011) gebeten, die vorhandene Information zu ergänzen und gegebenenfalls zu korrigieren. Die methodischen Standards

zur statistischen Analyse von Routinedaten (Kapitel 5.2) beruhen auf internationaler Methodenliteratur und methodischen Leitlinien zum Umgang mit Routinedaten (siehe Kapitel 5.2.4 und 5.2.6 für Details und Referenzen).

Das Kapitel erhebt keinen Anspruch auf Vollständigkeit der vorhandenen Datenquellen und der verwendeten Literatur zu den methodischen Standards. Allfällige weitere Datenquellen und methodische Anleitungen können im Prozess der Weiterentwicklung des Handbuchs ergänzt werden.

5.2.3 Definitionen und Abgrenzung des Themas

Die Literatur behandelt die für uns in diesem Kapitel relevanten Daten durchgängig unter dem Thema „Sekundärdatenanalyse“ (z. B. Boslaugh 2007, Swart und Ihle 2005). Tatsächlich ist das gemeinsame Merkmal der in Kapitel 6 und im Anhang (Kapitel 15) beschriebenen Datenquellen, dass es sich um Sekundärdaten handelt. Das heißt, im Gegensatz zu Primärdaten, die für den spezifischen Zweck der bearbeiteten Fragestellung erhoben werden, sind Sekundärdaten dadurch definiert, dass sie von anderen als den HTA-Erstellern / Erstellerinnen für einen ursprünglich anderen Zweck gesammelt wurden (Boslaugh 2007). Das Dokument „Gute Praxis Sekundärdatenanalyse“ (Arbeitsgruppe Erhebung und Nutzung von Sekundärdaten der DGSMF und Arbeitsgruppe Epidemiologische Methoden der DGEpi und der GMDS und der DGSMF 2008) versteht unter einer Sekundärdatenanalyse die „Analyse von Daten im Rahmen wissenschaftlicher Untersuchungen ohne direkten Bezug zum primären Erhebungsanlass“. Unter diese Definitionen fallen jedoch alle klassischen Daten, die im Rahmen von HTA verwendet werden, also etwa die Publikationen klinischer Studien oder die Daten aus klinischen Studien, die primär zum Zweck der Zulassung einer Technologie erhoben wurden. Diese Daten aus Zulassungsstudien oder Publikationen darüber sind jedoch nicht Gegenstand dieses Kapitels.

Hoffmann et al. (2008, s. 1118) definieren Sekundärdaten als jene Daten, die für einen anderen als einen *wissenschaftlichen* Zweck gesammelt wurden, in Abgrenzung zu Daten aus Feldstudien, die „gezielt für eine Forschungsfragestellung erhoben werden“. Diese Definition scheint im HTA-Kontext sinnvoller, weil sie die oben genannten klinischen Zulassungsstudien ausschließt. Die Abgrenzung zwischen wissenschaftlichem und „nicht wissenschaftlichem“ Zweck ist aber nicht immer eindeutig zu treffen. Beispielsweise können klinische Tumorregister eine relevante Sekundärdatenquelle für HTA darstellen, die daher in diesem Kapitel zu behandeln ist. Sie entsprechen aber vom Konzept her oftmals einer klinischen Beobachtungsstudie und werden neben anderen Gründen auch für wissenschaftliche Zwecke eingerichtet. Somit ist auch diese Definition unzulänglich.

Geeignet im Zusammenhang mit HTA erscheint der Begriff „Routinedaten“, den Raftery et al. (2005) folgendermaßen definieren: „Routine data are data that are collected regularly and continuously or periodically, that use standard definitions for all the population groups covered, that are characterised by some (usually considerable) degree of obligation to collect the data completely and regularly and that are collected at a national or regional level, including more than one centre, depending on the representativeness of the sample“ (Raftery, Roderick et al. 2005, s. 15). Wie die Autorinnen/Autoren ausführen, fallen unter diese Definition von Routinedaten nicht nur administrative Daten, sondern alle Krankheits- und „health technology“-Register, Systeme zur Registrierung von Nebenwirkungen und regelmäßig durchgeführte gesundheitsbezogene Befragungen.

Wesentlich enger definieren Hoffmann und Glaeske (2011) Routinedaten als prozessorientierte, umfangreiche Informationssammlungen, die im Rahmen der Verwaltung, Leistungserbringung bzw. Kostenerstattung (z. B. bei der gesetzlichen Kranken-, Renten- oder Unfallversicherung) anfallen und elektronisch erfasst sind. Bezogen auf die vorherige Definition von Raftery et al. (2005), beschreibt diese Definition eine bestimmte Gruppe von Routinedaten: die administrativen Daten.

In Anlehnung an diese Definitionen werden in diesem Kapitel unter **Routinedaten** alle *Sekundärdaten* verstanden, die *regelmäßig, verpflichtend* und zu primär *nicht wissenschaftlichen* Zwecken erhoben werden. **Administrative Daten** werden als eine *Gruppe von Routinedaten* verstanden, die im Rahmen der *Verwaltung, Leistungserbringung bzw. Kostenerstattung* im Gesundheitswesen anfallen.

Die in diesem Kapitel betrachteten Datenbestände umfassen demnach:

- » administrative Daten diverser Kostenträger oder Leistungserbringer,
- » Register (epidemiologische Register, Qualitätsregister, krankheits-, interventionsbezogene- oder risikobezogene Register),
- » regelmäßig durchgeführte Gesundheits-Surveys,
- » amtliche (nationale) Statistiken,
- » Daten, die primär zu Planungszwecken erhoben/verwaltet werden (z. B. Erreichbarkeit und natürliche Einzugsgebiete),
- » Datenbestände zu Versorgungsstrukturen (z. B. Rehakompass).

Daten aus klinischen Studien oder Publikationen in Fachzeitschriften, die auf Daten der oben genannten Quellen basieren, werden nicht berücksichtigt. Ebenso werden Gesundheitsberichte ausgeschlossen, da sie die Kriterien der regelmäßigen und verpflichtenden Datenerhebung nicht erfüllen und vielfach keine separaten Datenbestände darstellen, sondern diverse (bereits angeführte) Sekundärdatenquellen nutzen.

In jüngerer Vergangenheit nehmen Initiativen zu, mehrere der angeführten Datenbestände miteinander zu verknüpfen, um damit aussagekräftigere Daten zu erhalten

(z. B. Verlinkung von Daten über Systemgrenzen hinweg). Die auf diese Weise eingerichteten „Metadatenbanken diverser Sekundärdaten“ (z. B. GAP-DRG-Datenbank des Projekts zur Verknüpfung von Daten aus dem stationären und niedergelassenen Bereich; siehe Details im Anhang 9.2.7) könnten zwar als eigenständige Datenquelle betrachtet werden, sie werden in der vorliegenden Version aber lediglich beispielhaft angeführt. Bei Nutzung solcher Quellen im Rahmen von HTA erscheint eine Validitätsprüfung unter Mitbetrachtung und Mitberücksichtigung der Ausgangsdatensätze sinnvoll zu sein.

Als Hilfestellung, für welchen Schritt im HTA-Prozess welche Routinedaten geeignet sind, haben Raftery et al. (2005) eine Klassifikation von Routinedaten aus einer HTA-Perspektive vorgenommen, die für das vorliegende Handbuch in leicht adaptierter und erweiterter Form übernommen wird. Danach können die Datenquellen in vier Gruppen unterteilt werden:

- » Gruppe I: Daten enthalten Informationen zu der zu evaluierenden Technologie UND zu den relevanten Gesundheitszuständen/Patientendaten
 - a) klinische und Qualitätsregister (z. B. Chirurgie-Qualitätsregister)
 - b) klinisch-administrative Datensets (z. B. Krankengeschichten)
 - c) bevölkerungsbasierte Datensets (z. B. Zervixkarzinom-Screeningregister in Großbritannien)
 - d) sonstige
- » Gruppe II: Daten enthalten Informationen zu der zu evaluierenden Technologie, aber nicht zu den relevanten Gesundheitszuständen (z. B. Medizinprodukte-register, Arzneimittelverordnungen aus administrativen Daten der Sozialversicherungsträger)
- » Gruppe III: Daten enthalten Informationen zu den relevanten Gesundheitszuständen, aber nicht zu der zu evaluierenden Technologie
 - a) „adverse events reporting“ und vertrauliche Untersuchungen (z. B. „confidential enquiry into stillbirths and deaths in infancy“)
 - b) Krankheitsregister (z. B. Tumorregister)
 - c) Gesundheits-Surveys (z. B. HBSC-Survey)
 - d) sonstige
- » Gruppe IV: Daten enthalten Informationen für Kostenberechnungen
 - a) Ressourcenverbrauch UND Preise (z. B. Leistungsdokumentation in Krankenanstalten)
 - b) nur Ressourcenverbrauch (z. B. Blutverbrauchserhebung)
 - c) nur Preise (z. B. Pharmapreisinformationssystem)

Während die Gruppe-I-Daten beispielsweise für die Bewertung der Effektivität oder der Diffusion einer Technologie infrage kommen, können Gruppe-II- und Gruppe-III-Daten wesentlich eingeschränkter eingesetzt werden (etwa für die Darstellung der

Verbreitung einer Technologie bei Gruppe II oder für die Darstellung der Sicherheit bei Gruppe III). Gruppe-IV-Daten können als Grundlage für Kostenberechnungen herangezogen werden. Manche Datenbestände fallen in mehrere Gruppen. Beispielsweise entsprechen aus den administrativen Daten der Sozialversicherungsträger die Daten zu den Arzneimittelverordnungen Gruppe-II-Daten, die Unfallstatistikdaten fallen hingegen in Gruppe-III-Daten.

Routinedaten können in unterschiedlichem Aggregationslevel zugänglich sein. Sie können als individuelle (fallbezogene/versichertenbezogene/personenbezogene) Daten (z. B. Datenbanken der Sozialversicherungsträger) verfügbar sein oder in aggregierter Form (z. B. öffentlich zugängliche Standardauswertungen des Geburtenregisters). Während Erstere mehr Spielraum für Auswertungen beinhalten, ist deren Nutzung gleichzeitig mit erheblich größerem Aufwand verbunden. Mit unterschiedlichem Aggregationsniveau sind jeweils auch unterschiedliche Auswertungsverfahren verbunden (siehe Kapitel 5.2.3 für konkretes Beispiel zum Umgang mit „patient level data“ für Kostenberechnungen).

Zu unterscheiden ist ferner die direkte Nutzung der Daten durch Zugriff auf die anonymisierten oder pseudonymisierten Originaldaten und die indirekte Nutzung von Daten über Aggregatstatistiken oder Sonderauswertungen durch den Dateneigner.

5.2.4 Einsatz von Routinedaten im HTA-Prozess

5.2.4.1 Allgemeine Vor- und Nachteile von Routinedaten

Die Vorteile vieler der beschriebenen Routinedaten sind, dass sie unter Alltagsbedingungen entstehen und damit die Versorgungssituation abbilden. Dies trifft insbesondere auf administrative Daten zu. Bei der Nutzung von Routinedaten entsteht kein zusätzlicher Erhebungsaufwand, daher stehen die Daten zeitnah und im Vergleich zu umfangreichen Primärerhebungen vergleichsweise günstig zur Verfügung. Einige Verzerrungen, die bei Primärdaten bestehen können (z. B. durch Non-Responder oder selektives Erinnern) lassen sich bei vielen Routinedaten ausschließen. Außerdem ermöglichen sie Untersuchungen zu Personengruppen, die für eine Primärdatenerhebung nur schwer zugänglich oder von Primärstudien oft ausgeschlossen sind (z. B. Kinder, Bewohner von Pflegeheimen). Die Analysen sind mit vergleichsweise geringem Zeitaufwand und über große Populationen möglich. Meistens kann die „Nennerbevölkerung“ explizit definiert werden (z. B. alle Versicherten eines Sozialversicherungsträgers). Die Daten liegen häufig über längere Beobachtungszeiträume vor, sind aktuell, ermöglichen die Erfassung seltener Ereignisse oder lassen unter Umständen die

Darstellung von individuellen Krankheits- bzw. Leistungsverläufen zu (Andersohn und Garbe 2008, Boslaugh 2007, Hoffmann und Glaeske 2011). Diese Vorteile machen Routinedaten – vor allem wenn sie auf Fall- oder Versichertenebene vorliegen – für HTA höchst relevant, da sie eine für Entscheidungsträger relevante Wissensproduktion unterstützen (zeitnah, für Entscheidungskontext relevant).

Als Nachteile sind zu nennen, dass die Daten für andere Zwecke erhoben wurden und daher oft nicht alle für den Forschungszweck interessanten Variablen enthalten, dass die Variablen anders definiert oder kategorisiert wurden, als es für den eigenen Forschungszweck nötig wäre, oder dass große Unklarheiten darüber bestehen, wie die Variablen kategorisiert oder definiert wurden (fehlende Beschreibung). Dies betrifft oftmals klinische Informationen. Gewünschte Zielparameter (z. B. bestimmte Diagnosen) müssen häufig auf Basis der vorhandenen Information abgeleitet werden (z. B. auf Basis von Abrechnungen bestimmter Leistungen). Ein weiterer Nachteil ist, dass Details zur Datenerhebung und -qualität und somit die Datenvalidität und Präzision der Dokumentation häufig unbekannt sind. Eine Einschränkung der Datenvalidität ist zumeist bezüglich der Klassifikation von Diagnosen oder Prozeduren gegeben (Boslaugh 2007, Hoffmann, Bobrowski et al. 2008, Hoffmann und Glaeske 2011, Swart und Ihle 2008). Schließlich ist als Nachteil anzuführen, dass zwar kein zusätzlicher Zeitaufwand durch Datenerhebung besteht, dafür aber der Prozess der Datenaufbereitung vieler Routinedaten (insbesondere administrativer Daten) erstens sehr ressourcen- und zeitaufwendig sein kann und zweitens häufig in der Hand der Dateneigner liegt, sodass zeitliche Unwägbarkeiten und einer hoher Kommunikationsaufwand eingeplant werden müssen, bis ein entsprechender Auswertungsdatensatz vorliegt. Die Daten werden zwar häufig bezüglich ihres primären Erhebungszwecks einer Datenprüfung unterzogen, nicht aber bezüglich der Sekundärnutzung, sodass für eine solche Nutzung im Rahmen von HTA der Ressourcenaufwand für eine Validitätsprüfung (z. B. bezüglich Diagnosen, Vollständigkeit) berücksichtigt werden muss. Die Darstellung der Limitationen der verwendeten Daten für die jeweilige Fragestellung ist ein unverzichtbares Qualitätsmerkmal eines HTAs, die alleinige Angabe der Datenquelle reicht oft nicht aus, um eine korrekte Interpretation zu gewährleisten. Ein weiteres oft unverzichtbares Qualitätsmerkmal ist eine dokumentierte Rücksprache und Überprüfung der postulierten Schlussfolgerungen mit den entsprechenden Datenexperten/Datenexpertinnen.

5.2.4.2 Ethische und datenschutzrechtliche Aspekte

Die Verwendung von Routinedaten für wissenschaftliche Zwecke unterliegt prinzipiell den datenschutzrechtlichen Bestimmungen des Datenschutzgesetzes. Hierbei ist zu beachten, dass es sich bei vielen der genannten Daten um personenbezogene Daten handelt, das heißt um Angaben über Betroffene, deren Identität über die Daten bestimmt oder bestimmbar ist (z. B. Namen, Geburtsdatum, Anschrift). Laut Daten-

schutzgesetz ist die Verwendung personenbezogener Daten grundsätzlich verboten, außer es bestehen spezifische gesetzliche Regelungen, Verträge bzw. ein vertragsähnliches Vertrauensverhältnis oder es liegt eine ausdrückliche Einwilligung vor. Zudem handelt es sich bei zahlreichen Daten (z. B. Gesundheitsdaten) um sogenannte sensible Daten, die als besonders schutzwürdig definiert sind (Bundeskanzleramt 2011).

Anonymisierte Daten (personenbezogene Daten, die derart verändert worden sind, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmaren natürlichen Person zugeordnet werden können) fallen hingegen nicht unter personenbezogene Daten im Sinne des Datenschutzgesetzes (Arbeitsgruppe Erhebung und Nutzung von Sekundärdaten der DGSM und Arbeitsgruppe Epidemiologische Methoden der DGEpi und der GMDS und der DGSM 2008). Das wird zum Beispiel erreicht, indem die für die Anonymisierung verwendeten Schlüssel vernichtet werden. Zu beachten ist jedoch, dass der Anonymisierungsvorgang selbst noch den Bestimmungen des Datenschutzgesetzes unterliegt und nur von Personen vollzogen werden darf, die die Daten rechtmäßig nutzen (z. B. Krankenversicherungsträger) oder von den rechtmäßigen Nutzern beauftragt wurden (Ihle 2008).

Pseudonymisierte Daten (Daten mit indirektem Personenbezug, bei denen die Identität des Betroffenen mit rechtlich zulässigen Mitteln nicht bestimmt werden kann) gelten datenschutzrechtlich hingegen weiterhin als personenbezogen (Ihle 2008). Eine Pseudonymisierung erfolgt durch das Ersetzen des Namens und anderer Identifikationsmerkmale durch ein Kennzeichen. Eine Pseudonymisierung ist notwendig, wenn personenbezogene Daten über ein bekanntes Pseudonym bereits pseudonymisierten Daten zugeordnet werden sollen (z. B. bei der Verlinkung unterschiedlicher Datenquellen).

Die Abgrenzung zwischen anonymisierten und pseudonymisierten Daten ist fließend. Werden z. B. Daten durch eine Vertrauensstelle der Krankenversicherungsträger pseudonymisiert, und den Forschern/Forscherinnen bleibt der notwendige Schlüssel für die Re-Identifikation der Personen unbekannt, sind die Daten für die Forscher/innen anonymisiert. Aus datenschutzrechtlicher Sicht ist keine absolute, sondern nur eine faktische Anonymisierung gefordert.

Das österreichische Datenschutzgesetz erlaubt die Verwendung von pseudonymisierten (also indirekt personenbezogenen) Daten für wissenschaftliche Zwecke (Bundeskanzleramt 2011). Unter bestimmten Voraussetzungen sind auch direkt personenbezogene Daten für wissenschaftliche Verwertung zugänglich, es ist jedoch aus datenschutzrechtlicher Sicht von der Möglichkeit der Anonymisierung und Pseudonymisierung personenbezogener Daten jedenfalls Gebrauch zu machen, wenn das Auswertungsziel damit ebenfalls erreicht werden kann. Dies ist in den allermeisten Situationen der Fall (Ihle 2008). Wenn die datenschutzrechtlichen Vorschriften zur Pseudonymisie-

rung personenbezogener Daten erfüllt sind, kann die Konsultation einer Ethikkommission entfallen (Arbeitsgruppe Erhebung und Nutzung von Sekundärdaten der DGSMP und Arbeitsgruppe Epidemiologische Methoden der DGEpi und der GMDS und der DGSMP 2008). Empfohlen wird jedoch die Erstellung eines Datenschutzkonzeptes, in dem die datenschutzrechtlichen Aspekte schriftlich fixiert werden.

Wie eingangs angeführt, besteht eine zusätzliche Möglichkeit der unmittelbaren wissenschaftlichen Nutzung personenbezogener Daten (von Patienten/Patientinnen) darin, die Einwilligung der Versicherten oder Patienten/Patientinnen einzuholen. Hierbei sind aber oft besondere Schwierigkeiten (z. B. nicht einwilligungsfähige Patientengruppen) vorhanden und entsprechende gesetzliche Vorgaben (z. B. Ethikkommissionen) zu beachten.

5.2.4.3 Gute Praxis Routinedatenanalyse und Qualitätssicherung

Das Ausmaß notwendiger qualitätssichernder Maßnahmen hängt davon ab, zu welcher Fragestellung und in welchem Ausmaß Routinedaten eingesetzt werden. In jedem Fall sollen die vorgesehenen Routinedaten im Berichtsplan (siehe Kapitel 4.4. Methodenhandbuch) spezifiziert werden. Da die einer Routinedatenanalyse zugrunde liegenden Datenquellen jeweils unterschiedliche ursprüngliche Verwendungszecke haben, von diversen Dateneignern stammen und auf verschiedenen rechtlichen Grundlagen beruhen, ist Transparenz bezüglich der Datenentstehung und -übermittlung ein wesentliche Qualitätskriterium.

Für umfangreiche Analysen (z. B. wenn Register zur Bewertung der Sicherheit einer Technologie vorgesehen sind) wird empfohlen, wenn relevant die Studienpopulation zu definieren, die Vorgangsweise zur Vermeidung von Verzerrung (z. B. Confounding) zu beschreiben und die zeitlichen und personellen Ressourcen vorzusehen (Arbeitsgruppe Erhebung und Nutzung von Sekundärdaten der DGSMP und Arbeitsgruppe Epidemiologische Methoden der DGEpi und der GMDS und der DGSMP 2008). In der „Guten Praxis Sekundärdatenanalyse“ (Arbeitsgruppe Erhebung und Nutzung von Sekundärdaten der DGSMP und Arbeitsgruppe Epidemiologische Methoden der DGEpi und der GMDS und der DGSMP 2008) wird außerdem angeregt, ein Operationshandbuch zu erstellen, in dem organisatorische Maßnahmen zur Vorbereitung und Durchführung der Datenanalyse dokumentiert werden (Datenbereitstellung beim Dateneigner, Datenübermittlung, Aufbereitung).

Die „Gute Praxis Sekundärdatenanalyse“ empfiehlt zur Qualitätssicherung bei umfangreicher Nutzung von Routinedaten einen Prätest mit Probedaten, die transparente Dokumentation jeglicher Änderungen des Berichtsplans und der Datenaufbereitung in

Form eines Datenaufbereitungsprotokolls. In diesem Zusammenhang sollen sowohl der Weg der Daten von ihrer primären Erfassung bis zur Bereitstellung für die Sekundärnutzer/innen (Verwendungszweck der Datensammlung, Erfassungsregeln, Überprüfung der Konstanz von Erfassungsregeln, gesetzliche Rahmenbedingungen) dokumentiert werden als auch Umfang und Struktur der bereitgestellten Daten inklusive Hinweise auf fehlende Datensätze. Außerdem sollen für Qualitätssicherungszwecke ein Ausgangs- und ein Auswertungsdatensatz (der nach Plausibilitätsprüfung und Datentransformation überarbeitete Datensatz) in unveränderter Form verfügbar sein.

Für einzelne Routinedatentypen existieren Checklisten zur Überprüfung der Datenqualität allgemein bzw. der Datenqualität in wissenschaftlichen Studien, die auf Basis dieser Daten erstellt werden. Eine Prüfliste für Register befindet sich im Memorandum „Register für die Versorgungsforschung“ (Müller, Augustin et al. 2010), eine Checkliste für „Retrospective Database Studies“ wurde von der ISPOR Task Force on Retrospective Databases publiziert (Motheral, Brooks et al. 2003).

5.2.4.4 Routinedaten im HTA-Prozess

Der Einsatz von Routinedaten hängt zunächst von der zu bewertenden Technologie und vom Umfang des Assessments ab. Generell wird für stark kontextabhängige Technologien (z. B. Gesundheitsförderungsprogramme) die Verwendung von Routinedaten eher gefordert sein als für die Evaluierung der Wirksamkeit eines Arzneimittels oder eines Medizinprodukts. Ebenso wird ein Rapid Assessment, das ausschließlich die Wirksamkeit und Sicherheit einer Technologie bewertet, im Regelfall weniger Routinedaten beinhalten als ein umfangreicher HTA-Bericht, der auch Wirtschaftlichkeit und Tragweite einer Technologie (psychosoziale, organisatorische Effekte etc.) beurteilt.

Wie Raftery et al. (2008) anmerken, ist der Einsatz von Routinedaten im Rahmen eines Assessments außerdem von den Eigenschaften der Daten abhängig. Hierbei kommt wieder die Klassifizierung aus Kapitel 4.7 und deren Unterscheidung zum Einsatz, ob Informationen zu der zu evaluierenden Technologie, zu den relevanten Gesundheitszuständen oder zu beiden Faktoren bzw. ob Informationen für Kostenberechnungen enthalten sind.

Die im Anhang publizierte Auflistung der Routinedatenbestände zeigt, dass die Datenbestände am häufigsten (ca. 40 %) in die Kategorie IV fallen (enthalten Informationen zur Kostenberechnung). Knapp 30 Prozent der Datenbestände fallen in die Kategorie I (Daten enthalten Informationen zu der zu evaluierenden Technologie und zu den relevanten Gesundheitszuständen), und in jeweils knapp einem Viertel sind die Daten der Kategorie II zuzuordnen (Daten enthalten Informationen zu der zu evaluierenden Technologie, aber nicht zu den relevanten Gesundheitszuständen) und der

Kategorie III (Daten enthalten Informationen zu den Gesundheitszuständen, aber nicht zur zu evaluierenden Technologie)¹.

Aus dieser Information lässt sich schließen (wie auch Tabelle 9.1 im Anhang zeigt), dass die Datenbestände in der derzeit vorliegenden Form im Rahmen eines HTA-Prozesses am häufigsten für die Themenpriorisierung und bei der Erstellung des Assessments für die Beschreibung des Hintergrunds (Krankheitslast etc.) sowie für ökonomische Bewertungen einsetzbar und nötig sind, während sie für die Bewertung von Wirksamkeit und Sicherheit seltener geeignet sind, da nur „Kategorie-I-Daten“ die Voraussetzung dafür erfüllen, diese aber in Österreich häufig nur in aggregierter Form oder gar nicht zugänglich sind. Wie internationale Beispiele zeigen (z. B. Qualitätssicherung mit Routinedaten in Deutschland), besteht ein Ausweitungspotenzial bei administrativen Daten, wenn etwa durch Datenverlinkung verstärkt fallübergreifende Analysen von Abrechnungsdaten aus Krankenhäusern mit weiteren administrativen Versichertendaten möglich werden (Wissenschaftliches Institut der AOK 2011).

In Anlehnung an Perleth et al. (2005) soll nachfolgend entlang des gesamten HTA-Prozesses sowie für die inhaltlichen Elemente eines Assessments (wie sie in den Kapiteln 3 bis 11 in diesem Methodenhandbuch beschrieben werden) systematisch dargestellt werden, wo Routinedaten zum Einsatz kommen können oder sollen. Tabelle 5.2 bildet den HTA-Prozess ab und führt Beispiele für relevante Routinedaten an.

1

Aufgrund von Mehrfachzuordnungen ergibt die Prozentsumme mehr als 100 Prozent.

Tabelle 5.2:

Möglichkeiten für den Einsatz von Routinedaten im HTA-Prozess

Schritte im HTA-Prozess	Beispiele für relevante Routinedaten
A) Themenfindung und Priorisierung » Größe der betroffenen Personengruppe » Krankheitslast » Kosten » Öffentliches Interesse	» Epidemiologische Daten amtlicher Statistiken, epidemiologische Register » Administrative Daten der Krankenanstalten- und SV-Träger » Pharma-Preis-Informationssystem, administrative Daten der SV-Träger (z. B. Honorarordnungsdatenbank) » Gesundheits-Surveys
B) Scoping/Forschungsfrage	» Siehe Themenfindung/Priorisierung
C) Durchführung eines Assessments » Hintergrund: Statusbestimmung zu Krankheit und Technologie » Analyse der Wirksamkeit/Sicherheit » Wirtschaftlichkeit / ökonomische Bewertung » Abschätzung der Tragweite: ökonomische, organisatorische, juristische, ethische und psychosoziale Folgen	» Administrative Daten zur Verschreibungshäufigkeit, div. Register » Qualitätsregister » Administrative Daten zu Tarifen und konsumierten Einheiten einer Technologie, Register » Surveys, administrative Daten (zu Kosten, Anzahl und geographischer Lage der Leistungsanbieter)
D) Dissemination der Ergebnisse	» Administrative Daten zu Leistungsanbietern, Register (zur Identifikation relevanter Zielgruppen)
E) Umsetzung	-
F) Evaluierung des Impacts	» Administrative Daten (zur Evaluierung des Mengenverbrauchs oder der Ausgabenentwicklung nach HTA-Bericht)

Quelle: LBI; GÖG/BIQG-eigene Darstellung

5.2.4.4.1 Routinedaten für Themenpriorisierung und Scoping

Der HTA-Prozess beginnt mit der Identifikation einer zu evaluierenden Technologie, bei der ein Priorisierungsverfahren zum Einsatz kommt. Von den üblichen, bei einem Priorisierungsverfahren herangezogenen Kriterien liefern Routinedaten insbesondere Information zu

- » Größe der betroffenen Personengruppe (z. B. epidemiologische Daten aus amtlichen Statistiken),
- » Krankheitslast (z. B. administrative Daten der Krankenanstalten- und Sozialversicherungsträger),
- » Kosten (z. B. Pharma-Preis-Informationssystem, administrative Daten der Krankenanstalten- und Sozialversicherungsträger),
- » Fragen von öffentlichem Interesse (z. B. Surveys).

Ein Beispiel zum Einsatz von Routinedaten im Priorisierungsprozess ist in Box 5.1 dargestellt. Dieselben Informationen bzw. Datenquellen sind hilfreich auch beim Prozess des Scoping, der zur Entwicklung und Eingrenzung der Forschungsfrage dient.

Box 5.1:

Einsatz von Routinedaten für Priorisierung: Projektbeispiel „Hämocomplettan® und Fibrogammin® bei erworbenen Fibrinogenmangelzuständen“

Kostendaten aus der Krankenanstaltenkostenrechnung haben einen überdurchschnittlich hohen jährlichen Verbrauchszuwachs an Hämocomplettan® und Fibrogammin® gezeigt, sodass diese Produkte für eine Evidenzanalyse priorisiert wurden.

Quelle: (Warmuth und Mad 2009)

5.2.4.4.2 Routinedaten bei der Durchführung des Assessments

Bei der eigentlichen Durchführung des Assessments haben Raftery et al. (2008) zumindest für die HTA-Domänen „diffusion“, „effectiveness“, „equity“ und „costs“ Möglichkeiten identifiziert, mit Routinedaten zu arbeiten. Letztendlich können Routinedaten bei allen klassischen inhaltlichen Elementen eines Assessments (Hintergrund, Beschreibung der Technologie, Evaluierung von Wirksamkeit, Sicherheit und Wirtschaftlichkeit einer Gesundheitstechnologie sowie der ökonomischen, organisatorischen, juristischen, ethischen und psychosozialen Folgen der Einführung einer Technologie) notwendig oder hilfreich sein. Dies insbesondere deshalb, da der Technologiebegriff sehr breit definiert ist und sowohl Arzneimittel, Medizinprodukte, Prozeduren, Organisations- und Supportsysteme (z. B. Telematik) als auch komplexe Leistungen im Bereich Public Health (z. B. Vorsorgeprogramme) oder Behandlungsabläufe umfasst, die oft kontextbezogen sind und mit internationalen Daten nur eingeschränkt evaluiert werden können.

Hintergrund, Statusbestimmung von Technologie und Erkrankung

Bei der Beschreibung des Hintergrunds, der eine Statusbestimmung zur Technologie und zu der von ihr adressierten Erkrankung liefert, sind Routinedaten zur Darstellung der Verbreitung einer Technologie (z. B. zur Verschreibungshäufigkeit eines Arzneimittels) geeignet (Raftery, Roderick et al. 2005). Für ein Beispiel dazu siehe Box 5.2.

Box 5.2:**Einsatz von Routinedaten für die Statusbestimmung einer Technologie: Projektbeispiel „Halten Statine, was sie versprechen? Prognostizierte und tatsächliche Auswirkungen auf kardiovaskuläre Erkrankungen in Österreich“**

Routinedaten der Sozialversicherungsträger zur Anzahl der Verschreibung von Statinen sowie zu den unterschiedlichen Produkten und Packungsgrößen über einen Zeitraum von zehn Jahren wurden für die Beschreibung der Verbreitung dieses Arzneimittels im Zeitverlauf herangezogen. Die Daten wurden in anonymisierter und bereinigter Form vom Hauptverband der Sozialversicherungsträger zur Verfügung gestellt. Während von einer hohen Vollständigkeit der Daten ausgegangen werden kann, war die Aussage zur Verbreitung der Technologie eingeschränkt, da die verfügbaren Daten keine Information zu Alter und Geschlecht der behandelten Patienten/Patientinnen beinhalteten und einige zentrale Fragen (z. B. zur Compliance) aus den Daten nicht beantwortet werden konnten.

Quelle: (Zechmeister, Stollenwerk et al. 2008)

Amtliche Statistiken oder Register liefern Informationen zur Epidemiologie einer Erkrankung (z. B. Todesursachenstatistik, IVF-Register). Beim Einsatz von HTA für eine evidenzgestützte Bedarfsplanung nehmen die Beschreibung der Krankheitslast bzw. der Häufigkeit von Risikofaktoren – und demnach Routinedaten – eine zentrale Rolle ein. Ein Beispiel dazu ist in Box 5.3 dargestellt.

Box 5.3:**Einsatz von Routinedaten für die Beschreibung der Krankheitslast: Projektbeispiel „Eltern–Kind–Vorsorge neu. Teil I: Epidemiologie – Häufigkeiten von Risikofaktoren und Erkrankungen in Schwangerschaft und früher Kindheit“**

Zur Ermittlung des objektiven Bedarfs an Eltern–Kind–Vorsorgeleistungen wurden Routinedaten zur Darstellung der Evidenz zur Krankheitslast und zur Häufigkeit von Risikofaktoren aufgearbeitet. Genutzt wurden vor allem Register (z. B. Geburtenregister), Surveys (z. B. HBSC–Survey) und administrative Daten (z. B. Dokumentation der Krankenhausleistungen). In den meisten Fällen wurden die Sekundärdaten in aggregierter Form verwendet; in Einzelfällen (wie z. B. bei den Daten zur Leistungsdokumentation in Krankenanstalten) wurden fallbezogene Originaldaten herangezogen, die von den Dateneignern ausgewertet wurden.

Quelle: (Warmuth, Mad et al. 2011)

Wirksamkeit und Sicherheit

Die Evaluierung von Wirksamkeit und Sicherheit beruht in erster Linie auf klinischen Studien. In manchen Fällen können aber auch hier Routinedaten zum Einsatz kommen.

Einerseits, wenn es um die Evaluierung der Wirksamkeit unter Alltagsbedingungen geht – also um die Darstellung der Effektivität einer Technologie –, oder wenn die vergleichende Wirksamkeit (comparative effectiveness) dargestellt werden soll (Berger, Mamdani et al. 2009, Cox, Martin et al. 2009, Raftery, Roderick et al. 2005); andererseits, wenn bei der Analyse der Sicherheit einer Technologie seltene oder nach längerem Zeitraum auftretende Nebenwirkungen beschrieben werden sollen. Eine typische Quelle dafür sind Register. Für bestimmte Fragestellungen können aber auch administrative Daten relevant sein (siehe Box 5.4). Die derzeitige Qualität der Daten lässt aber häufig keine schlüssigen Aussagen oder gar die Herstellung von Kausalzusammenhängen zu.

Box 5.4:

**Einsatz von Routinedaten für die Evaluierung der Effektivität einer Technologie:
Projektbeispiel „Ambulante kardiologische Rehabilitation“**

Für eine retrospektive Kohortenstudie zum Vergleich der Effektivität zwischen Patienten/Patientinnen mit und ohne ambulante Phase-III-Rehabilitation wurden unterschiedliche administrative Daten der Sozialversicherungsträger zusammengeführt und die daraus konzipierte Datenbank für die Analyse mehrerer Outcomeparameter zur Verfügung gestellt (Summe aller medizinischen Events, einzelne Ereignisse wie Herzinfarkt, Schlaganfall etc., ökonomische Parameter wie Inanspruchnahme von Krankenversicherungsleistungen, Arbeitsunfähigkeit). Die verwendeten Datenbestände inkludierten Personenstammdaten der Patienten/Patientinnen, Befunddaten der Rehabilitationsdokumentation und Abrechnungsdaten (Verordnungsdaten, Krankenhausdaten, Umsatzdaten pro Behandlung, Krankenstandsdaten). Das Beispiel zeigt klassische Nachteile bei der Verwendung von Routinedaten für Effektivitätsanalysen auf, wie z. B. fehlende Werte bei den Befunddaten oder Confounding durch ungleiche Krankheitschwere in den Gruppen, die die Aussagekraft der Ergebnisse einschränken (siehe dazu auch 5.2.2 und Box 5.7).

Quelle: (Gyimesi, Piso et al. 2010)

Wirtschaftlichkeit

Wirtschaftlichkeitsprüfungen im Rahmen von ökonomischen Evaluationen verlangen jedenfalls bei der Kostenberechnung Routinedaten zur Ermittlung von Mengen und/oder Preisen der relevanten Leistungen (siehe Box 5.5). Darüber hinaus können sie bei ökonomischen Studien, die auf entscheidungsanalytischen Modellen basieren, für das Krankheitsmodell (z. B. epidemiologische Daten, Daten zu Todesursachen und Lebenserwartung aus amtlichen Statistiken) oder für die anschließenden Sensitivitätsanalysen notwendig sein (z. B. Impfstatistik für realistische Ober- und Untergrenzen von Teilnahmeraten bei einer Impfung).

Box 5.5:**Einsatz von Routinedaten für die Kostenberechnung:****Projektbeispiel „Ökonomische Evaluation der HPV-Impfung“**

Für die Kostenberechnung der Behandlung von Zervixkarzinomvorstufen und von invasivem Zervixkarzinom in unterschiedlichen Stadien wurden administrative Daten des Bundesministeriums für Gesundheit (Daten der leistungsorientierten Krankenanstaltenfinanzierung, insbesondere Punkte relevanter Diagnosegruppen und Einzelleistungen) und der Sozialversicherungsträger (Tarife zu Sozialversicherungsleistungen aus der Metahonorarordnung) verwendet. Die Daten dienen primär zur Information über Preise der verschiedenen Leistungen (z. B. Konisation, Bestrahlung, PAP-Abstrich). Dem verhältnismäßig raschen Informationsgewinn über administrative Ausgaben/Tarife steht die mangelnde „Kostenwahrheit“ dieser Daten als Nachteil gegenüber (siehe auch Kapitel 5.2.3).

Für die Berechnung der indirekten Kosten wurden ebenfalls Routinedaten aus diversen Quellen herangezogen (Pensionsversicherungsstatistik zur Berechnung des durchschnittlichen Pensionsantrittsalters, Einkommensstatistik zur Ermittlung des Durchschnittseinkommens, Krankenstandsstatistik zur Ermittlung der durchschnittlichen Krankenstandsdauer bei der Behandlung eines Zervixkarzinoms).

Quelle: (Zechmeister, Freiesleben de Blasio et al. 2007)

Abschätzung der Tragweite:**Psychosoziale, organisatorische, ethische und rechtliche Aspekte**

Bei der Abschätzung der Tragweite einer Technologie im Sinne von psychosozialen, organisatorischen, ethischen oder rechtlichen Aspekten sind Routinedaten ebenfalls hilfreich. Beispielsweise können Daten über die ärztlichen Vertragspartner der Sozialversicherung Auskunft geben über vorhandene oder zusätzliche nötige Personalkapazitäten (z. B. bei der Abschätzung organisatorischer Folgen nach Einführung eines Impf- oder Screeningprogramms).

5.2.4.4.3 Routinedaten bei der Dissemination der Ergebnisse und bei der Impact-Messung

Routinedaten können für die Verbreitung der Ergebnisse eines HTAs verwendet werden (z. B. Ärztelisten, Listen von Gesundheitsberufen). Schließlich kommen sie auch bei der Evaluierung des Impacts eines Assessments zum Einsatz, wenn etwa der Verlauf der Verschreibungshäufigkeit eines Medikaments vor und nach dem HTA-Bericht auf Basis von Arzneimittelverordnungen der Sozialversicherungsträger überprüft wird. Ein Beispiel dazu wird in Box 5.6 dargestellt.

Box 5.6.**Einsatz von Routinedaten für Impact-Evaluierung: Projektbeispiel
„Auswirkung der HTA-Forschung auf das Gesundheitswesen in Österreich“**

Für die Analyse ökonomischer Auswirkungen diverser HTA-Berichte wurden unterschiedliche administrative Daten verwendet, um Informationen über die Häufigkeit erbrachter Leistungen im Zeitverlauf, insbesondere über Veränderung nach Erscheinen eines HTA-Berichts, zu gewinnen (z. B. zu Immunglobulinen in der Transplantationsmedizin, zu Stents, Erythropoietin bei Tumoranämie, zu diversen monoklonalen Antikörpern in der Krebstherapie). Während die Daten rasch zur Verfügung standen, war deren Aussagekraft eingeschränkt, und zwar durch mangelnde Granularität (z. B. Mengenverbräuche pro Bundesland, aber nicht pro Krankenhaus), begrenzte Validität (z. B. bei der Zuordnung einer Leistung zu einer konkreten Indikation) und durch limitierte Aussagen zur Kausalität zwischen HTA-Bericht und Veränderung im Mengenverbrauch (andere, nicht bekannte Einflussfaktoren). Sie sind aber in Kombination mit anderen empirischen Daten (z. B. aus Interviews) gut für die Impact-Evaluierung geeignet.

Quelle: (Schumacher und Zechmeister 2011)

5.2.5 Methoden zur Analyse von Routinedaten im Rahmen von HTA

5.2.5.1 Datenvalidität

Zur Beurteilung der Qualität der Daten muss geprüft werden, ob diese als valide/gültig betrachtet werden können. Das heißt, es sollte beantwortet werden, ob das, was gemessen wurde, dem entspricht, was gemessen werden sollte (Holle, Behrend et al. 2005). Eine solche Prüfung ist immer im Zusammenhang mit der Verwendung der Daten im Assessment durchzuführen. Beispielsweise sind Vollständigkeitsverluste bei Kostenangaben von Krankenkassendaten bei der Verwendung für eine Budgetimpactanalyse anders zu beurteilen als bei einer Verwendung für die Beschreibung des Hintergrunds.

Zur Prüfung der Validität existieren keine allgemein gültigen Methoden für alle in dieses Kapitel einbezogenen Kategorien von Routinedaten. Empfehlungen gibt es allerdings für die Validierung von administrativen Daten (Hoffmann, Andersohn et al. 2008). Diese Methoden für Daten der deutschen gesetzlichen Krankenversicherung lassen sich auch auf österreichische administrative Daten umlegen. Eine Möglichkeit ist die interne Konsistenzprüfung mittels Abgleich der Daten mehrerer Krankenversicherungs- oder Krankenanstaltenträger (z. B. Vergleich von Ereignisraten zu bestimmten

Zeitpunkten). Eine interne Konsistenzprüfung kann außerdem über die Suche nach auffälligen Mustern bei einer deskriptiven Darstellung der Daten erfolgen (z. B. grafische Darstellung der Arzneimittelverordnungen im Zeitverlauf).

Ein besonderes Problem bei der Verwendung von administrativen Daten stellen die unspezifische Diagnoseverschlüsselung und die häufig gegebene Vielfalt der Codierungsmöglichkeiten dar. Eine Möglichkeit, die Validität der dokumentierten Diagnosen zu prüfen, ist die interne Diagnosevalidierung, bei der Diagnosen mit vorliegenden Informationen zu anderen erkrankungsspezifischen Aspekten (wie z. B. Verschreibungen) abgeglichen werden. Beispielsweise kann die Abgrenzung von akuten Ereignissen bei dokumentierten Beinvenenthrombosen und Lungenembolie durch gleichzeitiges Vorliegen einer dokumentierten Antikoagulanzen-Verschreibung erfolgen. Eine weitere Möglichkeit ist die Reproduktion eines bekannten Zusammenhangs. In einer deutschen Kohortenstudie wurde zum Beispiel überprüft, ob die Verordnungen eines Psychostimulans (z. B. Pemolin) mit höherem Risiko für stationäre Aufnahmen aufgrund von Verletzungen und Vergiftungen einhergehen (basierend auf der Hypothese, dass hyperkinetische Kinder, denen solche Arzneimittel verschrieben werden, zu Unfällen neigen). Durch die Bestätigung dieser Hypothese konnte aus der Verschreibung der Stimulanzen valide auf die Diagnose „F90 Hyperkinetische Störung“ geschlossen werden (Hoffmann, Andersohn et al. 2008).

5.2.5.2 Datenauswertung

5.2.5.2.1 Deskriptive Auswertung

In den allermeisten Fällen der in Kapitel 5.2.4 genannten Einsatzgebiete für Routinedaten im Rahmen von HTA (z. B. Beschreibung der Krankheitslast oder der Verbreitung einer Technologie) handelt es sich um eine deskriptive Darstellung von Daten. Je nach verwendeten Quellen sind hierbei spezifische Dateneigenschaften oder Unsicherheiten anzuführen (z. B. die verwendeten Ein- und Ausschlusskriterien). Bei der Verwendung von Leistungsabrechnungsdaten ist etwa auf die Validität der Diagnosen einzugehen. Werden Registerdaten verwendet, soll auf die Vollständigkeit der erhobenen Daten eingegangen werden. Wenn statistische Methoden zur Schätzung oder Substitution fehlender Werte eingesetzt werden (z. B. „conditional means“, „canonical correlation“), sollen die Resultate den vorhandenen Originaldaten gegenübergestellt werden (Müller, Augustin et al. 2010).

5.2.5.2.2 Methoden zur Bewertung von Effektivität und Sicherheit auf Basis von Routinedaten

Wenngleich in der Nutzenbewertung einer Technologie Routinedaten eine untergeordnete Stellung einnehmen, besteht bei manchen HTA-Themen ein Bedarf für die Auswertung solcher Daten. Für die Bewertung der Effektivität und Sicherheit kommen am ehesten Routinedaten der Kategorie I infrage, wobei klinische Register und Qualitätsregister eine zentrale Rolle einnehmen. Administrative Daten eignen sich nur dann für eine Bewertung der Effektivität, wenn sie Aussagen zu Technologie UND Outcome auf Ebene der Patienten und Patientinnen ermöglichen. Dies ist bei den administrativen Daten in der derzeitigen Form nur beschränkt möglich (siehe Beispiele in Box 4 + 5), denn die Daten liegen zwar auf Versichertenebene vor, relevante Outcomes lassen sich aber nur eingeschränkt daraus identifizieren bzw. operationalisieren. Problematisch sind etwa gänzlich fehlende, aber relevante Informationen (z. B. zum Raucherstatus bei der Frage nach der Wirksamkeit von Rehabilitationsmaßnahmen) oder getrennte Datenbestände aus dem stationären und niedergelassenen Bereich, sodass Patientenkarrerien nicht verfolgt werden können. Durch intensive Bemühung der Datenverlinkung (z. B. über das GAP-DRG-Projekt im Anhang 9.3.7) könnten zukünftig die Möglichkeiten der Datennutzung ausgeweitet werden.

Eine besondere Rolle spielen Routinedaten bei der Analyse der „comparative effectiveness“ (Cox, Martin et al. 2009, Johnson, Crown et al. 2009). Die Auswertung von Routinedaten hinsichtlich Effektivität oder Sicherheit unterscheidet sich bei den mathematisch-statistischen Methoden prinzipiell nicht von der Auswertung anderer wissenschaftlicher Untersuchungen, allerdings kommt – da es sich vom Studientyp her gesehen um eine Beobachtungsstudie handelt – der Berücksichtigung potenziell konfundierender oder effektmodifizierender Faktoren eine wichtige Bedeutung zu (Müller, Augustin et al. 2010). Häufig sind besondere, multivariate Auswertungsmethoden, Methoden zur Berücksichtigung fehlender Werte oder Sensitivitätsanalysen nötig. Die besonnene Anwendung dieser Auswertungsmethoden kann die Analyse von Kausalzusammenhängen („causal inference“) hinsichtlich der Effektivität von Technologien verbessern (Johnson, Crown et al. 2009). Dazu ist eine statistische Kontrolle unterschiedlicher Confounder und die Kenntnis diverser Verzerrungsformen (bias) notwendig (siehe dazu auch Kapitel 6 im Methodenhandbuch).

Die ISPOR Task Force for Comparative Effectiveness Research (Cox, Martin et al. 2009) führt folgende Biastypen und Confounder, die typisch für Routinedaten sind, an:

- » Klassifikationsbias: die fälschliche Zuordnung von Subjekten zu Exponierten oder nicht Exponierten
 - a) Falsche Klassifizierung der Technologieexposition, z. B.: Patienten/ Patientinnen, die Medikamente nicht über das öffentliche System beziehen (Selbstzahler), werden nicht erfasst
 - b) Falsche Klassifizierung des Outcomes: z. B. falsche Klassifizierung von Diagnosen infolge mangelnder Eindeutigkeit, Stigmata etc.

- » Confounding durch Indikation, z. B.: Wahl der Therapie ist abhängig von Krankheitsschwere, und diese beeinflusst Outcome
- » Residual Confounding: potenzielle Confounder, zu denen in den Daten keine Information vorliegt und für die daher nicht kontrolliert wird
- » Zeitabhängiges Confounding: Variablen, die gleichzeitig Confounder und Risikofaktoren sind (z. B. ist bei der Bewertung des Effekts von Aspirin ein vorangegangener Herzinfarkt einerseits ein Confounder, da er zur Einnahme von Aspirin führt, und gleichzeitig wird das Risiko für Herzinfarkt von Aspirin beeinflusst)

Eine grafische Darstellung der angenommenen biologischen Zusammenhänge und der möglichen konfundierenden Faktoren (kausale Graphen) kann hilfreich sein, die nötigen statistischen Verfahren zu identifizieren (Cox, Martin et al. 2009). Zusätzlich stehen verschiedene (analytische) Methoden zum Umgang mit den Verzerrungspotenzialen und Confoundern bei Routinedatenanalysen zur Verfügung (für ein Beispiel siehe Box 7).

Anwendung von Ein- und Ausschlusskriterien

Zur Minimierung von Confounding kann die untersuchte Patientenkohorte auf eine möglichst homogene Gruppe eingeschränkt werden. Beispielsweise können jene Fälle vor der Analyse ausgeschlossen werden, die aufgrund einer Vorerkrankung ein erhöhtes Risiko für den untersuchten klinischen Outcome und gleichzeitig eine höhere Wahrscheinlichkeit aufweisen, mit dem untersuchten Arzneimittel behandelt zu werden (z. B. vorangegangener Herzinfarkt). Diese Selektion ist aber nur möglich, wenn die gesuchten Merkmale aus den Daten hervorgehen. Bei den österreichischen administrativen Daten erfordert dies sehr wahrscheinlich die (zeitaufwendige) Verknüpfung diverser Datensätze. Zu beachten ist außerdem, dass durch die Anwendung von Ein- und Ausschlusskriterien die externe Validität eingeschränkt werden kann.

Für manche Fragestellungen (z. B. Vergleich eines alten mit einem neuen Cholesterinsenker) kann es auch sinnvoll sein, die untersuchte Kohorte auf jene Fälle einzuschränken, die jeweils *neu* mit einem der Präparate behandelt werden, da prävalente Fälle das Ergebnis verzerren können (z. B. sind prävalente Fälle eher Patienten/Patientinnen, die die Medikamente gut vertragen und einen Nutzen wahrnehmen). Allerdings ist die Unterscheidung zwischen inzidenten und prävalenten Fällen in den Daten nicht immer möglich (z. B. wenn keine fallbezogenen Langzeitinformationen vorhanden sind).

Empfohlen wird außerdem, jene Daten von Patienten/Patientinnen auszuschließen, in denen die untersuchten Interventionen kontraindiziert sind (Cox, Martin et al. 2009). Eine empirisch-statistische Methode, die das ermöglicht, sind Propensity Scores (Cox, Martin et al. 2009, Müller, Augustin et al. 2010).

In Abhängigkeit von der Fragestellung kann es auch sinnvoll sein, Fälle auszuschließen, die eine Therapie vorzeitig abbrechen, indem – wenn es die Daten erlauben – nur jene Patienten/Patientinnen in die Analyse einbezogen werden, die mehrfach Rezepte eingelöst haben.

Um die Generalisierbarkeit bzw. die externe Validität (einer der Vorteile von Routinedatenanalysen) auch in jenen Fällen zu erhalten, wo z. B. die Krankheitsschwere oder Komorbiditäten als Ausschlusskriterium herangezogen werden, kann die Analyse stattdessen in unterschiedliche Subgruppen stratifiziert werden, um statistisch zu testen, ob sich die Behandlungsergebnisse zwischen den Gruppen unterscheiden.

Stratifizierung

Bei der Stratifizierung wird das Datenkollektiv anhand mindestens einer potenziell konfundierenden bzw. effektmodifizierenden Variable in sogenannte Strata (Subgruppen) aufgeteilt (Müller, Augustin et al. 2010). Die statistische Beziehung zwischen der Gruppenzugehörigkeit und dem betrachteten Outcome wird dann in Abhängigkeit von dieser Stratifizierungsvariable ermittelt. Der Vorteil dieser Methode ist ihr intuitiver und praktischer Charakter sowie die verständliche Ergebnisdarstellung. Nachteile sind die oft schnell anwachsende Subgruppenzahl, wodurch die einzelnen Gruppen klein werden und mit geringer statistischer Präzision einhergehen oder zu Zufallsbefunden führen (Johnson, Crown et al. 2009, Müller, Augustin et al. 2010). Die Methode wird jedenfalls als Einstieg vor der Durchführung wesentlich komplexerer multivariater Regressionsmodelle empfohlen (Johnson, Crown et al. 2009).

Matching/Parallelisierung

Hierbei werden für jede Gruppe personenweise Mitglieder identifiziert, die in mindestens einer potenziell konfundierenden/effektmodifizierenden Variable identisch bzw. vergleichbar sind. Die gematchten Gruppen sind dann bezüglich der Matching-Kriterien äquivalent. Bei der Parallelisierung wird die Homogenität der Gruppen bezüglich mindestens einer potenziell konfundierenden/effektmodifizierenden Variable erreicht, indem die Gruppen in ihren deskriptiven Verteilungsparametern angeglichen werden. Dies erfolgt unter Ausschluss einzelner Gruppenteilnehmer, bis die Verteilungseigenschaften der Gruppen vergleichbar ausfallen (Müller, Augustin et al. 2010).

Standardisierung

Bei der Standardisierung werden bezüglich einer potenziell konfundierenden/effektmodifizierenden Variable stratifizierte Gruppen im Outcome nachträglich einer ebenso stratifizierten Vergleichs-(Standard-)Population angeglichen. Dies erfolgt,

indem das Outcome der Gruppenstrata prozentual gemäß der Strataverteilung der Vergleichs-(Standard-)Population gewichtet wird (Müller, Augustin et al. 2010).

Regressionsmodelle

Regressionsmodelle werden empfohlen, wenn multiple Subgruppen gebildet werden oder wenn für mehrere Confounder adjustiert werden muss (Johnson, Crown et al. 2009). Die Wahl des Regressionsmodells hängt vom theoretischen Modell ab, auf dem die Fragestellung beruht, insbesondere aber auch von der abhängigen Variablen. Beispielsweise können dichotome Outcomes (tot/lebendig; ja/nein) mit der logistischen Regression modelliert werden. Ein weiteres Regressionsmodell, das für Routinedaten infrage kommt, ist die „Cox proportional hazards“-Regression, die im Gegensatz zur logistischen Regression nicht von der Annahme ausgeht, dass kein „loss to follow-up“ oder konkurrierende Risiken bestehen und zudem die Aufnahme zeitabhängiger Kovariaten ermöglicht (Johnson, Crown et al. 2009). Beispielsweise kann die Medikamenteneinnahme über die Zeit variieren (z. B. Wechsel der Dosis), wodurch das Cox-Regressionsmodell ein realistischeres Modell für die Analyse des Zusammenhangs zwischen Medikamenteneinnahme und Outcome darstellen kann (Johnson, Crown et al. 2009).

Johnson et al. (2009) formulieren mehrere Qualitätssicherungsmaßnahmen bei der Anwendung von Regressionsmodellen. Dazu gehört, nicht nur die Effektmaße („rate ratio“, „odds ratio“ etc.) zu präsentieren, sondern auch das vollständige Regressionsmodell und dessen Adjustierung für Kovariaten. Unterschiedliche Annahmen, die den Regressionsmodellen zugrunde liegen, sollen vor Anwendung getestet werden (z. B. Annahme des linearen Zusammenhangs bei kontinuierlichen Variablen). Des Weiteren soll die Modellgüte berichtet werden (z. B. R^2 bei Kleinste-Quadrate-Schätzung, c-Statistik bei logistischer Regression).

Propensity-Score-Analyse

Propensity Scores geben die Wahrscheinlichkeit der Zugehörigkeit zu einer Gruppe an und werden über multiple Regression der Gruppenzugehörigkeit (nicht Outcome) auf potenziell konfundierende/effektmodifizierende Variablen durchgeführt (Müller, Augustin et al. 2010). Diese Methode wird zunehmend bei möglichem Selektionsbias oder „Confounding by indication“ in Beobachtungsstudien eingesetzt und eignet sich auch für Routinedatenanalysen (Müller, Augustin et al. 2010). Konfundierende Variablen können damit in Behandlungs- und Kontrollgruppen ausgeglichen werden. Die Scores können mit einer Matching-Methode, mit Stratifizierung oder mit Regressionsmodellen kombiniert werden und haben dieselben Limitationen wie diese (Johnson, Crown et al. 2009, Müller, Augustin et al. 2010).

Sensitivitätsanalysen im Zusammenhang mit „Residual Confounding“

Mit verschiedenen Formen der Sensitivitätsanalyse (z. B. „array approach“, „rule-out approach“) kann der Effekt von potenziellem „Residual Confounding“ quantifiziert werden (Johnson, Crown et al. 2009).

Weitere Methoden

Johnson et al. (2009) führen noch eine Reihe weiterer Methoden an, die für die Analyse von Routinedaten denkbar sind, allerdings bisher selten zum Einsatz kommen und daher nur der Vollständigkeit halber erwähnt werden. Darunter fallen die „Marginal Structural Models“ oder „G-estimation“, die sinnvoll sein können, wenn zeitabhängiges Confounding gegeben ist, das nicht mit den Standardmethoden (Stratifizierung, Regressionsmodelle) behoben werden kann (Johnson, Crown et al. 2009). Ein weiterer Ansatz ist die „Instrumental Variables Analysis“ (IVs Analysis), die aus der Ökonometrie stammt und dort im Zusammenhang mit dem Problem der Endogenität eingesetzt wird, das eng mit dem Problem des „Residual Confounding“ in der klinischen Epidemiologie verwandt ist. Der Ansatz wird als Methode beim Umgang mit „Sample Selection Bias“ angeführt, wobei die Autoren/Autorinnen betonen, dass er umstritten ist und empirische Beweise für seine Überlegenheit gegenüber gängigeren Verfahren bislang ausstehen. Schließlich wird das „Structural Equation Modeling“ erwähnt, mit dem sich differenzierte Information über Behandlungseffekte gewinnen lässt, die über erwartete Mittelwert-Unterschiede gängiger Methoden hinausgehen.

Box 5.7.**Umgang mit Confounding am Projektbeispiel
„Ambulante kardiologische Rehabilitation“**

Die in der Studie untersuchten Vergleichsgruppen waren durch unterschiedliche Ausgangswerte (z. B. in klinischen Parametern) in ihrer Vergleichbarkeit eingeschränkt. Sofern die Unterschiede in den Ausgangswerten der Vergleichsgruppen in den Routinedaten dokumentiert waren, wurden sie mittels logistischer Regression der Gruppenzugehörigkeit auf Basis verschiedener Merkmale (Alter, Geschlecht, Berufsstand etc.) und klinischer Parameter und mittels Bildung eines stichprobenspezifischen Propensity Scores berücksichtigt. Eine Adjustierung von Gruppenvergleichen unter Berücksichtigung des Propensity Scores wurde bei kontinuierlichen Merkmalen mittels multipler Regression und des stratifizierten Wilcoxon-Tests vorgenommen. Bei Vergleichen kategorialer Merkmale wurde eine logistische Regression unter Einschluss des Propensity Scores berechnet und neben p-Werten die rohen und adjustierten Odds Ratios angegeben. Für Überlebensdaten wurden Cox-Modelle einmal mit dem Propensity Score als Stratifizierungsmerkmal und einmal als linearem Prädiktor geschätzt. Stark limitierend ist jedoch die Tatsache, dass zahlreiche Gruppenunterschiede durch eine große Anzahl fehlender Werte (z. B. zum Raucherstatus) prinzipiell unbekannt blieben.

Quelle: (Gyimesi, Piso et al. 2010)

5.2.5.2.3 Methoden zur Kostenberechnung mit Routinedaten**Ermittlung von Preisen aus Routinedaten**

Häufig enthalten Routinedaten lediglich abgerechnete Häufigkeiten einzelner Leistungspositionen und eine dazugehörige Punktezahl, so wie dies beispielsweise bei den administrativen Daten der leistungsorientierten Krankenanstaltenfinanzierung (LKF) der Fall ist (siehe Beschreibung im Anhang 9.3.1). Zur Ableitung von Kostengrößen müssen diese Punkte monetär bewertet werden. Dies geschieht in aller Regel mithilfe eines einheitlichen Durchschnittswertes pro Punkt (Holle, Behrend et al. 2005). Ein klassischer Ansatz zur Ermittlung eines Bewertungsfaktors für LKF-Punkte ist die Division der stationären Endkosten durch die Summe der LKF-Punkte des jeweiligen Jahres. Limitierend bei dieser Methode ist, dass der tatsächlich ausgezahlte Punktwert von den unterschiedlichen Landesfonds festgelegt wird. Die angegebenen bewerteten LKF-Punkte sind daher nur eine grobe Schätzung für die Ausgaben für stationäre Leistungen. Zudem existiert im Krankenanstaltenbereich eine Reihe von nicht über das LKF-System abgegoltenen Kosten, die hier ebenfalls nicht enthalten sind und zu einer tendenziellen Unterschätzung der tatsächlichen Kosten führen.

Bei Routinedaten, die direkt Preisinformationen liefern (z. B. Tarife zu Leistungen der Sozialversicherungsträger aus der Metahonorarordnung), entfällt zwar der Berechnungsschritt zur Bewertung einer Leistung, allerdings sind diverse Einschränkungen zu beachten. Beispielsweise sind Tarife zu Metahonorarordnungspositionen (siehe Anhang 15.5.1.1.1) nur dann erfasst, wenn es sich um bundesweit gültige Gesamtverträge handelt (z. B. Mutter-Kind-Pass) oder wenn bei der Erarbeitung eines Musterkataloges Tarife kalkuliert wurden (Musterkataloge werden unter medizinischer Beratung erarbeitet für Bereiche mit hohem Honoraraufwand wie z. B. Laborleistungen von Experten/Expertinnen der KV-Träger und des Hauptverbandes und dienen den KV-Trägern als Verhandlungsunterlage und Empfehlung). Für andere Leistungspositionen muss bei Fragen zu Tarifen ein Durchschnittstarif aus den Trägerpositionen gerechnet werden. Die Leistungs- und Tarifaktualisierungen liegen in der Verantwortung der Kassen und erfolgen meist nach Gültigkeitsbeginn (keine Tagesaktualität). Durch Regelungen bei der Vertragsgestaltung mit den Vertragspartnern (Limitierungen bzw. Degression von bestimmten Tarifen bei Mehrfacherbringung einer Leistung) kann der in der Honorarordnungsdatenbank ausgewiesene Tarif höher sein als der in der Abrechnungsrealität gezahlte Tarif.

Kostenarten und Routinedaten

Im Zusammenhang mit Kostenmessungen für Wirtschaftlichkeitsanalysen wird immer wieder betont, dass diese möglichst weitgehend dem Opportunitätskostenprinzip¹ entsprechen sollen (Holle, Behrend et al. 2005). Nach der mikroökonomischen Theorie (die jedoch umstritten ist) spiegeln Marktpreise die Opportunitätskosten wider. Mit Verweis auf die erhebliche Bedeutung staatlicher Regulierung und verbandlicher Aushandlungsprozesse für die Preisbildung bei medizinischen Leistungen (z. B. Tarifverhandlungen zwischen Ärztekammer und Sozialversicherungsträger) wird diese Eigenschaft den administrativen Ausgabendaten meistens abgesprochen. Oft ist nicht einmal die Kostenkalkulation der Verhandlungspartner, die zur Bildung von Tarifen führen soll, vorhanden oder nachvollziehbar (Rechnungshof 2011). Statt der Verwendung von administrativen Preisen empfehlen einzelnen Experten/Expertinnen eine Mischkalkulation unter Einbeziehung von Marktpreisen (z. B. Preise für die Leistungen privater Krankenversicherungen) (z. B. Holle, Behrend et al. 2005). Klassische Methodenhandbücher (z. B. Drummond, Sculpher et al. 2005) betonen wiederum, dass bei einer ökonomischen Evaluation, die aus der Perspektive der Kostenträger durchgeführt wird, die administrativen Ausgabendaten relevanter sind als die tatsächlichen Kosten. Außerdem gibt es Möglichkeiten, die zur Kostenberechnung herangezogenen Tarife

1

Opportunitätskosten entsprechen dem entgangenen Nutzen, der dadurch entsteht, dass die vorhandenen nächstbesten alternativen Möglichkeiten zur Nutzung von Ressourcen nicht wahrgenommen werden.

mit weiteren Methoden (z. B. der Berechnung von hospital-level cost-to-charge ratios) zu adjustieren (Drummond, Sculpher et al. 2005).

Werden Kostenrechnungsdaten für Kosten- und Leistungsvergleiche verwendet, sind je nach Fragestellung unterschiedliche Kostenarten zu berücksichtigen, die unterschiedliche Routinedatenbedarfe nach sich ziehen. Eine grundlegende Entscheidung bei der Kostenbestimmung ist die Auswahl von Durchschnitts- oder Grenzkosten (Drummond, Sculpher et al. 2005, Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen 2009). Erstere entsprechen den Kosten pro produzierter Einheit, Letztere umfassen die entstehenden Zusatzkosten, wenn die Produktion um eine Mengeneinheit ansteigt (z. B. wenn eine zusätzliche Patientin behandelt wird). Bei der Berechnung gehen bei den Durchschnittskosten Fixkosten (Gebäude, Anlagen etc.) in die Kalkulation ein, bei den Grenzkosten hingegen nicht (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen 2009).

Die Auswahl des Kostenkonzepts – und demnach der benötigten Routinedaten – hängt von der Forschungsfrage ab. Vor der Verwendung von Routinedaten ist zu prüfen, ob die Dateneigenschaften dem ausgewählten Kostenkonzept Rechnung tragen. Bei der Verwendung von Kostenrechnungsdaten für Kosten- und Leistungsvergleiche werden etwa die Daten aus der Vollkostenrechnung empfohlen, wenn es sich um eine zusätzliche bauliche Leistung handelt, die umfangreiche bauliche, personelle und apparative Investitionen erfordert (z. B. radiologische Leistungen). Die Verwendung von Teilkostenrechnungsdaten wird hingegen dann empfohlen, wenn es sich um eine Leistung handelt, die im Rahmen der vorhandenen Infrastruktur, jedoch mit zusätzlicher Geräteinvestition, erbracht werden kann (z. B. endoskopische Leistung). Die Grenzkostenrechnung soll dann eingesetzt werden, wenn Leistungen im Rahmen der vorhandenen Infrastruktur ohne weitere Investitionen erbracht werden können (z. B. wenn ein vorhandenes Gerät besser ausgelastet werden kann) (Rechnungshof 2011). Aufgrund von Beschränkungen in der Datengranularität können Abweichungen von dieser Vorgangsweise erforderlich sein. Eine transparente Methodendokumentation ist daher wichtig.

Statistische Auswertung von individuellen Patientendaten

Für die statistische Auswertung von Kostendaten steht grundsätzlich der gleiche Methodenkanon statistischer Verfahren zur Verfügung wie für Daten anderer Art. Wenn Kostendaten basierend auf individuellen Beobachtungen bzw. „patient-level data“ verwendet werden, ergeben sich einige Besonderheiten, die bei der Methodenwahl zu berücksichtigen sind (Holle, Behrend et al. 2005). Dies ist durch die häufig vorliegende Verteilungsschiefe von Kostendaten und der damit zusammenhängenden Varianzheterogenität beim Vergleich mehrerer Gruppen bedingt. Gründe für die Verteilungsschiefe sind die Begrenzung der Daten nach unten durch die Null und das Vorhandensein

einzelner Ausreißer mit extrem hohen Kosten. Dadurch entsteht eine rechtsschiefe Verteilung.

Klassische Ansätze zum Umgang mit schiefen Verteilungen (z. B. Verwendung robuster Schätzer wie des Medians) sind bei der Auswertung von Kostendaten nicht geeignet, da jeder Wert zu den Gesamtkosten beiträgt. Demnach ist hier der Mittelwert einer Kostenverteilung ein aussagekräftigerer Parameter.

Eine weitere Möglichkeit ist die Anwendung geeigneter Transformationen, mit der eine Normalisierung der Daten erreicht wird (z. B. Log-Transformation). Diese Methode ist allerdings umstritten, da z. B. die logarithmische Transformation mit anschließender Rücktransformation einen Schätzer für das geometrische, nicht aber für das arithmetische Mittel der Kostenverteilung ergibt.

Außerdem können eine parametrische Verteilung (z. B. Gammaverteilung) und eine Schätzung des Erwartungswertes dieser Verteilung mit den jeweiligen Maximum-Likelihood-Schätzern zur Bestimmung des effizientesten Schätzers des Erwartungswertes erfolgen. Besonders bei kleinen Datensätzen kann allerdings die Entscheidung über die am besten passende Verteilung schwierig sein.

Zur Konstruktion von Vertrauensbereichen für Schätzgrößen im Zusammenhang mit Kostendaten auf „patient-level“-Niveau wird das Bootstrap-Verfahren empfohlen. Sollen Kostendaten in Abhängigkeit unterschiedlicher Einflussgrößen beschrieben werden, finden regressionsanalytische Ansätze Einsatz.

5.2.6 Routinedaten in Österreich

5.2.6.1 Kategorisierung

In diesem Kapitel werden die im Zuge der Recherche identifizierten in Österreich vorhandenen Routinedaten im Überblick dargestellt. Eine Aufstellung der Datenbestände findet sich im Anhang (Kapitel 15.4). Davon werden ausgewählte Datenquellen im Anhang detaillierter beschrieben (Kapitel 15.5). Tabelle 5.3 zeigt, welche Merkmale zu den Daten erhoben wurden. Diese inkludieren zunächst die Datenkategorien „administrative Daten“, „Register“, „Surveys“, „amtliche nationale und internationale Statistiken (auf Grundlage eines gesetzlichen Auftrags)“, „Daten zu Planungszwecken“, „Datenbestände zu Versorgungsstrukturen“, „sonstige Daten“. Die Datenbeschreibung enthält außerdem eine Kurzinformation zum Dateneigner, zum Originalzweck, zur Art und Weise der Datenerhebung/Datenregistrierung inklusive des Datenniveaus (fallbezogen, aggregiert) und zum Zugang zu den Daten inklusive der Information, ob ein

direkter Zugriff auf die Originaldaten möglich ist oder ein indirekter Zugang über Aggregatstatistiken oder Sonderauswertungen besteht. Außerdem werden die Daten hinsichtlich ihrer Eigenschaften nach der Klassifikation von Raftery et al. (2008) aus Kapitel 5.2.3 eingeordnet, und es wird beschrieben, in welchem Kontext eines HTA-Prozesses es denkbar ist, die einzelnen Datenbestände zu verwenden. Die Tabelle zu den Datenbeständen erhebt keinen Anspruch auf Vollständigkeit.

Tabelle 5.3:
Übersicht zur Datenbeschreibung

Datenkategorie
» Administrative Daten
» Register
» Surveys
» Amtliche nationale und internationale Statistiken
» Daten zu Planungszwecken
» Datenbestände zu Versorgungsstrukturen
» Sonstige Daten
Dateneigner (Institutionen, die die Daten erheben, speichern und nützen sowie die rechtliche Verfügungsgewalt darüber besitzen)
Originalzweck (seit wann, zu welchem Zweck, evt. gesetzliche Grundlage)
Datenerhebung / Generierung / Registrierung (Wie und zu welchen Zeitpunkten werden die Daten erhoben bzw. registriert, welche Daten werden erhoben bzw. registriert?)
Zugang zu Daten (direkt/indirekt*, wer, wie, rechtliche Aspekte, Kosten)
Datenverfügbarkeit (abgedeckter Zeitraum)
Klassifikation nach Raftery et al. (2008)
Einsatz im HTA-Prozess (siehe Kapitel 4)
Referenz (link zu Datenbestand oder zu Information über die Daten)
* direkte Nutzung: Zugriff auf die (anonymisierten oder pseudonymisierten) Originaldaten; indirekte Nutzung: Aggregatstatistiken oder Sonderauswertungen durch den Dateneigner

Quelle: LBI, Darstellung: GÖG/BIQG

5.2.6.2 Rechercheergebnisse

Datenkategorien (siehe Anhang Punkt 15.4 und 15.5)

Bei den administrativen Daten konnten Datenbestände auf allen Verwaltungsebenen bzw. Gebietskörperschaften identifiziert werden, die mehrheitlich bereits für HTA-Projekte verwendet wurden. Darunter fallen die Sozialversicherungsdaten, Daten des Bundesministeriums für Gesundheit (z. B. die administrativen Daten der Krankenanstaltenfinanzierung), Kostenrechnungsdaten der Krankenanstaltenträger bzw. der Krankenanstaltengesellschaften sowie Daten der Landessanitätsdirektionen (z. B. Impfstatistik).

In Österreich existieren außerdem eine Reihe von epidemiologischen Registern (z. B. Geburtenregister, Tumorregister), Qualitätsregister (z. B. Erwachsenen-Herzchirurgie-Register, Kinderkardiologie-Register) sowie risiko-, krankheits- oder interventionsbezogene Register (z. B. klinische Tumorregister, Medizinproduktregister).

Periodische Umfragen (insbesondere Gesundheits-Surveys) wurden nur vereinzelt identifiziert (z. B. Österreichische Befragung zur Patientenzufriedenheit, HBSC-Survey), hingegen existiert eine große Anzahl an amtlichen nationalen und internationalen Statistiken wie etwa die Gesundheitsstatistik der Statistik Austria, die Bevölkerungsstatistik oder die OECD Health Data.

Des Weiteren gibt es einige Datenbestände, deren ursprünglicher Zweck mit Planungsagenden zusammenhängt und die im Rahmen eines HTAs von Nutzen sein können wie etwa der Österreichische Strukturplan Gesundheit (ÖSG) und die diversen regionalen Strukturpläne (RSG) oder Daten zur Erreichbarkeit bzw. zu Einzugsgebieten.

Nicht zuletzt wurden einige Datenbestände identifiziert, die Versorgungsstrukturen bzw. das Leistungsangebot abbilden (z. B. Rehakompass, Spitalskompass).

Dateneigner

Die größten Datenbestände hinsichtlich Dateneigner befinden sich bei den Gebietskörperschaften (insbesondere beim Bundesministerium für Gesundheit und bei den Sozialversicherungsträgern) oder in Institutionen des Gesundheitssystems (z. B. Institut für Epidemiologie der Tiroler Landeskrankenanstalten GmbH, AGES PharmMed, Gesundheit Österreich GmbH), gefolgt von der Statistik Austria. Außerdem gibt es einige Datenbestände von medizinischen Fachgesellschaften (z. B. Dialyse- und Transplantationsregister der Österreichischen Gesellschaft für Nephrologie). Auch über internationale Institutionen (WHO, Eurostat) sind zahlreiche österreichischspezifische Daten abrufbar.

Originalzweck

Der Originalzweck, zu dem die Daten erhoben bzw. registriert werden, reicht von primär administrativen Zwecken bei den administrativen Datenbeständen über Qualitätssicherungsziele (z. B. Qualitätsregister) bis hin zu Zwecken der Politikgestaltung, Planung und Steuerung (z. B. diverse Gesundheitssurveys).

Datenerhebung/Generierung/Registrierung

Die Daten entstehen zum einen routinemäßig im Rahmen diverser administrativer Prozesse (z. B. elektronische Dokumentation der Leistungsabrechnung). Zum anderen werden sie zusätzlich erhoben. Dies kann in der Form von – teilweise verpflichtenden – Datenmeldungen an die Dateneigner erfolgen (z. B. Register, Pharmakovigilanzdatenbank) oder in Form von (Stichproben-)Befragungen entsprechender Bevölkerungsgruppen (z. B. Haushaltsangehörige für Mikrozensus, Patienten/Patientinnen für Patientenzufriedenheitssurvey, Schüler/innen für HBSC-Survey). Zunehmend werden mehrere Datenbestände kombiniert und miteinander verlinkt (z. B. Volkszählung in Form einer Verlinkung unterschiedlicher Verwaltungsregister/Registerzählung, OECD Health Data). Die Verlinkung kann sowohl auf Individualdatenniveau (z. B. Registerzählung) als auch auf Aggregatebene (z. B. ÖGIS) erfolgen.

Zugang zu den Daten

Die wenigsten Datenbestände sind vollständig öffentlich zugänglich, und auch im Fall eines öffentlichen Datenzugangs besteht dieser selten als direkter Zugriff auf die (pseudonymisierten oder anonymisierten) Originaldaten, sondern zumeist indirekt im Sinne eines Zugriffs auf Aggregatstatistiken oder Sonderauswertungen durch die Dateneigner. Gründe dafür sind vor allem Datenschutzbestimmungen und sonstige gesetzliche Regelungen, „Eigentumsverhältnisse“ (z. B. wenn die Dateneigner private Institutionen sind) und die generelle Kultur einer eingeschränkten Datentransparenz im österreichischen Gesundheitswesen.

Am häufigsten sind Auswertungen der Daten in Form von periodischen Berichten oder in aggregiert-tabellarischer Form öffentlich und online zugänglich (z. B. Sozialversicherungsstatistik, IVF-Registerbericht). Einige Bestände (z. B. Qualitätsregister an der GÖG) sind hingegen in keiner Form öffentlich zugänglich, sondern nur von den beteiligten Zentren einsehbar.

Sonderauswertungen für projektspezifische Fragen sind, sofern sie überhaupt möglich sind, fast immer kostenpflichtig. Dies trifft insbesondere auf die Datenbestände der Statistik Austria und der GÖG zu. Eine Ausnahme bilden die Register des Instituts für Epidemiologie der TILAK (IET), wo Auswertungen bislang kostenlos durchgeführt werden können. Bei den administrativen Datenbeständen sind Datenauswertungen dann möglich, wenn die entsprechende Rechtsgrundlage gegeben ist (nachvollziehbarer Rechtstitel wie z. B. universitäre Einrichtungen, Studierende, Einhaltung datenschutzrechtlicher Bestimmungen) und/oder wenn die Auswertung von öffentlichem Interesse ist. Daten können außerdem im Namen von Patienten/Patientinnen angefordert werden, sofern eine Zustimmung der Betroffenen vorliegt.

Verfügbare Zeitperiode

Die verfügbaren Zeitperioden variieren stark nach Datenbestand und Indikator. Einzelne Daten (z. B. Bevölkerungsdaten) oder Indikatoren (z. B. Gesundheitsausgaben) reichen bis in die 1950er/1960er Jahre zurück, in der weitaus größten Anzahl beginnt die Verfügbarkeit jedoch in den 1990er Jahren, was durch verbesserte technische Möglichkeiten und veränderte gesetzliche Rahmenbedingungen bedingt ist (z. B. Bundesgesetz über die Dokumentation im Gesundheitswesen (BGBl. Nr. 745/1996 idgF). Eine dritte Gruppe an Daten, worunter vor allem diverse Register fallen, ist erst seit wenigen Jahren verfügbar, oder die Erhebung wird überhaupt erst aufgebaut.

5.2.7 Fazit

Der Einsatz von Routinedaten im Rahmen eines HTA-Prozesses gewinnt zunehmend an Bedeutung und ist für einzelne Schritte im HTA-Prozess eine notwendige Voraussetzung für die Produktion von relevantem Wissen für Entscheidungsträger/innen. Sowohl die steigende Anzahl der Datenbestände als auch Initiativen zu deren Verlinkung beeinflussen diese Entwicklung. Nicht zuletzt werden Routinedaten aufgrund des steigenden Bedarfs zur Evaluation komplexer Interventionen zukünftig eine wichtige Rolle für die Weiterentwicklung von HTA spielen.

Nichtsdestotrotz sind nicht alle Routinedaten gleichermaßen für alle Schritte im HTA-Prozess geeignet. Während sie für die Priorisierung eines Themas, für die Hintergrundbeschreibung und für die ökonomische Analyse häufig eingesetzt werden (sollen), spielen sie für die Nutzenbewertung einer Technologie eine wesentlich geringere Rolle und können aufgrund ihrer Eigenschaften robuste klinische Studien oder Metaanalysen keinesfalls ersetzen. Die Qualität und die Eigenschaften der Daten erfüllen häufig die Erwartungen nicht und lassen bei der Ergebnisinterpretation oft keine schlüssigen Aussagen oder gar Beschreibungen von Kausalzusammenhängen zu. Für manche Datenbestände (z. B. Register) könnten durch Änderungen bei der Datenerhebung und verstärkte Zusammenarbeit mit HTA-Institutionen Verbesserungen erzielt werden.

Eine adäquate Auswertung von Routinedaten setzt gute Kenntnisse der jeweiligen Institutionen bzw. Dateneigner und von deren administrativen Strukturen bzw. Abläufen ebenso voraus wie Sorgfalt bei der Plausibilitätsprüfung der Daten und methodische Erfahrung bei der Auswahl und Anwendung geeigneter statistischer Verfahren (Holle, Behrend et al. 2005).

Da in der vorliegenden Version des Methodenhandbuchs etliche Datenbestände unberücksichtigt blieben bzw. unbekannt sind, ist ein regelmäßiges Update zu den Datenquellen sowie zu den Möglichkeiten statistischer Auswertungen empfehlenswert.

5.2.8 Danksagung

Zahlreiche Personen haben mich bei der Recherche und Beschreibung der österreichischen Datenbestände unterstützt. Mein Dank gilt insbesondere Willi Oberaigner, Gottfried Endel, Ingrid Wilbacher, Jutta Schröder, Rudolf Weissitsch, Monika Hackl, Heimo Wallenko, Andreas Egger sowie zahlreichen Mitarbeitern und Mitarbeiterinnen der GÖG. Ich bedanke mich außerdem bei Brigitte Piso und Claudia Wild für ihre Kommentare zu Vorversionen dieses Kapitels.

Autorin: Ingrid Zechmeister-Koss (LBI-HTA)

6 Beurteilung der medizinischen Studien und Extraktion der Daten

6.1 Beurteilung des Bias-Risikos (interne Validität)

Ein wesentlicher Aspekt von HTA-Berichten ist die kritische Evaluierung von Studien, die den Selektionskriterien entsprechen. Das Ziel dabei ist, das methodische Design und die Qualität der Ausführung einer Studie zu beurteilen, um abschätzen zu können, wie hoch das Bias-Risiko (interne Validität) bei jeder Studie ist (Glasziou, Vandenbroucke & Chalmers 2004; Lohr/Carey 1999). Eine gut durchgeführte methodische Evaluierung ermöglicht, die beste verfügbare Evidenz in Bezug auf definierte wissenschaftliche Forschungsfragen zu identifizieren. HTA-Berichte, die schlecht durchgeführte Studien mit hohem Bias-Risiko zusammenfassen, werden ebenfalls zu verzerrten oder falschen Ergebnissen führen. Die methodische Qualität der einzelnen Studien spielt auch eine wichtige Rolle bei der Synthese der Evidenz und bei der Erstellung von Empfehlungen.

Unter Bias versteht man systematische Fehler, die zu einer Verzerrung der Resultate und Schlussfolgerungen führen (Higgins/Green 2008). Vorhandensein und Ausmaß von Bias kann jedoch nicht direkt gemessen, sondern nur indirekt über die Beurteilung der Qualität des Studiendesigns und der Ausführung einer Studie bestimmt werden. Mit anderen Worten: Gutes Studiendesign und gute Ausführung einer Studie lassen darauf schließen, dass das Risiko für Bias gering ist und dass Resultate möglichst korrekt sind, wobei Zufallsvariabilität nicht beurteilt werden kann.

Die Möglichkeiten zur Beurteilung des Bias-Risikos hängen von der vorhandenen Information in den publizierten Artikeln ab. Eine schlechte Beschreibung der Methoden in den Publikationen bedeutet aber nicht automatisch ein hohes Bias-Risiko. Bei fehlender oder unklarer Information sollten die Studienautoren kontaktiert und um Klarstellung gebeten werden.

Kriterien zur Beurteilung des Bias-Risikos

Es gibt eine Vielzahl an Instrumenten zur Beurteilung des Bias-Risikos (West et al. 2002). Methodikstudien haben jedoch gezeigt, dass Punktesysteme zu wenig verlässlichen Ergebnissen führen (Juni et al. 1999). Eine bessere Methode ist, elementare Komponenten des Designs und der Ausführung von Studien zu beurteilen (Juni, Altman & Egger 2001). Solche Komponenten unterscheiden sich je nach Studientypen. Bei der Beurteilung des Bias-Risikos wird eine Studie in Bezug auf diese Komponenten evaluiert.

Das Bias-Risiko einzelner Studien ist immer von zwei Personen unabhängig zu beurteilen. Differenzen werden durch Diskussion und Konsens oder durch das Einbeziehen einer dritten Person gelöst. Standardisierte Formulare, die zur Beurteilung des Bias-Risikos verschiedener Studientypen verwendet werden können, sind im Anhang unter 15.6 zusammengefasst. Die Instrumente zur Beurteilung von RCTs und systematischen Reviews basieren auf Systemen, die von der Cochrane Collaboration (Higgins/Green 2008) und dem Centre for Reviews and Dissemination, University of York (NHS Centre for Reviews and Dissemination 2001) entwickelt wurden. Beobachtungsstudien werden nach Kriterien beurteilt, die von Deeks et al. beschrieben wurden (Deeks et al. 2003). Die interne Validität diagnostischer Studien wird unter Berücksichtigung von Kriterien beurteilt, die im Quality Assessment of Diagnostic Accuracy Studies (QUADAS) dargelegt wurden (Whiting et al. 2003).

Das Bias-Risiko hängt auch von zu beurteilenden Endpunkten ab. Beispielsweise erhöht eine fehlende Verblindung das Bias-Risiko für die meisten Endpunkte. Für Endpunkte, bei denen subjektive Einschätzungen keine Rolle spielen (z. B. Mortalität), hat fehlende Verblindung keinen Einfluss auf das Bias-Risiko. Die Beurteilung des Bias-Risikos sollte daher immer endpunktspezifisch durchgeführt werden. In Anlehnung an das Cochrane-Handbuch verwenden wir in den Formularen folgende Klassifizierung des Bias-Risikos für individuelle Studien (Tabelle 6.1):

Tabelle 6.1:
Klassifizierung des Bias-Risikos

Bias-Risiko	Voraussetzung	Interpretation
Geringes Bias-Risiko	Alle elementaren Komponenten zeigen ein geringes Bias-Risiko.	Es ist unwahrscheinlich, dass Bias das Ergebnis der Studie wesentlich verzerrt.
Unklares (mittleres) Bias-Risiko	Das Bias-Risiko ist für eine oder mehrere Komponenten unklar.	Es ist unklar, inwieweit Bias die Ergebnisse der Studie verzerren könnte. Bias ist möglich und könnte die Korrektheit der Resultate infrage stellen.
Hohes Bias-Risiko	Das Bias-Risiko ist für eine oder mehrere Komponenten hoch.	Es ist sehr wahrscheinlich, dass Bias das Ergebnis der Studie wesentlich verzerrt. Das Vertrauen in die Korrektheit der Resultate ist sehr gering.

Quelle: Cochrane-Handbuch (Higgins/Green 2008); GÖG/BIQG-eigene Darstellung

6.1.1 Therapie

6.1.1.1 RCTs

Randomized controlled trials sind der Goldstandard bei der klinischen Prüfung von Interventionen. Sie sind das beste Studiendesign, um kleine, aber klinisch wesentliche Behandlungseffekte zu untersuchen, da sie am besten in der Lage sind, Bias und Confounding zu minimieren. (Confounding liegt vor, wenn ein Faktor, der Confounder, sowohl mit der Intervention/Exposition als auch mit dem Studienoutcome in Zusammenhang steht und damit den Studienausgang beeinflusst bzw. verfälscht.) Bei der Beurteilung der Wirksamkeit von Interventionen in HTA-Berichten stellen RCTs daher zumeist die beste verfügbare Evidenz dar.

In RCTs werden Studienteilnehmer/innen durch verschiedene Randomisierungsmethoden per Zufall (at random) einer Interventions- oder Kontrollgruppe zugeteilt. Die Unvorhersehbarkeit der Zuteilung führt zu annähernd gleichen Studiengruppen mit gleichmäßig verteilten prognostischen Faktoren, die in nicht-randomisierten Studien als Confounder Ergebnisse verzerren könnten. RCTs werden immer prospektiv durchgeführt. Bei vielen RCTs ist durch die strenge Selektion der Studienpopulation die Anwendbarkeit der Ergebnisse auf die Normalbevölkerung (externe Validität) eingeschränkt. Pragmatische RCTs versuchen durch geringere Selektion, praxisrelevantere Durchführung und den Einsatz von patientenrelevanten Endpunkten eine bessere externe Validität als bei explanatorischen RCTs (Studien mit stark selektierten Populationen unter hochkontrollierten Bedingungen) zu erzielen (siehe Kapitel 6.2.1). Höhere Praxisrelevanz wird bei pragmatischen RCTs dadurch erreicht, dass beispielsweise allgemeinmedizinische Praxen als Studiensetting dienen, keine Überwachung der Medikamenteneinnahme erfolgt oder die Dosierung einer Medikation variabel ist und Patientenbedürfnissen angepasst wird.

In den folgenden Abschnitten werden Kriterien zusammengefasst, die bei der kritischen Evaluierung von RCTs in Betracht gezogen werden sollten (Higgins/Green 2008).

Kriterien zur Beurteilung von RCTs

Wurde eine adäquate Methode der Randomisierung verwendet, um Studienteilnehmer/innen den Behandlungsgruppen zuzuteilen?

Unter Randomisierung versteht man die auf Zufall beruhende Zuteilung einer Person zu einer Studiengruppe. Es gibt unterschiedliche Arten der Randomisierung (z. B. einfache Randomisierung, Blockrandomisierung, stratifizierte Randomisierung u. a.), die abhängig von Größe und Design einer jeweiligen Studie eingesetzt werden können.

Wichtig ist jedoch immer, dass das Zufallsprinzip und die Unvorhersehbarkeit der Zuteilung gewahrt bleiben. Inadäquat durchgeführte Randomisierung führt zu Selektionsbias. Die meisten Randomisierungen werden computergeneriert durchgeführt.

Beispiele für inkorrekte Randomisierungsmethoden wären:

- » Sortierung nach Geburtsdaten,
- » nach personenbezogenen Daten wie Ambulanz- oder Sozialversicherungsdaten,
- » abwechselnde Zuordnung (etwa ABAB..., die ersten fünf usw.).

Oft wird das Auswahlverfahren der Studienteilnehmer/innen als „Quasi-Randomisierung“ beschrieben. Dies ist meist nur die Bezeichnung für eine unvollständig durchgeführte Randomisierung, die das Zufallsprinzip oder die Unvorhersehbarkeit nicht gewährleistet. Quasi-Randomisierungen können zu Selektionsbias führen und sind einer korrekten Randomisierung nicht gleichzusetzen. Die Bezeichnung „Quasi-Randomisierung“ in einer Studie sollte dazu veranlassen, den Randomisierungsprozess einer kritischen Betrachtung zu unterziehen.

*War die Geheimhaltung der Randomisierungssequenz gewährleistet?
(Allocation Concealment)*

Neben dem auf dem Zufallsprinzip basierenden Erstellen einer Reihenfolge, nach der die Studienteilnehmer/innen den einzelnen Studienarmen zugeteilt werden (Randomisierungssequenz) ist der zweite wichtige Aspekt der Randomisierung das Geheimhalten dieser Sequenz (Allocation Concealment). Bei korrekt durchgeführter Geheimhaltung kann der Untersucher nicht erkennen, welcher Studiengruppe der nächste Proband angehören wird.

Weiß der Untersucher schon im Vorfeld, welcher Gruppe eine Patientin / ein Patient zugeordnet werden wird, so könnte er Einfluss darauf nehmen und so das Untersuchungsergebnis verfälschen. Methodenstudien haben außerdem gezeigt, dass mangelnde Geheimhaltung der Randomisierungssequenz zu einer Überschätzung von Behandlungseffekten führt (Wood et al. 2008).

Bei offener Zuweisung von Patientinnen/Patienten besteht die Gefahr der Manipulation durch deren bewusstes Ein- oder Ausschließen.

Beispiele für adäquate Geheimhaltung der Randomisierungssequenz wären:

- » sequenziell nummerierte, verschlossene, undurchsichtige Umschläge,
- » seriell nummerierte oder kodierte identische Container,
- » zentrale Randomisierung (Zuteilung über das Internet bzw. auch telefonisch),

- » jeder andere Prozess, der es unmöglich macht, die Reihenfolge der Gruppenzuordnung vorzusehen.

Beispiele für inadäquate Geheimhaltung der Randomisierungssequenz wären:

- » durchsichtige, unverschlossene Briefumschläge,
- » durchlaufend nummerierte Briefumschläge,
- » offene Randomisierungslisten,
- » spezifische, personenbezogene Daten (Ambulanznummern, Geburts- oder Sozialversicherungsdaten).

Waren die Behandlungsgruppen nach der Randomisierung ähnlich in Bezug auf wesentliche prognostische Merkmale oder Confounder?

Trotz gut durchgeführter Randomisierung kann es bei kleinen Studien zu einer Ungleichverteilung von wesentlichen prognostischen Faktoren kommen. Es sollte daher immer eine Überprüfung der Vergleichbarkeit der Gruppen stattfinden.

Waren Studienteilnehmer/innen verblindet?

Waren jene Personen, die die Interventionen verabreichten, verblindet?

Waren jene Personen, die die Endpunkte erhoben, verblindet?

Die Verblindung bei klinischen Studien dient dazu, eine eventuelle unabsichtliche Beeinflussung der Studienergebnisse durch Untersucher/in und/oder Teilnehmer/in auszuschalten. Adäquate Verblindung verhindert Messbias. Verblindung ist vor allem dann wesentlich, wenn Zielparameter gemessen werden, die durch Subjektivität beeinflussbar sind (Day, Altman 2000). Bei Endpunkten, die durch Subjektivität nicht beeinflussbar sind (z. B. Mortalität), ist eine Verblindung nicht nötig.

Studien können einfach oder doppelt verblindet sein. Bei einer einfach verblindeten Studie weiß die Patientin / der Patient nicht, in welcher Studiengruppe sie/er sich befindet bzw. weiß die/der Untersuchende nicht, welcher Intervention die Patientin / der Patient zugeteilt wurde. Bei einer doppelt verblindeten Studie wissen weder Untersucher/in noch Patient/in, welcher Interventionsgruppe die Patientin / der Patient angehört.

Eine doppelt verblindete Studie hat eine geringere Anfälligkeit für etwaige systematische Fehler als eine einfach verblindete, jedoch ist Doppel-Verblindung nicht immer möglich (z. B. bei chirurgischen Eingriffen). In solchen Fällen sollten zumindest die Untersucher/innen, die Endpunkte beurteilen, verblindet sein (Evaluator blinding).

Erhielten alle Studiengruppen neben der zu untersuchenden Intervention identische Behandlungen?

Um Unterschiede in Wirksamkeit und Sicherheit einer Intervention feststellen zu können, darf es keine Verschiedenheiten bei sonstigen Interventionen und Behandlungen zwischen den beiden Gruppen geben. Wird beispielsweise der Endpunkt einer Knieoperation in Bezug auf zwei verschiedene Operationsmethoden beurteilt, so würde das Ergebnis verfälscht werden, wenn eine Behandlungsgruppe postoperative Physiotherapie erhalten würden und die andere nicht.

Wurden die Endpunkte in allen Behandlungsgruppen zum selben Zeitpunkt erhoben?

Um Messbias zu vermeiden, müssen Endpunkte in allen Behandlungsgruppen zum selben Zeitpunkt gemessen werden.

Waren die Drop-out-Raten hoch?

Die Drop-out-Rate (Attrition, Loss to follow-up) ist der Prozentsatz an Personen, die randomisiert wurden, jedoch nicht bis zum Ende an der Studie teilgenommen haben.

Man unterscheidet die allgemeine von der differentiellen Drop-out-Rate.

Als **allgemeine Drop-out-Rate** wird der Prozentsatz aller Studienteilnehmer/innen bezeichnet, deren Daten nach erfolgter Randomisierung aus diversen Gründen (Übersiedelung, Nebenwirkungen, mangelnde Adherence etc.) nicht mehr verfügbar sind.

Es gibt derzeit keine klaren Richtlinien, wie hoch die allgemeine Drop-out-Rate sein darf. Der Richtwert wird häufig mit 20 Prozent angenommen, dieser Wert basiert jedoch nicht auf Methodikstudien. Aus pragmatischen Gründen wird bei speziellen Populationen (z. B. Patientinnen und Patienten mit psychiatrischen Erkrankungen) die Grenze auch oft höher angesetzt.

Eine zu hohe Drop-out-Rate führt zu Selektionsbias. Die Auswirkung der Drop-out-Rate auf eine einzelne Studie sollte aber von Fall zu Fall beurteilt werden, da sie auch von der zugrunde liegenden Ereignisrate des zu messenden Zielparameters abhängt. Drop-out-Raten können zu einer größeren Verzerrung führen, wenn die Ereignisrate des zu messenden Zielparameters niedrig ist. Bei Zielparametern mit hoher Ereignisrate ist der Einfluss der Drop-out-Raten meist geringer. Als allgemeiner Grundsatz gilt, dass Drop-out-Raten dann hohes Verzerrungspotenzial haben, wenn der Verhältniswert von Personen ohne Enddaten zu Personen, die bis zum Studienende teilgenommen haben, hoch ist (Higgins/Green 2008).

Die **differenzielle Drop-out-Rate** bezeichnet den Unterschied im Verhältnis zwischen den Drop-out-Raten der Behandlungsgruppen.

Auch bei der differenziellen Drop-out-Rate gibt es keine allgemeingültigen Standards. Als ungefähre Grenzwert werden häufig 15 Prozent angegeben.

Wurde eine Intention-to-treat-(ITT-)Analyse durchgeführt und war diese korrekt?

Eine Intention-to-treat-Analyse gilt als beste statistische Auswertung für RCTs, weil sie die Vorteile einer Randomisierung erhält. Dies bedeutet: Idealerweise sollten alle Patientinnen und Patienten, die in eine Gruppe randomisiert wurden, auch bei der Studienanalyse als dieser Gruppe zugehörig miteingerechnet werden. Dies sollte unabhängig von den Gründen geschehen, die dazu führten, dass eine Intervention gegebenenfalls abgebrochen wurde (Non-Adherence, Cross-Over, Fehldiagnose) (Heritier et al. 2003).

Andere statistische Auswertungen wie Per-protocol-Analysen, Completers-only-Analysen oder As-treated-Analysen können interessante Zusatzinformation liefern, sollten aber nur gemeinsam mit ITT-Analysen durchgeführt werden.

Kann angenommen werden, dass alle erhobenen Endpunkte auch erhoben wurden?

Publikations-Bias und Outcome-Reporting-Bias sind ein häufiges Problem in der medizinischen Literatur, können aber nur schwer überprüft werden. Bei publizierten Studien ist es wichtig zu klären, ob die primären Endpunkte, die in den Methoden angeführt wurden, auch in den Resultaten adäquat behandelt werden. Falls das Studienprotokoll registriert wurde, kann man im entsprechenden Register die a priori definierte Endpunkte einsehen.

6.1.1.2 Kohortenstudien

In einer Kohortenstudie wird eine definierte Gruppe von Menschen (Kohorte) einem bestimmten Einfluss (Exposition) ausgesetzt und über einen bestimmten Zeitraum beobachtet, um einen Zusammenhang zwischen Exposition und einem definierten Endpunkt darzustellen (Grimes/Schulz 2002c). Man unterscheidet zwischen prospektiven und retrospektiven Kohortenstudien. In einer prospektiven Kohortenstudie werden die Daten zur Exposition vor der Erfassung des Ereignisses erhoben (z. B. einer medizinischen Intervention). In einer retrospektiven Kohortenstudie werden die Teilnehmer/innen aus bereits vorhandenen Daten bezüglich der gewünschten Exposition

identifiziert und aus dem Verlauf bis in die Gegenwart bezüglich eines Endpunktes beurteilt (Grimes/Schulz 2002b).

Autorinnen und Autoren haben die Möglichkeit, Ergebnisse von Kohortenstudien in einen HTA-Bericht aufzunehmen, wenn die Fragestellung nicht durch Ergebnisse randomisiert-kontrollierter Studien (RCTs) zu beantworten ist (Higgins/Green 2008), z. B.:

- » zur Darstellung des positiven oder negativen Effekts von Interventionen, die nicht randomisiert werden können oder die mit hoher Wahrscheinlichkeit nicht in einem RCT behandelt werden können, z. B. aus ethischen Gründen (in diesem Zusammenhang kann eine Übersichtsarbeit, die über die Ergebnisse und Limitationen vorhandener nicht-randomisierter Studien berichtet, nützlich sein);
- » zur Untersuchung von Endpunkten, die nicht in RCTs behandelt werden können. Endpunkte, die weit in der Zukunft liegen oder extrem selten sind (z. B. seltene Nebenwirkungen eines Medikaments), werden durch RCTs meistens nicht erfasst.

Kohortenstudien werden schon seit vielen Jahrzehnten durchgeführt und leisten nach wie vor einen wichtigen Beitrag zu studienbasiertem Wissen (Mathis/Gartlehner 2008). Methodische Besonderheiten können allerdings zu verzerrten oder falschen Ergebnissen führen. Eine aufmerksame Evaluierung von Kohortenstudien nach bestimmten Kriterien ist notwendig, um vorhandene Mängel aufzudecken.

Kriterien zur Beurteilung von Kohortenstudien

Das methodologische Hauptproblem von Kohortenstudien ist Selektionsbias. Während bei RCTs die Zuteilung zu Untersuchungsgruppen durch Zufall bestimmt wird, kann die Zuteilung in Kohortenstudien durch unterschiedliche Faktoren beeinflusst werden. Dies kann zu offensichtlichen oder aber auch zu unbekanntem Unterschieden in prognostischen Faktoren zwischen den Vergleichsgruppen führen. Dieses Selektionsbias kann in der Folge zu systematischen Verzerrungen der Ergebnisse führen (Grimes/Schulz 2002b). Ein weiteres Problem, das bei Kohortenstudien auftreten kann, ist Confounding (Vandenbroucke et al. 2007). Darunter versteht man den Einfluss von Faktoren, die sowohl mit der Exposition als auch mit dem Endpunkt assoziiert sind, aber kein intermediärer Endpunkt sind (d. h. nicht direkt Gegenstand der Studienfragestellung sind), und die Ergebnisse einer Kohortenstudie verfälschen können. Deshalb sind bestimmte statistische Methoden (z. B. Stratifizierung, multiple Analyse) zur korrekten Auswertung von Kohortenstudien notwendig. Der Einfluss von unbekanntem Confounding auf die Ergebnisse einer Kohortenstudie kann jedoch auch durch statistische Methoden nicht verhindert werden.

Folgende Fragen sollten bei der kritischen Beurteilung von Kohortenstudien betrachtet werden (Deeks et al. 2003):

Selektion

Wurden die Studiengruppen aus derselben Population und während derselben Zeitperiode rekrutiert?

Um Selektionsbias zu verhindern, sollten Interventionsgruppe und Kontrollgruppe aus derselben Population stammen. Werden Studiengruppen zu unterschiedlichen Zeitpunkten oder aus unterschiedlichen Populationen rekrutiert, können prognostische Faktoren (z. B. Fortgeschrittenheit einer Erkrankung) einen unterschiedlichen Einfluss auf die Endpunkte haben und Confounding verursachen (Mathis/Gartlehner 2008, Deeks et al. 2003). Dies kann insbesondere beim Einsatz von historischen Kontrollgruppen (Kontrollgruppen, die aus zeitlich älteren Datensätzen generiert werden) problematisch werden.

Wurde durch die Autoren ausgeschlossen, dass ein definierter Endpunkt bereits zu Studienbeginn vorhanden war?

In der Beurteilung einer Kohortenstudie ist es wichtig festzustellen, ob ein definierter Endpunkt nicht schon zu Beginn der Studie vorhanden war (Mathis/Gartlehner 2008).

Wurden Interventionen in allen Gruppen auf gleiche Art und Weise beurteilt?

In einer prospektiven Kohortenstudie ist offensichtlich, welcher Teilnehmer / welche Teilnehmerin die Intervention erhalten hat. In einer retrospektiven Kohortenstudie muss dies durch Datenanalyse, strukturierte Interviews und Selbstbericht durch den Teilnehmer / die Teilnehmerin ermittelt werden. Die Methodik muss für Interventionsgruppe und Kontrollgruppe gleich sein. Datenanalyse und strukturierte Interviews sind dem Selbstbericht durch den Patienten / die Patientin vorzuziehen (Higgins/Green 2008, Wells et al.).

Vergleichbarkeit

Ist die Verteilung prognostischer Faktoren zwischen den Gruppen ausreichend beschrieben?

In vielen Publikationen werden in Tabelle 1 die Charakteristika der Studiengruppen beschrieben. Die wichtigsten prognostischen Faktoren bezüglich eines bestimmten Endpunkts sollten dort verglichen werden.

Ist die Verteilung der prognostischen Faktoren zwischen den Gruppen ähnlich?

Die prognostischen Faktoren zwischen den Studiengruppen sollten zu Beginn der Studie möglichst ähnlich sein. Große Unterschiede deuten auf Selektionsbias hin (Mathis/Gartlehner 2008). Unterschiede müssen in der statistischen Analyse berücksichtigt werden (Deeks et al. 2003).

Endpunkte

Wurden Endpunkte auf dieselbe Art und verblindet beurteilt?

Messfehler („measurement bias“ und „detection bias“) können durch eine Verblindung der Personen, welche die Endpunkte beurteilen, verhindert werden. Eine Verblindung ist allerdings nicht immer möglich. Die Beurteilung der Endpunkte muss in allen Gruppen auf gleiche Art und Weise durchgeführt werden (Beobachtungsgleichheit) (Grimes/Schulz 2002b).

Wurden potenzielle Confounder in der statistischen Auswertung berücksichtigt?

Es gilt festzustellen, ob die richtigen statistischen Methoden zur Korrektur eines Confoundings eingesetzt wurden. Unbekannte Confounder können allerdings auch trotz statistischer Verfahren zu einem systematischen Unterschied zwischen den Studiengruppen führen (Deeks et al. 2003, Vandenbroucke et al. 2007).

War die Studienlaufzeit adäquat und für alle Gruppen identisch?

Der Beobachtungszeitraum muss lange genug für den Eintritt des gewünschten Endpunktes gewählt werden und für beide Gruppen gleich sein. Zur prospektiven Festlegung eines adäquaten Zeitraumes soll eine klinische Expertin / ein klinischer Experte hinzugezogen werden. Unterschiedliche Studienlaufzeiten können zu Selektionsbias führen (Higgins/Green 2008).

Gab es eine hohe Drop-out-Rate?

Die Drop-out-Rate („attrition“, „loss to follow-up“) beinhaltet jene Studienteilnehmer/innen, die zu Studienbeginn in der Kohorte enthalten waren, aber nicht bis zum Studienende verblieben sind. Eine erhöhte Drop-out-Rate führt ebenfalls zu Selektionsbias (Higgins/Green 2008).

Gab es eine hohe differenzielle Drop-out-Rate?

Problematisch ist auch eine unterschiedlich hohe Drop-out-Rate in den unterschiedlichen Studiengruppen. Obwohl keine allgemeingültigen Standards existieren, wird ein Unterschied von 15 Prozentpunkten oft als Grenze gesehen (Mathis/Gartlehner 2008).

Wurde die Datenanalyse adäquat ausgeführt?

Da es aufgrund fehlender Randomisierung zu einer Ungleichverteilung von Confoundern kommen kann, muss die Datenanalyse dies berücksichtigen und für mögliche Confounder adjustieren.

6.1.1.3 Fall-Kontroll-Studien

In einer Fall-Kontroll-Studie wird die Häufigkeit eines potenziellen Risikofaktors in einer Fallgruppe (Personen, die an einer Erkrankung leiden) und in einer Kontrollgruppe (Personen, die nicht an einer Erkrankung leiden) verglichen, um einen Zusammenhang zwischen einer Krankheit oder Veränderung (dem Endpunkt) und einer vorhergegangenen Exposition zu überprüfen (Schulz/Grimes 2002). Der Begriff Exposition kann sich sowohl auf Noxen oder Schadstoffe beziehen als auch auf medizinische Interventionen wie die Einnahme eines bestimmten Medikaments. Die Relevanz von Fall-Kontroll-Studien bei der Durchführung von HTA-Berichten liegt weniger bei der Beurteilung der Effektivität als hauptsächlich bei der Beurteilung von Nebenwirkungen.

Fall-Kontroll-Studien sind besonders aufschlussreich,

- » wenn die untersuchte Krankheit sehr selten ist oder sich langsam entwickelt (Higgins/Green 2008);
- » wenn für seltene Nebenwirkungen randomisierte Studien zu klein sind, um einen signifikanten Unterschied zwischen den Gruppen entdecken zu können; um diese zu entdecken, wären sehr große langfristige Kohortenstudien erforderlich (Guyatt et al. 2008, Levine et al. 1994);
- » wenn Verbindungen zwischen seltenen Ereignissen und unterschiedlichen Expositionen untersucht werden sollen; Fall-Kontroll-Studien haben somit eine wichtige Rolle bei der Suche nach Evidenz über die Ursachen potenzieller Nebenwirkungen und unbeabsichtigter nützlicher Effekte von Interventionen (Schulz/Grimes 2002);
- » für die Evaluation von großen Public-Health-Interventionen, wie z. B. Unfallverhütung und Screening, die sehr schwierig oder durch randomisierte Versuche nur teuer zu evaluieren sind (Higgins/Green 2008; MacLehose et al. 2000).

Kriterien zur Beurteilung von Fall-Kontroll-Studien

Wie bei Kohortenstudien stellt Selektionsbias auch bei Fall-Kontroll-Studien eines der zentralen Probleme dar (Schulz, Grimes 2002). Um Selektionsbias zu minimieren, ist es wichtig, dass jede Person in der Kontrollgruppe theoretisch bei einer Erkrankung auch in der Gruppe der „Fälle“ eingeschlossen werden hätte können. Die Charakteristika der „Fall-Population“ und der „Kontroll-Population“ müssen aber nicht unbedingt identisch sein. Nachfolgende Kriterien sollten für die kritische Bewertung von Fall-Kontroll-Studien betrachtet werden (Deeks et al. 2003, Wells et al.):

Selektion

Wurde ein „Fall“ exakt definiert?

Es ist wichtig, dass die Autoren einer Studie einen „Fall“ detailliert beschrieben haben, am besten anhand einer international akzeptierten Definition. Autoren von systematischen Reviews müssen mit den verschiedenen internationalen Definitionen von Krankheiten vertraut sein und deren Verwendung in Fall-Kontroll-Studien überprüfen. Anzumerken ist, dass die akzeptierten Definitionen und Klassifizierungen von Erkrankungen laufend geändert werden (Higgins/Green 2008).

Wurden die Studiengruppen aus derselben Population rekrutiert?

Fall- und Kontrollgruppe müssen aus derselben Population rekrutiert werden, um Selektionsbias zu verhindern. Oft werden für Studien Nachbarn/Nachbarinnen oder Familienmitglieder als Kontrollpersonen genommen (Vandenbroucke et al. 2007). Trotzdem können sich Kontrollgruppen wesentlich in Begleiterkrankungen, zusätzlichen Risikofaktoren, prognostischen Faktoren usw. unterscheiden (Grimes/Schulz 2005).

Hatten die Personen in der Fallgruppe und in der Kontrollgruppe dasselbe Risiko, der Exposition ausgesetzt zu werden?

Die Personen in Fall- und Kontrollgruppen sollten dasselbe Risiko gehabt haben, der Exposition ausgesetzt zu werden. Die Reviewerin / der Reviewer muss sich fragen, ob die Personen in der Kontrollgruppe die Exposition vielleicht vermieden haben. Ein häufiges Beispiel dafür ist, wenn eine Kontrollgruppe in einem Spital ausgesucht wird. Fälle (mit Krankheit A) werden mit einer Gruppe von Patientinnen und Patienten von einer anderen Station verglichen (mit Krankheit B). Es kann sein, dass die Kontrollpersonen aufgrund ihrer Erkrankung der Exposition weniger oft ausgesetzt sind (z. B. seltenere Einnahme eines Medikaments). Dies führt zu dem Ergebnis, dass Personen

mit Krankheit A dem Medikament häufiger ausgesetzt waren, und zu der falschen Schlussfolgerung, dass dies ein Grund für Krankheit A ist. Tatsächlich hatten die beiden Gruppen nicht dasselbe Risiko, der Exposition ausgesetzt zu sein (Grimes/Schulz 2005).

Ist die Verteilung prognostischer Faktoren zwischen den Gruppen ausreichend beschrieben?

In vielen Publikationen werden in Tabelle 1 die Charakteristika der Studiengruppen beschrieben. In dieser Tabelle sollen die wichtigsten prognostischen Faktoren bezüglich eines bestimmten Endpunkts verglichen werden. Alle relevanten prognostischen Faktoren sollten von den Autorinnen und Autoren wie auch bei der kritischen Evaluierung von Kohortenstudien in Absprache mit klinischen Expertinnen und Experten bereits im Voraus bestimmt werden (Higgins/Green 2008).

*Ist die Verteilung der prognostischen Faktoren zwischen den Gruppen ähnlich?
Ist „Over-matching“ möglich?*

Bei der Auswahl der Kontrollpersonen wird bereits darauf geachtet, dass diese Faktoren gleich sind („matching“), z. B. werden Familienmitglieder oder Nachbarn als Personen in einer Kontrollgruppe rekrutiert. Wichtig ist, dass die wesentlichen Unterschiede bei der statistischen Analyse adjustiert werden (Schulz/Grimes 2002; Wells et al.).

Jedoch kann bei einer zu großen Ähnlichkeit der Fall- und Kontrollgruppen Bias als Folge auftreten. Man spricht von „Over-matching“, wenn sich in beiden Gruppen Faktoren, die kausal mit dem Endpunkt verbunden sind, sehr stark annähern. „Over-matching“ kann zu falsch-negativen Ergebnissen führen, d. h., die Studie liefert negative Ergebnisse, obwohl in der Realität ein Zusammenhang zwischen der Exposition und dem Endpunkt zu finden wäre (Higgins/Green 2008; Vandenbroucke et al. 2007).

Wurde die Exposition in Fall- und Kontrollgruppen auf gleiche Art und Weise und im besten Fall verblindet beurteilt?

Der Endpunkt ist bei einer Fall-Kontroll-Studie zu Beginn bekannt. Die Beurteilung der Exposition muss in beiden Gruppen auf identische, valide und verlässliche Art und Weise durchgeführt werden („Beobachtungsgleichheit“). Eine verblindete Befragung der Teilnehmer/innen durch die Studienautorinnen bzw. Studienautoren ist eine häufige Methode, um die Endpunkte zu beurteilen. Wenn die Befragung mittels Interview

durchgeführt wird, müssen dieselben Rahmenbedingungen (z. B. telefonisch oder persönlich) nach einem bestimmten Interview-Protokoll gelten (Wells et al.).

Da Kranke tendenziell ein besseres Erinnerungsvermögen betreffend Exposition gegenüber Risikofaktoren aufweisen als Kontrollpersonen (dieses Phänomen wird „Recall bias“ genannt), sind Methoden, die nicht auf das Erinnerungsvermögen von Patientinnen/Patienten und Kontrollpersonen angewiesen sind, zu bevorzugen (z. B. Krankenblatt oder Datenbanken) (Higgins/Green 2008; Schulz/Grimes 2002).

War die Rate der Personen, die eine Beteiligung an der Studie abgelehnt haben, in beiden Gruppen ähnlich?

Wie in RCTs und Kohortenstudien muss auch bei Fall-Kontroll-Studien aktiv rekrutiert werden. Sollten bei der Rekrutierung einer Gruppe Schwierigkeiten auftreten, kann dies zu ungleichen Studiengruppen führen. Eine große Bedeutung erfährt somit die Bekanntgabe der Anzahl der von den Studienautorinnen/-autoren befragten Personen in beiden Gruppen, die eine Teilnahme abgelehnt haben. Sollten viele Personen die Teilnahme an einer Studie abgelehnt haben, kann dies zu unbekanntem, aber wichtigen Unterschieden zwischen den beiden Gruppen führen (Wells et al.; Schulz/Grimes 2002).

Wurde eine adäquate statistische Analyse durchgeführt?

Unterschiede in prognostischen Faktoren sollten mittels multifaktorieller Methoden während der statistischen Analyse adjustiert werden. Trotz ausführlicher statistischer Korrekturen von bekannten prognostischen Faktoren bleibt immer ein Restrisiko, dass unbekanntes Confounder zu systematischen Unterschieden zwischen den Studiengruppen führen (unmeasured confounding) (Grimes/Schulz 2002a). Odds Ratios sind jenes Effektmaß, das bei Fall-Kontroll-Studien verwendet wird.

6.1.1.4 Systematische Übersichtsarbeiten

Systematische Übersichtsarbeiten gelten heute als das methodisch beste Instrument zur Beantwortung von klinischen und versorgungspolitischen Fragestellungen (Gartlehner et al. 2008). Eine gut durchgeführte systematische Übersichtsarbeit folgt standardisierten methodischen Kriterien, um eine objektive Zusammenfassung der besten verfügbaren Evidenz zu erstellen. Die systematische Beurteilung der Evidenz unterscheidet systematische Übersichtsarbeiten von konventionellen, nicht-systematischen Übersichtsarbeiten (narrative Reviews) (Mulrow 1994). Systematische Übersichtsarbeiten bieten jedoch anders als narrative Reviews einen objektiven und

aktuellen Zugang zur Evidenz, sie dienen auch als Instrument, um bestehende Lücken im Wissen hervorzuheben (Lang et al. 2007). Bestehende systematische Übersichtsarbeiten stellen daher eine wertvolle Ressource für HTA-Berichte dar (Whitlock et al. 2008; Hopewell et al. 2002). Die methodische Qualität publizierter systematischer Übersichtsarbeiten ist jedoch unterschiedlich. Eine kritische Beurteilung der Methoden ist daher immer vonnöten, um die Aussagekraft von Resultaten beurteilen zu können.

Kriterien zur Beurteilung von systematischen Übersichtsarbeiten

Das methodische Leitbild bei systematischen Übersichtsarbeiten besteht darin, nach vorab bestimmten Prinzipien systematisch vorzugehen, um Selektionsbias zu minimieren. Die folgenden Kriterien wurden im Anschluss an Vorschläge von Oxman, Guyatt 1993 sowie Balk et al. 2005 adaptiert.

Basiert die systematische Übersichtsarbeit auf einer klar definierten (PICO-)Forschungsfrage?

Die wissenschaftliche Frage, die eine systematische Übersichtsarbeit zu beantworten versucht, muss a priori formuliert und nicht post hoc der vorhandenen Evidenz angepasst werden (Counsell 1997). Sie muss präzise sein und soll Populationen, Interventionen, Kontrollinterventionen und Endpunkte klar definieren (Krupski et al. 2008).

Wurden Selektionskriterien klar definiert?

Selektionskriterien für Studien müssen ebenfalls a priori definiert werden und sollen klare Entscheidungen bezüglich Ein- oder Ausschluss von Studien ermöglichen. Eine systematische Übersichtsarbeit soll nicht eine Zusammenfassung der gesamten, jemals publizierten Evidenz sein, sondern soll die beste verfügbare Evidenz zusammenfassen.

Wurde eine umfangreiche systematische Literatursuche durchgeführt?

„Retrieval bias“ (Nichtauffinden von relevanter publizierter Literatur trotz umfangreicher Suche; Chapman et al. 2009) lässt sich nie gänzlich vermeiden, kann aber mit einer systematischen Literatursuche minimiert werden (Krupski et al. 2008). Suchbegriffe sollen dabei klar dargestellt werden und müssen die gesamte Fragestellung abdecken (Sampson et al. 2009).

Fehlende Systematik bei der Suche kann zusätzlich zu Selektionsbias führen. Eine umfangreiche systematische Literatursuche setzt sich aus zumindest drei Komponenten zusammen:

1. Literatursuche in mehreren elektronischen Datenbanken
Eine alleinige Suche in PubMed ist unzureichend, da nur eine limitierte Anzahl von Journals in PubMed indexiert ist (Betran et al. 2005). Weitere elektronische Datenbanken wie Embase, Cinahl (Cumulative Index to Allied and Nursing Health Literature), die Cochrane Library oder Datenbanken, die auf bestimmte Fachbereiche spezialisiert sind, müssen daher immer routinemäßig durchsucht werden.
2. Handsuche
Um das „Retrieval bias“ zu minimieren, muss die elektronische Suche durch Handsuche ergänzt werden (z. B. durch Durchsicht von Referenzlisten oder Citation tracking) (Hopewell et al. 2002).
3. Suche nach grauer Literatur
Eine Recherche nach nicht in Journalen publizierter „grauer“ Literatur soll eine Suche komplettieren (McAuley et al. 2000).

Wurde der Ein- oder Ausschluss von Studien von zumindest zwei Personen beurteilt?

Die Entscheidung, ob Studien die a priori definierten Auswahlkriterien erfüllen, ist zu einem gewissen Grad auch von der subjektiven Einschätzung der Reviewer/innen abhängig. Entscheidungen über Ein- und Ausschluss sowie über die methodische Qualität sollen daher von zwei Personen getroffen werden, um falsch-positive oder falsch-negative Resultate zu minimieren (Edwards et al. 2002). Bei Unterschieden in der Beurteilung, die nicht durch einen Konsens gelöst werden können, muss eine dritte Person involviert werden.

Wurde die methodische Qualität der eingeschlossenen Studien von zumindest zwei Personen beurteilt?

Unzureichende methodische Qualität (interne Validität) und hohes Bias-Risiko kann Resultate verzerren und zu einer Überschätzung des gemessenen Effekts führen. Die Beurteilung der internen Validität der Studien soll mithilfe eines validierten und verlässlichen Schemas (z. B. des Cochrane-Handbuchs) erfolgen.

Wurde die methodische Qualität der Studien bei der Evidenzsynthese berücksichtigt?

Studien mit guter interner Validität sollen bei der Synthese der Evidenz mehr Gewicht bekommen als Studien mit methodischen Problemen. Vor allem Studien mit hohem Bias-Risiko sollen nur in Ausnahmefällen bei der Evidenzsynthese berücksichtigt werden.

6.1.2 Qualitative Studien

Die meisten HTA-Berichte zielen darauf ab, Effektivität und Wirksamkeit von Interventionen sowie deren soziale, ökonomische und ethische Auswirkungen zu untersuchen. Evidenz von qualitativen Studien wird in HTA-Berichten selten eingeschlossen, kann aber wertvolle zusätzliche Information für Entscheidungsträger beinhalten, vor allem was ethische und soziale Auswirkungen betrifft. Qualitative Forschung liefert einen wesentlichen Beitrag zur Erforschung relevanter Themen im Bereich der Gesundheit und des Gesundheitswesens, beispielsweise zum Verständnis des subjektiven Erlebens von Gesundheit und Krankheit und zum Einfluss von Barrieren (z. B. Erreichbarkeit, Selbstbehalte oder Nebenwirkungen) auf die Inanspruchnahme von Gesundheitsdienstleistungen (Kuper et al. 2008).

Daher finden qualitative Methoden immer mehr Eingang in die Evaluation von Gesundheitsmaßnahmen und -interventionen (Pope, Mays 1995). Sie haben das Potenzial, die Relevanz und Nützlichkeit einer systematischen Übersichtsarbeit oder eines HTA deutlich zu verbessern (Higgins/Green 2008).

Der Stellenwert qualitativer Forschung zeigt sich unter anderem daran, dass die Food and Drug Administration (USA) empfiehlt, bei Studien zur Arzneimittelzulassung patientenrelevante Endpunkte auch mittels Interviews und Fokusgruppen zu erheben (U.S. Department of Health & Human Services), oder dass das Medical Research Council (UK) explizit die Verwendung von qualitativen Methoden bei der Erforschung komplexer Interventionen vorsieht (Campbell et al. 2000).

Definition von qualitativer Forschung

Was üblicherweise als quantitative bzw. als qualitative Forschung bezeichnet wird, lässt sich nicht lediglich dadurch unterscheiden, ob Zahlen fehlen oder vorkommen (Kuper et al. 2008).

Denzin & Lincoln definieren qualitative Forschung wie folgt: „... a situated activity that locates the observer in the world. It considers a set of interpretive, material practices that make the world visible ... [it] involves an interpretive, naturalistic approach to the world. This means qualitative researchers study things in their natural settings, attempting to make sense of, or to interpret, phenomena in terms of the meanings people bring to them“ (Denzin, Lincoln 2000).

Ziel der qualitativen Forschung ist es, ein Phänomen (z. B. Verhalten, Interaktionen, Bedeutungen, Werte) in einer bestimmten Gruppe zu beschreiben, es zu verstehen und zu erklären (Liamputtong/Ezzy 2005). Um das zu erreichen, werden viele verschiedene Methoden zur Datengewinnung (z. B. ethnographische Studien, Interviews, Fokusgrup-

pen, Tagebücher), zur Datenanalyse und zur Interpretation angewendet (z. B. konzeptionell-analytisch, historisch, kulturell, dekonstruktiv, grounded theory).

Die Flexibilität in der Wahl der Methoden wird als eine Stärke der qualitativen Forschung gesehen (Pyett 2001), denn diese ermöglicht eine dem Forschungsgegenstand angemessene Methode.

Kriterien zur Bestimmung der Güte qualitativer Forschungsarbeit

Die Vielfalt der verschiedenen Zugänge zu Erhebung, Analyse und Interpretation von qualitativen Daten führt dazu, dass Checklisten zur Beurteilung der Güte einer qualitativen Arbeit von vielen Forscherinnen und Forschern als unzureichend abgelehnt werden (z. B. Higgins/Green 2008, Green/Thorogood 2004, Popay et al. 1998, Pope/Mays 1995, Barbour 2001, Spencer 2003). Viele Autorinnen und Autoren (z. B. Kuper, Lingard/Levinson 2008, Kitto et al. 2008, Daly et al. 2007) sehen die Beurteilung von qualitativer Forschungsarbeit als eine interpretative Handlung, die einerseits Grundlagenwissen über qualitative Forschung und andererseits eine reflektierende Auseinandersetzung anstatt einer mechanischen Anwendung einer regelbasierten Checkliste erfordert. Ein Beispiel für einen fundierten und differenzierten Ansatz zur Bewertung von qualitativen Studien bietet Spencer 2003. Die folgenden Fragen beruhen auf verschiedenen Publikationen zur Beurteilung der Qualität qualitativer Forschung (z. B. Daly et al. 2007, Liamputtong/Ezzy 2005, Kuper et al. 2008, Kitto et al. 2008, Spencer 2003). Sie verstehen sich nicht als Checkliste, sondern als Leitlinien, mit deren Hilfe der Wert qualitativer Forschungsarbeiten beurteilt werden kann. Daher setzen sie im Gegensatz zu einem Abhaken formaler Kriterien eine inhaltliche Auseinandersetzung mit der Forschungsarbeit voraus:

Gibt es eine klare Forschungsfrage bzw. ist diese in einen theoretischen Rahmen eingebettet?

Auch qualitative Forschung beruht auf klaren, verständlichen und konkreten Fragestellungen. Diese sind in Bezug zur Forschungsliteratur zu setzen (Daly et al. 2007). Da es Ziel der qualitativen Forschung ist, Verhalten, Interaktionen, Bedeutungen, Werte ... von Individuen, Gruppen oder Organisationen zu beschreiben, zu verstehen und zu erklären (Liamputtong, Ezzy 2005), gilt es für die jeweilige Fragestellung Erklärungsmodelle für das Verhalten von Menschen bzw. Institutionen aus den jeweiligen wissenschaftlichen Disziplinen (z. B. Soziologie oder Psychologie) heranzuziehen (Daly et al. 2007).

Ist die Auswahl der Stichprobe geeignet, die Forschungsfrage zu beantworten?

Im Gegensatz zur quantitativen Forschung hat qualitative Forschung im Idealfall keine vordefinierte Stichprobengröße (Daly et al. 2007). Die Erhebung endet, wenn das Phänomen, das es zu untersuchen gilt, erklärt bzw. verstanden ist; dieser Zustand wird in der Literatur als „theoretische Sättigung“ bezeichnet. Das heißt, dass durch Hinzunahme weiterer Personen kein bedeutender zusätzlicher Erkenntnisgewinn erreicht werden kann. Die Stichprobenziehung sollte explizit und nachvollziehbar beschrieben werden (Kuper et al. 2008, Kitto et al. 2008).

Erfolgte die Datenerhebung durch jene Methode, die für das Setting am geeignetsten ist?

Die verwendete Methode der Datenerhebung (z. B. Beobachtung, Fokusgruppe, Interviews, Tagebücher) ist – idealerweise mit Bezug zur Literatur – zu begründen. Weiters sollte nachvollziehbar sein, wie die Methoden zur Anwendung kamen (Daly et al. 2007). Dabei gilt es spezielle Aufmerksamkeit auf methodische Artefakte zu legen, die das Ergebnis beeinflussen können (z. B. Hawthorne-Effekt: Inwieweit beeinflusst die Anwesenheit eines Beobachters / einer Beobachterin das Verhalten der Beobachteten?) (Kuper et al. 2008).

Ist die Art der Datenanalyse explizit und in Bezug zur Literatur dargestellt?

Auch bei der Datenanalyse gilt es, eine klare, nachvollziehbare Beschreibung des Vorgehens zu geben. Es ist darzulegen, wie genau die Analyse erfolgte und wer sie durchgeführt hat. Wenn mehr als ein Forscher die Daten analysiert hat, soll dargestellt werden, wie mit Widersprüchen in den Ergebnissen der Analysen umgegangen wurde (Kuper et al. 2008). Im Auswertungsprozess werden inhaltlich gleiche Ergebnisse (Aussagen, Beobachtungen ...), die aber in unterschiedlichen Worten formuliert sind, in Kategorien zusammengefasst. Diese „Kategorisierung“ sollte im Idealfall sowohl unter Einbeziehung von theoretischen Grundlagen (aus der einschlägigen Literatur) als auch auf Basis der gewonnenen Daten erfolgen. Für die Beurteilung einer qualitativen Studie ist des Weiteren wesentlich, ob die verwendeten Kriterien erklärt werden (Daly et al. 2007).

Sind die Schlussfolgerungen nachvollziehbar?

Es gilt, die Ergebnisse der Studie (z. B. Befragungs- bzw. Beobachtungsergebnisse) stimmig zu erklären (Daly et al. 2007), sodass das untersuchte Phänomen für die Leserin / den Leser nachvollziehbar ist.

Treten Widersprüche in den Ergebnissen auf, dann soll erklärt werden, warum es diese gibt und welche Forschungsschritte notwendig wären, um die Widersprüche aufzuklären. Die Schlussfolgerungen sind in Bezug zur Literatur zu setzen (Daly et al. 2007).

Reflektiert die Studie vorhandene ethische Aspekte?

Der Einfluss, den die Forscherin / der Forscher auf den jeweiligen Forschungsprozess und das Ergebnis haben, ist zu reflektieren und explizit darzulegen. Das sind z. B. Machtverhältnisse zwischen Forscherinnen/Forschern und Studienteilnehmerinnen/-teilnehmern, die die Datensammlung betreffen (wenn z. B. die Forscherin / der Forscher eine Ärztin / ein Arzt oder eine Lehrerin / ein Lehrer ist und die Studienteilnehmer/innen Patientinnen/Patienten oder Schüler/innen sind) (Kuper et al. 2008, Kitto et al. 2008).

6.1.3 Diagnostische Studien

Für die medizinische Abklärung von Erkrankungen dienen den Ärztinnen/Ärzten verschiedenste Untersuchungsverfahren (in der Folge kurz Tests genannt). Das Ziel dieser Tests ist, die Unsicherheit in der diagnostischen bzw. prognostischen Abschätzung zu reduzieren (Felder-Puig et al. 2009). Studien, welche die Leistungsfähigkeit dieser Tests untersuchen, helfen Ärztinnen/Ärzten dabei, diese Tests sinnvoll einzusetzen. Studien versuchen Fragen zu beantworten wie „Mit welcher Treffsicherheit werden pathologische Zustände durch den Test erkannt?“ oder „Mit welcher Wahrscheinlichkeit ist meine Patientin / mein Patient bei einem positiven Testergebnis tatsächlich krank bzw. bei einem negativen Testergebnis tatsächlich gesund?“ (Whiting et al. 2003, Felder-Puig et al. 2009). Leider sind diagnostische Studien oft von sehr geringer Qualität (Bossuyt et al. 2003a). Eine kritische Evaluierung ist daher entscheidend, um die Aussagekraft verschiedener Studien zu bewerten.

Nachstehend präsentieren wir eine Checkliste für die kritische Evaluierung diagnostischer Studien, basierend auf dem QUADAS-Instrument (Quality Assessment of Studies of Diagnostic Accuracy included in Systematic Reviews; Whiting et al. 2003, Whiting et al. 2006).

Kriterien zur Beurteilung diagnostischer Studien

Die QUADAS-Checkliste besteht aus 14 Fragen, die mit „Ja“, „Nein“ oder „Unklar“ zu beantworten sind (Whiting et al. 2003, Whiting et al. 2006). Dieses Tool wird dafür verwendet, um möglichen Bias und die Richtung des Bias in diagnostischen Studien

aufzudecken. Wie bei allen anderen Instrumenten zur Evaluierung der internen Validität sollen auch bei der QUADAS–Liste keine Punkte vergeben werden.

Repräsentiert die Studie jene Patientinnen und Patienten, die den Test in der Praxis erhalten werden?

Die Population einer diagnostischen Studie beeinflusst die Ergebnisse der Studie. Die Sensitivität und Spezifität diagnostischer Studien ist abhängig von der Prävalenz des zu testenden Parameters. Daten von diagnostischen Tests, die bei Patientinnen/Patienten mit hoher Prävalenz einer Erkrankung erhoben wurden, sind nicht unbedingt auf solche mit niedriger Prävalenz übertragbar (Felder–Puig et al. 2009). Zudem sollen diagnostische Tests bei aufeinanderfolgenden Patientinnen und Patienten durchgeführt werden, da eine Selektion der Patienten seitens der Studienautorinnen/–autoren Bias verursacht. Darüber hinaus müssen Reviewer/innen entscheiden, ob die Ergebnisse aus der untersuchten Population im Hinblick auf die im Review gestellten Fragen generalisierbar sind (Whiting et al. 2003).

Sind die Auswahlkriterien für Patientinnen und Patienten ausführlich definiert?

Die Eigenschaften der Studienpopulation üben großen Einfluss auf die Ergebnisse einer Studie aus. Daher wird empfohlen, eine ausführliche Definition und Beschreibung der Auswahlkriterien einer Studienpopulation vorzunehmen (Bossuyt et al. 2003b).

Wurde ein Referenztest verwendet, der als „Goldstandard“ angesehen werden kann?

In diagnostischen Studien wird ein neuer Test („Indextest“) mit dem besten verfügbaren Referenztest (dem sogenannten „Goldstandard“) verglichen. Wichtig ist, dass dieser Referenztest eine hohe Sensitivität und Spezifität aufweist, das heißt, dass Patientinnen/Patienten, die an einer Erkrankung leiden, auch wirklich als krank, bzw. dass gesunde Personen auch wirklich als gesund erkannt werden (Whiting et al. 2003). Referenztests können Laboruntersuchungen, radiologische Untersuchungen oder langfristige Follow–ups sein (Knottnerus et al. 2002). Diagnostische Studien sind nur anwendbar und interpretierbar, wenn das Ergebnis mit einem validierten Referenztest verglichen wird.

Ist das Zeitintervall zwischen Durchführung des Indextests und Durchführung des Referenztests der Erkrankung entsprechend kurz?

Idealerweise werden Index– und Referenztests gleichzeitig durchgeführt. Sollte dies nicht der Fall sein, könnte eine falsche Klassifizierung der Erkrankung die Folge sein und ein „Disease Progression Bias“ in der Studie entstehen (bis an Patientinnen oder

Patienten der Indextest durchgeführt wird, könnte deren Erkrankung bereits einen anderen Schweregrad erreicht haben) (Whiting et al. 2003). Die Auswirkungen, die unterschiedliche Zeitpunkte der Durchführung von Referenz- und Indextest auf das Ergebnis haben können, hängen vom klinischen Verlauf einer Erkrankung ab (Felder-Puig et al. 2009). Im Gegensatz zu einer chronischen Erkrankung kann sich der Schweregrad einer infektiösen Erkrankung sehr schnell verändern. Wenn klinische Follow-ups als Referenztest verwendet werden, kann das „Disease Progression Bias“ auch für Studien über chronische Krankheiten ein Problem darstellen (Whiting et al. 2003).

Wurde der Referenztest an der gesamten Studienpopulation ausgeführt?

Im besten Fall wird der Referenztest an der gesamten Studienpopulation durchgeführt. Wenn dies nicht möglich ist (z. B. bei großen Studien), sollte eine Teilpopulation, an der der Referenztest durchgeführt wird, per Randomisierung ermittelt werden. Mögliche Unterschiede bei Risikofaktoren und prognostischen Faktoren zwischen Gesamt- und Teilpopulation müssen dann sorgfältig beschrieben werden. Wenn die Ergebnisse des Indextests die Entscheidung beeinflussen, wer einen Referenztest erhält, ist mit Bias zu rechnen (Bossuyt et al. 2003b).

Wurden alle Patientinnen und Patienten demselben Referenztest unterzogen, unabhängig vom Resultat des Indextests?

Die Auswahl des Referenztests muss unabhängig vom Resultat des Indextests erfolgen. Erhielten beispielsweise Patientinnen und Patienten mit positivem Resultat im Indextest einen anderen Referenztest als Patientinnen und Patienten mit negativem Resultat, würde dies zu Bias führen. Die Teilpopulationen für die beiden Referenztests würden sich dann unterscheiden (Bossuyt et al. 2003b).

Wurde ausgeschlossen, dass der Indextest bereits Bestandteil des gewählten Referenztests ist?

Manchmal setzt sich der Referenztest aus mehreren Untersuchungsparametern zusammen. Wenn der Indextest bereits Bestandteil des Referenztests ist und dies nicht berücksichtigt wird, sind die Ergebnisse der Studie fälschlicherweise optimistisch. Beispiel: Der Referenztest kann mehrere radiologische, klinische und labortechnische Untersuchungen beinhalten, als Indextest wird eine dieser Untersuchungen gewählt (Whiting et al. 2003).

Wurde die Durchführung von Index- und Referenztest so detailgenau beschrieben, dass diese leicht zu reproduzieren wären?

Eine Studie sollte eine genaue Beschreibung von Index- und Referenztest beinhalten, um deren Einsatz in anderen Settings zu ermöglichen. Ebenso erlaubt diese Beschreibung den Vergleich mehrerer Studien derselben Verfahren (Bossuyt et al. 2003b).

Wurden Indextestergebnisse ohne Kenntnis der Ergebnisse des Referenzverfahrens interpretiert? Wurden die Ergebnisse des Referenzverfahrens ohne Kenntnis der Testergebnisse interpretiert?

Wie bei der „Verblindung“ in der Bewertung der RCTs ist es ausschlaggebend, dass die beiden Tests unabhängig voneinander und ohne Kenntnis der jeweiligen Resultate des anderen Verfahrens interpretiert und bewertet werden (Whiting et al. 2003).

Waren für die Interpretation des Testergebnisses dieselben klinischen Daten verfügbar, die auch vorliegen würden, wenn der Test in der klinischen Praxis durchgeführt wird?

Demografische und medizinische Faktoren wie Alter, Geschlecht und Symptome können bei der Interpretation eines Tests großen Einfluss haben. Deshalb ist es wichtig zu erfahren, welche demografischen und medizinischen Faktoren dem Begutachter während der Auswertung der Tests bewusst waren (Whiting et al. 2003).

Wurde der Anteil an nicht interpretierbaren Resultaten berichtet?

Die meisten Tests werden zeitweise Resultate liefern, die nicht interpretiert werden können. In vielen Fällen werden nicht interpretierbare Tests bei den statistischen Analysen nicht berücksichtigt (Whiting et al. 2003). Eine Information über solche Resultate ist jedoch von großer Bedeutung für Reviewer/innen, um eine Studie richtig interpretieren zu können. Es besteht zurzeit kein Konsens zur Frage, ab welchem Prozentsatz die Rate an nicht interpretierbaren Resultaten die Validität der Ergebnisse negativ beeinflusst (Felder-Puig et al. 2009).

Wurde die Höhe der Drop-out-Rate während der Tests genannt?

Eine zu hohe Drop-out-Rate kann zu Selektionsbias führen. Es gibt keine einheitlichen Empfehlungen, wie hoch die Drop-out-Rate sein darf (Felder-Puig et al. 2009).

6.1.4 Präventionsstudien

Prävention ist neben der kurativen Medizin eine der wesentlichen Säulen eines modernen Gesundheitssystems. Unter Prävention (= Primärprävention) versteht man in der Regel eine Intervention, die darauf abzielt, in einer bestimmten Risikopopulation eine bestimmte Erkrankung zu vermeiden oder durch frühzeitiges Erkennen einer Erkrankung eine wirksamere, frühere Behandlung zu ermöglichen. Früherkennungsuntersuchungen an gesunden, symptomfreien Personen werden als Screening bezeichnet. Präventionsstudien können ähnlich wie Interventionsstudien evaluiert werden. Die „Intervention“ in diesem Fall ist die Präventionsmaßnahme, die zum Ziel hat, eine bestimmte Erkrankung zu vermeiden. Wie therapeutische Studien erfordern auch Präventionsstudien eine sehr genaue Abschätzung der Vor- und Nachteile der Maßnahmen, um zu gewährleisten, dass mehr Nutzen als Schaden (Netto-Nutzen) erzielt werden kann.

Präventionsstudien werden nach denselben Designs durchgeführt wie therapeutische Studien. Um Bias und Confounding zu minimieren, gilt der RCT als das beste Studiendesign. Die gemessenen Endpunkte beziehen sich auf die zu vermeidende Erkrankung oder auf Risikofaktoren, die damit in Zusammenhang stehen. Bei Präventionsmaßnahmen für Erkrankungen, die selten auftreten, werden auch kontrollierte Beobachtungsstudien eingesetzt. Die Vor- und Nachteile dieser Studiendesigns entsprechen denen von therapeutischen Studien. Die Evaluierung des Bias-Risikos für Präventionsstudien sollte sich daher an Schematas der einzelnen Studiendesigns in Kapitel 6.1.1.1, 6.1.1.2 und 6.1.1.3 orientieren. Allerdings ist insbesondere mit der Verblindung anders umzugehen, insbesondere wenn es um die Beurteilung von Präventionsmaßnahmen (z. B. Beratungsgespräche) geht, wo eine Verblindung nicht möglich ist. In solchen Fällen sind die Checklisten zu adaptieren.

Cluster-RCTs

Bei gesundheitsfördernden Maßnahmen ist die Einheit der Randomisierung manchmal nicht die einzelne Person, sondern eine Gruppe von Personen. Diese sogenannten Cluster-RCTs randomisieren jene organisatorische Einheit (z. B. Schulen, Gemeinden, Spitäler etc.), an der präventive oder gesundheitsfördernde Maßnahmen durchgeführt werden. Idealerweise sollten Endpunkte dann nicht in Bezug auf einzelne Personen ausgewertet werden, sondern in Bezug auf jene Einheit, die zur Randomisierung herangezogen wurde. Mit anderen Worten: Die Einheit der Randomisierung sollte auch die Einheit der Analyse sein (Donner/Klar 2004).

Bei Cluster-RCTs sollten zusätzlich zu den allgemeinen Punkten bei der Beurteilung von RCTs (Kapitel 4.1.1.1) noch folgende Kriterien einbezogen werden (Higgins/Green 2008):

Könnte Rekrutierungs-Bias eine Rolle spielen?

Rekrutierungs-Bias ist eine Form von Selektionsbias, das entstehen kann, wenn Individuen auch nach der Randomisierung noch in einzelne Cluster aufgenommen werden. Es kann dann durch selektive Rekrutierung zu einer Ungleichverteilung von prognostischen Faktoren kommen (z. B. könnten Personen mit geringem Risiko in die Interventions-Cluster nachrekrutiert werden).

Sind die Eigenschaften der Cluster nach der Randomisierung ähnlich?

Die Anzahl der Cluster, die randomisiert werden, ist manchmal gering, daher besteht die Möglichkeit, dass es zu Ungleichheiten bei der Verteilung von prognostischen Faktoren kommt. Dieses Problem kann zum Beispiel durch stratifizierte Cluster-Randomisierung gelöst werden.

Gab es Drop-out von ganzen Clustern?

Wie bei hohen Drop-out-Raten von RCTs, die Individuen randomisieren, kann Drop-out auch bei Cluster-RCTs zu Selektionsbias führen.

Wurde der Cluster-RCT statistisch korrekt ausgewertet?

Cluster-Effekte (z. B.: Personen innerhalb einer Gruppe sind meist ähnlicher als Personen zwischen zwei Gruppen) müssen bei der statistischen Analyse einbezogen werden. Wenn Cluster-Effekte nicht einbezogen werden, kann dies zu Ergebnissen führen, die zu präzise sind. In anderen Worten: Die Konfidenzintervalle sind zu eng, um vorhandene Unsicherheiten adäquat darzustellen.

Wie bei therapeutischen Studien ist auch bei Präventionsstudien die Übertragbarkeit der Ergebnisse ein kritischer Faktor, der immer in Bezug auf die eigene Population bewertet werden muss. Der Netto-Nutzen einer Präventionsmaßnahme kann sich mit veränderndem Basisrisiko auch ändern. Dies bedeutet, dass Präventionsmaßnahmen, die in einer Population mit hohem Risiko effektiv waren, nicht unbedingt auch in Populationen mit geringerem Risiko effektiv sein müssen. Zum Beispiel ist durch Studien gut belegt, dass niedrig dosiertes Aspirin bei Patientinnen und Patienten mit erhöhtem kardiovaskulärem Risiko die Rate von kardiovaskulären Ereignissen senken kann. Die Einnahme von Aspirin führt zwar zu einem vermehrten Auftreten von gastrointestinalen Blutungen, der Benefit ist aber deutlich größer als das zusätzliche Risiko. Dieser Netto-Nutzen ist aber in einer Population mit niedrigem kardiovaskulärem Risiko umstritten. Das Risiko für gastrointestinale Blutungen bleibt gleich, die Reduktion des absoluten Risikos für kardiovaskuläre Ereignisse ist in einer solchen

Population aber geringer. Die Bilanz zwischen Nutzen und Schaden von präventiven Maßnahmen kann sich daher abhängig vom Basisrisiko einer Population ändern.

6.2 Beurteilung der externen Validität

Das übergeordnete Ziel eines HTA ist es, Lesern dabei zu helfen, informierte Entscheidungen in Bezug auf die beste verfügbare Evidenz zu treffen. HTA-Berichte müssen daher immer sowohl das Bias-Risiko der eingeschlossenen Studien als auch Faktoren, die die Übertragbarkeit der Ergebnisse beeinflussen, evaluieren (Gartlehner et al. 2009).

Um klinischen Nutzen zu erhalten, muss neben der internen Validität einer Studie auch die Relevanz der Ergebnisse für spezifische Populationen in spezifischen Settings gegeben sein (Rothwell 2005).

Der Begriff „externe Validität“ bezieht sich auf die Anwendbarkeit von Studienergebnissen auf Patientinnen/Patienten außerhalb der Studienpopulation. Die externe Validität hängt daher in erster Linie von der Population und vom Gesundheitssystem einer Studie ab und ist somit eine subjektive Beurteilung (Rothwell 2009).

Die externe Validität hängt jedoch nicht nur von der Ähnlichkeit der Studienpopulation mit einer externen Population ab, sondern beispielsweise auch von zusätzlichen Faktoren und Aspekten des Studiendesigns, der Durchführung der Studien und von den Nebenwirkungen (Rothwell 2005, Green/Glasgow 2006). Voraussetzung bei der Beurteilung der externen Validität ist immer, dass eine ausreichende interne Validität gegeben ist. Studien mit hohem Bias-Risiko führen unabhängig von externer Validität zu verzerrten Ergebnissen und müssen daher mit Vorsicht interpretiert werden.

Bei der Beurteilung der externen Validität gilt es zwei wesentliche Aspekte zu berücksichtigen:

- » Sind Populationen und Gesundheitssysteme, in denen eine Studie durchgeführt wurde, auf jene Populationen und Gesundheitssysteme übertragbar, die für das HTA von Interesse sind?
- » Kann das Studiendesign als pragmatisch eingestuft werden? (siehe unten)

6.2.1 Pragmatische Studien

Studien mit guter externer Validität werden in der Literatur häufig als „pragmatische Studien“ („pragmatic studies“ oder „effectiveness studies“ im Gegensatz zu explanato-

rischen oder „efficacy studies“) bezeichnet (Godwin et al. 2003, Brook/Lohr 1985, Helms 2002, Glasgow et al. 2005, Zwarenstein et al. 2008).

Der Ausdruck „pragmatische Studien“ wurde im Jahr 1967 geprägt (Schwartz/Lellouch 1967). Die Autoren unterschieden bei RCTs zwischen pragmatischen und explanatorischen Studien. Die Termini „pragmatisch“ und „explanatorisch“ sind aber nicht nur auf RCTs beschränkt, sondern können auch auf Beobachtungsstudien angewandt werden. Pragmatische Studien haben aufgrund der wenig selektierten Patienten und Patientinnen und des natürlichen Settings hohe externe Validität, während explanatorische Studien hochselektierte Teilnehmer/innen einschließen und mitunter nur sehr beschränkt für die Normalbevölkerung anwendbar sind. Tabelle 6.2 fasst die wesentlichsten Unterschiede zwischen pragmatischen und explanatorischen Studien zusammen.

Tabelle 6.2:
Unterschiede zwischen explanatorischen und pragmatischen Studiendesigns

Studiendesign	Explanatorisch (Efficacy)	Pragmatisch (Effectiveness)
Frage	Ist die Intervention unter Idealbedingungen wirksam?	Wirkt die Intervention, wenn sie unter „Alltagsbedingungen“ angewandt wird?
Setting	Ideales Setting	„Alltagsbedingungen“
Studienteilnehmer/innen	Hochselektiert	Wenig oder keine Selektion
Intervention	Die Durchführung der Intervention wird forciert und die Einnahme streng überwacht, um die Adherence-Rate zu erhöhen	Flexible Durchführung der Intervention, wie unter „Alltagsbedingungen“ Variable Adhärenz wird in Kauf genommen
Endpunkte	Nicht immer patientenrelevante Endpunkte, oft Surrogat-Endpunkte	Patientenrelevante Endpunkte
Praxisrelevanz	Indirekt: Das Studiendesign reflektiert die tatsächlichen Anforderungen für Entscheidungen im klinischen Alltag nicht	Direkt: Das Studiendesign ist so gestaltet, dass von Ärzten und anderen Entscheidungsträgern im Gesundheitswesen praxisnahe Entscheidungen getroffen werden können.

Quelle: modifiziert nach Zwarenstein et al. 2008; GÖG/BIQG-eigene Darstellung

Die nachfolgenden Kriterien zur Identifikation von pragmatischen Studien wurden vom Evidence-based Practice Center des Research Triangle Institute an der University of North Carolina (RTI-UNC) definiert (Gartlehner et al. 2006). Es sollte jedoch angemerkt werden, dass es selten typisch pragmatische oder typisch explanatorische Studien gibt, vielmehr sind die Grenzen oft fließend.

Handelt es sich bei der Studienpopulation um eine Population in der Primärversorgung?

Die meisten klinischen Studien werden an Universitätskliniken durchgeführt. Die Bedingungen dort sind jedoch mit Gegebenheiten in der Primärversorgung nicht vergleichbar. Patientinnen und Patienten in universitären Einrichtungen sind meist stark selektiert, das behandelnde Personal ist sehr spezialisiert, und die eingesetzten medizinischen Geräte sind hochmodern. Die Generalisierbarkeit von Resultaten ist daher häufig limitiert. In seltenen Fällen dienen Universitätskliniken aber auch der Primärversorgung (Transplantationen).

Waren die Auswahlkriterien der Studienpopulation wenig restriktiv?

Die Teilnehmer/innen klinischer Studien sind oft sehr ausgesucht. Die Auswahl kann durch strenge Ausschlusskriterien (z. B. Ausschluss wegen Multimorbidität) erfolgen, aber auch durch spezielle Studiendesigns wie z. B. Placebo- oder Treatment-Run-in-Perioden, die dazu dienen, bereits vor tatsächlichem Beginn der klinischen Studie jene Patientinnen und Patienten auszuschließen, die entweder auf Placebo oder auf die zu testende Medikation in unerwünschter Form und Weise reagieren. Die Anwendbarkeit der Ergebnisse auf durchschnittliche Patientinnen/Patienten wird dadurch reduziert. Pragmatische Studien haben weniger strenge Auswahlkriterien, um repräsentative Patientenpopulationen zu erhalten.

Wurden patientenrelevante Endpunkte (health outcomes) untersucht?

Patientenrelevante Endpunkte sind solche, die Patientinnen/Patienten selbst erfahren oder erspüren können, wie etwa ein Rezidiv oder die Remission einer Erkrankung. Schmerzen oder Lebensqualität stellen ebenfalls typische patientenrelevante Endpunkte dar. Die Mortalität ist der am besten definierte patientenrelevante Endpunkt. Im Gegensatz dazu sind Surrogatparameter, die auch häufig als Endpunkte herangezogen werden, oft nicht patientenrelevant. Sie sind meist biochemische Marker und werden als Ersatz für patientenrelevante klinische Endpunkte herangezogen (Biomarkers Definitions Working Group 2001). Bei nachgewiesenem kausalem Zusammenhang durch einen RCT kann jedoch auch ein Surrogatparameter als Studien-Endpunkt akzeptiert werden (z. B. HbA1c, HDL, Blutdruck).

Waren Studiendauer und Behandlungsmodalitäten klinisch relevant?

Studiendauer und Behandlungsmodalitäten sollten den tatsächlichen Therapiebedingungen entsprechen. Bei Medikamentenstudien entspricht es beispielsweise eher der realen Situation, wenn Dosierungen flexibel und nicht fixiert sind. Auch die Dauer der Therapie sollte mit der Länge der tatsächlichen Anwendung vergleichbar sein.

Wurden Nebenwirkungen der Intervention adäquat erhoben?

Eine korrekte Erhebung der Nebenwirkungen beinhaltet, diese im Vorhinein zu definieren und mithilfe eines validen Instrumentes zu eruieren (z. B. WHO-Klassifikation). Von besonderer Bedeutung ist eine ausreichend lange Studiendauer, da sonst seltene Nebenwirkungen leichter übersehen werden können.

Ist die Studiengröße adäquat, um einen minimal-wesentlichen Unterschied aus Patientenperspektive erheben zu können?

Die Größe einer Studie und deren statistische Signifikanz stehen in engem Zusammenhang. Statistische Signifikanz ist jedoch nicht gleichbedeutend mit klinischer Relevanz. Meist sind Veränderungen, die statistisch signifikant sind, aber von Patientinnen/Patienten nicht wahrgenommen werden können, auch klinisch nicht relevant. Der „minimal-wesentliche Unterschied“ („the minimal clinically important difference“), also der kleinste Unterschied auf einer Skala, der von Patientinnen/Patienten als Verbesserung ihres Zustandes erlebt werden kann (Jaeschke et al. 1989), sollte in pragmatischen Studien bei Berechnungen der Studiengröße unbedingt berücksichtigt werden.

Ist eine Intention-to-treat-(ITT)-Analyse durchgeführt worden?

Neben der Gewährleistung der Randomisierung reflektiert eine ITT-Analyse die realen Verhältnisse einer Intervention an der Patientin / am Patienten, das heißt, es werden auch die Daten jener mit einberechnet, die nicht bis zum Schluss an der Studie teilgenommen haben. Unvorhergesehene Probleme wie Verstöße gegen das Studienprotokoll, falsche Diagnosestellung oder fehlende Adherence können durch eine ITT-Analyse in das Studienergebnis mit einbezogen werden.

6.3 Datenextraktion

Eine valide Synthese der Evidenz setzt voraus, dass Daten der eingeschlossenen Studien korrekt interpretiert und wiedergegeben werden. Für dieses Kapitel definieren wir „Daten“ als jegliche Information, die Studien entnommen wird; dies bedeutet, dass auch Publikationsjahr, Information über Studienteilnehmer/innen, Information über Sponsoren und Methoden etc. als „Daten“ bezeichnet werden.

Es muss schon am Beginn des HTA geplant werden, welche Daten wesentlich sind und im Rahmen der Studie zusammengefasst werden sollen. Das Übertragen von Information aus einzelnen Studien in Formulare und Tabellen eines HTA wird als Datenextraktion bezeichnet. Datenextraktion stellt einen wichtigen, aber extrem fehleranfälligen

Prozess dar. Methodikstudien haben gezeigt, dass Daten, die von nur einer Person extrahiert wurden, eine um 21 Prozent höhere Fehlerrate aufwiesen als Daten, die von zwei Personen extrahiert wurden (Buscemi et al. 2006). Datenextraktion sollte daher immer durch zwei Personen unabhängig voneinander erfolgen (Horton et al. 2009). Dies kann durch Parallel-Extraktion erfolgen, das heißt, beide Reviewer/innen extrahieren alle Daten einer Studie und vergleichen die Ergebnisse. Eine effizientere, aber ungenauere Methode ist, extrahierte Daten der ersten Reviewerin / des ersten Reviewers von einer zweiten Reviewerin / einem zweiten Reviewer kontrollieren zu lassen. Auf jeden Fall müssen unterschiedliche Ergebnisse immer diskutiert und per Konsens gelöst werden. Es gibt derzeit jedoch keinen allgemein anerkannten Standard für Datenextraktion.

Die extrahierte Information wird zumeist in Tabellen (sogenannten Evidenztabelle) dargestellt. Evidenztabelle sollen allgemeine Informationen zu Studien kompakt darstellen und jene Daten präsentieren, die für die Fragestellung relevant sind. Evidenztabelle können sich daher von Projekt zu Projekt unterscheiden. Evidenztabelle erfüllen zwei wesentliche Aufgaben:

- » Sie bieten eine bündige Zusammenfassung einzelner Studien und können beim Schreiben des Berichts verwendet werden.
- » Leser/innen finden in den Evidenztabelle Details zu einzelnen Studien.

Der Inhalt, der in Evidenztabelle zusammengefasst wird, ändert sich von Projekt zu Projekt. Folgende Informationen sind bei Primärstudien jedoch meistens enthalten:

- » Studienidentifikationsnummer,
- » Autoren und Studienjahr,
- » Studiendesign,
- » Studiendauer,
- » Studiengröße,
- » Methoden,
- » Auswahlkriterien der Population,
- » Charakteristika der Studienpopulation,
- » Intervention/Exposition,
- » Endpunkte,
- » Resultate,
- » Nebenwirkungen,
- » Bias-Risiko,
- » Sponsoren,
- » Kommentare.

Sekundärstudien sollten folgende Informationen enthalten:

- » Studienidentifikationsnummer,
- » Autoren und Studienjahr,

- » Studiendesign,
- » Literatursuche,
- » Selektionskriterien,
- » Intervention/Exposition,
- » Endpunkte,
- » Resultate,
- » Nebenwirkungen,
- » Bias-Risiko,
- » Sponsoren,
- » Kommentare.

Um die Konsistenz der Evidenztabelle zu gewährleisten, hat es sich bewährt, detaillierte Anleitungen und Definitionen zu erstellen, welche Daten in welcher Form extrahiert werden müssen. Dies verhindert, dass Evidenztabelle unterschiedlich werden und später mit viel Aufwand korrigiert und editiert werden müssen. Es hat sich außerdem bewährt, Evidenztabelle in einer Pilotphase zu testen, um Unterschiede in der Art und Weise der Datenextraktion zwischen einzelnen Reviewerinnen und Reviewern zu minimieren.

Publizierte Studien beinhalten nicht immer alle Daten, die für einen HTA-Bericht von Interesse sind. Fehlende Daten sind immer problematisch und können zu Bias führen (Burdett et al. 2003). Manche Daten können errechnet oder von anderen Daten abgeleitet werden (z. B. Drop-out-Raten). Häufig ist dies jedoch nicht möglich. In einem solchen Fall sollten die Studienautorinnen/-autoren kontaktiert werden. Daten sollen keinesfalls in Abbildungen (z. B. Diagrammen) „geschätzt“ werden.

Autorinnen/Autoren: Gerald Gartlehner, Ludwig Grillich, Angela Kaminski, Kylie Thaler (Donau-Universität Krems)

7 Synthese der medizinischen Evidenz

Ziel eines jeden HTA ist es, eine Grundlage für evidenzbasierte Entscheidungen im Gesundheitssystem zu liefern. Die a priori definierten HTA-Fragestellungen sollen durch eine Synthese der besten verfügbaren Evidenz beantwortet werden. Lücken und Unsicherheiten im Wissen müssen klar dargestellt werden.

Die Synthese der Evidenz kann prinzipiell qualitativ (beschreibend) oder quantitativ (z. B. durch Metaanalysen) durchgeführt werden. Die Möglichkeiten der quantitativen Synthese der Evidenz werden in Kapitel 7.3 beschrieben.

7.1 Qualitative Zusammenfassung

Nicht immer können Studien durch quantitative Methoden zusammengeführt werden, um ein gemeinsames Resultat (z. B. durch Metaanalyse) zu erzielen. Meist kann die vorhandene Evidenz nur deskriptiv zusammengefasst werden. Lesbarkeit und Objektivität stellen dabei wesentliche Eckpfeiler eines gut geschriebenen HTA dar und sind eine grundlegende Voraussetzung dafür, dass ein HTA für gesundheitspolitische Entscheidungen herangezogen wird.

Eine gute qualitative Synthese sollte eine klare Beantwortung der HTA-Fragestellungen bieten und keine „Katalogisierung“ der vorhandenen Studien sein. Bei HTA-Berichten, die Interventionen evaluieren, muss in einem Ausmaß Information beschrieben werden, dass das Verhältnis von Nutzen zu Risiko daraus beurteilt werden kann. Eine gute Synthese der Evidenz sollte daher Parameter beinhalten, die für die Beurteilung der internen und externen Validität essenziell sind. Folgende Bereiche sind dabei von Bedeutung:

- » interne Validität der Studien,
- » Studiengrößen,
- » Studiendauer,
- » relative und absolute Effektmaße mit Konfidenzintervallen (wenn vorhanden) und P-Werten (wenn vorhanden),
- » Darstellung der Konsistenz der Resultate der Studien,
- » Faktoren, die die externe Validität einschränken könnten,
- » andere Faktoren, die das Vertrauen in die vorhandene Evidenz beeinflussen könnten (z. B. Verdacht auf Publikationsbias, finanzielle Unterstützung der Studien etc.).

Sind viele Studien vorhanden, sollte der Fokus auf der besten Evidenz liegen. Es reicht dann meist, wenn weniger gute oder kleine Studien nur kurz zusammengefasst oder in

Tabellen dargestellt werden und ein Zusammenhang mit der besten vorhandenen Evidenz hergestellt wird.

Eine gute Möglichkeit zur Zusammenfassung und Darstellung der vorhandenen Evidenz bieten Evidenzprofile nach der GRADE (The Grading of Recommendations Assessment, Development and Evaluation) Working Group (Atkins et al. 2004a). GRADE-Profile stellen sowohl erwünschte als auch unerwünschte Effekte dar und setzen sie in Bezug mit der Stärke der vorhandenen Evidenz (vgl. 7.4). Bei wichtigen Endpunkten, für die keine Evidenz vorhanden ist, wird das Fehlen der Evidenz ebenfalls in den Evidenzprofilen dargestellt. Tabelle 7.1 stellt ein fiktives Beispiel eines Evidenzprofils dar.

Tabelle 7.1:
Beispiel eines Evidenzprofils

Anzahl der Studien (Patientinnen/Patienten)	Design	Bias-Risiko	Konsistenz der Ergebnisse	Direktheit	Größe des Effekts	Andere modifizierende Faktoren	Stärke der Evidenz
Endpunkt: Ansprechen auf die Therapie							
3 (272)	RCT	Niedrig	Ja	Ja	Behandlungseffekte im Vergleich zu Placebo sind ähnlich. RR: 1,5 (1,2-1,7)	Keine	Hoch
Endpunkt: Remission							
1 (98)	RCT	Unklar	N/A	Ja	Kein wesentlicher Unterschied zu Placebo. RR: 1,05 (0,95-1,15)	Keine	Niedrig
Endpunkt: Mortalität							
<i>Keine Evidenz</i>							
Endpunkt: Lebensqualität							
<i>Keine Evidenz</i>							
Endpunkt: Nebenwirkungen							
6 (468)	1 RCT, 5 Fallserien	Hoch	Ja	Ja	Kopfschmerzen bei 30-50 % aller Patientinnen und Patienten	Ja*	Sehr niedrig

* Möglicher Publikationsbias

Quelle: GRADE Working Group (Atkins et al. 2004a); GÖG/BIQG-eigene Darstellung

Jede Synthese der Evidenz muss in Bezug auf die Stärke der Evidenz erfolgen (Guyatt et al. 2006). Die Stärke der Evidenz reflektiert das Vertrauen, dass die vorhandene Evidenz die richtige Antwort gibt (Brozek et al. 2009).

7.2 Effektgrößen

Der Unterschied in Ergebnissen zwischen zwei unterschiedlich behandelten Gruppen wird Effekt, Behandlungseffekt oder Interventionseffekt genannt (Higgins/Green 2008). Die Maßzahl, um die Größe eines Effektes darzustellen, wird Effektmaß bezeichnet.

Gemessene Daten in Studien können grundsätzlich in solche mit diskreten und solche mit stetigen (kontinuierlichen) Merkmalen unterteilt werden. Merkmale sind diskret, wenn sie abzählbar viele Werte annehmen können (z. B. Blutgruppe, Anzahl der Schwangerschaften). Stetige Merkmale können alle Werte innerhalb eines definierten Intervalls annehmen (z. B. Gewicht, Blutdruck).

Diskrete und stetige Merkmale können auf unterschiedlichen Skalen gemessen werden und werden in Skalenniveaus eingeteilt.

1. Nominalskala

Eine Nominalskala entspricht dem niedrigsten Skalenniveau. Die Ausprägungen auf einer Nominalskala unterscheiden sich nur begrifflich voneinander, haben aber keine unterschiedliche Wertigkeit (z. B. Blutgruppen). Eine spezielle Form auf einer Nominalskala stellen Alternativmerkmale (dichotome oder binäre Ausprägungen) dar (z. B. Geschlecht, Fieber – kein Fieber).

2. Ordinalskala

Die Ausprägungen der Merkmale auf einer Ordinalskala lassen sich nach einer Rangfolge anordnen (z. B. Schulnoten, Tumorstadien).

3. Metrische Skalen

Unter metrischen Skalen versteht man Intervall- und Verhältnisskalen. Beide bieten einen höheren Informationsgrad als Nominal- oder Ordinalskalen.

Bei Intervallskalen unterscheiden sich die Ausprägungen zahlenmäßig (z. B. Temperatur in Grad Celsius). Bei Intervallskalen können sowohl negative als auch positive Werte auftreten. Bei Verhältnisskalen gibt es einen Nullpunkt (z. B. Körpergewicht, Cholesterinspiegel).

Um Effekte vergleichbar zu machen oder um eine Synthese von Effekten aus unterschiedlichen Studien (in einer Metaanalyse) herzustellen, müssen diese auf standardi-

sierte Art definiert, gemessen und zusammengefasst werden. Im folgenden Abschnitt werden die wesentlichsten Effekt- und Risikomaße, die in medizinischen Studien verwendet werden, dargestellt und deren Interpretation diskutiert.

7.2.1 Dichotome Effektmaße

Dichotome Effektmaße werden verwendet, wenn ein Endpunkt nur zwei Möglichkeiten zulässt (z. B. Tod oder Überleben, Remission oder keine Remission).

Fünf Effektmaße werden häufig verwendet, um dichotome Endpunkte zu beschreiben:

1. relatives Risiko (RR; auch als Risk Ratio oder relativer Benefit bezeichnet),
2. relative Risikoreduktion (RRR; auch als relativer Effekt bezeichnet),
3. Odds Ratio (OR),
4. absolute Risikoreduktion (ARR; auch als Risiko-Differenz RD bezeichnet),
5. Number needed to treat (NNT) / Number needed to harm (NNH).

Meistens bieten publizierte Artikel genügend Informationen, um mithilfe einer Vierfeldertafel einzelne Effektmaße zu berechnen (Abbildung 7.2).

Tabelle 7.2:

Vierfeldertafel zur Berechnung von Effektmaßen für dichotome Zielvariablen

	<i>Intervention</i>	<i>Keine Intervention</i>	<i>Gesamt</i>
<i>Endpunkt vorhanden</i>	<i>a</i>	<i>b</i>	<i>a + b</i>
<i>Endpunkt nicht vorhanden</i>	<i>c</i>	<i>d</i>	<i>c + d</i>
<i>Gesamt</i>	<i>n der Interventionsgruppe</i>	<i>n der Kontrollgruppe</i>	<i>n der Studie</i>

Quelle und Darstellung: Donau-Universität Krems

1. Relatives Risiko (RR)

Das relative Risiko beschreibt, um wie viel höher die Wahrscheinlichkeit für jemanden mit Exposition (Intervention) ist, einen bestimmten Endpunkt zu erreichen, verglichen mit jemandem ohne Exposition. Das relative Risiko beschreibt auch die Stärke des Zusammenhangs von Exposition zu einem bestimmten Endpunkt. Wenn das Konfidenzintervall eines relativen Risikos 1 kreuzt, besteht kein statistisch signifikanter Unterschied.

» Formel: $[a / (a + c)] / [b / (b + d)]$

Folgende Interpretationen gelten für das relative Risiko:

- » RR = 1: kein Unterschied
- » RR = 2: Die Wahrscheinlichkeit für das Auftreten eines bestimmten Endpunkts ist mit Exposition doppelt so groß (100 %) wie ohne Exposition.
- » RR = 0,5: Die Wahrscheinlichkeit für das Auftreten eines bestimmten Endpunkts ist mit Exposition halb so groß (50 %) wie ohne Exposition.

2. Relative Risikoreduktion (RRR, relativer Effekt)

Die relative Risikoreduktion beschreibt, um welchen Prozentsatz eine Intervention das Risiko relativ reduziert.

- » Formel: $RR - 1$ (oder $1 - RR$)
- » RRR = 0,1 bedeutet, dass die Intervention zu einer relativen Verringerung des Risikos um 10 Prozent führt.

Ohne Kenntnis des Basisrisikos können relatives Risiko und relative Risikoreduktion jedoch keine Auskunft über das absolute Risiko geben. Dies birgt die Gefahr, dass bei der Interpretation der relativen Risikoreduktion der Effekt gelegentlich überschätzt wird. Jedes relative Effektmaß sollte daher immer in Zusammenhang mit einem absoluten Effektmaß dargestellt werden.

3. Odds Ratio (Chancenverhältnis; OR)

Odds Ratios wurden für Fall-Kontroll-Studien entwickelt und werden aufgrund von mathematischen Vorteilen in Studien häufig statt relativen Risiken verwendet. Odds Ratios bezeichnen das Verhältnis der Chance, dass ein Ereignis oder Endpunkt in der experimentellen Gruppe eintritt, zur Chance, dass ein Ereignis in der Kontrollgruppe eintritt. Eine Odds Ratio von 1 bedeutet keinen Unterschied zwischen den Gruppen.

Da Odds Ratios weniger intuitiv als relative Risiken sind, werden sie von den meisten Personen wie relative Risiken (d. h. als Verhältnis von Wahrscheinlichkeiten, einen Endpunkt zu erreichen) interpretiert (Altman et al. 1998, Sackett et al. 1996). Dies kann jedoch problematisch sein, wenn die zugrunde liegende Ereignisrate hoch ist. Bei niedriger Ereignisrate (unter 20 %) führen Odds Ratio und relatives Risiko annähernd zu einem gleichen Ergebnis. Bei zunehmender Ereignisrate divergieren Odds Ratio und relatives Risiko (Deeks 2002). Als Konsequenz hat eine Odds Ratio bei hoher Ereignisrate einen numerisch größeren Wert als das korrespondierende relative Risiko. Da beide meist identisch interpretiert werden, besteht die Gefahr, dass mit einer Odds Ratio der Effekt überschätzt wird.

Odds Ratios werden dazu verwendet, um zu beschreiben, wie stark das Verhältnis einer Intervention (Exposition) zu einem bestimmten Endpunkt ist. Wenn das Konfidenzintervall einer Odds Ratios 1 kreuzt, besteht kein statistisch signifikanter Unterschied.

» Formel: $(a/c) / (b/d)$

4. Absolute Risikoreduktion (ARR; Risiko-Differenz: RD)

Die absolute Risikoreduktion gibt an, um wie viel das Risiko durch eine Intervention reduziert werden kann. Die absolute Risikoreduktion hat gegenüber dem relativen Risiko und der Odds Ratio den Vorteil, dass sie auch kalkulierbar ist, wenn die Ereignisrate in einer der beiden Gruppen 0 ist. Wie auch bei der relativen Risikoreduktion ist die Interpretation der absoluten Risikoreduktion an das Basisrisiko gebunden.

» Formel: $[a / (a + b)] - [b / (b + d)]$

Die absolute Risikoreduktion ist die Grundlage zur Berechnung der Number needed to treat (NNT) und der Number needed to harm (NNH), und sie ist klinisch aussagekräftiger als relative Effektmaße.

» Interpretation: absolute Risikoreduktion = 0,1: Die Intervention führt zu einer absoluten Verringerung des Risikos um 10 Prozentpunkte (die absolute Risikoreduktion kann im Gegensatz zur relativen Risikoreduktion auch in Bezug auf die klinische Relevanz beurteilt werden).

5. Number needed to treat (NNT) / Number needed to harm (NNH)

Das relative Risiko und die Odds Ratio sind relative Effektmaße, die nur beschränkt Auskunft über die klinische Relevanz von Ergebnissen geben können. Die NNT und die NNH geben an, wie viele Personen behandelt werden müssen, um bei einer Person innerhalb eines bestimmten Zeitraumes einen bestimmten Endpunkt zu erzielen bzw. zu vermeiden.

- » Formel: $1/ARR$
- » Interpretation: $NNT = 10$ bedeutet, dass zehn Personen die Intervention erhalten müssen, damit eine Person innerhalb eines bestimmten Zeitraumes den Endpunkt erreicht (z. B. von der Intervention profitiert).

7.2.2 Stetige Effektmaße

Stetige Effektmaße werden für Daten verwendet, die auf metrischen Skalen gemessen wurden. Beispiele sind Gewicht, Größe oder medizinische Parameter wie Leukozytenzahl. Um stetige Endpunkte von verschiedenen Studien vergleichbar zu machen, werden bei Metaanalysen folgende zwei Parameter verwendet: die gewichtete mittlere Differenz („weighted mean difference“) und die standardisierte mittlere Differenz („standardized mean difference“).

1. Gewichtete mittlere Differenz („weighted mean difference“)

Die gewichtete mittlere Differenz ist ein Effektmaß für Endpunkte, die auf derselben Skala gemessen wurden. Sie wird zur Beschreibung eines Gesamteffektes bei Metaanalysen eingesetzt. Dabei erhalten Einzelstudien ein unterschiedliches Gewicht, um die Präzision der Ergebnisse der Einzelstudien zu berücksichtigen (Higgins/Green 2008).

2. Standardisierte mittlere Differenz („standardized mean difference“; SMD)

Wenn ein Endpunkt in verschiedenen Studien auf unterschiedlichen Skalen gemessen wurde, können die Ergebnisse mithilfe der standardisierten mittleren Differenz vergleichbar gemacht werden.

Diese wird berechnet, indem die gemessene Veränderung in jeder Studie durch die Standardabweichung dividiert wird. Dabei wird die Änderung (normalerweise eine Verbesserung) relativ zu der Variabilität in der Studie dargestellt. Beispiel: Die Ergebnisse der Patientinnen/Patienten aus der Interventionsgruppe haben sich um eine halbe Standardabweichung im Vergleich zu den Ergebnissen der Patientinnen/Patienten aus der Kontrollgruppe verbessert. Wenn derselbe Endpunkt von Interesse ist, können verschiedene Skalen kombiniert werden.

- » Formel: $SMD = \frac{\text{Unterschied der Mittelwerte der Gruppen}}{\text{Standardabweichung}}$

In Bezug auf die Größe einer klinisch relevanten Verbesserung kann man davon ausgehen, dass eine Verbesserung um 0,2 Standardabweichungen einen kleinen Effekt, eine Verbesserung um 0,5 Standardabweichungen einen moderaten Effekt und eine

Verbesserung um 0,8 Standardabweichungen oder mehr einen großen Effekt darstellt (Cohen 1988).

7.2.3 Inzidenzmaße

Die Inzidenz entspricht dem Risiko einer Population, einen bestimmten Endpunkt zu erfahren. Manche Endpunkte können jedoch bei derselben Person mehrmals auftreten (z. B. Nebenwirkungen wie Kopfschmerzen). Außerdem können Personen eine Studie abbrechen, bevor sie einen bestimmten Endpunkt erfahren. Um dieses Herausforderung zu überwinden, wird in Studien zur Darstellung von Nebenwirkungen oft die „Inzidenzdichte“ verwendet. Bei der Inzidenzdichte bezieht sich der Zähler auf die neu aufgetretenen Ereignisse, während sich der Nenner auf die Summe der Beobachtungszeiten aller Individuen bezieht (Personenzeit). Die Inzidenzdichte wird dementsprechend als Fälle pro Personenzeit angegeben (z. B. 50 Fälle pro 100 Patientenjahre).

7.2.4 Überlebenszeitanalysen

Überlebenszeitanalysen untersuchen die Zeit zwischen einem definierten Anfangs- und einem Endereignis. Überlebenszeitanalysen können sich auf den Tod als Endpunkt beziehen, können aber auch für jedes andere Ereignis, bei dem eine zeitliche Komponente wesentlich ist, durchgeführt werden (z. B. Zeit bis zur nächsten Hospitalisierung, Zeit bis zur nächsten Krankheitsepisode). Überlebenszeitdaten werden in medizinischen Publikationen oft in Form einer Kaplan–Meier–Überlebenskurve dargestellt (Lagakos 2006).

Statistisch werden diese Daten häufig mittels der Hazard Ratio ausgewertet, die ein Maß für das relative Risiko ist. Die Hazard Ratio beschreibt, wie wahrscheinlich es ist, dass eine Person in der Interventionsgruppe in einer bestimmten Zeit (der Studiendauer) den Endpunkt erlebt im Vergleich zu einer Person in der Kontrollgruppe. Die Hazard Ratio wird wie das relative Risiko interpretiert.

Autor/in: Gerald Gartlehner, Kylie Thaler (Donau–Universität Krems)

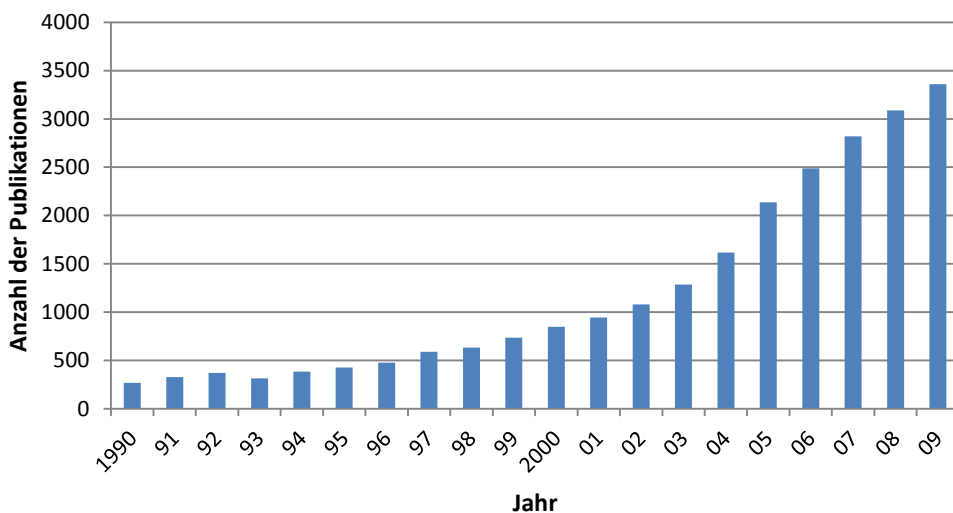
7.3 Quantitative Zusammenfassung

7.3.1 Metaanalysen

7.3.1.1 Einleitung

Die Zahl der Metaanalysen, die jährlich durchgeführt werden, wächst stetig. Abbildung 7.1 zeigt die Anzahl der veröffentlichten Metaanalysen im jeweiligen Jahr, die in MEDLINE/PubMed als Metaanalysen definiert wurden.

Abbildung 7.1:
Anzahl der Publikationen im Bereich Metaanalyse in MEDLINE/PubMed



Quelle und Darstellung: UMIT

Der Begriff „Metaanalyse“ wird für unterschiedliche Sachverhalte verwendet. So werden gelegentlich bereits systematische Übersichtsarbeiten ohne quantitative Zusammenfassung der Ergebnisse als Metaanalyse bezeichnet. In der Regel wird darunter jedoch eine quantitative Zusammenfassung von Effektmaßen mehrerer Studien zur gleichen Forschungsfrage mittels statistischer Methoden verstanden (Grouven 2010). In einem umfassenden Verständnis setzt diese Auswertung eine systematische Literaturübersicht voraus (Sutton, Abrams et al. 2000), bei enger gefasster Bedeutung wird darunter allein die quantitative Synthese verstanden, unabhängig vom Prozess der Identifikation der Studien (z. B. (Glass 1976)). Außerdem ist noch die quantitative Analyse von gepoolten Originaldaten auf individueller Patientenebene zu erwähnen (sog. „IPD“-

Metaanalyse), die ebenfalls eine Form der Metaanalyse darstellt. Im vorliegenden Handbuch wird entsprechend dem im gesundheitlichen Bereich vorherrschenden Allgemeinverständnis der Begriff Metaanalyse stets als das quantitative Zusammenfassen von Effektmaßen aus verschiedenen Studien mithilfe statistischer Methoden im Rahmen einer systematischen Übersichtsarbeit verstanden, was sowohl auf aggregierten Daten als auch auf individuellen Patientendaten basieren kann.

Auch rein narrative systematische Reviews haben das Ziel, die Ergebnisse der einzelnen Studien im Kontext aller relevanten Studien darzustellen, um die Bewertung der Effektivität einer Intervention zu beurteilen. Dem narrativen Review, der auf der getrennten Darstellung der Ergebnisse aus einer Reihe von Studien beruht, fehlt ein systematisches Verfahren, die Daten quantitativ zu synthetisieren. Stattdessen muss in einem narrativen Review eine Art „Arithmetik in Worten“ betrieben werden (Borenstein, Hedges et al. 2009). Die statistische Signifikanz in Form von p-Werten dient häufig dem Nachweis, dass ein mehr als zufälliger Effekt existiert oder nicht. In einem narrativen Review werden die p-Werte aus verschiedenen Studien i. d. R. nicht quantitativ zusammengeführt, sondern es wird häufig lediglich bewertet, in wie vielen Studien der Effekt statistisch signifikant ist („positive“ Studien) oder nicht („negative“ Studien). Dies kann leicht zu Fehleinschätzungen führen, insb. wenn viele kleine Studien mit geringer Power vorhanden sind, die alle das Signifikanzniveau verfehlen. Im Unterschied dazu kann eine Metaanalyse die Effektgrößen aus allen Studien und deren Streuung in einer übergreifenden statistischen Analyse untersuchen und Aussagen über die Präzision des Gesamtschätzers machen (Borenstein, Hedges et al. 2009)

Mit quantitativen Methoden wird

- » das Vorhandensein eines Behandlungs- oder Expositionseffektes getestet,
- » ein Gesamtmaß für die Assoziation oder die Effektgröße geschätzt,
- » der Effekt des Zufallsfehlers reduziert und die Präzision der Effektschätzung erhöht,
- » die Heterogenität der Studienergebnisse untersucht und ggf. erklärt,
- » die Sensitivität der Ergebnisse für eine Gefährdung der internen Validität etwa durch Publication Bias oder durch eine mangelhafte Studienqualität geprüft.

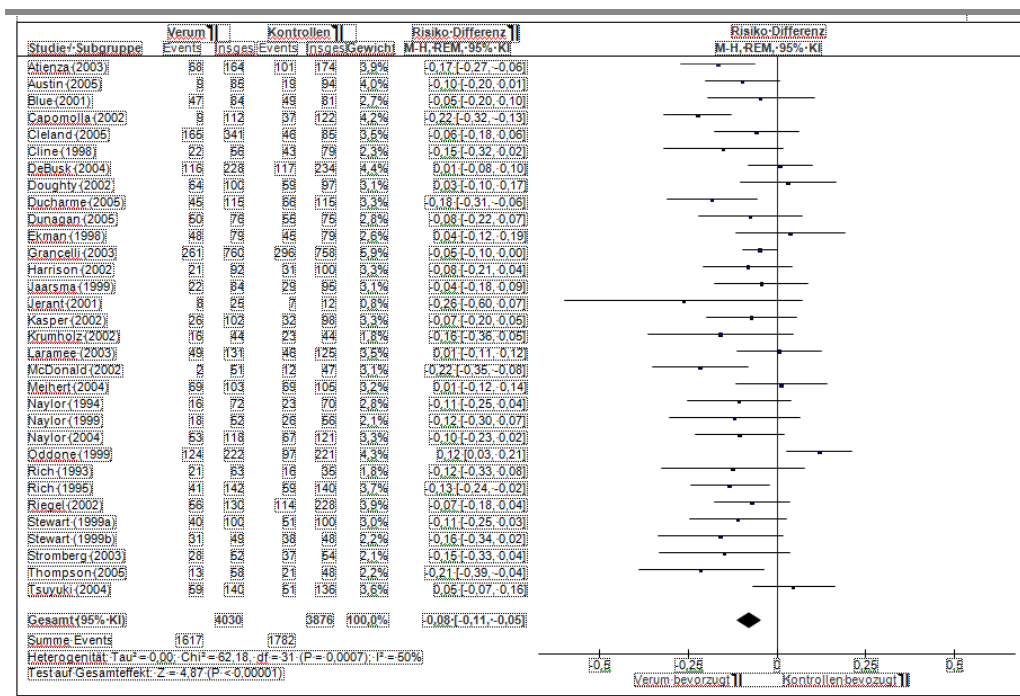
Grundsätzlich wird eine Metaanalyse mit aggregierten Daten in zwei Schritten vollzogen: Im ersten Schritt wird für jede Studie das Effektmaß ermittelt (für binäre bzw. dichotome Daten sind dies üblicherweise Risikodifferenz, relatives Risiko oder Odds Ratio; s. Kap. 7.3.1.4.1). Im zweiten Schritt wird ein gewichteter Mittelwert ermittelt.

Zuerst ist für die Beurteilung allerdings wichtig, ob die Effektgröße, also z. B. der Behandlungseffekt, in allen Studien konsistent ist. Wenn ja, soll die Effektgröße möglichst genau bestimmt und eine Aussage zu deren Robustheit über alle Studien hinweg gemacht werden. Wenn sich die Effektgröße hingegen von Studie zu Studie

substanziell unterscheidet, dann soll das Ausmaß dieser Varianz quantifiziert, und die daraus resultierenden Implikationen betrachten werden.

Im Folgenden werden anhand eines Beispiels in Anlehnung an (Borenstein, Hedges et al. 2009) die verschiedenen Elemente einer Metaanalyse kurz vorgestellt.

Abbildung 7.2:
Rehospitalisierungen bei Disease-Management-Programmen
im Vergleich zur Standardbehandlung



REM: „random effects“-Modell, M-H: Mantel-Haenszel-Schätzer, 95%-KI: 95%-Konfidenzintervall

Quelle und Darstellung: adaptiert nach Göhler et al. (2006)

Tabelle 7.4 stellt die Metaanalyse von (Göhler, Januzzi et al. 2006) grafisch dar. Sie untersucht die Effektivität von Disease-Management-Programmen (DMP) im Vergleich zur Standardbehandlung bei Herzinsuffizienz hinsichtlich der Häufigkeit erneuter Rehospitalisierungen. 32 Studien wurden in die Analyse einbezogen. In Tabelle 7.4 wird jede Studie durch eine Zeile repräsentiert. Die erste Spalte enthält die Bezeichnung der Studie (hier repräsentiert durch die Namen der Erstautoren), die zweite Spalte gibt die Anzahl der Ereignisse in der Interventionsgruppe an (hier DMP), die dritte die Anzahl der Patienten dieser Gruppe. In den Spalten 4 und 5 folgen die entsprechenden Zahlen für die Kontrollgruppe. Spalte 6 informiert über das relative Gewicht, das die Studie bei der Berechnung des Gesamtschätzers hat, und in Spalte 7 wird die Effektgröße mit dem 95%-Konfidenzintervall mitgeteilt. Daneben sind Effektschätzer und

Konfidenzintervall für jede Studie grafisch dargestellt. Das relative Gewicht wird durch die Größe des Punktes für die Effektgröße angegeben. Die senkrechte Linie gibt den Wert für die Effektgröße an, bei der Behandlung und Kontrollintervention gleich gut sind. Im vorliegenden Beispiel zeigen alle Werte, die links von der Linie liegen, dass die Behandlung, und alle rechts davon, dass die Kontrollintervention besser ist.

Die Effektgröße gibt die Größe des Behandlungseffekts bzw. ganz allgemein die Stärke der Assoziation zwischen zwei Variablen an. In der Metaanalyse wird für jede einzelne Studie die Effektgröße berechnet. Anhand der Effektgrößen wird die Konsistenz des Effekts über alle Studien hinweg bewertet und der gemeinsame Gesamteffekt berechnet (Borenstein, Hedges et al. 2009). Im vorliegenden Fall wurde die Risikodifferenz als Effektgröße verwendet. Hier repräsentiert eine Differenz von 0 keinen Effekt (sog. Nulleffekt), bei relativen Risiken oder Hazard Ratios repräsentiert der Wert 1 den Nulleffekt.

Die Präzision bzw. die statistische Unsicherheit, mit der die Effektgröße geschätzt wurde, wird durch das Konfidenzintervall angegeben. Beispielsweise hat die Studie von Riegel im oben genannten Beispiel ein engeres Konfidenzintervall als die Studie von Jerant (Jerant, Azari et al. 2001; Riegel, Carlson et al. 2002). Wenn das Konfidenzintervall vollständig auf einer Seite der Nulleffektlinie liegt, ohne diese zu überdecken, dann entspricht dies einem statistisch signifikanten Ergebnis auf dem entsprechenden Alpha-Niveau (0,05 bei 95%-Konfidenzintervall). Im vorliegenden Beispiel weisen sieben Studien ein statistisch signifikantes Ergebnis auf.

In der Metaanalyse ist festzulegen, nach welchen Kriterien **die Gewichtung** der einzelnen Studien durchgeführt wird. Die Gewichtung erfolgt über zum Teil unterschiedliche Verfahren, in die insbesondere die Präzision der einzelnen studienspezifischen Effektschätzer einfließt. Die Präzision der einzelnen Effektschätzer hängt von der Effektgröße, Fallzahl und der Streuung der Werte zwischen den Beobachtungseinheiten (in der Regel Studienteilnehmer) ab. Studien mit höherer Präzision erhalten ein höheres Gewicht. Die Präzision des Gesamtschätzers hängt außerdem ab von den Verteilungsannahmen über den wahren Effekt bei den angewendeten Gewichtungsverfahren (siehe Kap. 7.3.1.2.4).

Eines der Hauptziele der Metaanalyse ist die Berechnung eines gemeinsamen Effektschätzers für alle Studien, d. h. des gewichteten Mittelwerts der Effekte der Einzelstudien. Dieser Mittelwert wird in Abbildung 7.2 am unteren Ende der Studien als Raute dargestellt. Im Beispiel beträgt der Mittelwert der Risikodifferenz $-0,08$, was bedeutet, dass durch DMP das Risiko für eine erneute Einweisung in eine Krankenanstalt um acht absolute Prozentpunkte verringert werden kann. Dies entspricht einer Number needed to treat (NNT) für die Vermeidung eines Ereignisses von 12 bis 13 Patienten. Die Breite der Raute gibt das 95%-Konfidenzintervall des gepoolten Effektschätzers an. Im

Beispiel reicht das 95%-Konfidenzintervall für die gepoolte Risikodifferenz von $-0,11$ bis $-0,05$.

Im Beispiel variieren die Effektschätzer der einzelnen Studien stark. So gibt es einige Studien, in denen die Patienten mit Standardbehandlung weniger Rehospitalisierungen aufweisen als die mit DMP. Inwiefern die Streuung durch Zufallsfehler zustande kommt oder auf einer Heterogenität des wahren Effektes beruht, ist weiter zu analysieren.

7.3.1.2 Prinzipien der Metaanalyse

7.3.1.2.1 Gepoolter Schätzer

Ein wesentliches Ziel der Metaanalyse ist die Berechnung eines zusammenfassenden Punktschätzers der Größe des Effekts mit einer möglichst geringen statistischen Unsicherheit (= hohe Präzision, kleines Konfidenzintervall). Die damit verbundene Erhöhung der statistischen „Power“ bedeutet, dass die Wahrscheinlichkeit, einen real vorhandenen Effekt als statistisch signifikant zu erkennen, zunimmt. Dies spielt vor allem bei kleineren, aber klinisch relevanten Effekten eine Rolle. Oft werden Studien zur medizinischen Effektivität mit Stichproben durchgeführt, die zu klein sind, um statistisch signifikante Ergebnisse zu liefern. Von den 32 Studien in der Metaanalyse von Göhler (Tabelle 7.4) zeigen nur sieben einen statistisch signifikanten Effekt, der Gesamtschätzer aller Studien hingegen ist mit hoher Präzision statistisch signifikant (Göhler, Januzzi et al. 2006). Nur durch die quantitative Zusammenfassung wird dies deutlich. Das Zusammenfassen (Pooling) der Studien geschieht in einem zweistufigen Verfahren. Im ersten Schritt wird für jede Studie das Effektmaß ermittelt (für binäre bzw. dichotome Daten sind dies üblicherweise Risikodifferenz, relatives Risiko oder Odds Ratio). Im zweiten Schritt wird ein gewichteter Mittelwert ermittelt. Die Gewichte sollen dabei den Informationsgehalt der einzelnen Studien wiedergeben. Häufig wird für das Gewicht die Präzision (d. h. der Kehrwert der Varianz) der Effektgröße verwendet. Die Varianz hängt stark von der Stichprobengröße ab und hat einen starken Einfluss auf die Präzision, mit der eine Effektgröße gemessen werden kann. Präzision (angegeben durch Konfidenzintervalle) und Signifikanz des Gesamtschätzers werden ebenfalls ermittelt. Alle üblicherweise verwendeten Methoden für Metaanalysen folgen diesem Schema. Andere Aspekte jedoch können zwischen verschiedenen statistischen Methoden der Metaanalyse variieren (Deeks, Altman et al. 2001; Higgins and Green 2011).

7.3.1.2.2 Heterogenität

Ein wichtiger Aspekt einer Metaanalyse ist es, die Konsistenz des Behandlungseffekts zwischen den Primärstudien zu untersuchen bzw. die Heterogenität aufzudecken und

zu erklären. Unterschiedliche Studien zur Effektivität werden oft nicht nach einem gemeinsamen Protokoll durchgeführt. Dadurch entstehen zahlreiche Unterschiede hinsichtlich der eingeschlossenen Patientengruppen, der klinischen Einrichtungen, der möglichen Zusatzbehandlungen sowie der Verabreichung der Intervention. Die Variabilität der Schätzer zwischen den Studien kann größer sein als die Zufallsvariabilität. Dies wird als statistische Heterogenität bezeichnet. Sie ist eine Folge klinischer und methodischer Heterogenität (Higgins and Green 2011).

Klinische Heterogenität beinhaltet Unterschiede bei den Patienten, bei der Intervention und den Zielgrößen der Primärstudien. Methodische Heterogenität umfasst Unterschiede im Studiendesign und dem Risiko für Bias (Higgins and Green 2011). Heterogenität kann auch von der Wahl des Effektmaßes abhängen; so können beispielsweise Effekte als relatives Risiko über alle Studien der Metaanalyse hinweg gemessen homogen und gemessen als Risikodifferenz heterogen sein. Das Vorliegen von Heterogenität sollte statistisch geprüft und falls bestätigt formal untersucht werden. Hierzu existieren zahlreiche Verfahren, u. a. das Stratifizieren und Aufteilen des Studienpools in Subgruppen oder das Durchführen einer Metaregression zur Erklärung der Heterogenität. Diese Verfahren und der Umgang mit Heterogenität werden in Kapitel (7.3.1.2.4.2) im Detail erläutert.

7.3.1.2.3 Verzerrungsmöglichkeiten

Reporting/Publication Bias

In den vorangehenden Kapiteln des Methodenhandbuchs wurden bereits die Möglichkeiten für eine Verzerrung des wahren Effekts auf Ebene der einzelnen Studie angesprochen. Beim Zusammenfassen von Studien für einen systematischen Review, aber besonders für eine Metaanalyse können durch die Studienselektion weitere Verzerrungen bei der Schätzung von Effekten auf der Ebene der gesamten Metaanalyse auftreten.

Metaanalysen sind auf die Publikation von Studienergebnissen angewiesen und sollten alle zum jeweiligen Zeitpunkt zur Verfügung stehenden Daten beinhalten. Zahlreiche Studien werden jedoch aufgrund von unerwünschten bzw. nicht signifikanten Ergebnissen nicht oder nur zum Teil veröffentlicht. Der Großteil der publizierten Studien umfasst Studien mit positiven Effekten oder Studien, die bei einem Register gemeldet sind. Die dadurch entstehende Verzerrung wird u. a. als Reporting/Publication Bias bzw. als „File Drawer“-Problem bezeichnet (als bildhafter Vergleich des Schicksals der Studienergebnisse mit dem Umgang mit unerwünschten Befunden durch Wegsperrern in Schubladen) (Egger, Dickersin et al. 2001).

(Egger, Smith et al. 2001) beschreibt mehrere Arten von Reporting Bias:

- » „Publication bias“: das Publizieren oder Nicht-Publizieren von Studienergebnissen, abhängig von Art und Richtung der Ergebnisse;
- » „Time lag bias“: die beschleunigte oder verzögerte Publikation von Studienergebnissen, abhängig von Art und Richtung der Ergebnisse;
- » „Multiple (duplicate) publication bias“: mehrfache oder einfache Publikation von Studienergebnissen, abhängig von Art und Richtung der Ergebnisse;
- » „Citation bias“: Zitierung oder Nicht-Zitierung von Studienergebnissen, abhängig von Art und Richtung der Ergebnisse;
- » „Language bias“: Publikation von Studienergebnissen in einer bestimmten Sprache, abhängig von Art und Richtung der Ergebnisse
- » „Outcome reporting bias“: selektives Berichten bestimmter Outcomes, nicht jedoch anderer Outcomes in Abhängigkeit von Art und Richtung der Ergebnisse.

Wie man systematische Reviews auf das Vorhandensein solcher Verzerrungen untersuchen kann, wurde von (Dwan, Gamble et al. 2010) in einem Tutorium ausführlich dargestellt.

Um valide systematische Reviews bzw. Metaanalysen durchführen zu können, bedarf es gewissenhaft durchgeführter Studien sowie der Publikation aller Ergebnisse, seien sie positiv oder negativ (Egger, Dickersin et al. 2001). Sonst besteht die Gefahr, dass die Ergebnisse von Metaanalysen verzerrt sind. Mögliche Lösungsansätze sind die Etablierung von Studienregistern, in denen alle Studien registriert werden müssen, und die Verpflichtung, alle Studienergebnisse zu veröffentlichen. Verbindliche Leitlinien für Mindeststandards, was und wie zu berichten ist, wären in diesem Zusammenhang ebenfalls hilfreich. Derzeit werden Studien gelistet u. a. in der MEDLINE-Datenbank, im Cochrane Controlled Trials Register, auf Current Controlled Trials (www.controlled-trials.com) und in nationalen Registern. Dies ist jedoch z. T. noch nicht verpflichtend. Um die methodische Qualität von Studien sowie die Berichtsqualität zu gewährleisten, wurden bereits zahlreiche Leitlinien erstellt; siehe Kapitel 7.3.1.2.5.

7.3.1.2.4 Statistische Methoden

7.3.1.2.4.1 Ermittlung des gepoolten Schätzer

Gewichtungsverfahren

Die **Inverse-Varianz-Methode** wird sowohl für binäre als auch für stetige Daten genutzt. Effektgrößen können log OR, log RR, RD, MD etc. sein. Der gepoolte Effekt ist ein gewichtetes Mittel aller Behandlungseffekte der individuellen Studien. Die Gewichte sind dabei der Kehrwert des quadrierten Standardfehlers, d. h., große Studien mit kleinem Standardfehler haben mehr Gewicht als kleine Studien mit großem Standardfehler. Stärke dieses Ansatzes ist die weite Anwendbarkeit, d. h., jede Studie, die für

den Effekt einen Standardfehler angibt, kann in die Metaanalyse eingehen. Dies gilt für standardisierte Alterssterblichkeiten, Hazard Ratios und Ergebnisse von Querschnittsstudien. Ebenso kann diese Methode verwendet werden, wenn keine 2x2-Kreuztabellen, jedoch Konfidenzintervalle und der Effektschätzer für jede Studie vorhanden sind. (Deeks, Altman et al. 2001)

Die **Mantel-Haenszel**-Methode wird bei spärlichen Daten angewandt. Sind die Ereignisraten gering, die Stichproben relativ klein und der Standardfehler groß, ist die Inverse-Varianz-Methode nicht die beste Wahl. Bei geringer Datenmenge ist die Mantel-Haenszel-Methode zu bevorzugen, da eine andere Gewichtung zu robusteren Ergebnissen führt. Bei großen Studien liefert sie ähnliche Ergebnisse wie die Inverse-Varianz-Methode. Die Mantel-Haenszel-Methode ist jedoch nur für binäre Outcomes geeignet. (Deeks, Altman et al. 2001)

Das **Peto**-Verfahren ist eine alternative Methode zur Mantel-Haenszel-Gewichtung, auch „Peto Odds Ratio“ genannt. Die Odds Ratios werden über die Strata zusammengefasst. Diese Methode kann jedoch Bias verursachen, vor allem bei großen Unterschieden zwischen Interventions- und Kontrollgruppengröße und bei sehr großen Behandlungseffekten (Deeks, Altman et al. 2001).

Die drei genannten Verfahren sind Gewichtungsmethoden für Modellansätze mit festen Effekten (Erläuterung im folgenden Abschnitt). Eine gängige Methode bei der Annahme zufälliger Effekte ist das DerSimonian-and-Laird-Modell. Eine genaue Beschreibung findet sich bei (Deeks, Altman et al. 2001).

Modellansätze mit festen Effekten und zufälligen Effekten

Die meisten Metaanalysen basieren auf zwei statistischen Modellen, dem Modell mit festen Effekten („**Fixed-Effect-Model**“, FEM) und dem Modell mit zufälligen Effekten („**Random-Effects-Model**“, REM).

Beim FEM geht man bei der Schätzung des durchschnittlichen Effekts von der Annahme aus, dass ein einzelner, „wahrer“ Effekt allen Studien zugrunde liegt. Die Studien streuen um diesen Effekt zufällig. Dabei wird den größeren Studien größeres Gewicht zugerechnet als den kleinen Studien, da Erstere präzisere Informationen über den wahren Effekt enthalten (Borenstein, Hedges et al. 2009).

Im Gegensatz dazu besteht beim REM die Modellvorstellung, dass der wahre Effekt in den Studien verschieden sein kann und die beobachteten Effekte der Studien einer Verteilung des wahren Effekts folgen. Der durchschnittliche Effekt bildet dabei das Zentrum, und die Unterschiede der einzelnen Studien basieren auf der Variation des wahren Effekts und der zusätzlichen zufälligen Variation der einzelnen Studien (Deeks,

Altman et al. 2001). Unabhängig von der Größe des Effektschätzers erhalten damit große (d. h. umfangreiche) Studien im Vergleich zum FEM geringere Gewichte und damit auch einen geringeren Einfluss, sodass auch extreme Studienergebnisse einzelner Studien einen geringeren Einfluss auf den Gesamtschätzer erhalten. Die einzelne Studie erhält relativ gesehen eine größere Bedeutung (Borenstein, Hedges et al. 2009).

Diese Annahmen führen in der Konsequenz in den meisten Fällen auch zu breiteren Konfidenzintervallen, sodass REM meist etwas konservativer sind und ggf. bei ungeklärter Heterogenität eher zum Einsatz kommen können.

Die Wahl des geeigneten Modells wird in der Literatur allerdings kontrovers diskutiert. Ein FEM sollte Borenstein zufolge nur gewählt werden, wenn die eingeschlossenen Studien in ihrer Durchführung identisch sind und wenn mithilfe des gepoolten Schätzers nicht auf eine Gesamtbevölkerung geschlossen werden soll (Borenstein, Hedges et al. 2009). Das bedeutet, dass im Rahmen eines HTA-Berichts im Zweifelsfall eher ein REM verwendet werden sollte. Es wird empfohlen, die Wahl des Modells nicht im Nachhinein vom Ergebnis eines Heterogenitätstests abhängig zu machen, sondern vom Zweck der Untersuchung. Durch die Wahl eines FEM wird die Präzision des Effektschätzers häufig überschätzt (d. h., die Breite des Konfidenzintervalls fällt zu klein aus). Hunter und Schmidt raten daher zu einer bevorzugten Verwendung von REM (Hunter and Schmidt 2000). Wenn die Anzahl der in die Metaanalyse einbezogenen Studien sehr klein ist, besteht jedoch das Problem, dass die für das REM benötigte Varianz zwischen den Studien nur sehr ungenau geschätzt und auch die Verwendung des REM nicht empfohlen werden kann. Egger (2001) sieht bei der Verwendung von REM die Gefahr, dass großen Studien, die keinen protektiven Effekt zeigen, zu wenig Gewicht zukommt (Egger and Smith 2001). Die Wahl des Modells und die der Entscheidung zugrunde liegenden Überlegungen sollten daher im Methodenteil der Metaanalyse ausführlich beschrieben und begründet werden.

Bayes-Verfahren

Bayes-Verfahren stellen einen alternativen statistischen Denkansatz dar. Dabei wird von der Grundannahme ausgegangen, dass vorbestehendes Wissen oder Annahmen (prior beliefs) durch neue Informationen bzw. Daten aktualisiert werden und zu modifiziertem Wissen führen (posteriors). Hier wird der Behandlung der Unsicherheit von Variablenwerten große Bedeutung beigemessen. Die Szenarien bzw. Variablen werden dabei weniger durch Häufigkeiten als durch Wahrscheinlichkeitsverteilungen charakterisiert. Die Berechnungen erfolgen meist mit rechenintensiven Simulationsverfahren, die erst aufgrund der Entwicklung leistungsfähiger Computer möglich wurden und die diesem Ansatz zu Popularität und breiter Durchführbarkeit verholfen haben.

Das Vorwissen kann auf empirischen Daten basieren oder durch Annahmen modelliert werden. Letztere Möglichkeit beinhaltet wesentliche Anteile an Subjektivität, die wiederholt Anlass zu Kritik gegeben und zu Misstrauen gegenüber dem Ansatz geführt haben. Darüber hinaus besteht die Möglichkeit, Vorwissen zu ignorieren, in diesem Fall sind die Ergebnisse (posterior distribution) nur von den neuen Daten abhängig.

Für die Durchführung von Metaanalysen nach Bayes ist spezielle Software erforderlich, die jedoch frei verfügbar ist (WinBUGS). Spezifische methodische Kenntnisse vorausgesetzt, sind auch komplexere Modelle einfach umsetzbar und intuitiv interpretierbar, wenngleich auswertungsintensiv (Egger and Smith 2001; Sutton and Abrams 2001; Borenstein, Hedges et al. 2009).

Exkurs: Kumulative Metaanalyse

Die kumulative Metaanalyse verfolgt weniger einen statistischen als einen eher methodischen Ansatz, um insbesondere zeitliche Entwicklungen aufzuzeigen. Die kumulative Metaanalyse stellt wiederholt Metaanalysen im zeitlichen Verlauf unter Einschluss jeweils neu publizierter Studien zur Fragestellung dar. Das heißt, jedes Mal, wenn eine neue Studie veröffentlicht wird, wird erneut eine Metaanalyse durchgeführt, und die gepoolten Schätzer und Konfidenzintervalle aller Metaanalysen werden zeitlich geordnet grafisch dargestellt. Dies ermöglicht (retrospektiv) den Zeitpunkt festzustellen, ab dem sich ein statistisch signifikanter Effekt zeigt, und hat insbesondere Bedeutung in der Versorgungsforschung und bei der Formulierung von (Therapie-)Empfehlungen z. B. in Leitlinien. Die Ergebnisse kumulativer Metaanalysen können dazu führen, dass ab einem gewissen Zeitpunkt redundante Studien vermieden und über zeitnahe Implementierung in Behandlungsempfehlungen relevante Gesundheitseffekte auf Public-Health-Ebene erzielt werden (Petitti 2000; Egger, Smith et al. 2001).

7.3.1.2.4.2 Untersuchung von Heterogenität

Umgang mit Heterogenität

Bei der Annahme oder dem statistisch nachgewiesenen Vorliegen von Heterogenität in relevantem Ausmaß ist nach Durchführung qualitätssichernder Maßnahmen (Kontrolle, ob bei der Datenextraktion oder -übertragung keine Fehler gemacht wurden) zu untersuchen, inwieweit die Heterogenität durch Faktoren, durch die sich die Studien unterscheiden, erklärt werden kann. Dies kann dazu führen, einzelne Studien von der Metaanalyse auszuschließen. Die Berechnung eines gepoolten Effektschätzers bei relevanter Heterogenität ist unangemessen und irreführend. Welches Ausmaß an Heterogenität toleriert wird, hängt jedoch auch von der Fragestellung ab (Sutton and Higgins 2008). Die Analyse der Heterogenität stellt einen zentralen Aufgabenbereich der Metaanalyse dar, der zu Generierung weiterführender Hypothesen hinsichtlich

zugrunde liegender Wirkmechanismen führen kann. Verschiedene Techniken zur Untersuchung von Heterogenität stehen zur Verfügung, diese werden nachfolgend kurz dargestellt.

Statistische Tests/Kenngrößen

Zusätzlich zu den bereits erwähnten Verfahren zur Analyse von Heterogenität bestehen verschiedene Kenngrößen und statistische Tests zur Quantifizierung der Heterogenität des tatsächlichen Effekts bzw. zur Differenzierung der Anteile der Heterogenität, die auf den tatsächlichen Effekt und auf den zufälligen Fehler der einzelnen Studien entfallen. Im Folgenden soll ein knapper inhaltlicher Überblick vermittelt werden; hinsichtlich statistischer Details muss auf die Primärliteratur bzw. auf Standardlehrbücher verwiesen werden.

- » **Q (Cochrane's Q)** ist die gewichtete Summe der quadrierten Unterschiede zwischen den einzelnen Studieneffekten und dem gepoolten Effekt zwischen den Studien. Gewichtet wird mit der jeweiligen Gewichtungsmethode, die für den gepoolten Effekt verwendet wurde. Die Q-Statistik ist als standardisiertes Maß unabhängig von der Skalierung der Effektgröße und stellt die Basis für weitere Testgrößen dar.
- » Der Vergleich der Q-Statistik mit χ^2 -Prüfwerten (**χ^2 -Test**) ermöglicht Aussagen zur statistischen Signifikanz der beobachteten Unterschiede in den Effektgrößen. Bei einem kleinen p-Wert wird von Heterogenität ausgegangen. Testergebnisse müssen jedoch vorsichtig interpretiert werden, da bei kleinen Stichproben und geringer Studienanzahl die Aussagekraft eingeschränkt ist. Daher wird meist schon ab einem p-Wert kleiner 0,1 von statistischer Signifikanz ausgegangen, um Heterogenität nicht zu leichtfertig auszuschließen (Borenstein, Hedges et al. 2009; Higgins and Green 2011).
- » **I²** gibt den Anteil der Varianz über die Studien hinweg wieder, der auf Heterogenität des tatsächlichen Effekts zurückzuführen ist, bezogen auf die gesamte Streuung. Diese Testgröße ist von der Q-Statistik abgeleitet (Higgins, Thompson et al. 2003). Im Cochrane-Handbuch werden vier Kategorien als grobe Richtlinie zur Interpretation der Werte von I² angegeben: Wert zwischen 0 und 40 Prozent können als Heterogenität ohne Bedeutung, Werte zwischen 30 und 60 Prozent als moderate, Werte zwischen 50 und 90 Prozent als substantielle und Werte zwischen 75 und 100 Prozent als beträchtliche Heterogenität interpretiert werden. Die Verwendung fixer Schwellenwerte wurde vermieden, weil die Bedeutung von I²-Werten von der Größe und Richtung des Effekts und der Stärke der Evidenz für die Heterogenität, also z. B. dem Konfidenzintervall des I²-Werts, abhängt. (Higgins, Thompson et al. 2003; Higgins and Green 2011).
- » **Tau²** wird im Rahmen des REM berechnet und beschreibt das absolute Ausmaß der Varianz des tatsächlichen Effekts. Die Größe ist nicht standardisiert und damit von der Skalierung der Effektgröße und deren absolutem Betrag abhängig.

Subgruppen-Analysen

Das Hauptziel einer Metaanalyse ist das Generieren eines durchschnittlichen Punktschätzers für den Effekt einer Intervention. Ausmaß und Richtung des durchschnittlichen Effekts sollen helfen, Entscheidungen für die klinische Praxis für ein breites Patientenspektrum zu treffen. Häufig interessiert jedoch auch, ob der Effekt zwischen verschiedenen Patientengruppen variiert. Um dies herauszufinden, werden in Metaanalysen ähnlich wie in klinischen Studien Subgruppenanalysen durchgeführt. Subgruppenanalysen unterliegen jedoch der Gefahr zufälliger Auffälligkeiten (McAlister 2001; Smith and Egger 2001). Nach Möglichkeit sollten Subgruppenanalysen bereits bei der Planung der Metaanalyse berücksichtigt und entsprechend beschrieben werden. Bei ungeplanten, nachträglich durchgeführten Subgruppenanalysen ist diese Tatsache offenzulegen und zu begründen. Die hierbei gewonnenen Erkenntnisse sind aufgrund der Fehler- und Verzerrungsmöglichkeiten zurückhaltend zu interpretieren und können im Wesentlichen der Generierung weitergehender Forschungshypothesen dienen.

Sensitivitätsanalysen

Um die Robustheit der Ergebnisse zu überprüfen, ist es auch bei Metaanalysen empfehlenswert, Sensitivitätsanalysen durchzuführen. So werden vor dem Hintergrund divergierender Ansichten zur Wahl des statistischen Modells mit Sensitivitätsanalysen Unterschiede zwischen FEM und REM untersucht. Bei Unterschieden in Studienpopulation oder -design bzw. in der Bewertung der methodischen Qualität eingeschlossener Studien sollten in Subgruppenanalysen (z. B. Art der Randomisierung, Studiengröße, Follow-up-Dauer etc.) die Auswirkungen der entsprechenden Faktoren auf den gepoolten Schätzer geprüft werden. Geringe Schwankungen der Effektgrößen zeigen robuste Ergebnisse (Egger and Smith 2001).

Metaregression

Die Metaregression bedient sich eines (meist linearen) Regressionsmodells zur Analyse und Erklärung der Heterogenität sowie der Bedeutung und Richtung verschiedener Einflussvariablen. Die Beobachtungseinheiten in der Regression sind die in die Metaanalyse eingeschlossenen Studien bzw. Studienarme oder ausgewählte Patientengruppen. Die abhängige Variable der Metaregression könnte im Falle einer dichotomen Variable das logarithmierte (log) Odds Ratio der Zielgröße der einzelnen Studien sein. Die Prädiktoren (unabhängige Variablen) der Regression sind Merkmale auf Studienebene und könnten zum Beispiel folgende Variablen beinhalten: klinisches Protokoll, Charakteristika der Studienpopulation (Durchschnittsalter, ethnische Herkunft, Geschlechtsverteilung) oder auch Variablen, die das Studiensetting beschreiben (Größe

des Krankenhauses, Lage bzw. Typ etc.) (Thompson and Higgins 2002; Morton, Adams et al. 2004).

7.3.1.2.4.3 Untersuchung von Publication Bias

Um Hinweise auf ein mögliches Publication Bias zu erhalten, gibt es verschiedene statistische und grafische Methoden. Mit diesen können sowohl der Einfluss auf das Ergebnis quantifiziert als auch korrigierte Effektgrößen berechnet werden (Sterne, Egger et al. 2001). Im Folgenden werden einige Verfahren zur Aufdeckung und Korrektur von Publication Bias behandelt.

Zu den grafischen Explorationsmöglichkeiten zählt der sogenannte Funnel Plot (Trichter-Diagramm). Die Effektgrößen der Studien werden dabei auf der x-Achse aufgetragen, dem gegenübergestellt wird die Stichprobengröße oder die Varianz der Studie. Diese Grafik basiert auf der Annahme, dass die Präzision der Ergebnisse mit der Größe der Stichprobe bzw. mit sinkender Varianz des Effektschätzers zunimmt. Ohne Bias ist der Funnel Plot symmetrisch; falls Bias vorhanden sind, nimmt der Funnel Plot eine asymmetrische bzw. schräge Form an. Asymmetrische Verteilungen können jedoch auch andere Ursachen als Publication Bias haben (Egger, Davey Smith et al. 1997). In der Literatur wird daher auch häufig allgemeiner von „Small Study Effects“ gesprochen (Sterne, Gavaghan et al. 2000). Weiterführende Informationen zum Funnel Plot sind in Kapitel 7.3.1.4.2.2 aufgeführt.

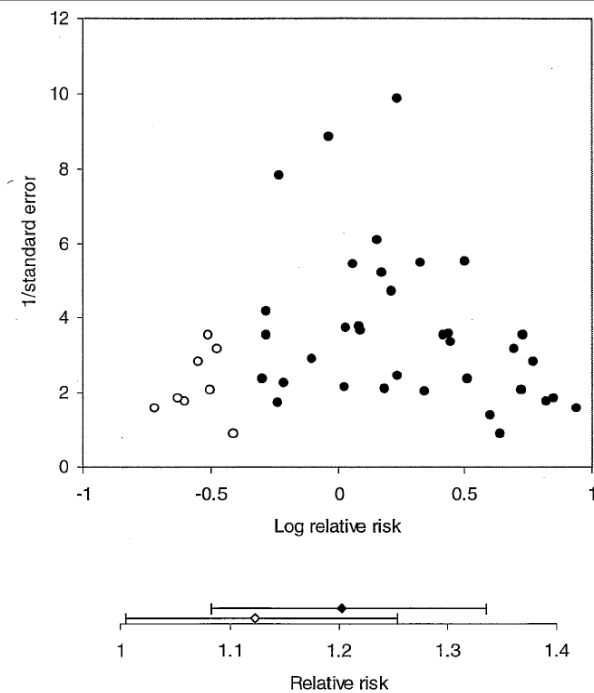
Mithilfe des „Trim & Fill“-Ansatzes von Duval und Tweedie soll ein korrigierter gepoolter Effektschätzer ermittelt werden. Die Anzahl der fehlenden Studien wird durch Symmetrie-Annahmen ermittelt. Fehlende Werte werden dabei durch den „Trim & Fill“-Ansatz berechnet, und in weiterer Folge wird die Symmetrie des Diagramms hergestellt (grafisches Auffüllen). Tabelle 7.5 zeigt einen Funnel Plot, bei dem die Symmetrie durch Hinzufügen von zusätzlichen, fiktiven Studien (Kreise ohne Füllung) wiederhergestellt wurde. Darunter sieht man die Auswirkungen auf den gepoolten Punktschätzer vor (voller Punkt) und nach dem Anwenden des „Trim & Fill“-Ansatzes. (Duval and Tweedie 2000)

Zusätzlich zu den grafischen Verfahren existieren statistische Tests und Kenngrößen, u. a. der Begg-Test und der Egger-Test. Der Begg-Test ist ein adjustierter Rang-Korrelations-Test, der analog zum Funnel Plot durchgeführt wird. Folgende wesentliche Faktoren beeinflussen die Teststatistik: (i) die Anzahl der einbezogenen Studien, (ii) die zugrunde liegenden Effektgrößen sowie (iii) die Spannweite der Studienvarianzen (Begg and Mazumdar 1994).

Beim Egger-Test wird die Asymmetrie des Funnel Plots mithilfe einer linearen Regression berechnet. Die abhängige Variable stellt hierbei die Standardnormalabweichung (SND = standard normal deviation) dar. Die SND berechnet sich durch die Odds Ratio,

dividiert durch den Standardfehler. Als unabhängige Variable wird die Präzision (= Kehrwert des Standardfehlers) eingesetzt. Die daraus resultierende Formel lautet: $SND = a + b * \text{Präzision}$. Führt diese nicht durch den Koordinatenursprung, bedeutet dies eine Asymmetrie. Statistische Signifikanz wird auch hier bereits ab einem p-Wert kleiner als 0,1 unterstellt (Egger, Davey Smith et al. 1997).

Abbildung 7.3:
Funnel Plot und gepoolter Punktschätzer
vor und nach Anwendung des „Trim & Fill“-Ansatzes



Quelle und Darstellung: Duval and Tweedie 2000

Obere Grafik zeigt einen Funnel Plot der logarithmierten relativen Risiken für die Assoziation von Lungenkrebs und umweltbezogener Tabakrauchexposition. Kreise ohne Füllung wurden durch den „Trim & Fill“-Ansatz hinzugefügt. Der untere Teil der Grafik zeigt das gesamte relative Risiko und das 95%-Konfidenzintervall vor und nach Berücksichtigung von Publication Bias (Duval and Tweedie 2000).

7.3.1.2.4.4 Indirekte Vergleiche und Mixed Treatment Comparisons

Indirekte Vergleiche

Die Bewertung zweier konkurrierender medizinischer Technologien sollte wenn möglich in einem direkten Vergleich erfolgen. Ein Großteil der randomisierten kontrollierten Studien, die für Zulassungszwecke durchgeführt werden, vergleicht neue Medikamente jedoch meist mit einem Placebo oder mit dem Versorgungsstandard. So ist es nicht selten, dass Studien mit direkten Vergleichen zweier aktiver Interventionen (sogenannte Head-to-head-Vergleiche) nicht existieren und somit auch für eine Metaanalyse nicht genutzt werden können. Gibt es jedoch Studien, die die jeweils interessierende Intervention und einen gemeinsamen Komparator untersuchen, kann durch Metaanalysen aus indirekten Vergleichen die Evidenz aus diesen Studien zur Beantwortung der Forschungsfragen herangezogen werden. Das Bias-Risiko für diese Art der Vergleiche ist in der Regel jedoch größer als bei Metaanalysen von direkt vergleichenden Studien. Wenn keine direkte Evidenz vorhanden ist, können indirekte Vergleiche aber zusätzliche, analytische Möglichkeiten bieten, zwei konkurrierende Interventionen zu vergleichen.

Angenommen, es sollen zwei Interventionen A vs. B miteinander verglichen werden, es gibt aber nur Studien zu A vs. C und B vs. C. Mittels indirekter Vergleiche kann die Größe des Unterschieds von A vs. B statistisch geschätzt werden. Es gibt dazu mehrere Möglichkeiten. Die einfachste Art ist die Durchführung von zwei Metaanalysen, die die Studien A vs. C und B vs. C poolen. Die Durchführung von Metaanalysen ist deshalb wichtig, weil die Randomisierung zwischen den Studienarmen aufrechterhalten, d. h. die Komparatorintervention in der Analyse berücksichtigt wird. Ein nicht-adjustiertes Zusammenfassen nur der interessierenden Interventionsarme kann wegen der nicht mehr vorhandenen Randomisierung eine Effektverzerrung durch Confounding nicht ausschließen und sollte deshalb nicht durchgeführt werden. Bei einem adjustierten Vergleich werden in einem ersten Schritt die Effekte der Interventionen gegenüber dem Komparator berechnet (z. B. Placebo). Aus beiden resultierenden Summeneffektmaßen (A vs. C und B vs. C) wird die Differenz gebildet (Gartlehner and Moore 2008). Voraussetzung für die Validität adjustierter indirekter Vergleiche ist neben der Annahme ausreichender Homogenität wie in der Standard-Metaanalyse, dass der durchschnittliche relative Effekt des Komparators C in den beiden direkten Vergleichspaaren ähnlich ist. Ob diese Ähnlichkeit gegeben ist, hängt sowohl von den klinischen Charakteristika wie Patienten, Interventionen und Settings als auch vom Risiko für Bias durch das Studiendesign ab. Zur Untersuchung der Ähnlichkeit können die gleichen Methoden angewendet werden wie zur Untersuchung der Heterogenität. Die Annahmen über Heterogenität und Ähnlichkeit sollten bei der Verwendung indirekter Vergleiche explizit offengelegt werden (Song, Loke et al. 2009). Allerdings ist es schwierig festzustellen und größtenteils nicht überprüfbar, ob diese Voraussetzungen erfüllt

sind. Die Validität der Resultate von indirekten Vergleichen beruht daher zum Teil auf Annahmen, die nicht verifizierbar sind.

Ein weiterer Nachteil indirekter Vergleiche ist die deutlich geringere statistische Power, was häufig statistisch nicht signifikante Ergebnisse mit weiten Konfidenzintervallen nach sich zieht, wenn nur wenige Studien verfügbar sind. Drittens existiert kein Konsens, wie die Ergebnisse indirekter Vergleiche zu interpretieren sind, wenn sie im Widerspruch zu direkten Vergleichen stehen (Gartlehner and Moore 2008).

Falls keine direkte Evidenz vorliegt und dennoch eine Entscheidung über eine Technologie getroffen werden muss, erscheint dieser quantitative Ansatz jedoch transparenter und expliziter als weniger formalisierte Ansätze (Lu and Ades 2004; Sutton, Ades et al. 2008). Grundsätzlich sollten indirekte Vergleiche bereits im Voraus im Studienprotokoll einer Metaanalyse geplant sein, und bei der Interpretation sollte der höheren Unsicherheit der Ergebnisse Rechnung getragen werden.

Mixed Treatment Comparisons

Wenn in einem HTA mehr als zwei Interventionen verglichen werden sollen, um die beste Therapieoption für eine Indikation zu identifizieren, wie dies bei Fragestellungen der „comparative effectiveness research“ der Fall ist, stößt die Methode der paarweisen Metaanalyse an ihre Grenzen. Es wurden Metaanalyse-Methoden entwickelt, die auch mehr als zwei Interventionen gleichzeitig vergleichen können und die sowohl die Evidenz aus direkten als auch aus indirekten Vergleichen in die Analyse einbeziehen. Diese Methoden werden häufig als „mixed treatment comparison“ (MTC) oder Netzwerk-Metaanalysen bezeichnet (Sutton, Ades et al. 2008). Da hier Netzwerke aus Studien gebildet werden, die über verschiedene Komparatoren verknüpft werden, muss hier eine weitere Annahme über die Konsistenz des Behandlungseffekts über mehrere indirekte und direkte Vergleiche getroffen werden, und die Inkonsistenz in den Netzwerken muss überprüft werden. Welche Studien in die Analyse ein- oder ausgeschlossen werden sollen und wie weit das Netzwerk ausgedehnt werden soll, ist derzeit unklar (Gartlehner and Moore 2008; Sutton, Ades et al. 2008). Die Methodik der MTC ist derzeit noch in Entwicklung, und die Validität der Ergebnisse ist noch zu unklar, um Netzwerk-Metaanalysen bereits regelmäßig in systematischen Reviews und HTA-Berichten verwenden zu können (Salanti, Higgins et al. 2008). Derzeit werden sie am ehesten dazu verwendet, die Effektivitätsdaten für Kosten-Effektivitäts-Analysen mehrerer konkurrierender Interventionen zu generieren (Song, Loke et al. 2009). Eine Weiterentwicklung der Methoden für indirekte Vergleiche und MTC ist speziell für HTA-Berichte höchst relevant, in denen mehrere Behandlungsmethoden miteinander verglichen werden sollen. Mittlerweile gibt es einige HTA-Berichte, die sich ausführlich mit den Methoden indirekter Vergleiche und/oder MTC beschäftigen (Glenny, Altman et al. 2005; Schöttker, Lühmann et al. 2009; Wells, Sultan et al. 2009).

7.3.1.2.4.5 Exkurs: Metaanalyse individueller Patientendaten

Die Analyse individueller Patientendaten stellt eine spezielle Art der Metaanalyse dar (sogenannte IPD-Metaanalyse). Sie bedient sich im Gegensatz zur klassischen Metaanalyse, die die aggregierten Effektschätzer (z. B. Mittelwerte und Standardabweichungen) publizierter Studien verwendet, der Originaldaten der jeweiligen Studien. Jedoch ist es eher selten, dass den Autoren der Metaanalyse die Studiendaten aller einzuschließenden Studien verfügbar gemacht werden. Eine derartige Datenbasis bietet über die Auswertungen einfacher Metaanalysen weit hinausgehende Möglichkeiten der Analyse, wie die Definition neuer bzw. anderer Kategorien von Einfluss- oder Zielvariablen oder von Effektmaßen, die Durchführung verschiedener statistischer Tests und die Untersuchung von Heterogenität z. B. über Regressionsansätze unter Berücksichtigung von Daten auf Patientenebene (Patientencharakteristika). Neben der direkten Auswertung des gepoolten Datensatzes ist auch eine Auswertung auf Studienebene mit sukzessiver Kombination der Ergebnisse anhand metaanalytischer Techniken möglich (Clarke and Stewart 2001; Borenstein, Hedges et al. 2009).

7.3.1.2.5 Leitlinien

7.3.1.2.5.1 Leitlinien zur Optimierung der Berichtsqualität von Metaanalysen

Leitlinien zur Berichtsqualität haben das Ziel, die Klarheit und Transparenz von Veröffentlichungen systematischer Reviews und MA zu erhöhen. Sie bewerten nicht unmittelbar die methodische Qualität von SR und MA, beschreiben jedoch detailliert alle Elemente, die für die methodische Qualität wichtig sind, und geben insofern auch direkte Hinweise darauf, auf welche Punkte bei der Durchführung von SR und MA geachtet werden sollte. Eine der wichtigsten Berichtsleitlinien ist das „PRISMA Statement“ (Liberati, Altman et al. 2009; Moher, Liberati et al. 2009) (früher: QUOROM Statement). PRISMA steht für „Preferred Reporting Items for Systematic Reviews and Meta-Analysis“ und besteht aus einer Checkliste mit 27 Punkten und einem Diagramm zur Studienselektion in einer Metaanalyse. Entlang der Struktur einer wissenschaftlichen Veröffentlichung wird beschrieben, welche Punkte in Titel, Abstract, Einleitung, Methodenteil, Ergebnissen und Diskussion berichtet werden sollen. In einer ausführlichen Anleitung werden positive Beispiele für die einzelnen Items aus der Literatur erläutert. PRISMA bezieht sich auf SR und MA von RCTs und andere Studientypen.

Auf die Besonderheiten von Beobachtungsstudien geht MOOSE (Stroup, Berlin et al. 2000) ein. MOOSE steht für „Meta-Analysis of Observational Studies in Epidemiology“ und unterstützt eine systematische und transparente Beschreibung von Metaanalysen von Beobachtungsstudien. Die Items der Checkliste sind ähnlich aufgebaut wie in PRISMA.

7.3.1.2.5.2 Empfehlungen zur Durchführung von Metaanalysen

Der umfassendste Ratgeber zur Methodik, Planung, Durchführung, Darstellung und Interpretation systematischer Reviews und Metaanalysen ist anerkannterweise nach wie vor das Cochrane-Handbuch, das online verfügbar ist (www.cochrane-handbook.org/) (Deeks, Altman et al. 2001; Higgins and Green 2011).

Darüber hinaus bietet der Artikel von Riley und Kollegen eine Hilfestellung zur Durchführung von IPD-Metaanalysen (Riley, Lambert et al. 2010). Hier spielt die Berichtsqualität eine untergeordnete Rolle, vielmehr wird hier auf PRISMA und MOOSE verwiesen. In erster Linie werden Sinn und Zweck, Vor- und Nachteile von MA mit IPD dargestellt und wie sie durchgeführt werden können. Zur Bewertung der methodischen Qualität von systematischen Reviews und MA gibt es eine Reihe von Bewertungsinstrumenten. Diese sind beispielsweise im DIMDI-HTA-Bericht von Dreier (Vergleich von Bewertungsinstrumenten für die Studienqualität von Primär- und Sekundärstudien zur Verwendung für HTA-Berichte im deutschsprachigen Raum) aufgeführt und dargestellt (Dreier, Borutta et al. 2010). Ein validiertes Bewertungsinstrument ist z. B. AMSTAR (Assessment of Multiple SysTematic Reviews) (Shea, Grimshaw et al. 2007). Informationen und Links zu zahlreichen Berichtsleitlinien für diverse Studientypen, Originalstudien wie systematische Reviews, auch zu solchen, die hier nicht erwähnt wurden, finden sich im Internet im EQUATOR-Netzwerk (<http://www.equator-network.org/>) (Simera, Moher et al. 2010).

7.3.1.2.6 Interpretation

Die Interpretation der Effektschätzer hat die der Methode innewohnenden Einschränkungen zu berücksichtigen. So ist der Gesamteffekt ein gewichteter Durchschnittswert, der je nach Gewichtungsmethode variieren kann. Ergänzende Evidenz aus qualitativen Daten (z. B. aus Fallberichten), die in die quantitativen Auswertungen keinen Eingang gefunden hat, kann v. a. im hypothesengenerierenden Sinne die Interpretation der Ergebnisse der Metaanalyse unterstützen. Dabei sind die jeweiligen Verzerrungsmöglichkeiten angemessen zu berücksichtigen. So könnten beispielsweise Fallberichte Hinweise auf eine möglicherweise unterschiedliche Wirkung einer Intervention in bestimmten Untergruppen geben.

Von wesentlicher Bedeutung ist es, die medizinische Relevanz (der Größenordnung) eines Ergebnisses von dessen statistischer Signifikanz zu unterscheiden und ggf. vor dem Hintergrund der einbezogenen Datenmasse kritisch in die Interpretation einzubeziehen. Hierzu ist neben der inhaltlichen Beurteilung der medizinischen Relevanz auch die Bedeutung des untersuchten Endpunkts für die Betroffenen von Wichtigkeit. Beispielhaft sei die Erfassung der Wirksamkeit von Medikamenten gegen Demenzerkrankungen angeführt. Diese werden häufig anhand von klinischen Skalen beurteilt. Eine Veränderung um einen einzigen Skalenpunkt kann bei Einbezug ausreichender

Patientenzahlen als statistisch signifikant belegt werden, dürfte jedoch weder eine medizinische Bedeutung für den Grad der Demenz noch für das Befinden der Patienten haben.

Die wichtigste Einschränkung von Metaanalysen besteht in der Abhängigkeit von der Daten-, Studien- und Berichtsqualität. So kann die Qualität der Daten und Methodik eingeschlossener Studien nur insoweit evaluiert werden, als diese transparent und vollständig berichtet werden. Eine eingeschränkte Berichtsqualität kann Schlussfolgerungen der Metaanalyse selbst deutlich beeinflussen („garbage in, garbage out“). Bei der Durchführung der Metaanalyse ist zu prüfen, ob die zwischen den Primärstudien vorhandene Heterogenität ein sinnvolles Zusammenfassen zulässt oder ob es sich um eigenständige Entitäten bzw. Kontexte handelt („Vergleich von Äpfeln und Birnen“). Gegebenenfalls ist eine Synthese auf ausreichend homogene Untergruppen zu beschränken. Der häufig geäußerten Kritik an einem aggregierten Schätzer („Zusammenfassung eines gesamten Forschungsfelds in einer Zahl“) ist zu entgegnen, dass in der klinischen und wissenschaftlichen Praxis Entscheidungen häufig anhand eines Entscheidungsparameters getroffen werden bzw. getroffen werden müssen. Die Notwendigkeit, die Fülle verfügbarer Informationen in korrekter Weise zusammenzufassen, kann als generell akzeptiert betrachtet werden (Borenstein, Hedges et al. 2009).

Um die für Metaanalysen spezifischen Verzerrungen weitestgehend zu vermeiden, wird häufig empfohlen, in Metaanalysen auch nicht in Fachzeitschriften publizierte Studienergebnisse einzubeziehen (sog. graue Literatur wie z. B. Berichte von Institutionen, Kongresspräsentationen oder Berichte auf Internetseiten), eine Handsuche durchzuführen, unterschiedliche Datenbanken heranziehen sowie regelmäßige Aktualisierungen durchzuführen (Tricco, Tetzlaff et al. 2010). Ziel dabei ist es, einer bevorzugten Publikation positiver bzw. statistisch signifikanter Ergebnisse und der damit verbundenen Überschätzung von Effekten entgegenzuwirken. Neben den genannten Empfehlungen gibt es zusätzliche methodische Ansätze zur Prüfung bzw. Korrektur eines „publication bias“ (s. Abschnitt 7.3.1.2.4.3). Dem Verwenden von nicht in Fachzeitschriften mit „peer review“-Verfahren veröffentlichter Literatur steht jedoch oft die Forderung zum Ausschluss qualitativ eingeschränkter Studien entgegen. Häufig sind unveröffentlichte Studien von geringerer Qualität, zumindest jedoch wurden sie nicht den im Rahmen eines Veröffentlichungsprozesses üblichen Qualitätskontrollen in zufriedenstellender Weise unterzogen. Eine systematische Suche grauer Literatur z. B. mit Internetsuchmaschinen ist wegen der Fülle der Information und der fehlenden Transparenz der Suchalgorithmen faktisch unmöglich. Eine verzerrte Selektion kann deshalb hier nicht ausgeschlossen werden. Daher sind die oben genannten Empfehlungen gegeneinander abzuwägen. Die Abwägung sollte offengelegt, die Gründe sollten berichtet und das Ergebnis diskutiert werden. Vorgegebene Ein- und Ausschlusskriterien können das Ergebnis stark beeinflussen. So könnten selbst wichtige Studien ignoriert werden. Die Auswahl der Studien ist daher transparent zu berichten und zu begründen.

7.3.1.3 Besonderheiten verschiedener Studientypen

Nachfolgend werden Besonderheiten und Limitationen besonderer Studientypen zusammenfassend dargestellt. Diese Zusammenstellung erfolgt vor dem Hintergrund, dass die Besonderheiten, Fehlermöglichkeiten und Limitationen der jeweiligen Studientypen bei der Beurteilung der Studienqualität im Rahmen der Erstellung von Metaanalysen sowie bei der Interpretation der Ergebnisse angemessen berücksichtigt werden müssen.

7.3.1.3.1 Studien zur Effektivität

7.3.1.3.1.1 Randomisierte kontrollierte klinische Studien

Randomisierte kontrollierte klinische Studien (RCTs) gelten als Goldstandard zur Ermittlung von kausalen Interventionseffekten. Gut durchgeführte doppelblinde RCTs unterliegen aufgrund der Randomisierung, der Verblindung, starker Standardisierung und zahlreichen Richtlinien nur in geringem Ausmaß Verzerrungen durch Confounding, Selektions- und Informationsfehler. Die überwiegende Mehrheit publizierter Metaanalysen wird mit RCTs durchgeführt. Dies mag darin begründet liegen, dass der Bereich therapeutischer Intervention Gegenstand des größten Forschungsinteresses ist, die metaanalytische Methodik in diesem Bereich ist am weitesten ausgereift und etabliert, außerdem gelten in den Primärstudien die wohl höchsten Qualitätsstandards. Zudem kann der Fokus der Cochrane Collaboration auf Interventionsstudien als prägend angesehen werden.

Dennoch weisen auch RCTs Einschränkungen auf. So ist aufgrund der starken Selektion (Ein-/Ausschlusskriterien) und des artifiziellen Studienumfelds die Generalisierbarkeit der Ergebnisse häufig problematisch. Insbesondere Aussagen zur Wirksamkeit unter Realbedingungen sind häufig nur sehr eingeschränkt möglich, selten auftretende unerwünschte Wirkungen werden nur unzureichend erfasst (Schöffski and Schulenburg 2008; Schumacher and Schulgen 2008).

7.3.1.3.1.2 Studien mit Cross-over-Design

Cross-over-Studien stellen ein Studiendesign dar, das Studienteilnehmer sowohl mit dem Prüfmedikament als auch mit der Placebokontrolle exponiert. Dies erfolgt (selten) entweder gleichzeitig (z. B. an verschiedenen Körperstellen bei Dermatologika) oder typischerweise zeitlich nacheinander. Bei zeitlicher Folge sind die Applikationen durch Phasen ohne Intervention, sog. „wash-out“-Phasen, getrennt, die je nach Arzneimittelwirkung lange genug sein müssen, um die Effekte sicher zu trennen. Die Reihenfolge der Interventionen erfolgt bei qualitativ hochwertigen Studien randomisiert. Jeder Studienteilnehmer dient als eigene Kontrolle. Die dadurch reduzierte Varianz führt

dazu, dass dieses Design statistisch sehr effizient ist und geringere Stichprobenumfänge erfordert. Jedoch sind bei der statistischen Auswertung die besonderen Varianz- und Korrelationsstrukturen abhängiger Daten angemessen zu berücksichtigen. Verzerrungen können durch eine inadäquate Länge der „wash-out“-Perioden entstehen. Dieser Studientyp setzt eine chronische Grunderkrankung bzw. (häufig) rezidivierende Schübe bzw. Anfälle bei allenfalls symptomatisch wirksamer Intervention voraus (Whitehead 2002; Schumacher and Schulgen 2008).

Prinzipiell ist eine Kombination von Studien mit Cross-over-Design und Parallel-Gruppen-Design in einer Metaanalyse möglich, wenn jeweils der Effektschätzer und der entsprechend dem jeweiligen Studiendesign korrekt berechnete Standardfehler berichtet werden (Elbourne, Altman et al. 2002). Diese Ergebnisse sind in Publikationen oft in auswertbarer Weise enthalten. Ein weiterer Ansatz besteht darin, Cross-over-Studien wie Studien mit parallelen Kontrollgruppen auszuwerten. Hierbei wird die geringere Varianz aufgrund mehrfacher Messung im gleichen Individuum ignoriert. Dies führt tendenziell zur Unterschätzung eines Studieneffekts. Ein korrekterer Ansatz beschränkt die Auswertung auf die Daten der ersten Behandlungsphase (vor dem „cross-over“). Mit diesem Ansatz können auch Übertragungseffekte ausgeschlossen werden. Ein Nachteil besteht darin, dass nicht alle vorliegenden Informationen genutzt werden. Soweit Ergebnisse nur für einen Teil der Primärstudien in dieser Weise dargestellt werden, ist zu prüfen, inwieweit es sich hier aufgrund der eher ungewöhnlichen Präsentationsweise um eine verzerrte Untergruppe von Studien beispielweise aufgrund von Carry-over-Effekten handeln könnte. Darüber hinaus können bei diesem Auswertungsansatz die Behandlungs- bzw. Kontrollgruppen strukturunterschiedlich sein, da unter Umständen nur hinsichtlich der Behandlungsreihenfolge, nicht jedoch nach Patientencharakteristika (stratifiziert) randomisiert wird (Curtin, Altman et al. 2002; Curtin, Elbourne et al. 2002; Curtin, Elbourne et al. 2002; Elbourne, Altman et al. 2002; Schumacher and Schulgen 2008; Higgins and Green 2011).

7.3.1.3.1.3 Beobachtungsstudien

Metaanalysen zu Beobachtungsstudien werden seltener durchgeführt als Metaanalysen zu RCTs. Bedingt durch den zugrunde liegenden Studientyp, können Verzerrungen häufiger auftreten (Confounding, Selektionsbias). Beobachtungsstudien sind aufgrund der geringeren Standardisierung in Methodik, Struktur und Datenbasis deutlich heterogener, umfassen jedoch meist höhere Fallzahlen. Bei der Durchführung der Metaanalyse sollen die für Confounding adjustierten Effektschätzer verwendet werden. Hierbei ist zu berücksichtigen, dass in der Regel in unterschiedlichen Studien nach unterschiedlichen Confoundern adjustiert wurde.

Die Generalisierbarkeit auf die Zielpopulation ist aufgrund eines naturalistischen Studenumfelds in aller Regel relativ gut gegeben. Sehr häufig sind Beobachtungsstudien Fragestellungen vorbehalten, die aus methodischen, ethischen, logistischen

und/oder wirtschaftlichen Gründen klinischen Studien nicht zugänglich sind (z. B. gesundheitsgefährdende Noxen wie Nikotin oder Asbest). Jedoch zeigen seltene Situationen, in denen beide Studientypen vorliegen, dass gleiche Effektschätzer ermittelt werden können, wie dies beispielweise bei der Assoziation zwischen Hormonersatztherapie und Thromboembolien der Fall war (Canonic, Plu-Bureau et al. 2008). Für Fragestellungen der Versorgungsforschung sind Beobachtungsstudien aufgrund des stärker an den Alltagsverhältnissen orientierten naturalistischen Designs trotz weniger ausgereifter Qualitätsstandards und geringerer Standardisierung besonders geeignet.

Methodisch ist besonderes Augenmerk auf die Untersuchung der Heterogenität zu legen, ggf. ist ein Pooling nicht möglich, sodass Ergebnisse stratifiziert zu berichten sind. Daten aus Beobachtungsstudien und RCTs sollten nicht zusammengefasst, sondern getrennt berichtet werden. Oft ist es auch geboten, Fall-Kontroll- und Kohortenstudien getrennt zu poolen und zu berichten (Egger, Smith et al. 2001).

7.3.1.3.2 Metaanalysen zu Prävalenz- und Inzidenzstudien

Eine Zusammenfassung von Prävalenz- und Inzidenzdaten ist mit dem üblichen metaanalytischen Instrumentarium möglich. Metaanalysen zu Prävalenz- bzw. Inzidenzdaten sind schon aufgrund der selteneren Durchführung zugrunde liegender Inzidenz- bzw. Prävalenzstudien geringer verbreitet. Bei den Analysen ist zu berücksichtigen, dass Inzidenzen, kumulative Inzidenzen oder Prävalenzen nahezu regelmäßig altersabhängig sind und oft von zahlreichen weiteren Parametern beeinflusst werden, die in die Auswertung einzubeziehen sind, um Verzerrungen zu vermeiden (wie beispielsweise Geschlecht, geographische Lage). Beispielhaft für eine Metaanalyse von Inzidenzdaten kann auf die Studie von Kotsopoulos et al. zur Inzidenz epileptischer Anfälle bzw. von Krampfanfällen hingewiesen werden (Kotsopoulos, van Merode et al. 2002).

7.3.1.3.3 Metaanalysen zu diagnostischen Studien

Diagnostische Studien weisen methodische Besonderheiten auf, sodass auch Metaanalysen diagnostischer Studien über die Standardmethodik hinausgehen. Diagnostische Studien werden fast ausschließlich als Querschnittsstudien zur Bestimmung der Testgüte (Sensitivität und Spezifität) durchgeführt und seltener als sog. diagnostische RCTs (v. a. zur Bestimmung des Einflusses des diagnostischen Tests auf die klinische Entscheidungsfindung und auf patientenrelevante Endpunkte). Unter diagnostischen RCTs werden Studiendesigns verstanden, bei denen primär ein diagnostischer Test zusammen mit einer anschließenden (therapeutischen) Intervention durchgeführt wird. Häufig erfolgt die Randomisierung nach dem diagnostischen Test, jedoch sind auch Studiendesigns mit randomisierter Zuweisung verschiedener diagnostischer Tests selbst möglich (Knottnerus 2002; Fearon, Tonino et al. 2007). Zielsetzung ist dabei die

Evaluation des Effekts der verbesserten diagnostischen Testgüte auf patientenrelevante Endpunkte (Knottnerus 2002; Knottnerus, van Weel et al. 2002; Lijmer and Bossuyt 2009).

In diagnostischen Studien erfolgt ein Vergleich des zu prüfenden diagnostischen Verfahrens mit einem Referenzstandard, der jedoch nicht selten diagnostisch unzulänglich ist, ohne dass dies exakt quantifiziert werden könnte. Häufig wird eine Fallzahlplanung nicht durchgeführt und fehlt eine Abschätzung der statistischen Power der Studie. Es werden jeweils zwei Parameter des Prüfverfahrens ermittelt, die Sensitivität und die Spezifität. Bei kontinuierlichen Messgrößen wie laborchemischen Untersuchungen ergibt sich darüber hinaus die Klassifikationsfrage, ab welchem Wert ein Test als positiv bzw. als negativ einzustufen ist, bzw. die Frage nach der Wahl des richtigen Schwellenwerts.

Weitere Besonderheiten ergeben sich aus der Tatsache, dass die Testcharakteristiken in mehr oder weniger großem Umfang von der Zusammensetzung der Population, in der diese untersucht werden, beeinflusst werden können. Zahlreiche Testverfahren unterliegen darüber hinaus subjektiven Einflüssen mit einer relevanten Variabilität zwischen einzelnen Untersuchern (sog. „interobserver variability“) und sogar zwischen wiederholten Beurteilungen des gleichen Untersuchers („intraobserver variability“), wie beispielsweise bei der Beurteilung von Ultraschall- und Röntgenbildern oder pathologischen Biopsiebefunden. Verzerrungen können durch unterschiedliche Behandlung der Gruppen mit den verschiedenen Tests (bzw. Teststrategien) entstehen oder auch durch zusätzliche Patienteninformationen (Symptome), die eine Diagnosewahrscheinlichkeit unabhängig vom Test beeinflussen können. Insbesondere beeinflusst und u. U. verzerrt werden diagnostische Tests mit subjektiver Beurteilungskomponente durch zusätzliches Vorwissen (entsprechend einem zusätzlichen diagnostischen Test), wie z. B. Schmerzangaben eines Patienten (observer bias).

Als Ergebnisparameter werden neben Sensitivität und Spezifität als typische Testcharakteristika auch Likelihood Ratios oder Diagnostic Odds Ratios (DOR: Odds der Sensitivität dividiert durch das Odds von 1-Spezifität) berichtet. Zum Pooling der Daten sind besondere statistische Verfahren erforderlich. Die Verwendung sog. ROC-Kurven (Darstellung von Sensitivität und Spezifität der einzelnen Studien in einem Streudiagramm) ermöglicht Aussagen zur gesamten Testgenauigkeit, da Sensitivität und Spezifität eines Tests über den angesetzten Schwellenwert indirekt voneinander abhängig sind. Ansätze zur metaanalytischen Zusammenfassung und Quantifizierung der Testgenauigkeit beinhalten die Berechnung gewichteter Mittelwerte von Sensitivität und Spezifität (sog. „naives Pooling“), die Ermittlung der Fläche unter der ROC-Kurve, Regressionsanalysen nach logarithmischer Transformation der Wertepaare anhand linearer oder hierarchischer Regression sowie den Einsatz bivariater Meta-Regressionsmodelle (Knottnerus 2002; Reitsma, Glas et al. 2005).

Bei der Erstellung der Metaanalyse ist ein besonderes Augenmerk auf die Vergleichbarkeit der Studienpopulationen in den Primärstudien, auf die Vergleichbarkeit der Tests bzw. Teststrategien und die spezifischen Arten von Bias zu legen (Deeks 2001; Knottnerus, van Weel et al. 2002; Jones and Athanasiou 2009).

7.3.1.3.4 Metaanalysen zu prognostischen Studien

Prognostische Studien umfassen klinische und epidemiologische Studien zu Prädiktoren des Krankheitsverlaufs und zu Risiko- bzw. ätiologischen Faktoren. Untersucht wird der Vorhersagewert bestimmter Variablen. Bei der metaanalytischen Analyse ergeben sich häufig Schwierigkeiten bei der Analyse der Zielparameter, da sehr unterschiedliche Outcomes berichtet werden. So werden häufig Ereigniszeiten, Odds- oder Hazard Ratios berichtet. Kategorisierungen erfolgen meist anhand von zwischen den Studien unterschiedlichen Schwellenwerten. Auf diese Weise kann eine quantitative Synthese aufgrund von fehlenden Informationen oder unterschiedlichen Ergebnisparametern unmöglich werden. Darüber hinaus können starke Korrelationen zwischen einzelnen Prädiktor-Variablen eine Auswertbarkeit einschränken. Im Vergleich zu RCTs ist von einem höheren Anteil unveröffentlichter Studien auszugehen (Publication Bias). Insgesamt zeigt sich, dass Metaanalysen prognostischer Studien aufgrund der Qualität der Primärstudien, aufgrund methodischer Probleme und stark uneinheitlichen Studiendesigns unbefriedigend sein können (Altman 2001).

7.3.1.4 Praktischer Teil

7.3.1.4.1 Datenextraktion

Wenn eine Metaanalyse geplant ist, muss dies bereits bei der Datenextraktion berücksichtigt werden. Je nachdem, welche Zielgrößen (dichotom, kontinuierlich, ordinal, Überlebenszeit etc.) analysiert werden sollen, mit welchen Studiendesigns die Daten erhoben wurden (RCT, Beobachtungsstudien) und welche Metaanalyse-Modelle herangezogen werden sollen, müssen die entsprechenden Daten aus den Studien extrahiert und zum Teil, soweit dies möglich ist, auch berechnet werden. Grundsätzlich sollten dabei die Daten zur Effektivität (z. B. Anzahl der Ereignisse) für jeden Studienarm erhoben werden, sowie ggf. Variablen, die zur Interpretation, zur Durchführung von Subgruppenanalysen oder zur weiterführenden Untersuchung eventueller Heterogenität von Bedeutung sein können (z. B. Geschlechterverhältnis, Confounding-Variablen etc.). Ersatzweise können auch bereits in der Studie berechnete Effektmaße (mit dazu gehörigem Standardfehler) extrahiert und ausgewertet werden.

Aus den extrahierten Daten werden Effektmaße berechnet, wobei für die unterschiedlichen Skalenniveaus der Daten unterschiedliche Effektmaße berechnet werden können:

- » **Risikodifferenz:** Diese wird als Differenz zwischen den relativen Häufigkeiten der Zielgröße zwischen den untersuchten Gruppen berechnet.
- » **Relatives Risiko:** Hier werden die relativen Häufigkeiten der Zielgröße der beiden untersuchten Gruppen ins Verhältnis gesetzt.
- » **Odds Ratio** (Chancen- oder Quoten-Verhältnis): Beim Odds Ratio werden zuerst die relativen Häufigkeiten der Zielgröße in den jeweiligen Gruppen in Odds umgerechnet. Dabei werden die entsprechenden Wahrscheinlichkeiten ins Verhältnis zu ihren Gegenwahrscheinlichkeiten gesetzt. Die beiden Odds werden dann gleichfalls ins Verhältnis gesetzt.
- » **Mittelwertdifferenz (Differenz der Mittelwerte):** Gemessen wird die absolute Differenz des Mittelwerts einer Zielgröße zwischen zwei Gruppen. Dieses Effektmaß kann als Summenmaß für Metaanalysen verwendet werden, wenn die Zielgrößen in allen Studien gleich skaliert sind, also beispielsweise die Lebensqualität immer mit dem gleichen Fragebogen (Messinstrument) gemessen wurde.
- » **Die standardisierte Differenz der Mittelwerte** ist die Differenz der Mittelwerte einer Zielgröße zwischen zwei Gruppen, dividiert durch die Standardabweichung für die Zielgröße bei den Studienteilnehmern. Sie wird verwendet, wenn die Zielgröße in verschiedenen Skalen gemessen wird, also z. B. die Lebensqualität mit unterschiedlichen Fragebögen (Messinstrumenten) erhoben wurde.
- » **Hazard Ratio:** Hazard Ratios werden aus den ereignisfreien Zeiten in beiden Gruppen ermittelt.

Bei dichotomen Daten können als Effektmaße die Risikodifferenz, das relative Risiko oder das Odds Ratio verwendet werden, bei stetigen Daten Mittelwertdifferenz und standardisierte Mittelwertdifferenz sowie bei Überlebenszeitdaten das Hazard Ratio. Dabei bezeichnet der Begriff „Überlebenszeitdaten“ allgemein die Situation, wenn die Zeit bis zum Eintritt eines bestimmten Ereignisses von Interesse ist. Häufig werden die Effektmaße auch in logarithmierter Form verwendet, um damit günstigere statistische Eigenschaften für weitere Auswertungen auszunützen.

Ausführliche Hinweise zur Datenextraktion und insbesondere auch dazu, wie nicht direkt berichtete Größen teilweise aus anderen berechnet werden können, finden sich im Cochrane-Handbuch Version 5.1.0 im Unterkapitel 7.7 „extracting study results and converting to the desired format“ (Deeks, Altman et al. 2001; Higgins and Green 2011).

7.3.1.4.2 Präsentation der Ergebnisse

7.3.1.4.2.1 Zielsetzungen der Reporting Guidelines

Empfehlungen

In Übereinstimmung mit dem „PRISMA Statement“ sollte eine Metaanalyse den Selektionsprozess und die Charakteristika eingeschlossener Studien beschreiben, das Risiko für Bias innerhalb der Studien und über die Studien hinweg abschätzen sowie die Ergebnisse der einzelnen Studien, der Synthese der Effekte und ggf. zusätzlicher Analysen vollständig und nachvollziehbar berichten.

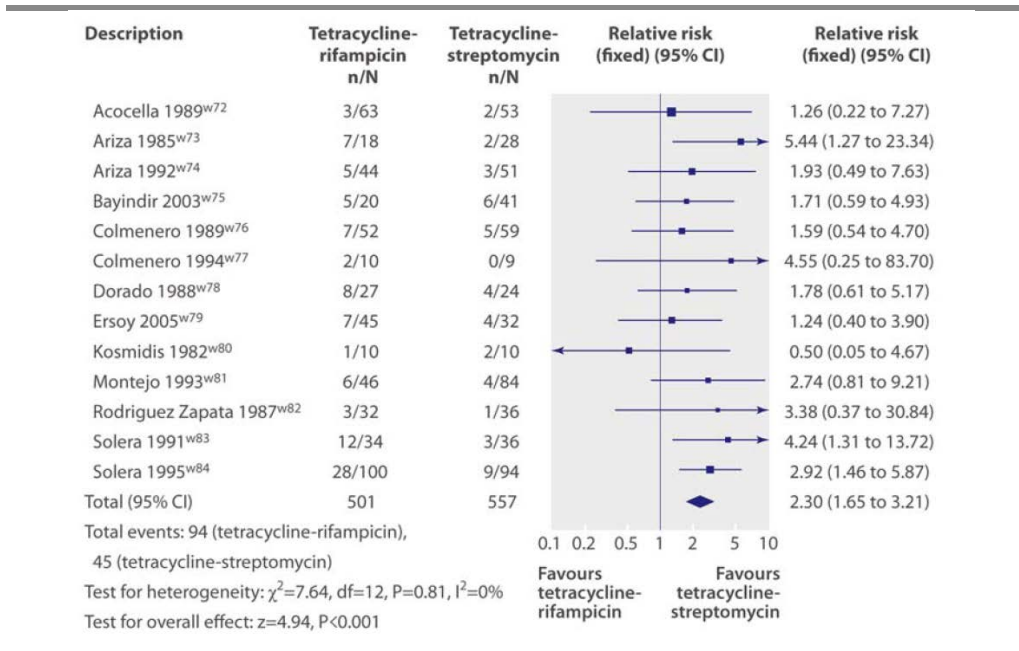
Konkret ist aufzuschlüsseln, wie viele Studien identifiziert, als geeignet betrachtet und eingeschlossen wurden. Ausschlüsse sind zu begründen. Zur Darstellung wird die Verwendung von Flussdiagrammen empfohlen. Wesentliche Charakteristika eingeschlossener Studien wie die Studiengröße, Follow-up-Zeiten, Patientenpopulationen mit Alter und Geschlecht, Ein- und Ausschlusskriterien sollten tabellarisch zusammengestellt werden. Die Abschätzung möglicher Verzerrungen unter Berücksichtigung von Design und Zielsetzung der Primärstudien wie auch auf der Ebene der Metaanalyse (Publication Bias, selektive Ergebnispräsentation, Overreporting) ist strukturiert und systematisch durchzuführen und transparent zu beschreiben. Ergebnisse einzelner Studien sollten zumindest in einer einfachen Zusammenfassung wiedergegeben werden. Für alle untersuchten Zielgrößen sind gruppenspezifische Ergebnisse mit Stichprobengrößen wie auch Interventionseffekte (OR, relative Risiken, Risikodifferenz) mit den entsprechenden Varianzschätzern bzw. Konfidenzintervallen anzugeben und der gepoolte Schätzer zu berichten. Eine Präsentation in grafischer Form mittels Forest Plot (s. Tabelle 7.3) ist üblich. Anzugeben sind weiters die Ergebnisse der Analyse bestehender Heterogenität. Ergänzende Untersuchungen, wie Sensitivitätsanalysen, Subgruppenanalysen oder Meta-Regressionsanalysen, sind im Einzelfall aufschlussreich.

7.3.1.4.2.2 Grafische Darstellungsmöglichkeiten

Forest Plot

Der Forest Plot stellt die wesentlichen Ergebnisse der eingeschlossenen Studien mit Punktschätzer und Konfidenzintervallen zusammen mit dem gepoolten Schätzer übersichtlich in einer Grafik zusammen. Tabelle 7.3 zeigt ein Beispiel für einen Forest Plot.

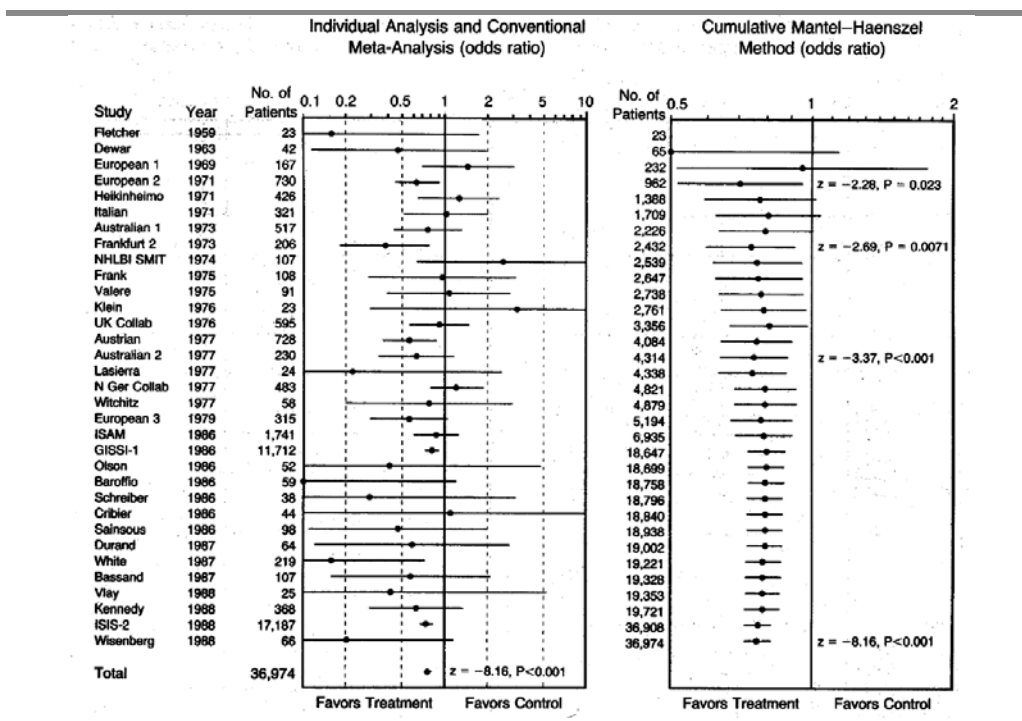
Tabelle 7.3:
Beispiel für einen Forest Plot



Quelle und Darstellung: Liberati et al. (2009)

Bei der Durchführung kumulativer Metaanalysen wird der Forest Plot durch eine zusätzliche Grafik ergänzt. Zu den zeitlich aufsteigend angeordneten Primärstudien wird in einem weiteren Forest Plot jeweils der gepoolte Schätzer angegeben, der für jede weitere eingeschlossene Studie neu berechnet wird. Aufgrund der zunehmenden Gesamtfallzahl der eingeschlossenen Studien steigt die Präzision des Schätzers sukzessive an und werden die Konfidenzintervalle entsprechend enger. Tabelle 7.4 zeigt ein Beispiel für eine kumulative Metaanalyse. Diese wurde von Lau et al. durchgeführt und demonstriert, dass retrospektiv gesehen bereits zu einem sehr frühen Zeitpunkt ein signifikanter durchschnittlicher Effektschätzer belegt gewesen wäre (Lau, Antman et al. 1992).

Tabelle 7.4:
Beispiel für eine kumulative Metaanalyse



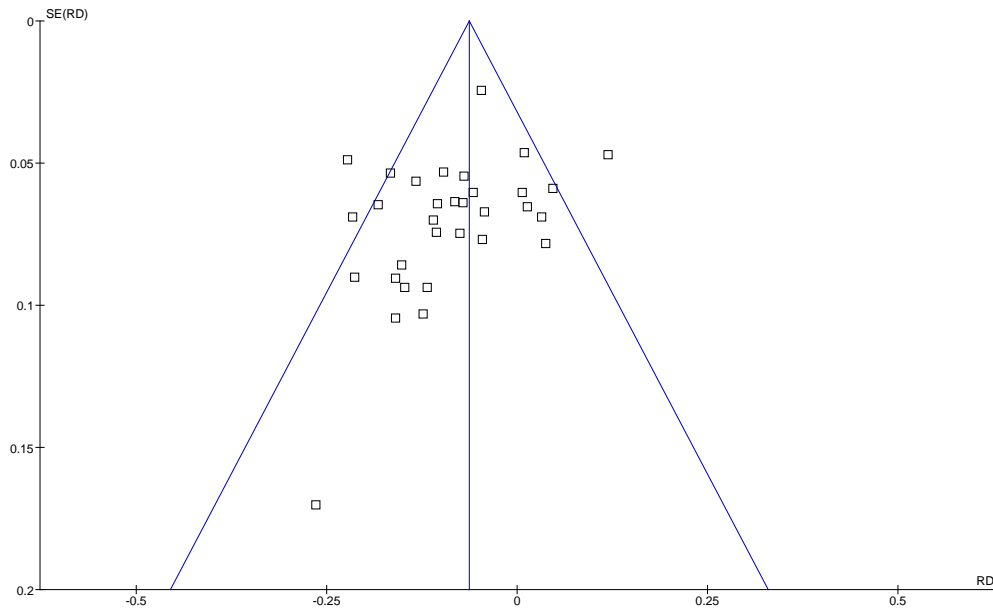
Quelle und Darstellung: Lau et al. (1992)

Funnel Plot

Funnel Plots (Trichter-Grafiken) sind Streudiagramme mit Darstellung der Größe des Studieneffekts (auf der horizontalen Achse) gegen ein Maß der Studiengröße (häufig Standardfehler bzw. dessen Kehrwert). Funnel Plots dienen v. a. der Abschätzung des Ausmaßes von Publication Bias. Im (unverzerrten) Idealfall stellt sich die Grafik als symmetrischer, umgestülpter Trichter dar. Mit Zunahme der Studiengröße steigt die Präzision des Effekts, daher ist die Streuung im unteren Bereich der Grafik (geringere Studiengröße) größer. Stärkere Asymmetrien insbesondere in den unteren Bereichen mit geringeren Studiengrößen weisen auf Publication Bias bzw. verzerrte Effektschätzungen kleinerer Studien hin, ohne diese zu beweisen. So können auch Subgruppeneffekte (bzw. Heterogenitätseffekte), methodische Mängel der Studien, Artefakte oder Zufallseinflüsse Asymmetrien hervorrufen (Sterne, Egger et al. 2001; Liberati, Altman et al. 2009). Abbildung 7.4 zeigt ein Beispiel für einen Funnel Plot. Dieser zeigt eine leichte Asymmetrie mit einem Schwerpunkt bei größeren Therapieeffekten (v. a. bei den unpräziseren, d. h. kleineren Studien), sodass vermutet werden könnte, dass hier Studien, die keinen oder sogar einen negativen (schädlichen) Effekt zeigten, seltener

publiziert wurden. Dieser Effekt könnte als leichtes Publication Bias interpretiert werden, ohne dass dieses Ergebnis zwingend wäre.

Abbildung 7.4:
Beispiel für einen Funnel Plot



Quelle und Darstellung: adaptiert nach Göhler et al. (2006)

7.3.1.4.3 Software

Es existieren zahlreiche verschiedene Softwarepakete unterschiedlichen Umfangs, die unterschiedliche Anforderungen erfüllen und unterschiedlichen Einarbeitungsaufwand bzw. Programmierkenntnisse erfordern. Für einfache Metaanalysen reichen oft schon frei verfügbare Programme.

Für die Durchführung von Metaanalysen von Interventionsstudien hat sich das frei verfügbare Softwareprogramm der Cochrane Collaboration Review-Manager als wesentlichstes Hilfsmittel etabliert. Für Metaanalysen diagnostischer Studien ist die gleichfalls frei verfügbare Software MetaDisc eine wertvolle Hilfe. Für statistisch anspruchsvollere Fragestellungen wird sehr häufig das Softwarepaket STATA herangezogen, dessen Stärke in der Ausgabe hochwertiger Grafiken liegt. Hier sind zahlreiche Zusatzprogramme für speziellere Zwecke verfügbar und in einem gesonderten Handbuch erläutert (Sterne 2009). Spezielle Fragestellungen (z. B. hierarchische Regressionsmodellierung bei diagnostischen Metaanalysen) erfordern ggf. den Einsatz der

Statistikprogramme SAS, SPSS oder R, für die auch einige zusätzliche Metaanalyse-Routinen verfügbar sind. In diesem Zusammenhang sind insbesondere die grafischen Darstellungsmöglichkeiten von R herauszustellen. Bayesianische Metaanalysen verwenden das Standardprogramm WinBUGS.

Eine auszugsweise Übersicht möglicher Software findet sich in Tabelle 7.5.

Tabelle 7.5:
Übersicht verfügbarer Software zur Durchführung von Metaanalysen

Name	Aktuelle Version *	Link	Preispolitik	Schwerpunkt	Besondere Vorteile
Review Manager	5.0.255	http://www.cc-ims.net/revman	frei verfügbar	Cochrane Reviews	Einfache Bedienung
Comprehensive Meta-Analysis	2.2.050	http://www.meta-analysis.com	\$ 1.295 Corporate \$ 795 Academic \$ 395 Student	Multiple Studien	Verschiedene Studiendesigns können in einer Analyse kombiniert werden
Meta-Analysis for STATA	(2009)	http://www.stata.com/support/faqs/stat/meta.html	frei verfügbar	-	Ständige Weiterentwicklung, Integration in STATA
Meta-Analysis for R	1.60	http://cran.r-project.org/web/packages/meta/index.html	frei verfügbar	-	Integration in R
Excel - MIX Addin	2.0.0.9	http://www.meta-analysis-made-easy.com/	\$ 210 Normal \$ 155 Academic \$ 75 Student	-	Integration in Office, Bearbeitung in Excel
MetaDisc	1.4	http://www.hrc.es/investigacion/metadisc_en.htm	frei verfügbar	Diagnostische Metaanalysen	-
SPSS, STATA & SAS Makro	(2002)	http://mason.gmu.edu/~dwilsonb/ma.html	frei verfügbar	-	Integration in SPSS, SAS, STATA

* Stand: Oktober 2010; ** Kostenfrei ist lediglich das Metaanalyse-Modul, die zugrunde liegende Statistik-Software ist jedoch kostenpflichtig.
Anmerkung: Historisch gesehen sind noch die DOS-basierten Programme Meta-Analysis by Ralf Schwarzer (http://userpage.fu-berlin.de/~health/meta_e.htm) und Meta-Analysis Easy to Answer (META, <http://davidakenny.net/meta.htm>) erwähnenswert, die heute jedoch kaum noch praktische Relevanz besitzen dürften.

Quelle: UMIT 2011; Darstellung: GÖG/BIQG

7.3.1.5 Ausblick

Die hohe Bedeutung von Metaanalysen zur Evidenzsynthese ist unbestritten und wird in Zukunft aller Wahrscheinlichkeit nach weiterhin zunehmen. Neue Methoden, wie beispielsweise der Einsatz indirekter Vergleiche, sind derzeit in Entwicklung bzw. noch nicht generell verbreitet und in allen Bereichen akzeptiert. Eine der größten Einschränkungen der Metaanalyse besteht im potenziellen Publication Bias, für das wissenschaftliche Verlage mit in die Verantwortung zu nehmen sind, um eine selektive Veröffentlichung von Studien mit statistisch signifikant nachgewiesenen Effekten zu vermeiden. Weitere Lösungsansätze, insbesondere die Etablierung verpflichtender Studienregister, werden noch einige Zeit in Anspruch nehmen und sind im Wesentlichen für Interventionsstudien bedeutsam. In diese Register werden zwar hohe Erwartungen gesetzt, ob diese gerechtfertigt sind, werden jedoch erst künftige Erfahrungen zeigen.

7.3.2 Entscheidungsanalytische Nutzenmodellierung

Eine weitere Form der systematischen und quantitativen Evidenzsynthese ist die medizinische Entscheidungsanalyse. Diese wird zur Bestimmung des medizinischen Nutzens eingesetzt (Nutzenmodellierung). Die medizinische Entscheidungsanalyse umfasst die Modellierung von Kurz- und Langzeitoutcomes zum medizinischem Nutzen und schließt dabei auch die Bestimmung des Zusatznutzens (im Vergleich zu anderen Technologien, wie z. B. Standard of Care) und der Nutzen-Schaden-Balance ein (Weinstein, Fineberg et al. 1980). Die Entscheidungsanalyse basiert auf dem Prinzip der Nutzenmaximierung. Der Nutzen ist als der patientenrelevante Gesamtnutzen einer Intervention zu verstehen und schließt sämtliche positiven und negativen, kurz- und langfristigen Wirkungen der Interventionen auf Lebenszeit, Morbidität und Lebensqualität ein. Während die Metaanalyse die Synthese der Ergebnisse gleichartiger Studien (z. B. randomisierter Wirksamkeitsstudien) darstellt, erfordert die entscheidungsanalytische Nutzenmodellierung die Kombination von empirischen Daten aus unterschiedlichen Studientypen (z. B. die Synthese aus Wirksamkeitsstudien, epidemiologischen Studien, Studien zur Sicherheit, Lebensqualitätsstudien). Entscheidungsanalytische Modellierungen werden einerseits in der reinen Nutzenbewertung eingesetzt, häufiger jedoch noch in der gesundheitsökonomischen Evaluation, d. h. dann, wenn eine Abwägung zwischen dem gesundheitlichen Zusatznutzen und den Zusatzkosten einer Technologie erfolgt (Siebert 2003a). Die Modellierung aller gesundheitlichen Konsequenzen einer Technologie bildet die Basis der gesundheitsökonomischen Bewertung. Die Vorgehensweise bei der Modellierung sowie die eingesetzten Methoden werden daher im Kapitel 8.3 „Gesundheitsökonomische Modellierung“ ausführlich beschrieben. Es soll hier jedoch betont werden, dass die entscheidungsanalytische Modellierung auch dann eine geeignete Methode darstellt, wenn die Kosten nicht von Interesse sind und

lediglich ein Vergleich der gesundheitlichen Konsequenzen erfolgen soll (wie beispielsweise der Effektivität und der Nebenwirkungen verschiedener Technologien) (Siebert und Kurth 2000, Hunink, Glasziou et al. 2001). Im vorliegenden Kapitel werden Situationen dargestellt, in denen der Einsatz der Entscheidungsanalyse unumgänglich oder besonders wertvoll ist. Außerdem werden integrierte Nutzenmaße vorgestellt, die sich für die Zusammenfassung der Auswirkungen einer Technologie in einer einzigen Größe eignen, auch wenn diese in unterschiedlichen Dimensionen wie Lebenszeit und Lebensqualität liegen.

Ein entscheidungsanalytisches Modell ist eine formale, transparente und quantitative Methode, die Konsequenzen von Interventionen unter Unsicherheit systematisch abzuschätzen (Weinstein, Fineberg et al. 1980; Hunink, Glasziou et al. 2001; Siebert 2003). Unsicherheit über die Bewertung von Interventionen kann z. B. dann bestehen, wenn verschiedene Endpunkte klinischer Studien wie Testgüte, Ansprechraten, Mortalität und Lebensqualität zusammengefasst werden müssen, um den Gesamtnutzen der Interventionen zu ermitteln. Oder aber es liegen direkte Vergleiche aus klinischen Studien nur für Kurzzeitendpunkte vor und patientenrelevante Auswirkungen der Interventionen fallen zum Teil erst zu einem späteren Zeitpunkt an, so dass eine Synthese der Evidenz aus Studien mit kurzem und längerem Zeithorizont erforderlich ist. Der Vergleich von Interventionen kann auch dann ein Modell erfordern, wenn das Abwägen zwischen Nutzen und Risiken komplex ist. Zudem kommen entscheidungsanalytische Modellierungen bei Fragestellungen zum Einsatz, die anhand empirischer Studien nicht, nicht im benötigten Zeitrahmen oder nur mit unverhältnismäßig großem Aufwand zu beantworten wären.

Auf der anderen Seite ist zu betonen, dass in jedes Modell Annahmen eingehen; dies gilt auch für entscheidungsanalytische Modelle. Diese Annahmen können in Parameterwerten liegen, für die keine vollständigen empirischen Daten vorliegen, oder aber auch in strukturellen Annahmen, z. B. bei der Festlegung der relevanten Gesundheitszustände und Ereignisse oder der Extrapolation der Effekte einer Behandlung über eine randomisierte klinische Studie hinaus. In Sensitivitätsanalysen muss untersucht werden, ob diese Annahmen entscheidenden Einfluss auf die Ergebnisse einer Studie haben. Ist das Ergebnis auch bei Variation der Annahmen robust, erhöht das das Vertrauen in die Schlussfolgerungen der Studie. Zeigt sich, dass das Ergebnis stark von den Annahmen beeinflusst wird, ist unter Umständen weitere Forschung in Bezug auf diese Annahmen und Parameter notwendig, um die empirische Basis zu verbessern. Die Methoden der Sensitivitätsanalyse werden im Kapitel 8.3.4 ausführlich beschrieben. Wichtig ist auch die Überprüfung der strukturellen und numerischen Annahmen auf Plausibilität durch klinische Experten. Siehe dazu auch das Kapitel 8.3.2.7 zur Validierung von Modellen, in dem auch eine weitere wichtige Prüfung, die externe Validierung, beschrieben wird.

Im folgenden Abschnitt werden Situationen bzw. beispielhafte Fälle beschrieben, in denen eine entscheidungsanalytische Nutzenmodellierung sinnvoll sein kann: Der Abschnitt folgt der Darstellung in (Siebert 2003 c) und führt u. a. die für das Methodenhandbuch relevanten Teile aus diesen Publikationen an..

7.3.2.1 Gründe für die Datensynthese in Form eines medizinischen entscheidungsanalytischen Modells

7.3.2.1.1 Synthese von Kurz- und Langzeitergebnissen

Bei manchen Erkrankungen vergehen Jahre oder Jahrzehnte bis zum Auftreten nennenswerter gesundheitlicher Beeinträchtigungen. Ein Beispiel ist die chronische Hepatitis C, die zunächst keine Symptome zeigt, aber nach Jahrzehnten zu Leberzirrhose und Leberkrebs führen kann und dann mit erhöhter Mortalität und reduzierter Lebensqualität einhergeht. In diesen Fällen erfassen klinische Studien meist nur intermediäre Endpunkte, die als Surrogat für patientenrelevante Endpunkte wie Langzeitmorbidity und Mortalität dienen. Bei der antiviralen Therapie von Patienten mit chronischer Hepatitis C ist das Lanzeitansprechen (engl. „Sustained Virological Response“, Abk.: SVR) ein solcher intermediärer Endpunkt. Die Ergebnisse epidemiologischer Studien, häufig Beobachtungsstudien, die Auskunft über die langfristige Entwicklung einer Krankheit geben, können in einem Modell mit den Ergebnissen aus klinischen Studien kombiniert werden, um die langfristigen Folgen von Interventionen abzuschätzen (Siebert 2003c).

7.3.2.1.2 Zusammenfassen verschiedener Zielgrößen

Die Entscheidung zur Einführung einer Intervention beinhaltet oft ein Abwägen erwünschter und unerwünschter Auswirkungen auf Lebenserwartung und Lebensqualität (s. folgenden Abschnitt 7.3.2.2). Alle Wirkungen einer Intervention auf die Lebensdauer und auf die Lebensqualität können in einem entscheidungsanalytischen Modell gemeinsam erfasst und verknüpft werden. Für das systematische quantitative Abwägen ist es hilfreich, ein Nutzenmaß zu definieren, das sowohl den Nutzen als auch den Schaden abbildet. Ein solches kombiniertes Nutzenmaß ist das qualitätsadjustierte Lebensjahr (engl. Quality-adjusted life-year, Abk.: QALY) (s. Abschnitt 7.3.2.2). (Siebert 2003c) Es sei hier hervorgehoben, dass der Einsatz dieses Nutzenmaßes auch dann sinnvoll ist, wenn keine gesundheitsökonomische Evaluation durchgeführt wird und ausschließlich Nutzen und Schaden abgewogen und modelliert werden sollen. Insbesondere, wenn die innerhalb einer Indikation zu vergleichenden Interventionen verschiedene Komponenten der Gesundheit der Patienten beeinflussen, ist der Endpunkt QALY dazu geeignet, die Evidenz zu diesen Komponenten in einem Maß zusammenzuführen (IQWiG 2009).

7.3.2.1.3 Verknüpfung der Ergebnisse aus diagnostischen Studien mit Ergebnissen aus Therapiestudien

Insbesondere die Beurteilung medizinischer Testverfahren erfordert häufig ein Modell, um Daten zur diagnostischen Genauigkeit der Tests mit Daten zu den erwünschten und unerwünschten Folgen einer infolge der Testergebnisse durchgeführten Behandlung zu verbinden. Die Ergebnisgrößen diagnostischer Studien sind meist Spezifität, Sensitivität und negativer und positiver Vorhersagewert. Um den patientenrelevanten Gesamtnutzen eines Tests abzuschätzen, muss sowohl der Nutzen der Behandlung tatsächlich Erkrankter (richtig-positiv) als auch der Schaden für fälschlich behandelte nicht erkrankte Personen (falsch-positiv) oder für fälschlich nicht behandelte erkrankte Personen (falsch-negativ) einbezogen werden. Ebenso müssen die Konsequenzen weiterer, möglicherweise invasiver Tests berücksichtigt werden. Der Gesamtnutzen ist zudem abhängig von der Prävalenz der Erkrankung in der Zielpopulation (Siebert 2003c). Mithilfe eines Modells kann anschließend auch der optimale Schwellenwert eines Tests hinsichtlich des Gesamtnutzens bestimmt werden. Sensitivität und Spezifität werden so gegeneinander abgewogen, dass die langfristigen Auswirkungen auf Lebenserwartung und Lebensqualität optimiert werden (Trikalinos, Siebert et al. 2009). Modellierung bei diagnostischen Tests kann auch dann eingesetzt werden, wenn nicht alle relevanten Kombinationen von diagnostischen Tests in klinischen Studien getestet werden können. Durch Modellierung lassen sich dagegen beliebig viele Testverfahren und Testkombinationen sowie unterschiedliche Screeningstrategien (z. B. in Abhängigkeit von Alter, Screeningintervallen, Risikofaktoren etc.) miteinander vergleichen (Siebert 2003c).

7.3.2.1.4 Untersuchung des Einflusses von Routinebedingungen

Kontrollierte klinische Studien finden meist unter künstlichen klinischen Bedingungen statt und spiegeln häufig nicht die Routinesituation wider. Ein entscheidungsanalytisches Modell kann hier Aspekte der Routinesituationen aufnehmen und ihren Einfluss auf den medizinischen Nutzen einer Intervention untersuchen. Ein Beispiel hierfür ist die Einbeziehung der Compliance der Patienten unter Routinebedingungen (Siebert 2003c).

7.3.2.1.5 Untersuchung des Einflusses von abweichenden situativen Bedingungen

Populationscharakteristika, Versorgungsmuster oder Nutzwerte für Krankheitszustände können je nach Land und Umfeld der Evaluation (z. B. ambulante oder stationäre Anwendung) unterschiedlich sein. Ein entscheidungsanalytisches Modell ermöglicht das Abschätzen des Einflusses konkreter Bedingungen, die von denen der klinischen Studie abweichen (Siebert 2003c).

7.3.2.1.6 Systematische Untersuchung der Unsicherheiten

Ein entscheidungsanalytisches Modell bietet die Möglichkeit der systematischen Analyse von Unsicherheiten. Sensitivitätsanalysen ermöglichen Aussagen darüber, welche Parameter das Ergebnis der Modellierung entscheidend beeinflussen und für welche Größen unzureichende Evidenz vorliegt. Liegen z. B. Daten über die kurzfristige Wirksamkeit eines Medikaments vor, jedoch nur unzureichende Evidenz über die Dauer der Wirkung, dann kann eine Entscheidungsanalyse systematisch verschiedene Szenarien für die Wirkungsabnahme untersuchen und die potenziellen Auswirkungen auf den Gesamtnutzwert der Intervention abschätzen (Siebert 2003c).

Bei vollkommen fehlender Evidenz in einem Bereich, zum Beispiel im Frühstadium der Entwicklung von Interventionen, kann ein Modell dazu dienen, Parameterbereiche anzugeben, die zu einer effektiven Intervention führen würden. Ein Beispiel wären die für einen patientenrelevanten Gesamtnutzen notwendige Sensitivität und Spezifität eines Tests, der sich noch in Entwicklung befindet (Siebert 2003c).

Im Rahmen multivariater Sensitivitätsanalysen kann die gesamte Unsicherheit der Modellergebnisse bestimmt werden (Briggs, Sculpher et al. 1994; Briggs 2000).

7.3.2.2 Qualitätsadjustierte Lebensjahre (QALY)

Der medizinische Nutzen einer Technologie kann in natürlichen klinischen Zielgrößen (z. B. Therapieansprechen, Reduktion der Inzidenz, verhinderte Ereignisse, gewonnene

Lebensjahre) oder nicht-klinischen Endpunkten (z. B. verhinderter Arbeitsausfall) oder kombinierten Endpunkten (z. B. qualitätsadjustierten Lebensjahren (QALYs) gemessen werden (Siebert und Kurth 2000, Siebert 2003a).

Klein und Koautoren (Klein et al. 2010) fassen die wichtigsten Aspekte wie folgt zusammen: „Die Lebensqualität kann mittels standardisierter Verfahren (z. B. Standard Gamble, Time-trade-off, visuelle Analogskala) direkt oder mit Hilfe sogenannter Indexfragebögen (z. B. SF-6D, EuroQoL etc.) indirekt gemessen werden“ (Siebert und Kurth 2000). Die so ermittelten Nutzwerte für bestimmte Gesundheitszustände gehen dann in die Kosten-Nutzen-Bewertung ein. Ein QALY wird durch die Multiplikation der Zeit, die eine Person in einem bestimmten Gesundheitszustand verbringt, mit dem Nutzwert dieses Gesundheitszustandes berechnet. Ein QALY entspricht somit einem Jahr in perfekter Gesundheit bzw. einem längeren Zeitraum in eingeschränkter Gesundheit.

Der Vorteil des QALY-Ansatzes liegt in der Möglichkeit einer Synthese der Bewertung von Lebenszeit und Lebensqualitätseffekten. Damit wird ermöglicht, verschiedenartigste tödliche und nicht tödliche Krankheits- oder Technologiefolgen mittels eines Maßes zu quantifizieren. Die Verwendung eines solchen integrativen Nutzenmaßes ist eine wichtige Voraussetzung für die systematische Abschätzung des Gesamtnutzens von Technologien, die nicht nur eine Wirksamkeit, sondern auch Risiken beinhalten. (Klein et al. 2010)

Klein und Kollegen führen das folgende Beispiel an: „Beispielsweise können bei der Nutzenbewertung einer aggressiven Chemotherapie therapiebedingte Gewinne an Lebenszeit mit therapiebedingten Verlusten an Lebensqualität über QALYs miteinander verrechnet werden. Zur umfassenden Information des Entscheidungsträgers sollten die medizinischen Konsequenzen einer Technologie aber nicht nur in aggregierter Form des QALYs, sondern stets auch disaggregiert in ihren natürlichen Einheiten berichtet werden. Dadurch wird transparent, in welchem Ausmaß sich eine Intervention auf die Lebenserwartung und auf die Lebensqualität auswirkt, was wiederum zur Erleichterung ethischer Abwägungen beiträgt.“ (Klein et al. 2010) .

Zur Verwendung des QALY in der Ressourcenallokation siehe auch Kapitel 8.4.2.9.

Autorinnen und Autoren: Petra Schnell-Inderst, Bernhard Bornschein, Annette Conrads-Frank, Nikolai Mühlberger, Gaby Sroczynski, Uwe Siebert (UMIT).

7.4 Stärke der Evidenz

Die Beantwortung einer wissenschaftlichen Frage in einem HTA muss Lesern auch einen Eindruck davon vermitteln, wie gut die zugrunde liegende Evidenz ist. Eine explizite Darstellung der Stärke (oder Qualität) der Gesamtheit der Evidenz wird auch vom „PRISMA Statement“ gefordert (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Moher et al. 2009), wobei es wichtig ist, zwischen Evidenzhierarchien von Studien und der Stärke der Evidenz zu unterscheiden. Eine umfassende Beurteilung der Stärke der Evidenz berücksichtigt nicht nur Unterschiede in Studiendesigns, sondern auch die interne Validität der einzelnen Studien, die Konsistenz der Ergebnisse zwischen den Studien, die Größe des ermittelten Effektes, die Präzision des Effektes, mögliche Confounder und die Übertragbarkeit der Ergebnisse.

Es gibt eine Vielzahl von Möglichkeiten und Systemen, die Stärke der Evidenz darzustellen (West et al. 2002). Nur wenige garantieren jedoch ein explizites und transparentes Vorgehen (Atkins et al. 2004b). Eine systematische und explizite Methode zur Beurteilung der Stärke der Evidenz verhindert Fehler und erleichtert die Reproduzierbarkeit und kritische Evaluierung der Entscheidungen durch Außenstehende (Lohr 2004). Der Vorteil ist nicht, dass subjektive Beurteilungen und Empfehlungen vermieden werden, sondern dass sie transparent werden.

Das folgende System zur Beurteilung der Stärke der Evidenz wurde von der GRADE Working Group entwickelt (Atkins et al. 2004a). Der GRADE-Ansatz beurteilt die Stärke der Evidenz endpunktspezifisch für eine bestimmte wissenschaftliche Fragestellung. Dies bedeutet, dass für jeden wesentlichen Endpunkt die Stärke der Evidenz neu ermittelt wird. Die Stärke der Evidenz kann sich dadurch aber auf Basis derselben Studien von Endpunkt zu Endpunkt auch ändern (z. B. geringe Stärke der Evidenz für Lebensqualität, aber hohe Stärke für Reduktion des Blutdrucks). Ein unverzichtbarer Schritt bei der Anwendung des GRADE-Systems ist daher immer die Priorisierung von Endpunkten. Patientenrelevante Endpunkte wie Mortalität oder Lebensqualität haben dabei höhere Priorität als Surrogatparameter (vgl. 4.1.2.2).

Das GRADE-Konzept zieht bei der Beurteilung der Stärke der Evidenz unterschiedliche Faktoren in Betracht, die in der Folge diskutiert werden und deren Auswirkung auf die Stärke der Evidenz in Tabelle 7.2 zusammengefasst ist (Brozek et al. 2009).

Tabelle 7.6:
Faktoren, die die Stärke der Evidenz beeinflussen

Faktoren, die die Stärke der Evidenz verringern
Beobachtungsstudien Erhöhtes Bias-Risiko Inkonsistenz der Resultate Fehlende Direktheit Unpräzise Resultate Publikationsbias
Faktoren, die die Stärke der Evidenz erhöhen
Große Effekte Dosis-Wirkungs-Gradient Alle plausiblen Confounder würden den beobachteten Effekt vermindern.

Quelle: Brozek et al. 2009; GÖG/BIQG-eigene Darstellung

Bias-Risiko

Die Beurteilung des Bias-Risikos bezieht sich sowohl auf das Studiendesign und die Anfälligkeit für systematische Fehler als auch auf methodische Mängel der Studien. Die Einbeziehung des Studiendesigns berücksichtigt, dass manche Studiendesigns eine geringere Anfälligkeit für systematische Fehler haben als andere. RCTs werden dabei höher bewertet als Beobachtungsstudien. Fehler in der Ausführung von Studien führen zu einem erhöhten Risiko, dass Resultate von Bias verzerrt werden (z. B. fehlerhafte Randomisierung, fehlende Verblindung, hohe Drop-out-Raten).

Konsistenz der Ergebnisse

Heterogene und nicht geklärte Unterschiede in Resultaten einzelner Studien reduzieren das Vertrauen in die Richtigkeit eines einzelnen Studienergebnisses. Eine gewisse Heterogenität der Ergebnisse wird in einem HTA immer vorhanden sein. Dies ist einerseits auf klinische Heterogenität der Populationen in den Studien, andererseits auch auf Zufallsfehler zurückzuführen. Wenn die Heterogenität von Resultaten nicht erklärt werden kann, insbesondere wenn Ergebnisse konträr sind (Resultate in manchen Studien zeigen Nutzen, in anderen Studien keinen Nutzen oder sogar Schaden), ist das Vertrauen in die Ergebnisse gering.

Direktheit

Direktheit bezieht sich auf die Übertragbarkeit der Studienergebnisse auf andere Populationen beziehungsweise auf das Vorhandensein oder Nicht-Vorhandensein von

direkten Vergleichen. Indirekte Vergleiche von Interventionen haben ein höheres Bias-Risiko als direkte Vergleiche.

Präzision der Ergebnisse

Studien, die sehr wenige Patientinnen/Patienten einschließen oder Endpunkte mit wenigen Ereignissen untersuchen, haben häufig sehr breite Konfidenzintervalle, die sowohl wichtigen Nutzen als auch irrelevante Effekte mit einschließen können.

Zusätzliche Faktoren

Bei der Beurteilung der Stärke der Evidenz zieht das GRADE-Konzept weitere Faktoren in Betracht, wie das mögliche Vorhandensein von Publikationsbias, große oder sehr große Effekte, Dosis-Wirkungs-Beziehungen oder die Richtung der Verzerrungen durch Confounder.

Das GRADE-Konzept benutzt folgende Klassifizierungen und Definitionen, um die Stärke der Evidenz zu beurteilen (Schünemann 2009):

- » **Hoch:** Es ist unwahrscheinlich, dass weitere Forschung das Vertrauen in den beobachteten Behandlungseffekt verändert.
- » **Moderat:** Weitere Forschung wird sich vermutlich erheblich auf unser Vertrauen in den beobachteten Behandlungseffekt auswirken. Möglicherweise ändert sich der Behandlungseffekt.
- » **Niedrig:** Weitere Forschung wird sich sehr wahrscheinlich auf unser Vertrauen in den beobachteten Behandlungseffekt auswirken. Möglicherweise ändert sich der Behandlungseffekt.
- » **Sehr niedrig:** Der beobachtete Behandlungseffekt ist mit sehr großer Unsicherheit behaftet.

Eine kostenfreie Software zur Anwendung des GRADE-Systems kann unter <http://www.ims.cochrane.org/revman/gradepro> heruntergeladen werden.

Autor/in: Gerald Gartlehner, Kylie Thaler (Donau-Universität Krems)

8 Ökonomische Bewertung

8.1 Ökonomische Übersichtsarbeit

8.1.1 Hintergrund

Die steigende Nachfrage nach (zunehmend teuren) Leistungen bedingt, dass die Mittel im Gesundheitssystem knapper werden. Das bedeutet gleichzeitig: Wenn eine neue Intervention öffentlich finanziert wird und diese mehr Ressourcen beansprucht als die derzeitige Praxis, stehen diese für andere Leistungen nicht zur Verfügung. Die Rahmenbedingung der begrenzten Ressourcen bei gleichzeitig steigenden Kosten trägt dazu bei, dass neben der Bewertung des Nutzen-Risiko-Verhältnisses von Interventionen im Gesundheitssystem auch Unterstützung gefragt ist, wenn entschieden werden soll, wie die knapper werdenden Gelder sinnvoll – also möglichst effizient – eingesetzt werden können. Abgesehen davon, dass mit diesem ökonomischen Denken die Verschwendung öffentlicher Gelder verhindert werden soll, kann umgekehrt ein ineffizienter Ressourceneinsatz indirekt mit einem Schaden für die Gesellschaft und die einzelnen Individuen einhergehen. Auch der Ansatz mancher gesundheitspolitischer Akteure, dass bei zu wenig Ressourcen einfach der Geldtopf zu vergrößern sei, greift zu kurz, denn um die verfügbaren Ressourcen in einer Volkswirtschaft konkurrieren neben dem Gesundheitswesen auch andere gesellschaftliche Bereiche (z. B. Bildung, Soziales.), sodass das Gesundheitsbudget nicht endlos erhöht werden kann.

Bei der Frage, wie die Mittel innerhalb des Gesundheitssystems rational zu verteilen sind, bieten gesundheitsökonomische Studien Unterstützung. Darunter fallen gesundheitsökonomische Evaluationen (auch als vollständige gesundheitsökonomische Evaluationen oder vergleichende Studien der gesundheitsökonomischen Evaluation bezeichnet). Das sind Studien, die Kosten und Gesundheitseffekte von zwei oder mehreren „Technologien“ vergleichend analysieren. Das Ziel ist, die Kosten und die gesundheitlichen Konsequenzen der alternativen Interventionen zu identifizieren, zu messen und vergleichend zu bewerten (Drummond, Sculpher et al. 2005). Daneben gibt es aus dem Feld der Gesundheitsökonomie weitere Studien, die von Relevanz für Entscheidungsträger sind. Dazu gehören etwa Krankheitskostenstudien, die die volkswirtschaftlichen Auswirkungen von Erkrankungen abschätzen (in Abgrenzung zu den oben beschriebenen gesundheitsökonomischen Evaluationen auch als nicht-vergleichende gesundheitsökonomische Studien bezeichnet) oder Budgetauswirkungsanalysen (siehe Kapitel 8.2), die die direkten finanziellen Auswirkungen (Folgekosten bzw. -erlöse) bewerten, die aufgrund der Einführung bzw. Erstattung einer neuen Gesundheitstechnologie bzw. Änderung der Nutzung einer bestehenden Gesundheitstechnologie in einem Gesundheitssystem entstehen.

Da die gesundheitsökonomische Evidenz allein aufgrund der möglichen Studientypen meistens heterogen ist und die Übertragbarkeit von derartigen Studien vom Studienland auf den eigenen Entscheidungskontext häufig eingeschränkt ist, wurde die Sinnhaftigkeit einer gesundheitsökonomischen Übersichtsarbeit prinzipiell hinterfragt (Anderson 2009). Jedoch sind gerade in Ländern, in denen wenige länderspezifische gesundheitsökonomische Studien zur Verfügung stehen, Entscheidungsträger auf das anderswo generierte Wissen angewiesen. In vielen Fällen existiert zu einer Fragestellung/Technologie mehr als eine gesundheitsökonomische Studie, sodass der Bedarf besteht, diese oder eine begründete Auswahl einzelner Studien in einer systematischen Form transparent darzustellen. Außerdem kann eine Übersichtsarbeit wertvolle Information für die Durchführung einer eigenen Studie liefern und/oder unnötige Duplikationen (z. B. von Modellen) verhindern.

Das Kapitel beschreibt den derzeitigen Stand der Wissenschaft zur Erstellung einer gesundheitsökonomischen Übersichtsarbeit von der Definition der Fragestellung bis hin zur Darstellung der Studienergebnisse. Wo Kontroversen bestehen, werden alternative Vorgehensweisen optional angeführt.

8.1.2 Exkurs: Grundlagen gesundheitsökonomischer Evaluation

Die am häufigsten für Fragen der Ressourcenallokation eingesetzten gesundheitsökonomischen Studien sind gesundheitsökonomische Evaluationen. In der Literatur werden drei generische Typen von gesundheitsökonomischer Evaluation genannt, die sich darin unterscheiden, wie der Gesundheitseffekt von Interventionen dargestellt wird (s. Box 8.1) (Centre for Reviews and Dissemination 2009). Am meisten verbreitet sind die Kosten-Effektivitäts-Analyse und die Kosten-Nutzwert-Analyse, wobei die Kosten-Nutzwert-Analyse in manchen Lehrbüchern als Spezialform der Kosten-Effektivitäts-Analyse verstanden wird (Drummond, Sculpher et al. 2005). Der Einfachheit halber wird in den folgenden Kapiteln der Begriff Kosten-Effektivitäts-Analyse stellvertretend für alle gesundheitsökonomischen Studien verwendet.

Gesundheitsökonomische Evaluationen wurden bisher am häufigsten zu pharmazeutischen Interventionen durchgeführt, aber auch alle anderen Formen von Interventionen im Gesundheitssystem, insbesondere Public-Health-Maßnahmen (Impfungen, Lebensstilmaßnahmen etc.), werden zunehmend gesundheitsökonomisch evaluiert (Drummond, Weatherly et al. 2007).

Box 8.1:**Studientypen von gesundheitsökonomischen Evaluationen**

Alle Studientypen stellen die Kosten (Ressourcenverbrauch), die mit alternativen Interventionen verbunden sind, in Relation zu den Konsequenzen (klinische Effekte, Nebenwirkungen etc.) dar. Die Kosten werden immer in der gleichen Weise ermittelt, während die Effekte unterschiedlich gemessen bzw. bewertet werden.

Kosten-Effektivitäts-Analyse (cost-effectiveness analysis) (CEA)

Bei diesem Studientyp wird ein einzelner klinischer Outcome in Form einer natürlichen Einheit dargestellt (z. B. gewonnene Lebensjahre, vermiedene postoperative Infektionen, depressionsfreie Tage) und den Kosten gegenübergestellt. Wenn multiple klinische Outcomes berichtet werden, können diese nebeneinander in Form einer Kosten-Konsequenzen-Analyse (cost-consequence analysis) dargestellt werden.

Kosten-Nutzwert-Analyse (cost-utility analysis) (CUA)

Bei einer CUA werden die klinischen Outcomes in einen Nutzwert-Score konvertiert. Dazu sind Messinstrumente zur Erfassung der Lebensqualität wie SF-6D oder EuroQol (EQ-5D) erforderlich, auf welcher Basis generische Nutzwerte wie etwa „quality-adjusted life years“ (QALY) berechnet werden. Alternativ können andere Nutzwerte wie „healthy years equivalents“ (HYE) oder „disability-adjusted life years“ (DALY) eingesetzt werden. Der generische Charakter der Nutzwerte ermöglicht es, Gesundheitseffekte über Indikationen hinweg zu vergleichen.

Kosten-Nutzen-Analyse (cost-benefit analysis) (CBA)

Bei der CBA werden klinische Outcomes in monetäre Einheiten konvertiert. Somit kann ein monetärer Nettonutzen (bzw. Nettokosten) berechnet werden. Methoden zur Konvertierung klinischer Outcomes in monetäre Werte sind u. a. „willingness to pay“ (WTP) oder die „Humankapitalmethode“.

Kosten-Minimierungs-Analyse (cost-minimisation analysis) (CMA)

Bei der CMA (auch als Kosten-Kosten-Analyse bezeichnet) werden die Kosten zweier ergebnisgleicher Maßnahmen bezüglich eines Kostenvorteils verglichen. Die Situation der Ergebnisgleichheit ist allerdings selten anzutreffen.

Quelle: adaptiert nach (Centre for Reviews and Dissemination 2009), (Shemilt, Mugford et al. 2008) und (Schöffski and Graf von der Schulenburg 2007); Darstellung: GÖG/BIQG

8.1.3 Übersicht zu vorhandenen Methodenhandbüchern

Im Gegensatz zur Synthese der medizinischen Evidenz sind die Methoden zur Erstellung von Übersichtsarbeiten gesundheitsökonomischer Evidenz bedeutend weniger standardisiert und erfahren erst in jüngerer Zeit zunehmend Aufmerksamkeit. Eine Analyse zur Qualität publizierter systematischer Übersichtsarbeiten von gesundheitsökonomischen Evaluationen zeigte, dass der Prozess der Erstellung wie auch die Qualität insgesamt große Unterschiede aufweisen (z. B. Unterschiede in der Verwendung einschlägiger Datenbanken bei der Literatursuche, Anwendung unterschiedlicher Qualitätsprüfungsinstrumente etc.) (Jefferson, Demicheli et al. 2002).

Es existieren einige Handbücher von etablierten Institutionen mit Anleitungen zur Durchführung gesundheitsökonomischer Übersichtsarbeiten. Es handelt sich zumeist um einzelne Kapitel in allgemeinen Handbüchern zu systematischen Übersichtsarbeiten. So enthält das Cochrane-Handbuch seit 2008 im Kapitel 15 eine Anleitung zur Einbeziehung gesundheitsökonomischer Evidenz in Cochrane Reviews (Shemilt, Mugford et al. 2009). Das Handbuch zur Erstellung systematischer Reviews des Center für Reviews and Dissemination (CRD) enthält ein Kapitel zur Erstellung systematischer Übersichtsarbeiten gesundheitsökonomischer Evaluationen (Centre for Reviews and Dissemination 2009). Eine weitere Anleitung zu einzelnen Teilen einer Übersichtsarbeit steht über die U.S. Taskforce on Community Preventive Services zur Verfügung (in der Folge CPS Task Force genannt) (Task Force on Community Preventive Services 2010). Das Handbuch der NHS Economic Evaluation Database (Craig and Rice 2007) und das „NICE guidelines manual“ (National Institut for Health and Clinical Excellence 2009) bieten ebenfalls Hilfestellung zu einzelnen Schritten an (z. B. zur Literatursuche).

Gesundheitsökonomische Studien weisen einige Charakteristika auf, die spezielle Herausforderungen bei Übersichtsarbeiten mit sich bringen. Im Unterschied zu klinischen Studien spielt das Land, in dem eine gesundheitsökonomische Studie durchgeführt wurde, eine bedeutende Rolle für das Ergebnis. Beispielsweise sind die Preise einzelner Leistungen sowie die Preisniveaus länderspezifisch unterschiedlich. Außerdem werden gesundheitsökonomische Evaluationen oft nach länderspezifischen methodischen Leitlinien durchgeführt, die sich von jenen im eigenen Entscheidungskontext unterscheiden können. Auf diese Besonderheiten gehen die Handbücher in unterschiedlicher Weise ein, sodass daraus kein einheitlicher Standard abzuleiten ist.

Weitgehende Übereinstimmung besteht jedoch beim grundsätzlichen Prozedere zur Durchführung einer gesundheitsökonomischen Übersichtsarbeit. Dieser Prozess wird daher als derzeitiger Standard angenommen und im folgenden Kapitel auch als solcher für das vorliegende Handbuch vorgeschlagen. Der Prozess ist eng an die Durchführung einer medizinischen Evidenzsynthese angelehnt, sodass primär auf die Besonderheiten

bei gesundheitsökonomischen Arbeiten eingegangen wird und ansonsten auf das entsprechende Kapitel im Methodenhandbuch verwiesen wird.

8.1.4 Durchführung gesundheitsökonomischer Übersichtsarbeiten

8.1.4.1 Fragestellung

Systematische Übersichtsarbeiten wurden entwickelt, um verlässliche und rezente Evidenz zum Nutzen und Schaden von alternativen Interventionen in transparenter und systematischer Art und Weise darzustellen. Es geht dabei um eine Evidenzsynthese der Ergebnisse verschiedener (klinischer) Studien. Obwohl der methodische Ansatz des systematischen Arbeitens auf gesundheitsökonomische Evidenz prinzipiell übertragbar ist, ist der Zweck der gesundheitsökonomischen Übersichtsarbeit nur selten die Synthese der Endergebnisse gesundheitsökonomischer Studien, wie z. B. die Synthese von Kosten-Effektivitäts-Verhältnissen. Das Ziel einer gesundheitsökonomischen Übersichtsarbeit ist viel häufiger ein exploratives, nämlich zu analysieren, wie und warum Interventionen mehr oder weniger effektiv, ressourcen- bzw. kostenintensiv und kosteneffektiv sind. Die Frage, ob Alternative A im Vergleich zu Alternative B kosteneffektiv ist, wird hingegen aufgrund vielfältiger Limitationen (z. B. aufgrund mangelnder Übertragbarkeit der Studien) in den seltensten Fällen sinnvoll im Rahmen einer gesundheitsökonomischen Übersichtsarbeit zu beantworten sein.

In der Literatur werden drei geeignete Zwecke für gesundheitsökonomische Übersichtsarbeiten genannt (Anderson and Shemilt 2010): 1) um die Entwicklung eines neuen entscheidungsanalytischen Modells zu unterstützen, 2) um jene gesundheitsökonomischen Studien zu identifizieren, die für den eigenen Entscheidungskontext am meisten relevant sind und 3) um die wichtigsten ökonomischen Auswirkungen oder Zielkonflikte zu identifizieren, die mit einer Entscheidung verbunden sind. Beispielsweise könnte eine Frage in diesem Zusammenhang lauten: „Wie beeinflussen Level und Zusammensetzung der Ressourcen von Behandlung A und B die Höhe und Art der Outcomes, und welche Kontextfaktoren beeinflussen diesen Zusammenhang?“ Diese Form einer eher explorativen und theoriegenerierenden Übersichtsarbeit wird auch als „realistische Übersichtsarbeit“ bezeichnet (Anderson and Shemilt 2010).

Vor der Konkretisierung der Fragestellung ist daher zu Beginn zu klären, zu welchem Zweck die gesundheitsökonomische Übersichtsarbeit erstellt wird. Box 8.2 nennt als Hilfestellung beispielhaft konkrete Ziele einer gesundheitsökonomischen Übersichtsarbeit.

Box 8.2:**Mögliche Ziele gesundheitsökonomischer Übersichtsarbeiten**

- » Ökonomische Unterschiede zwischen zwei oder mehreren effektiven Interventionen für dieselbe Indikation quantifizieren (z. B.: Welche Darmkrebs-Screening-Strategie zeigt unter welchen Rahmenbedingungen die größte Kosteneffektivität?)
- » Ökonomische Auswirkungen einer Intervention für verschiedene Risikogruppen, Altersgruppen, Intervalle etc. illustrieren (z. B.: Ab welcher Altersgruppe ist ein Zervixkarzinom-Screening unter ökonomischen Gesichtspunkten sinnvoll?)
- » Potenzial neuer Technologien bewerten (z. B.: Ist die virtuelle Koloskopie effizienter, als es die traditionellen Ansätze sind?)
- » Kriterien identifizieren, die Voraussetzung für ein erwünschtes ökonomisches Ergebnis einer Intervention sind (z. B.: In welchem Ausmaß beeinflusst die Inanspruchnahme einer Koloskopie nach einem positiven Test auf okkultes Blut im Stuhl die Kosteneffektivität dieses Tests?)
- » Einfluss verschiedener Präferenzen der Patienten und Patientinnen auf die Kosteneffektivität identifizieren (z. B.: Wie wirkt sich die unterschiedliche Toleranz von Nebenwirkungen auf die Kosteneffektivität einer Chemotherapie aus?)
- » Leistungen nach ihren Kosten und erwarteten Effekten reihen (z. B.: Was ist der relative Nutzwert der HPV-Impfung im Vergleich zu anderen Impfprogrammen?)
- » Einfluss unterschiedlicher Rahmenbedingungen auf die Kosteneffektivität identifizieren (z. B.: Wie beeinflussen unterschiedliche Teilnehmeraten die Kosteneffektivität der HPV-Impfung?)
- » Umfassende Information für ein zu erstellendes entscheidungsanalytisches Modell generieren (z. B.: Welche Modelltypen wurden für die gesundheitsökonomische Analyse der Statintherapie angewendet?)

Quelle: (Pignone, Saha et al. 2005; Centre for Reviews and Dissemination 2009);
Darstellung: GÖG/BIQG

Zur Konkretisierung der Forschungsfrage sind folgende Orientierungsfragen hilfreich (Shemilt, Mugford et al. 2009):

- » Welche volkswirtschaftlichen Auswirkungen (z. B. für das Gesundheitssystem, für Anbieter von Leistungen, Individuen, Familien, Arbeitgeber) sind mit der Krankheit, auf die die Intervention abzielt, verbunden?
- » Welche zusätzlichen Ressourcen sind notwendig, um die neue Intervention im Vergleich mit den Alternativen zu implementieren? (z. B. Personal, Ausstattung, Infrastruktur)
- » Welche Auswirkungen auf Ressourcen sind mit der Implementierung einer neuen Technologie langfristig im Vergleich zu den Alternativen zu erwarten? (z. B. Komplikationen, Ambulanzbesuche, Arbeitszeit)
- » Welche zusätzlichen Kosten entstehen durch einen mit der Intervention verbundenen veränderten Ressourcenbedarf im Vergleich zu den Alternativen? (z. B. direkte medizinische Kosten, private Selbstbehalte)

- » Welche ökonomischen Auswirkungen entstehen durch die zusätzlichen Gesundheitseffekte oder Nebenwirkungen im Vergleich zu den Alternativen?
- » Welche potenziellen „trade-offs“ zwischen Kosten und Effekten können für die Entscheidung, eine neue Leistung zu finanzieren oder abzulehnen, von Relevanz sein?

Zusätzlich sind folgende Kriterien in Betracht zu ziehen:

- » Größenordnung: Welche Arten von Ressourcen (Ressourcenbedarf und Konsequenzen) und welche Kosten sind voraussichtlich relevant für die Entscheidung zwischen alternativen Interventionen?
- » Zeithorizont: Über welchen Zeithorizont sind die Kosten und Gesundheitseffekte zu erwarten? (Der Zeithorizont bei gesundheitsökonomischen Übersichtsarbeiten kann möglicherweise vom Zeithorizont bei Wirksamkeits-Übersichtsarbeiten differieren.)
- » Perspektive: Wer wird voraussichtlich die Kosten, die mit einer neuen Intervention verbunden sind, tragen (im Vergleich zu den Alternativen), und wer profitiert von den Gesundheitseffekten (Patienten/Patientinnen bzw. bestimmte Subgruppen, Angehörige, Leistungsanbieter, Kostenträger, Gesellschaft)?

Klinische Ereignispfade können helfen, die Rolle und Relevanz ökonomischer Aspekte eines speziellen Themas zu konzeptualisieren (Centre for Reviews and Dissemination 2009; Shemilt, Mugford et al. 2009) (s. Tabelle 8.1).

Tabelle 8.1:
Klinischer Ereignispfad

• Ereignispfad	• Beispiel
Klinisches Ereignis	Schlaganfall
↓ Management des Ereignisses und nachfolgende Ereignisse	Akutversorgung + Rehabilitation + Spätkomplikationen
↓ Ressourcen für Versorgung und Effekte	Dauer Spitalsaufenthalt, Intensität der Rehabilitation, Management der Komplikationen und Gesundheitseffekte in jedem dieser Stadien
↓ Kosten und Nutzwerte	Bewertung des Ressourcenverbrauchs mittels administrierter Preisen und Bewertung der Outcomes z. B. mittels quality-adjusted life years (QALYs)

Quelle: Shemilt, Mugford 2009; Darstellung: GÖG/BIQG

Das PICO-Prinzip (s. 4.1.2.1) kann bei gesundheitsökonomischen Übersichtsarbeiten für die Operationalisierung der Forschungsfrage hilfreich sein, das Ausmaß der direkten Anwendbarkeit hängt aber vom Ziel der Arbeit ab. Für Beispielfrage 1 in Tabelle 8.2 (Beispielfragen) ist das PICO-Prinzip für die Operationalisierung der Forschungsfrage anwendbar, für Beispielfrage 2 ist es nur teilweise anwendbar.

Tabelle 8.2:
Beispielfragen

1. Welche Primärpräventionsstrategie zeigt bei Patienten und Patientinnen mit erhöhtem Risiko für koronare Herzerkrankungen unter welchen Bedingungen aus gesellschaftlicher Perspektive die größte Kosteneffektivität?	
Population	Patienten mit erhöhtem Risiko für koronare Herzerkrankung
Intervention	Primärprävention mit Cholesterinsenkern
Control	Nicht-medikamentöse Prävention
Outcome	Mortalität -> Kosten pro gewonnenes Lebensjahr (LYG)
2. In welchem Ausmaß beeinflusst die Inanspruchnahme einer Koloskopie nach einem positiven Test auf okkultes Blut im Stuhl die Kosteneffektivität dieses Tests?	
Population	Bevölkerung
Intervention	Darmkrebscreening mit Test auf okkultes Blut im Stuhl und nachfolgende Koloskopie
Control	Nicht relevant
Outcome	Mortalität, Lebensqualität -> Kosten pro QALY, Kosten pro LYG

Quelle: LBI-HTA 2011; Darstellung: GÖG/BIQG

8.1.4.2 Definition der Selektionskriterien

Wie in Kapitel 4.2 für medizinische Übersichtsarbeiten beschrieben, sind auch für eine gesundheitsökonomische Übersichtsarbeit Selektionskriterien zu definieren, die als Grundlage für die Suchstrategie und für die Selektion der Literatur verwendet werden. Zusätzlich zu den beschriebenen Kriterien in Kapitel 4.2 (Sprache, Intervention etc.) sind in einer gesundheitsökonomischen Übersichtsarbeit folgende Selektionskriterien relevant (National Institut for Health and Clinical Excellence 2009; Shemilt, Mugford et al. 2009):

Studientyp (z. B. Kosten-Effektivitäts-Analyse, Kosten-Nutzwert-Analyse)

Die Entscheidung über die einzuschließenden Studientypen hängt von der gesundheitsökonomischen Frage ab. Für viele Fragestellungen ist es ratsam, die Studientypen

nicht zu stark einzuschränken. Zum Beispiel wäre der geeignete Studientyp für die Frage, welche ökonomischen Auswirkungen langfristig mit der Refundierung einer neuen Technologie verbunden sind, eine Ausgaben-Einfluss-Analyse, solche Studien sind aber selten vorhanden und außerdem im Ergebnis sehr kontextabhängig. Kosten-Effektivitäts-Studien können für diese Frage ebenso wertvolle Information liefern, wenn darin die Ressourcenkategorien und die Menge des Ressourcenverbrauchs, die bei den untersuchten Interventionen anfallen, detailliert dargestellt werden.

Quelle für die Gesundheitseffektkomponente

Es ist zu definieren, welche Quellen klinischer Evidenz für die Effektkomponente der gesundheitsökonomischen Evaluationen akzeptiert werden. Gesundheitsökonomische Evaluationen können entweder primär auf *einer* Quelle klinischer Evidenz basieren, wenn sie parallel zu einer klinischen Studie erstellt werden (Piggy-back-Studien) oder mit multiplen klinischen (Sekundär-)Quellen erstellt werden, die in entscheidungsanalytische Modelle integriert werden. In jedem Fall ist zu entscheiden, welcher Studientyp zur Bewertung der Effektivität einer Maßnahme akzeptiert wird (z. B. ausschließlich RCTs bzw. Metaanalysen).

Perspektive

Es ist zu definieren, ob gesundheitsökonomische Studien mit jeglicher Perspektive oder nur solche mit einer vorab spezifizierten Perspektive eingeschlossen werden (z. B. nur solche aus der Kostenträgerperspektive oder aus der gesellschaftlichen Perspektive).

Outcomeparameter

Es ist zu definieren, nach welchen Outcomeparametern gesundheitsökonomische Evaluationen selektiert werden sollen (z. B. QALYs, LYG, krankheitsspezifische natürliche Einheiten wie depressionsfreie Tage).

Zeithorizont

Ein weiteres Ausschlusskriterium kann der Zeithorizont einer Studie sein. Wenn es beispielsweise um die Behandlung einer chronischen Erkrankung geht, könnte definiert werden, dass alle Studien mit einem Zeithorizont von weniger als zwei Jahren ausgeschlossen werden.

Land/Setting

Je nach definiertem Zweck der Übersichtsarbeit können Studien aus bestimmten Ländern oder Settings, die vom eigenen Entscheidungskontext sehr stark abweichen, ausgeschlossen werden.

Tabelle 8.3:
Beispiel für Ein- und Ausschlusskriterien

Welche Primärpräventionsstrategie zeigt bei Patienten und Patientinnen mit erhöhtem Risiko für koronare Herzerkrankungen unter welchen Bedingungen aus gesellschaftlicher Perspektive die größte Kosteneffektivität?	
Studientyp	Kosten-Effektivitäts-Analysen, Kosten-Nutzwert-Analysen
Quelle für Gesundheitseffektkomponente	RCTs, Übersichtsarbeiten, Metaanalysen; (Piggy-back-Studien und entscheidungsanalytische Modelle)
Perspektive	Gesellschaft
Outcomekategorie	QALYs, LYG
Zeithorizont	Mindestens 2 Jahre bis lebenslang
Land/Setting	„Westliche Länder“ mit vergleichbaren Präventionssettings

Quelle: LBI-HTA 2011; Darstellung: GÖG/BIQG

Wie für klinische Fragestellungen ist auch für gesundheitsökonomische Übersichtsarbeiten ein Berichtsplan zu erstellen (siehe Kapitel 4.4).

8.1.4.3 Literatursuche

Grundsätzlich gelten die gleichen Anforderungen wie für eine systematische Literatursuche zu medizinischen Übersichtsarbeiten (siehe Kapitel 4.5). Für gesundheitsökonomische Fragestellungen existieren jedoch sowohl zusätzliche Datenbanken als auch spezifische Hilfestellungen für die Erstellung der Suchstrategie.

Eine Liste zu sämtlichen Quellen für gesundheitsökonomische Studien (elektronische Literaturdatenbanken, Internetseiten, die relevante graue Literatur enthalten etc.) findet sich bei (Napper and Varney 2005), bei (Aguar-Ibanez, Nixon et al. 2005) oder auf der CRD-Website (www.york.ac.uk/inst/crd/econ4.htm).

Die Literatursuche kann sowohl in den klassischen biomedizinischen bibliographischen Datenbanken (Medline, Embase etc.) erfolgen als auch in Datenbanken, die auf gesundheitsökonomische Studien spezialisiert sind. Dazu gehören die frei zugängliche „CRD NHS Economic Evaluation Database“ (NHS EED), die Kurzbewertungen von vollständigen gesundheitsökonomischen Evaluationen und bibliographische Details zu einigen weiteren gesundheitsökonomischen Studien enthält, und die kostenpflichtige

„Health Economics Evaluation Database“ (HEED), die von John Wiley und Söhnen produziert wird. Die „NHS-EED-Datenbank“ ist sowohl über das CRD zugänglich als auch über die Cochrane Library, wobei die CRD-Website monatlich aktualisiert wird, während die NHS-EED-Version der Cochrane Library vierteljährlich aktualisiert wird. Der Wunsch, die in Großbritannien verankerte NHS-EED-Datenbank auf eine supra-national verortete europäische Datenbank zu erweitern, hat zum European Network of Health Economic Evaluation Databases (EURONHEED) geführt, die ebenfalls frei zugänglich ist (<http://infodoc.inserm.fr/euronheed/>) und ergänzend durchsucht werden kann.

Sowohl die „NHS-EED-Datenbank“ als auch die „HEED-Datenbank“ enthalten ausschließlich gesundheitsökonomische Studien, daher muss bei der Suchstrategie nur das inhaltliche Themengebiet spezifiziert werden. Theoretisch enthält die „NHS-EED-Datenbank“ alle publizierten gesundheitsökonomischen Evaluationen inklusive einer kritischen Begutachtung. In dieser Datenbank soll daher auf jeden Fall gesucht werden. Die größte Sensitivität und Spezifität ist jedoch in der MEDLINE-Suche zu erreichen (Sassi, Archard et al. 2002), da der Eintrag in die gesundheitsökonomischen Datenbanken zeitverzögert und möglicherweise selektiv erfolgt. Eine ergänzende Suche in biomedizinischen Datenbanken (insbesondere in MEDLINE und EMBASE) oder weiteren bibliographischen Datenbanken (z. B. PsycINFO) wird daher empfohlen. Eine zusätzlich relevante Datenbank ist „EconLit“ (<http://www.econlit.org/>), die allerdings nicht auf gesundheitsökonomische Literatur eingeschränkt ist, sondern ökonomische Literatur insgesamt enthält. Für die Suchstrategie in den genannten bibliographischen Datenbanken müssen sowohl das inhaltliche Themengebiet als auch die gesundheitsökonomischen Begriffe spezifiziert werden.

Als Unterstützung für die Suche in biomedizinischen Datenbanken sind spezielle Suchfilter und Suchstrategien vorhanden. Beispielsweise wurden von der McMaster-Universität (finanziert von der US National Library of Medicine) Suchfilter für große biomedizinische Datenbanken erstellt, die auch gesundheitsökonomische Filter oder Kostenfilter beinhalten (http://hiru.mcmaster.ca/hiru/HIRU_Hedges_home.aspx). Die „InterTASC Information Specialists' Subgroup“ enthält ebenfalls einen Abschnitt zu ökonomischen Filtern (www.york.ac.uk/inst/crd/intertasc/econ.htm). Das „NHS Economic Evaluation Database“-Handbuch enthält Suchstrategien für die gängigen biomedizinischen Datenbanken (Craig and Rice 2007). Die Suchstrategien sind sehr breit angelegt und können je nach Fragestellung angepasst werden. Sie lassen sich auch für andere Datenbanken adaptieren.

8.1.4.4 Datenextraktion

Für die Datenextraktion sind zwei Kategorien von Daten zu unterscheiden: Daten zu den Studiencharakteristika und Daten zu den Ergebnissen der Studien. Welche Daten aus den Studien zu extrahieren sind, hängt von der Fragestellung der Übersichtsarbeit ab, einige Schlüsselparameter werden jedoch wiederholt genannt (Centre for Reviews and Dissemination 2009; Shemilt, Mugford et al. 2009). Sie beziehen sich auf gesundheitsökonomische Evaluationen. Für andere gesundheitsökonomische Studien existieren keine spezifischen Empfehlungen für die Datenextraktion.

Für die Darstellung der Studiencharakteristika sind folgende Kategorien relevant:

- » Autor/Autorin, Jahr,
- » Studienland,
- » Intervention und Vergleichsintervention(en),
- » Studientyp, Modellierungstechnik,
- » Perspektive,
- » Studienpopulation,
- » Zeithorizont,
- » Quellen/Qualität von Effektivitäts- und Kostendaten inkl. Währung, Bezugsjahr,
- » verwendeter Outcomeparameter + angewendete Methode,
- » Diskontierung,
- » sonstige methodische Aspekte (z. B. Umgang mit Unsicherheit),
- » Finanzierung der Studie.

Für die Darstellung der Studienergebnisse sind folgende Kategorien relevant:

- » Gesundheitseffekte,
- » Ressourcenverbrauch, Preise und Kosten,
- » Synthese von Effekten und Kosten
(z. B. inkrementelles Kosten-Effektivitäts-Verhältnis),
- » Ergebnisse der Sensitivitätsanalysen.

Je nach Fragestellung der Übersichtsarbeit können weitere oder andere Parameter erforderlich sein. Wenn die Übersichtsarbeit als Basis für die Erstellung einer eigenen Studie dienen soll, sind mitunter detaillierte Angaben zur methodischen Ausführung notwendig.

Als Hilfestellung enthält die ältere Version des „CRD-Handbuchs“ (Khan, Riet ter et al. 2001) einen Vorschlag für ein Datenextraktionsformular (siehe Anhang). Auch die Abstract-Strukturierung in der NHS Economic Evaluation Database ist eine mögliche Hilfestellung zur Strukturierung der Datenextraktion. Das dazugehörige Handbuch enthält für jede Kategorie eine ausführliche Beschreibung inklusive einer genauen

Definition der verwendeten Begriffe (Craig and Rice 2007). Eine Hilfestellung für sehr detaillierte Datenextraktion bieten die Abschnitte I–III in der „Economic Evaluation Abstraction Form“ der „CPS Task Force“ (Task Force on Community Preventive Services 2010). Bei gesundheitsökonomischen Evaluationen, die bereits in der „CRD–Economic Evaluation Database“ enthalten sind, kann sowohl auf die Struktur als auch auf den Inhalt der Datenextraktion in den dokumentierten Abstracts zurückgegriffen werden.

8.1.4.5 Qualitätsprüfung

Die Qualität einer gesundheitsökonomischen Studie kann sehr unterschiedlich sein. Das Ziel einer Qualitätsprüfung besteht primär darin, zu evaluieren, ob die Methoden, Annahmen, Modelle, Ergebnisse und mögliche Verzerrungen transparent beschrieben wurden, sodass die Studie für die Leserschaft so nachvollziehbar wie möglich ist. Die nachfolgend angeführten Elemente und Instrumente einer Qualitätsprüfung beziehen sich primär auf gesundheitsökonomische Evaluationen. Für andere gesundheitsökonomische Studien fehlen standardisierte Instrumente. Als Hilfestellung können aber methodische Leitlinien herangezogen werden, z. B. ISPOR good practice for budget impact analysis (Mauskopf, Sullivan et al. 2006).

8.1.4.5.1 Elemente einer Qualitätsprüfung

Die Qualitätsprüfung umfasst insbesondere folgende Elemente:

Methode zur Generierung der Effektivitätsdaten

Die Qualität einer gesundheitsökonomischen Evaluation hängt stark von der Verwendung verlässlicher Wirksamkeits- bzw. Effektivitätsdaten der evaluierten Technologien ab. Entsprechend der Hierarchie der Evidenz in Kapitel 4.8 sollten, wann immer möglich, Effektivitätsdaten von systematischen Übersichtsarbeiten bzw. Metaanalysen stammen. Die Palette für den Einsatz der Effektivitätsdaten reicht von einzelnen Wirksamkeitsparametern aus einer Metaanalyse bis hin zu epidemiologischen Daten für die Darstellung des natürlichen Krankheitsverlaufs. Die Qualität wird davon abhängen, welche Art von Daten wofür verwendet wurde.

Wurden die Effektivitätsdaten von *einer* Quelle bezogen, sollte diese entsprechend der Qualitätsprüfung in Kapitel 8.1.4.5 geprüft werden. Wurden *mehrere* Quellen verwendet, ist zu prüfen, ob eine systematische Suche angewendet wurde, in welchen Datenbanken recherchiert wurde, ob transparente Selektionskriterien angewendet wurden und ob ausreichend Information über die Qualität der inkludierten Studien gegeben wurde.

Messung und Bewertung des Ressourcenverbrauchs

Die Qualitätsprüfung untersucht, welche Arten von Kosten in der Studie erfasst und wie diese berechnet wurden, d. h., wie die Mengen an Ressourcen, die für die Durchführung von Leistungen nötig ist, erfasst wurde und wie diese bewertet wurden (z. B. mit Marktpreisen, Tarifen).

Alle Kosten, die für die Forschungsfrage der Studie und für die gewählte Perspektive relevant sind, sollen erfasst sein. Die benötigten Ressourcen sollen in physischen Einheiten dargestellt sein (z. B. Krankenhausaufenthalt in Tagen). Zu prüfen ist, ob die Quellen für die Daten zum Ressourcenbedarf transparent berichtet werden, ob der Bedarf an Ressourcen getrennt von den Preisen dargestellt wird, ob Annahmen zum Ressourcenbedarf explizit beschrieben werden und – falls Experten/Expertinnen zur Definition des Ressourcenbedarfs herangezogen wurden – ob die Methode der Befragung korrekt dargestellt wurde.

Die Qualitätsprüfung zur Bewertung des Ressourcenverbrauchs beinhaltet, ob die Quellen für die Preise angegeben wurden, ob die Kosten auf ein einheitliches Jahr angepasst wurden und ob dieses Jahr genannt wird, ob die Kosten bei einem mehr als einjährigen Zeithorizont diskontiert wurden und ob angegeben wurde, ob die verwendeten Daten die tatsächlichen Kosten widerspiegeln (z. B. wird die Verwendung von Tarifen als Preiskomponente die tatsächlichen Kosten mitunter nicht exakt widerspiegeln).

Messung und Bewertung der Gesundheitseffekte

Wenn in der Studie Nutzwerte (z. B. QALYs, DALYs, HYE) verwendet wurden, können für ihre Ermittlung entweder generische oder krankheitsspezifische Instrumente herangezogen werden. Die Bewertung der Gesundheitszustände kann mit unterschiedlichen Methoden (z. B. „standard gamble“, „time trade-off“) erfolgen. Diese Methoden unterscheiden sich erheblich voneinander, und die Wahl des Instruments kann sich auf das Ergebnis auswirken. Bei der Qualitätsprüfung ist daher zu kontrollieren, ob die Methoden transparent dargestellt werden.

Außerdem soll darauf geachtet werden, von wem die Bewertung der Gesundheitseffekte stammt (Patienten/Patientinnen, Gesellschaft, Kliniker etc.), welche Instrumente benützt wurden (EQ-5D, SF-6D etc.) und wann die Werte erhoben wurden (baseline, während der Behandlung, nach der Behandlung etc.).

Methode zur Synthese der Kosten und Effekte

Der ökonomische Wert einer Intervention im Vergleich zu einer anderen hängt davon ab, wie hoch die zusätzlichen Kosten und die zusätzlichen Gesundheitseffekte sind. Die Synthese dieser beiden Ergebnisse wird als „inkrementelles Kosten-Effektivitäts-Verhältnis“ (IKEV) oder „incremental cost-effectiveness ratio“ (ICER) ausgedrückt. Außer in den Fällen, wo die Kosten einer untersuchten Intervention geringer und die Effekte gleichzeitig besser sind als die der Vergleichsintervention, soll die Synthese von Kosten und Effekten immer in Form des IKEV erfolgen.

Die Qualitätsprüfung untersucht außerdem, ob die Ergebnisse sowohl in nicht-aggregierter als auch in aggregierter Form dargestellt wurden. Das heißt, neben dem IKEV als Endergebnis sollen Gesundheitseffekte und Kosten der Nachvollziehbarkeit wegen immer separat und in undiskontierter und diskontierter Form präsentiert sein.

Analyse der Unsicherheit

Bei der Qualitätsprüfung wird untersucht, ob alle relevanten Formen von Unsicherheit in der Studie berücksichtigt und mit adäquaten Methoden überprüft wurden.

Folgende Formen von Unsicherheit können in einer gesundheitsökonomischen Evaluation auftreten:

- » Parameter-Unsicherheit: entsteht dadurch, dass Parameter von Stichproben ermittelt werden und ihr wahrer Wert dadurch unbekannt ist;
- » methodische Unsicherheit: entsteht durch die Verwendung zahlreicher analytischer Methoden, deren Anwendung kontrovers ist (z. B. Berechnung indirekter Kosten, Diskontierung der Gesundheitseffekte, Diskontierungsraten);
- » Modell-Unsicherheit: entsteht durch simplifizierende Annahmen, die häufig für die Modellierung notwendig sind.

Für den Umgang mit Unsicherheit existieren verschiedene Methoden, wie etwa statistische Vergleiche, Bootstrapping, Sensitivitätsanalysen (univariat oder multivariat), Grenzwertanalysen (Thresholdanalysen), Extremwertanalysen bzw. Worst/best-case-Szenarien und probabilistische Sensitivitätsanalysen. Die Wahl der Methode(n) hängt von der Fragestellung und den verwendeten Daten für die Inputparameter ab.

- » Statische Vergleiche: Statistische Tests zum Vergleich von Effekten, Kosten oder Kosteneffektivität sind dann adäquat, wenn die Effektivitäts- und Kostendaten auf Patientenebene erhoben wurden.
- » Bootstrapping: Diese statistische Methode kann angewendet werden, wenn Daten auf Patientenebene erhoben wurden. Da das IKEV ein Quotient ist (z. B. Kosten/QALY), können parametrische statistische Methoden, die auf dem Standardfehler-Prinzip beruhen, nicht angewendet werden. Nicht-parametrisches Bootstrapping

ist eine Methode, die den Vergleich eines arithmetischen Mittels ohne Annahmen über die Stichprobenverteilung ermöglicht. Prinzipiell kann aber statt eines IKEV auch ein statistischer Netto-Benefit berechnet werden, wodurch die Statistik-Problematik, die mit einem Quotienten verbunden ist, umgangen wird.

- » Sensitivitätsanalysen bei Parameterunsicherheit: Diese Analysen sind typisch für gesundheitsökonomische Evaluationen, die Daten aus systematischen Übersichtsarbeiten verwenden. Es wird analysiert, wie sich die Ergebnisse durch die Veränderung der Parameter ändern. Dazu können univariate und multivariate Sensitivitätsanalysen oder Grenzwertanalysen eingesetzt werden, bei denen sich aber nur wenige Parameter gleichzeitig variieren lassen. Mit einer weiteren Form, der probabilistischen Sensitivitätsanalyse, können alle Parameter gleichzeitig variiert werden (siehe übernächster Absatz). Es ist zu prüfen,
 - » ob die für die Sensitivitätsanalyse gewählten Parameter gerechtfertigt sind,
 - » ob der gewählte Variationsbereich sinnvoll ist,
 - » ob die Robustheit der Ergebnisse nach einem vorher definierten akzeptierten Schwankungsbereich ermittelt wurde.
- » Sensitivitätsanalyse bei methodologischer Unsicherheit; hierunter fällt etwa der Einfluss unterschiedlicher Diskontierungsraten auf die Ergebnisse.
- » Probabilistische Sensitivitätsanalysen (PSA): Diese Methode kann nur für das Testen von Parameterunsicherheit in modellbasierten Studien angewendet werden. PSA berücksichtigt die Unsicherheit, die mit den Parameter-Werten verbunden ist, indem jedem Parameter eine Wahrscheinlichkeitsverteilung (anstatt eines Punktschätzers) zugeordnet wird. Es ist zu prüfen,
 - » ob den Modellparametern bei der PSA eine adäquate Verteilung zugeordnet wurde,
 - » ob relevante Annahmen getestet wurden (z. B. Modellstruktur).

Generalisierbarkeit der Ergebnisse

Generalisierbarkeit bezieht sich auf die Frage, ob die Ergebnisse in den Studien auf einen anderen Kontext übertragbar sind. Die Generalisierbarkeit kann durch die Relevanz der Intervention, der Population der Patienten bzw. Patientinnen und durch die beschriebenen Ressourcen in einer Studie eingeschränkt sein. Die Generalisierbarkeit kann mit Sensitivitätsanalysen geprüft werden (z. B. kann durch eine univariate Sensitivitätsanalyse der Kosten geprüft werden, in welchem Ausmaß die Preise der notwendigen Interventionen das Ergebnis beeinflussen). Details zum Thema „Übertragbarkeit von Studien“ sind in Kapitel 8.1.5 angeführt.

- » Instrumente zur Qualitätsprüfung

Zur Qualitätsprüfung existiert eine Reihe von Checklisten, die eine kohärente Qualitätsprüfung unterstützen. Keine der Checklisten wurde bisher formal validiert, aller-

dings wurden zwei besonders häufig überprüft und empfohlen (Shemilt, Mugford et al. 2009):

- » British Medical Checklist for authors and peer reviewers of economic submissions (Drummond and Jefferson 1996): Diese Liste ist in unterschiedlich ausführlichen Versionen erhältlich (10-Elemente-Version, 35-Elemente-Version und 36-Elemente-Version) und überprüft vorwiegend, ob zentrale methodische Charakteristika und Ergebnisse in adäquater Form berichtet werden.
- » CHEC Criteria list for assessment of methodological quality of economic evaluations (Evers, Goossens et al. 2005): Der Schwerpunkt dieser Liste ist primär auf die Überprüfung der methodischen Korrektheit ausgerichtet.

Es existiert auch ein deutschsprachiger Kriterienkatalog zur Beschreibung und Bewertung gesundheitsökonomischer Evaluationsstudien, der im Konsensusverfahren von den „gesundheitsökonomischen Projektgruppen München, Hannover, Ulm“ entwickelt wurde (Siebert, Behrend et al. 1999).

Eine spezifische Checklist zur Prüfung von Modellen in gesundheitsökonomischen Studien wurde von (Philips, Ginnelly et al. 2004) publiziert. Alle genannten Checklisten finden sich im Anhang.

Da die Ergebnisse der Qualitätsprüfung weniger vom verwendeten Instrument als von den Evaluatoren beeinflusst zu sein scheinen (Gerken, Crott et al. 2008), ist für eine möglichst gute „Interrater-Validität“ (Übereinstimmung der Bewertungsergebnisse bei unterschiedlichen Beobachtern) die Qualitätsprüfung durch zwei unabhängige Experten bzw. Expertinnen empfehlenswert.

Für die Qualitätsprüfung jener gesundheitsökonomischen Evaluationen, die in der „CRD-Economic Evaluation Database“ enthalten sind, können die Informationen der Abstracts, insbesondere die kritischen Kommentare zu den Studien, hilfreich sein. Die Informationen sind jedoch nicht immer gleich ausführlich, sodass sie eher als Ergänzung denn als Ersatz für eine standardisierte Qualitätsprüfung im Rahmen einer Übersichtsarbeit einsetzbar sind.

Es gibt keine standardisierten Vorgaben, wie mit den Ergebnissen einer Qualitätsprüfung umzugehen ist (z. B. Ausschluss von Studien aus der Übersichtsarbeit). Die Vorgangsweise wird von der Fragestellung der Übersichtsarbeit abhängen, sie soll aber in jedem Fall vorab definiert und transparent beschrieben werden.

8.1.4.5.2 Punktesysteme

Zur Bewertung der Studienqualität gesundheitsökonomischer Evaluationen wurden mehrere Punktesysteme entwickelt. Sie funktionieren nach dem Prinzip, den verschiedenen Elementen eines Qualitätskatalogs Werte zuzuordnen und diese zu einem

Summenscore zu addieren, der eine bestimmte Qualität widerspiegeln soll. Eine Übersichtsarbeit zu Punktesystemen (Thurston, Craig et al. 2008) weist sechs publizierte Methoden aus (Gerard, Seymore et al. 2000; Gonzalez-Perez 2002; Wallace, Weingarten et al. 2002; Chiou, Hay et al. 2003).

Da keine der Methoden als ausreichend valide und verlässlich beschrieben wurde, wird von der Anwendung von Punktesystemen abgeraten (Shemilt, Mugford et al. 2009). Sinnvoller ist die einfache Darstellung von Studienstärken und -schwächen auf Basis der beschriebenen Checklisten und die Illustration des Einflusses dieser Information auf die Verlässlichkeit des Gesamtergebnisses einer Studie.

8.1.4.6 Datensynthese

Anders als bei medizinischen Übersichtsarbeiten existieren für gesundheitsökonomische Übersichtsarbeiten keine allgemein gültigen Instrumente zur Evidenzsynthese (wie z. B. Evidenzprofile nach GRADE; siehe Kapitel 7.1), sodass dazu kein standardisiertes Vorgehen empfohlen werden kann. Die Variationsbreite der Forschungsfragen und die Studienheterogenität erfordern ein entsprechendes Maß an Kreativität, die manchmal sehr umfangreiche Menge an Information zusammenzufassen. Wesentlich ist dabei, so transparent wie möglich vorzugehen.

Da gesundheitsökonomische Studien in einer Übersichtsarbeit zumeist sehr heterogen sind, wird primär eine qualitative Datensynthese empfohlen. Es besteht Uneinigkeit darüber, ob von einer quantitativen Synthese prinzipiell abzusehen ist (Centre for Reviews and Dissemination 2009) oder ob in bestimmten Fällen Metaanalysen (Shemilt, Mugford et al. 2009) oder Metaregressionsmodelle (Stanley 2010) sinnvoll sein können.

Im Zusammenhang mit der Datensynthese besteht mitunter Bedarf für eine Anpassung bestimmter Daten (z. B. Währungs- und Preisanpassungen) aus den einzelnen Studienkontexten an den eigenen Entscheidungskontext. In der Realität werden Datensynthese und Anpassung von Daten – je nach Fragestellung und Zielgruppe – ineinander übergreifen. In diesem Handbuch werden die Themen zunächst getrennt behandelt. Zu Details zum Thema Übertragbarkeit gesundheitsökonomischer Ergebnisse sowie methodische Ansätze zur Anpassung von Studienergebnissen siehe Kapitel 8.1.5.

8.1.4.6.1 Qualitative Synthese

Als „good practice“ einer narrativen Zusammenfassung werden folgende Inhalte vorgeschlagen (Shemilt, Mugford et al. 2009):

- » Angabe der Anzahl ausgewählter und ausgeschlossener Studien;
- » zusammenfassende Beschreibung der gesundheitsökonomischen Forschungsfrage in den Studien,
- » der Studiendesigns,
- » der Perspektive,
- » des Zeithorizonts in den Studien;
- » Diskussion der Methoden zur Darstellung von inkrementellem Ressourcenverbrauch, inkrementellen Kosten und inkrementeller Kosteneffektivität in den Studien;
- » Beschreibung der Methoden zum Umgang mit Unsicherheit bei Ressourcenverbrauch, Kosten und Kosteneffektivität;
- » Beschreibung der Währung und des Jahres, auf die sich die extrahierten Kostendaten beziehen;
- » Anpassung der Kostendaten in den Studien auf eine gemeinsame Währung und ein gemeinsames Jahr, falls möglich (siehe Kapitel 4);
- » Diskussion der Qualität der Effektivitätsdaten;
- » Beschreibung der wichtigsten Merkmale der durchgeführten Analysen zur Unsicherheit und der Konsistenz der Ergebnisse innerhalb sowie zwischen den Studien;
- » Diskussion der Gesamtqualität und der Limitationen der Studien;
- » Diskussion der Relevanz und Generalisierbarkeit der Studien für den eigenen Anwendungskontext.

8.1.4.6.2 Darstellungsformen

Die Anwendbarkeit einer Übersichtsarbeit wird durch geeignete Darstellungsformen wesentlich verbessert. Nachfolgend werden einige Beispiele aus existierenden Übersichtsarbeiten angeführt.

Tabellarische Zusammenschau

Für die qualitative Synthese bietet sich als erster Schritt an, Schlüsselinformationen aus den Datenextraktionstabellen und aus der Qualitätsprüfung in Übersichtstabellen zusammenzufassen. Ein Beispiel dafür ist in Tabelle 8.4 dargestellt. Ein weiteres Beispiel findet sich im Anhang (siehe Tabelle 15.7).

Tabelle 8.4:

Beispiel für Übersichtstabelle; Übersichtsarbeit aus dem Bereich der Kinder- und Jugendpsychiatrie

• Autor, Jahr, Land	• Indikation	• Intervention	• Outcome	• Kostenarten/Perspektive	• Ergebnis
Donnelly et al., 2004, AUS	ADHS	Dexamphetamine/ DEX vs. Methylphenidate/ MPH	Behinderungs- bereinigtes Lebens- jahr/DALY	Kosten öffentlicher Einrichtungen und Kosten der Patienten, jeweils für Serviceleistungen und Medikamente	Sowohl DEX (4.100 A\$/DALY) als auch MPH (15.000 A\$/DALY) kosteneffektiv
Dretzke J. et al., 2005, UK	Störung des Sozial- verhaltens mit oppositionellem, aufsässigem Verhalten	Trainings und Schulungen für Eltern vs. diverse Formen des Trainings für Kinder vs. Trainingsformen für Eltern untereinander	QALY	Operationale Kosten (Personal, Ausstattung ...), Opportunitätskosten (aufgewandte Zeit, Ressourcen ...), Kosten für Klienten (Reisezeit, Pflegezeit ...)	Gesamtkosten für Training von Eltern/Familien höher als bis dato angenommen. Bei einer angenommenen Lebensqualitätssteigerung von 20 % beträgt das IKEV zumindest 20.000 £.
Edwards et al., 2007, UK	Verhaltensstörung	Webster-Stratton Incredible Years Erziehungshilfeprogramm für Eltern von Kindern, die zu Verhaltensstörungen tendieren, vs. sechsmonatige „waiting list control“	Veränderungen auf der Intensitäts- und Problemskala zum Eyberg-Erhebungsbogen von kindlichem Verhalten	Kosten aus Sicht des öffentlichen Gesundheits-, Sozial- und Bildungswesens inklusive Zeit- und Reisekosten	IKEV: 73 £ pro verbesserten Punkt in der Eyberg-Skala
Faber et al., 2008, NED	ADHS	lang anhaltende orale, osmotische Abgabe von Methylphenidaten/ OROS an Jugendliche, bei denen schnell wirksame Methylphenidate suboptimal wirken, vs. schnell wirksame Methylphenidate	QALY	Kosten für Medikamente, Beratung, Behandlungsinterventionen und Zusatzkosten für die Teilnahme an speziellen Ausbildungen	2.004 € per QALY für OROS; über 10 Jahre ergibt OROS-Variante 0,13 QALYs (1,5 Monate) mehr

Quelle: Radlberger, Zechmeister et al. 2009; Darstellung: GÖG/BIQG

Neben der tabellarischen Darstellung existieren weitere Möglichkeiten, die Ergebnisse von gesundheitsökonomischen Übersichtsarbeiten im Überblick darzustellen (Nixon, Khan et al. 2001). Ob und wann diese Formen sinnvoll sind, hängt von Zweck und Fragestellung der Arbeit ab:

Permutationsmatrix

Diese Matrix zeigt die neun möglichen Ergebniskonstellationen in Bezug auf Kosten (höher, unverändert, geringer) und Effekte (besser, unverändert, schlechter) von untersuchten Alternativen (Felder a bis i), denen die Studien aus der Übersichtsarbeit zugeordnet werden. Mit unterschiedlicher Schattierung wird zusätzlich zwischen verschiedenen Stärken einer Empfehlung differenziert. Die Empfehlung, eine neue Technologie zu akzeptieren, wird etwa dann stark unterstützt, wenn die Intervention bessere Gesundheitseffekte bei gleichzeitig niedrigeren Kosten mit sich bringt als die Vergleichsintervention (g), während bei einer Intervention, die zwar mehr Effektivität, aber auch höhere Kosten mit sich bringt (a), keine eindeutige Empfehlung abzugeben ist.

Tabelle 8.5:
Permutationsmatrix

		Inkrementelle Effektivität		
		+	0	-
Inkrementelle Kosten	+	a (n)	b (n)	c (n)
	0	d (n)	e (n)	f (n)
	-	g (n)	h (n)	i (n)

Legende

	Effektivität	Kosten
+	besser	höher
0	unverändert	unverändert
-	schlechter	geringer

	Eindeutige Empfehlung für eine Entscheidung
	g = Intervention akzeptieren c = Intervention ablehnen
	Schwache Empfehlung für eine Entscheidung
	d = Intervention akzeptieren b = Intervention ablehnen f = Intervention ablehnen h = Intervention akzeptieren
	Keine Empfehlung für eine Entscheidung
	a = Ist der zusätzliche Effekt die zusätzlichen Kosten wert? j = Ist der verringerte Effekt akzeptabel aufgrund ebenfalls verringerter Kosten? e = Keine zusätzlichen Effekte, keine zusätzlichen Kosten; gibt es andere Gründe für Befürwortung/Ablehnung der Intervention?

Quelle: Nixon, Khan 2001; Darstellung: GÖG/BIQG

Hierarchische Entscheidungsmatrix

Diese Matrix ist eine Fortführung der Permutationsmatrix, bei der die Entscheidungsmöglichkeiten hierarchisch dargestellt werden. Die farbige Darstellung unterstützt die Orientierung in der Matrix.

Tabelle 8.6:
Hierarchische Matrix

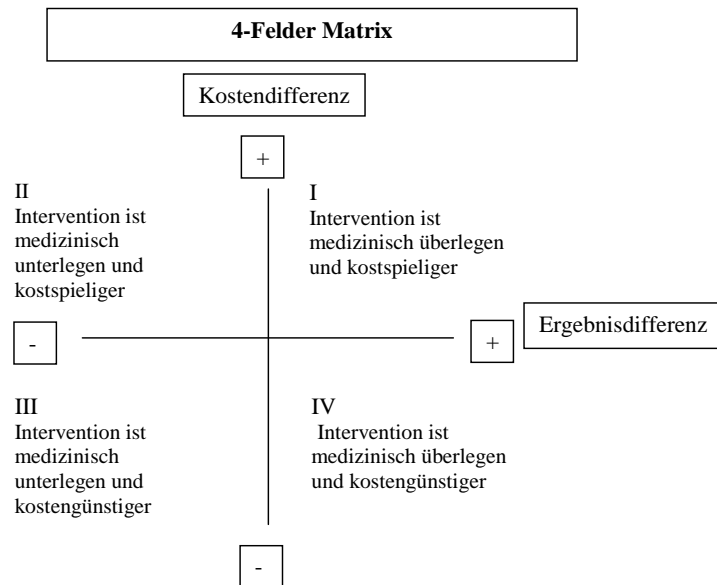
Kosten	Anzahl Studien	Gesundheits- effekte	Entscheidung
+		-	Intervention ablehnen
0		-	Intervention ablehnen
+		0	Intervention ablehnen
-		-	Inkrementelle Analyse notwendig
0		0	Neutral
+		+	Inkrementelle Analyse notwendig
-		0	Intervention akzeptieren
0		+	Intervention akzeptieren
-		+	Intervention akzeptieren

Quelle: (Nixon, Khan et al. 2001); Darstellung: GÖG/BIQG

Vier-Felder-Matrix (cost-effectiveness plane)

In der Vier-Felder-Matrix werden die Ergebnisse zu Kosten- und Effektdifferenzen in ein Koordinatensystem übertragen. Der Vorteil dieser Darstellung ist, dass nicht nur abgebildet wird, ob Kosten und Effekte höher, niedriger oder unverändert sind, sondern sichtbar gemacht wird auch die Größendimension dieser Veränderung. Der Nachteil ist, dass diese Darstellung etwas „technischer“ ist als die vorher angeführten Vorschläge und dass bei verschiedenen gesundheitsökonomischen Studientypen in einer Übersichtsarbeit (z. B. Kosten-Effektivitäts-Analyse, Kosten-Nutzwert-Analyse) deren unterschiedliche Outcomeparameter (z. B. LYG, QALYs, depressionsfreie Tage) nicht in ein und derselben Matrix dargestellt werden können.

Abbildung 8.1:
Vier-Felder-Matrix



Quelle und Darstellung: Schöffski, Graf von der Schulenburg 2007

Die Limitation bei den unterschiedlichen Matrizen besteht darin, dass die meisten gesundheitsökonomischen Evaluationen im Feld „höhere Kosten bei höherem Gesundheitseffekt“ angesiedelt sind und somit selten einfache Empfehlungen abgeleitet werden können, denn die Frage, ob der zusätzliche Effekt die zusätzlichen Kosten wert ist, ist eine politische, die nicht über eine Entscheidungsmatrix gelöst werden kann.

Integration von gesundheitsökonomischer Evidenz in das GRADE-Profil

GRADE (Grading of Recommendations Assessment, Development and Evaluation) ist eine anerkannte Methode zur Synthese der Evidenz von erwünschten und unerwünschten Effekten von Interventionen (siehe Kapitel 7.1 und 7.4). Es existieren erste Ansätze, gesundheitsökonomische Evidenz in die GRADE-Evidenzprofile zu integrieren (Brunetti, Ruiz et al. 2010). Allerdings sind davon laut derzeitigem Standard Studien, die auf entscheidungsanalytischen Modellen basieren, ausgeschlossen. Der methodische Ansatz ist so gestaltet, dass das Evidenzprofil um die Evidenz zu Ressourcenverbrauch und Kosten erweitert wird. Dazu wird folgende Vorgangsweise empfohlen:

1. Identifikation jener Ressourcen und Kosten, die potenziell relevant sind und möglicherweise zwischen den betrachteten Alternativen differieren;
2. Evidenz zu Ressourcenunterschieden zwischen den verglichenen Alternativen identifizieren;

3. Qualitätsprüfung der Evidenz;
4. monetäre Bewertung der Ressourcen,
adaptiert für den eigenen Entscheidungskontext.

Wie bei der klinischen Evidenz erfolgt die Qualitätsprüfung der gesundheitsökonomischen Evidenz im Rahmen der GRADE-Methode nach den Kriterien Studienlimitationen (z. B. fehlende Verblindung etc.), Konsistenz der Ergebnisse, Direktheit, Präzision und Publikationsbias. Bezüglich Konsistenz der Ergebnisse sind bei Ressourcendaten Variationen aufgrund unterschiedlicher Settings etc. möglich, die Qualität der Evidenz sinkt jedoch, wenn diese nicht erklärbar sind. Die Variabilität der Ressourcendaten führt nicht zuletzt dazu, dass größere Stichproben notwendig sind, um eine entsprechende Präzision der Ergebnisse zu erreichen. Ein Beispiel dazu ist in Tabelle 8.7 abgebildet.

Tabelle 8.7:

Beispiel für ein gesundheitsökonomisches Evidenzprofil auf Basis der GRADE-Methodologie bei Drogensatztherapiestudien

Question: Should buprenorphine maintenance flexible doses vs methadone maintenance flexible doses be used for opioid maintenance treatment? Patient or population: Opiate dependents Setting: Outpatients in USA, Australia, Austria, Switzerland, UK Viewpoint: Societal									
Studies (follow-up)	Quality assessment					Summary of resources and costs			Overall quality
	Design	Limitations	Inconsistency	Indirectness	Imprecision	No. of patients	Resources Costs per patient (1999 AU \$)		
							Methadone ^b	Buprenorphine ^b	
Drugs									
Harris 2005 (1 year) ^{a23}	RCT	Serious limitations ^g	No	Some uncertainty ^g	Small sample size	139	<i>Resources (mean daily)</i> 50 mg 14 mg <i>Costs (annual)</i> 1122 (85 SE) 1785 (204 SE)		Very low
Doran 2003 (6 months) ^{a24}	RCT	No	No	Some uncertainty ^g	–	405	<i>Resources (mean daily)</i> 57 mg 11 mg <i>Costs (6 months)</i> 37 (33 SD) 459 (461 SD)		Moderate
Other health care costs									
Harris 2005 (1 year) ^{b23}	RCT	Serious limitations ^g	No	Some uncertainty ^g	Small sample size	139	<i>Resources</i> NA NA <i>Costs (annual)</i> 2500 (489 SE) 3316 (667 SE)		Very low

Doran 2003 (6 months) ^{c24}	RCT	No	No	Some uncertainty ^g	-	405	Resources NA	NA	Moderate
Crime Costs							Costs (6 months) 1378 (NA)	1270 (NA)	
Harris 2005 (1 year) ^{d23}	RCT	Very serious limitations ^f	No	Some uncertainty ^g	Small sample size	139	Resources NA	NA	Very low
							Costs (annual) ^j 13,223 (10,209 SE)	6265 (2,028 SE)	

In this example we decided not to pool resource data from different studies due to insufficient information provided.
NA, not available; SD, standard deviation; SE, standard error.

^a Including dispensing fee
^b Includes other prescription and OTC drugs, prescriber, inpatient, outpatient, emergency, ambulance, counseling, allied health and pathology services
^c Include staff time (i.e. face-to-face contact and preparation time), diagnostic procedures and facility level (supplies, consumables, capital, equipment, ancillary support including administration, management, security, etc.)
^d Health care costs from assault, loss of income by the victims of crime, depreciated value of property damaged, stolen or obtained fraudulently, detection, prosecution and imprisonment
^e Some limitations because of incomplete outcome data
^f Some limitations because of incomplete outcome data and crucial limitations for self-reported crime data
^g All the studies were conducted within the Australian health system (while the recommendation was global)
^h Doses for methadone and buprenorphine derived from Doran 2003 study, at the 10th week²⁴
ⁱ The average cost of crime was substantial across the sample but these reported costs were associated with just a few participants. 90% of the sample randomized to methadone and 96% of that randomized on buprenorphine reported non-involvement in property crime. Indeed, the majority of patients reported no criminal activity during the trial (6/66 patients for methadone and 3/73 for buprenorphine)²³

Quelle und Darstellung: (Brunetti, Ruiz et al. 2010)

Das National Institute for Health and Clinical Excellence (NICE) hat einen eigenen Ansatz entwickelt, bei dem auch modellbasierte gesundheitsökonomische Evaluationen in das GRADE-Profil integriert werden (National Institut for Health and Clinical Excellence 2009; Brunetti, Ruiz et al. 2010). Hierfür wurde das GRADE-Profil adaptiert. Der NICE-Ansatz reduziert die Qualitätsprüfung auf die relevanten Kriterien Anwendbarkeit für den Entscheidungskontext und methodische Qualität. Für beide Kriterien existieren Prüflisten, die eine Differenzierung der Studien erlauben einerseits in „direkt anwendbar“, „teilweise anwendbar“ und „nicht anwendbar“ und andererseits in „Studien mit geringen Limitationen“, „Studien mit potenziell erheblichen Limitationen“ und „Studien mit erheblichen Limitationen“. Korrespondierend dazu werden die zentralen Studienergebnisse dargestellt (für ein Beispiel siehe Tabelle 8.8).

Tabelle 8.8:

Beispiel für NICE-Evidenzsynthese nach GRADE-Methodologie für gesundheitsökonomische Studien zu Omega-3-Säurenethylester-Ergänzungen nach Herzinfarkt

Quality assessment						Summary of findings			
Outcome	Design (No. studies)	Study limitations	Inconsistency	Indirectness	Imprecision	Supplements	Control	Relative risks	Quality
All-cause mortality	RCT (3)	Serious limitations ^a	No	Some indirectness ^b	No	581/6830 (8.5%)	755/6830 (11.1%)	0.83 (0.75 to 0.93)	Low
Combined CV events	RCT (3)	Serious limitations ^a	No	Some indirectness ^b	No	755/6830 (11.1%)	839/6826 (12.3%)	0.90 (0.82 to 0.99)	Low
Cancers	RCT (3)	Serious limitations ^a	No	Some indirectness ^b	No	150/6830 (2.2%)	138/6826 (2.0%)	1.09 (0.86 to 1.36)	Low

Quality assessment				Summary of findings			
Study	Applicability	Limitations	Other comments	Incremental cost ^c	Incremental effect	ICER	Uncertainty
Franzosi 2001 ⁷¹	Partially applicable ^d	Potentially serious limitations ^e	Based on measured resource use and survival in 3.5 years follow-up in GISSI-P trial	£871	0.0332 LY	£26,243 per LY	£16,769 to £56,025 per LY gained (best /worst case SA)
Lamotte 2006 ⁷²	Partially applicable ^f	Very serious limitations ^g	Based on measured resource use and survival over 3.5 years in GISSI-P, plus longer term survival benefits attributed to non-fatal events using Canadian database. Belgian results presented	£1090	0.282 LY	£3860 per LY	>98% probability ICER less than 120,000 per QALY gained (PSA)

NCC analysis ⁷³	Directly applicable ^h	Minor limitations ⁱ	Based on morbidity and mortality estimated from Markov model using pooled effectiveness data from GISSI-P and DART. Results sensitive to the size of treatment effects and over their assumed duration	£1073	0.09 QALY	£12,480 per QALY	£3912 to £130,705 per QALY gained (range in one-way SA)
----------------------------	----------------------------------	--------------------------------	--	-------	-----------	------------------	---

LY, life-years; QALY, quality-adjusted life-years; ICER, incremental cost-effectiveness ratio; SA, sensitivity analysis; PSA, probabilistic sensitivity analysis.

^a Increase in statin use over follow-up in GISSI-P trial differed between the groups (from 4.4% to 46.0% in the omega-3 group and from 5.1% to 44.4% in the control group)

^b High baseline rate of fish consumption in GISSI-P (more than 70%) – may not reflect usual diet in UK

^c Costs estimated in 2006 UK pounds – converted using PPP exchange rates (www.oecd.org/std/ppp) and UK HCHS inflation index (www.psr.u.ac.uk/pdf/uc/uc2006/uc2006.pdf)

^d Uses Italian estimates of resource use and unit costs. Some uncertainty over the applicability of Italian effectiveness data to UK

^e This study is relatively conservative, as it does not impute any quality of life or longer-term survival benefit to supplements. Conversely, it omits GI side effects

^f Uses Belgian estimates of resource use and unit costs. 5% discount rate for costs and effects. Some uncertainty over the applicability of Italian effectiveness data to UK

^g Methods and data used to estimate life expectancy are questionable, and were not subjected to sensitivity analysis. This is likely to have biased the results

^h Costs estimated following the NICE reference case, using UK resource use and unit cost estimates. Some uncertainty over applicability of Italian effectiveness estimates in UK population

ⁱ Analysis is based on best available effectiveness estimates and follows NICE methodological guidance. Minor limitations in reporting (e.g. for inputs taken from NICE statins appraisal)

Quelle und Darstellung: (Brunetti, Ruiz et al. 2010)

Ergänzend zur allen genannten Formen der Darstellung ist jedenfalls eine narrative Beschreibung und eine kontextbezogene Diskussion bzw. Interpretation notwendig. Darunter fällt die Angabe der Variationsbreite der Ergebnisse (z. B. der Kosten-Nutzwert-Verhältnisse), ihrer internen Validität und ihrer Generalisierbarkeit (siehe auch Kapitel 5). Außerdem erforderlich ist eine Erläuterung der unterschiedlichen analytischen Ansätze, die in den Studien angewendet wurden, und ihrer Robustheit (z. B. ob in den Studien gleiche Modelltypen und Modellstrukturen verwendet wurden).

8.1.4.6.3 Quantitative Synthese

Derzeit existieren keine methodischen Standards zum Poolen von Ergebnissen zur Kosteneffektivität (z. B. verschiedener inkrementeller Kosten-Effektivitäts-Verhältnisse) aus unterschiedlichen Studien mithilfe von Metaanalysen oder anderen quantitativen Methoden.

Laut Cochrane-Handbuch (Shemilt, Mugford et al. 2009) kann eine Metaanalyse unter Umständen für Studiendaten zu Ressourcenmengen oder Kosten in Betracht gezogen werden, sofern eine entsprechende Vergleichbarkeit gewährleistet ist. Da jedoch gerade die Daten zum Ressourcenverbrauch, zu den Preisen und somit zu den Kosten besonders kontextabhängig sind, ist sowohl beim Einsatz einer quantitativen Analyse als auch bei der Interpretation ihrer Ergebnisse große Vorsicht angesagt. Vor einer Metaanalyse von Kostendaten sind diese jedenfalls auf ein gemeinsames Preisjahr und eine gemeinsame Währung anzupassen (siehe Kapitel 5).

Jeder Metaanalyse gesundheitsökonomischer Daten soll grundsätzlich eine kritische Analyse der Daten, die in die Metaanalyse eingehen, vorausgehen. Sonstige Standards bei der Durchführung einer Metaanalyse wie etwa statistische Tests zur Ermittlung der Heterogenität zwischen den Studien (Chi-Quadrat-Test, „random-effect models“ etc.) sind ebenfalls zu berücksichtigen (siehe Kapitel 7.3.1). Ergänzend zu den Ergebnissen ist eine narrative Zusammenfassung zur Größenordnung, Richtung und Präzision der Ergebnisse notwendig.

8.1.5 Übertragbarkeit und Adaption von Ergebnissen vom Studien- auf den Entscheidungskontext

Die Ergebnisse gesundheitsökonomischer Studien hängen von verschiedenen Einflussfaktoren ab, die oft kontextspezifisch geprägt sind. Beispielsweise kann eine länderspezifische Leitlinie zur Erstellung gesundheitsökonomischer Evaluationen einen bestimmten Diskontierungsfaktor für Kosten und Gesundheitseffekte vorschreiben, der

einen großen Einfluss auf das Ergebnis hat und sich vom Diskontierungsfaktor im eigenen Entscheidungskontext unterscheidet. Daher können Ergebnisse gesundheitsökonomischer Studien meistens nicht einfach von einem Land auf ein anderes übertragen werden (Welte and Leidl 1999). Anders ausgedrückt: Die externe Validität einer gesundheitsökonomischen Studie ist häufig eingeschränkt.

Je nach Fragestellung der Übersichtsarbeit kann es entweder notwendig sein, die Übertragbarkeit bereits als Ausschlusskriterium bei der Studienselektion anzuwenden, oder die Studienergebnisse sind in unterschiedlichem Umfang zu adaptieren, um die Relevanz für den eigenen Entscheidungskontext zu erhöhen.

8.1.5.1 Faktoren, die Übertragbarkeit beeinflussen

Es existiert eine Reihe von Faktoren, die die Übertragbarkeit von gesundheitsökonomischen Studien beeinflussen. Welte et al. (2004) haben auf Basis einer Literaturrecherche eine Liste an Übertragbarkeitsfaktoren erstellt und nach unterschiedlichen Charakteristika klassifiziert.

Tabelle 8.9:
Übertragbarkeitsfaktoren von (Welte, Feenstra et al. 2004)

• Übertragbarkeitsfaktoren
Methodologische Charakteristika
» Perspektive
» Diskontierungsrate
» Ansatz zur Berechnung der medizinischen Kosten
» Ansatz zur Berechnung der Produktivitätskosten
Gesundheitssystemcharakteristika
» Absolute und relative Preise im Gesundheitssystem
» Praxisunterschiede
» Verfügbarkeit der Technologie
Bevölkerungscharakteristika
» Inzidenz und Prävalenz der Erkrankung
» Fallmischung
» Lebenserwartung
» Präferenzen für Gesundheitszustände
» Akzeptanz, Compliance, Anreize für Patienten
» Produktivitätskosten und Arbeitszeitverluste
» Ausbreitung der Erkrankung

Quelle: Welte, Feenstra 2004; LBI-HTA-eigene Darstellung 2011

Perspektive: Die gewählte Perspektive in einer gesundheitsökonomischen Evaluation hat einen zentralen Einfluss auf die Kosten in einer Studie. Aus der Perspektive der Krankenversicherungsträger sind etwa nur Kosten jener Leistungen relevant, für die

Krankenversicherungen zahlen, während aus einer gesellschaftlichen Perspektive zusätzlich sowohl Kosten außerhalb des Gesundheitssektors (z. B. Produktivitätsverluste) als auch von Patientinnen und Patienten selbst zu tragende (private) Kosten, Zeitverluste etc. in die Analyse einzubeziehen sind. Beispielsweise sind Evaluierungsergebnisse zur HPV-Impfung, die gesellschaftliche Kosten berücksichtigen, um bis zu 87 Prozent niedriger als Berechnungen mit ausschließlich direkten Kosten (Damm, Nocon et al. 2009). Sogar bei gleicher Perspektive zwischen Studienland und eigenem Entscheidungskontext kann die Übertragbarkeit eingeschränkt sein, wenn etwa Sozialversicherungsträger in Deutschland andere Kosten übernehmen als in Österreich.

Diskontierung: Bei der Diskontierung wird dem Phänomen Rechnung getragen, dass gegenwärtige Kosten und zukünftiger Nutzen weniger präferiert werden als zukünftige Kosten bzw. gegenwärtiger Nutzen. Zukünftige Kosten und Effekte werden mit einer Diskontrate auf die Gegenwart „abgezinst“. Die Diskontrate beeinflusst das Ergebnis oft maßgeblich, vor allem bei Technologien, deren Kosten oder Nutzen langfristig anfallen.

Ansatz zur Berechnung der medizinischen Kosten: Zur Berechnung der medizinischen Kosten werden länderspezifisch oft unterschiedliche Quellen herangezogen (Tarife, Marktpreise etc.), die zu Unterschieden im Ergebnis führen.

Ansatz zur Berechnung der Produktivitätskosten: Produktivitätskosten sind die monetär bewerteten Produktivitätsausfälle durch Krankheit oder frühzeitigen Tod. Die (mitunter in Leitlinien vorgeschriebene) Methode zur Berechnung der Produktivitätskosten (Humankapitalansatz, Friktionskostenmethode) kann unterschiedlich sein und führt gelegentlich zu sehr großen Unterschieden im Gesamtergebnis.

Absolute und relative Preise: Das Verhältnis zwischen den Kosten für mehrere Produkte (relative Kosten) kann variieren, und zwar sowohl zwischen den Ländern – z. B. ist die Behandlung von Gebärmutterhalskrebsvorstufen in Norwegen mit geringeren Kosten verbunden als in Österreich (Zechmeister, Freiesleben de Blasio et al. 2007) – als auch über die Zeit – z. B. sinken Kosten für Bandscheibenprothesen über die Zeit, während die Personalkosten konstant bleiben.

Praxisunterschiede: Unter die Praxisunterschiede fallen etwa die organisatorische Verortung einer Technologie oder die Berufsgruppen, die eine Leistung erbringen. Wenn beispielsweise geburtshilfliche Leistungen in einem Land vorrangig von Hebammen erbracht werden, sind die Kosten wahrscheinlich niedriger, als wenn dieselbe Leistung von Fachärzten oder Fachärztinnen erbracht wird. Auch wenn eine Leistung in einem Land stationär, in einem anderen aber ambulant erbracht wird, wird das die Kosten und somit die Übertragbarkeit der Ergebnisse beeinflussen.

Verfügbarkeit der Technologie: Die Diffusion einer Technologie beeinflusst maßgeblich das Ergebnis einer gesundheitsökonomischen Evaluation und kann länderspezifisch stark variieren. Zum Beispiel sind in einem Land mit weniger Standorten für Computertomographie längere Anfahrtszeiten für Patienten und Patientinnen, in weiterer Folge höhere Arbeitszeitverluste und somit höhere Produktivitätskosten verbunden. Außerdem können mit der Einführung einer neuen Technologie Anfangsprobleme verbunden sein, die sich nach längerer Anwendung reduzieren (Lerneffekte) und die ebenso zwischen Studienland und Entscheidungskontext unterschiedlich sein können.

Inzidenz und Prävalenz der Erkrankung: Inzidenz und Prävalenz können sowohl auf der Kosten- als auch auf der Effektseite die Ergebnisse beeinflussen (z. B. bei Screeningprogrammen).

Fallmischung: Erkrankungsrisiken, Komplikationsrisiken, Restlebenszeit etc. sind oft vom Alter der jeweiligen Individuen abhängig. Unterscheidet sich die Altersstruktur im Studienland wesentlich von der im Entscheidungskontext, kann es zu einer Über- oder Unterschätzung der Kosten-Effektivitäts-Verhältnisse kommen. Dasselbe gilt für die Geschlechterstruktur, für die Verteilung von Schweregraden einer Erkrankung oder für die Prävalenz von Komorbidität.

Lebenserwartung: Die Lebenserwartung beeinflusst alle Outcomeparameter, die die verbleibende Restlebenszeit (z. B. gewonnene Lebensjahre, QALYs) integrieren. Unterschiede in der Lebenserwartung beeinflussen somit die Übertragbarkeit der Studienergebnisse.

Präferenzen: Bei manchen Studientypen (z. B. Kosten-Nutzwert-Analysen) werden für die Ermittlung der Gesundheitseffekte einer Intervention Präferenzen für bestimmte Gesundheitszustände ermittelt. Diese können zwischen Ländern schwanken.

Akzeptanz, Compliance, Anreizwirkungen: Diese Faktoren haben einen Einfluss auf die Inanspruchnahme von Leistungen und somit sowohl auf die Kosten als auch auf die Gesundheitseffekte. Wenn die Compliance zu Cholesterinhemmern im Studienland höher ist als im eigenen Entscheidungskontext, werden die Kosten im Studienland höher sein, mitunter aber auch die Effektivität des Arzneimittels.

Produktivität und Arbeitszeitverlust: Die Produktivitätskosten hängen von der Wertschöpfung im jeweiligen Land ab. In einem Land mit hoher Wertschöpfung wird eine Technologie, die Krankenstände vermindert, zu einem besseren Kosten-Effektivitäts-Verhältnis führen als in einem Land mit niedriger Wertschöpfung.

Krankheitsausbreitung: Bei Infektionskrankheiten stellt die Krankheitsausbreitung, die wiederum durch weitere Faktoren bedingt ist (z. B. Lebensstil, Umweltfaktoren, geneti-

sche Prädisposition), einen Schlüsselindikator für den Erfolg von Präventionsmaßnahmen dar, der länderspezifisch variieren kann.

Manche dieser Faktoren (z. B. Perspektive) sind leicht zu überprüfen, andere kaum (z. B. Verbreitung der Technologie). Nach Prüfung der Faktoren auf ihre jeweilige Relevanz in den Studien der Übersichtsarbeit kann als weiterer Schritt eine Adaption besonders kritischer Faktoren erfolgen. Wenn dies nicht möglich ist, so ist es bei Übersichtsarbeiten, die auf einen bestimmten Entscheidungskontext ausgerichtet sind, zumindest sinnvoll, die identifizierten relevanten Faktoren qualitativ zu beschreiben und darzustellen, in welche Richtung sich das Ergebnis durch die Rahmenbedingungen im eigenen Land verändern würde (für ein Beispiel dazu siehe Box 8.3). Hilfreich ist auch, auf Basis der Sensitivitätsanalysen aus den Originalstudien das Ausmaß der Auswirkung einzelner Faktoren auf das Studienergebnis zu identifizieren und damit den Einfluss dieser Faktoren auf das Ergebnis im eigenen Entscheidungsland abzuschätzen.

Box 8.3:

Beispiel für die qualitative Beschreibung der Übertragbarkeit (Faktor „Produktivität“)

„... Da die Statinbehandlung eine Verminderung des Produktionsausfalls nach sich zieht, sind die potenziellen indirekten Kosteneinsparungen in Ländern mit hoher Wertschöpfung höher, das IKEV ist daher niedriger (also günstiger). Die [...] relevante Studie rechnet mit schwedischen Wirtschaftsdaten aus dem Jahr 1995. Das schwedische BIP-pro-Kopf zu laufenden Preisen und Kaufkraftparitäten ist seit 1995 geringfügig niedriger als das österreichische (Statistik Austria 2005). Die indirekte Kosteneinsparung könnte daher in Österreich marginal höher sein, was ein günstigeres IKEV bedeuten würde.“ (Zechmeister and Wild 2006)

Eine Möglichkeit für die systematische Überprüfung der Übertragbarkeit ist,

- » die in **Fehler! Verweisquelle konnte nicht gefunden werden.** angeführten Übertragbarkeitsfaktoren im Hinblick auf ihre Relevanz in der behandelten Studie zu analysieren,
- » die Übereinstimmung des jeweiligen Faktors zwischen Studienland und Entscheidungsland systematisch abzuschätzen,
- » daraus die Übereinstimmung, Über- oder Unterschätzung der Kosten-Effektivitäts-Verhältnisse zwischen Studienland und Entscheidungsland für die Übertragbarkeitsfaktoren zu beschreiben.

Ein Beispiel dafür ist in Tabelle 8.11 angeführt.

Tabelle 8.10:

Beispiel für systematische Überprüfung der Übertragbarkeit: Transferability of the Belgium Netherlands Stent Study II to Germany

■

Quelle und Darstellung: Welte, Feenstra 2004

Ein ähnlicher Ansatz wird derzeit in den Niederlanden entwickelt (Grutters, Seferina et al. 2010).

8.1.5.2 Adaptionmethoden

Der Bedarf für die tatsächliche Anpassung von Studienergebnissen in gesundheitsökonomischen Übersichtsarbeiten hängt von Forschungsfrage und Zielgruppe der Arbeit ab. Beispielsweise wird eine gesundheitsökonomische Übersichtsarbeit im Rahmen eines Cochrane-Berichts für eine internationale Zielgruppe erstellt. Eine Anpassung von Studiendaten wird nur in geringem Ausmaß sinnvoll sein, da es keinen einheitlichen Entscheidungskontext gibt, auf den die Anpassung zu beziehen ist. Bei einer gesundheitsökonomischen Übersichtsarbeit für länder-, regionen- oder institutenspezifische Entscheidungsträger kann umgekehrt eine Anpassung von Studiendaten im Rahmen der Datensynthese die Anwendbarkeit der Übersichtsarbeit im tatsächlichen Entscheidungsprozess wesentlich verbessern.

Eine Adaption kann in unterschiedlichem Umfang erfolgen und reicht von relativ einfachen Methoden der Inflations- und Währungsanpassung bis hin zur Adaption von Ressourcen- bzw. Kostendaten oder ganzen Modellstrukturen. Die Durchführung einer Adaption hängt ab von der Detailliertheit der präsentierten Daten in den Studien, von der Verfügbarkeit von Daten für den eigenen Entscheidungskontext und vom Gesamtaufwand der Adaption. Wenn zum Beispiel in den Studien nur die Gesamtkosten angegeben werden, nicht aber die Mengen- und Preisgerüste der einzelnen Kostenelemente, können die Kosten nicht adaptiert werden. Umgekehrt kann auch bei fehlenden Kostendaten aus dem eigenen Entscheidungskontext keine Adaption erfolgen. In manchen Fällen wird es überhaupt sinnvoll sein, statt einer aufwendigen Adaption eine eigene Studie durchzuführen.

Inflations- und Währungsanpassung

Die einfachste Form der Anpassung, die eine Vergleichbarkeit der Studien einer Übersichtsarbeit wesentlich erleichtert und in allen Handbüchern erwähnt wird, ist eine Inflations- und Währungsanpassung. Diese Anpassung ist deshalb notwendig, da die Studien oft aus unterschiedlichen Jahren und Währungssystemen mit unterschiedlichen Preisniveaus stammen oder – auch wenn sie aus gleichen Währungssystemen (z. B. Eurozone) stammen – die Kaufkraft in den Ländern unterschiedlich ist. Zur genauen Methode der Inflations- und Preisanpassung enthalten die meisten der in Kapitel 2 zitierten Methodenhandbücher keine Information. Für Österreich wird folgende Vorgangsweise vorgeschlagen:

Die Inflationsanpassung ist erforderlich, wenn das Basisjahr, auf das sich die Kosten in den untersuchten Studien bezieht, nicht mit dem Jahr übereinstimmt, für das man die Ergebnisse in der Übersichtsarbeit benötigt. Die Anpassung erfolgt mittels sogenannter Deflatoren, die unter Verwendung von Preisindizes berechnet werden können. Die OECD erstellt Preisindizes gesamtwirtschaftlich (z. B. GDP-Preisindex), sektorenspezifisch (z. B. expenditure on in-patient-care, pharmaceutical expenditure) und aggregiert (z. B. total expenditure on health care). Die sektorenspezifischen Preisindizes haben zwar den theoretischen Vorteil, dass sie die Preisentwicklung medizinischer Güter besser widerspiegeln, sie beruhen jedoch oft auf einem kleinen Teil der interessierenden Ausgaben im jeweiligen Sektor und sind nicht ausreichend standardisiert (Welte and Leidl 1999). Für den gesamtwirtschaftlichen Index (GDP-Preisindex) gelten diese Nachteile nicht, er ist für alle OECD-Länder verfügbar und für aktuellere Jahre vorhanden. Ein pragmatischer Ansatz (Welte and Leidl 1999) geht daher davon aus, den gesamtwirtschaftlichen Index zu verwenden. Die „CPS Task Force“ (Task Force on Community Preventive Services 2010) hat hingegen ein differenziertes System entwickelt, in dem die Wahl des Preisindex von der Art der Kosten in den jeweiligen Studien abhängig ist. Wenn etwa die Programmkosten und die vermiedenen Kosten primär nicht-medizinisch sind (z. B. in Public-Health-Programmen), ist dort der „Consumer Price Index Urban“ (CPI-U) zu verwenden. Für Österreich wird empfohlen, die gesamtwirtschaftlichen Indizes (GDP-Indizes) einzusetzen.

Den Wert für das gesuchte Jahr (W_x) erhält man, indem der Wert aus der Studie (W_s) mit dem Quotienten aus dem Index des gesuchten Jahres (PI_x) und dem Index des Studienjahrs (PI_a) multipliziert wird: $W_x = W_s (PI_x / PI_a)$. Ein Beispiel dazu ist in Box 11 dargestellt. Eine Liste länderspezifischer GDP-Preisindizes findet sich im Anhang Tabelle 15.8.

Box 8.4:
Inflationsbereinigung mit Preisindizes

In der Studie ist das Kosten-Effektivitäts-Verhältnis in CAD pro LYG zu 2004-Preisen angegeben. Gesucht wird der inflationsbereinigte Wert für 2007 auf Basis der GDP-Indizes.

Angegebenes Kosten-Effektivitäts-Verhältnis (W_s): CAD 100.000 / LYG

Preisindex 2004 (PI_a): 109

Preisindex 2007 (PI_x): 119

$$W_x = W_s (PI_x / PI_a)$$

$$W_x = 100.000 (119/109)$$

$$W_x = 109.174/\text{LYG}$$

Das inflationsbereinigte Kosten-Effektivitäts-Verhältnis beträgt CAD 109.174/LYG

Für die Währungskonversion wird die Verwendung von Kaufkraftparitäten (purchasing power parities, PPP) empfohlen, da Wechselkurse zu verzerrten Resultaten führen. Kaufkraftparitäten geben an, wie viele Einheiten der jeweiligen Währung notwendig sind, um den gleichen repräsentativen Güterkorb zu kaufen, den man für einen US-Dollar in den USA erhält. Es gibt gesamtwirtschaftliche (GDP-PPP) und sektorenspezifische Kaufkraftparitäten (Health PPP, Medical Services PPP, Pharmaceutical PPP). Die OECD Health Data beinhalten nur die gesamtwirtschaftlichen und die pharmazeutischen Kaufkraftparitäten. In Anlehnung an Welte et al. (2004) wird ein pragmatischer Weg vorgeschlagen und empfohlen, für alle Konversionen die GDP-PPPs zu verwenden, da die sektorenspezifischen Paritäten oft unrepräsentativ sind. Wenn sowohl eine Inflations- als auch eine Währungsanpassung durchgeführt wird, soll der Konsistenz wegen jedenfalls dasselbe Bezugssystem für die Kaufkraftparitäten und die Preisindizes verwendet werden (z. B. GDP-Preisindex und GDP-PPP) (Welte and Leidl 1999).

Da die Kaufkraftparitäten in USD angegeben sind, erfolgt die konkrete Berechnung des gesuchten Wertes (W_E) im ersten Schritt durch Umrechnung des Betrages in der Studienwährung (W_S) in USD (W_{US}) mit der PPP-Rate für das Studienland (PPP_S): $W_{US} = W_S * 1 / PPP_S$; der damit erhaltene Betrag in USD muss dann mit der PPP-Rate des jeweiligen „Entscheidungslandes“ (PPP_E) multipliziert werden: $W_E = W_{US} * PPP_E$ (für ein Beispiel siehe Box 12). Eine Liste länderspezifischer Kaufkraftparitäten ist im Anhang dargestellt.

Box 8.5:

Beispiel für Währungskonversion

In der Studie ist das Kosten-Effektivitäts-Verhältnis in CAD/LYG zu 2007-Preisen angegeben. Gesucht wird der währungskonvertierte Wert in Euro, bezogen auf österreichische Kaufkraft.

Angegebenes Kosten-Effektivitäts-Verhältnis (W_C): CAD 109.174/LYG

GDP-PPP Canada 2007 (PPP_S): 1,21

GDP-PPP Österreich 2007 (PPP_E): 0,88

$$W_{US} = W_S * 1 / PPP_S = 109.174 * 0,83 = \text{US\$ } 90.614/\text{LYG}$$

$$W_E = W_{US} * PPP_E = 90.614 * 0,88 = \text{€ } 79.741/\text{LYG}$$

Der kaufkraftbereinigte und währungskonvertierte Wert für das Kosten-Effektivitäts-Verhältnis beträgt € 79.741/LYG.

Zur Frage, ob zuerst die Inflationsanpassung oder die Währungsanpassung erfolgen soll, existieren keine gültigen Standards. Welte et al. (2004) empfehlen, als ersten Schritt die Inflationsanpassung der Kosten mit den im jeweiligen Studienland gültigen Preisindizes und dann die Währungskonversion durchzuführen. Dies wird damit begründet, dass möglichst viele Anpassungen innerhalb des Studienlandkontextes erfolgen sollen.

Im Formular zur Datenadaption der CPS Task Force (Task Force on Community Preventive Services 2010) ist hingegen die Währungsanpassung vor der Inflationsanpassung vorzunehmen. Eine Begründung dafür wird nicht genannt. Für die Durchführung der Währungskonversion vor der Inflationsanpassung spricht, dass primär die Kaufkraftentwicklung und somit auch die Inflationsentwicklung des Entscheidungslandes interessieren. Wird die Währungskonversion vor der Inflationsanpassung durchgeführt, sind zuerst die Schritte unter Box 10 durchzuführen und der erhaltene Betrag W_E mit den entsprechenden Preisindizes des Entscheidungslandes anzupassen. Im obigen Beispiel würde das zuerst konvertierte Kosten-Effektivitäts-Verhältnis 70.470 Euro pro LYG betragen ($100.000 \cdot 0,81 \cdot 0,87$). Nach der Inflationsbereinigung beträgt das Kosten-Effektivitäts-Verhältnis 74.698 Euro pro LYG ($70.470 \cdot 1,06$).

Sonstige Adaptionismethoden

Neben der Inflations- und Währungsanpassung existieren aufwendigere Adaptionismethoden, die bis hin zur Adaption ganzer Modelle reichen. Da die Kosten in einer gesundheitsökonomischen Studie häufiger in ihrer Übertragbarkeit eingeschränkt sind als die Effekte (Sculpher, Pang et al. 2004; Welte, Feenstra et al. 2004), ist es sinnvoll, zunächst die Kostenparameter (also die Mengen und Preise der eingesetzten Ressourcen) an den eigenen Entscheidungskontext anzupassen. Damit kann die Übertragbarkeit fast aller angeführten Faktoren wesentlich verbessert werden. Konkret bedeutet das, zu überprüfen, wo die in den Studien beschriebenen Mengen für die im Rahmen einer Intervention benötigten Ressourcen (z. B. Häufigkeit bestimmter Chemotherapien, Operationsmethoden, Laboruntersuchungen etc.) von den Häufigkeiten im eigenen Entscheidungskontext abweichen (z. B. durch länderspezifisch unterschiedliche Behandlungsleitlinien) und bei Differenzen die eigenen Mengenangaben für die Berechnung zu verwenden. Dasselbe gilt für die Preise der einzelnen Leistungen. Voraussetzung dafür ist jedoch eine detaillierte und in disaggregierter Form präsentierte Information zu den Kosten in den Studien und im eigenen Entscheidungskontext.

Eine technische Form der Adaption gibt die CPS Task Force vor (Task Force on Community Preventive Services 2010). Der Adaptionsprozess ist genau festgelegt. Er beginnt mit der Währungsanpassung, gefolgt von der Inflationsbereinigung. Schritt drei ist die Adaption der Diskontierungsrate, danach folgt die Umrechnung der Programmkosten sowie der Kosteneinsparungen und der Produktivitätsverluste. Die Vorlage beschreibt anschließend die Konvertierung diverser Outcomeparameter in QALYs. Unter Angabe weiterer, nicht adaptierbarer Unterschiede wird schließlich für jede Studie ein Wert für das adaptierte Ergebnis der gesundheitsökonomischen Evaluation ermittelt. Zu allen Adaptionsschritten werden die mathematischen Formeln für die Berechnung angegeben. Die Originaldaten sowie die adaptierten Werte für Kosten und Outcomes bzw. für das Endergebnis werden abschließend für jede Studie tabellarisch dargestellt.

Eine solche Vorgangsweise ist dann geeignet, wenn die gesundheitsökonomische Evidenz zu unterschiedlichsten Interventionen und zu diversen Indikationen standardisiert dargestellt werden soll, um eine Vergleichbarkeit zwischen verschiedenen Interventionen über Indikationen hinweg zu gewährleisten.

8.1.6 Resümee

Qualität und Anwendbarkeit gesundheitsökonomischer Übersichtsarbeiten hängen von einer relevanten Forschungsfrage ab. Ob eine Intervention im Vergleich mit Alternativen kosteneffektiv ist oder nicht, wird aufgrund der Charakteristika gesundheitsökonomischer Studien kaum im Rahmen einer Übersichtsarbeit valide zu beantworten sein. Wohl aber kann eine gesundheitsökonomische Übersichtsarbeit mit einem explorativen Ansatz Information über die wichtigsten Einflussfaktoren auf die Kosteneffektivität oder Basisinformation für die Erstellung einer eigenen gesundheitsökonomischen Studie liefern.

Eine gesundheitsökonomische Übersichtsarbeit kann aufgrund der Heterogenität der Fragestellungen und der Studien weniger standardisiert werden als eine Übersichtsarbeit zur Wirksamkeit. Die zentralen Merkmale einer Übersichtsarbeit – Transparenz und Reproduzierbarkeit – gelten jedoch uneingeschränkt auch für die gesundheitsökonomische Übersichtsarbeit.

Die angeführten Inhalte zum Prozess und zu den methodischen Details einer gesundheitsökonomischen Übersichtsarbeit sind eine Hilfestellung, basierend auf dem derzeitigen Stand der Wissenschaft. Die konkrete Durchführung muss auf das individuelle Projekt und den Zweck der Übersichtsarbeit abgestimmt werden. In diesem Zusammenhang kann der Austausch im Rahmen bestehender Methodennetzwerke (z. B. Cochrane subgroup) eine zusätzliche wertvolle Hilfe bei spezifischen methodischen Fragen sein. Im Anhang unter Punkt 15.6.1 finden sich Tabellen und Checklisten zu diesem Kapitel.

Autorin: Ingrid Zechmeister (Ludwig Boltzmann Institut für HTA)

8.2 Budgetauswirkungsanalyse

8.2.1 Definition

Neben gesundheitsökonomischen Studien wie der Kosten–Kosten–Analyse, der Kosten–Nutzen–Analyse, der Kosten–Wirksamkeits–Analyse, der Kosten–Nutzwert–Analyse, der Kostenanalyse sowie der Krankheitskostenanalyse können auch Budgetauswirkungsanalysen (engl.: budget impact analysis, BIA) zur gesundheitsökonomischen Bewertung einer medizinischen Technologie oder Strukturveränderung durchgeführt werden.

Eine international einheitliche Definition bzw. Methodik zur Erstellung der Budgetauswirkungsanalyse existiert (bislang) nicht, die Inhalte variieren je nach Einsatz von (länderspezifischen) Richtlinien und gesundheitspolitischen Anforderungen (Garattini 2011, Mauskopf 2007). Beispielsweise finden sich für Deutschland allgemeine methodische Grundlagen einer Budgetauswirkungsanalyse im IQWiG–Handbuch „Allgemeine Methoden zur Bewertung von Verhältnissen zwischen Nutzen und Kosten“ (IQWiG 2009).

Generell werden im Rahmen einer Budgetauswirkungsanalyse die direkten finanziellen Auswirkungen (Folgekosten bzw. –erlöse) bewertet, die aufgrund der Einführung bzw. Erstattung einer neuen Gesundheitstechnologie bzw. Änderung der Nutzung einer bestehenden Gesundheitstechnologie in einem Gesundheitssystem entstehen (Garattini 2011, Mauskopf 2007, Nuijten 2011, Patented Medicine Prices Review Board 2007, Trueman 2001). Darunter fallen beispielsweise Finanzierungsflüsse, die für den potenziellen Anteil der Patientinnen und Patienten, die die neue bzw. veränderte Therapie erhalten werden, anfallen, Finanzierungsflüsse, die der Verbreitung der Therapie im Gesundheitssystem dienen, sowie Finanzierungsflüsse, die aufgrund der Anwendung der Therapie bei zuvor unbehandelten Patientinnen und Patienten anfallen werden (Mauskopf 2007).

Die begrifflichen Grundlagen sowie die Ermittlung und Berechnung von Folgekosten bzw. –erlösen sind in Österreich in der „Verordnung betreffend Richtlinien für die Ermittlung und Erstellung und Darstellung der finanziellen Auswirkungen neuer rechtsetzender Maßnahmen“ (BGBl. II Nr. 50/1999 idgF) gesetzlich verankert. Unter Folgekosten bzw. –erlösen werden Vollzugskosten (Kosten, die beim Vollzug von Rechtsnormen oder Maßnahmen anfallen, z. B. Personalkosten, Verwaltungssachkosten) und Nominalkosten (Transferzahlungen, materielle bzw. immaterielle Leistungen eines öffentlichen Rechtsträgers an Einzelpersonen, Personengruppen oder andere öffentliche Rechtsträger und Institutionen) subsumiert, die bei Einführung einer neuen rechtssetzenden Maßnahme entstehen. Entstehungskosten (Erstanschaffungs-, Erwei-

terungs-, Ergänzungs- oder Ersatzinvestitionen für neue rechtssetzende Maßnahmen) bleiben im Regelfall bei der Berechnung finanzieller Auswirkungen außer Acht (BGBl. II Nr. 50/1999 idgF). Weiters ist die Erstellung einer Budgetauswirkungsanalyse eine Möglichkeit, den gesetzlichen Regelungen des Allgemeinen Sozialversicherungsgesetzes (ASVG § 133, Ökonomiegebot), des Erstattungskodex nach § 351 ASVG sowie der Richtlinie über die Berücksichtigung ökonomischer Grundsätze bei der Krankenversicherung (RöK; gemäß § 31 Abs 5 Z 10 ASVG) Rechnung zu tragen (ASVG, GÖG 2009; siehe auch Kapitel 8.4.).

Eine Budgetauswirkungsanalyse kann die Unsicherheit bei der Planung und Organisation des zur Verfügung stehenden Budgets bzw. bei Prognosen für das Gesundheitssystem für die Entscheidungs- und Finanzierungsträger verringern, d. h., dass eine Budgetauswirkungsanalyse zeigt, mit welchen direkten finanziellen Auswirkungen der Kostenträger rechnen muss. Der Umfang einer BIA ist von Faktoren wie dem Anwendungsbereich bzw. Setting, der Fragestellung, der zu untersuchenden Technologie (z. B. Arzneimittel, Geräte) und dem vorhandenen Budget für die Durchführung der Analyse abhängig und deshalb zuvor mit der Auftraggeberin / dem Auftraggeber entsprechend den jeweiligen Anforderungen abzustimmen (Garattini 2011, Mauskopf 2007, Orlewska 2004).

8.2.2 Abgrenzung der Budgetauswirkungsanalyse zu anderen gesundheitsökonomischen Studien

Generell sollte eine Budgetauswirkungsanalyse als Ergänzung zu gesundheitsökonomischen Studien von Gesundheitstechnologien vor Zulassungs- bzw. Erstattungsentscheidungen bzw. bei einer veränderten Nutzung einer bestehenden Gesundheitstechnologie durchgeführt werden (Garattini 2011, Mauskopf 2007, Orlewska 2004, Truman 2001). Jedoch kann eine Budgetauswirkungsanalyse auch als alleinstehendes Produkt durchgeführt werden, wenn bestimmte Informationen bereits vorliegen (siehe Kapitel 2.4).

Bei beispielhafter Gegenüberstellung einer Budgetauswirkungsanalyse und einer Kosten-Effektivitäts-Analyse zeigt sich, dass die Ergebnisse der Analysen unterschiedlich ausfallen können, obwohl beide Studientypen auf ähnlichen Dateninhalten bzw. methodischen Anforderungen basieren. Dies begründet sich darin, dass die Budgetauswirkungsanalyse primär die Sicht des Kostenträgers beinhaltet und die Bezahlbarkeit einer Intervention, einer neuen Therapie oder eines neuen Medikaments bewertet. Im Gegensatz dazu wird bei der Kosten-Effektivitäts-Analyse ein einzelner klinischer Outcome in Form einer natürlichen Einheit dargestellt (z. B. gewonnene Lebensjahre, vermiedene postoperative Infektionen, depressionsfreie Tage) und den Kosten gegenübergestellt, um die ökonomische Wirksamkeit einer Gesundheitstechnologie zu

bewerten (Garattini 2011, Mauskopf 2007, Nuijten 2011, Orlewska 2004; siehe auch Kapitel 8.1).

Bei Vergleich einer Budgetauswirkungsanalyse mit anderen gesundheitsökonomischen Studien liegen die Besonderheiten einer Budgetauswirkungsanalyse in der Studienperspektive (Sicht des Kostenträgers), in der Betrachtung eines kurzen Zeithorizonts, in der ausschließlichen Betrachtung der Kosten ohne Berücksichtigung des Nutzens einer Technologie bzw. eines Technologiemicx sowie in einer eingeschränkten Fragestellung bzw. in einem abgegrenzten Setting (Garattini 2011, Nuijten 2011, Mauskopf 2007). In folgender Tabelle werden die Unterschiede explizit dargestellt (siehe Tabelle 8.11).

Tabelle 8.11:
Unterschiede zwischen Budgetauswirkungsanalysen
und gesundheitsökonomischen Studien

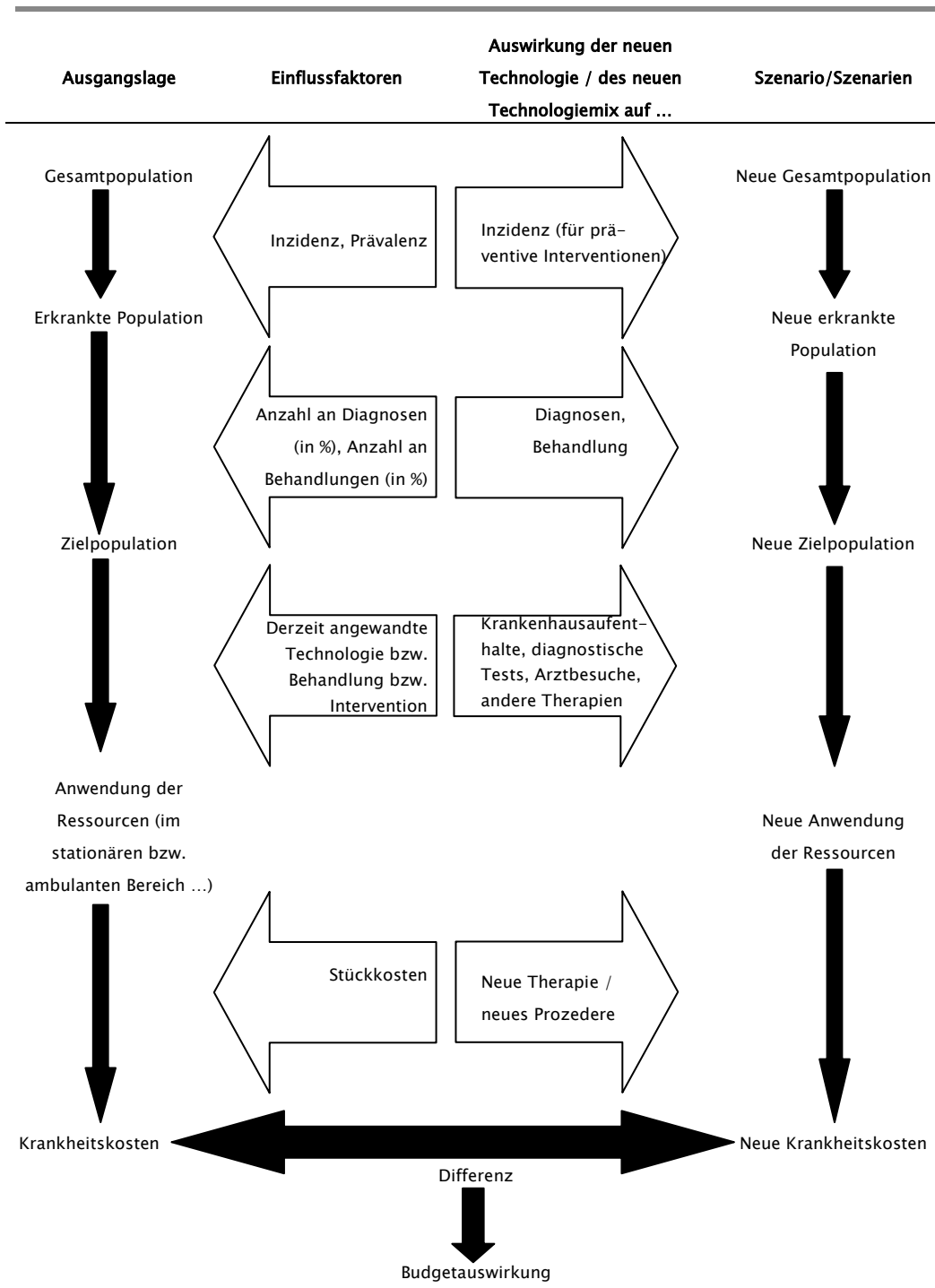
	Budgetauswirkungsanalyse	Gesundheitsökonomische Studien
Studienperspektive	Primär: Kostenträger	Unterschiedlich (siehe ökonomische Empfehlung) – Sicht der gesamten Gesellschaft, des gesamten oder öffentlichen Gesundheitssystems, der Sozialversicherungen, Krankenkassen, Krankenanstalten, Ärzte und Patienten
Zeithorizont	Üblicherweise kurz (≤ 3 Jahre) bis mittelfristig (≤ 5 Jahre) – abhängig von den Erfordernissen des Kostenträger sowie der Haushaltsperiode	Unterschiedlich (siehe ökonomische Empfehlung) – abhängig von Forschungsfrage bzw. vom Studiengegenstand
Kosten	Direkte Kosten (meist Behandlungskosten)	Unterschiedlich (siehe ökonomische Empfehlung) – direkte Kosten, indirekte Kosten oder Opportunitätskosten
Fragestellung/Setting	Eingeschränkt bzw. abgegrenzt	Breit
Effektivität	Betrachtung der Kosten einer Technologie bzw. eines Technologiemicx (Nutzen wird nicht berücksichtigt)	Betrachtung von Kosten und Nutzen (siehe ökonomische Empfehlung)
Ergebnisse	Differenz zwischen den Kosten, inkrementelle Gesundheitskosten	Differenz des Kosten-Nutzen-Verhältnisses, z. B. QALY (siehe ökonomische Empfehlung)
Sensitivitätsanalyse	Analyse von Extremen / Analyse von verschiedenen Szenarien	Unterschiedliche Durchführung (siehe Kapitel 8.4 bzw. 8.3)

Quelle und Darstellung: GÖG/BIQG

8.2.3 Erstellung und Durchführung einer Budgetauswirkungsanalyse

Für die Erstellung einer BIA werden die zukünftigen finanziellen Auswirkungen einer Gesundheitstechnologie unter Verwendung von Szenarien abgebildet (Ausgangslage versus ein Szenario oder mehrere Szenarien). Zur Erstellung werden benötigt u. a. Informationen und Daten zur betroffenen (derzeitigen und zukünftigen) Patientenpopulation und Patientencharakteristika, zu derzeit angewandten Interventionen und zu deren Kosten, zu neuen Interventionen und zu deren Kosten sowie zur Anwendung anderer Therapien und Prozesse sowie deren Kosten (siehe Tabelle 8.13, adaptiert nach Mauskopf 2007).

Abbildung 8.2:
Zu berücksichtigende Aspekte in einer Budgetauswirkungsanalyse



Quelle: adaptiert nach Mauskopf 2007; Darstellung: GÖG/BIQG

In Tabelle 8.12 und in Tabelle 8.13 wird ein vereinfachtes Beispiel zur Berechnung der Budgetauswirkung ein Jahr nach Einführung eines neuen Arzneimittels (C) anhand von zwei Szenarien für die öffentliche Versicherung dargestellt.

Tabelle 8.12:

Vereinfachtes Beispiel zur Berechnung der Budgetauswirkung (Angabe zur Berechnung)

	Arzneimittel A	Arzneimittel B	Arzneimittel C
Derzeitiger Marktanteil (vor Einführung von Arzneimittel C)	60 %	40 %	-
Voraussichtlicher Marktanteil (ein Jahr nach Einführung von Arzneimittel C)	59 %	39 %	2 %
Derzeitiger Anteil an Verordnungen pro Jahr (vor Einführung von Arzneimittel C)	6 Millionen	4 Millionen	-
Voraussichtlicher Anteil an Verordnungen pro Jahr (ein Jahr nach Einführung von Arzneimittel C)	5,9 Millionen	3,9 Millionen	200.000
Kosten pro Verordnung	500 €	150 €	1.000 €

Quelle und Darstellung: GÖG/BIQG

Tabelle 8.13:

Vereinfachtes Beispiel zur Berechnung der Budgetauswirkung (Berechnung)

Arzneimittel	Szenario 1 (ohne Einführung Arzneimittel C)	Szenario 2 (mit Einführung Arzneimittel C)
A	Gesamtkosten für Arzneimittel A = 6 Millionen x 500 € = 3 Milliarden €	Gesamtkosten für Arzneimittel A = 5,9 Millionen x 500 € = 2,95 Milliarden €
B	Gesamtkosten für Arzneimittel B = 4 Millionen x 150 € = 600 Millionen €	Gesamtkosten für Arzneimittel B = 3,9 Millionen x 150 € = 585 Millionen €
C	Gesamtkosten für Arzneimittel C = 0 € , da es in diesem Szenario kein C gibt	Gesamtkosten für Arzneimittel C = 200.000 x 1.000 € = 200 Millionen €
Gesamtkosten	= 3,6 Milliarden €	= 3,74 Milliarden €
Budgetauswirkung	Differenz (3,74 Milliarden € - 3,6 Milliarden €) = 135 Millionen € Das heißt: Wenn das Arzneimittel C eingeführt wird, kostet es den Kostenträger um 135 Millionen € mehr als bei Nichteinführung (Anmerkung: Weiters sollte eine Sensitivitätsanalyse durchgeführt werden).	

Quelle und Darstellung: GÖG/BIQG

Im Anhang findet sich ein ausführliches Beispiel zur Berechnung der Budgetauswirkung anhand epidemiologischer Daten (siehe Anhang 15.8).

Heranzuziehende Datenquellen für die Erstellung einer Budgetauswirkungsanalyse umfassen beispielsweise Primärdaten aus publizierten klinischen Untersuchungen bzw. Einschätzungen, Vergleichsstudien für Wirksamkeit und Sicherheit der derzeitigen bzw. neuen Technologie, Daten des Entscheidungsträgers zur Population, lokale bzw. nationale statistische Informationen, publizierte bzw. nicht publizierte epidemiologische Daten, Registerdaten sowie im Bedarfsfall Expertenmeinungen (Mauskopf 2007).

Inhaltlich sowie formal soll sich eine Budgetauswirkungsanalyse mit folgenden Punkten auseinandersetzen: Forschungsfrage, Studiendesign, Perspektive, Patientenpopulation, Szenarien, Intervention und Alternativen, Zeithorizont, Ressourceneinsatz und Kosten, Variabilität und Unsicherheit sowie Verallgemeinerbarkeit und Übertragbarkeit der Ergebnisse (Mauskopf 2007, Patented Medicine Prices Review Board 2007). Die einzelnen Aspekte werden in der ökonomischen Empfehlung (siehe Kapitel 8.4) erläutert und gelten – in Abstimmung mit den jeweiligen Anforderungen – auch für die Erstellung einer Budgetauswirkungsanalyse. Neben den genannten Punkten ist auch auf die Berichtsform und Publikation der Ergebnisse zu achten, die in der ökonomischen Empfehlung näher betrachtet werden (siehe Kapitel 8.4).

Autorin: Daniela Pertl (GÖG/BIQG)

8.3 Gesundheitsökonomische Modellierung

Zunehmend werden in klinischen Studien auch ökonomische Daten erhoben (Ramsey, Willke et al. 2005). Allerdings sind diese Daten häufig entweder zeitlich begrenzt, oder die Bedingungen einer klinischen Studie erlauben es nicht, den Ressourcenverbrauch im Versorgungsalltag realistisch wiederzugeben. So kann der Ressourcenverbrauch protokollinduziert erhöht sein, oder die Versorgung der Patienten spiegelt nicht die Situation im gewünschten regionalen Kontext wider. Das Heranziehen von Daten aus anderen Studien im Rahmen eines entscheidungsanalytischen Modells ist oft unumgänglich (Drummond and Davies 1991; Sculpher, Claxton et al. 2006; Weinstein 2006).

Ein gesundheitsökonomisches entscheidungsanalytisches Modell stellt Nutzen und Schaden einer Intervention in Beziehung zu deren Gesamtkosten dar. Die Gesamtkosten beinhalten nicht nur die Kosten der Intervention, sondern auch die Kosten der weiteren Behandlung und Versorgung des Patienten, der Nebenwirkungen und ihrer Behandlung sowie die Kostenersparnis durch Verminderung von Krankheitsfolgen (Drummond, O'Brien et al. 1997).

In Abhängigkeit vom verwendeten Nutzenmaß unterscheidet man verschiedene Typen der gesundheitsökonomischen Evaluation (Drummond, O'Brien et al. 1997) (Leidl 1998; Siebert 2003). Diese Typen wurden bereits im Kapitel 8.1 dargestellt.

In Folgenden werden die wichtigsten Konzepte und Methoden entscheidungsanalytischer Modelle in der gesundheitsökonomischen Evaluation beschrieben. Viele der hier angeführten Konzepte gelten gleichermaßen für die rein medizinische Entscheidungsanalyse, werden aber zur Vermeidung von Redundanz nur hier genannt.

8.3.1 Rahmenbedingungen des Modells

Ausgangspunkt der Modellentwicklung ist die präzise Formulierung der Fragestellung und der Rahmenbedingungen, die das Modell erfüllen muss.

In der Formulierung der Fragestellung sollten die Zielsetzung der Studie, die Auftraggeber und Adressaten und das Spektrum der betrachteten Krankheiten genau definiert werden (siehe auch 8.4.2). Diese Formulierung sollte nicht durch die Verfügbarkeit von Daten beschränkt werden (Roberts, Chambers et al. 2011).

Als Rahmenbedingungen des Modells sind zu definieren:

- » Zielpopulation,
- » Intervention und Vergleichstechnologien,
- » Zielgrößen,
- » Perspektive,
- » Zeithorizont,
- » Modelltyp und statistischer Analyseansatz.

Einzelheiten dazu beschreiben die ökonomischen Empfehlungen (siehe 8.4.2).

8.3.2 Vorgehen bei der Modellentwicklung

Die folgenden Abschnitte beschreiben die einzelnen Schritte, die üblicherweise bei der Entwicklung eines Modells durchlaufen werden:

- » Entwurf des Modells,
- » Auswahl der Modellierungstechnik,
- » Auswahl der Daten,
- » Definition der Abhängigkeiten von Modellparametern,
- » Programmierung,
- » Kalibrierung (falls erforderlich),

- » Validierung,
- » Dokumentation der Modellannahmen und Limitationen.

8.3.2.1 Entwurf des Modells

Nach Festlegung der Rahmenbedingungen wird ein Modell entworfen, das den Ereignisablauf und die Gesundheitszustände der untersuchten Erkrankung bzw. Interventionen beschreibt. Dabei sollte das Modell einerseits eine vereinfachte Darstellung der Wirklichkeit anstreben und die wesentlichen Aspekte der Krankheit, Technologie und Zielpopulation berücksichtigen. Andererseits sollte das Modell so differenziert sein, dass alle Zustände und Ereignisse beschrieben werden, in denen die Handlungsalternativen wichtige Unterschiede aufweisen. Diese Unterschiede können z. B. in den Mortalitäts- oder Progressionsraten liegen, in der Lebensqualität oder in den Kosten.

Die Einfachheit des Modells ist wünschenswert im Hinblick auf die Transparenz des Modells (Axelrod 1997; Weinstein, O'Brien et al. 2003). Das Modell muss jedoch komplex genug sein, um die Fragestellung zu beantworten und um die Augenscheinvalidität („face validity“) für Experten sicherzustellen (Roberts, Chambers et al. 2011).

Die ISPOR-SMDM Modeling Good Research Practice Task Force (Roberts, Chambers et al. 2011) empfiehlt einen festgelegten Prozess der Umsetzung des Konzepts zur Untersuchung der Fragestellung in ein Modellkonzept. In diesen Prozess sollten medizinische Experten eingebunden sein. Das Modell sollte mittels einer geeigneten strukturierten grafischen Entwurfsmethode wiedergegeben werden, welche die Diskussion und das Verständnis erleichtert (z. B. Entscheidungsbäume, Einflussdiagramme, concept maps).

8.3.2.2 Auswahl der Modellierungstechnik

Für die entscheidungsanalytische Modellierung stehen unterschiedliche Modellierungstechniken zur Verfügung, so z. B. Entscheidungsbäume, Markov-Modelle, diskrete Ereignissimulationsmodelle und dynamische Übertragungsmodelle. Die Wahl der Modellierungstechnik sollte nicht von vornherein eingeschränkt werden. Auch gibt es keine Modellierungstechnik, die generell als Methode der Wahl herangezogen werden kann (Drummond, Schwartz et al. 2008; IQWiG 2009; Roberts, Chambers et al. 2011). Vielmehr ist die Modellierungstechnik festzulegen in Abhängigkeit von der Fragestellung, von den Rahmenbedingungen des Modells und von der Struktur des Entscheidungsproblems. Die ausgewählte Modellierungstechnik muss ausreichend flexibel sein (Drummond, Schwartz et al. 2008; IQWiG 2009). Die Verfügbarkeit von Daten und Software sollte hier nicht zu Einschränkungen führen, die das Ergebnis verzerren. Die

Vorteile und Anwendungsbereiche der einzelnen Modellierungsmethoden werden in Abschnitt 8.3.3 im Einzelnen beschrieben. Weitere Übersichten und Hilfen geben die folgenden Publikationen: (Barton, Bryan et al. 2004; Brennan, Chick et al. 2006; Cooper, Brailsford et al. 2007; Kim and Goldie 2008; Stahl 2008; Roberts, Chambers et al. 2011).

8.3.2.3 Auswahl der Daten

Die Auswahl der Daten für die Inputparameter des Modells sollte gut begründet sein. Daten für die Wirksamkeit einer Behandlung werden dabei vorwiegend aus klinischen Studien stammen, Daten zur Güte eines diagnostischen Tests aus Diagnosestudien, prognostische Parameter zu Morbidität und Mortalität aus Prognosestudien und Inzidenzen von zukünftigen Ereignissen aus Risikofaktorstudien. Die Beurteilung von medizinischen Studien und die Extraktion und Synthese von Effektdaten wird in den Kapiteln 6 und 7 dieses Handbuchs ausführlich beschrieben. Darüber hinaus wird in der Regel eine Reihe weiterer Datenquellen notwendig sein, um das Modell vollständig zu beschreiben. Ereignishäufigkeiten zur Beschreibung des Krankheitsverlaufs über den Zeitraum der klinischen Studien hinaus stammen oft aus epidemiologischen Beobachtungsstudien oder Krankheitsregistern. Die Vergleichbarkeit der Studienpopulation mit der Zielpopulation des Modells muss sorgfältig geprüft werden. Studien zur Therapietreue sollten einbezogen werden, um Routinebedingungen in die Bewertung der Intervention so weit wie möglich einzubeziehen. Daten zum Ressourcenverbrauch können aus Abrechnungsdaten von Krankenhäusern und Krankenkassen stammen. Hier ist insbesondere die Eignung für den regionalen Kontext zu überprüfen. Nationale und regionale Gesundheitsstatistiken geben demografische Parameter wie Alters- und Geschlechtsverteilung, Mortalitätsdaten und in vielen Fällen die Inzidenz von Krankheiten. Stehen für gewisse Modellparameter keinerlei empirische Daten zur Verfügung, dann kann auch die Meinung von Experten herangezogen werden. In diesem Fall ist insbesondere für diese Parameter eine umfangreiche Sensitivitätsanalyse notwendig.

8.3.2.4 Definition der Abhängigkeiten von Modellparametern

Bei den Modellparametern handelt es sich häufig nicht um fixe Werte, sondern um Funktionen von Patientencharakteristika oder anderen Parametern. Ein Beispiel sind Risikofunktionen, in die eine Reihe von Risikofaktoren eingehen. Die in einem Modell verwendeten mathematischen Beziehungen und ihre Herkunft sollten in der Modellbeschreibung dokumentiert werden. Werden vom Modellentwickler Regressionsgleichungen oder andere mathematische Beziehungen nicht aus Veröffentlichungen

entnommen, sondern selbst bestimmt, sollte die Auswahl der verwendeten statistischen Verfahren begründet und ebenso wie die benutzten Datenquellen dokumentiert werden.

8.3.2.5 Programmierung des Modells

Nach Festlegung der Modellierungstechnik wird die Modellstruktur weiter spezifiziert und das Modell unter Zuhilfenahme entsprechender Software programmiert. Für jeden Modelltyp ist speziell angepasste Software erhältlich. Die Details der Programmierung werden daher vom Modelltyp und der verwendeten Software bestimmt.

Während und nach Abschluss der Programmierung erfolgt die Überprüfung des Programmcodes. Dabei wird die korrekte Umsetzung der logischen Verknüpfungen und Algorithmen getestet. Mathematische Gleichungen, die in das Modell eingehen, sollten schon vorab, z. B. in Excel, getestet werden, bevor sie ins Modell aufgenommen werden. Kann die Programmentwicklung modular erfolgen, ist es sinnvoll, die Module unabhängig zu testen. Der Test des gesamten Programmcodes erfolgt vor allem durch Variation der Inputparameter und Vergleich der vom Modell prädizierten Ergebnisse mit den Erwartungen. Aufschlussreich sind hier insbesondere Extremwerte und Sonderfälle, für die die Ergebnisse bekannt sind. Auch die statische Überprüfung des Codes und der eingehenden Parameterwerte kann zum Programmtest eingesetzt werden. Dabei liest ein unabhängiger Modellierer den Programmcode und versucht ihn nachzuvollziehen. Bei komplexen Modellen wird gelegentlich eine unabhängige Zweitprogrammierung, oft auf einer anderen Softwarebasis, durchgeführt. Dieses Vorgehen führt zu größerer Sicherheit, ist aber wegen des enormen Arbeitsaufwandes in manchen Situationen nicht durchführbar.

8.3.2.6 Kalibrierung

In der Regel werden für jeden Parameter in gesundheitsökonomischen Studien getrennt Daten identifiziert. Aus der gefundenen Quelle wird ein Erwartungswert für diesen Parameter entnommen und möglichst auch der Standardfehler dieses Werts. In komplexen Modellen gibt es aber häufig Parameter, für die keine Daten vorliegen. Das könnte zum Beispiel die Inzidenz eines Krankheitsfrühstadiums sein, das noch keine Symptome verursacht. Die Inzidenz des nicht diagnostizierten Frühstadiums geht im Modell aber ein in die Berechnung der insgesamt anfallenden Krankheitsfälle in einem bestimmten Zeitraum. Liegen für diese Endgröße empirische Daten vor, dann kann der unbekannte Parameter so angepasst werden, dass das Modellergebnis für die Zielgröße den empirischen Daten möglichst nahe kommt. Diese Parameter-Anpassung wird auch Kalibrierung genannt. Es handelt sich hierbei um einen Optimierungsansatz: Ein Satz

von Parametern wird variiert mit dem Ziel, die Abweichung der Modellvorhersage für zeitlich nachgelagerte Zielgrößen von empirischen Zielwerten zu minimieren.

Für die Kalibrierung gesundheitsökonomischer Modelle werden zunehmend Methoden aus anderen Feldern wie der Produktionsplanung (engl. „Operations Research“) erfolgreich eingesetzt (Kim, Kuntz et al. 2007; Kong, McMahon et al. 2009; Rutter, Miglioretti et al. 2009; Karnon and Vanni 2011; Vanni, Karnon et al. 2011). Die Entwicklung ist hier noch nicht abgeschlossen. Karnon und Vanni 2011 geben eine vergleichende Übersicht über verwendete Verfahren. (Stout, Knudsen et al. 2009) empfehlen eine Checkliste zur Dokumentation der Elemente einer Kalibrierung wie Zielwerte, Suchalgorithmus, Maß für die Anpassungsgüte, Akzeptanz- und Abbruchkriterium. Derzeit gibt es noch keine internationale Empfehlungen oder Standards zum Vorgehen bei der Kalibrierung.

8.3.2.7 Validierung

Unter der Validierung eines Modells versteht man eine Konsistenzprüfung des Modells auf einer Reihe von Ebenen:

- » Augenscheinvalidität („face validity“, Plausibilität) des Modellkonzepts,
- » Verifikation des Programmcodes und interne Validierung,
- » externe und prädiktive Validierung,
- » Kreuzvalidierung („model cross validation“).

Bereits beim Modellentwurf sollte eine Überprüfung der Augenscheinvalidität des Modells durch klinische Experten erfolgen. Dabei wird beurteilt, ob Struktur und Datenbasis des Modells konsistent sind mit dem aktuellen Stand der Erkenntnis zum natürlichen Krankheitsverlauf, der Versorgung und zu Konsequenzen von Behandlungen (Kleijnen 1995; Weinstein, Toy et al. 2001; Banks, Carson et al. 2005). Bereits in dieser Phase ist auch ein Vergleich mit anderen Modellen zu empfehlen.

Die Verifikation des Programmcodes wurde bereits in Abschnitt 8.3.2.5 beschrieben. Augenscheinvalidität und technischer Test sind Voraussetzung für die eigentliche Validierung des Modells. Von interner Validierung spricht man, wenn das Modell in der Lage ist, Daten (z. B. die Anzahl der Todesfälle im Zeitverlauf) aus Studien zu reproduzieren, aus denen auch die Modellparameter entnommen wurden (z. B. relative Risiken). Verifikation und interne Validierung werden häufig zusammengefasst (Wong, Eddy et al. 2011). Interne Validierung wird generell erwartet.

Wo externe Daten zur Verfügung stehen, sollte eine externe Validierung durchgeführt werden. Von externer Validierung spricht man, wenn sich das Modell in der Vorhersage von Daten bewährt, die aus Studien stammen, die nicht für die Auswahl der Modell-

parameter benutzt wurden (Eddy and Schlessinger 2003). Dazu können Langzeitstudien, Registerdaten oder nationale Statistikdatenbanken verwendet werden. Da in ein Modell häufig der größte Teil der verfügbaren Evidenz eingeht, ist eine externe Validierung oft erst möglich, wenn neue Studien zu einer Erkrankung veröffentlicht werden. Dann spricht man auch von prädiktiver Validierung. Weicht die Modellvorhersage von den Ergebnissen neuer Studien ab, kann dies bedeuten, dass neue Erkenntnisse vorliegen, die die Erweiterung oder Überarbeitung des Modells notwendig machen (Weinstein, O'Brien et al. 2003). Einige Autoren (Sculpher, Fenwick et al. 2000) fordern für entscheidungsanalytische Modelle eine andere Form der externen Validierung als den Vergleich der Modellvorhersage mit unabhängigen epidemiologischen Daten, da diese Modelle nicht für die Prädiktion entwickelt wurden. Sie schlagen vor, diese Modelle an der Nützlichkeit für die Entscheidungsfindung zu messen. Die ISPOR-SMDM Modeling Good Research Practice Task Force empfiehlt jedoch ausdrücklich die oben beschriebene externe Validierung (Wong, Eddy et al. 2011).

Aufschlussreich ist auch die Kreuzvalidierung. Dabei wird ein Modell in Struktur und Ergebnissen mit anderen Modellen verglichen. Zunehmend werden komplexe Modelle verschiedener Arbeitsgruppen systematisch miteinander verglichen, und die Gründe für abweichende Ergebnisse werden aufgeklärt (CISNET im Bereich Cancer-Modelle (Berry, Cronin et al. 2005), Mount Hood 4 Modeling Group im Bereich Diabetes (Group 2007)). Koopman schlägt für den Bereich der Infektionsmodelle ein Verfahren vor, das die Robustheit der Entscheidungsempfehlung unter Lockerung wichtiger Modellannahmen anhand des Vergleichs mit komplexeren Modellen untersucht (Koopman 2005). Ein Vergleich der Modellergebnisse mit denen anderer Modelle und eine Diskussion der Abweichungen und ihrer Ursachen sollten im Modellreport enthalten sein.

Detaillierte Empfehlungen zu allen Stufen der Validierung werden in Kürze von der ISPOR-SMDM Modeling Good Research Practice Task Force herausgegeben (Wong, Eddy et al. 2011).

8.3.2.8 Dokumentation des Modells

Die Dokumentation eines Modells im Rahmen eines HTA-Auftrags unterliegt speziellen Bedingungen und Vorgaben, die in den gesundheitsökonomischen Empfehlungen festgehalten sind (siehe Kapitel 8.4.2).

Abgesehen davon sollten die Empfehlungen der ISPOR-SMDM Modeling Good Research Practice Task Force (Wong, Eddy et al. 2011) zur Darstellung des Modells befolgt werden. Diese sehen vor, jedes Modell in einem öffentlich zugänglichen nicht-technischen Dokument zu beschreiben und zusätzlich in einem technischen Dokument, dessen Veröffentlichung eingeschränkt sein kann, um das geistige Eigentum des

Modellentwicklers zu schützen. Die nicht-technische Dokumentation sollte nach diesen Empfehlungen in nicht-technischer Formulierung die folgenden Punkte beinhalten:

- » Typ des Modells und beabsichtigte Verwendung,
- » Finanzierung des Modells,
- » Struktur des Modells und beabsichtigter Gebrauch,
- » Parameter, Zielgrößen, andere Komponenten, die die Funktion des Modells bestimmen, und ihre Abhängigkeiten,
- » Datenquellen,
- » Validierungsmethoden und Ergebnisse,
- » Limitationen.

Da ein Modell immer ein vereinfachtes Abbild der Wirklichkeit darstellt, muss die Interpretation der Modellergebnisse immer vor dem Hintergrund der Modellannahmen und Limitationen erfolgen. Für die Dokumentation eines Modells ist daher die umfassende Darstellung dieser Annahmen von herausragender Wichtigkeit. Numerische Annahmen betreffen z. B. die Abhängigkeiten zwischen Modellparametern oder die Extrapolation von Effekten der Interventionen. Strukturelle Annahmen betreffen den Aufbau des Modells, beispielsweise die Reihenfolge von diagnostischen Tests oder die Unabhängigkeit dieser Tests.

Die technische Dokumentation sollte so detailliert sein, dass ein Leser mit ausreichender Expertise das Modell bewerten und potenziell reproduzieren kann. Ziel der Dokumentation eines Modells sollte immer größtmögliche Transparenz sein.

8.3.3 Modelltypen

8.3.3.1 Entscheidungsbaum

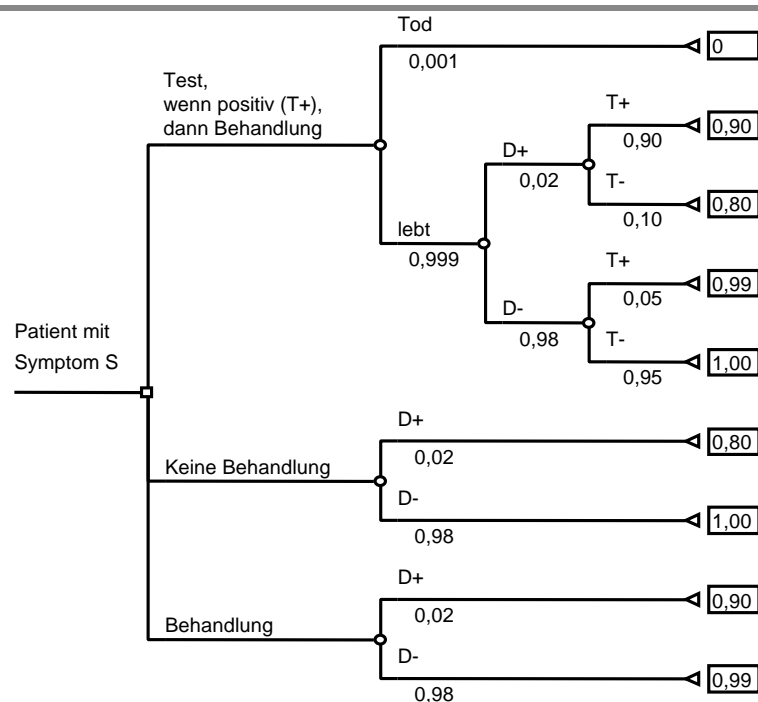
Klein et al. beschreiben einen Entscheidungsbaum wie folgt (Klein 2010):

„Ein Entscheidungsbaum ist eine visuelle Darstellung von möglichen Handlungsoptionen und Ereignisabläufen sowie deren medizinischen und ökonomischen Konsequenzen (Hunink, Glasziou et al. 2001). Auf jede mögliche medizinische Handlungsalternative folgen Pfade, die alle möglichen Ereignisabläufe mit ihren jeweiligen Eintrittswahrscheinlichkeiten abbilden. Diese Wahrscheinlichkeiten können dabei sowohl von der gewählten Handlungsalternative, als auch von Patientenmerkmalen wie soziodemographischen Kenngrößen, Risikoprofilen oder dem Krankheitsstadium abhängen. Am Ende des Entscheidungsbaumes führt jeder Pfad zu einem Outcome wie beispielsweise komplikationsfreies Überleben, Überleben nach Komplikation, oder Tod. Für jede einzelne der untersuchten Handlungsalternativen kann der Erwartungswert des medi-

zinischen Nutzens als gewichteter Mittelwert aller möglichen Outcomes berechnet werden. Dabei stellen die Pfadwahrscheinlichkeiten der einzelnen Pfade im Entscheidungsbaum die Gewichte dar. Das Entscheidungsbaumverfahren eignet sich gut für die Untersuchung von Situationen, bei denen keine wiederholten Ereignisse vorkommen und der Zeithorizont der Analyse begrenzt ist und exakt festgelegt werden kann.“

Ein Beispiel für einen Entscheidungsbaum, der die Strategie eines diagnostischen Tests mit Behandlung bei positivem Test bewertet (Siebert, Mühlberger et al. 2000), zeigt Abbildung 8.3..

Abbildung 8.3:
Entscheidungsbaum zur Anwendung eines diagnostischen Tests vor Behandlung



Eingetragen sind Ereigniswahrscheinlichkeiten und Überlebensraten nach einem Jahr.
D+/- kennzeichnet das Vorliegen der Krankheit, T+/- das Testergebnis.

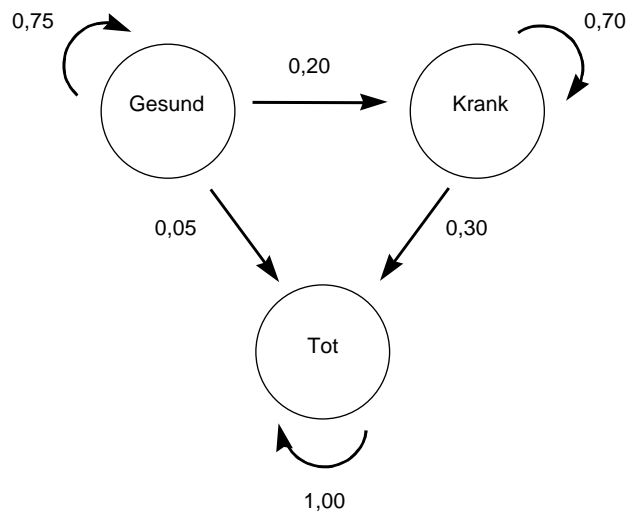
Quelle und Darstellung: (Siebert, Mühlberger et al. 2000)

8.3.3.2 Markov-Modelle und Zustands- übergangsmodelle

Markov-Modelle bzw. Zustandsübergangsmodelle sind mathematische Modelle der Entscheidungsanalyse, die aus Gesundheitszuständen und möglichen Übergängen bestehen (Sonnenberg and Beck 1993).

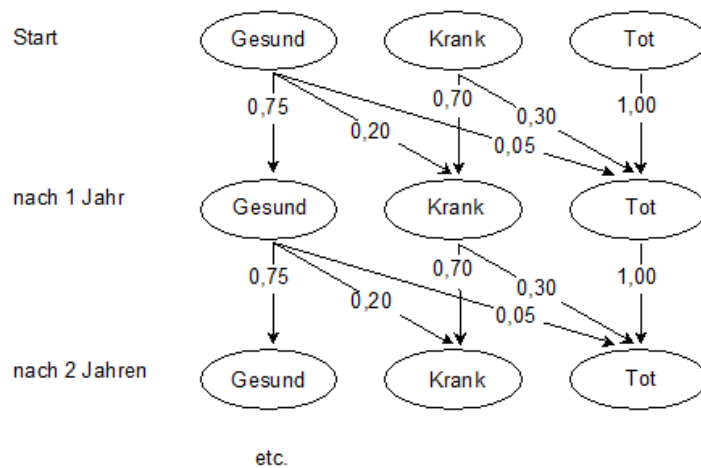
Analysiert werden Markov-Modelle nach Kein und Kollegen folgendermaßen (Klein et al. 2010, 243 f.): „In einer Computersimulation durchlaufen hypothetische Patienten die Gesundheitszustände in bestimmten gleich großen Intervallen (Markov-Zyklen). Nach jedem Zyklus können die Mitglieder der Kohorte mit bestimmten Übergangswahrscheinlichkeiten (transition probabilities) entweder in andere Gesundheitszustände übertreten, im aktuellen Gesundheitszustand verweilen oder an der untersuchten Erkrankung bzw. aus anderer Ursache sterben [Abbildung 8.4]. Über die Gesamtzeit der Analyse werden so Lebenszeit, Lebensqualität und Kosten entsprechend der verbrachten Zeit in den spezifischen Gesundheitszuständen kumuliert [Abbildung 8.5]. Diese Ergebnisse können dann zwischen den verschiedenen Strategien verglichen werden. In einem Markov-Modell hängen die Übergangswahrscheinlichkeiten in einem Zyklus ausschließlich vom momentanen Gesundheitszustand und nicht von der Geschichte der bereits durchlaufenen Zustände ab (Markov-Annahme) (Beck and Pauker 1983; Sonnenberg and Beck 1993; Siebert, Mühlberger et al. 2000). Diese Markov-Annahme nennt man auch die ‚Gedächtnislosigkeit‘ der Markov-Modelle. Markov-Modelle sind geeignet, (a) wenn das Entscheidungsproblem veränderliche Risiken im Zeitverlauf beinhaltet, (b) wenn der Zeitpunkt des Ereignisauftritts eine Rolle spielt (Time-to-event) oder (c) wenn relevante Ereignisse mehr als einmal auftreten können.“ (Klein 2010)

Abbildung 8.4:
Zustandsübergangsdiagramm für das Markov-Modell
einer chronischen Krankheit mit Übergangswahrscheinlichkeiten



Quelle und Darstellung: (Siebert, Mühlberger et al. 2000)

Abbildung 8.5:
Darstellung des Ablaufs eines Markov-Modells für eine chronische Erkrankung



Dargestellt sind Gesundheitszustände, Übergangswege und Übergangswahrscheinlichkeiten für zwei Zyklen.

Quelle und Darstellung: (Siebert, Mühlberger et al. 2000)

Die statistische Analyse eines Modells kann entweder als Kohortensimulation oder als Mikrosimulation durchgeführt werden. Bei der Kohortensimulation durchläuft eine Gruppe gleichartiger Patienten das Modell, und als Ergebnis des Modells erhält man einen durchschnittlichen Erwartungswert ohne Schätzung einer Streubreite. Die Kohortensimulation ist eine deterministische Analyse. Wie im Folgenden an einem Beispiel erläutert wird, kann die Gedächtnislosigkeit des Modells bei einer Kohortensimulation zur Zustandsexplosion führen. In einem Modell für Erkrankungen der Herzkranzgefäße könnte es z. B. einen Anteil der Patientenkohorte geben, der Diabetes hat. Dieser Anteil der Kohorte hat ein erhöhtes Risiko für einen Infarkt. Auch nach dem Infarkt hat dieser Teil der Kohorte eine höhere Progressionswahrscheinlichkeit. Ein Kohortenmodell muss hier zur Beschreibung zwei Gesundheitszustände modellieren: einen Zustand nach dem Infarkt für Patienten mit Diabetes und einen Zustand nach dem Infarkt für Patienten ohne Diabetes. Gäbe es nur einen Zustand „nach Infarkt“, wäre die Information über das Vorliegen von Diabetes in diesem Zustand nicht mehr verfügbar, und die weitere Progression könnte nicht in Abhängigkeit von dieser Eigenschaft modelliert werden. Ein weiterer Risikofaktor dieser Art würde die Zahl der Zustände auf vier erhöhen. Bestimmen eine Reihe von Eigenschaften den Verlauf der Krankheit, dann kann die Zahl der Zustände so groß werden, dass das Modell unhandlich wird. In diesen Fällen empfiehlt sich die Analyse als Mikrosimulation (ISPOR-SMDM Modeling Good Research Practices Task Force 2011). Bei der Mikrosimulation durch-

laufen viele individuell verschiedene Patienten nacheinander das Modell, wobei je nach Eigenschaften (Alter, Geschlecht, Risikofaktoren) des Patienten die Ergebnisse verschieden wahrscheinlich sind. Aus den Ergebnissen wird dann ein Mittelwert gebildet (Weinstein 2006; Zaidi, Harding et al. 2009; Zuccelli, Jones et al. 2010; Rutter, Zaslavsky et al. 2011). Die individuellen Patienteneigenschaften und die Geschichte des Patienten stehen während der Simulation zur Verfügung und bestimmen deren Verlauf. Patienteneigenschaften können sich während der Simulation ändern. Der Einfluss der Patientengeschichte auf den Weg des Patienten durch das Modell bedeutet die Aufgabe der Markov-Annahme (siehe oben). Daher spricht man bei Mikrosimulation oft nicht mehr von Markov-Modellen, sondern allgemeiner von Zustandsübergangsmodellen. Mikrosimulationen werden oft auch Monte-Carlo-Simulationen genannt, da die Patientencharakteristika mithilfe eines Zufallszahlengenerators aus vorgegebenen statistischen Verteilungen gezogen werden.

Die Vorzüge der Kohortensimulation sind Übersichtlichkeit, Effizienz, einfacheres Auffinden von Programmierungsfehlern und die einfachere Durchführung von „Value of Information“-Analysen (Siebert, Alagoz et al. 2011). (Zur Value-of-Information-Analyse s. Abschnitt 8.3.4.2.) Bei der Entscheidung zwischen Kohorten- und Mikrosimulation ist jedoch zu bedenken, dass die Beschreibung des Krankheitsverlaufs im Modell nicht vereinfacht werden sollte, um eine technisch einfachere Analyseverfahren verwenden zu können. Empfehlungen zu Entwurf, Parametern und Dokumentation eines Markov-Modells werden zurzeit von der ISPOR-SMDM Modeling Good Research Practices Task Force erarbeitet (Siebert, Alagoz et al. 2011).

Im Folgenden werden weitere weniger häufig eingesetzte, aber immer wichtiger werdende Techniken der Modellierung beschrieben. Diese sind zum Teil in den Texten zu Modelltypen im Lehrbuch von Schöffski et al. 2007 dargestellt, auf die der Leser / die Leserin an dieser Stelle verwiesen wird (Siebert, Mühlberger et al. 2000). Auf diesem Lehrtext basieren die folgenden Kapitel „Diskrete ereignisorientierte Simulation“, „Agentenbasierte Modelle“ und „Systems-Dynamics-Modelle“.

8.3.3.3 Diskrete ereignisorientierte Simulation

Im Folgenden wird der Modelltyp mit diskreter ereignisorientierter Simulation beschrieben. Die Beschreibung folgt dem Lehrtext zu Modelltypen aus dem Lehrbuch von Schöffski et al 2007.

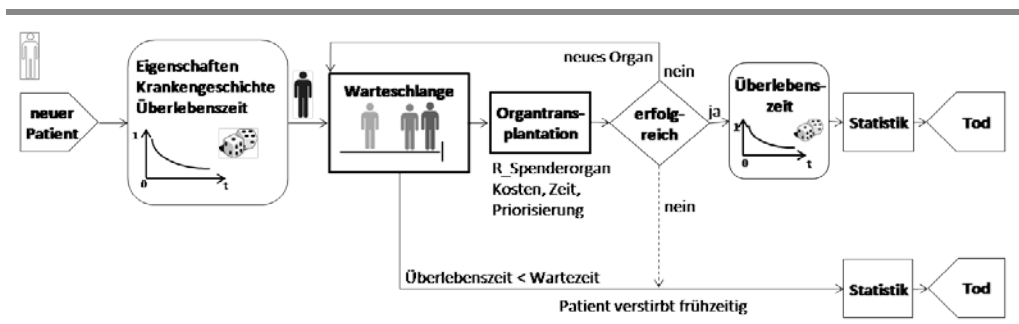
Die ereignisorientierte Simulation wurde ursprünglich für die Produktionsoptimierung entwickelt. Sie ermöglicht die Modellierung von begrenzten Ressourcen und kann die Dynamik von Warteschlangen (engl.: queue) abbilden. Im Gesundheitswesen kann es sich bei den begrenzten Ressourcen z. B. um Organe für die Transplantation handeln,

um Operationssäle, Ärztinnen/Ärzte und anderes Personal oder um diagnostische Verfahren und Interventionen (Caro 2005; Stahl 2008). Das Optimieren von organisatorischen Abläufen und das Vermeiden von Verzögerungen kann, z. B. beim Schlaganfall, einen wesentlichen Beitrag zu einer erfolgreichen Behandlung leisten (Stahl, Furie et al. 2003).

Tabelle 8.19 zeigt als Beispiel das Flussdiagramm eines Organtransplantationsmodells nach (Ratcliffe, Young et al. 2001; Shechter, Bryce et al. 2005). Mit diesem Modell können Priorisierungsregeln für die Vergabe von Spenderorganen bewertet werden. In diesem Beispiel stellt das Spenderorgan eine begrenzte Ressource dar. In diesem Modell erfahren einzelne Patienten die modellierten Ereignisse. Es handelt sich also um eine Mikrosimulation. Ereignisse sind z. B. die Ankunft einer Patientin / eines Patienten in einer Warteschlange oder Anfang und Ende einer Behandlung. Dabei verändert sich der Zustand des Patienten. Der Zustand des Patienten wird durch Attribute charakterisiert, die sich bei Ereignissen verändern können. Beispiele für solche Attribute sind Alter, Geschlecht, Vorliegen von Diabetes oder Zeit seit dem letzten Herzinfarkt. Die Geschichte des Patienten ist in diesen Attributen während der Simulation verfügbar und beeinflusst den Weg des Patienten durch das Modell (das Modell hat ein Gedächtnis). Die Patienten beanspruchen Ressourcen wie ärztliches Personal, Medikamente, Katheter, Krankenhausbetten. Steht die begrenzte Ressource nicht zur Verfügung, kommt es zu Verzögerungen, oder es werden Alternativbehandlungen erforderlich. Die Modellierung dynamischer Warteschlangen erlaubt das Beschreiben dieser Vorgänge.

Abbildung 8.6:

Diskrete Ereignissimulation, Flussdiagramm für ein Organtransplantationsmodell

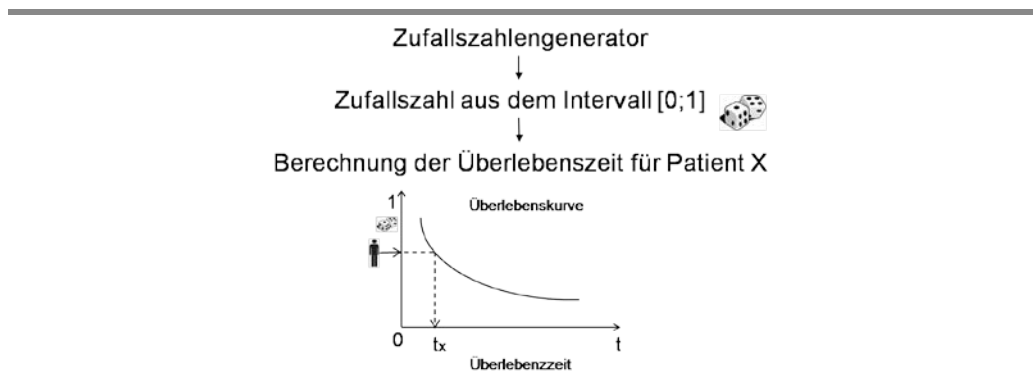


Quelle und Darstellung: (Siebert, Mühlberger et al. 2000)

Die Zeit verläuft in der ereignisorientierten Simulation nicht in gleichbleibenden Zeitintervallen wie in einem Markov-Modell, sondern schreitet von einem Ereignis (engl.: event) zum nächsten fort. Ereignisse können dabei dynamisch neue Ereignisse hervorrufen, sodass komplexe Ereignisabläufe nachgebildet werden können. Die Zeit bis zum nächsten Ereignis kann probabilistisch gewählt sein. Während bei einer Mikrosimulation in einem Zustandsübergangsmodell die Übergangswahrscheinlichkeiten innerhalb eines Zeitintervalls aus einer Verteilung gezogen werden, werden bei der

ereignisorientierten Simulation die Zeiten bis zu einem Ereignis gezogen. Dazu werden zunächst gleich verteilte Zufallszahlen aus einem Intervall von 0 bis 1 gezogen. Diese Zufallszahlen werden dann z. B. in Überlebenszeiten der einzelnen Patienten übersetzt. Diese Übersetzung erfolgt auf solche Weise, dass sich für die Überlebenszeiten vieler Patienten eine bekannte Survival-Kurve ergibt (s. Abbildung 8.7).

Abbildung 8.7:
Diskrete Ereignissimulation, Ermittlung der Überlebenszeit für Patienten in der Simulation



Quelle und Darstellung: (Siebert, Mühlberger et al. 2000)

Während des Ereignisablaufs werden außer der in bestimmten Zuständen verbrachten Zeit und den angelaufenen Kosten auch Größen akkumuliert, die die Effizienz des Systems beschreiben, wie Wartezeiten oder Systemauslastung.

Der Einsatz diskreter Ereignissimulationen ist geeignet, wenn knappe Ressourcen und die Dynamik von Warteschlangen nachgebildet werden sollen. In anderen Fällen kann sie als Alternative zur Mikrosimulation in einem Zustandsübergangsmodell angesehen werden. Da sie nicht zyklusgebunden ist, bietet sie im Vergleich zu Zustandsübergangsmodellen größere Flexibilität in der Behandlung der Zeit (Stahl, Karnon et al. 2011). Die ISPOR-SMDM Modeling Good Research Practices Task Force (Stahl, Karnon et al. 2011) gibt Empfehlungen zu Details der Modellierung als diskrete Ereignissimulation.

8.3.3.4 Dynamische Übertragungsmodelle

Bei der gesundheitsökonomischen Evaluation von Impfprogrammen spielt die Herdenimmunität eine große Rolle (Anderson and May 1991; Fine 1993; Brisson and Edmunds 2003; Drummond, Chevat et al. 2007; Welte, Leidl et al. 2010). Klein und Koautoren schreiben hierzu (Klein et al. 2010): „Darunter versteht man die indirekten Auswirkun-

gen einer Impfung auf nicht geimpfte bzw. nicht immune Individuen infolge der mit der Impfung einhergehenden Abnahme des Infektionsdrucks in der Population, die durch den Rückgang der Zahl infektiöser Individuen verursacht wird. Besonders zu beachten ist, dass Herdenimmunitätseffekte nicht ausschließlich gesundheitlichen Nutzen, sondern auch gesundheitlichen Schaden bewirken können (Edmunds, Medley et al. 1999; Brisson and Edmunds 2003; Beutels, Scuffham et al. 2008; Welte, Leidl et al. 2010). Schaden entsteht beispielsweise durch die Erhöhung des durchschnittlichen Infektionsalters bei Infektionskrankheiten, deren Schweregrad mit dem Alter zunehmen kann (z. B. Mumps, Röteln, Polio, HAV, Windpocken).“ (Klein 2010)

„Die Berücksichtigung von Herdenimmunitätseffekten im Rahmen ökonomischer Evaluationen erfordert den Einsatz dynamischer Übertragungsmodelle (Edmunds, Medley et al. 1999; Beutels, Van Doorslaer et al. 2003; Brisson and Edmunds 2003; WHO 2008; Walker, Hutubessy et al. 2010; Welte, Leidl et al. 2010). Übertragungsmodelle bezeichnet man als dynamisch, wenn die Infektionsrate (force of infection) zeitlich veränderlich als Funktion des Anteils infektiöser Individuen in der Population modelliert wird. Neben der Verteilung der Infektion in der Population gehen Interaktionen zwischen den Individuen (Kontaktverhalten) und die Übertragbarkeit der Infektion in die Funktion der Infektionsrate ein. Im Unterschied zu dynamischen Übertragungsmodellen nehmen statische Übertragungsmodelle eine vom Anteil infektiöser Individuen in der Population unabhängige Infektionsrate an. Diese kann über die Zeit konstant gehalten oder in Abhängigkeit von individuellen Eigenschaften deterministisch verändert werden.“ (Klein 2010) Statische Übertragungsmodelle lassen keine Modellierung von Herdenimmunitätseffekten zu.

Die Verwendung eines dynamischen Modells ist auch dann notwendig, wenn eine Intervention die Proportionen verschiedener Typen von Erregern (z. B. Genotypen oder Serotypen) verändert (ISPOR-SMDM Modeling Good Research Practices Task Force 2011).

Dynamische Modelle sind häufig Differentialgleichungsmodelle, die nicht individuelle Patienten modellieren, sondern z. B. die Gruppen der Empfänglichen (Suszeptiblen), Infizierten und Resistenten (SIR-Modell) als Einheit (Kompartiment).

Die mithilfe von dynamischen Transmissionsmodellen ermittelte Infektionsrate kann in ein Markov-Modell eingesetzt werden, um Langzeit-Effektivität und Kosten-Effektivität einer Intervention abzuschätzen (Kim and Goldie 2008). Darüber hinaus gibt es komplexere Modelltypen, die gleichzeitig sowohl die Transmission als auch Gesundheitseffekte und Kosten einbeziehen.

Diskrete ereignisorientierte Modelle und agentenbasierte Modelle (Abschnitt 8.3.3.5) werden zur Modellierung der Transmissionsdynamik eingesetzt, wenn die Interaktion von Individuen dabei von Bedeutung ist.

Detailliertere Empfehlungen zur Entwicklung, Analyse und Dokumentation von dynamischen Transmissionsmodellen werden zurzeit von der ISPOR-SMDM Modeling Good Research Practices Task Force (Pitman, Edmunds et al. 2011) zusammengestellt.

8.3.3.5 Agentenbasierte Modelle

Die agentenbasierte Modellierung ist eine Simulation, bei der Entitäten autonom auf ihre Umwelt reagieren. Sie werden daher Agenten genannt. Während die diskrete Ereignissimulation durch einen Ablauf von Ereignissen angetrieben wird, beruht die agentenbasierte Simulation auf Verhaltensregeln für die individuellen Agenten unter gegebenen Bedingungen in ihrer Umwelt. Agenten besitzen die Fähigkeit zu lernen. Relativ einfache Verhaltensregeln für den individuellen Agenten können zu komplexem und unerwartetem Verhalten des Gesamtsystems führen (engl: emergence) (Axelrod 1997; Bonabeau 2002; Macal and North 2005; Billari, Fent et al. 2006; Macal and North 2006; North and Macal 2007). Agentenbasierte Modelle werden im biomedizinischen Bereich eingesetzt, zur Simulation von Epidemien oder für soziale Netzwerke (Bonabeau 2002; Koopman 2005; Macal and North 2006; Lee, Brown et al. 2010; Auchincloss, Riolo et al. 2011; Brown, Price et al. 2011). Es handelt sich dabei um probabilistische Mikrosimulationen.

Die Anwendung der agentenbasierten Simulation in der Gesundheitsökonomie ist derzeit noch in Entwicklung. Beispiele sind (Rein, Saaddine et al. 2007; Sander, Nizam et al. 2009). Die agentenbasierte Simulation bildet oft eine Alternative zur ereignisgesteuerten Simulation und bietet eine noch größere Flexibilität als Letztere. (Schöffski et al. 2011)

8.3.3.6 Systems-Dynamics-Modelle

Systems-Dynamics-Modelle sind deterministische Kohortenmodelle, die ein System mithilfe zusammenfassender Größen beschreiben. Charakteristisch für Systems-Dynamics-Modelle (SD-Modelle) ist die Beschreibung des Systems durch Flussdiagramme und Feedback Loops. Diese Diagramme werden in Differenzialgleichungen übersetzt. Systems-Dynamics-Modelle können unerwartetes nichtlineares Verhalten von Systemen beschreiben. Die Zeit wird dabei kontinuierlich behandelt. Im Gesundheitswesen werden SD-Modelle für ähnliche Probleme eingesetzt wie die diskrete ereignisorientierte Modellierung, d. h. insbesondere für operationale Optimierung, aber auch für die Modellierung von Infektionskrankheiten. Einen Vergleich von SD- und DES-Modellen (s. Abschnitt 8.3.3.3) im Gesundheitswesen geben die Publikationen (Brailford 2007 und Schöffski et al. 2011).

8.3.4 Basisfall- und Sensitivitätsanalyse

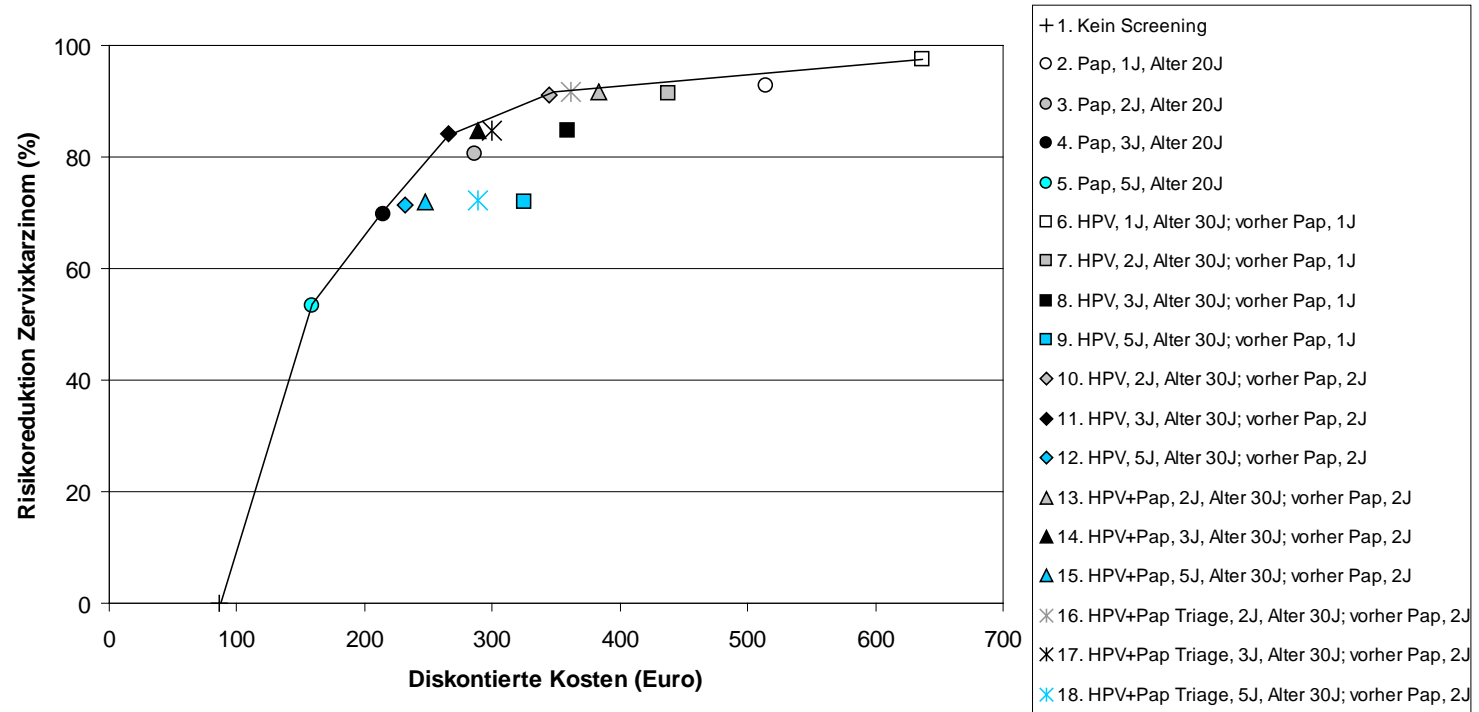
Neben der Basisfallanalyse sind zur Beurteilung der Unsicherheit ausreichende Sensitivitätsanalysen durchzuführen (Siebert 2003; Weinstein, O'Brien et al. 2003; Drummond, Schwartz et al. 2008; Briggs, Fenwick et al. 2011).

In der entscheidungsanalytischen Modellierung wird zwischen Basisfallanalysen und Sensitivitätsanalysen unterschieden. In der Basisfallanalyse eines Modells werden die wahrscheinlichsten Struktur- und Parameterannahmen verwendet. Die Basisfallanalyse stellt damit eine Art Primäranalyse dar. Eine Variation der Annahmen ist im Rahmen von (sekundären) Sensitivitätsanalysen möglich, deren Ergebnisse dann mehr oder weniger vom Ergebnis der Basisfallanalyse abweichen (Siebert, Mühlberger et al. 2000; Siebert, Mühlberger et al. 2000).

8.3.4.1 Basisfallanalyse

Bei der Basisfallanalyse werden unter Annahme der wahrscheinlichsten Parameter die Zielgrößen der Evaluation berechnet. Bei einer Kosten-Effektivitäts-Analyse sind das die erwartete Effektivität und die Kosten sowie das inkrementelle Kosten-Effektivitäts-Verhältnis für den untersuchten Zeithorizont und für jede Intervention. In einem Diagramm der Kosten-Effektivitäts-Ebene können diese Ergebnisse als Punkte eingetragen werden. Die Effizienz einer Intervention wird anhand des inkrementellen Kosten-Effektivitäts-Verhältnisses, kurz IKEV (engl. ICER: „incremental cost-effectiveness ratio“), im Vergleich zu einer geeigneten Vergleichsintervention gemessen. Je geringer die zusätzlichen Kosten pro erzielter Einheit zusätzlicher Gesundheit, desto effizienter ist eine Strategie. Werden mehr als zwei Strategien bewertet, werden zunächst diejenigen Strategien eliminiert, die aus medizinischer und ökonomischer Sicht keinen Sinn ergeben, da sie im engeren oder erweiterten Sinne durch eine andere Strategie dominiert werden. Eine Strategie bezeichnet man als dominiert, wenn es eine mindestens ebenso effektive, aber kostengünstigere Strategie gibt. Man spricht von erweiterter Dominanz, wenn eine Strategie von einer Linearkombination zweier anderer Strategien dominiert wird. Das Eliminierungsverfahren ist dabei iterativ. Die Eliminierung wird so oft wiederholt, bis keine dominierten oder erweitert dominierten Strategien mehr enthalten sind. Das Ergebnis dieses Verfahrens ist ein Satz von Strategien, die auf der sogenannten Effizienzgrenze liegen. Für jede dieser verbleibenden Strategien wird das inkrementelle Kosten-Effektivitäts-Verhältnis zur jeweils nächsten weniger effektiven Strategie gebildet. Wenn man in einem Diagramm der Kosten-Effektivitäts-Ebene die Strategien auf der Effizienzgrenze in der Ordnung zunehmender Effektivität miteinander verbindet, bilden die IKEV die Steigungen der Segmente zwischen je zwei Strategien. Ein Beispiel für die Effizienzgrenze von unterschiedlichen Zervixkrebs-Screeningstrategien gibt Tabelle 8.21. (Sroczynski 2010)

Abbildung 8.8:
Zervixkrebs-Screeningstrategien, Effizienzgrenze



Quelle und Darstellung: Sroczyński 2010

Alle Ergebnisse sollten bei Darstellung als Kosten–Effektivitäts–Diagramm auch in numerischer Form, z. B. als Tabelle, berichtet werden. Außer den absoluten und inkrementellen Ergebnissen für Lebensjahre bzw. QALYs, Kosten und IKEV werden auch die Ergebnisse für alle übrigen evaluierten Nutzenparameter angegeben. Die IKEV sollen dabei mit diskontierten Kosten und Effekten berechnet werden. Die gesundheitlichen Outcomes werden zusätzlich undiskontiert berichtet. Bei Verwendung von aggregierten Nutzenmaßen wie dem QALY ist es sinnvoll, die Effekte auch disaggregiert zu berichten.

Wird die Basisfallanalyse deterministisch als Kohortensimulation durchgeführt, dann sind die Ergebnisse jeweils die Erwartungswerte. Bei Mikrosimulation erhält man als Ergebnis nicht nur einen Erwartungswert, sondern auch eine Verteilung der Ergebnisse, die darauf beruht, dass die Pfade der einzelnen Patienten durch das Modell zufällig sind und zu unterschiedlichen Ergebnissen führen (engl. „first–order uncertainty“). Diese Streuung ist wichtig, wenn für individuelle Patienten die Frage gestellt wird, wie wahrscheinlich ein Behandlungsergebnis ist, das vom Erwartungswert abweicht. Im Rahmen von HTA ist diese Streuung in der Regel weniger bedeutsam, kann aber für Planungsaspekte bspw. für Krankenhäuser wichtig sein.

Die Variation der Basisfallergebnisse aufgrund der natürlichen Heterogenität der Patienten (Variabilität) ist dagegen ein wichtiges Ergebnis, da sie zu unterschiedlichen Empfehlungen für bestimmte Patientengruppen führen kann (Hunink, Glasziou et al. 2001). Unter Heterogenität versteht man die Unterschiede in der Ausprägung von Variablen bei den Patienten, beispielsweise verschiedene Risikoprofile. Soweit Kenntnis über die Heterogenität der Zielpopulation vorliegt, wird sowohl in der Kohorten– als auch in der Mikrosimulation die Heterogenität durch Subgruppenanalysen untersucht. Die Simulation wird dabei für jede der relevanten Gruppen wiederholt.

8.3.4.2 Sensitivitätsanalyse

Die Unsicherheit in einem Modellergebnis resultiert aus den Unsicherheiten der Modellstruktur und der Werte der Modellparameter.

Sensitivitätsanalysen sind Verfahren, die die Auswirkung unsicherer struktureller und parametrischer Modellannahmen auf das Ergebnis der Modellierung (z. B. IKEV) untersuchen (Briggs, Sculpher et al. 1994; Siebert, Mühlberger et al. 2000; Siebert, Mühlberger et al. 2000). „Strukturelle und numerische Annahmen werden dabei über einen möglichen Bereich variiert. Wichtige Ziele dabei sind die Prüfung der Stabilität der Entscheidung über die Kosteneffektivität, die Identifizierung einflussreicher Parameter und die Ermittlung von Schwellenwerten.“ (Klein 2010)

Zeigt sich, dass die Modellergebnisse, insbesondere die Entscheidung über Kosteneffektivität, sensitiv auf bestimmte Parameter reagieren, dann muss für diese Parameter die Datenqualität überprüft werden. Sollte nicht genügend Evidenz vorliegen, um eine verlässliche Aussage über die Kosteneffektivität zu ermöglichen, sind weitere Studien zur Verbesserung der Datenqualität zu empfehlen. Die Value-of-Information-Analyse stellt eine formale Methode dar, den Aufwand für die Erhebung neuer Daten mit dem Nutzen der dadurch gewonnenen Information in Beziehung zu setzen. Eine übersichtliche Beschreibung der Value-of-Information-Analyse und zahlreiche Referenzen bietet die Publikation (Groot Koerkamp, Weinstein et al. 2010).

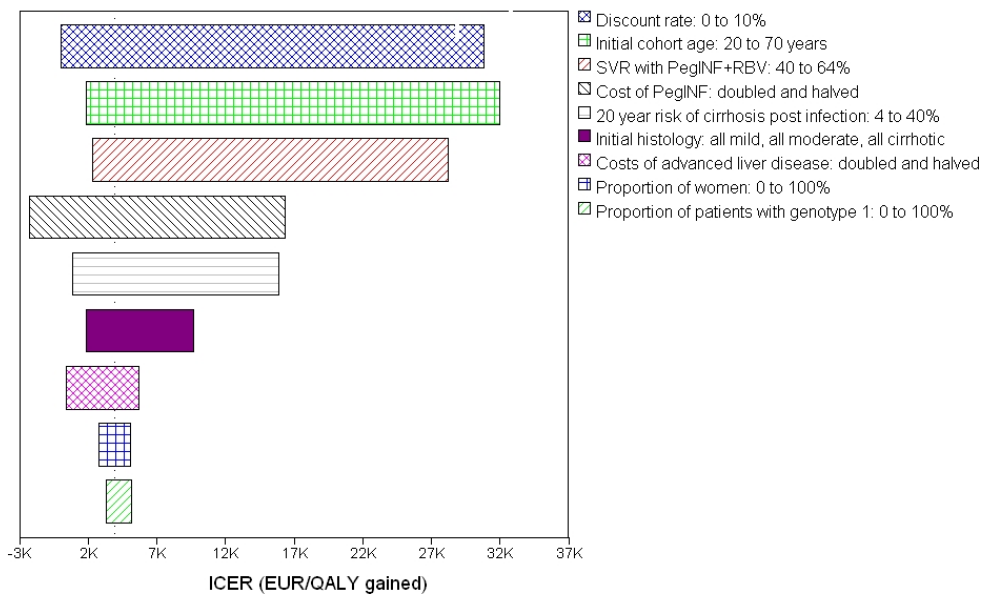
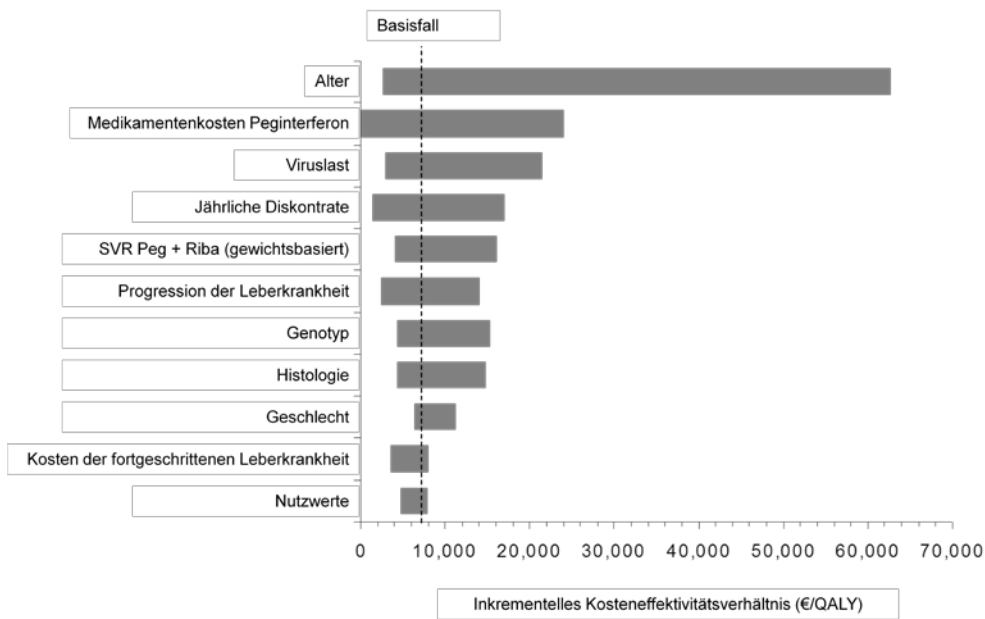
Sensitivitätsanalysen können vom Ansatz her deterministisch oder probabilistisch konzipiert sein.

8.3.4.2.1 Deterministische Sensitivitätsanalyse

In deterministischen Sensitivitätsanalysen werden unsichere Einflussparameter über einen Bereich variiert, und die Ergebnisgrößen werden in Abhängigkeit vom Wert des Parameters dargestellt. Das kann getrennt für jeden Parameter erfolgen (Einweg-Sensitivitätsanalysen), oder es können mehrere Parameter gleichzeitig variiert werden (Mehrweg-Sensitivitätsanalysen). Die Ergebnisse aller Einweg-Sensitivitätsanalysen können in einem Tornado-Diagramm dargestellt werden (Abbildung 8.9). Dabei werden die Parameter nach der Stärke ihres Einflusses geordnet. Man erhält auf diese Weise eine Vorstellung davon, in welchem Umfang das IKEV variiert, wenn Parameter variiert werden. Die Möglichkeiten der grafischen Darstellung von deterministischen Mehrweganalysen sind beschränkt. Bei in der Praxis anzutreffenden Mehrweganalysen werden in der Regel nicht mehr als drei Parameter gleichzeitig variiert.

Abbildung 8.9:

Tornado-Diagramm von Einweg-Sensitivitätsanalysen zum inkrementellen Kosten-Effektivitäts-Verhältnis einer Behandlung mit pegyliertem Interferon und Ribavirin im Vergleich zur Behandlung mit Interferon und Ribavirin



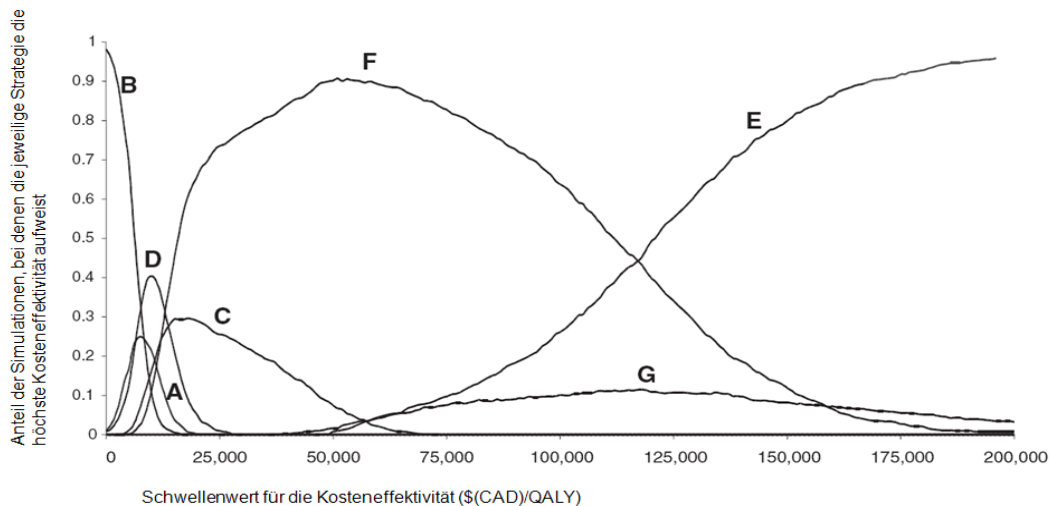
Quelle und Darstellung: (Siebert, Sroczynski et al. 2003)

Die Wahl des Bereichs, in dem ein Parameter variiert wird, ist weniger kritisch, wenn nur festgestellt werden soll, ob das Modellergebnis auf Variation eines Parameters empfindlich reagiert oder nicht. Geht es um die Quantifizierung der Unsicherheit innerhalb des Modells, dann muss der gewählte Bereich gut begründet sein und möglichst auf Daten basieren, die die Unsicherheit der Parameterwerte wiedergeben (ISPOR–SMDM Modeling Good Research Practices Task Force 2011). Liegen für einen Parameter zum Zeitpunkt der Analyse keine oder nur sehr unsichere empirische Daten vor, kann eine Modellierung mit eingehender Sensitivitätsanalyse für diesen Parameter Schwellenwerte ermitteln, an denen das IKEV gewisse Grenzen überschreitet.

8.3.4.2.2 Probabilistische Sensitivitätsanalyse

Zu den probabilistischen Sensitivitätsanalysen schreiben Klein und Kollegen (Klein et al. 2010): „In probabilistischen Sensitivitätsanalysen (synonym: verteilungsorientierte Sensitivitätsanalyse) werden den einzelnen Parametern des Modells Verteilungen zugeordnet. Unter Verwendung eines Zufallszahlengenerators können dann für jeden unsicheren Einflussparameter Werte zufällig und wiederholt aus den zugehörigen Verteilungen gezogen werden. Nach jeder Ziehung erfolgt eine erneute Berechnung der Erwartungswerte. Daraus resultiert eine Wahrscheinlichkeitsverteilung der erwarteten Ergebnisparameter. Es lässt sich dann angeben, mit welcher Wahrscheinlichkeit das IKEV einer Intervention unterhalb vorgewählter Zahlungsbereitschaftsgrenzen liegt. Das Ergebnis der probabilistischen Analyse wird häufig in Form eines Kosteneffektivitäts-Diagramms oder einer Kosteneffektivitäts-Akzeptanz-Kurve dargestellt.“ (Klein 2010). Die in probabilistischen Sensitivitätsanalysen verwendeten Verteilungen sollten gut begründet sein. In der Dokumentation des Modells sollten sie beschrieben werden.

Abbildung 8.10:
 Kosten–Nutzwert–Akzeptanzkurven zu drei diagnostischen Strategien für die
 Langzeitbehandlung von Patienten mit Sodbrennen



Quelle und Darstellung: (Goeree, O'Brien et al. 2002)

Bei einer probabilistischen Sensitivitätsanalyse können alle unsicheren Parameter gleichzeitig variiert werden. Dadurch erhält man Ergebnisse, die durch deterministische Sensitivitätsanalysen nicht zu erzielen sind. Andererseits erhält man dabei keine Aussage über den Einfluss einzelner Parameter auf das Ergebnis. Deshalb ist es ratsam, neben einer probabilistischen Sensitivitätsanalyse auch deterministische Analysen durchzuführen. Das Ergebnis einer probabilistischen Sensitivitätsanalyse wird häufig zusammengefasst in einer Kosteneffektivitäts–Akzeptanzkurve. Ein Beispiel dafür (Goeree et al. 2002) wird in Abbildung 8.10 gegeben.

Für einen Überblick über die Quellen von Unsicherheit und den Umgang damit sei auf die einschlägige Literatur verwiesen (Briggs, Sculpher et al. 1994; Briggs 2000). Empfehlungen für den Umgang mit Unsicherheit werden von der neuen ISPOR–SMDM Modeling Good Research Practices Task Force entwickelt (Briggs, Fenwick et al. 2011).

8.3.5 Qualitätskriterien für die Modellierung

Im Kapitel 8.1 wurden bereits Publikationen genannt, die sich mit der Qualitätsbewertung von gesundheitsökonomischen Evaluationen befassen. Ebenso wurden zwei Publikationen genannt, die speziell die Modellierung betreffen (Weinstein, O'Brien et al.

2003; Philips, Bojke et al. 2006). Diese Studien geben Checklisten an, nach denen Modelleigenschaften und Beschreibung eines Modells beurteilt werden können.

Für den Entscheidungsträger wird der Nutzen eines Modells mit guter Validität vor allem von folgenden Kriterien bestimmt, die in den Empfehlungen des IQWiG angeführt werden (IQWiG 2009):

Transparenz

Die vollständige Beschreibung des verwendeten Dateninputs und klare Darstellung aller verwendeten funktionalen Beziehungen ist Voraussetzung für die Nachvollziehbarkeit des Modells. Modellannahmen müssen umfassend beschrieben werden, um Grenzen für die Interpretation der Modellergebnisse aufzuzeigen. (IQWiG 2009)

Ausreichende Tiefe/Detaillierungsgrad

Ein Modell sollte die modellierte Erkrankung und die mit ihr assoziierten Kosten und Versorgungsmaßnahmen in hinreichender Tiefe und mit hinreichendem Detaillierungsgrad abbilden. Die Verfügbarkeit von Daten sollte die Detailliertheit nicht in unangemessener Weise einschränken. (IQWiG 2009)

Flexibilität

Flexibilität zur Berechnung multipler Szenarien bei variierenden Annahmen und Settings erlaubt die Beantwortung erweiterter Fragestellungen. (IQWiG 2009)

Möglichkeit der Bestimmung von Unsicherheit bei Kosten und Nutzen

Die Möglichkeit der Bestimmung von Unsicherheit bei den vorhergesagten Kosten- und Nutzenkomponenten ist als Entscheidungsgrundlage ebenso wichtig wie die Angabe eines Erwartungswerts für das Kosten-Effektivitäts-Verhältnis. (IQWiG 2009)

Verwendung von Daten, die für den Kontext relevant sind

Der Nutzen eines Modells innerhalb eines bestimmten Gesundheitssystems ist umso größer, je besser es den regionalen Kontext beschreibt. Dabei sollten nicht nur Kosten, sondern auch praxisübliche Behandlungsmuster, Demografie und Epidemiologie berücksichtigt werden. (IQWiG 2009)

Autorinnen/Autoren: Annette Conrads-Frank, Gaby Sroczynski, Beate Jahn, Nikolai Mühlberger, Petra Schnell-Inderst, Uwe Siebert (UMIT/ONCOTYROL)

8.4 Ökonomische Empfehlung

8.4.1 Einleitung

Health Technology Assessments (HTA) nehmen eine Bewertung des medizinischen Nutzens, der Nutzen-Schadens-Abwägung, der Wirtschaftlichkeit sowie sozialer, ethischer, rechtlicher und organisatorischer Aspekte von gesundheitsrelevanten Verfahren, Leistungen, Technologien oder Prozeduren vor. Das Methodenhandbuch für Health Technology Assessment erläutert in seiner ersten Version 1.2010 (GÖG 2011) die Methodik zur Bewertung der Wirksamkeit medizinischer Verfahren und stellt mögliche Herangehensweisen zur Bewertung sozialer, organisatorischer, ethischer und rechtlicher Aspekte und Wirkungen dieser Technologien dar. Die Methodik der Bewertung der Wirtschaftlichkeit von Gesundheitsleistungen blieb bislang ausgeklammert.

Die im Gesundheitswesen vorhandenen beschränkten Geldmittel werden größtenteils öffentlich verwaltet; der Einsatz dieser Mittel ist sorgfältig abzuwägen, um den größtmöglichen Nutzen für die Patientinnen und Patienten zu erzielen. Bei dieser Entscheidung können gesundheitsökonomische Studien zu verschiedenen medizinischen Maßnahmen unterstützen. Darunter fallen einerseits gesundheitsökonomische Evaluationen (d. h. Studien, die Kosten und Gesundheitseffekte von zwei oder mehreren Gesundheitstechnologien vergleichend analysieren), Krankheitskostenstudien und Budget-Impact-Analysen.

International kommt aufgrund der beschränkten Ressourcen gesundheitsökonomischen Studien zur Entscheidungsunterstützung zunehmend größere Bedeutung zu. Konzepte, wie diese Entscheidungsunterstützung aussehen soll, differieren allerdings stark (z. B. hinsichtlich der Kosten), da sie sich an den lokalen Gegebenheiten zu orientieren haben und nicht einfach zwischen verschiedenen Ländern übertragbar sind. Zu berücksichtigen sind u. a. die Epidemiologie, die organisatorischen Strukturen und Versorgungspfade des Gesundheitswesens, die Vergütung der Gesundheitsleistungen, aber auch die Wertehaltung in einer Gesellschaft.

Ökonomische Empfehlungen sollen Erkenntnisse vorhandener Richtlinien sammeln, diese gegebenenfalls ergänzen und einen Bezug zu Österreich herstellen. Dazu liegen verschiedene Arbeiten vor: drei deutschsprachige („Guidelines zur gesundheitsökonomischen Evaluation“ des Institut für Pharmaökonomische Forschung / IPF 2006, „Deutsche Empfehlungen zur gesundheitsökonomischen Evaluation“ des Hannoveraner Konsens und „Allgemeine Methoden zur Bewertung von Verhältnissen zwischen Nutzen und Kosten“ des IQWiG) und drei englischsprachige („Guidelines for the Economic Evaluation of Health Technologies: Canada“ der Canadian Agency for Drugs and Technologies in Health, „The Guideline Manual“ des NICE und „Key principles for the

improved conduct of health technology assessments for resource allocation decisions“ von Drummond et al.).

Die Inhalte der ökonomischen Empfehlung werden einem öffentlichen Stellungnahmeverfahren unterzogen.

Den Rahmen für die gesundheitsökonomische Evaluation von Gesundheitsleistungen bilden die gesetzlichen Regelungen über die Gesundheitsversorgung in Österreich. Das Allgemeine Sozialversicherungsgesetz (ASVG § 133) hält fest, dass die Krankenbehandlung ausreichend und zweckmäßig sein sollte, das Maß des Notwendigen aber nicht zu überschreiten hat (Ökonomiegebot). Bei der Bewertung von Gesundheitsleistungen sollen Interventionen verglichen werden, wobei hier auch der Vergleich unterschiedlicher Ansätze heranzuziehen ist.

Obwohl der rechtliche Rahmen wiederholt auf die Wirtschaftlichkeit von Gesundheitsleistungen verweist, haben gesundheitsökonomische Studien und insbesondere gesundheitsökonomische Modellrechnungen in Österreich keine lange Tradition. Nur in einzelnen Bereichen werden sie systematisch gefordert. So führt die Verfahrensordnung zur Herausgabe des Erstattungskodex nach § 351 ASVG (VO-EKO) aus, dass für Arzneimittel mit zusätzlichem oder wesentlichem zusätzlichem Nutzen (sogenannte innovative Arzneimittel) gesundheitsökonomische Evaluationen vorzulegen sind (§ 25 VO-EKO). Als Teil dieser gesundheitsökonomischen Evaluation ist die Vorlage von pharmakoökonomischen Studien vorgesehen. Anzusetzen sind hierbei *„die direkten Kosten der Pflichtleistungen der Sozialversicherungsträger der Krankenbehandlung (Ärztliche Hilfe, Heilmittel, Heilbehelfe), der Anstaltspflege (auf Basis der LKF-Punkte) sowie der medizinischen Maßnahmen der Rehabilitation auf Basis der tatsächlich verrechneten Preise“* (§ 25 VO-EKO).

Vorliegende ökonomische Empfehlung sollen Erstellerinnen und Erstellern von ökonomischen Studien als Leitlinie dienen, die Ergebnisse dieser Studien so weit wie möglich vergleichbar machen und damit auch aus ökonomischer Perspektive informierte Entscheidungen ermöglichen.

Die an der Erstellung dieser Empfehlung beteiligten HTA-Institutionen verpflichten sich, die beschriebenen Methoden in allen vom BMG und von der Bundesgesundheitsagentur beauftragten gesundheitsökonomischen Studien anzuwenden. Im Rahmen der nationalen HTA-Strategie wird das mittelfristige Ziel verfolgt, dass auch andere (öffentliche) Einrichtungen bzw. Auftraggeber/innen sowie HTA-Anbieter/innen diese Empfehlung ihren Projekten zugrunde legen. Damit soll es gelingen, in Österreich einheitliche methodische Standards zu etablieren.

Außerdem soll diese Empfehlung einen Anstoß geben für einen grundsätzlichen Diskurs in Österreich auch hinsichtlich einer stärkeren rechtlichen Verankerung gesundheitsökonomischer Studien für die effizientere Verwendung knapper Ressourcen.

8.4.2 Ökonomische Empfehlung im Detail

8.4.2.1 Forschungsfrage

Jeder wissenschaftlichen Arbeit geht eine Fragestellung voraus. Die Fragestellung bestimmt die Bandbreite einer Studie, das Studiendesign und die Berichtsform. Da die Forschungsfrage der Ausgangspunkt jeder wissenschaftlichen Arbeit ist und viele weitere Schritte beeinflusst, können Versäumnisse bei der Formulierung die Aussagekraft der Studie abschwächen.

Bevor über eine Forschungsfrage nachgedacht wird, müssen die Autorinnen und Autoren über die Adressaten der Studie und ihre Bedürfnisse ausreichend informiert sein. Adressaten von ökonomischen Studien (Politiker, Entscheidungsträger, Patientengruppen etc.) wollen mithilfe einer Studie Fragen beantworten oder Entscheidungen treffen. Nur wenn die Fragestellung auf die individuellen Bedürfnisse der Adressaten abgestimmt ist, können die Erkenntnisse auch verarbeitet werden. Weiters soll die Fragestellung nie zu weit, sondern präzise formuliert sein.

Die wichtigsten Bestandteile einer Forschungsfrage sind die Patientenpopulation, die Studienperspektive, Outcomes sowie die untersuchte Intervention und ihre Alternative(n). In einzelnen Fällen kann die Fragestellung um weitere Aspekte erweitert werden. Die oben angeführten Punkte werden im Anschluss genauer erläutert und sollten immer Bestandteil der Forschungsfrage sein.

8.4.2.2 Studiendesign

Die Wahl des Studiendesigns soll nachvollziehbar sein und begründet werden. Die gängigen Studiendesigns für ökonomische Fragestellungen sind die Kosten-Effektivitäts-Analyse, Kosten-Nutzwert-Analyse, Kosten-Nutzen-Analyse und Kosten-Minimierungsanalyse. Die Studiendesigns werden in Kapitel 4.7 und 6 detailliert dargestellt.

8.4.2.3 Perspektive

In einer ökonomischen Studie können Kosten und Nutzen aus verschiedenen Perspektiven betrachtet und bewertet werden. Die Perspektive orientiert sich primär am Adressaten und an deren Bedürfnissen, Fragen und Zielen. Mögliche Perspektiven sind die Sicht der gesamten Gesellschaft, des gesamten oder öffentlichen Gesundheitssystems, der Sozialversicherungen, einzelner Krankenkassen, Krankenanstalten, Ärztinnen/Ärzte und Patientinnen/Patienten. In vielen ökonomischen Richtlinien werden die gesellschaftliche Sicht oder die Sicht des öffentlichen Gesundheitssystems als primäre Perspektiven empfohlen bzw. vorgegeben. Im Idealfall muss die Perspektivenwahl logisch aus der Fragestellung herausgearbeitet werden und sich an den Bedürfnissen der Adressaten orientieren. Auf jeden Fall muss die Perspektivenwahl begründet werden. Betrachtet die Studie mehrere Perspektiven, müssen die Inputs und Outputs jedenfalls für jede Perspektive getrennt angegeben und begründet werden. Unabhängig von der Perspektivenwahl muss gewährleistet sein, dass inkludierte Kosten und Nutzen stets aus der gleichen Sicht betrachtet und bewertet bzw. soweit möglich synchronisiert werden. Kann die gleiche Sichtweise aufgrund von Datenschwächen nicht gewährleistet werden, muss auf diese Mängel im Bericht aufmerksam gemacht werden.

8.4.2.4 Zielpopulation

Als Zielpopulation bezeichnet man die Gruppe von Patienten, die in einer Studie betrachtet wird. Die Zielpopulation kann anhand vieler Merkmale identifiziert werden. Merkmale können sein: demografische Daten (Alter, Geschlecht, sozialer Status, geographische Unterschiede etc.), Krankheitsmerkmale (ob erkrankt, in welchem Stadium, spezielle Risikofaktoren etc.), Behandlungsmerkmale (Einhalten von Therapien etc.) oder Umfeld (Krankenhaus, Pflegeheim, ambulante Versorgung, Hauskrankenpflege etc.).

Die Zielpopulation muss im Rahmen der Forschungsfrage (dem ersten Schritt jeder Studie) präzise festgelegt werden. Eine Betrachtung von mehr als einer Patientengruppe ist möglich. Die Wahl einer oder mehrerer Patientengruppen muss nachvollziehbar und begründet sein. Bei der Betrachtung mehrerer Patientengruppen müssen die Input-Parameter der gesundheitsökonomischen Studien und die Ergebnisse jedenfalls für jede Zielpopulation getrennt voneinander angegeben werden.

Die in die Studie einfließenden Wirksamkeits-/Effektivitäts-Daten müssen auf die gewählte(n) Zielpopulation(en) anwendbar sein. Liegen dazu nicht ausreichende Daten vor, kann eine nachvollziehbare und begründete Anpassung geschehen. Bei angepas-

ten oder nicht relevanten Daten für die gewählte Patientenpopulation muss in der Studie auf jeden Fall darauf hingewiesen werden.

Im Idealfall ermöglichen die vorliegenden Daten eine Analyse der gesamten gewünschten Patientenpopulation. Bei hoher Heterogenität in der Patientenpopulation wird den Autorinnen/Autoren einer ökonomischen Studie jedoch nahegelegt, die Patientengruppe in angemessene Subgruppen aufzuteilen, um Unterschiede zu identifizieren, valide Ergebnisse zu erhalten und damit eine subgruppenspezifische Entscheidungsunterstützung zu ermöglichen.

8.4.2.5 Intervention und Alternativen

Bei ökonomischen Problemstellungen sollen wie auch bei Aussagen über Wirksamkeit oder Nebenwirkungen von Maßnahmen verschiedene Alternativen (Technologien) bewertet und gegenübergestellt werden. Interventionen und Alternativen sind beispielsweise verschiedene Arzneimittel, diagnostische Verfahren oder Behandlungsarten. Die Auswahl der untersuchten Intervention(en) und Alternative(n) muss stets nachvollziehbar sein und begründet werden.

8.4.2.6 Validität der Datenquellen

Zur Erstellung gesundheitsökonomischer Studien werden klinische, ökonomische und epidemiologische Daten herangezogen. Alle Datenquellen müssen genau beschrieben werden, und ihre Auswahl ist zu begründen. Außerdem sollte die interne und externe Validität bewertet werden (IPF 2006) (siehe dazu 6). Zum Evidenzgrad der Daten siehe 4.8.

8.4.2.7 Zeithorizont

Die Wahl des Zeithorizonts hängt von der Forschungsfrage bzw. vom Studiengegenstand ab. Der Zeithorizont kann reichen von wenigen Stunden (z. B. bei der Evaluation von Krankenhausmanagementstrategien) über wenige Wochen (akute Erkrankungen wie Influenza), mehrere Jahre (Krebserkrankung mit hoher Mortalität) bis zu einem lebenslangen Zeitraum (chronische Erkrankungen wie Diabetes). Weicht der gewählte Zeithorizont von der für den Studiengegenstand üblichen Dauer ab, so muss diese Entscheidung ausreichend begründet werden.

Grundsätzlich gilt bei der Wahl des Zeithorizonts, dass der Einfluss der untersuchten Interventionen auf den Ressourcenverbrauch und die Gesundheitseffekte (Effektivität, medizinischer Schaden, Nutzwert, Lebensqualität, Lebenserwartung, QALYs etc.) beobachtbar bzw. durch Daten belegt ist. In vielen Fällen liegen keine Langzeitdaten vor, sodass eine ökonomische Studie nicht möglich wäre. In diesen Fällen können entscheidungsanalytische Modelle (s. Kapitel 8.3) herangezogen werden, um dieses Problem zu umgehen.

8.4.2.8 Modellierung

Auch wenn klinische Studien ökonomische Daten erfassen, reichen diese häufig nicht aus, um die Kosteneffektivität einer Intervention abzuschätzen. Oft sind diese Studien zeitlich begrenzt, oder die Versorgung der Patientinnen/Patienten und der Ressourcenverbrauch entsprechen nicht den Bedingungen im Versorgungsalltag des gewünschten Kontexts. In diesen Fällen stellt die entscheidungsanalytische Modellierung die Methode der Wahl dar. Die entscheidungsanalytische Modellierung (kurz: Entscheidungsanalyse) ist ein systematisches und quantitatives Verfahren der Evidenzsynthese, das es ermöglicht, die Evidenz aus verschiedenen Studien (verschiedener Studientypen) zusammenzuführen (Hunink, Glasziou et al. 2001) (Siebert 2003). Das Ziel der Entscheidungsanalyse ist die Auswahl der optimalen Handlungsalternative unter Berücksichtigung medizinischer und gesundheitsökonomischer Aspekte (Hunink, Glasziou et al. 2001) (Siebert 2003). Die Ergebnisse der Entscheidungsanalyse sollen die Entscheidungsträger bei der Frage nach der Einführung oder Erstattung medizinischer Technologien unterstützen. Eine Entscheidungsanalyse kann sich auf rein medizinische Aspekte beziehen (medizinische Entscheidungsanalyse) oder die Ermittlung der Kosteneffektivität von Technologien zum primären Ziel haben (gesundheitsökonomische Entscheidungsanalyse). Die gesundheitsökonomische Entscheidungsanalyse (gesundheitsökonomische Modellierung) wird in Kapitel 8.3 eingehend beschrieben. Der folgende Absatz enthält eine Zusammenfassung von Empfehlungen zur Modellierung.

Der für die Entscheidungsanalyse einzusetzende Modelltyp wird insbesondere durch die präzise, explizit und detailliert ausgearbeitete Fragestellung und die Rahmenbedingungen bestimmt. Das Modell sollte detailliert genug sein, um die Forschungsfrage beantworten zu können, und nicht wegen unzureichender Datenlage eingeschränkt werden. Andererseits ist unnötige Komplexität zu vermeiden, um die Transparenz und Verständlichkeit des Modells zu erhöhen.

Schon beim Modellentwurf sollten Modellierer mit klinischen Experten zusammenarbeiten, die die Augenscheinvalidität des Modells überprüfen. Das Modell sollte den gegenwärtigen Stand der Erkenntnis über eine Krankheit widerspiegeln. Die Nutzung

grafischer Darstellungsmöglichkeiten für das Modellkonzept sollte genutzt werden, um die Verständlichkeit zu erhöhen und die Kommunikation zu erleichtern. Strukturelle Annahmen, die in den Modellentwurf eingehen, sollten festgehalten und berichtet werden.

Die Auswahl des Modelltyps (z. B. Entscheidungsbaum, Markov-Modell etc.) sollte nicht a priori festgelegt werden. Der Modelltyp sollte ausreichend flexibel sein, um die Fragestellung zu beantworten. Technische Probleme und Mangel an Daten sollten nicht zu unangemessenen Vereinfachungen führen. Die Art des Problems und das Modellkonzept legen in der Regel bestimmte Modelltypen nahe, die sich zur Umsetzung eignen. Empfehlungen, welche Modelltypen jeweils geeignet sind, sind im Kapitel 8.3 angegeben.

Die Auswahl der Modellparameter sollte den Prinzipien der evidenzbasierten Medizin folgen (Briggs, Fenwick et al. 2011). Die gesamte relevante Evidenz sollte einfließen. Bei der Auswahl der Datenquellen sollte die Studienqualität beachtet werden. Die formale Synthese von Daten in Metaanalysen sollte, sofern angebracht, verwendet werden. In Kapitel 6 ist die Qualitätsbewertung von Studien zur Wirksamkeit medizinischer Interventionen detailliert beschrieben.

Zur Ermittlung von Nutzwerten (Lebensqualitätsindizes) siehe Abschnitt 8.4.2.9. Bei Kostendaten ist insbesondere darauf zu achten, dass sie dem gewünschten Kontext entstammen, da hier regionale und regulatorische Unterschiede besonders stark sein können.

Alle Input-Parameter sollten mit einem Punkt- und Intervallschätzer sowie der Quellenangabe im Modellbericht aufgeführt werden. Die Qualität der Daten sollte beschrieben und die Auswahl der Datenquellen begründet werden. Werden Expertenmeinungen herangezogen, muss auf jeden Fall eine umfassende Sensitivitätsanalyse erfolgen. Auch für alle anderen relevanten Parameter sollte eine Sensitivitätsanalyse erfolgen (s. Abschnitt 8.4.2.12 dieses Kapitels und Kapitel 8.3.4.2).

Es ist eine Verifikation (interne Validierung) des Modells durchzuführen, und die Methoden sind detailliert zu beschreiben. Das Modell sollte nach Möglichkeit auch einer externen Validierung unterzogen werden, und die Validierungsmethoden sollten berichtet werden. Zur Validierung von Modellen s. auch Kapitel 8.3.2.7.

Die Modellannahmen sind möglichst vollständig zu erfassen und im Modellbericht klar zu beschreiben.

8.4.2.9 Bewertung des Gesundheitseffekts und Wahl der Zielgrößen

Effektivität

Wirksamkeitsdaten sollten wenn möglich aus randomisierten klinischen Studien bzw. Metaanalysen dieser Studien ermittelt werden. Ein systematischer Review sollte der Ausgangspunkt der Evaluation sein. Die Wirksamkeit (engl. „efficacy“) einer Intervention, die unter kontrollierten Bedingungen innerhalb eines Studienprotokolls in einer selektierten Patientenpopulation ermittelt wird und in der Regel kurz- und mittelfristige Ergebnisparameter misst, genügt jedoch nicht zur Schätzung der Effektivität in einer Kosten-Effektivitäts-Studie. Effektivität ist zu verstehen als Gesamtnutzen einer Intervention unter Routinebedingungen. Darin sollen alle wesentlichen Konsequenzen der Intervention enthalten sein, insbesondere auch langfristige Folgen und Risiken bzw. negative Folgen bzw. Schäden der Behandlung. Um die Effektivität abzuschätzen, muss häufig die in klinischen Studien ermittelte Wirksamkeit mit Daten aus anderen Studien kombiniert werden. Surrogatparameter, die Ergebnisgrößen in klinischen Studien sind, können mit epidemiologischen Studien kombiniert werden, um auf patientenrelevante Endpunkte wie Mortalität, Morbidität und Lebensqualität zu extrapolieren (Siebert 2003). Der kausale Zusammenhang zwischen Surrogatparameter und diesen Endpunkten sollte dabei allerdings belegt sein. Die Studienqualität und das höhere Verzerrungspotenzial in epidemiologischen Studien sollte in Sensitivitätsanalysen berücksichtigt werden und muss bei Interpretation der Ergebnisse in Betracht gezogen werden.

Bei der Bewertung von diagnostischen Tests müssen in die Bewertung der Effektivität sowohl Testgenauigkeit als auch Konsequenzen der nachfolgenden Behandlung in die Bewertung einbezogen werden.

Der Unterschied zwischen Routinebedingungen und den Bedingungen einer klinischen Studie bestehen in der Adhärenz der Patienten, der Therapietreue der Ärzte, in der Erfahrung und in den Fähigkeiten der behandelnden Ärzte sowie in der Auswahl der Studienpopulation, z. B. dem Ausschluss von Komorbiditäten. Außerdem können durch das Studienprotokoll Bedingungen geschaffen werden, die im klinischen Alltag anders gehandhabt werden.

Zum Teil liegen Daten hierzu vor, und Routinebedingungen sollten so weit wie möglich berücksichtigt werden. Liegen keine Daten vor, können in Sensitivitätsanalysen Szenarien mit unterschiedlichen Annahmen (z. B. zur Adhärenz) untersucht werden.

Bei der Verwendung von Beobachtungsstudien oder Registerdaten ist darauf zu achten, dass adäquat auf Confounding kontrolliert wird.

Nutzwerte

Häufig haben Behandlungen nicht nur Auswirkungen auf die Lebensdauer, sondern auch auf die Lebensqualität. Dabei kann es sich um erwünschte oder unerwünschte Wirkungen handeln. Um alle Effekte in einem einheitlichen Maß zu integrieren, kann in Kosten-Nutzwert-Analysen das qualitätsadjustierte Lebensjahr verwendet werden (QALY). Wird das QALY verwendet, sollten Ergebnisse jedoch auf jeden Fall auch die dem QALY zugrunde liegenden klinischen Ereignisse und die hierfür verwendeten Nutzwerte berichtet werden. Zur Bestimmung der Nutzwerte sollten präferenzbasierte Maße verwendet werden. Zu Verfahren für die Bestimmung der Nutzwerte siehe Kapitel 7.3.2. Die Verwendung indirekter standardisierter Verfahren (HUI, EQ-5D, SF-6D, 15D) ist dabei durchaus wünschenswert, da Nutzwerte mit diesen Verfahren relativ einfach zu ermitteln und die Ergebnisse gut vergleichbar sind. Werden Nutzwerte direkt bestimmt, sollten Standard-Gamble oder Time trade-off verwendet werden. Visuelle Analogskalen werden eingesetzt, sind aber aus theoretischer Sicht problematisch.

Verwendung des QALY zur Ressourcenallokation

Neben der Integration verschiedener Zielgrößen bietet ein generisches Nutzenmaß wie das QALY die Möglichkeit, (a) innerhalb einer Indikation die verschiedenen Dimensionen der Konsequenzen der untersuchten Technologie in einem Maß zusammenzuführen und (b) indikationsübergreifende Nutzen- bzw. Kosten-Nutzen-Vergleiche durchzuführen. Ursprünglich wurde das QALY sogar dafür entwickelt, Entscheidungsträger bei der nutzenmaximalen Allokation begrenzter Ressourcen über verschiedene Bereiche des Gesundheitswesens zu unterstützen (Weinstein, Torrance et al. 2009). Das QALY-Konzept unterliegt allerdings auch besonderen Annahmen bzw. Limitationen, die im Rahmen eines HTA-Berichts angesprochen bzw. diskutiert werden müssen. Siehe hierzu auch die 2009 publizierten Ergebnisse einer Konsensuskonferenz der International Society for Pharmacoeconomics and Outcomes Research (ISPOR) zur Weiterentwicklung und Verbesserung des QALY (Drummond, Brixner et al. 2009; Garrison 2009; Johnson 2009; Kind, Lafata et al. 2009; Lipscomb, Drummond et al. 2009; Nord, Daniels et al. 2009; Smith, Drummond et al. 2009; Weinstein, Torrance et al. 2009). Wesentliche Kritikpunkte am QALY als Grundlage für Allokationsentscheidungen sind kontroverse und zum Teil empirisch widerlegte Annahmen des QALY-Konzepts, die im Falle der Verletzung zu verzerrten Kosten-Nutzen-Bewertungen führen können. Kritisiert werden vor allem die Annahme der Risikoneutralität der Individuen im Bezug auf ihre Lebensdauer und die Annahme der Additivität der individuellen Nutzwerte über die Zeit (Weinstein, Torrance et al. 2009; Schlander 2010). Erstere geht davon aus, dass der Anteil der Restlebenserwartung, den ein Individuum bereit wäre gegen eine Verbesserung der Lebensqualität einzutauschen, nicht von der Länge der Restlebenserwartung abhängig ist. Letztere besagt, dass die Bewertung eines Gesundheitszustands weder von vorausgehenden oder nachfolgenden

Gesundheitszuständen noch von der in den jeweiligen Zuständen verbrachten Zeit beeinflusst wird. Weitere Kritik am QALY als Grundlage für indikationsübergreifende Allokationsentscheidungen resultiert aus der nachgewiesenen Abhängigkeit der Nutzwerte von den angewandten Messinstrumenten und Ermittlungsverfahren und dem Fehlen eines international einheitlichen Erhebungsstandards (Weinstein, Torrance et al. 2009; Schlander 2010). Schließlich steht das QALY-Konzept als Grundlage für indikationsübergreifende Allokationsentscheidungen in der Kritik, weil es sich auf die Integration und Quantifizierung gesundheitlicher Konsequenzen beschränkt und dabei ethische Aspekte wie Gleichheit und Gerechtigkeit außer Acht lässt, obwohl diese in der Realität von großer Bedeutung für eine Allokationsentscheidung sind. Einige Studien weisen darauf hin, dass die Maximierung von QALYs kein vordringliches Ziel der Gesellschaft sei (Schlander 2010).

Auf der anderen Seite stellt das QALY-Konzept die derzeit einzige international umfassend eingesetzte und erprobte Methode dar, die explizit und mit einem theoretisch begründeten Konstrukt die Bewertung der Effektivität von Technologien ermöglicht, auch wenn diese auf mehrere bzw. unterschiedliche Endpunkte wirken (z. B. Mortalität und Lebensqualität, Lebensqualität bezüglich verschiedener Endpunkte). Damit ist das QALY-Konzept auch in vielen Fällen die Methode der Wahl bei Nutzen-Schadens-Abwägungen.

Die Effekte auf klinische Endpunkte, Nebenwirkungen, auf die Lebenserwartung und die gesundheitsbezogene Lebensqualität sollten aus Transparenzgründen auch disaggregiert dargestellt werden.

8.4.2.10 Ressourceneinsatz und Kosten

Die Perspektive bestimmt die Art der zu berücksichtigenden Kosten. Aus der Kostenträger-Perspektive werden alle direkten medizinischen Kosten erfasst, z. B. für Krankenhausaufenthalte, diagnostische Tests, Medikamente. Sind direkte nicht-medizinische Kosten relevant, z. B. Transportkosten, dann werden auch sie mit einbezogen. Aus der gesellschaftlichen Perspektive sind auch die Produktivitätsausfallskosten zu berücksichtigen. Bei Kosten-Effektivitäts-Analysen mit dem Maß Kosten pro gewonnenes Lebensjahr sind Produktivitätsverluste durch Mortalität bereits auf der Effektivitätsseite im Nenner des inkrementellen Kosten-Effektivitäts-Verhältnisses (Lebensjahre) enthalten. Auf der Kostenseite darf daher nur der Produktivitätsverlust durch Morbidität erfasst werden. Bei Kosten-Nutzwert-Analysen (Kosten pro gewonnenes QALY) sind in vielen Fällen die Gesundheitseffekte bereits auf der Nutzwertseite im Nenner des inkrementellen Kosten-Nutzwert-Verhältnisses enthalten und dürfen nicht nochmals bei den Kosten berücksichtigt werden (Doppelzählung). Da die mit einer Technologie verbundenen Effekte auf die Produktivität ein wichtiges Ergebnis darstel-

len, sind die Produktivitätskosten getrennt zu berichten. Der unscharfe Begriff „indirekte Kosten“ sollte vermieden werden (Gold, Siegel et al. 1996).

Im HTA-Bericht sollten die diskontierten und undiskontierten Gesamt- und inkrementellen Kosten beschrieben werden. Die wichtigen Kostenblöcke sollten disaggregiert berichtet werden. Ferner ist das diskontierte inkrementelle Kosten-Effektivitäts-Verhältnis (IKEV bzw. IKNV) anzugeben. Produktivitätsausfallskosten sind getrennt zu berichten.

Jedenfalls müssen die Kosten in einem Mengen-Preis-Gerüst nach Art des Ressourcenverbrauchs (Arzneimittel, Erstversorgung, Nachbehandlung, Rehabilitation usw.) und der Marginalbetrachtung (Kosten jeder zusätzlichen Einheit) dargestellt werden. Ist eine detaillierte Darstellung (Micro-Costing-Ansatz) nicht möglich, muss der Grund für die aggregierte Darstellung (Macro-Costing-Ansatz) genannt oder müssen die Durchschnittswerte angeführt werden.

Oft sind Daten für Kosten nur aus verschiedenen Jahren und Ländern vorhanden. In diesem Fall müssen die Zahlen an das Referenzjahr (meistens das Jahr der Erstellung) angepasst (inflationbereinigt) und gegebenenfalls mithilfe von Kaufkraftparität (KKP) (engl. purchasing power parity, PPP) in Euro umgerechnet werden.

8.4.2.11 Diskontierung

Wenn eine ökonomische Studie einen Zeitraum betrachtet, der ein Jahr übersteigt, sollten zukünftige Kosten und Nutzen in ihren Gegenwartswert umgerechnet (diskontiert) werden. Die Wahl des Diskontierungssatzes muss auf jeden Fall begründet werden und für Kosten und Nutzen gleich hoch sein. Diskontierungssätze liegen üblicherweise zwischen 0 Prozent und 5 Prozent. In vergleichbaren Ländern wird am häufigsten ein Diskontierungssatz von 3 Prozent gewählt, der auch für Österreich empfohlen wird. Die Wahl abweichender Diskontierungssätze muss stets ausführlich begründet werden.

Es besteht auch die Möglichkeit, den Diskontierungssatz an den aktuellen Kapitalmarktzinssatz anzupassen. Um die Auswirkungen der Diskontierung auf die Ergebnisse zu überprüfen, soll eine Sensitivitätsanalyse mit abweichenden Sätzen durchgeführt werden (0 %, 5 % und 10 %).

8.4.2.12 Variabilität und Unsicherheit

Für Variabilität und Unsicherheit gibt es vier zu unterscheidende Ursachen. Wir übernehmen im Folgenden die Terminologie, die die ISPOR-SMDM Good Modeling Practice Task Force in ihrem neuen (noch vorläufigen) Dokument empfiehlt. Zum einen kann es eine natürliche Heterogenität in der Zielpopulation geben, die durch Unterschiede in der Ausprägung von Merkmalen bzw. Variablen (z. B. Risikofaktoren) bei den Individuen entsteht und zu Unterschieden im Ergebnis führt. Die Unsicherheit ist hier erster Ordnung (engl. „first-order uncertainty“), d. h., diese Heterogenität wird – soweit sie bekannt ist und gemessen wurde – in Subgruppenanalysen untersucht. Dabei wird die Zielpopulation in annähernd homogene Gruppen aufgeteilt und die Analyse für jede Gruppe wiederholt.

Diese Heterogenität ist zu unterscheiden von der Unsicherheit ersten Grades, die in der zufälligen Variabilität der Ergebnisgrößen bei (zumindest augenscheinlich) identischen Individuen besteht. Die Ergebnisse streuen um einen Erwartungswert. Diese Unterschiede werden gelegentlich auf die ungeklärte Heterogenität der Patienten zurückgeführt.

Von herausragender Bedeutung für die entscheidungsanalytische Modellierung ist die Unsicherheit in Bezug auf die Parameterwerte (engl. „parameter uncertainty“ oder „second-order uncertainty“). Diese Parameterwerte, z. B. Wahrscheinlichkeiten, sind geschätzte Werte aus Studien mit endlichen Stichproben, die damit einen Zufallsfehler aufweisen, der häufig in den Quellenstudien auch in Form des 95-%-Intervalls berichtet wird. Eine Untersuchung der Auswirkungen dieser Unsicherheit auf die Modellergebnisse ist erforderlich, um die Unsicherheit der Gesamtergebnisse (IKEV oder IKNV) zu bestimmen. Zur Untersuchung dieser Unsicherheit sind deterministische und probabilistische Sensitivitätsanalysen geeignet. Die Durchführung von deterministischen Sensitivitätsanalysen der wichtigsten einzelnen Parameter und Parameterkombinationen ist wichtig, um die Robustheit der Ergebnisse zu beurteilen. Deterministische Mehrweg-Sensitivitätsanalysen können in der Regel nur für zwei oder drei Variablen gleichzeitig übersichtlich dargestellt werden. Will man mehrere oder alle Variablen in die Unsicherheitsanalyse einbeziehen, so kann eine probabilistische Sensitivitätsanalyse durchgeführt werden. Diese ersetzen jedoch nie deterministische Ein- und Zweiweg-Sensitivitätsanalysen.

Mittels probabilistischer Sensitivitätsanalyse(n) kann eine Aussage darüber gemacht werden, mit welcher Wahrscheinlichkeit bei der gegebenen Unsicherheit ein Kosten-Effektivitäts-Verhältnis oberhalb einer gewählten Grenze der Zahlungsbereitschaft liegt; die Darstellung der Ergebnisse in einem Scatterplot kann außerdem eine Vorstellung von der Streubreite des IKEV geben.

Zu empfehlen ist, in jedem Fall deterministische Einweganalysen durchzuführen und diese in Form eines Tornado-Diagramms darzustellen (CADTH Guidelines). Sollte der HTA-Bericht auch die Analyse der gesamten Unsicherheit als Ziel haben, so ist auch eine probabilistische Sensitivitätsanalyse durchzuführen. Bei allen Analysen ist darauf zu achten, dass die Vorinformation über Fehlerintervalle und Korrelationen zwischen Parametern berücksichtigt werden. Wenn nicht alle Parameter in die Sensitivitätsanalyse einbezogen werden, sollte die Auswahl gut begründet sein.

Eine weitere Art der Unsicherheit liegt in der Struktur des Modells. Die strukturelle Unsicherheit eines Modells kann schwierig zu untersuchen sein. In einigen Fällen, z. B. bei Annahmen über den Verlauf des Behandlungseffekts nach Ende der Behandlung, lassen sich strukturelle Annahmen innerhalb des Modells durch Verwendung verschiedener Funktionsverläufe nachbilden und mithilfe von Sensitivitätsanalysen untersuchen. Sofern möglich, sollte dies geschehen. In anderen Fällen würde eine Untersuchung der strukturellen Annahmen einen Umbau des Modells erfordern, was aus Gründen des Arbeitsaufwands oft unmöglich ist. In diesem Fall sollten die strukturellen Annahmen wie alle anderen Annahmen und Limitationen klar beschrieben werden.

8.4.2.13 Verteilungsgerechtigkeit

Die Ergebnisse von ökonomischen Studien können je nach Zielpopulation variieren. So können bei einer Diagnose unterschiedliche Altersgruppen verschieden hohe Kosten verursachen. Um den Aspekt der Verteilungsgerechtigkeit zu berücksichtigen, sollen Ergebnisse nicht nur aggregiert, sondern nach allen relevanten Gruppen aufgeschlüsselt dargestellt werden. Diese Gruppen können beispielsweise nach Alter, Geschlecht, sozialem Status, geographischer Lage usw. oder aber nach Krankheitsstadium und klinischen Merkmalen differenziert werden. Kann eine Studie nicht auf Aspekte der Verteilungsgerechtigkeit eingehen, muss der Grund dafür erläutert werden.

8.4.2.14 Verallgemeinerbarkeit und Übertragbarkeit der Ergebnisse

In vielen Fällen beziehen sich ökonomische Studien nicht auf die betroffene Gesamtpopulation, sondern unterscheiden nach sozioökonomischen und demografischen Faktoren. In solchen Fällen muss immer festgehalten werden, ob Parameter wie Kosten und Wirksamkeiten bzw. die Ergebnisse verallgemeinert und übertragen werden können. Ein Hinweis auf die Verallgemeinerbarkeit und Übertragbarkeit der Ergebnisse sollte in jeder Studie erwähnt und ausreichend diskutiert werden.

8.4.2.15 Bericht und Publikation der Ergebnisse

Ökonomische Studien sollen wie alle anderen wissenschaftlichen Arbeiten unabhängig von ihren Ergebnissen und Auftraggebern publiziert, der Öffentlichkeit zugänglich gemacht und verbreitet werden. Eine Veröffentlichung in einem Fachjournal ist wünschenswert, aber nicht zwingend notwendig. Jedenfalls soll eine Kurzfassung (Executive Summary oder Abstract) auch in englischer Sprache verfasst werden.

Im Sinne größtmöglicher Transparenz kann – falls es der Auftraggeber wünscht – auch bei ökonomischen Studien ein öffentliches Stellungnahmeverfahren durchgeführt werden. Wenn kein Stellungnahmeverfahren durchgeführt wird, bekommt der Auftraggeber vor Publikation die Möglichkeit, eine Stellungnahme zu den Ergebnissen abzugeben (siehe Prozesshandbuch Version 1.2012, Kapitel 2).

Für die Publikation ist darauf zu achten, dass eine lückenlose Darstellung aller Schritte von der Erläuterung der Methodik bis zur Präsentation der Ergebnisse sichergestellt ist. Dabei sollen alle Daten so dargestellt werden, dass die Ergebnisse nachvollziehbar und nachprüfbar sind. Um eine umfangreiche Darstellung nicht auf Kosten der Lesefreundlichkeit zu erreichen, können Tabellen, Grafiken usw. im Anhang oder auf einer Webseite dargestellt werden. Sinnvoll ist die Angabe von Kontaktdaten der Autorinnen/Autoren, die bei Anfrage Daten zur Verfügung stellen können.

Autorinnen/Autoren: Johannes Zsifkovits, Barbara Fröschl (GÖG), Annette Conrads-Frank, Nikolai Mühlberger, Petra Schnell-Inderst, Gaby Sroczynski, Uwe Siebert (UMIT)

9 Psychosoziale, ethische, rechtliche und organisatorische Aspekte

Obwohl die Definition von HTA neben der Bewertung der medizinischen Wirksamkeit, der Sicherheit und der ökonomischen Auswirkungen auch die Berücksichtigung sozialer, organisatorischer, ethischer oder rechtlicher Aspekte beinhaltet (im Folgenden „soziokulturelle“ Aspekte genannt), kommt diesen Bereichen in Technologiebewertungen zumeist deutlich weniger Aufmerksamkeit zu. Diese soziokulturellen Faktoren können aber einen wesentlichen Beitrag zu den Fragen leisten, ob eine Technologie oder eine medizinische Intervention die unter idealen Studienbedingungen gezeigte Wirksamkeit auch unter realen Alltagsbedingungen entfaltet, welche die organisatorischen Grundvoraussetzungen sind für eine optimale Wirksamkeit oder welche psychosozialen oder ethischen Hindernisse der Entfaltung der Wirksamkeit entgegenstehen. Darüber hinaus verändern diese Aspekte – auch Kontextinformationen genannt – die Perspektive auf die Technologie und können neben dem rein medizinischen Blickwinkel auch eine breitere Betrachtungsweise der Intervention bieten.

Die Ursachen für die seltenere Berücksichtigung soziokultureller Aspekte (Gerhardus/Perleth 2008) liegen

- » zum einen in dem (Irr-)Glauben oder Missverständnis, dass diese Aspekte nicht systematisch in die Analyse einbezogen werden können und sich einer objektiven Synthese von Informationen entziehen und daher immer subjektiv bleiben müssen,
- » zum anderen in der tatsächlich geringeren Erkenntnisbasis, d. h. in dem nur lückenhaft zur Verfügung stehenden Wissen. Dieser Mangel macht aufwendige Primärerhebungen (Interviews, teilnehmende Beobachtung, partizipative Verfahren etc.) notwendig und lässt – unter der Bedingung begrenzter zeitlicher und materieller Ressourcen – die Berücksichtigung soziokultureller Aspekte zumeist zu kurz kommen.
- » Drittens aber sind die Fragestellungen – etwa für Refundierungsentscheidungen – häufig auf die medizinische Wirksamkeit und die gesundheitsökonomischen Auswirkungen beschränkt, und die Berücksichtigung soziokultureller Aspekte ist nicht gefragt, nicht erwünscht, zuweilen auch nicht möglich. Sie haben eine „nachgeordnete Entscheidungsrelevanz“ (Gerhardus/Stich 2008).

Obwohl Konsens darüber besteht, dass die Berücksichtigung von soziokulturellen Aspekten (vgl. Tabelle 9.1) in HTA von großer Bedeutung sein kann, liegt keine einheitliche Methodik zur Durchführung vor. Es werden unterschiedliche Herangehensweisen akzeptiert, denen allen der Grundsatz der Transparenz und Nachvollziehbarkeit der Analysen zugrunde liegt (Lühmann/Raspe 2008, Droste 2008, Busse 2001). Kein Konsens besteht hingegen, inwieweit diese Art von Information tatsächlich auch

systematisch aufgearbeitet werden soll. Viele HTA-Berichte enthalten Hinweise auf psychosoziale Problemlagen, rechtliche oder organisatorische Aspekte im Einleitungs- bzw. im Diskussionsteil, wobei dann aber nicht klar wird, inwieweit zu ihrer Gewinnung ein systematischer methodischer Ansatz zu Anwendung kam (Lühmann et al. 2007). Bei systematischem Vorgehen (vgl. dazu Kapitel 5 des Methodenhandbuchs) gilt das Prozedere wie bei anderen Aspekten der evidenzbasierten Medizin: Definition der Fragestellung, Auswahl relevanter Datenbanken, Festlegen einer (manchmal auch iterativen) Suchstrategie, Recherche, systematisches und transparentes Zusammenfassen von bestehendem Wissen.

Dieses Kapitel soll die Berücksichtigung soziokultureller Aspekte in HTA-Berichten unterstützen und dazu beitragen, deren objektive und systematische Darstellung zu erleichtern. Für alle Bereiche gilt, dass die Inhalte einer Analyse von psychosozialen, ethischen, rechtlichen und organisatorischen Aspekten einer medizinischen Technologie völlig unabhängig vom methodischen Vorgehen einer systematischen oder unsystematischen Wissensgenerierung sind.

Tabelle 9.1 soll die „breite“ Sichtweise von HTA und die Möglichkeiten eines systematischen Vorgehens vor Augen führen. Die Betrachtung soziokultureller Aspekte beinhaltet typischerweise die Perspektiven unterschiedlicher Stakeholder, und die betrachteten Zielgrößen sind spezifisch für die jeweilige Dimension. Am Beispiel Gendiagnostik kann das systematische Durchdenken der Perspektiven und Dimensionen etwa folgendermaßen aussehen:

- a) Psychosozial: Auswirkungen auf soziale Integration und Interaktionen einschließlich der determinierenden psychologischen Faktoren;
- b) Ethik: Vereinbarkeit mit allgemein akzeptierten Grundwerten;
- c) Recht: Kompatibilität mit rechtlichen Vorschriften;
- d) Organisatorisch: Einpassung in die Gegebenheiten des Versorgungssystems;

Tabelle 9.1 ist ein Hilfsmittel zur Unterstützung systematischen Herangehens und weniger als tatsächlicher Bestandteil in einem HTA gedacht.

Tabelle 9.1:

Framework zur systematischen Analyse soziokultureller Implikationen
medizinischer Technologien (beispielhaft)

Interessengruppen/ Stakeholder	Psychosoziale Aspekte	Ethische Aspekte	Rechtliche Aspekte	Organisatorische Aspekte
Patientinnen/Patienten	Compliance/ Adherence, Akzeptanz Zufriedenheit	Grenzen gesund/krank Diskriminierung	Informierte Entscheidung Versicherungs- schutz	Wartezeiten Aufenthaltsdauer
Angehörige	Akzeptanz	Recht auf Wissen/ Nicht- wissen in Generationen	Informierte Entscheidung	Wartezeiten Aufenthaltsdauer
Dienstleister/ medizinische & pflegerische Professo- nen	Akzeptanz Präferenzen	Berufsethos	Aufklärungs- pflicht	Ausbildung, Qualifikationsbe- darf Praxisvaria- bilitäten
Hersteller/ Marktanbieter/in Cluster von Technolo- gieentwicklern		Ethik der Markt- bearbeitung	Haftungsrecht	Praxisvariabilitäten Marktbearbeitung
Kostenträger		Angemessenheit, Bedarfs- gerechtigkeit	Leistungsrecht	Bedarf vs. Nachfrage Kapazitäts- und Investitionsplanung
Gesellschaft		Verteilungs- gerechtigkeit	Zulassungs- recht	Mindestmengen Standortfragen

Quelle: LBI-HTA; Darstellung: GÖG/BIQG

9.1 Psychosoziale Aspekte

Analog zur systematischen Analyse organisatorischer Aspekte, die Organisationseinheiten ins Zentrum der Betrachtung stellen, stellt die Analyse psychosozialer Aspekte das Individuum, alle Beteiligten und Anwenderinnen/Anwender (Patientinnen/Patienten, Angehörige, Angehörige der gehobenen medizinisch-technischen Dienste, Ärztinnen/Ärzte, Pflegepersonal etc.) ins Zentrum der Untersuchung, um Aussagen zu Implikationen von Technologien zu machen. Die Wahrnehmung einer Technologie (etwa Akzeptanz oder Widerstand gegen eine medizinische Technologie) hat laut Sozialwissenschaften einen wesentlichen Einfluss auf deren Anwendung und damit auf die Entfaltung der Wirksamkeit von Technologien.

Die Methodik (EUnetHTA 2008), psychosoziale Aspekte der Wirkungen von medizinischen Technologien systematisch zu evaluieren, basiert auf dem Konzept der

- » sozialen Funktionsanalyse: Die Betroffenen (Anwender/innen und Nutzer/innen einer medizinischen Technologie) werden in ihren sozialen Rollen betrachtet (im Berufsleben, im Familienleben, in der Freizeit etc.), und analysiert werden die Auswirkungen der Technologie auf das Funktionieren in den Rollen. Direkte und indirekte ökonomische Auswirkungen (etwa Verlust von Einkommen, lange Wegzeiten), individuelle psychische Implikationen (etwa Angst vor Schmerz, Angst, die Partnerschaft zu belasten), ethische und politische Auswirkungen (etwa Erwartung von Stigmatisierung, Diskriminierung), physische und biologische Wirkungen (etwa Erwartung von Behinderung, Bettlägerigkeit), Kommunikation und soziale Beziehungen (etwa Hoffnung auf Anerkennung, verändertes Berufsbild) stehen im Zentrum der systematischen Analyse zu möglichen Wirkungen, Hindernissen oder Potenzialen von medizinischen Interventionen.
- » Als pragmatische Herangehensweise bei der Analyse psychosozialer Implikationen einer medizinischen Technologie stehen primär qualitative Methoden zur Erhebung der Wahrnehmung der Anwender/innen und Patientinnen/Patienten (Interviews, teilnehmende Beobachtung oder aber auch Literaturanalyse) zur Verfügung oder – sofern vorhanden – die Auswertung von Sekundärliteratur.

Die Analyse psychosozialer Aspekte einer medizinischen Technologie betreffen folgende Bereiche und werfen beispielsweise folgende Fragen auf:

Akzeptanz, „Compliance“/Therapietreue, Adhärenz

- » Was behindert die Therapietreue/Compliance und Adhärenz bei Patientinnen und Patienten: Therapieschemata, Nebenwirkungen, Angst vor sozialer Stigmatisierung?
- » Was behindert die Akzeptanz einer (neuen) Technologie unter den Anwenderinnen und Anwendern: konfligierende Interessen, mangelnde Refundierung etc.?

Zufriedenheit, Präferenzen

- » Was sind die Hoffnungen und Erwartungen an die medizinische Technologie? Was sind die Charakteristika, die einer Technologie gegenüber einer anderen den Vorzug geben: geringere Wartezeit, mehr Lebensqualität, besseres Funktionieren im Alltag, mehr Kommunikation und Verständnis?
- » Welche Faktoren bedingen die Zufriedenheit/Unzufriedenheit mit der medizinischen Technologie?

Die ethischen, rechtlichen, organisatorischen und psychosozialen Fragestellungen überschneiden sich zum Teil und sind nicht immer leicht getrennt zu bearbeiten. Ein systematisches Herangehen verhindert aber, dass wesentliche Aspekte, die zur Wirksamkeit einer medizinischen Technologie beitragen, vergessen werden. Die HTA-Fragestellung und die Auftraggeberin / der Auftraggeber eines Assessments bestimm-

men letztendlich, ob eine enge Perspektive oder eine wie in diesem Kapitel 9 beschriebene breite Sichtweise auf die zu bewertende Technologie eingenommen wird bzw. welche Aspekte besondere Aufmerksamkeit erfahren.

9.2 Ethische Aspekte

Ethische Aspekte spielen insbesondere bei der Bewertung solcher medizinischer Technologien eine Rolle, die informierte Entscheidungen in stark von Werthaltungen geprägten Versorgungssituationen erforderlich machen. Hierzu gehören z. B. die „end-of-life decisions“ und die prädiktive Medizin mit Technologien und Gebieten wie Gen- und Pränataldiagnostik, Screeningverfahren, Onkologie und Transplantationsmedizin. Ebenso sind aber auch medizinethische vs. Public-Health-ethische Fragen in der Bewertung von Technologien von Bedeutung, die ressourcenintensiv sind und daher auch Fragen der Verteilungsgerechtigkeit aufwerfen (Wild 2005, Wild/Zechmeister 2007).

In die Analyse ethischer Aspekte fließt Wissen zu prävalenten gesellschaftlichen Werthaltungen, Normen und moralischen Prinzipien ein, die soziale Geltung haben und die den Kontext für die Anwendung medizinischer Technologien bilden. Vorherrschende kulturelle (z. B. religiöse) und ökonomische Bedingungen (z. B. Bruttoinlandsprodukt und Ausgaben für die Gesundheitsversorgung) sind wesentliche Ursachen für z. B. länderspezifische, aber auch regionale oder populationsspezifische Unterschiede in der Verwendung medizinischer Verfahren.

Eine multidimensionale Bewertung, wie sie in dem pragmatischen Herangehen der Technikfolgenabschätzung (TA) geschieht, hat eine Bestandsaufnahme der verschiedenen Regelwerke, die Einfluss auf die Anwendung medizinischer Technologien haben, zum Inhalt. In einer herkömmlichen Technikfolgenabschätzung können – je nach Fragestellung – Normen (Gesetzestexte wie z. B. das Krankenanstaltengesetz) und Praxisregeln (Anwendungsrichtlinien oder Standard Operating Procedures) ebenso analysiert werden wie Dokumente zu Erwartungen, Befürchtungen und Meinungen von Stakeholdern zur Technologie oder Zielsetzungen und Visionen, die mithilfe der medizinischen Intervention erfüllt werden sollen. Es werden also alle die Technologie beeinflussenden Faktoren identifiziert und auch normative Widersprüchlichkeiten benannt.

Die Methodiken (INAHTA 2003, EUnetHTA 2008, Lühmann/Raspe 2008) und die konzeptionellen Ansätze, um verschiedene ethische Aspekte in die jeweilige Fragestellung zu integrieren, können unterschiedlich sein.

- » Kasuistiken/Präzedenzmethode: Anhand von Einzelfällen werden „typische“ Problematiken modellhaft beschrieben und die zugrunde liegenden ethischen Prinzipien offengelegt. Das Herangehen ist ein mehrstufiges Prozedere: Ein „Fall“ wird ausgewählt und einer Kategorie zugeordnet (Beispiel 1: Verteilung rarer Ressourcen – Intensivbetten, Organe; Beispiel 2: Prognostik unter Wahrscheinlichkeiten – Gendiagnostik, Polymorphismenscreening) und standardisiert beschrieben. Danach folgt die Analyse der zugrunde liegenden Maximen/Regeln. Zuletzt wird dieser Fall mit weiteren ähnlichen Fällen und zugrunde liegenden Regelwerken verglichen, und die gesellschaftlichen Werthaltungen, Normen und moralischen Prinzipien werden dargestellt.
- » Problemzentrierter Ansatz: Der Analyse von Dokumenten geht eine systematische Identifikation und Formulierung von relevanten ethischen Aspekten und Kernfragen voraus; fokussierter als der pragmatische Ansatz der Literaturerhebung.
- » Interaktive, partizipative Methode: In einem Diskurs werden die Werthaltungen, Erwartungen, Bedürfnisse etc. aller Beteiligten und Interessengruppen, also Patientinnen/Patienten, Ärztinnen/Ärzte, Pflegekräfte, Marktanbieter/in etc., erhoben und auf konfligierende Werthaltungen analysiert. Ziel ist die Identifikation der relevanten Parameter, die in die Beurteilung einer medizinischen Technologie einfließen.
- » Prinzipalismus: Die Analyse erfolgt in Anwendung der vier fundamentalen ethischen Dimensionen oder Prinzipien (Autonomie – Fürsorge – Nichtschädigung – Gerechtigkeit) als analytischem „Framework“. Diese Prinzipien müssen dabei der Fragestellung entsprechend konkretisiert werden.

Sehr häufig kommen in HTA-Berichten Kombinationen von verschiedenen methodischen Herangehensweisen zum Einsatz, wobei die in Publikationen gefundenen Werthaltungen und Normen (pragmatischer Ansatz) anhand konkreter Beispiele in einem analytischen Framework (Prinzipalismus) dargestellt werden. Diese Herangehensweise ist die Mindestanforderung und kann zur Vervollständigung – bei entsprechendem zeitlichen Ressourceneinsatz – in Betroffenenengruppen diskutiert werden (interaktive, partizipative Methode).

Datenbanken und Quellen zum Recherchieren ethischer Implikationen medizinischer Technologien finden sich bei Droste 2008.

Als Systematik der Darstellung wird eine „perspektivensensitive“ zweidimensionale Matrix vorgeschlagen, die zum einen die relevanten Interessenvertreter (Patientinnen/Patienten, Angehörige, Professionen, Hersteller/Anbieter, Kostenträger, Gesellschaft), zum anderen die möglichen Werthaltungen und ethischen Kernfragen aus der jeweiligen Perspektive abbildet.

Ziel in herkömmlichen HTA-Berichten ist die systematische Darlegung von Wertekonflikten, nicht aber deren Lösung. Erst in einem moderierten interaktiven weiteren Arbeitsschritt in sog. „partizipativen Technikfolgenabschätzungen“ in Form von „Citizen Councils“, Bürgerforen, Konsensuskonferenzen (Gerhardus/Stich 2008) etc. sollen auch Lösungen erarbeitet werden. Diese Methodiken eignen sich dazu, in gesellschaftlichen Priorisierungsdebatten als Basis für grundlegende Rationierungsentscheidungen eingesetzt zu werden.

Ethische Analysen setzen sich mit dem normativen Wertegerüst sowie mit individual- und sozialetischen Problembereichen auseinander. Beispielhaft sind folgende Bereiche betroffen:

Verteilungsgerechtigkeit

- » Bindet der Einsatz der Technologie Ressourcen, die anderswo fehlen?
- » Ist der Zugang zur Technologie durch soziale Barrieren behindert?
- » Hat der Einsatz der Technologie Auswirkungen auf die Beziehung zwischen den Generationen?

Wertkonformität, –konfliktbereiche & Werteveränderungen

- » Gibt es unterschiedliche Werthaltungen, Perspektiven unter den Betroffenen/Beteiligten/Interessengruppen?
- » Welches sind die Konfliktbereiche, und was sind die möglichen Konsequenzen der Konfliktbereiche?
- » Hat die Technologie einen Einfluss auf die Beziehung zwischen den Betroffenen/Beteiligten/Interessengruppen (Arzt/Ärztin–Patient/in, Angehörige/r–Patient/in etc.)?
- » Kommt es zu nachhaltigen Werteveränderungen aufgrund des Einsatzes der Technologie? (Verschiebung der Definition gesund/krank, Diskriminierung am Arbeitsmarkt, Betroffenheit vulnerabler sozialer Gruppen, Veränderung sozialer Rollenverteilungen)

Autonomie, Würde, Integrität, Bedarfsgerechtigkeit

- » Beeinträchtigt der Einsatz der Technologie die Autonomie und Selbstbestimmung, die Integrität oder Würde der Patientinnen und Patienten?
- » Sind mit dem Einsatz der Technologie Risiken verbunden, über die die Patientinnen/Patienten informiert werden müssen?

- » Sind mit dem Einsatz der Technologie nicht beabsichtigte Folgen auf berufliche Rollen oder tradierte Werte zu erwarten?

9.3 Rechtliche Aspekte

Auch rechtliche Aspekte werden in HTA-Berichten selten systematisch berücksichtigt. Die methodischen und konzeptionellen Ansätze zur Bewertung von rechtlichen Implikationen medizinischer Technologien sind analog zu der Analyse ethischer Aspekte durchführbar: Der systematischen Recherche folgt die Analyse von Dokumenten nach Rechtsfragestellungen.

Die Herangehensweisen sind:

- » Darlegung rechtlicher Vorschriften,
- » Kasuistiken/Präzedenzmethode: Analyse von Einzelfällen und Schlussfolgerungen daraus,
- » problemzentrierter Ansatz: systematische Erhebung von relevanten legistischen Aspekten und Kernfragen,
- » Interessengruppen: systematische Analyse von legistischen Regelwerken der Beteiligtegruppen (etwa Berufskodizes),
- » Prinzipalismus: Analyse der Einhaltung der Menschenrechte (z. B. Autonomie).

Rechtliche Aspekte betreffen die Steuerung und Regulierung medizinischer Technologien und stellen damit die Rahmenbedingungen für deren Einsatz dar (Francke 2008). Rechtsfragen lassen sich nach vier Bereichen kategorisieren (EUnetHTA 2008), die Gewichtung ist dabei von der jeweiligen Technologie abhängig:

1. Fragestellungen, die Patientinnen/Patienten betreffen: grundlegende Rechte wie das Selbstbestimmungsrecht, das Recht auf informierte Entscheidung und ärztliche Aufklärung, Recht auf würdevolle Behandlung, Recht auf vertrauliche Behandlung von medizinischen Informationen (Privacy und Confidentiality), aber auch Versicherungsschutz, Berufspflichten etc.;
2. Fragestellungen, die die Technologie betreffen: z. B. Patentrecht, Arzneimittelrecht, Medizinprodukterecht, Sozialversicherungs- und Leistungsrecht, Produkthaftung etc.;
3. Fragestellungen, die den Prozess der Anwendung der Technologie betreffen (z. B. Berufsrechte und -pflichten, Ärztegesetz, Ethikkommissionen, Richtlinien, länderspezifische Rechtsnormen etc.);

4. Fragestellungen, die die Gesundheitspolitik auf lokaler, nationaler und EU-Ebene betreffen.

Die Inhalte einer Analyse rechtlicher Aspekte einer medizinischen Technologie betreffen folgende Bereiche und werfen etwa folgende Fragen auf (beispielhaft):

Autonomie, informierte Entscheidung

- » Wird die Patientin / der Patient über mögliche therapeutische Optionen, deren Wirksamkeitswahrscheinlichkeiten und Risiken aufgeklärt?
- » Wird die Technologie in Patientenverfügungen geregelt?

Daten- und Versicherungsschutz

- » Hat die Anwendung der Technologie nicht nur auf das Individuum, sondern auch auf Folgegenerationen rechtliche Auswirkungen?
- » Wer hat Zugang zu heiklen medizinischen Daten von Patientinnen/Patienten, und welchen Einfluss hat das auf den Versicherungsschutz?

Rechtskonformität (Sozial-, Produktrecht, Leistungsrecht)

- » Hat die Technologie eine Zulassung?
- » In welchen Ländern wird sie auch refundiert?
- » Handelt es sich um eine experimentelle Technologie, die nur unter Forschungsbedingungen Anwendung finden darf?
- » Entspricht die Anwendung der Technologie den Rechtsnormen (off-label, compassionate use, offene Haftungsfragen, Ausnahmebestimmungen)?
- » Liegen leistungsrechtliche Bestimmungen, die die Inanspruchnahme beeinflussen, vor? Welche rechtlich relevanten Perspektiven nehmen verschiedene Interessengruppen ein?
- » Wie wurde in Präzedenzfällen entschieden?
- » Bestehen in der EU unterschiedliche rechtliche Regelungen, die zu „Medizin-Tourismus“ führen?

9.4 Organisatorische Aspekte

Die systematische Analyse organisatorischer Aspekte bezweckt Aussagen zu Implikationen von Technologien auf Organisationen und Institutionen, aber auch vice versa zu Wirkung von Organisationen und Institutionen auf die Anwendung medizinischer

Technologien. Die wachsende Aufmerksamkeit in HTA-Berichten für organisatorische Aspekte entspricht der Einsicht, dass das organisatorische Umfeld und die Anwendungsbedingungen einen wesentlichen Beitrag zur Entfaltung der Wirksamkeit von Technologien haben. Dennoch haben sich international weder eine einheitliche Definition von „Organisation“ noch methodische Standards herausgebildet, nach denen organisatorische Auswirkungen durch medizinische Technologien analysiert werden. Die Ursache dafür dürfte zum einen in der Komplexität von Organisationen, zum anderen aber auch in deren Kontextabhängigkeit und damit in der geringen Übertragbarkeit der Ergebnisse in andere (Medizin-)Kulturen liegen.

Unter Organisation wird gemeinhin eine koordinierte soziale Einheit verstanden, die durch gemeinsame Aktivitäten und Zielsetzungen definiert ist. Faktoren, die eine Organisation ausmachen, sind etwa physische Strukturen, soziale Beziehungen, Mechanismen der Interaktion und Koordination und eine Organisationskultur (EUnetHTA 2008, Perleth 2008).

Die Methodiken, organisatorische Aspekte der Wirkungen von medizinischen Technologien systematisch zu evaluieren, basieren letztendlich auf zwei Konzeptionen (EUnetHTA 2008, Perleth 2008). Beide Ansätze gehen davon aus, dass „Organisationen“ die Anwendung einer Technologie beeinflussen und diese auch gestalten können, dass also die Wirksamkeit einer Technologie durch sogenannte „Ko-Produktion“ zustande kommt. Organisatorische Aspekte betreffen somit die aufgewendeten Ressourcen, die eingesetzten Prozesse und ihre Interdependenzen.

- » Translationsmodell: Technologien werden als endogener Bestandteil von Organisationen und Anwendern gesehen und betrachtet. Die Forschungsfragen im Translationsmodell sind, welche Ressourcen (Personal, Zeit, Material), Strukturen und Prozesse mobilisiert werden müssen, um die gewünschten Ergebnisse der Anwendung einer medizinischen Technologie zu erzielen. Damit ist die Annahme verbunden, dass medizinische Technologien nicht per se in jeglichem Umfeld wirksam sind. Das Translationsmodell als Konzept zur Analyse organisatorischer Auswirkungen steht in Kontrast zum „linearen Diffusionsmodell“, das aber inzwischen als obsolet gilt: Technologien werden hier als ein exogener zusätzlicher Bestandteil gesehen, die außerhalb der Organisation stehen und „uni-direktional“ wirken.
- » Input-Output-Modell: Input, Prozesse und Output werden anhand von Indikatoren auf verschiedenen Ebenen analysiert (Mikro-Ebene: Individuen/Anwender; Meso-Ebene: Organisationseinheiten: Stationen, Kliniken, Krankenhäuser; Makro-Ebene: Regionen, Nationen, internationale Systeme). Zum Input zählen etwa Gesundheitsstand der Bevölkerung und damit Bedarf, Strukturen und anbieterinduzierte Nachfrage sowie finanzielle Ressourcen. Prozessindikatoren sind Krankenhausaufenthalte, Liegezeiten, Leistungsdichte. Outputindikatoren sind klinische Ergebnisse, Komplikationsraten, Patientenzufriedenheit oder prozedurbedingte Effekte.

- » Als pragmatische Herangehensweise in der Analyse der Implikationen einer medizinischen Technologie bietet sich daher eine Prozessanalyse an, die die Arbeitsabläufe betrachtet, den eingesetzten materiellen Aufwand, das beteiligte Personal und deren benötigte Fachkenntnisse und Qualifikationen, die Kooperationen ebenso wie Kommunikationsschemata und -kulturen. Aber auch Fragen des Standortes und der Zugänglichkeit (De-/Zentralisierung), Fragen zu Hierarchien von Entscheidungen (Management) und zu Anwendungsmustern (Praxisvariabilität) sowie zu Verschiebungen von Kompetenzen und Verantwortlichkeiten aufgrund einer Technologie sind Teil der Analyse organisatorischer Einflussfaktoren. Diese Analyse kann erfolgen über Datenerhebungen (Ist-Analysen zu Strukturdaten, Leistungsdaten, Wartelistendaten; vgl. dazu Kapitel 5), quantitative Controllingdaten, qualitative Methoden wie Interviews, Inhaltsanalysen und teilnehmende Beobachtung.

Sehr häufig kommen in HTA-Berichten Kombinationen von verschiedenen methodischen Herangehensweisen zum Einsatz (Mathis 2009), die etwa den Bedarf nach medizinischen Interventionen aus internationalen epidemiologischen Daten generieren und den nationalen Nachfragedaten (d. h. den Leistungsdaten) gegenüberstellen. Es muss hier nicht eigens ausgeführt werden, dass die Analyse organisatorischer (Standortfragen, Mindestmengen, regionale Zugänglichkeit, Wartezeiten etc.) und psychosozialer Aspekte (Akzeptanz, Zufriedenheit, Kommunikation) eng miteinander verbunden ist.

Die Analyse von organisatorischen Aspekten einer medizinischen Technologie betrifft folgende Bereiche und wirft beispielsweise folgende Fragen auf:

Nutzungsgrad, Auslastung und Wartezeiten

- » Wodurch wird die Nachfrage nach einer Technologie beeinflusst (Zuweisungsrichtlinien, exakte Indikationsstellungen)? Hat die Technologie Einfluss auf Veränderungen bei Wartezeiten? Liegen administrative Zugangsbarrieren vor?
- » Wer profitiert von der Einführung der Technologie? Welche Interessenkonflikte treten auf?

Standort, Lokalisation, Zugang

- » Welche Auswirkung hat die medizinische Technologie auf regionale De-/Zentralisierung (z. B. durch Mindestmengenregelungen)?
- » Kommt es zu Kompetenzverschiebungen innerhalb einer Institution infolge (des Einsatzes) von (neuen) Technologien?
- » Welche Auswirkung hat die medizinische Technologie auf Zugänglichkeit (z. B. niedrigschwellige Dienstleistungen)?

Aufenthaltsdauer

- » Kommt es zu einer Veränderung der Dauer des Krankenhausaufenthaltes oder zu einer Verschiebung der Patienten und Patientinnen in einen anderen Sektor im Gesundheitssystem?
- » Gibt es große Anwendungsunterschiede bezüglich Häufigkeiten in den Regionen? Worauf sind diese zurückzuführen (Anwendungsrichtlinien, Auslastungsnotwendigkeiten)?

Infrastruktur und Personal, Mindestmengen, Ausbildung und Qualitätssicherung

- » Welche Voraussetzungen bezüglich Anforderungen (an Personal, Verfahrensabläufe, Infrastruktur) bestehen, um das gewünschte Behandlungsergebnis zu erhalten?
- » Gibt es Kommunikationskulturen und Kooperationsformen, die mit geringerem Aufwand zu den gewünschten Ergebnissen führen? Gibt es Kommunikationskulturen und Kooperationsformen, unter denen häufiger „Fehler“ (unerwünschte Ereignisse) stattfinden?

Autorin: Claudia Wild (LBI-HTA)

10 Qualitätssicherung

Um eine hohe wissenschaftliche Qualität der HTA-Produkte zu gewährleisten, werden vor, während und nach dem gesamten Erstellungsprozess der HTA-Produkte Strategien der Qualitätssicherung angewandt. Viele Qualitätssicherungsstrategien wurden bereits in vorangegangenen Kapiteln behandelt. Diese umfassen das Festlegen von Prozessen und Methoden der Themenfindung und Priorisierung (vgl. Kapitel 3), das Entwickeln der Fragestellung nach dem PICO-Prinzip (vgl. Kapitel 4.1), das Definieren von Selektionskriterien (vgl. Kapitel 4.2), das Durchführen einer systematischen Literatursuche inkl. der Entwicklung einer Suchstrategie sowie Überprüfung der Validität der Suchstrategie, das Überprüfen von ergänzenden Suchen wie Handsuchen, das Suchen nach nicht publizierter (grauer) Literatur bzw. das Identifizieren von Literatur durch Expertinnen und Experten (vgl. Kapitel 4.5), die Literaturselektion (vgl. Kapitel 4.6), das Überprüfen der internen und externen Validität von medizinischen und ökonomischen Studien sowie das Anwenden von Datenextraktions- und Qualitätsbewertungstabellen (vgl. Kapitel 6.3).

Darüber hinaus können ein Lektorat, das interne und externe Begutachtungsverfahren sowie das schriftliche Stellungnahmeverfahren zur Qualitätssicherung beitragen. Diese Maßnahmen können durchgeführt werden, wenn es die auftraggebende Institution bzw. die Auftraggeberin / der Auftraggeber des HTA-Berichts, Rapid Assessments bzw. Updates zu einem bestehenden Bericht wünscht. Für die hierbei zu setzenden Prozessschritte kann das GÖG/BIGQ-Prozesshandbuch Version 1.2012 herangezogen werden.

In jedem Fall wird die auftraggebende Institution bzw. die Auftraggeberin / der Auftraggeber des HTA-Berichts, Rapid Assessment bzw. Update zu einem bestehenden Bericht im entsprechenden Bericht namentlich als solche/r genannt.

Autorin: Daniela Pertl (GÖG/BIGQ)

11 Implementierung und Impact

Eine vordringliche Aufgabe von HTA-Produkten ist das fachgerechte Aufarbeiten und Bereitstellen wissenschaftlicher Evidenz in Form einer Entscheidungshilfe für gesundheitspolitische Entscheidungsträger. Die Relevanz von HTA-Produkten bemisst sich somit vorrangig daran, in welchem Ausmaß HTA-Produkte im Rahmen gesundheitspolitischer Entscheidungsprozesse berücksichtigt werden und welchen Beitrag HTA-Produkte langfristig zur Verbesserung der Gesundheitsversorgung leisten können.

11.1 Implementierungsvorschläge

Unter Implementierung wird die konkrete Umsetzung von HTA-Ergebnissen verstanden (Drummond et al. 2000). Die beabsichtigte Verwertung der HTA-Ergebnisse sollte nach Möglichkeit bereits am Beginn der HTA-Erstellung mit den potenziellen Entscheidungsträgern (z. B. Bund, Ländern, Sozialversicherungen) erörtert werden. Wenn von der Auftraggeberin / vom Auftraggeber gewünscht, können HTA-Produkte Vorschläge zur Implementierung der HTA-Ergebnisse enthalten; Ausführlichkeit und inhaltliche Tiefe der Implementierungsvorschläge sind auf die jeweilige Fragestellung sowie auf die verfügbaren Zeit- und Personalressourcen zur Erstellung des HTA-Produkts abzustimmen. Auch für den Fall, dass zu einer Fragestellung keine ausreichende Evidenz gefunden wird, sollte das weitere Vorgehen mit den Entscheidungsträgern und der Auftraggeberin / dem Auftraggeber geklärt werden. Beispielsweise ist dann der erforderliche Forschungsbedarf aufzuzeigen und zu präzisieren.

Die Wahrscheinlichkeit einer Implementierung von HTA-Ergebnissen wird erwartungsgemäß von zahlreichen Faktoren beeinflusst (Drummond et al. 2000). Sie steigt unter anderem, wenn HTA-Ergebnisse auf sehr zuverlässiger Evidenz beruhen (Gerhardus et al. 2008), wenn sie wenig etablierte Technologien betreffen (Sheldon et al. 2004) und wenn gesundheitspolitische Entscheidungen durch transparente (und bestenfalls evidenzbasierte) Prozesse erfolgen (Oortwinjn 2001).

11.2 Impact

Der Impact bezeichnet den Einfluss von HTA-Produkten auf das Gesundheitssystem (Gerhardus et al. 2008). Die erzielbaren Effekte durch HTA-Produkte sind vielfältig und können sich auf unterschiedliche Adressaten beziehen (z. B. auf gesundheitspolitische Institutionen und Entscheidungsträger, Erstattungseinrichtungen, Hersteller von Arzneimitteln und Medizinprodukten, Ärzteschaft, einzelne Patientinnen und Patienten, Patientenvertretungen). Geeignete Beschreibungsmodelle zum HTA-Impact sowie

valide Methoden zur Messung des Impacts von HTA-Produkten befinden sich derzeit noch in Entwicklung (Gerhardus et al. 2008). Für einige Faktoren und Maßnahmen konnte die einschlägige Forschung jedoch bereits aufzeigen, dass sie mit einer Verbesserung des HTA-Impacts assoziiert sind. So gilt eine hohe methodische Qualität der HTA-Produkte grundsätzlich als Voraussetzung für einen (positiven) Impact (Gerhardus et al. 2008).

Maßnahmen zur Qualitätssicherung (siehe Kapitel 1) sind unter diesem Aspekt angezeigt. Als förderlich ist auch eine gute Verfügbarkeit von HTA-Produkten zu betrachten (McGregor 2006). Korrespondierende Maßnahmen hierzu sind z. B. die Verbesserung der Zugänglichkeit zu nationalen und internationalen HTA-Ergebnissen, die Entwicklungen einer HTA-Datenbank der österreichischen HTA-Ersteller, der HTA-Internetführer der GÖG oder die Link-Liste des LBI zu HTA im Web. Des Weiteren zeigt sich, dass die Erhöhung der Akzeptanz von HTA bei gesundheitspolitischen Entscheidungsträgern und anderen Stakeholdern den Impact verbessern kann (Oortwijn 2001). Förderliche Maßnahmen hierzu sind – neben der bereits genannten Qualitätssicherung für HTA-Ersteller – unter anderem die Einbindung von Stakeholdern bei der HTA-Themenfindung, transparente und standardisierte Prozesse bei der HTA-Erstellung, Öffentlichkeitsarbeit sowie einschlägige Informationsangebote zu HTA für Stakeholder (z. B. Schulungen, Workshops, Symposien).

Einen wesentlichen Faktor zur Erhöhung des HTA-Impacts bildet – im Kontext bereits bestehender Regelungen – die strukturelle und rechtliche Verankerung von HTA im Rahmen gesundheitspolitischer Entscheidungsprozesse (Gerhardus 2008). Die Entwicklung einer nationalen HTA-Strategie in Österreich ist hierzu eine vorbereitende Maßnahme. Vorschläge zur rechtlichen Verankerung von HTA sollen von der HTA-Arbeitsgruppe in Zukunft erarbeitet werden.

Die GÖG beabsichtigt in den folgenden Jahren die Durchführung weiterer konkreter Projekte und Maßnahmen zur Erhöhung des HTA-Impacts sowie mittelfristig eine standardisierte Evaluierung des HTA-Impacts im österreichischen Gesundheitswesen. Im LBI-HTA wird im Jahr 2010 in Kooperation mit der UMIT eine standardisierte Evaluierung von österreichischen HTA-Projekten durchgeführt, die als Vorlage für weitere Impact-Evaluierungen dienen kann (Schumacher/Zechmeister 2010).

Autor: Alexander Eisenmann (GÖG/BIQG)

12 Ausblick

Das Methodenhandbuch wird nach Bedarf überarbeitet und aktualisiert. Sollten Fehler im Handbuch oder wesentliche Entwicklungen eine Aktualisierung erfordern, so wird diese frühestmöglich durchgeführt.

13 Glossar

Adaption bestehender Assessments	Eine Adaption bestehender Berichte kann entweder in Form eines umfassenden Updates oder in Form einer Abänderung von einzelnen Berichtsteilen erfolgen.
Administrative Daten	Gruppe von Routinedaten, die in Form von prozessorientierten, umfangreichen Informationssammlungen vorliegen und im Rahmen der Verwaltung, Leistungserbringung bzw. Kostenerstattung (z. B. bei der gesetzlichen Kranken-, Renten- oder Unfallversicherung) anfallen und elektronisch erfasst sind
As-treated-Analyse	Die As-treated-Analyse berücksichtigt nur die Daten jener Patientinnen und Patienten, die die im Rahmen der Studie zu prüfende Intervention bzw. Medikation tatsächlich erhalten haben, unabhängig von ihrer ursprünglichen Zuordnung zu einer Studiengruppe.
Bias	Unterschiedliche Einflussgrößen können zu einer Fehleinschätzung der untersuchten Assoziation zwischen Exposition und Ereignis führen. Diese systematischen Fehler werden in folgende Kategorien eingeteilt: Auswahlverzerrung der Studienpopulation, Fehler bei der Messung von Exposition, Ereignis oder Störfaktoren, Verzerrung durch einen in der Analyse nicht berücksichtigten dritten Faktor und fälschlich umgekehrte Schlussfolgerung hinsichtlich der Kausalität von Exposition und Ereignis.
Carry-over-Effekt	In Cross-over-Studien, in denen die Teilnehmer sowohl die Intervention(en) als auch Placebo nacheinander erhalten, kann es zu einem Übertragungseffekt der Wirkung aus der einen Interventionsphase in die andere kommen, wenn der Abstand zwischen den Interventionen zu gering ist.
Completers-only-Analyse	Bei der Completers-only-Analyse werden nur die Daten jener Patientinnen und Patienten berücksichtigt, die bis zum Schluss an der Studie teilgenommen haben.
Confounder	Störfaktor, der (unabhängig von der Exposition) einen Risikofaktor für ein Ereignis („Outcome“) darstellt und mit der Exposition korreliert, ohne Teil der Ereigniskette zwischen Exposition und Outcome zu sein
Daten	logisch gruppierte Informationseinheiten
Delphi-Verfahren	Beschreibt eine interaktive Umfragemethode, bei der eine ausgesuchte Gruppe von Expertinnen/Experten gebeten wird, zu einem Thema wiederholt Stellung zu nehmen. Ein Delphi-Verfahren umfasst mehrere schriftliche Befragungsrunden, wobei nach jeder Runde die

	Antworten zusammengefasst und den Teilnehmerinnen /Teilnehmern zugesandt werden.
(Clinical) Effectiveness / Effektivität	Grad für das Erreichen eines bestimmten (therapeutischen) Ergebnisses in einer allgemeinen Patientenpopulation, die eine bestimmte Intervention unter tatsächlichen oder durchschnittlichen Gebrauchs bzw. Alltagsbedingungen im Falle einer bestimmten medizinischen Fragestellung erhält.
Efficacy/Effektivität	Beschreibt die klinische Wirksamkeit unter kontrollierten Bedingungen.
Efficiency/Effizienz	Als effizient wird jene Alternative bezeichnet, die bei einem vorgegebenen Mitteleinsatz den höchsten Zielerreichungsgrad erlangt oder die bei einem vorgegebenen Ziel die geringsten Mittel benötigt (Minimal- bzw. Maximalprinzip). Die Effizienz beinhaltet die Gegenüberstellung von Kosten (Miteinsatz) und Nutzen (medizinischer Output) im Gegensatz zur Effektivität.
Exazerbation	Verschlimmerung, Verschlechterung einer Erkrankung
Explorative Forschung	Bezweckt, in einem relativ unerforschten Bereich Hypothesen zu entwickeln oder theoretische und begriffliche Voraussetzungen zu schaffen. Unterscheidet sich von der Forschung zur Beantwortung von bereits vorab festgelegten Forschungsfragen.
Externe Validität	Kriterium zur Bewertung einer Studie, orientiert an Anwendbarkeit bzw. Übertragbarkeit der Ergebnisse (Umfang und Präzision des Behandlungseffekts, Berücksichtigung aller relevanten Zielkriterien) auf andere Gegebenheiten/Populationen etc.
Gesamtschätzer	Im Kontext einer Metaanalyse das quantitativ zusammengefasste Effektmaß aus allen in die Metaanalyse eingeschlossenen Studien. Wird auch als „gepoolter Schätzer“ bezeichnet.
Heterogenität	Uneinheitlichkeit eines oder mehrerer mathematischer Elemente; Gegenteil von Homogenität (Gleichheit)
HTA-Bericht	Ein umfassendes Health Technology Assessment inkludiert Recherche und Aufarbeitung des aktuellen, international verfügbaren Wissensstandes zur festgelegten Fragestellung mithilfe einer systematischen Übersichtsarbeit und ggf. einer Metaanalyse. Untersucht werden Wirksamkeit, Sicherheit, Wirtschaftlichkeit sowie soziale, ethische, rechtliche und organisatorische Aspekte. Im Rahmen der Berichtserstellung wird entschieden, ob eine ökonomische Modellrechnung erforderlich ist.
Horizon Scanning	Verfahren zur Identifizierung neuer, auf den Markt drängender Technologien.
Iatrogen	(Adjektiv): durch die ärztliche Diagnose oder Versorgung verursacht

Indexierung	Verschlagwortung
Interne Validität	Kriterium zur Bewertung einer Studie im Sinne der Frage, ob die Arbeit in Design und Durchführung glaubwürdig ist (Ausmaß, in dem Verzerrungen vermieden werden, via Randomisierung, Verblindung, Auswahl vergleichbarer Gruppen, gleiche medizinische Behandlung der Gruppen abseits der Studie)
Inzidenz	Maßzahl der Epidemiologie, die die Anzahl der Neuerkrankungen an einer bestimmten Krankheit in einer bestimmten Population während einer bestimmten Zeit angibt
Intention-to-treat-Analyse (ITT)	Die ITT-Analyse reflektiert die realen Verhältnisse einer Intervention an der Patientin / am Patienten, das heißt, es werden auch die Daten jener mit einberechnet, die die zugeordnete Therapieform partiell oder gar nicht erhalten haben, die in die alternative Behandlungsgruppe übergewechselt sind oder die nicht bis zum Schluss an der Studie teilgenommen haben.
Konfidenzintervall	Wertebereich, innerhalb dessen mit einer bestimmten Irrtumswahrscheinlichkeit (zumeist 95 %, wenn $p < 0,05$) der tatsächlich gemessene Wert liegt
Kosten-Wirksamkeits-Analyse	Die Kosten einer Maßnahme in monetären Einheiten werden dem Nutzen in nichtmonetären Einheiten gegenübergestellt (auch Kosten-Effektivitäts-Analyse genannt). Um nichtmonetäre Einheiten darzustellen, kommen je nach Einsatzgebiet verschiedene Ansätze (gerettete Leben, gewonnene Lebensjahre ...) zur Anwendung.
Kosten-Nutzen-Analyse	Der wesentliche Unterschied zur Kosten-Wirksamkeits-Analyse liegt darin, dass sowohl Kosten als auch Nutzen in monetären Einheiten bewertet werden. Das ermöglicht eine zweidimensionale Nutzenmessung. Warum diese Methode eher selten angewandt wird: wegen der als unethisch empfundenen Reduktion eines Nutzens im Gesundheitswesen auf seinen monetären Aspekt.
Kosten-Nutzwert-Analyse	Wie bei der Kosten-Wirksamkeits-Analyse werden bei der Kosten-Nutzwert-Analyse die Kosten in monetären Einheiten dem Nutzen in nichtmonetären Einheiten, hier jedoch als QALY (s. u.), gegenübergestellt. Manchmal werden auch beide Analysearten unter dem Begriff Kosten-Wirksamkeits-Analyse zusammengefasst.
Kosten-Minimierungs-Analyse	Spezialfall der Kosten-Wirksamkeits-Analyse, ermittelt unter der Voraussetzung gleicher Wirksamkeit die kostengünstigere Maßnahme.

Machbarkeitsanalyse (für ein HTA, RA bzw. Update)	Abklärung der genauen Fragestellung und des gesundheitspoli- tischen Hintergrundes, Abklärung der Literaturverfügbarkeit (Litera- tursuche und grobe Sichtung), Abschätzung der sinnvollen (methodi- schen) Vorgehensweise
MeSH (Medical Subject Headings)	polyhierarchischer Thesaurus mit Schlagwörtern, die für die Indexie- rung biowissenschaftlicher Literatur herangezogen werden
MeSH Term	normiertes Vokabular, das von der National Library of Medicine zur Verschlagwortung der Artikel in Medline herangezogen wird
Metaanalyse	Zusammenfassung von Ergebnissen von Primäruntersuchungen mithilfe quantitativer Mittel
Morbidität	Häufigkeit nicht-tödlicher Krankheitsereignisse
Mortalität	Sterblichkeit; als Mortalitätsziffer: Verhältnis der Anzahl von Sterbe- fällen zum Durchschnittsbestand der Population
Ökonomische Modellrechnung	Gegenüberstellung gesundheitsbezogener und wirtschaftlicher Ergebnisse in Bezug auf Gesundheitstechnologien für Patientinnen /Patienten oder eine ganze Bevölkerung unter verschiedenen Szenarien
Per-protocol-Analyse	Bei der Per-protocol-Analyse werden nur die Daten jener Patientin- nen und Patienten berücksichtigt, die die laut Prüfplan zugewiesene Intervention bzw. Medikation tatsächlich erhalten haben. Daten von Patientinnen und Patienten, die vorzeitig aus der Studie ausgeschie- den sind, die Intervention abgesetzt haben, in eine andere Studien- gruppe gewechselt sind oder auf sonstige Weise das Studienprotokoll nicht zur Gänze erfüllt haben, fließen nicht in die Analyse mit ein.
Placebo	Sog. Scheinintervention. Wird in der Regel im Rahmen der klinischen Prüfung neuer Arzneimittel als pharmakologisch unwirksame Substanz in Arzneimittelform angewendet.
Placebo-run-in- Periode/Treatment- run-in-Periode	In Placebo- oder Treatment-run-in-Perioden werden bereits vor tatsächlichem Beginn der klinischen Studie jene Patientinnen und Patienten ausgesiebt, die entweder auf Placebo oder auf die zu testende Medikation in unerwünschter Form und Weise reagieren. Die Anwendbarkeit der Ergebnisse auf durchschnittliche Patientin- nen/Patienten wird dadurch reduziert.
Prävalenz	Kennzahl der Epidemiologie, die aussagt, wie viele Menschen einer bestimmten Population an einer bestimmten Krankheit erkrankt sind
Primärstudien	In Primärstudien werden Originaldaten (Primärdaten) von Patientin- nen und Patienten erhoben.
Publikationsbias	Unter Publikationsbias versteht man einen Bias aufgrund einer selektiven Publikationspraxis, bei der Studien mit positiven und

	signifikanten Ergebnissen eine größere Chance haben, publiziert zu werden, als Studien mit negativen und nicht-signifikanten Resultaten. Ein systematischer Review oder eine Metaanalyse, die sich ausschließlich auf publizierte Studien stützt, läuft Gefahr, den Effekt der untersuchten Intervention zu überschätzen.
QALY	Qualitätsbereinigtes Lebensjahr (<i>Quality Adjusted Life-Year</i>). Kennzahl für die Bewertung eines Lebensjahres in Relation zur Gesundheit. Bewertet wird die Lebensqualität auf einer Skala von 0 bis 1 (Tod bis bestmöglicher Gesundheitszustand).
Rapid Assessment	Ein Berichtsverfahren, das innerhalb eines kurzen Zeitrahmens (ca. zwei Monate) die Evidenz zu einer fokussierten Fragestellung zusammenfasst, die gezielt auf den Informationsbedarf der Nutzer/innen zugeschnitten ist. Der verkürzte Zeitrahmen führt zu einer geringeren Detailtiefe von RA im Vergleich zu HTA-Berichten; aufgrund der vereinfachten Methodik besteht außerdem ein erhöhtes Risiko für Verzerrungen.
Randomisierte kontrollierte Studie (randomized controlled trials, RCT)	Die Patientinnen und Patienten (mit einem bestimmten Krankheitsbild) werden durch verschiedene Randomisierungsmethoden per Zufall (at random) in zwei Gruppen geteilt (Interventions- oder Kontrollgruppe), von denen eine die Therapie bekommt, die andere nicht.
Relatives Risiko	Verhältnis der Wahrscheinlichkeit des Eintretens eines Ereignisses in der exponierten Gruppe im Vergleich zu einer nicht-exponierten Gruppe
Reliabilität	Zuverlässigkeit bzw. Genauigkeit eines Verfahrens, gibt den Grad der Genauigkeit der Messung an. Eine Messung mit perfekter Reliabilität wäre ein Verfahren ohne jeglichen Messfehler oder ohne Fehlerquelle. In der Praxis ist dies leider nicht möglich, es können aber unterschiedliche Schritte unternommen werden, um eine möglichst hohe Reliabilität in der Forschung zu gewährleisten.
Routinedaten	Sekundärdaten, die regelmäßig, verpflichtend und zu primär nicht wissenschaftlichen Zwecken erhoben werden
Sekundärdaten	Daten, die von anderen als den HTA-Erstellern / Erstellerinnen für einen ursprünglich anderen (als wissenschaftlichen) Zweck gesammelt werden. Der Begriff ist in Abgrenzung zu Primärdaten zu verstehen, die für den spezifischen Zweck der bearbeiteten Fragestellung erhoben werden.
Sekundärdatenanalyse	Analyse von Daten im Rahmen wissenschaftlicher Untersuchungen ohne direkten Bezug zum primären Erhebungsanlass

Selektionsbias	systematische Verzerrung (Über-, Unterschätzung) des interessierenden Effekts infolge unterschiedlicher Auswahlkriterien in den zu vergleichenden Gruppen
Sensitivität	Wahrscheinlichkeit, einen tatsächlich positiven Sachverhalt durch ein positives Testergebnis zu erkennen; z. B.: Wie viele tatsächlich Erkrankte werden durch einen Test als Erkrankte diagnostiziert?
Spezifität	Wahrscheinlichkeit, einen tatsächlich negativen Sachverhalt durch ein negatives Testergebnis zu erkennen; z. B.: Wie viele tatsächlich Gesunde werden durch einen Test als Personen ohne Erkrankung diagnostiziert?
Systematische Übersichtsarbeit	Mithilfe einer systematischen Literatursuche in internationalen Datenbanken, einer systematischen Selektion der aufgefundenen Literatur nach vorab definierten Kriterien, einer systematischen Datenextraktion und Qualitätsbewertung der ausgewählten Studien sowie einer qualitativen (und ggf. quantitativen = Metaanalyse) Datensynthese wird der aktuelle Wissensstand zu einer definierten Fragestellung in einem Bericht zusammengetragen und übersichtlich dargestellt, um Schlussfolgerungen daraus zu ziehen.
Subject heading	standardisierte Schlagwörter
Thesaurus	Ordnungssystem, in dem Begriffe systematisch dargestellt und durch synonyme Bezeichnungen beschrieben werden. Wörter mit gleicher Bedeutung werden durch Schlagwörter zusammengefasst, wodurch ein kontrolliertes Vokabular entsteht.
Update bestehender Assessments	Bei Erstellung eines Updates wird neue Evidenz identifiziert und in einem bereits vorhandenen und vollständigen Bericht eingearbeitet. Dabei kann das Update als Erweiterung oder Einschränkung der ursprünglichen Fragestellung dienen, je nachdem ob Fragestellung und (oder) Selektionskriterien überarbeitet werden oder nicht.

14 Literaturverzeichnis

14.1 Literatur zu Kapitel 2 HTA-Produkte

GÖG/BIQG 2010: Prozesshandbuch für Health Technology Assessment Version 1.2010. Gesundheit Österreich GmbH / Geschäftsbereich BIQG. Wien

14.2 Literatur zu Kapitel 3 Themenfindung und Priorisierung

GÖG/BIQG 2008: Health Technology Assessment: Organisationsstrukturen in ausgewählten Ländern und HTA-Aktivitäten in Österreich (Langfassung). Gesundheit Österreich GmbH / Geschäftsbereich BIQG. Wien,
<http://www.goeg.at/de/Bereich/Nationale-HTA-Strategie.html>

GÖG/BIQG 2010: Prozesshandbuch für Health Technology Assessment Version 1.2010. Gesundheit Österreich GmbH / Geschäftsbereich BIQG. Wien

GÖG/BIQG 2009: Nationale HTA-Strategie. Gesundheit Österreich GmbH / Geschäftsbereich BIQG. Wien

Perleth M, Busse R, Gerhardus A, Gibis B, Lüthmann D (2008): Health Technology Assessment Konzepte, Methoden, Praxis für Wissenschaft und Entscheidungsfindung. Berlin

14.3 Literatur zu Kapitel 4 Methodik der Durchführung eines Health Technology Assessment

Biomarkers Definitions Working Group 2001: Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. In: Clinical pharmacology and therapeutics, vol. 69, no. 3, pp. 89–95

Centers for Disease Control. The Guide to Community Preventive Services. Verfügbar unter: www.thecommunityguide.org/methods/abstractionform.pdf. Zitiert am 23.07.2006.

Counsell C: Formulating questions and locating primary studies for inclusion in systematic reviews. In: *Ann. Intern. Med.* 127 (5):380–387, 1997

CRD 2001: Undertaking Systematic Reviews of Research on Effectiveness. CRD's Guidance for those Carrying Out or Commissioning Reviews. NHS Center for Reviews and Dissemination. York. <http://www.york.ac.uk/inst/crd/report4.htm> (2. 10. 2008)

CRD 2009: CRDs guidance for Undertaking Systematic Reviews in Health Care. Center for Reviews and Dissemination. University of York. 2009. http://www.york.ac.uk/inst/crd/systematic_reviews_book.htm

DACEHTA 2008: Health Technology Assessment Handbook. Danish Centre for Health Technology Assessment. Copenhagen. http://www.sst.dk/Planlaegning_og_behandling/Medicinsk_teknologivurdering.aspx?lang=en (5. 9. 2008)

Drummond M, Weatherly H (2000): Implementing the findings of health technology assessments. If the CAT got out of the bag, can the TAIL wag the dog. In: *Int J Technol Assess Health Care*, 16(1):1–12

Eng J: Sample size estimation: how many individuals should be studied? *Radiology* 2003, 227: 309–13

EUnetHTA 2008: HTA Core Model Handbook for diagnostic technologies 1.0R. <http://www.eunethta.eu/upload/WP4/Final%20Deliverables/HTA%20Core%20Model%20for%20Diagnostic%20Technologies%201%20r.pdf>. accessed 7.6.2010.

EUnetHTA 2008: HTA Core Model Handbook for medical and surgical interventions. <http://www.eunethta.net/upload/WP4/Final%20Deliverables/HTA%20Core%20Model%20for%20Medical%20and%20Surgical%20Interventions%201%20r.pdf>. accessed 7.6.2010

Garritty, Chantelle; Tsertsvadze, Alexander; Tricco, Andrea; Sampson, Margaret; Moher, David (2010): Updating Systematic Reviews: An International Survey. In: *PLoS ONE* 2010/5/4, 1–11

Gartlehner G (2009): Internes Manual. Abläufe und Methoden. Teil 2 (2. Aufl.). Ludwig Boltzmann Institut. Health Technology Assessment. HTA-Projektbericht 06

GÖG/BIQG 2008: Health Technology Assessment: Organisationsstrukturen in ausgewählten Ländern und HTA-Aktivitäten in Österreich (Langfassung). Gesundheit Österreich GmbH / Geschäftsbereich BIQG. Wien.
<http://www.goeg.at/de/Bereich/Nationale-HTA-Strategie.html>

GÖG/BIQG 2010: Prozesshandbuch für Health Technology Assessment Version 1.2010. Gesundheit Österreich GmbH / Geschäftsbereich BIQG. Wien
GÖG/BIQG 2009: Nationale HTA-Strategie. Gesundheit Österreich GmbH / Geschäftsbereich BIQG. Wien

Guba 2008: Systematische Literatursuche. In: Medizinische Wochenschrift, 158/1-2, Wien

Higgins JPT, Green S (2011): Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. www.cochrane-handbook.org (9. 7. 2011)

IPF 2006: Guidelines zur gesundheitsökonomischen Evaluation (Konsenspapier). Institut für Pharmaökonomische Forschung. Wien.
www.ispor.org/PEguidelines/source/Guidelines_Österreich.pdf (5. 9. 2008)

IQWiG 2008a: Allgemeine Methoden. Version 3.0. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Köln.
<http://www.iqwig.de/allgemeine-methoden.428.html> (5. 9. 2008)

IQWiG 2008b: Methodik für die Bewertung von Verhältnissen zwischen Nutzen und Kosten im System der deutschen gesetzlichen Krankenversicherung. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Köln.
<http://www.iqwig.de/kosten-nutzen-bewertung.736.html> (5. 9. 2008)

Kunz R, Khan KS, Kleijnen J, Antes G (Hg.): Systematische Übersichtsarbeiten und Metaanalysen. 2. Auflage 2009. Hogrefe, Bern.

LBI-HTA 2007a: (Internes) Manual: Abläufe und Methoden. Ludwig Boltzmann Institut für Health Technology Assessment. Wien.
<http://eprints.hta.lbg.ac.at/713/> (5. 9. 2008)

Liberati A, Altman D, Tetzlaff J, Mulrow C, Götzsche P, Ioannidis J, Clarke M, Devereaux P, Kleijnen J, Moher D (2009): The PRISMA Statement for Reporting Systematic Reviews and Meta-Analysis of Studies that Evaluate Health Care Interventions: Explanation and Elaboration: In: PLoS Medicine 6/7

McGregor M (2006): What decision-makers want and what they have been getting. Value in Health, 9: 181-185

Machin D, Campbell MJ: Design of studies for medical research. Chichester: Wiley 2005; 1-286

Mangiapane S, Velasco Garrido M (2009): Surrogatendpunkte als Parameter der Nutzenbewertung. DIMDI HTA-Bericht 91. <http://www.dimdi.de> (23.11.2009)

Moher, David; Tsertsvadze, Alexander; Tricco, Andrea; Eccles, Martin; Grimshaw, Jeremy; Sampson, Margaret; Barrowman, Nick (2008): When and how to update systematic reviews. In: Cochrane Database of Systematic Reviews 2008/1, o. S.

Müllner Marcus: Erfolgreich wissenschaftlich arbeiten in der Klinik. Springer Wien New York, 2002

National Institute for Health and Clinical Excellence (April 2006): The guidelines manual. London: National Institute for Health and Clinical Excellence. Verfügbar unter: www.nice.org.uk. Zitiert am 25.05.2006.

Randomization.com: Welcome to randomization.com. www.randomization.com/; letzte Version: 16. 7. 2008.

Röhrig B, Prel J-B-du, Wachtlin D, Blettner M.: Studientypen in der medizinischen Forschung: Teil 3 der Serie zur Bewertung wissenschaftlicher Publikationen. In: Dtsch Arztebl Int 2009; 106(15): 262-8

Schug L, Federspiel B, Brügger U (2006): Synthesebericht. Ökonomische Evaluation in den Kernthemen. Gesundheitsförderung Schweiz. Bern

Shojania, Kaveh; Sampson, Margaret; Ansari, Mohammed; Ji, Jun; Garritty, Chantelle; Rader, Tamara; Moher, David (2007): Updating Systematic Reviews. Technical Review No. 16. (Prepared by the University of Ottawa Evidence-based Practice Center under Contract No. 290-02-0017.) AHRQ Publication No. 07-0087. Rockville

The Cochrane Collaboration 2008: Cochrane handbook for systematic reviews of interventions. Version 5.01. The Cochrane Collaboration. Chichester.
<http://www.cochrane-handbook.org/> (2. 10. 2008)

WHO 2007: World Health Organization Family of International Classifications: definition, scope and purpose. <http://www.who.int/> (27. 12. 2010)

14.4 Literatur zu Kapitel 5 Zusätzliche Datengenerierung und Datenanalyse

14.4.1 Literatur zu Kapitel 5.1 Grundlagen

Bohnsack Ralf (2000): Gruppendiskussion. In: Flick Uwe, Kardoff Ernst von & Steinke, Ines (Hg). Qualitative Forschung – ein Handbuch. Reinbek: Rowohlt. 369–384

Bortz Jürgen, Döring Nicola (1984/2002): Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Berlin: Springer-Verlag

Diekmann Andreas (1995/2003): Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. Reinbek bei Hamburg: Rowohlt

Diekmann Andreas (Hg.), (2004): Methoden der Sozialforschung. Kölner Zeitschrift für Soziologie und Sozialpsychologie. Wiesbaden: VS Verlag für Sozialwissenschaften.

Flick Uwe (1995/2006): Qualitative Sozialforschung. Eine Einführung. Reinbek bei Hamburg: Rowohlt

Froschauer Ulrike, Lueger Manfred (2003): Das qualitative Interview. Wien: Facultas Verlag

Geyer S, Siegrist J (1998): Sozialwissenschaftliche Verfahren in den Gesundheitswissenschaften. In: Hurrelmann, Klaus, Laaser, Ulrich (Hg.): Handbuch Gesundheitswissenschaften, 1998. München: Juventa Verlag: 259–279

Häder Michael, Häder Sabine (Hg), (2002): Die Delphi-Technik in den Sozialwissenschaften. Methodische Forschungen und innovative Anwendungen. Wiesbaden: Westdeutscher Verlag

König René (1962/1972): Das Interview. Köln: Kiepenheuer und Witsch

Linstone Harold A, Turoff Murray (Hg), (1975): The Delphi Method. Techniques and Applications. Massachusetts: Addison-Wesley Publishing Company

Loos Peter, Schäffer Burkhard (2001): Das Gruppendiskussionsverfahren. Theoretische Grundlagen und empirische Anwendung. Opladen: Leske & Budrich

Morgan David L (1988): Focus groups as qualitative research. Newbury Park, CA: Sage

Rubin Herbert, Rubin Irene (2005)
Qualitative Interviewing. The art of hearing data. London: Sage.

Schnell Rainer, Hill Paul, Esser Elke (1995)
Methoden der empirischen Sozialforschung. München: Oldenburg.

Schulz Marlen (HglN) (2009)
Das Gruppendelphi. Konzept und Fragebogenkonstruktion. Wiesbaden: VS Verlag für Sozialwissenschaften.

Stone DH (1993): Design a questionnaire. In: BMJ Volume 307: 1264-1266

14.4.2 Literatur zu Kapitel 5.2. Einsatz von Routinedaten im Rahmen von Health Technology Assessments

Abuzahra M, Piso B. Schweregraddifferenzierung in der Neuro- und Traumarehabilitation. Projektbericht Nr. 23c. Wien: Ludwig Boltzmann Institut für Health Technology Assessment; 2010.

Andersohn F, Garbe E. Pharmakoepidemiologische Forschung mit Routinedaten des Gesundheitswesens. Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz 2008:1135-44.

Arbeitsgemeinschaft Medikamentöse Tumortherapie (AGMT). Register der AGMT. 2011 [cited 2011 July 2011]; Available from:
<http://www.agmt.at/index.php?m=sz&s=kontakt>

Arbeitsgruppe Erhebung und Nutzung von Sekundärdaten der DGSMP, Arbeitsgruppe Epidemiologische Methoden der DGEpi und der GMDS und der DGSMP. GPS- Gute Praxis Sekundärdatenanalyse: Revision nach grundlegender Überarbeitung. 2008.

Arbeitsgruppe Erhebung und Nutzung von Sekundärdaten der DGSMP, Arbeitsgruppe Epidemiologische Methoden der DGEpi und der GMDS und der DGSMP. GPS- Gute Praxis Sekundärdatenanalyse: Revision nach grundlegender Überarbeitung. Das Gesundheitswesen. 2008;70:54–60.

Arbeitsgruppe Interventionelle Kardiologie. Austrian National CathLab Registry. 2011 [cited 2011 July 2011]; Available from: <http://iik.i-med.ac.at/index.php>

Berger ML, Mamdani M, Atkins D, Johnson ML. Good Research Practice for Comparative Effectiveness Research: Defining, Reporting and Interpreting Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The ISPOR Good Research Practice for Retrospective Database Analysis Task Force Report – Part I. Value in Health 2009;10:44–52.

Boslaugh S. Secondary Data Sources for Public Health. A Practical Guide. University Press 2007.

Bundeskanzleramt. Datenschutzgesetz 2000. Fassung 2011. Wien: Bundeskanzleramt; 2011.

Bundeskanzleramt. Epidemiegesetz 1950 i. d. g. F. BGBl Nr. 76/2008. Wien: Bundeskanzleramt; 2008.

Bundeskanzleramt. IVF-Fondsgesetz. Wien: Bundeskanzleramt; 2010.

Bundesministerium für Gesundheit. Dokumentation im ambulanten Bereich: Bericht zum Projekt "Dokumentation im ambulanten Bereich". Wien: Bundesministerium für Gesundheit; 2010.

Bundesministerium für Gesundheit. Handbuch zur Dokumentation in landesgesundheitsfondsfinanzierten Krankenanstalten: 2004+. Wien: Bundesministerium für Gesundheit; 2010.

Bundesministerium für Gesundheit. Leistungsorientierte Krankenanstaltenfinanzierung (LKF) Systembeschreibung. Wien: Bundesministerium für Gesundheit; 2009.

Bundesministerium für Gesundheit Familie und Jugend. Epidemiologisches Meldesystem. Benutzerhandbuch fachlicher Teil. Wien: Bundesministerium für Gesundheit, Familie und Jugend; 2008.

Bundesministerium für Gesundheit Familie und Jugend. Handbuch zur Dokumentation von Kostendaten in landesgesundheitsfondsfinanzierten Krankenanstalten. Wien:

Bundesministerium für Gesundheit, Familie und Jugend; 2007.

Cox E, Martin BC, Van Staa T, Garbe E, Siebert U, Johnson ML. Good Research Practices for Comparative Effectiveness Research: Approaches to Mitigate Bias and Confounding in the Design of Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The International Society for Pharmacoeconomics and Outcome Research Good Research Practices for Retrospective Database Analysis Task Force Report – Part II. *Value in Health*. 2009;12(8):1053–61.

Cox E, Martin BC, Van STaa T, Garbe E, Siebert U, Johnson ML. Good Research Practices for Comparative Effectiveness Research: Approaches to Mitigate Bias and Confounding in the Design of Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The International Society for Pharmacoeconomics and Outcome Research Good Research Practices for Retrospective Database Analysis Task Force Report – Part II. *Value in Health* 2009:1053–61.

Drummond M, Sculpher M, Torrance GW, O'Brien B, Stoddart GL, eds. *Methods for the Economic Evaluation of Health Care Programmes*. 3rd ed. New York: Oxford University Press 2005.

Geiersbach A. Dokumentation im ambulanten Bereich: Fortsetzung des Projektes "Katalog ambulanter Leistungen". *Soziale Sicherheit*. 2009;Mai 2009:256–66.

Geiersbach A. Erstellung eines Kataloges ambulanter Leistungen. *Soziale Sicherheit*. 2008;Februar 2008:78–84.

Gesundheit Österreich GmbH. Blutverbrauchserhebung. 2011 [cited 2011 August]; Available from: <http://www.goeg.at/de/Bereich/Blutverbrauchs-Erhebung.html>

Glaeske G, Augustin M, Abholz H, Banik N, Brüggengjürgen B, Hasford J, et al. *Epidemiologische Methoden für die Versorgungsforschung. Das Gesundheitswesen*. 2009;71:685–93.

Gliklich R, Dreyer N, eds. *Registries for Evaluating Patient Outcomes: A User's Guide: Outcome DEClDE Center* 2007.

Gyimesi M, Piso B, Tüchler H. *Ambulante Kardiologische Rehabilitation Teil IIIa: Retrospektive Kohortenstudie (mit/ohne Phase III-Rehabilitation), explorative Analyse und Entwicklung eines Auswertungskonzepts*. HTA-Projektbericht Nr. 31a. Wien: Ludwig Boltzmann Institut für Health Technology Assessment; 2010.

Hauptverband der österreichischen Sozialversicherungsträger. *Handbuch der österreichischen Sozialversicherung 2011*. Wien: Hauptverband der österreichischen Sozialversicherungsträger; 2011.

Hauptverband der Sozialversicherungsträger. Statistisches Handbuch der österreichischen Sozialversicherung. Wien: Hauptverband der Sozialversicherungsträger; 2010.

Hoffmann F, Andersohn F, Giersiepen K, Scharnetzky E, Garbe E. Validierung von Sekundärdaten. Bundesgesundheitsblatt–Gesundheitsforschung–Gesundheitsschutz. 2008;51(1118–1126).

Hoffmann F, Bobrowski C, Fendrich K. Sekundärdatenanalyse in der Versorgungsepidemiologie. Bundesgesundheitsblatt–Gesundheitsforschung–Gesundheitsschutz. 2008;51:1193–201.

Hoffmann F, Glaeske G. Analyse von Routinedaten. In: Pfaff H, Neugebauer EAM, Glaeske G, Schrappe M, eds. Lehrbuch Versorgungsforschung. Stuttgart: Schattauer 2011.

Holle R, Behrend C, Reitmeir P, John J. Methodenfragen der Nutzung von GKV–Routinedaten für Kostenanalysen. In: Swart E, Ihle P, eds. Routinedaten im Gesundheitswesen Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und Perspektiven. Bern: Huber 2005.

Ihle P. Datenschutzrechtliche und methodische Aspekte beim Aufbau einer Routinedatenbasis aus der Gesetzlichen Krankenversicherung zu Forschungszwecken. Bundesgesundheitsblatt–Gesundheitsforschung–Gesundheitsschutz. 2008;51:1127–34.

Institut für klinische Epidemiologie der TILAK. Diabetesregister Tirol. 2011 [cited 2011 June 2011]; Available from: <http://www.iet.at/index.php?id=4127>

Institut für klinische Epidemiologie der TILAK. Geburtenregister Tirol und Österreich. 2011 [cited 2011 June 2011]; Available from: <http://www.iet.at/index.php?geburtenregisterhom>

Institut für klinische Epidemiologie der TILAK. Prothesenregister Tirol. 2011 [cited 2011 June 2011]; Available from: <http://www.iet.at/index.php?id=1617>

Institut für klinische Epidemiologie der TILAK. Tumorregister Bericht der Diagnosejahre 2004–2006. Innsbruck: Institut für klinische Epidemiologie der TILAK; 2008.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Arbeitspapier Kostenbestimmung. Köln: IQWiG; 2009.

Johnson ML, Crown W, Martin BC, Dormuth CR, Siebert U. Good Research Practices for Comparative Effectiveness Research: Analytic Methods to Improve Causal Inference

from Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report – Part III. *Value in Health*. 2009;12(8):1062–73.

Johnson ML, Crown W, Martin BC, Dormuth CR, Siebert U. Good Research Practices for Comparative Effectiveness Research: Analytic Methods to Improve Causal Inference from Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report – Part III. *Value in Health* 2009:1062–73.

KAGES. Geburtenregister Steiermark. Graz: KAGES, IES; 2008.

Koordinationsbüro für das Transplantationswesen. Transplant–Jahresbericht 2010. Wien: Österreichisches Bundesinstitut für Gesundheitswesen; 2010.

Last JM. *A dictionary of epidemiology*. 4 ed. Oxford: University Press 2001.

Leitner H. Geburtenregister Österreich heute. Innsbruck: Institut für klinische Epidemiologie der TILAK; 2009.

Mathis S, Wild C. Register für klinische und gesundheitsökonomische Fragestellungen: Einsatzbereiche von kardiovaskulären, wirbelsäulenspezifischen und neurologischen Registern und Good Practice Strategien für die Arbeit mit Registern. HTA–Projektbericht Nr. 11. Wien: Ludwig Boltzmann Institut für HTA.; 2008.

Motheral B, Brooks J, Clark MA, Crown W, Davey P, Hutchins D, et al. A Checklist for Retrospective Database Studies – Report of the ISPOR Task Force on Retrospective Databases. *Value in Health*. 2003;6(2):90–7.

Müller D, Augustin M, Banik N, Baumann W, Bestehorn K, Kieschke J, et al. Memorandum Register für die Versorgungsforschung. *Das Gesundheitswesen* 2010:824–39.

Müller D, Augustin M, Banik N, Baumann W, Bestehorn K, Kieschke J, et al. Memorandum Register für die Versorgungsforschung. *Das Gesundheitswesen*. 2010;72:824–39.

Oberaigner W, Labek G. Prothesenregister Tirol. Bericht über Operationsjahre 2004–2006. Innsbruck: Institut für klinische Epidemiologie der TILAK; 2007.

Oberaigner W, Leitner H. Geburtenregister Tirol. Bericht über die Geburtshilfe in Tirol 2007. Innsbruck: Institut für klinische Epidemiologie der TILAK; 2008.

Österreichische Hämophilie Gesellschaft. Österreichisches Hämophilie Register. 2011 [cited 2011 June 2011]; Available from: <https://www.studienregister.at/web/guest/home>

Perleth M, Antes G, Busse R, Gibis B, Lühmann D, Meyer V, et al. Sekundärdatennutzung im Health Technology Assessment (HTA). In: Swart E, Ihle P, eds. *Routinedaten im Gesundheitswesen Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und Perspektiven*. Bern: Hans Huber 2005.

Perleth M, Busse R. Health Technology Assessment (HTA) – Teil und Methode der Versorgungsforschung. *Gesundheitsökonomie Qualitätsmanagement*. 2004;9:172–6.

Qualitätsmanagement ÖGfQu, in der Medizin GmbH. CIRS medical. 2011 [cited 2011 July 2011]; Available from: <http://www.cirsmedical.at/index.php?id=71>

Raftery J, Bryant J, Powell J, Kerr C, Hawker S. Payment to healthcare professionals for patient recruitment to trials: systematic review and qualitative study. *Health Technology Assessment* 2008;1–128.

Raftery J, Roderick P, Stevens A. Potential use of routine databases in health technology assessment. *Health Technology Assessment*. 2005(20):1–106.

Rechnungshof. Finanzierung und Kosten von Leistungen in Spitalsambulanzen und Ordinationen. Wien: Rechnungshof; 2011.

Schumacher I, Zechmeister I. Auswirkungen der HTA-Forschung auf das Gesundheitswesen in Österreich. Teil 2: Ergebnisse der empirischen Erhebung. Wien: Ludwig Boltzmann Institut für Health Technology Assessment; 2011.

Statistik Austria. Krebsinzidenz und Krebsmortalität in Österreich. Wien: Statistik Austria; 2010.

Swart E, Ihle P. Der Nutzen von GKV-Routinedaten für die Versorgungsforschung. *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz*. 2008;51:1093–4.

Swart E, Ihle P, eds. *Routinedaten im Gesundheitswesen. Handbuch Sekundärdatenanalyse: Grundlagen, Methoden, Perspektiven*. Bern: Hans Huber 2005.

Van Emmerich C, Metzinger B. Qualitätssicherung mit Routinedaten aus der Sicht der deutschen Krankenhausgesellschaft. *Das Krankenhaus*. 2010;12:1177–82.

Warmuth M, Mad P. Haemocomplettan alone or in combination with Fibrogammin in acquired hypofibrinogenemia. Wien: Ludwig Boltzmann Institut für Health Technology Assessment; 2009.

Warmuth M, Mad P, Piso B, Wild C. Eltern-Kind-Vorsorge neu. Teil I: Epidemiologie – Häufigkeiten von Risikofaktoren und Erkrankungen in Schwangerschaft und früher Kindheit. Wien: Ludwig Boltzmann Institut für Health Technology Assessment; 2011.

Wilbacher I, Schröder J. Zusammenführung verschiedener Dokumentationssysteme. Neuchâtel 2006.

Wild C, Mad P, Kvas E, Reichelt T, Adlbrecht C. Rolle und Positionierung der Ambulanzen von Universitätskliniken im Gesundheitswesen. Wien: Ludwig Boltzmann Institut für HTA; 2007.

Wissenschaftliches Institut der AOK. Qualitätssicherung mit Routinedaten. 2011 [cited 2011 August]; Available from: <http://www.qualitätssicherung-mit-routinedaten.de/>

Zechmeister I, Freiesleben de Blasio B, Radlberger P, Wild C, Kvas E, Garnett GP, et al. Economic Evaluation of HPV vaccination in Austria. Vienna: Ludwig Boltzmann Institute for Health Technology Assessment; 2007.

Zechmeister I, Stollenwerk B, Langley T. Have statins met our expectations? A comparison of expected health gains from statins with epidemiological trends in Austria. Wien: Ludwig Boltzmann Institut für Health Technology Assessment; 2008.

14.5 Literatur zu Kapitel 6 Beurteilung der medizinischen Studien und Extraktion der Daten und Kapitel 7 Synthese der medizinischen Evidenz

"Reviews: making sense of an often tangled skein of evidence", 2005, Ann Intern Med, vol. 142, no. 12 Pt 1, pp. 1019–20

Altman DG, Deeks JJ, Sackett DL (1998): Odds ratios should be avoided when events are common. In: BMJ, vol. 317, no. 7168, pp. 1318

Altman, D. G. (2001). Systematic reviews of evaluations of prognostic variables. Systematic Reviews in Health Care: Meta-analysis in context. Second edition. M. Egger, G. D. Smith and D. G. Altman. London, BMJ Publishing Group.

Andrea LC, Laura CM, Gerald G (2009): Semi-automating the manual literature search for systematic reviews increases efficiency. In: *Health Information & Libraries Journal*, vol. 9999, no. 9999

Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, Guyatt GH, Harbour RT, Haugh MC, Henry D, Hill S, Jaeschke R, Leng G, Liberati A, Magrini N, Mason J, Middleton P, Mrukowicz J, O'Connell D, Oxman AD, Phillips B, Schunemann HJ, Edejer TT, Varonen H, Vist GE, Williams JW Jr, Zaza S (2004a): Grading quality of evidence and strength of recommendations. In: *BMJ*, vol. 328, no. 7454, pp. 1490

Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, Liberati A, O'Connell D, Oxman AD, Phillips B, Schunemann H, Edejer TT, Vist GE, Williams JW Jr (2004b): Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. In: *BMC Health Serv Res*, vol. 4, no. 1, pp. 38

Balk EM, Lau J, Bonis PA (2005): Reading and critically appraising systematic reviews and meta-analyses: a short primer with a focus on hepatology. In: *J Hepatol*, vol. 43, no. 4, pp. 729-36

Barbour RS (2001): Checklists for improving rigour in qualitative research: a case of the tail wagging the dog? In: *BMJ (Clinical research ed.)*, vol. 322, no. 7294, pp. 1115-1117

Begg, C. B. and M. Mazumdar (1994). "Operating characteristics of a rank correlation test for publication bias." *Biometrics* 50(4): 1088-1101.

Betran AP, Say L, Gulmezoglu AM, Allen T, Hampson L (2005): Effectiveness of different databases in identifying studies for systematic reviews: experience from the WHO systematic review of maternal morbidity and mortality. In: *BMC Med Res Methodol*, vol. 5, no. 1, pp. 6

Biomarkers Definitions Working Group 2001: Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. In: *Clinical pharmacology and therapeutics*, vol. 69, no. 3, pp. 89-95

Borenstein, M., L. V. Hedges, et al. (2009). *Introduction to Meta-Analysis*. Chichester, West Sussex, UK, John Wiley & Sons, Ltd.

Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HC (2003a): Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. In: *BMJ*, vol. 326, no. 7379, pp. 41-4

Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Moher D, Rennie D, de Vet HC, Lijmer JG (2003b): The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. In: *Clin Chem*, vol. 49, no. 1, pp. 7–18

Brook RH, Lohr KN (1985): Efficacy, effectiveness, variations, and quality. Boundary-crossing research. In: *Med Care*, vol. 23, no. 5, pp. 710–22

Brozek JL, Akl EA, Alonso-Coello P, Lang D, Jaeschke R, Williams JW, Phillips B, Lelgemann M, Lethaby A, Bousquet J, Guyatt GH, Schunemann HJ, GRADE Working Group (2009): Grading quality of evidence and strength of recommendations in clinical practice guidelines. Part 1 of 3. An overview of the GRADE approach and grading quality of evidence about interventions. In: *Allergy*, vol. 64, no. 5, pp. 669–677

Burdett S, Stewart LA, Tierney JF (2003): Publication bias and meta-analyses: a practical example. In: *Int J Technol Assess Health Care*, vol. 19, no. 1, pp. 129–34

Buscemi N, Hartling L, Vandermeer B, Tjosvold L, Klassen TP (2006): Single data extraction generated more errors than double data extraction in systematic reviews. In: *J Clin Epidemiol*, vol. 59, no. 7, pp. 697–703

Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, Tyrer P (2000): Framework for design and evaluation of complex interventions to improve health. In: *BMJ (Clinical research ed.)*, vol. 321, no. 7262, pp. 694–696

Canonico, M., G. Plu-Bureau, et al. (2008). "Hormone replacement therapy and risk of venous thromboembolism in postmenopausal women: systematic review and meta-analysis." *BMJ* 336(7655): 1227–1231.

Clarke, M. J. and L. A. Stewart (2001). Obtaining individual patient data from randomised controlled trials. *Systematic Reviews in Health Care: Meta-analysis in context*. Second edition. M. Egger, G. D. Smith and D. G. Altman. London, BMJ Publishing Group.

Counsell C (1997): Formulating questions and locating primary studies for inclusion in systematic reviews. In: *Ann Intern Med*, vol. 127, no. 5, pp. 380–7

Curtin, F., D. G. Altman, et al. (2002). "Meta-analysis combining parallel and cross-over clinical trials. I: Continuous outcomes." *Stat Med* 21(15): 2131–2144.

Curtin, F., D. Elbourne, et al. (2002). "Meta-analysis combining parallel and cross-over clinical trials. II: Binary outcomes." *Stat Med* 21(15): 2145–2159.

Curtin, F., D. Elbourne, et al. (2002). "Meta-analysis combining parallel and cross-over clinical trials. III: The issue of carry-over." *Stat Med* 21(15): 2161–2173.

Daly J, Willis K, Small R, Green J., Welch N, Kealy M, Hughes E (2007): A hierarchy of evidence for assessing qualitative health research. In: *J Clin Epidemiol*, vol. 60, no. 1, pp. 43–9

Day SJ, Altman DG. 2000, "Statistics notes: blinding in clinical trials and other studies", *BMJ (Clinical research ed.)*, vol. 321, no. 7259, pp. 504.

Deeks JJ (2002): Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. In: *Stat Med*, vol. 21, no. 11, pp. 1575–600

Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, Petticrew M, Altman, DG (2003): Evaluating non-randomised intervention studies. In: *Health Technol Assess*, vol. 7, no. 27, pp. iii–x, 1–173

Deeks, J. J., D. G. Altman, et al. (2001). *Systematic Reviews in Health Care: Meta-analysis in context*. London, BMJ Publishing Group.

Deeks, J. J. (2001). Systematic reviews of evaluations of diagnostic and screening tests. *Systematic Reviews in Health Care: Meta-analysis in context*. Second edition. M. Egger, G. D. Smith and D. G. Altman. London, BMJ Publishing Group.

Deeks, J. J., D. G. Altman, et al. (2001). Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. *Systematic Reviews in Health Care: Meta-analysis in context*. Second edition. M. Egger, G. D. Smith and D. G. Altman. London, BMJ Publishing Group.

Dreier, M., B. Borutta, et al. (2010). Vergleich von Bewertungsinstrumenten für die Studienqualität von Primär- und Sekundärstudien zur Verwendung für HTA-Berichte im deutschsprachigen Raum. Schriftenreihe Health Technology Assessment, Bd. 102. Köln, DAHTA@DIMDI.

Denzin NK, Lincoln YS (2000): *The SAGE Handbook of Qualitative Research*, SAGE Publications Ltd., London

Dissemination, NHS Centre for Reviews and 2001, *Understating systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews*. 2nd edn

Donner A, Klar N (2004): Pitfalls of and controversies in cluster randomization trials. In: *Am J Public Health*, vol. 94, no. 3, pp. 416–22

Dreier, M., B. Borutta, et al. (2010). Vergleich von Bewertungsinstrumenten für die

Studienqualität von Primär- und Sekundärstudien zur Verwendung für HTA-Berichte im deutschsprachigen Raum. Schriftenreihe Health Technology Assessment, Bd. 102. Köln, DAHTA@DIMDI

Duval, S. and R. Tweedie (2000). "Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis." *Biometrics* 56(2): 455-463.

Dwan, K., C. Gamble, et al. (2010). "Assessing the potential for outcome reporting bias in a review: a tutorial." *Trials* 11: 52.

Edwards P, Clarke M, DiGuseppi C, Pratap S, Roberts I, Wentz R (2002): Identification of randomized controlled trials in systematic reviews: accuracy and reliability of screening records. In: *Stat Med*, vol. 21, no. 11, pp. 1635-40

Egger, M., G. Davey Smith, et al. (1997). "Bias in meta-analysis detected by a simple, graphical test." *BMJ* 315(7109): 629-634.

Egger, M., K. Dickersin, et al. (2001). Problems and limitations in conducting systematic reviews. *Systematic Reviews in Health Care: Meta-analysis in context*. Second edition. M. Egger, G. D. Smith and D. G. Altman. London, BMJ Publishing Group.

Egger, M. and G. D. Smith (2001). Principles of and procedures for systematic reviews. *Systematic Reviews in Health Care: Meta-analysis in context*. Second edition. M. Egger, G. D. Smith and D. G. Altman. London, BMJ Publishing Group.

Egger, M., G. D. Smith, et al. (2001). Rationale, potentials, and promise of systematic reviews. *Systematic Reviews in Health Care: Meta-analysis in context*. Second edition. M. Egger, G. D. Smith and D. G. Altman. London, BMJ Publishing Group.

Egger, M., G. D. Smith, et al. (2001). Systematic reviews of observational studies. *Systematic Reviews in Health Care: Meta-analysis in context*. Second edition. M. Egger, G. D. Smith and D. G. Altman. London, BMJ Publishing Group.

Elbourne, D. R., D. G. Altman, et al. (2002). "Meta-analyses involving cross-over trials: methodological issues." *Int J Epidemiol* 31(1): 140-149.

Fearon, W. F., P. A. Tonino, et al. (2007). "Rationale and design of the Fractional Flow Reserve versus Angiography for Multivessel Evaluation (FAME) study." *Am Heart J* 154(4): 632-636.

Felder-Puig R, Mad P, Gartlehner G (2009): Diagnostic studies. In: Wiener medizinische Wochenschrift (1946), vol. 159, no. 13-14, pp. 359-366

Gartlehner G, Hansen RA, Nissman D, Lohr KN, Carey TS (2006): A simple and valid tool distinguished efficacy from effectiveness studies. In: J Clin Epidemiol, vol. 59, no. 10, pp. 1040-8

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

Gartlehner G, Thieda P, Hansen RA, Morgan LC, Shumate JA, Nissman DB (2009): Inadequate reporting of trials compromises the applicability of systematic reviews. In: *International Journal of Technology Assessment in Health Care*, vol. 25, no. 3, pp. 323-330

Gartlehner G, Wild C, Mad P (2008): Systematic reviews and meta-analysis. In: Wiener medizinische Wochenschrift (1946), vol. 158, no. 3-4, pp. 127-133

Glasgow RE, Magid DJ, Beck A, Ritzwoller D, Estabrooks PA (2005): Practical clinical trials for translating research to practice: design and measurement recommendations. In: *Med Care*, vol. 43, no. 6, pp. 551-7

Glass, G. V. (1976). "Primary, Secondary, and Meta-Analysis of Research." *Educational Researcher* 5(10): 3-8.

Glasziou P, Vandenbroucke JP, Chalmers I (2004): Assessing the quality of research. In: *BMJ*, vol. 328, no. 7430, pp. 39-41

Glenny, A. M., D. G. Altman, et al. (2005). "Indirect comparisons of competing interventions." *Health Technol Assess* 9(26): 1-134, iii-iv.

Godwin M, Ruhland L, Casson I, MacDonald S, Delva D, Birtwhistle R, Lam M, Seguin R (2003): Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. In: *BMC Med Res Methodol*, vol. 3, no. 1, pp. 28

Gohler, A., J. L. Januzzi, et al. (2006). "A systematic meta-analysis of the efficacy and heterogeneity of disease management programs in congestive heart failure." *J Card Fail* 12(7): 554-567.

Green J & Thorogood N (2004): *Qualitative methods for health research*. SAGE Publications Ltd., London

Green LW, Glasgow RE (2006): Evaluating the relevance, generalization, and applicability of research: issues in external validation and translation methodology. In: *Eval Health Prof*, vol. 29, no. 1, pp. 126–53

Grimes DA, Schulz KF (2005): Compared to what? Finding controls for case–control studies. In: *Lancet*, vol. 365, no. 9468, pp. 1429–33

Grimes DA, Schulz KF (2002a): Bias and causal associations in observational research. In: *Lancet*, vol. 359, no. 9302, pp. 248–52

Grimes DA, Schulz KF (2002b): Cohort studies: marching towards outcomes. In: *Lancet*, vol. 359, no. 9303, pp. 341–5

Grimes DA, Schulz KF (2002c): An overview of clinical research: the lay of the land. In: *Lancet*, vol. 359, no. 9300, pp. 57–61

Grouven, U. (2010). *Crashkurs Meta-Analysen*. 55. GMDs Jahrestagung. Mannheim.

Guyatt G, Rennie D, Meade MO, Cook DJ (2008): In a case–control study, did the cases and control group have the same risk (chance) for being exposed in the past? In: *Users' Guide to the Medical Literature*, 2nd edn, *JAMAevidence*, , pp. 151–154

Guyatt G, Vist G, Falck–Ytter Y, Kunz R, Magrini N, Schunemann H (2006): An emerging consensus on grading recommendations? In: *ACP J Club*, vol. 144, no. 1, pp. A8–9

Helms PJ (2002): Real world! pragmatic clinical trials: what are they and what do they tell us? In: *Pediatr Allergy Immunol*, vol. 13, no. 1, pp. 4–9

Heritier SR, GebSKI VJ, Keech AC (2003): Inclusion of patients in clinical trial analysis: the intention–to–treat principle. In: *Med J Aust*, vol. 179, no. 8, pp. 438–40

Higgins, J. P., S. G. Thompson, et al. (2003). "Measuring inconsistency in meta-analyses." *BMJ* 327(7414): 557–560.

Higgins J., Green S (2008): *Cochrane Handbook for Systematic Reviews of Interventions– Version 5.0.1*

Higgins, J. P. T. and S. Green (2011). *Cochrane Handbook for Systematic Reviews of Interventions*, The Cochrane Collaboration.

Hopewell S, Clarke M, Lusher A, Lefebvre C, Westby M (2002): A comparison of handsearching versus MEDLINE searching to identify reports of randomized controlled trials. In: *Stat Med*, vol. 21, no. 11, pp. 1625–34

Horton J, Vandermeer B, Hartling L, Tjosvold L, Klassen TP, Buscemi N (2009): Systematic review data extraction: cross-sectional study showed that experience did not increase accuracy. In: *Journal of clinical epidemiology*

Hunter, J. E. and F. L. Schmidt (2000). "Fixed Effects vs. Random Effects Meta-Analysis Models: Implications for Cumulative Research Knowledge." *Int J Sel Assess* 8(4): 275-292.

Jaeschke R, Singer J, Guyatt GH (1989): Measurement of health status. Ascertaining the minimal clinically important difference. In: *Controlled clinical trials*, vol. 10, no. 4, pp. 407-415

Jerant, A. F., R. Azari, et al. (2001). "Reducing the cost of frequent hospital admissions for congestive heart failure: a randomized trial of a home telecare intervention." *Med Care* 39: 1234-1245.

Jones, C. M. and T. Athanasiou (2009). "Diagnostic accuracy meta-analysis: review of an important tool in radiological research and decision making." *Br J Radiol* 82(978): 441-446.

Juni P, Altman DG, Egger M (2001): Systematic reviews in health care: Assessing the quality of controlled clinical trials. In: *BMJ*, vol. 323, no. 7303, pp. 42-6

Juni P, Witschi A, Bloch R, Egger M (1999): The hazards of scoring the quality of clinical trials for meta-analysis. In: *JAMA*, vol. 282, no. 11, pp. 1054-60

Kitto SC, Chesters J, Grbich C (2008): Quality in qualitative research. In: *The Medical journal of Australia*, vol. 188, no. 4, pp. 243-246

Knottnerus, J. A., Ed. (2002). *The evidence base of clinical diagnosis*. London, BMJ Books.

Knottnerus JA, van Weel C, Muris JW (2002): Evaluation of diagnostic procedures. In: *BMJ (Clinical research ed.)*, vol. 324, no. 7335, pp. 477-480

Kotsopoulos, I. A., T. van Merode, et al. (2002). "Systematic review and meta-analysis of incidence studies of epilepsy and unprovoked seizures." *Epilepsia* 43(11): 1402-1409.

Krupski TL, Dahm P, Fesperman SF, Schardt CM (2008): How to perform a literature search. In: *J Urol*, vol. 179, no. 4, pp. 1264-70

- Kuper A, Lingard L, Levinson W (2008): Critically appraising qualitative research. In: *BMJ*, vol. 337, pp. a1035
- Lagakos SW (2006): Time-to-event analyses for long-term treatments--the APPROVe trial. In: *N Engl J Med*, vol. 355, no. 2, pp. 113-7
- Lang A, Edwards N, Fleischer A (2007): Empty systematic reviews: hidden perils and lessons learned. In: *J Clin Epidemiol*, vol. 60, no. 6, pp. 595-7
- Lau, J., E. M. Antman, et al. (1992). "Cumulative meta-analysis of therapeutic trials for myocardial infarction." *N Engl J Med* 327(4): 248-254.
- Levine M, Walter S, Lee H, Haines T, Holbrook A, Moyer V (1994): Users' guides to the medical literature. IV. How to use an article about harm. Evidence-Based Medicine Working Group. In: *JAMA : the journal of the American Medical Association*, vol. 271, no. 20, pp. 1615-1619
- Liamputtong P, Ezzy D (2005): *Qualitative research methods*. In: Oxford University Press
- Liberati, A., D. G. Altman, et al. (2009). "The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration." *PLoS Med* 6(7): e1000100.
- Lijmer, J. G. and P. M. M. Bossuyt (2009). "Various randomized designs can be used to evaluate medical tests." *J Clin Epidemiol* 62: 364-373.
- Lohr KN (2004): Rating the strength of scientific evidence: relevance for quality improvement programs. In: *Int J Qual Health Care*, vol. 16, no. 1, pp. 9-18
- Lohr KN, Carey TS (1999): Assessing "best evidence": issues in grading the quality of studies for systematic reviews. In: *Jt Comm J Qual Improv*, vol. 25, no. 9, pp. 470-9
- Lu, G. and A. E. Ades (2004). "Combination of direct and indirect evidence in mixed treatment comparisons." *Stat Med* 23: 3105-3124.
- MacLehose RR, Reeves BC, Harvey IM, Sheldon TA, Russell IT, Black AM (2000): A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies. In: *Health technology assessment (Winchester, England)*, vol. 4, no. 34, pp. 1-154
- Mathis S, Gartlehner G (2008): Cohort studies. In: *Wiener medizinische Wochenschrift (1946)*, vol. 158, no. 5-6, pp. 174-179

McAuley L, Pham B, Tugwell P, Moher D (2000): Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? In: *Lancet*, vol. 356, no. 9237, pp. 1228–31

McAlister, F. A. (2001). *Applying the results of systematic reviews at the bedside. Systematic Reviews in Health Care: Meta-analysis in context*. Second edition. M. Egger, G. D. Smith and D. G. Altman. London, BMJ Publishing Group.

Moher D, Liberati A, Tetzlaff J, Altman DG & PRISMA Group (2009): Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. In: *BMJ (Clinical research ed.)*, vol. 339, pp. b2535

Morton, S. C., J. L. Adams, et al. (2004). *Meta-regression Approaches: What, Why, When, and How? Technical Review 8* (Prepared by Southern California–RAND Evidence-based Practice Center, under Contract No 290–97–0001). Rockville, MD, Agency for Healthcare Research and Quality.

Mulrow CD (1994): Rationale for systematic reviews. In: *BMJ*, vol. 309, no. 6954, pp. 597–9

Oxman AD, Guyatt GH 1993, "The science of reviewing research", *Ann N Y Acad Sci*, vol. 703, pp. 125–33; discussion 133–4

Petitti, D. B. (2000). *Meta-Analysis, Decision Analysis, and Cost-Effectiveness Analysis*. New York, Oxford University Press, Inc.

Popay J, Rogers A, Williams G (1998): Rationale and standards for the systematic review of qualitative literature in health services research. In: *Qualitative health research*, vol. 8, no. 3, pp. 341–351

Pope C, Mays N (1995): Reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. In: *BMJ (Clinical research ed.)*, vol. 311, no. 6996, pp. 42–45

Pyett PM (2001): Innovation and compromise: responsibility and reflexivity in research with vulnerable groups. pp. 136–150

Reitsma, J. B., A. S. Glas, et al. (2005). "Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews." *J Clin Epidemiol* 58: 982–990.

Riegel, B., B. Carlson, et al. (2002). "Effect of a standardized nurse case-management telephone intervention on resource use in patients with chronic heart failure." *Arch Intern Med* 162: 705–712.

Riley, R. D., P. C. Lambert, et al. (2010). "Meta-analysis of individual participant data: rationale, conduct, and reporting." *BMJ* 340: c221.

Rothwell PM (2009): Commentary: External validity of results of randomized trials: disentangling a complex concept. In: *International journal of epidemiology*

Rothwell PM (2005): External validity of randomised controlled trials: "to whom do the results of this trial apply?". In: *Lancet*, vol. 365, no. 9453, pp. 82–93

Sackett DL, Deeks JJ, Altman DG (1996): Down with odds ratios! In: *Evidence-based medicine*, vol. 1, no. 6, pp. 164–166

Salanti, G., J. P. T. Higgins, et al. (2008). "Evaluation of networks of randomized trials." *Stat Methods Med Res* 17: 279–301.

Sampson M, McGowan J, Cogo E, Grimshaw J, Moher D, Lefebvre C (2009): An evidence-based practice guideline for the peer review of electronic search strategies. In: *Journal of clinical epidemiology*, vol. 62, no. 9, pp. 944–952

Schöffski, O. and J. M. Schulenburg (2008). *Gesundheitsökonomische Evaluationen*. Berlin Heidelberg, Springer-Verlag.

Schöttker, B., D. Lühmann, et al. (2009). *Indirekte Vergleiche von Therapieverfahren*. Schriftenreihe Health Technology Assessment, Bd. 88. Köln, DAHTA@DIMDI.

Schulz KF, Grimes DA (2002): Case-control studies: research in reverse. In: *Lancet*, vol. 359, no. 9304, pp. 431–4

Schumacher, M. and G. Schulgen (2008). *Methodik klinischer Studien. Methodische Grundlagen der Planung, Durchführung und Auswertung*. Berlin Heidelberg, Springer-Verlag.

Schünemann H (2009): GRADE: Von der Evidenz zur Empfehlung Beschreibung des Systems und Lösungsbeitrag zur Übertragbarkeit von Studienergebnissen. In: *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* Jahrgang 103, Heft 6, 09–2009, vol. 103, no. 6, pp. 391–400

Schwartz D, Lellouch J (1967): Explanatory and pragmatic attitudes in therapeutical trials. In: *Journal of chronic diseases*, vol. 20, no. 8, pp. 637–648

Shea, B. J., J. M. Grimshaw, et al. (2007). "Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews." *BMC Med Res Methodol* 7: 10.

Simera, I., D. Moher, et al. (2010). "A catalogue of reporting guidelines for health research." *Eur J Clin Investig* 40(1): 35–53.

Smith, G. D. and M. Egger (2001). *Going beyond the grand mean: subgroup analysis in meta-analysis of randomised trials. Systematic Reviews in Health Care: Meta-analysis in context. Second edition.* M. Egger, G. D. Smith and D. G. Altman. London, BMJ Publishing Group.

Song, F., Y. K. Loke, et al. (2009). "Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews." *BMJ* 338: b1147.

Spencer L (2003): *Quality in Qualitative Evaluation: A Framework for Assessing Research Evidence* [Homepage of Government Chief Social Researcher's Office, Cabinet Office], [Online]. Available: http://www.gsr.gov.uk/downloads/evaluating_policy/a_quality_framework.pdf

Sterne, J. (2009). *Meta-Analysis: An Updated Collection from the Stata Journal*, Stata Press.

Sterne, J. A. C., M. Egger, et al. (2001). *Investigating and dealing with publication and other biases. Systematic Reviews in Health Care: Meta-analysis in context. Second edition.* M. Egger, G. D. Smith and D. G. Altman. London, BMJ Publishing Group.

Sterne, J. A. C., D. Gavaghan, et al. (2000). "Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature." *J Clin Epidemiol* 53(11): 1119–1129.

Stroup, D. F., J. A. Berlin, et al. (2000). "Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group." *JAMA* 283(15): 2008–2012.

Sutton, A., A. E. Ades, et al. (2008). "Use of indirect and mixed treatment comparisons for technology assessment." *Pharmacoeconomics* 26(9): 753–767.

Sutton, A. J. and K. R. Abrams (2001). "Bayesian methods in meta-analysis and evidence synthesis." *Stat Methods Med Res* 10(4): 277–303.

Sutton, A. J., K. R. Abrams, et al. (2000). *Methods for Meta-Analysis in Medical Research.* Chichester, West Sussex, UK, John Wiley & Sons, Ltd.

Sutton, A. J. and J. P. Higgins (2008). "Recent developments in meta-analysis." *Stat Med* 27(5): 625–650.

Thompson, S. G. and J. P. T. Higgins (2002). "How should meta-regression analyses be undertaken and interpreted? ." *Stat Med* 21: 1559–1573.

Tonino, P. A., B. De Bruyne, et al. (2009). "Fractional flow reserve versus angiography for guiding percutaneous coronary intervention." *N Engl J Med* 360(3): 213–224.

Tricco, A. C., J. Tetzlaff, et al. (2010). "The art and science of knowledge synthesis." *J Clin Epidemiol*.

U.S. Department of Health & Human Services 10.12.2009–last update. Guidance for Industry on Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Available: <http://www.fda.gov/ohrms/dockets/dockets/06d0044/06d0044/06d044.htm>

Vandenbroucke JP, von Elm E, Altman DG, Gotzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M (2007): Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. In: *Ann Intern Med*, vol. 147, no. 8, pp. W163–94

Wells GA, Shea B, Peterson J, Welch V, Losos M, Tugwell P. The Newcastle Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Available: http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm Dezember, 2009

Wells, G. A., S. A. Sultan, et al. (2009). Indirect Evidence: Indirect Treatment Comparisons in Meta-Analysis. Ottawa, Canadian Agency for Drugs and Technologies in Health (CADTH).

West S, King V, Carey TS, Lohr KN, McKoy N, Sutton SF, Lux L (2002): Systems to rate the strength of scientific evidence. In: *Evid Rep Technol Assess (Summ)*, , no. 47, pp. 1–11

Whitehead, A. (2002). *Meta-Analysis of Controlled Clinical Trials*. Chichester, West Sussex, UK, John Wiley & Sons, Ltd.

Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J (2003): The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. In: *BMC Med Res Methodol*, vol. 3, pp. 25

Whiting PF, Weswood ME, Rutjes AW, Reitsma JB, Bossuyt PN, Kleijnen J (2006): Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. In: *BMC medical research methodology*, vol. 6, pp. 9

Whitlock EP, Lin JS, Chou R, Shekelle P, Robinson KA (2008): Using existing systematic reviews in complex systematic reviews. In: *Ann Intern Med*, vol. 148, no. 10, pp. 776–82

Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, Gluud C, Martin RM, Wood AJ, Sterne JA (2008): Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. In: *BMJ (Clinical research ed.)*, vol. 336, no. 7644, pp. 601–605

Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, Oxman AD, Moher D, CONSORT group & Pragmatic Trials in Healthcare (Practihc) group (2008): Improving the reporting of pragmatic trials: an extension of the CONSORT statement. In: *BMJ (Clinical research ed.)*, vol. 337, pp. a2390

14.6 Literatur zu Kapitel 8 Ökonomische Bewertung

14.6.1 Literatur zu Kapitel 8.1. Ökonomische Übersichtsarbeit

Aguiar-Ibanez, R., J. Nixon, et al. (2005). "Economic evaluation databases as an aid to healthcare decision-makers and researchers." *Pharmacoconomics and Outcomes Research* 5: 721–722.

Altman, D. G. (2001). *Systematic reviews of evaluations of prognostic variables. Systematic Reviews in Health Care: Meta-analysis in context. Second edition.* M. Egger, G. D. Smith and D. G. Altman. London, BMJ Publishing Group.

Anderson, R. (2009). "Systematic Reviews of Economic Evaluations: Utility or Futility." *Health Economics* DOI/10.000/hec.1486.

Anderson, R. and I. Shemilt (2010). *The role of economic perspectives and evidence in systematic review. Evidence-based decisions and economics: health care, social welfare, education and criminal justice.* I. Shemilt, M. Mugford, L. Vale, K. Marsh and C. Donaldson. Oxford, Wiley-Blackwell.

Anderson, R. M. and R. M. May (1991). *Infectious diseases of humans: dynamics and control.* Oxford, UK, Oxford University Press.

Auchincloss, A. H., R. L. Riolo, et al. (2011). "An agent-based model of income inequalities in diet in the context of residential segregation." *Am J Prev Med* 40(3): 303–311.

Axelrod, R. (1997). *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*. Princeton, Princeton University Press.

Banks, J., J. S. Carson, et al. (2005). *Discrete-Event System Simulation*. Upper Saddle River, New Jersey, Pearson Prentice Hall.

Barton, P., S. Bryan, et al. (2004). "Modelling in the economic evaluation of health care: selecting the appropriate approach." *J Health Serv Res Policy* 9(2): 110–118.

Beck, J. R. and S. G. Pauker (1983). "The Markov Process in Medical Prognosis." *Medical Decision Making* 3: 419–458.

Begg, C. B. and M. Mazumdar (1994). "Operating characteristics of a rank correlation test for publication bias." *Biometrics* 50(4): 1088–1101.

Berry, D. A., K. A. Cronin, et al. (2005). "Effect of screening and adjuvant therapy on mortality from breast cancer." *N Engl J Med* 353(17): 1784–1792.

Beutels, P., P. A. Scuffham, et al. (2008). "Funding of drugs: do vaccines warrant a different approach?" *The Lancet Infectious Diseases* 8(11): 727–733.

Beutels, P., E. Van Doorslaer, et al. (2003). "Methodological issues and new developments in the economic evaluation of vaccines." *Expert Rev Vaccines* 2(5): 649–660.

Billari, F. C., T. Fent, et al. (2006). *Agent-Based Computational Modelling: Introduction*. Agent-Based Computational Modelling. Heidelberg, Physica – Verlag: 1 – 16.

Bonabeau, E. (2002). "Agent-based modeling: methods and techniques for simulating human systems." *Proc Natl Acad Sci U S A* 99 Suppl 3: 7280–7287.

Bonabeau, E. (2002). "Predicting the unpredictable." *Harv Bus Rev* 80(3): 109–116, 134.

Borenstein, M., L. V. Hedges, et al. (2009). *Introduction to Meta-Analysis*. Chichester, West Sussex, UK, John Wiley & Sons, Ltd.

Brailford, S. (2007). Tutorial: Advances and challenges in healthcare simulation modeling. Winter Simulation Conference.

Brennan, A., S. E. Chick, et al. (2006). "A taxonomy of model structures for economic evaluation of health technologies." *Health Econ* 15(12): 1295–1310.

Briggs, A., E. Fenwick, et al. (2011). DRAFT – Model Parameter Estimation and Uncertainty: A Report of the ISPOR–SMDM Modeling

Good Research Practices Task Force Working Group – Part 3.

Briggs, A., M. Sculpher, et al. (1994). "Uncertainty in the Economic Evaluation of Health Care Technologies: The Role of Sensitivity Analysis." *Health Economics* 3: 95–104.

Briggs, A. H. (2000). "Handling uncertainty in cost-effectiveness models." *Pharmacoeconomics* 17(5): 479–500.

Brisson, M. and W. J. Edmunds (2003). "Economic evaluation of vaccination programs: The impact of herd-immunity." *Medical Decision Making* 23(1): 76–82.

Brisson, M. and W. J. Edmunds (2003). "Varicella vaccination in England and Wales: cost-utility analysis." *Arch Dis Child* 88(10): 862–869.

Brown, B. N., I. M. Price, et al. (2011). "An agent-based model of inflammation and fibrosis following particulate exposure in the lung." *Math Biosci.*

Brunetti, M., F. Ruiz, et al. (2010). Grading economic evidence. Evidence-based decisions and economics: health care, social welfare, education and criminal justice. I. Shemilt, M. Mugford, L. Vale, K. Marsh and C. Donaldson. Oxford, Wiley-Blackwell.

Canonica, M., G. Plu-Bureau, et al. (2008). "Hormone replacement therapy and risk of venous thromboembolism in postmenopausal women: systematic review and meta-analysis." *BMJ* 336(7655): 1227–1231.

Caro, J. J. (2005). "Pharmacoeconomic analyses using discrete event simulation." *Pharmacoeconomics* 23(4): 323–332.

Centre for Reviews and Dissemination (2009). *Systematic Reviews*. York, CRD, University of York.

Chiou, C., J. W. Hay, et al. (2003). "Development and validation of a grading system for the quality of cost-effectiveness studies." *Medical Care* 41: 32–44.

Clarke, M. J. and L. A. Stewart (2001). Obtaining individual patient data from random-

ised controlled trials. *Systematic Reviews in Health Care: Meta-analysis in context*. Second edition. M. Egger, G. D. Smith and D. G. Altman. London, BMJ Publishing Group.

Cooper, K., S. C. Brailsford, et al. (2007). "Choice of Modelling Technique for Evaluating Health Care Interventions." *The Journal of the Operational Research Society* 58(2): 168–176.

Craig, D. and S. Rice, Eds. (2007). *NHS Economic Evaluation Database Handbook*. York, Centre for Reviews and Dissemination.

Curtin, F., D. G. Altman, et al. (2002). "Meta-analysis combining parallel and cross-over clinical trials. I: Continuous outcomes." *Stat Med* 21(15): 2131–2144.

Curtin, F., D. Elbourne, et al. (2002). "Meta-analysis combining parallel and cross-over clinical trials. II: Binary outcomes." *Stat Med* 21(15): 2145–2159.

Curtin, F., D. Elbourne, et al. (2002). "Meta-analysis combining parallel and cross-over clinical trials. III: The issue of carry-over." *Stat Med* 21(15): 2161–2173.

Damm, O., M. Nocon, et al. (2009). *Impfung gegen humane Papillomaviren (HPV) zur Prävention HPV 16/18 induzierter Zervixkarzinome und derer Vorstufen*. Köln, DIMDI.

Deeks, J. J. (2001). *Systematic reviews of evaluations of diagnostic and screening tests*. *Systematic Reviews in Health Care: Meta-analysis in context*. Second edition. M. Egger, G. D. Smith and D. G. Altman. London, BMJ Publishing Group.

Deeks, J. J., D. G. Altman, et al. (2001). *Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis*. *Systematic Reviews in Health Care: Meta-analysis in context*. Second edition. M. Egger, G. D. Smith and D. G. Altman. London, BMJ Publishing Group.

Deeks, J. J., D. G. Altman, et al. (2001). *Systematic Reviews in Health Care: Meta-analysis in context*. London, BMJ Publishing Group.

Dreier, M., B. Borutta, et al. (2010). *Vergleich von Bewertungsinstrumenten für die Studienqualität von Primär- und Sekundärstudien zur Verwendung für HTA-Berichte im deutschsprachigen Raum*. Schriftenreihe Health Technology Assessment, Bd. 102. Köln, DAHTA@DIMDI.

Drummond, M., C. Chevat, et al. (2007). "Do we fully understand the economic value of vaccines?" *Vaccine* 25(32): 5945–5957.

Drummond, M. and T. O. Jefferson (1996). "Guidelines for authors and peer reviewers of economic submissions to the BMJ." *British Medical Journal* 313: 275–283.

Drummond, M., M. Sculpher, et al., Eds. (2005). *Methods for the Economic Evaluation of Health Care Programmes*. New York, Oxford University Press.

Drummond, M., H. Weatherly, et al. (2007). *Assessing the Challenges of Applying Standard Methods of Economic Evaluation to Public Health Interventions*. York, Public Health Research Consortium.

Drummond, M. F. and L. Davies (1991). "Economic Analysis Alongside Clinical Trials." *International Journal of Technology Assessment in Health Care* 7(4): 561–573.

Drummond, M. F., B. O'Brien, et al. (1997). *Methods for the Economic Evaluation of Health Care Programmes*. New York, Oxford University Press Inc.

Drummond, M. F., J. S. Schwartz, et al. (2008). "Key principles for the improved conduct of health technology assessments for resource allocation decisions." *Int J Technol Assess Health Care* 24(3): 244–258; discussion 362–248.

Duval, S. and R. Tweedie (2000). "Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis." *Biometrics* 56(2): 455–463.

Dwan, K., C. Gamble, et al. (2010). "Assessing the potential for outcome reporting bias in a review: a tutorial." *Trials* 11: 52.

Eddy, D. M. and L. Schlessinger (2003). "Validation of the archimedes diabetes model." *Diabetes Care* 26(11): 3102–3110.

Edmunds, W. J., G. F. Medley, et al. (1999). "Evaluating the cost-effectiveness of vaccination programmes: A dynamic perspective." *Statistics in Medicine* 18(23): 3263–3282.

Egger, M., G. Davey Smith, et al. (1997). "Bias in meta-analysis detected by a simple, graphical test." *BMJ* 315(7109): 629–634.

Egger, M., K. Dickersin, et al. (2001). *Problems and limitations in conducting systematic reviews. Systematic Reviews in Health Care: Meta-analysis in context*. Second edition. M. Egger, G. D. Smith and D. G. Altman. London, BMJ Publishing Group.

Egger, M. and G. D. Smith (2001). *Principles of and procedures for systematic re-*

views. *Systematic Reviews in Health Care: Meta-analysis in context*. Second edition. M. Egger, G. D. Smith and D. G. Altman. London, BMJ Publishing Group.

Egger, M., G. D. Smith, et al. (2001). Rationale, potentials, and promise of systematic reviews. *Systematic Reviews in Health Care: Meta-analysis in context*. Second edition. M. Egger, G. D. Smith and D. G. Altman. London, BMJ Publishing Group.

Egger, M., G. D. Smith, et al. (2001). Systematic reviews of observational studies. *Systematic Reviews in Health Care: Meta-analysis in context*. Second edition. M. Egger, G. D. Smith and D. G. Altman. London, BMJ Publishing Group.

Elbourne, D. R., D. G. Altman, et al. (2002). "Meta-analyses involving cross-over trials: methodological issues." *Int J Epidemiol* 31(1): 140–149.

Evers, S., M. Goossens, et al. (2005). "Criteria for assessment of methodological quality of economic evaluations: Consensus on Health Economic Criteria." *International Journal of Technology Assessment in Health Care* 21: 240–245.

Fearon, W. F., P. A. Tonino, et al. (2007). "Rationale and design of the Fractional Flow Reserve versus Angiography for Multivessel Evaluation (FAME) study." *Am Heart J* 154(4): 632–636.

Fine, P. E. (1993). "Herd immunity: history, theory, practice." *Epidemiol Rev* 15(2): 265–302.

Gartlehner, G. and C. G. Moore (2008). "Direct versus indirect comparisons: a summary of the evidence." *Int J Technol Assess Health Care* 24(2): 170–177.

Gerard, K., J. Seymore, et al. (2000). "A tool to improve quality of reporting published economic analyses." *International Journal of Technology Assessment in Health Care* 16: 100–110.

Gerkens, S., R. Crott, et al. (2008). "Comparison of three instruments assessing the quality of economic evaluations: A practical exercise on economic evaluations of the surgical treatment of obesity." *International Journal of Technology Assessment in Health Care* 24(3): 318–325.

Glass, G. V. (1976). "Primary, Secondary, and Meta-Analysis of Research." *Educational Researcher* 5(10): 3–8.

Glenny, A. M., D. G. Altman, et al. (2005). "Indirect comparisons of competing interventions." *Health Technol Assess* 9(26): 1–134, iii–iv.

Gohler, A., J. L. Januzzi, et al. (2006). "A systematic meta-analysis of the efficacy and heterogeneity of disease management programs in congestive heart failure." *J Card Fail* 12(7): 554–567.

Göhler, A., J. L. Januzzi, et al. (2006). "A systematic meta-analysis of the efficacy and heterogeneity of disease management programs in congestive heart failure." *J Card Fail* 12(7): 554–567.

Goldstein, M., Tsevat, J. (2003). *Interactive Textbook on Clinical Symptom Research Chapter 24: Assessing Desirability of Outcomes States for Medical Decision Making and Cost-effectiveness Analysis*. M. Mitchell, Lynn, J., NIH. Chapter 24.

Gonzalez-Perez, J. (2002). "Developing a scoring system to quality assess economic evaluations." *European Journal of Health Economics* 3: 131–136.

Groot Koerkamp, B., M. C. Weinstein, et al. (2010). "Uncertainty and patient heterogeneity in medical decision models." *Med Decis Making* 30(2): 194–205.

Group, T. M. H. M. (2007). "Computer modeling of diabetes and its complications: a report on the Fourth Mount Hood Challenge Meeting." *Diabetes Care* 30(6): 1638–1646.

Grouven, U. (2010). *Crashkurs Meta-Analysen*. 55. GMDS Jahrestagung. Mannheim.

Grutters, J., S. Seferina, et al. (2010). *Bridging trial and decision: a checklist to determine the generalizability of HTAs for resource allocation decisions*. Health Technology International Conference. Dublin.

Higgins, J. P., S. G. Thompson, et al. (2003). "Measuring inconsistency in meta-analyses." *BMJ* 327(7414): 557–560.

Higgins, J. P. T. and S. Green (2011). *Cochrane Handbook for Systematic Reviews of Interventions*, The Cochrane Collaboration. Version 5.1.0 [updated March 2011].

Hunink, M. G., P. P. Glasziou, et al. (2001). *Decision making in health and medicine. Integrating evidence and values*. Cambridge, Cambridge University Press.

Hunter, J. E. and F. L. Schmidt (2000). "Fixed Effects vs. Random Effects Meta-Analysis Models: Implications for Cumulative Research Knowledge." *Int J Sel Assess* 8(4): 275–292.

IQWiG (2009). *Allgemeine Methoden zur Bewertung von Verhältnissen zwischen*

Nutzen und Kosten. Version 1.0 vom 12.10.2009., Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG).

ISPOR-SMDM Modeling Good Research Practices Task Force, I. (2011). "Presentations."

Jefferson, T., V. Demicheli, et al. (2002). "Quality of Systematic Reviews of Economic Evaluations in Health Care." *Journal of the American Medical Association* 287(2): 2809–2812.

Jerant, A. F., R. Azari, et al. (2001). "Reducing the cost of frequent hospital admissions for congestive heart failure: a randomized trial of a home telecare intervention." *Med Care* 39: 1234–1245.

Jones, C. M. and T. Athanasiou (2009). "Diagnostic accuracy meta-analysis: review of an important tool in radiological research and decision making." *Br J Radiol* 82(978): 441–446.

Karnon, J. and T. Vanni (2011). "Calibrating models in economic evaluation: a comparison of alternative measures of goodness of fit, parameter search strategies and convergence criteria." *Pharmacoeconomics* 29(1): 51–62.

Khan, K. S., G. Riet ter, et al., Eds. (2001). *Undertaking Systematic Reviews of Research on Effectiveness – CRD's Guidelines for those Carrying out or Commissioning Reviews*. York, NHS Centre for Reviews and Disseminations.

Kim, J. J., K. M. Kuntz, et al. (2007). "Multiparameter calibration of a natural history model of cervical cancer." *Am J Epidemiol* 166(2): 137–150.

Kim, S. Y. and S. J. Goldie (2008). "Cost-effectiveness analyses of vaccination programmes : a focused review of modelling approaches." *Pharmacoeconomics* 26(3): 191–215.

Kleijnen, J. P. C. (1995). "Verification and Validation of simulation models." *European Journal of Operational Research* 82(1): 145–162.

Klein, T., Scholz, Höer, Albrecht, Häussler, Schmidt, Jahn, Staudt, Farrenkopf, Lux, Weegen, Thomas, Wasem, Sussex, Shah, Butler, Mühlberger, Gothe, Schwarzer, Schnell-Inderst, Sroczynski, Siebert (2010). *Gutachten zur Verbesserung der Wirtschaftlichkeit von Impfstoffen in Deutschland B. f. G. (BMG)*. Berlin, Germany.

Knottnerus, J. A., Ed. (2002). *The evidence base of clinical diagnosis*. London, BMJ Books.

Knottnerus, J. A., C. van Weel, et al. (2002). "Evaluation of diagnostic procedures." *BMJ* 324(7335): 477–480.

Kong, C. Y., P. M. McMahon, et al. (2009). "Calibration of disease simulation model using an engineering approach." *Value Health* 12(4): 521–529.

Koopman, J. S. (2005). "Infection transmission science and models." *Jpn J Infect Dis* 58(6): S3–8.

Kotsopoulos, I. A., T. van Merode, et al. (2002). "Systematic review and meta-analysis of incidence studies of epilepsy and unprovoked seizures." *Epilepsia* 43(11): 1402–1409.

Lau, J., E. M. Antman, et al. (1992). "Cumulative meta-analysis of therapeutic trials for myocardial infarction." *N Engl J Med* 327(4): 248–254.

Lee, B. Y., S. T. Brown, et al. (2010). "A computer simulation of vaccine prioritization, allocation, and rationing during the 2009 H1N1 influenza pandemic." *Vaccine* 28(31): 4875–4879.

Leidl, R. (1998). *Der Effizienz auf der Spur: Eine Einführung in die ökonomische Evaluation. Das Public Health Buch: Gesundheit und Gesundheitswesen.* F. W. Schwartz, et al., Urban & Schwarzenberg. 18: 346–369.

Liberati, A., D. G. Altman, et al. (2009). "The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration." *PLoS Med* 6(7): e1000100.

Lijmer, J. G. and P. M. M. Bossuyt (2009). "Various randomized designs can be used to evaluate medical tests." *J Clin Epidemiol* 62: 364–373.

Lu, G. and A. E. Ades (2004). "Combination of direct and indirect evidence in mixed treatment comparisons." *Stat Med* 23: 3105–3124.

Macal, C. M. and M. J. North (2005). Tutorial on agent-based modeling and simulation. Proceedings of the 37th conference on Winter simulation. Orlando, Florida, Winter Simulation Conference: 2–15.

Macal, C. M. and M. J. North (2006). Tutorial on agent-based modeling and simulation part 2: how to model with agents. Proceedings of the 38th conference on Winter simulation. Monterey, California, Winter Simulation Conference: 73–83.

Mauskopf, J., S. Sullivan, et al. (2006). "Principles of Good Practice for Budget Impact Analysis: Report of the ISPOR Task Force on Good Research Practice – Budget Impact Analysis." *Value Health* 10(5).

McAlister, F. A. (2001). *Applying the results of systematic reviews at the bedside. Systematic Reviews in Health Care: Meta-analysis in context. Second edition.* M. Egger, G. D. Smith and D. G. Altman. London, BMJ Publishing Group.

Moher, D., A. Liberati, et al. (2009). "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement." *PLoS Med* 6(7): e1000097.

Morton, S. C., J. L. Adams, et al. (2004). *Meta-regression Approaches: What, Why, When, and How? Technical Review 8* (Prepared by Southern California-RAND Evidence-based Practice Center, under Contract No 290-97-0001). Rockville, MD, Agency for Healthcare Research and Quality.

Napper, M. and J. Varney (2005). *Etext on Health Technology Assessment (HTA) Information Resources. Chapter 11: Health Economics Information.*, U.S. National Library of Medicine.

National Institut for Health and Clinical Excellence (2009). *The guidelines manual.* London, National Institute for Health and Clinical Excellence.

Nixon, J., K. S. Khan, et al. (2001). "Summarising economic evaluations in systematic reviews: a new approach." *British Medical Journal* 322: 1596-1598.

North, M. J. and C. M. Macal (2007). *Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation*, Oxford University Press, Inc.

Petitti, D. B. (2000). *Meta-Analysis, Decision Analysis, and Cost-Effectiveness Analysis.* New York, Oxford University Press, Inc.

Philips, Z., L. Bojke, et al. (2006). "Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment." *Pharmacoeconomics* 24(4): 355-371.

Philips, Z., L. Ginnelly, et al. (2004). "A review of guidelines for good practice in decision-analytical modelling in health technology assessment." *Health Technology Assessment* 8(36): 1-158.

Pignone, M., S. Saha, et al. (2005). "Challenges in Systematic Reviews of Economic

Analyses." *Annals of Internal Medicine* 142: 1073–1079.

Pitman, R., J. Edmunds, et al. (2011). DRAFT – Dynamic Transmission Modelling: A Report of the ISPOR–SMDM Modelling Good Research Practices Task Force Working Group – Part 7.

Ramsey, S., R. Willke, et al. (2005). "Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT–CEA Task Force report." *Value Health* 8(5): 521–533.

Ratcliffe, J., T. Young, et al. (2001). "A simulation modelling approach to evaluating alternative policies for the management of the waiting list for liver transplantation." *Health Care Manag Sci* 4(2): 117–124.

Rein, D. B., J. B. Saaddine, et al. (2007). "Cost-effectiveness of vitamin therapy for age-related macular degeneration." *Ophthalmology* 114(7): 1319–1326.

Reitsma, J. B., A. S. Glas, et al. (2005). "Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews." *J Clin Epidemiol* 58: 982–990.

Riegel, B., B. Carlson, et al. (2002). "Effect of a standardized nurse case-management telephone intervention on resource use in patients with chronic heart failure." *Arch Intern Med* 162: 705–712.

Riley, R. D., P. C. Lambert, et al. (2010). "Meta-analysis of individual participant data: rationale, conduct, and reporting." *BMJ* 340: c221.

Roberts, M., M. Chambers, et al. (2011). DRAFT – Conceptual Modeling: A Report of the ISPOR–SMDM Modeling Good Research Practices Task Force Working Group – Part 2.

Rutter, C. M., D. L. Miglioretti, et al. (2009). "Bayesian Calibration of Microsimulation Models." *J Am Stat Assoc* 104(488): 1338–1350.

Rutter, C. M., A. M. Zaslavsky, et al. (2011). "Dynamic microsimulation models for health outcomes: a review." *Med Decis Making* 31(1): 10–18.

Salanti, G., J. P. T. Higgins, et al. (2008). "Evaluation of networks of randomized trials." *Stat Methods Med Res* 17: 279–301.

Sander, B., A. Nizam, et al. (2009). "Economic evaluation of influenza pandemic mitigation strategies in the United States using a stochastic microsimulation trans-

mission model." *Value Health* 12(2): 226–233.

Sassi, F., L. Archard, et al. (2002). "Searching Literature Databases for Health Care Economic Evaluations." *Medical Care* 40(5): 387–394.

Schöffski, O. and J. M. Graf von der Schulenburg (2007). *Gesundheitsökonomische Evaluation*. Berlin, Springer.

Schöffski, O. and J. M. Schulenburg (2008). *Gesundheitsökonomische Evaluationen*. Berlin Heidelberg, Springer-Verlag.

Schöttker, B., D. Lühmann, et al. (2009). *Indirekte Vergleiche von Therapieverfahren*. Schriftenreihe Health Technology Assessment, Bd. 88. Köln, DAHTA@DIMDI.

Schumacher, M. and G. Schulgen (2008). *Methodik klinischer Studien. Methodische Grundlagen der Planung, Durchführung und Auswertung*. Berlin Heidelberg, Springer-Verlag.

Sculpher, M., E. Fenwick, et al. (2000). "Assessing quality in decision analytic cost-effectiveness models. A suggested framework and example of application." *Pharmacoeconomics* 17(5): 461–477.

Sculpher, M., F. Pang, et al. (2004). "Generalisability in economic evaluation studies in health care. a review and case studies." *Health Technology Assessment* 8: 1–192.

Sculpher, M. J., K. Claxton, et al. (2006). "Whither trial-based economic evaluation for health care decision making?" *Health Econ* 15(7): 677–687.

Shea, B. J., J. M. Grimshaw, et al. (2007). "Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews." *BMC Med Res Methodol* 7: 10.

Shechter, S. M., C. L. Bryce, et al. (2005). "A clinically based discrete-event simulation of end-stage liver disease and the organ allocation process." *Med Decis Making* 25(2): 199–209.

Shemilt, I., M. Mugford, et al. (2009). *Incorporating Economic Evidence*. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.0.2. J. P. T. Higgins and S. Green, The Cochrane Collaboration.

Shemilt, I., M. Mugford, et al. (2008). Chapter 15: Incorporating economics evidence. *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.0.1 (updated September 2008). J. P. T. Higgins and C. Green, The Cochrane Collaboration.

Siebert, U. (2003). Transparente Entscheidungen in Public Health mittels systematischer Entscheidungsanalyse. Das Public Health Buch. Gesundheit fördern, Krankheit verhindern. F. W. Schwartz, B. Badura, R. Leidlet al. München, Urban & Fischer.

Siebert, U. (2003). "The use of decision analysis in health economic evaluation [German]." Informatik, Biometrie und Epidemiologie in Medizin und Biologie 34(2): 112-124.

Siebert, U. (2003). "When should decision-analytic modeling be used in the economic evaluation of health care? [Editorial]." European Journal of Health Economics 4(3): 143-150.

Siebert, U. (2009). Antiviral Therapy of Chronic Hepatitis C –

A Pilot Study Applying the "Methods for Assessment of the Relation of Benefits to Costs in the German Statutory Health Care System".

Siebert, U., O. Alagoz, et al. (2011). DRAFT – State-Transition Modeling: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group – Part 5.

Siebert, U., C. Behrend, et al. (1999). Entwicklung eines Kriterienkatalogs zur Beschreibung und Bewertung ökonomischer Evaluationsstudien in Deutschland. Ansätze und Methoden der ökonomischen Evaluation. Eine internationale Perspektive. R. Leidl, J. M. Graf von der Schulenburg and J. Wasem. Baden-Baden, Nomos Verlag.

Siebert, U. and T. Kurth (2000). Lebensqualität als Parameter von medizinischen Entscheidungsanalysen. Lebensqualität und Gesundheitsökonomie in der Medizin. U. Ravens-Sieberer, A. Cieza, N. von Steinbüchel and M. Bullinger. Landsberg, Ecomed: 365-392.

Siebert, U., N. Mühlberger, et al. Evidenzsynthese: Meta-Analysen und Entscheidungsanalysen – Draft for 4th Edition. Gesundheitsökonomische Evaluationen

O. Schöffski and J.-M. Graf von der Schulenburg. Berlin, Springer.

Siebert, U., N. Mühlberger, et al. (2000). Desk Research. Gesundheitsökonomische Evaluationen. O. Schöffski and J.-M. Graf v. d. Schulenburg. Berlin, Springer.

Siebert, U., N. Mühlberger, et al. (2000). Formale Entscheidungsanalyse.

Gesundheitsökonomische Evaluationen. O. Schöffski and J.-M. Graf v. d. Schulenburg. Berlin, Springer: 88–120.

Simera, I., D. Moher, et al. (2010). "A catalogue of reporting guidelines for health research." *Eur J Clin Invest* 40(1): 35–53.

Smith, G. D. and M. Egger (2001). *Going beyond the grand mean: subgroup analysis in meta-analysis of randomised trials. Systematic Reviews in Health Care: Meta-analysis in context. Second edition.* M. Egger, G. D. Smith and D. G. Altman. London, BMJ Publishing Group.

Song, F., Y. K. Loke, et al. (2009). "Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews." *BMJ* 338: b1147.

Sonnenberg, F. A. and J. R. Beck (1993). "Markov models in medical decision making: a practical guide." *Med Decis Making* 13(4): 322–338.

Stahl, J. E. (2008). "Modelling methods for pharmacoeconomics and health technology assessment: an overview and guide." *Pharmacoeconomics* 26(2): 131–148.

Stahl, J. E., K. L. Furie, et al. (2003). "Stroke: Effect of implementing an evaluation and treatment protocol compliant with NINDS recommendations." *Radiology* 228(3): 659–668.

Stahl, J. E., J. Karnon, et al. (2011). *DRAFT- Modeling Discrete Event Simulation: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group – Part 6.*

Stanley, T. D. (2010). *Meta-regression models of economics and medical research. Evidence-based decisions and economics: health care, social welfare, education and criminal justice.* I. Shemilt, M. Mugford, L. Vale, K. Marsh and C. Donaldson. Oxford, Wiley-Blackwell.

Statistik Austria (2005). "Bruttoinlandsprodukt je Einwohner in EUR." Retrieved November, 2005, from http://www.statistik.at/statistische_uebersichten/deutsch/k16.shtml.

Sterne, J. (2009). *Meta-Analysis: An Updated Collection from the Stata Journal*, Stata Press.

Sterne, J. A. C., M. Egger, et al. (2001). *Investigating and dealing with publication and other biases. Systematic Reviews in Health Care: Meta-analysis in context. Second*

edition. M. Egger, G. D. Smith and D. G. Altman. London, BMJ Publishing Group.

Sterne, J. A. C., D. Gavaghan, et al. (2000). "Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature." *J Clin Epidemiol* 53(11): 1119–1129.

Stout, N. K., A. B. Knudsen, et al. (2009). "Calibration methods used in cancer simulation models and suggested reporting guidelines." *Pharmacoeconomics* 27(7): 533–545.

Stroup, D. F., J. A. Berlin, et al. (2000). "Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group." *JAMA* 283(15): 2008–2012.

Sutton, A., A. E. Ades, et al. (2008). "Use of indirect and mixed treatment comparisons for technology assessment." *Pharmacoeconomics* 26(9): 753–767.

Sutton, A. J. and K. R. Abrams (2001). "Bayesian methods in meta-analysis and evidence synthesis." *Stat Methods Med Res* 10(4): 277–303.

Sutton, A. J., K. R. Abrams, et al. (2000). *Methods for Meta-Analysis in Medical Research*. Chichester, West Sussex, UK, John Wiley & Sons, Ltd.

Sutton, A. J. and J. P. Higgins (2008). "Recent developments in meta-analysis." *Stat Med* 27(5): 625–650.

Task Force on Community Preventive Services (2010). *Economic Evaluation Abstraction Form*. Atlanta, Community Guide Branch, National Center for Health Marketing (NCHM), Centers for Disease Control and Prevention.

Thompson, S. G. and J. P. T. Higgins (2002). "How should meta-regression analyses be undertaken and interpreted? ." *Stat Med* 21: 1559–1573.

Thurston, S., D. Craig, et al. (2008). "Increasing decision-makers' access to economic evaluations: Alternative methods of communicating information." *International Journal of Technology Assessment in Health Care* 24(2): 151–157.

Tricco, A. C., J. Tetzlaff, et al. (2010). "The art and science of knowledge synthesis." *J Clin Epidemiol*.

Trikalinos, T. A., U. Siebert, et al. (2009). "Decision-Analytic Modeling to Evaluate Benefits and Harms of Medical Tests—Uses and Limitations."

- Vanni, T., J. Karnon, et al. (2011). "Calibrating models in economic evaluation: a seven-step approach." *Pharmacoeconomics* 29(1): 35–49.
- Walker, D. G., R. Hutubessy, et al. (2010). "WHO Guide for standardisation of economic evaluations of immunization programmes." *Vaccine*.
- Wallace, J., S. Weingarten, et al. (2002). "The limited incorporation of economic analyses in clinical practice guidelines." *J Gen Intern Med* 17: 210–220.
- Weinstein, M. C. (2006). "Recent developments in decision-analytic modelling for economic evaluation." *Pharmacoeconomics* 24(11): 1043–1053.
- Weinstein, M. C., H. V. Fineberg, et al. (1980). *Clinical Decision Analysis*. Philadelphia, W. B. Saunders Company.
- Weinstein, M. C., B. O'Brien, et al. (2003). "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* 6(1): 9–17.
- Weinstein, M. C., G. Torrance, et al. (2009). "QALYs: the basics." *Value Health* 12 Suppl 1: S5–9.
- Weinstein, M. C., E. L. Toy, et al. (2001). "Modeling for health care and other policy decisions: uses, roles, and validity." *Value Health* 4(5): 348–361.
- Wells, G. A., S. A. Sultan, et al. (2009). *Indirect Evidence: Indirect Treatment Comparisons in Meta-Analysis*. Ottawa, Canadian Agency for Drugs and Technologies in Health (CADTH).
- Welte, R., T. Feenstra, et al. (2004). "A Decision Chart for Assessing and Improving the Transferability of Economic Evaluation Results Between Countries." *Pharmacoeconomics* 22(13): 857–876.
- Welte, R. and R. Leidl (1999). *Übertragung der Ergebnisse ökonomischer Evaluationsstudien aus dem Ausland auf Deutschland: Probleme und Lösungsansätze. Ansätze und Methoden der ökonomischen Evaluation. Eine internationale Perspektive*. R. Leidl, J. M. Graf von der Schulenburg and J. Wasem. Baden-Baden, Nomos.
- Welte, R., R. Leidl, et al. (2010). *Health Economics of Infectious Diseases. Modern Infectious Disease Epidemiology: Concepts, Methods, Mathematical Models, and Public Health (Statistics for Biology and Health)*. A. Krämer, M. Kretzschmar and K. Krickeberg. Berlin, Springer.

Whitehead, A. (2002). *Meta-Analysis of Controlled Clinical Trials*. Chichester, West Sussex, UK, John Wiley & Sons, Ltd.

WHO (2008). *WHO guide for standardization of economic evaluations of immunization programmes*. Geneva, WHO Press.

Wong, J., D. Eddy, et al. (2011). *DRAFT – Model Transparency and Validation: A Report of the ISPOR-SMDM Modeling Good Research Practices*

Task Force Working Group – Part 4.

Zaidi, A., A. Harding, et al. (2009). *New Frontiers in Microsimulation*. Vienna, Ashgate.

Zechmeister, I., B. Freiesleben de Blasio, et al. (2007). *Economic Evaluation of HPV vaccination in Austria*. Vienna, Ludwig Boltzmann Institute for Health Technology Assessment.

Zechmeister, I. and C. Wild (2006). *Einsatz von Statinen zur Sekundärprävention kardio-vaskulärer Erkrankungen. Ein systematischer Review gesundheitsökonomischer Analysen*. Wien, Institut für Technikfolgenabschätzung.

Zuccelli, A., M. Jones, et al. (2010). *The Evaluation of Health Policies through Microsimulation Methods*, HEDG, c/o Department of Economics, University of York.

14.6.2 Literatur zu Kapitel 8.2 Bugdetauswirkungsanalyse

BGBI. 189/1955 in der geltenden Fassung (idgF): Bundesgesetz vom 9. September 1955 über die Allgemeine Sozialversicherung (Allgemeines Sozialversicherungsgesetz – ASVG.). <http://www.ris.bka.gv.at/> (9. 7. 2011)

BGBI. II Nr. 50/1999 in der geltenden Fassung (idgF): Bundesgesetz. Verordnung des Bundesministers für Finanzen betreffend Richtlinien für die Ermittlung und Erstellung und Darstellung der finanziellen Auswirkungen neuer rechtsetzender Maßnahmen. <http://www.ris.bka.gv.at/> (9. 7. 2011)

Garattini, Livio; Van de Vooren, Katelijne (2011): Budget impact analysis in economic evaluation: a proposal for a clearer definition. In: Eur J Health Econ 2011, o. S.

GÖG / BIQG (2009): Health Technology Assessment. Einsatz und Bedarf in Österreich und Implikationen für die Verankerung in Österreich. Gesundheit Österreich GmbH / Geschäftsbereich BIQG. Wien

IQWiG (2009): Allgemeine Methoden zur Bewertung von Verhältnissen zwischen Nutzen und Kosten. Version 1.0 vom 12.10.2009. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Köln

Mauskopf, Josephine; Sullivan, Sean; Annemans, Lieven; Caro, Jaime; Mullins, Daniel; Nuijten, Mark; Orlewska, Ewa; Watkins, John; Trueman, Paul (2007): Principles of Good Practice for Budget Impact Analysis: Report of the ISPOR Task Force on Good Research Practices – Budget Impact Analysis. In: Value in Health 2007/10/5, 336–347

Milne, Richard; Grimwood, Keith (2009): Budget Impact and Cost-Effectiveness of Including a Pentavalent Rotavirus Vaccine in the New Zealand Childhood Immunization Schedule. In: Value in Health 2009/12/6, 888–898

Nuijten, Mark; Mittendorf, Thomas; Persson U. (2010): Practical issues in handling data input and uncertainty in a budget impact analysis. In: Eur J Health Econ 2011/12, 231–241

Orlewska, Ewa; Mierzejewski, Piotr (2004): Proposal of Polish Guidelines for Conducting Financial Analysis and Their Comparison to Existing Guidance on Budget Impact in Other Countries. In: Value in Health 2004/7/1, 1–10

Patented Medicine Prices Review Board (2007): Budget Impact Analysis Guidelines. Guidelines for Conducting Pharmaceutical Budget Impact Analyses for Submission to Public Drug Plans in Canada. Kanada

Trueman, Paul; Drummond, Michael; Hutton, John. (2001): Developing guidance for budget impact analysis. In: Pharmacoeconomics 2001/19/6, 609–621

14.6.3 Literatur zu Kapitel 8.3 Gesundheitsökonomische Modellierung

Anderson, R. M. and R. M. May (1991). *Infectious diseases of humans: dynamics and control*. Oxford, UK, Oxford University Press.

Auchincloss, A. H., R. L. Riolo, et al. (2011). "An agent-based model of income inequalities in diet in the context of residential segregation." *Am J Prev Med* 40(3): 303–311.

Axelrod, R. (1997). *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*. Princeton, Princeton University Press.

Banks, J., J. S. Carson, et al. (2005). *Discrete-Event System Simulation*. Upper Saddle River, New Jersey, Pearson Prentice Hall.

Barton, P., S. Bryan, et al. (2004). "Modelling in the economic evaluation of health care: selecting the appropriate approach." *J Health Serv Res Policy* 9(2): 110–118.

Beck, J. R. and S. G. Pauker (1983). "The Markov Process in Medical Prognosis." *Medical Decision Making* 3: 419–458.

Berry, D. A., K. A. Cronin, et al. (2005). "Effect of screening and adjuvant therapy on mortality from breast cancer." *N Engl J Med* 353(17): 1784–1792.

Beutels, P., P. A. Scuffham, et al. (2008). "Funding of drugs: do vaccines warrant a different approach?" *The Lancet Infectious Diseases* 8(11): 727–733.

Beutels, P., E. Van Doorslaer, et al. (2003). "Methodological issues and new developments in the economic evaluation of vaccines." *Expert Rev Vaccines* 2(5): 649–660.

Billari, F. C., T. Fent, et al. (2006). *Agent-Based Computational Modelling: Introduction*. Agent-Based Computational Modelling. Heidelberg, Physica – Verlag: 1 – 16.

Bonabeau, E. (2002). "Agent-based modeling: methods and techniques for simulating human systems." *Proc Natl Acad Sci U S A* 99 Suppl 3: 7280–7287.

Bonabeau, E. (2002). "Predicting the unpredictable." *Harv Bus Rev* 80(3): 109–116, 134.

Brailford, S. (2007). Tutorial: Advances and challenges in healthcare simulation modeling. Winter Simulation Conference.

Brennan, A., S. E. Chick, et al. (2006). "A taxonomy of model structures for economic evaluation of health technologies." *Health Econ* 15(12): 1295–1310.

Briggs, A., E. Fenwick, et al. (2011). DRAFT – Model Parameter Estimation and Uncertainty: A Report of the ISPOR–SMDM Modeling

Good Research Practices Task Force Working Group – Part 3.

Briggs, A., M. Sculpher, et al. (1994). "Uncertainty in the Economic Evaluation of Health Care Technologies: The Role of Sensitivity Analysis." *Health Economics* 3: 95–104.

Briggs, A. H. (2000). "Handling uncertainty in cost-effectiveness models." *Pharmacoeconomics* 17(5): 479–500.

Brisson, M. and W. J. Edmunds (2003). "Economic evaluation of vaccination programs: The impact of herd-immunity." *Medical Decision Making* 23(1): 76–82.

Brisson, M. and W. J. Edmunds (2003). "Varicella vaccination in England and Wales: cost-utility analysis." *Arch Dis Child* 88(10): 862–869.

Brown, B. N., I. M. Price, et al. (2011). "An agent-based model of inflammation and fibrosis following particulate exposure in the lung." *Math Biosci.*

Caro, J. J. (2005). "Pharmacoeconomic analyses using discrete event simulation." *Pharmacoeconomics* 23(4): 323–332.

Cooper, K., S. C. Brailsford, et al. (2007). "Choice of Modelling Technique for Evaluating Health Care Interventions." *The Journal of the Operational Research Society* 58(2): 168–176.

Drummond, M., D. Brixner, et al. (2009). "Toward a consensus on the QALY." *Value Health* 12 Suppl 1: S31–35.

Drummond, M., C. Chevat, et al. (2007). "Do we fully understand the economic value of vaccines?" *Vaccine* 25(32): 5945–5957.

Drummond, M. F. and L. Davies (1991). "Economic Analysis Alongside Clinical Trials." *International Journal of Technology Assessment in Health Care* 7(4): 561–573.

Drummond, M. F., B. O'Brien, et al. (1997). *Methods for the Economic Evaluation of Health Care Programmes*. New York, Oxford University Press Inc.

Drummond, M. F., J. S. Schwartz, et al. (2008). "Key principles for the improved conduct of health technology assessments for resource allocation decisions." *Int J Technol Assess Health Care* 24(3): 244–258; discussion 362–248.

Eddy, D. M. and L. Schlessinger (2003). "Validation of the archimedes diabetes model." *Diabetes Care* 26(11): 3102–3110.

Edmunds, W. J., G. F. Medley, et al. (1999). "Evaluating the cost-effectiveness of vaccination programmes: A dynamic perspective." *Statistics in Medicine* 18(23): 3263–3282.

Fine, P. E. (1993). "Herd immunity: history, theory, practice." *Epidemiol Rev* 15(2): 265–302.

Goldstein, M., Tsevat, J. (2003). *Interactive Textbook on Clinical Symptom Research Chapter 24: Assessing Desirability of Outcomes States for Medical Decision Making and Cost-effectiveness Analysis*. M. Mitchell, Lynn, J., NIH. Chapter 24.

Groot Koerkamp, B., M. C. Weinstein, et al. (2010). "Uncertainty and patient heterogeneity in medical decision models." *Med Decis Making* 30(2): 194–205.

Group, T. M. H. M. (2007). "Computer modeling of diabetes and its complications: a report on the Fourth Mount Hood Challenge Meeting." *Diabetes Care* 30(6): 1638–1646.

Hunink, M. G., P. P. Glasziou, et al. (2001). *Decision making in health and medicine. Integrating evidence and values*. Cambridge, Cambridge University Press.

IQWiG (2009). *Allgemeine Methoden zur Bewertung von Verhältnissen zwischen Nutzen und Kosten*. Version 1.0 vom 12.10.2009., Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG).

ISPOR–SMDM Modeling Good Research Practices Task Force, I. (2011). "Presentations."

Karnon, J. and T. Vanni (2011). "Calibrating models in economic evaluation: a comparison of alternative measures of goodness of fit, parameter search strategies and convergence criteria." *Pharmacoeconomics* 29(1): 51–62.

Kim, J. J., K. M. Kuntz, et al. (2007). "Multiparameter calibration of a natural history model of cervical cancer." *Am J Epidemiol* 166(2): 137–150.

- Kim, S. Y. and S. J. Goldie (2008). "Cost-effectiveness analyses of vaccination programmes : a focused review of modelling approaches." *Pharmacoeconomics* 26(3): 191–215.
- Kleijnen, J. P. C. (1995). "Verification and Validation of simulation models." *European Journal of Operational Research* 82(1): 145–162.
- Klein, T., Scholz, Höer, Albrecht, Häussler, Schmidt, Jahn, Staudt, Farrenkopf, Lux, Weegen, Thomas, Wasem, Sussex, Shah, Butler, Mühlberger, Gothe, Schwarzer, Schnell-Inderst, Sroczynski, Siebert (2010). Gutachten zur Verbesserung der Wirtschaftlichkeit von Impfstoffen in Deutschland B. f. G. (BMG). Berlin, Germany.
- Kong, C. Y., P. M. McMahon, et al. (2009). "Calibration of disease simulation model using an engineering approach." *Value Health* 12(4): 521–529.
- Koopman, J. S. (2005). "Infection transmission science and models." *Jpn J Infect Dis* 58(6): S3–8.
- Lee, B. Y., S. T. Brown, et al. (2010). "A computer simulation of vaccine prioritization, allocation, and rationing during the 2009 H1N1 influenza pandemic." *Vaccine* 28(31): 4875–4879.
- Leidl, R. (1998). *Der Effizienz auf der Spur: Eine Einführung in die ökonomische Evaluation. Das Public Health Buch: Gesundheit und Gesundheitswesen.* F. W. Schwartz, et al., Urban & Schwarzenberg. 18: 346–369.
- Macal, C. M. and M. J. North (2005). Tutorial on agent-based modeling and simulation. Proceedings of the 37th conference on Winter simulation. Orlando, Florida, Winter Simulation Conference: 2–15.
- Macal, C. M. and M. J. North (2006). Tutorial on agent-based modeling and simulation part 2: how to model with agents. Proceedings of the 38th conference on Winter simulation. Monterey, California, Winter Simulation Conference: 73–83.
- North, M. J. and C. M. Macal (2007). *Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation*, Oxford University Press, Inc.
- Philips, Z., L. Bojke, et al. (2006). "Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment." *Pharmacoeconomics* 24(4): 355–371.

Pitman, R., J. Edmunds, et al. (2011). DRAFT – Dynamic Transmission Modelling: A Report of the ISPOR–SMDM Modelling Good Research Practices Task Force Working Group – Part 7.

Ramsey, S., R. Willke, et al. (2005). "Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT–CEA Task Force report." *Value Health* 8(5): 521–533.

Ratcliffe, J., T. Young, et al. (2001). "A simulation modelling approach to evaluating alternative policies for the management of the waiting list for liver transplantation." *Health Care Manag Sci* 4(2): 117–124.

Rein, D. B., J. B. Saaddine, et al. (2007). "Cost-effectiveness of vitamin therapy for age-related macular degeneration." *Ophthalmology* 114(7): 1319–1326.

Roberts, M., M. Chambers, et al. (2011). DRAFT – Conceptual Modeling: A Report of the ISPOR–SMDM Modeling Good Research Practices Task Force Working Group – Part 2.

Rutter, C. M., D. L. Miglioretti, et al. (2009). "Bayesian Calibration of Microsimulation Models." *J Am Stat Assoc* 104(488): 1338–1350.

Rutter, C. M., A. M. Zaslavsky, et al. (2011). "Dynamic microsimulation models for health outcomes: a review." *Med Decis Making* 31(1): 10–18.

Sander, B., A. Nizam, et al. (2009). "Economic evaluation of influenza pandemic mitigation strategies in the United States using a stochastic microsimulation transmission model." *Value Health* 12(2): 226–233.

Sculpher, M., E. Fenwick, et al. (2000). "Assessing quality in decision analytic cost-effectiveness models. A suggested framework and example of application." *Pharmacoeconomics* 17(5): 461–477.

Sculpher, M. J., K. Claxton, et al. (2006). "Whither trial-based economic evaluation for health care decision making?" *Health Econ* 15(7): 677–687.

Shechter, S. M., C. L. Bryce, et al. (2005). "A clinically based discrete-event simulation of end-stage liver disease and the organ allocation process." *Med Decis Making* 25(2): 199–209.

Siebert, U. (2003). *Transparente Entscheidungen in Public Health mittels systematischer Entscheidungsanalyse. Das Public Health Buch. Gesundheit fördern,*

Krankheit verhindern. F. W. Schwartz, B. Badura, R. Leidlet al. München, Urban & Fischer.

Siebert, U. (2003). "The use of decision analysis in health economic evaluation [German]." *Informatik, Biometrie und Epidemiologie in Medizin und Biologie* 34(2): 112–124.

Siebert, U. (2003). "When should decision-analytic modeling be used in the economic evaluation of health care? [Editorial]." *European Journal of Health Economics* 4(3): 143–150.

Siebert, U. (2009). *Antiviral Therapy of Chronic Hepatitis C –*

A Pilot Study Applying the "Methods for Assessment of the Relation of Benefits to Costs in the German Statutory Health Care System".

Siebert, U., O. Alagoz, et al. (2011). *DRAFT – State-Transition Modeling: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group – Part 5.*

Siebert, U. and T. Kurth (2000). *Lebensqualität als Parameter von medizinischen Entscheidungsanalysen. Lebensqualität und Gesundheitsökonomie in der Medizin.* U. Ravens-Sieberer, A. Cieza, N. von Steinbüchel and M. Bullinger. Landsberg, Ecomed: 365–392.

Siebert, U., N. Mühlberger, et al. *Evidenzsynthese: Meta-Analysen und Entscheidungsanalysen – Draft for 4th Edition. Gesundheitsökonomische Evaluationen*

O. Schöffski and J.-M. Graf von der Schulenburg. Berlin, Springer.

Siebert, U., N. Mühlberger, et al. (2000). *Desk Research. Gesundheitsökonomische Evaluationen.* O. Schöffski and J.-M. Graf v. d. Schulenburg. Berlin, Springer.

Siebert, U., N. Mühlberger, et al. (2000). *Formale Entscheidungsanalyse. Gesundheitsökonomische Evaluationen.* O. Schöffski and J.-M. Graf v. d. Schulenburg. Berlin, Springer: 88–120.

Sonnenberg, F. A. and J. R. Beck (1993). "Markov models in medical decision making: a practical guide." *Med Decis Making* 13(4): 322–338.

Sroczyński, G., P. Schnell-Inderst, et al. (2010). *Entscheidungsanalytische Modellierung zur Evaluation der Langzeit-Effektivität und Kosten-Effektivität des*

Einsatzes der HPV-DNA-Diagnostik im Rahmen der Zervixkarzinomfrüherkennung in Deutschland

[Decision-analytic modeling to evaluate the long-term effectiveness and cost-effectiveness of HPV-DNA testing in primary cervical cancer screening in Germany]. Köln.

Stahl, J. E. (2008). "Modelling methods for pharmacoeconomics and health technology assessment: an overview and guide." *Pharmacoeconomics* 26(2): 131-148.

Stahl, J. E., K. L. Furie, et al. (2003). "Stroke: Effect of implementing an evaluation and treatment protocol compliant with NINDS recommendations." *Radiology* 228(3): 659-668.

Stahl, J. E., J. Karnon, et al. (2011). DRAFT- Modeling Discrete Event Simulation: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group - Part 6.

Stout, N. K., A. B. Knudsen, et al. (2009). "Calibration methods used in cancer simulation models and suggested reporting guidelines." *Pharmacoeconomics* 27(7): 533-545.

Trikalinos, T. A., U. Siebert, et al. (2009). "Decision-Analytic Modeling to Evaluate Benefits and Harms of Medical Tests-Uses and Limitations."

Vanni, T., J. Karnon, et al. (2011). "Calibrating models in economic evaluation: a seven-step approach." *Pharmacoeconomics* 29(1): 35-49.

Walker, D. G., R. Hutubessy, et al. (2010). "WHO Guide for standardisation of economic evaluations of immunization programmes." *Vaccine*.

Weinstein, M. C. (2006). "Recent developments in decision-analytic modelling for economic evaluation." *Pharmacoeconomics* 24(11): 1043-1053.

Weinstein, M. C., H. V. Fineberg, et al. (1980). *Clinical Decision Analysis*. Philadelphia, W. B. Saunders Company.

Weinstein, M. C., B. O'Brien, et al. (2003). "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* 6(1): 9-17.

Weinstein, M. C., G. Torrance, et al. (2009). "QALYs: the basics." *Value Health* 12 Suppl 1: S5-9.

Weinstein, M. C., E. L. Toy, et al. (2001). "Modeling for health care and other policy decisions: uses, roles, and validity." *Value Health* 4(5): 348–361.

Welte, R., R. Leidl, et al. (2010). *Health Economics of Infectious Diseases. Modern Infectious Disease Epidemiology: Concepts, Methods, Mathematical Models, and Public Health (Statistics for Biology and Health)*. A. Krämer, M. Kretzschmar and K. Krickeberg. Berlin, Springer.

WHO (2008). *WHO guide for standardization of economic evaluations of immunization programmes*. Geneva, WHO Press.

Wong, J., D. Eddy, et al. (2011). *DRAFT – Model Transparency and Validation: A Report of the ISPOR–SMDM Modeling Good Research Practices*

Task Force Working Group – Part 4.

Zaidi, A., A. Harding, et al. (2009). *New Frontiers in Microsimulation*. Vienna, Ashgate.

Zuccelli, A., M. Jones, et al. (2010). *The Evaluation of Health Policies through Microsimulation Methods*, HEDG, c/o Department of Economics, University of York.

14.7 Literatur zu Kapitel 9 Soziale, ethische, rechtliche und organisatorische Aspekte

Busse R, Orvain J, Drummond M, Gurtner F, Jorgensen T, Jovell A, Malone J, Perleth M, Wild C (2001): Best Practice in undertaking and reporting HTA. ECHTA-final report. http://www.oeaw.ac.at/ita/ebene5/WG4_FinalReport_010719.pdf. accessed 4.1.2010.

Droste S (2008): Systematische Gewinnung von Informationen zu ethischen Aspekten in HTA-Berichten zu medizinischen Technologien bzw. Interventionen. In: *Z Evid Fortbild Qual Gesundheitswesen (ZEFQ)* 102 (2008) 329–343

EUnetHTA 2008: HTA Core Model Handbook for medical and surgical interventions.
<http://www.eunetha.net/upload/WP4/Final%20Deliverables/HTA%20Core%20Model%20for%20Medical%20and%20Surgical%20Interventions%201%200r.pdf>. accessed 4.1.2010.

- » Ethical Analysis: 92–115
- » Organisational Aspects: 116–127
- » Social Aspects: 128–141
- » Legal Aspects: 142–156

Francke R, Hart D (2008): Einführung in die rechtlichen Aspekte bei HTAs. In: Z Evid Fortbild Qual Gesundheitswesen (ZEFQ) 102 (2008) 63–68

Gerhardus A, Perleth M (2008): Health Technology Assessment: Mehr als die Bewertung von Kosten und Nutzen? Editorial In: Z Evid Fortbild Qual Gesundheitswesen (ZEFQ) 102 (2008) 61–62

Gerhardus A, Stich AK (2008): Sozio-kulturelle Aspekte in Health Technology Assessment. In: Z Evid Fortbild Qual Gesundheitswesen (ZEFQ) 102 (2008) 77–83

INAHTA Ethics Working Group 2003: Report on handling ethical issues in HTA.
<http://www.inahta.org/upload/Final%20report%20Ethics%20in%20HTA%20Nov%2007.pdf>. accessed 4.1.2010.

Lühmann D, Raspe H (2008): Ethik im Health Technology Assessment – Anspruch und Umsetzung. In: Z Evid Fortbild Qual Gesundheitswesen (ZEFQ) 102 (2008) 69–76

Lühmann D, Bartel C, Raspe H (2007): Ethische Aspekte und gesellschaftliche Wertvorstellungen in HTA-Berichten zu genetischen Testverfahren. In: Brand A, Schröder P, Bora A, Dabrock P, Kälble K, Ott N, Wewetzer C, Brand H: Genetik in Public Health, Teil 2. Wissenschaftliche Reihe, Bielefeld, 349–415

Mathis S, Wild C, Piso B, Piribauer F (2009): Evidenzgestützte Versorgungsplanung. Vergleichende Analyse ausgewählter Methoden zur Identifikation von Potenzialen der Evidenzbasierung der Versorgungsplanung. A Projektbericht 021.
http://eprints.hta.lbg.ac.at/843/1/HTA-Projektbericht_Nr21.pdf. accessed 4.1.2010

Perleth M (2008): Berücksichtigung organisatorischer Aspekte in HTA-Berichten. In: Z Evid Fortbild Qual Gesundheitswesen (ZEFQ) 102 (2008) 85–91

Wild C (2007): (Externes) Manual. Selbstverständnis und Arbeitsweise. LBI-HTA/
Ludwig Boltzmann Institut für Health technology Assessment
http://eprints.hta.lbg.ac.at/714/1/HTA-Projektbericht_003.pdf. accessed 4.1.2010

Wild C (2005): Ethics of Resource Allocation: Instruments for rational decision-making in support of a sustainable health care. invited paper. In: Poiesis & Praxis. International Journal of Ethics of Science and Technology Assessment, 3(4): 296–309

Wild C, Zechmeister I (2007): Gestaltung von Gesundheitspolitik: Rationalitäten einer gerechten Verteilungspolitik – HTA als Mittel zum Zweck. In: Kurswechsel (2): 87–98

14.8 Literatur zu Kapitel 10 Qualitätssicherung

GÖG/BIQG 2010: Prozesshandbuch für Health Technology Assessment Version 1.2010. Gesundheit Österreich GmbH / Geschäftsbereich BIQG. Wien

14.9 Literatur zu Kapitel 11 Implementierung und Impact

Gerhardus A, Dorendorf E, Rottingen J–A, Santamera A: What are the effects of HTA reports on the health system? Evidence from the research literature. In: Health Technology Assessment and Health Policy–Making in Europe. Current status, challenges and potential. M. V. Garrido, F. B. Kristensen, C. P. Nielsen, R. Busse (Hg.). European Network for HTA, 2008.

Oortwijn W, Banta HD, Cranovsky R (2001): Introduction: mass screening, health technology assessment, and health policy in some European countries. In: Int J Technol Assess Health Care. 17(3):269–74

Schumacher I, Zechmeister I (2010): Auswirkungen der HTA–Forschung auf das Gesundheitswesen in Österreich. Teil 1: Methodenübersicht – Update. HTA–Projektbericht. Ludwig Boltzmann Institut für Health Technology Assessment. Wien

Sheldon TA, Cullum N, Dawson D, Lankshear A, Lowson K, Watt I, West P, Wright D, Wright J (2004): What's the evidence that NICE guidance has been implemented? Results from a national evaluation using time series analysis, audit of patients' notes, and interviews. In: BMJ. 329(7473):999

15 Anhang A

15.1 Methodenarbeitsgruppe

15.1.1 Department für Evidenzbasierte Medizin und Klinische Epidemiologie an der Donau-Universität Krems

Das Department für Evidenzbasierte Medizin und Klinische Epidemiologie wurde am 1. Jänner 2008 an der Donau-Universität Krems gegründet. Das Department beruht auf einer Stiftungsprofessur für Evidenzbasierte Medizin und Klinische Epidemiologie des Niederösterreichischen Gesundheits- und Sozialfonds (NÖGUS), die an Prof. Dr. Gerald Gartlehner, MPH vergeben wurde.

Ziel des Departments ist, Entscheidungen im österreichischen Gesundheitssystem mit evidenzbasierten Methoden wie Metaanalysen, systematischen Übersichtsarbeiten und Health Technology Assessments zu unterstützen. Ein Kernstück ist das evidenzbasierte Informationszentrum für niederösterreichische Ärztinnen und Ärzte, das Anfragen von niederösterreichischen Krankenhausärztinnen und Krankenhausärzten in kompakter und evidenzbasierter Form beantwortet.

Das Department ist unter www.donau-uni.ac.at/de/department/evidenzbasiertemedizin auf der Webseite der Donau-Universität präsent.

Das Department wurde als weisungsfreie und unabhängige Institution mit uneingeschränkten Möglichkeiten zur Drittmittelerwerbung gegründet. Arbeits- und Forschungsschwerpunkte liegen im Bereich der relativen Effektivitätsbewertungen von medizinischen Interventionen und der Evaluation von gesundheitsfördernden und präventiven Maßnahmen. Das Department hat ein offizielles Arbeitsabkommen mit dem Sheps Center for Health Services Research der University of North Carolina at Chapel Hill, USA und nutzt diese Zusammenarbeit intensiv zum methodischen und fachlichen Austausch. Der Schwerpunkt der Lehrtätigkeit liegt bei Workshops über Methoden evidenzbasierter Medizin und dem Einsatz von Evaluationen im Gesundheitssystem.

15.1.2 EBM Review Center Graz

Das EBM Review Center an der Medizinischen Universität Graz wurde mit 1. April 2005 an der Universitätsklinik für Innere Medizin unter der Leitung von Priv. Doz. Dr. Karl Horvath und Univ. Prof. Dr. Andrea Siebenhofer-Kroitzsch eingerichtet und ist seit 2009 Research Unit am Institut für Medizinische Informatik, Statistik und Dokumentation (IMI).

Erklärte Absicht des EBM Review Centers ist die Erstellung von systematischen Übersichtsarbeiten und Metaanalysen zu bestimmten medizinischen Fragestellungen. Die Ergebnisse sollen nicht nur eine Grundlage für Versorgungsentscheidungen im Gesundheitsbereich, für Health Technology Assessments, Disease Management Programme und Leitlinien bilden, sondern sie sollen auch praktisch tätige Ärztinnen und Ärzte bei diagnostischen und therapeutischen Entscheidungen unterstützen und als Basis für die Erstellung von Patienteninformationen dienen.

Grundlage sämtlicher Aktivitäten bildet ein systematisches, transparentes Erarbeiten der Ergebnisse auf Basis von prospektiv festgelegten und nachvollziehbaren Methoden sowie die Unabhängigkeit von Interessengruppen.

Vom EBM RC wurden bisher systematische Übersichtsarbeiten für das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen in Deutschland und für mehrere Abteilungen österreichischer Spitäler und die medizinischen Behandlungspfade für das DMP Diabetes mellitus 2 „Therapie aktiv“ erstellt. Eine Zusammenarbeit mit der Arzneimittelkommission am LKH Graz und weiteren in- und ausländischen Partnern besteht. Weiters wurden von der Autorengruppe des EBM RC bisher drei Cochrane Reviews veröffentlicht und ein Review befindet sich aktuell im Editorialprozess, weitere sind geplant. Ebenso sind als konkrete Aufgaben die Weiterentwicklung von Methoden zur Studienbewertung im Sinne einer Methoden-Grundlagenforschung sowie Fortbildung und Lehre ("Evidenzbasierte Medizin für die Praxis") zu sehen. Im Rahmen des Medizinstudiums an der Medizinischen Universität Graz ein Special Study Modul implementiert, in dem Studentinnen und Studenten in den grundlegenden Aspekten Evidenzbasierter Medizin unterrichtet werden.

15.1.3 Gesundheit Österreich GmbH / Geschäftsbereich BIQG

In seiner Rolle als bundesweite Koordinationsstelle in Bereichen der Qualität im Gesundheitswesen erarbeitet der Geschäftsbereich BIQG die nationale HTA-Strategie, gemeinsam mit Vertreterinnen und Vertretern von Bund, Ländern und Sozialversiche-

rung und unter Einbeziehung nationaler und internationaler Expertinnen und Experten. Diese HTA-Strategie schafft einen Rahmen für die Verankerung und den verstärkten Einsatz von HTA als Entscheidungsgrundlage im österreichischen Gesundheitswesen und legt gemeinsame Ziele fest. Sie beschreibt organisatorische und regulatorische Rahmenbedingungen sowie inhaltliche Schwerpunkte des HTA-Prozesses (Themenauswahl, Qualitätssicherung der Berichte, Transparenz, Veröffentlichung und Umsetzung der Ergebnisse).

15.1.4 Ludwig Boltzmann Institut Health Technology Assessment

Das Ludwig Boltzmann Institut für Health Technology Assessment (LBI-HTA) wurde im März 2006 für sieben Jahre gegründet. Von Priv.-Doz. Dr. Claudia Wild geleitet, teilt sich das Institut inhaltlich in drei Schwerpunktressorts: Public Health und Versorgungsforschung (Ressortleiterin Dr. Brigitte Piso, MPH), Gesundheitsökonomie (Ressortleiterin Dr. Ingrid Zechmeister, M.A.) und High Tech in Krankenanstalten (Ressortleiterin Priv.-Doz. Dr. Claudia Wild).

Als unabhängige Instanz der wissenschaftlichen Entscheidungsunterstützung im Gesundheitswesen bietet das LBI-HTA wissenschaftliche Grundlagen für Entscheidungen zugunsten eines effizienten und angemessenen Ressourceneinsatzes. Dabei wird eine breite gesellschaftsrelevante Sicht auf medizinische Interventionen eingenommen. Das Institut sieht sich als innovative und kreative interdisziplinäre Denkwerkstatt, die sich zu einem qualitativen Fortschrittsbegriff verpflichtet und sich die Weiterentwicklung der HTA-Methodik zum Ziel gesetzt hat. Die Nachvollziehbarkeit der Ergebnisse (systematisches Arbeiten und das Offenlegen der Methoden) ist ebenso Prämisse wie die uneingeschränkte Distanz zu Interessengruppen. Darüber hinaus ermöglicht intensive internationale Vernetzung, auf dem letzten Stand des Wissens zu bleiben und Redundanzen zu vermeiden.

Entsprechend der Forschungspolitik der Ludwig Boltzmann Gesellschaft steht die Fokussierung auf „Translational Research“, also die Anwend- und die Umsetzbarkeit der Forschungsergebnisse, im Zentrum des Institutsarbeitsprogramms. Als Grundlage dafür wurde eine Synergie von forschungsdurchführenden und -anwendenden Institutionen geschaffen, die den raschen Wissenstransfer ermöglicht. Institutionelle Partner des LBI-HTA sind Akteurinnen und Akteure in der Gesundheitsadministration (Bundesministerium für Gesundheit, Hauptverband der österreichischen Sozialversicherungsträger), Krankenanstaltenträger (Tiroler Landeskrankenanstalten, Steiermärkische Krankenanstalten) und Privatuniversitäten (Private Universität für Gesundheitswissenschaften, Medizinische Informatik und Technik, Institut für Public Health, Medical Decision Making und HTA).

Die Finanzierung des Ludwig Boltzmann Instituts für Health Technology Assessment wird zu 60 Prozent von der Ludwig Boltzmann Gesellschaft und zu 40 Prozent von den Partnern bestritten.

15.1.5 UMIT

Das Institut für Public Health, Medical Decision Making und Health Technology Assessment (IPH) an der UMIT - University for Health Sciences, Medical Informatics and Technology in Hall i.T., Österreich wurde im Jahr 2005 gegründet und wird von Univ.-Prof. Dr. Uwe Siebert, MPH, MSc. geleitet. Seit 2008 ist Univ.-Prof. Dr. Siebert auch Leiter des Forschungsbereichs Public Health, Decision Modeling, Health Technology Assessment und Gesundheitsökonomie in "ONCOTYROL - Center for Personalized Cancer Medicine" in Österreich. Univ.-Prof. Dr. Siebert ist außerdem Adjunct Professor an der Harvard School of Public Health in Boston, MA, USA und Leiter des "Cardiovascular Research Program" am Institute for Technology Assessment (ITA) und Department of Radiology am Massachusetts General Hospital, Harvard Medical School in Boston, MA, USA.

Ziel der Forschungsarbeit am Institut ist die Entwicklung und Anwendung interdisziplinärer Methoden zur umfassenden, systematischen und praxisorientierten Evaluation und Bewertung von Maßnahmen und Verfahren in Medizin und Public Health. Die Forschung unterstützt Entscheidungsträger, um die Qualität, die Effektivität und Kosteneffektivität im Gesundheitswesen zu verbessern. Risiken sollen minimiert werden und die Gesundheit von Individuum und Gesellschaft erhalten und gefördert werden.

Der wissenschaftliche Fokus am Institut liegt im interdisziplinären Public-Health-Ansatz. Er bedient sich vor allem der Methoden der entscheidungsanalytische Modellierung, evidenz-basierten Medizin, Versorgungsforschung, Gesundheitsökonomie, Biostatistik, Epidemiologie und Lebensqualitätsforschung. Das IPH umfasst mehrere Forschungsgruppen, darunter ein eigenes Programm für Health Technology Assessment (Leitung: Dr. Petra Schnell-Inderst, MPH). Ziel dabei ist die Bewertung der Kurz- und Langzeit-Effektivität von Gesundheitstechnologien sowie der Kosten und Kosteneffektivität für das Gesundheitssystem und die Gesellschaft. Besondere Schwerpunkte stellen hierbei kardiovaskuläre Krankheiten, Krebsvorsorge und -behandlung, Infektionskrankheiten, Diabetes sowie neurologische Funktionsstörungen dar.

Das Institut ist auf der Webseite der UMIT unter http://www.umat.at/page.cfm?vpath=departments/public_health erreichbar.

15.2 Nationale und internationale Methodenhandbücher

Im Folgenden sind Methodenhandbücher, die für die Erstellung herangezogen wurden, aufgelistet und verlinkt. Die meisten dieser Handbücher sind ausführlicher als vorliegendes Handbuch und bieten sich deshalb für eine vertiefende Auseinandersetzung mit HTA-Methoden an.

- » Center for Reviews and Dissemination (CRD):
„CRD’s guidance for undertaking reviews in health care“,
http://www.york.ac.uk/inst/crd/systematic_reviews_book.htm
- » Cochrane Collaboration:
„Cochrane Handbook for Systematic Reviews of Interventions“,
<http://www.cochrane-handbook.org/>
- » Danish Centre for Health Technology Assessment (DACEHTA):
Handbücher der DACEHTA:
http://www.sst.dk/English/DACEHTA/Toolbox.aspx?sc_lang=en
- » EUnetHTA :
„Core Model“,
http://www.eunetha.eu/Public/EUnetHTA_Deliverables_project_2006-2008/
- » Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG):
„Allgemeine Methoden (Nutzenbewertung)“,
<http://www.iqwig.de/index.428.html>
- » Ludwig Boltzmann Institut für Health Technology Assessment (LBI-HTA):
„Internes Manual. Abläufe und Methoden“,
<http://eprints.hta.lbg.ac.at/713/>
- » National Institute for Health and Clinical Excellence (NICE):
„Guide to the methods of technology appraisal“,
http://www.nice.org.uk/aboutnice/howwework/devnicetech/technologyappraisalprocessguides/guide_to_the_methods_of_technologyappraisal_reference_n0515.jsp

15.3 Quellen für Datenbanken

15.3.1 Quellen für die systematische Literatursuche

15.3.1.1 Literaturdatenbanken für Primärstudien

EMBASE: <http://www.embase.com/>

Medline: Suche über PubMed: <http://www.ncbi.nlm.nih.gov/sites/entrez/>

15.3.1.2 Literaturdatenbanken für Sekundärstudien

Cochrane: <http://www3.interscience.wiley.com/cgi-bin/mrwhome/106568753/HOME?CRETRY=1&SRETRY=0>

Center for Reviews and Dissamination: <http://www.crd.york.ac.uk/crdweb/>

NICE: <http://www.nice.org.uk/Guidance/Topic>

DAHTA: <http://www.dimdi.de/dynamic/de/hta/db/index.htm>

The Cumulative Index to Nursing and Allied Health Literature (CINAHL):
<http://www.ebscohost.com/cinahl/>

Gerontologische Literatur (GeroLit): http://www.ub.fu-berlin.de/digibib_neu/datenbank/metalib/titel/KOB12499.html

Literaturdatenbankplattform OVID (unterschiedliche Pakete möglich, z. B. Medline, Embase und Cochrane: <http://ovidsp.ovid.com/>

Genauere Infos zu den Datenbanken und weitere Links zu Literaturdatenbanken finden sich im HT-Guide des BIQG unter <http://hta-guide.biqq.at/>.

Eine umfassende Linkliste zu den Datenbanken findet sich auch auf der Homepage des LBI unter <http://hta.lbg.ac.at/de/links.php?iMenuID=61>.

15.3.2 Quellen für die ergänzende Suche

SCOPUS: <http://info.scopus.com/>

Studienregister: www.clinicaltrials.gov

EMA: <http://www.ema.europa.eu/>

FDA: www.fda.gov

15.4 Tabellarische Übersicht zu Datenbeständen

Tabelle 15.1:
Übersicht zu Datenbeständen

Datenkategorie	Dateneigner*	Originalzweck	Datenerhebung	Zugang zu Daten/ Datenniveau ^α	Verfügbarkeit	Klassifikation#	Einsatz im HTA- Prozess##	Referenz
1) Administrative Daten								
Sozialversicherungsdaten	Sozialversicherungsträger, Hauptverband der Sozialversicherungsträger (HVB)	administrative Zwecke, insb. Leistungsabrechnung	div. administrative Prozesse (Personenstandsdaten, Daten zu Beziehungen zu Leistungserbringern, Leistungsdaten)	tlw. öffentlich, ansonsten projektabhängig (d, id) ^α	abhängig von Art der benötigten Daten	II, III, IV	A, B, C1, C3, C4, D, F	http://www.sozialversicherung.at/portal27/portal/esvportal/channel_content/cmSWindow?action=2&p_menuid=60847&p_tabid=5
Beispiel Vorsorgeuntersuchungen	HVB (an GÖG übermittelt)	administrative Zwecke, insb. Leistungsabrechnung	fallbezogene Daten über div. administrative Prozesse	projektabhängig (d); Sonderauswertung an GÖG gegen Kostenersatz (id)	ab 1991	II, IV	A, B, C1, C3, F	
Klinische Dokumentation des Rehabilitationsverlaufs	Zuständige Sozialversicherungsträger Rehasentren	Qualitätssicherung	fallbezogene Daten; Erhebung elektronisch u. manuell, div. Berufsgruppen, div. Zeitpunkte	projektabhängig (d)	abhängig von Träger/Zentren	Ib, IVb	C1, C2	
Dokumentation ambulanter Leistungen über den Katalog ambulanter Leistungen	Sozialversicherungsträger, Hauptverband der Sozialversicherungsträger (HVB), Bundesministerium für Gesundheit	Dokumentation, Planung, Steuerung	fallbezogene elektronische Erfassung auf Basis multiaxialer Struktur	derzeit nicht zugänglich	derzeit noch nicht verfügbar	II, IV	A, B, C3, C4, F	http://bmg.gv.at/home/Schwerpunkte/Gesundheitssystem_Qualitaetsicherung/Dokumentation/Dokumentation_im_ambulantem_Bereich
administrative Daten des Bundesministeriums für Gesundheit								
Daten der leistungsorientierten Krankenanstaltenfinanzierung, minimal basic data set (MBDS)	Bundesministerium für Gesundheit (BMG)	Leistungsabrechnung, Dokumentation, Planung/Steuerung	fallbezogene elektronische Erfassung von Krankenanstaltenträgern	aggregierte Auswertungen öffentlich, ansonsten projektabhängig (id)	ab 1997	Ib	A, B, C1, C3, C4, F	http://bmg.gv.at/cms/home/attachment/5/7/5/CH1166/CMS1128348439539/Anhang1_-_aufbau_und_inhalt_der_datenmeldungen_neu_20100330.pdf http://www.bmg.gv.at/home/Schwerpunkte/Krankenanstalten/LKF_Modell_2011/

Datenkategorie	Dateneigner*	Originalzweck	Datenerhebung	Zugang zu Daten/ Datenniveau ^α	Verfügbarkeit	Klassifi- kation#	Einsatz im HTA- Prozess##	Referenz
Impfstatistik	BMG	Impfdokumentation	Dokumentation der Län- der/Landessanitäts- direktionen, Registrierung des Impfstoffverbrauchs	projektabhängig (id)	mind. ab 2004	II, IVa	A, B, C1, C3, C4, F	
administrative Daten der Krankenanstalten- träger								-
Dokumentation spitalsambulanter Leistungen	Krankenanstalten- träger	administrative Zwecke / Leistungs- abrechnung	elektronische Dok. unterschiedlicher Daten von Kranken- anstaltenträgern (z. B. Art und Frequenz erbrachter Leistungen)	projektabhängig (d, id)	abhängig von Träger	(Ic), II	A, B, C1, C3, C4, F	
Kostendaten aus der Krankenanstal- ten- Kostenrechnung	Leistungsanbieter (z. B. Krankenan- staltenverbände, Eigentümer)	Kostenrechnung, Planung, Steuerung	gemäß Handbuch für Fonds- Krankenanstalten	aggregierte Auswertungen öffentlich, ansonsten projektabhängig (id)	abhängig von Träger	IV	A, B, C3, C4, F	
Krankengeschich- ten	Leistungsanbieter, ELGA	Dokumentation	Dokumentationssys- teme bei Leistungs- anbietern	projektabhängig (d)	fortlaufend; retrospekt. entsprechend Aufbewahrungs- pflicht zw. 10 u. 30 Jahren	Ib, IV	A, B, C1, C2, C3, C4, F	
Administrative Daten der Landessanitäts- direktionen								
Impfstatistik für öffentlich finanzierte Impfungen am Beispiel Kärnten	Landessanitätsdi- rektions Kärnten	Impfdokumentation	Einspeisung fallbezogener Impfdaten von niedergelassenen Ärzten überwiegend über Web- Applikation	derzeit nicht zugänglich	ab 1998	II, IVa	A, B, C1, C3, C4, F	
Sonstige Gebiets- körperschaften								

Datenkategorie	Dateneigner*	Originalzweck	Datenerhebung	Zugang zu Daten/ Datenniveau ^α	Verfügbarkeit	Klassifikation#	Einsatz im HTA- Prozess##	Referenz
Daten der Stellungs- untersuchung	Verteidigungs- ministerium (übermit- telt an GÖG)	Dokumentation	fallbezogene Primärdatenerhebung bei Stellungsunters- uchung	nicht öffentlich; Auswertung an der GÖG gegen Kostenersatz möglich (id)	ab 1990	III	A, B, C1	
2) Register								
Epidemiologische Register								
Krebsregister Österreich	Statistik Austria	Dokumentation der Krebsfälle	fallbezogene Meldung von Krankenanstalten	aggregierte Auswertungen öffentlich, Sonderauswertung gegen Kostenersatz (id)	ab 1983	IIIb	A, B, C1, C4, F	http://www.statistik.at/web_de/statistiken/gesundheit/krebserkrankungen/index.html
Regionale Tumorregister am Beispiel Tiroler Tumorregister	Institut für klinische Epidemiologie der TILAK (IET)	Dokumentation der Krebsfälle	fallbezogene Meldung von Krankenanstalten, Recherche, Patholo- giebefunde	aggregierte Auswertungen öffentlich, Sonderauswertun- gen projektabhän- gig möglich (id)	ab 1986	IIIb	A, B, C1, C4, F	http://www.iet.at/index.php?id=1618
Geburtenregister Österreich	IET	Dokumentation der Geburten, Qualitäts- sicherung	fallbezogene elektronische Dokumentation in Krankenanstalten	aggregierte Auswertungen öffentlich, Sonderauswertun- gen projektabhän- gig möglich (id)	ab 1998	Ia	A, B, C1, C2, C3, C4, F	http://www.iet.at/fileadmin/download/geburtenregister/groe-handbuch-2005.pdf
Register der anzeigepflichtigen Erkrankungen / epidemiologisches Melde-system	BMG, Bezirksver- waltungsbehörden	Überwachung und Bekämpfung von Infektionskrankhei- ten gemäß bundesgesetzlicher Vorschriften und EU-Vorgaben	fallbezogene elektronische Erfassung über das epidemiologische Melde-system	seit 2010 monatli- che Statistiken über Erkrankungen u. Todesfälle pro Bundesland online verfügbar; Sonderauswertun- gen bei Bundesmi- nisterium für Gesundheit projektabhängig, Ausweitung des Datenzugangs geplant (id)	ab 2000; online zugänglich seit 2010; über GÖG ab 1990	IIIb	A, B, C1, C4	http://bmg.gv.at/home/Schwerpunkte/Krankheiten/Uebertragbare_Krankheiten/Statistiken/
Qualitätsregister								

Datenkategorie	Dateneigner*	Originalzweck	Datenerhebung	Zugang zu Daten/ Datenniveau ^α	Verfügbarkeit	Klassifikation#	Einsatz im HTA- Prozess##	Referenz
Erwachsenen- Herzchirurgie- Register	GÖG	Qualitätssicherung	fallbezogene Erfassung in Krankenanstalten, 1x jährlich Übermittlung an GÖG	für teilnehmende Zentren eigene Daten jederzeit einsehbar; jährlicher anonymisierter, nicht öffentlicher Gesamtbericht; Auswertungen projektabhängig gegen Kostenersatz (id)	ab 2006	Ia	A, B, C1, C2, C3, C4, F	http://www.goeg.at/de/Bereich/Erwachsenen-Herzchirurgie-Register.html
Herzschrittma- cher- und Loop- Recorder-Register	GÖG	Qualitätssicherung	produktbezogene Erfassung direkt online oder dezentral und Übermittlung mittels Web-Service	jährliche Auswertung für das BMG, kein unmittelbarer öffentlicher Zugang (id)	ab 2007	II	A, B, C1	http://www.goeg.at/de/Bereich/Herzschrittma-cher-und-Loop-Recorder-Register.html
Hüftendoprothesen Register	GÖG	Qualitätssicherung	fallbezogene Online-Erhebung in den Zentren	Daten und Auswertungen nur für teilnehmende Zentren zugänglich	derzeit Pilotbetrieb	Ia	A, B, C1, C2, C3, C4, F	http://www.goeg.at/de/Bereich/Hueften-doprothesen-Register.html
Prothesenregister Tirol	IET	Qualitätssicherung	fallbezogene Patientenfragebögen, Operationsdaten (Eingabemaske od. Fragebogen)	aggregierte Auswertungen öffentlich, Sonderauswertungen projektabhängig möglich (id)	ab 2003	Ia	A, B, C1, C2, C3, C4, F	http://www.iet.at/index.php?id=1617
Kinderkardiologie- Register	GÖG	Qualitätssicherung	fallbezogene Online-Erhebung in den Zentren	Daten und Auswertungen nur für teilnehmende Zentren zugänglich	derzeit Pilotbetrieb	Ia	A, B, C1, C2, C3, C4, F	http://www.goeg.at/de/Bereich/Kinderkardiologie-Register.html
Chirurgie- Qualitätsregister	GÖG	Qualitätssicherung	fallbezogene Online-Erhebung in den Zentren	Daten und Auswertungen nur für teilnehmende Zentren zugänglich	ab 2007	Ia	A, B, C1, C2, C3, C4, F	http://www.goeg.at/cxdata/media/download/goeg_qualitaetsregister_2.pdf
Stroke-Unit- Register	GÖG	Qualitätssicherung	fallbezogene Online-Erhebung in den Zentren	Daten und Auswertungen nur für teilnehmende Zentren zugänglich, evt. Veröffentlichung österreichweiter Auswertungen (id)	ab 2003	Ia	A, B, C1, C2, C3, C4, F	http://www.goeg.at/de/Bereich/Stroke-Unit-Register.html

Datenkategorie	Dateneigner*	Originalzweck	Datenerhebung	Zugang zu Daten/ Datenniveau ^α	Verfügbarkeit	Klassifikation#	Einsatz im HTA- Prozess##	Referenz
Diabetesregister Tirol	IET	Qualitätssicherung	fallbezogene elektronische Dokumentation in Institutionen, pseudonymisierte Übermittlung	aggregierte Auswertungen öffentlich, Sonderauswertungen projektabhängig möglich (id)	ab 2006	Ia	A, B, C1, C2, C3, C4, F	http://www.iet.at/index.php?id=1360
Implantatregister für Silikonimplantate	Med. Univ. Wien	Qualitätssicherung	Implantatpass	k. A.	ab 1996	Ia	C2	http://webappl1.imc.meduniwien.ac.at/implant/ImplantatregisterLoqon.jsp
Pharmakovigilanz	AGES PharmMed	Qualitätssicherung	fallbezogene verpflichtende Meldung von Angehörigen der Gesundheitsberufe und Firmen an die AGES PharmMed	öffentlich zugängliche Datenbank auf nationaler und europäischer Ebene (EMA) für nächstes Jahr geplant (id)	ab 2006	Ia	C2	http://www.basg.at/pharmakovigilanz/
Qualitätsmangelmeldung Arzneimittel	AGES PharmMed	Qualitätssicherung	fallbezogene verpflichtende Meldung von Ärzten, Apotheken, Gewerbetreibenden und Zulassungsinhabern	chronologische Veröffentlichung der Meldungen öffentlich zugänglich, keine Datenbank	jeweils aktuelle Meldungen	(Ia)	C2	http://www.basg.at/arsneimittel/arsneimittel-qualitaet/qualitaetsmangelmeldung/
Risiko-, krankheits- oder interventionsbezogene Register								
Klinische Tumorregister (Beispiele)								
Kopf-Halstumorregister	Arbeitsgemeinschaft medikamentöse Tumorthherapie	Forschung, Behandlungsverbesserung	fallbezogene elektronische Datenerhebung an Kliniken	Zugang nur über Publikationen in wissenschaftlichen Zeitschriften (id)	ab 2011	Ia	A, B, C1, C2, C3, C4, F	http://www.agmt.at/index.php?m=register&s=allregistries#3 http://www.agmt.at/files/hno/Hintergrundinformation_Kopf-Halstumor-Register.doc
Registry and Biobank for the collection of clinical data and biomaterial from adult ALL patients	Arbeitsgemeinschaft medikamentöse Tumorthherapie	Forschung, Behandlungsverbesserung	fallbezogene elektronische Datenerhebung an Kliniken	Zugang nur über Publikationen in wissenschaftlichen Zeitschriften (id)	ab 2011	Ia	A, B, C1, C2, C3, C4, F	http://www.agmt.at/index.php?m=register&s=allregistries#7

Datenkategorie	Dateneigner*	Originalzweck	Datenerhebung	Zugang zu Daten/ Datenniveau ^α	Verfügbarkeit	Klassifikation#	Einsatz im HTA-Prozess##	Referenz
Register zu „Surgical Resection of Solitary Metastases in Patients with Metastatic Breast Cancer“	Arbeitsgemeinschaft medikamentöse Tumorthherapie	Forschung, Behandlungsverbesserung	fallbezogene elektronische Datenerhebung an Kliniken	Zugang nur über Publikationen in wissenschaftlichen Zeitschriften (id)	ab 2011	Ia	A, B, C1, C2, C3, C4, F	http://www.aqmt.at/index.php?m=register&s=allregistries#8
Register zu „Screening for human epidermal growth factor receptor 2 (HER2) positivity in patients with inoperable locally advanced or metastatic gastric or gastro-esophageal junction (GEJ) cancer“	Arbeitsgemeinschaft medikamentöse Tumorthherapie	Forschung, Behandlungsverbesserung	fallbezogene elektronische Datenerhebung an Kliniken	Zugang nur über Publikationen in wissenschaftlichen Zeitschriften (id)	ab 2011	Ia	A, B, C1, C2, C3, C4, F	http://www.aqmt.at/index.php?m=register&s=allregistries#2
Register zu „Biomarkers in Metastatic Colorectal Cancer“	Arbeitsgemeinschaft medikamentöse Tumorthherapie	Forschung, Behandlungsverbesserung	fallbezogene elektronische Datenerhebung an Kliniken	Zugang nur über Publikationen in wissenschaftlichen Zeitschriften (id)	ab 2010	Ia	A, B, C1, C2, C3, C4, F	http://www.aqmt.at/index.php?m=register&s=allregistries#4
Register zu „VIDAZA® in patients with MDS, CMML, or AML“	Arbeitsgemeinschaft medikamentöse Tumorthherapie	Forschung, Behandlungsverbesserung	fallbezogene elektronische Datenerhebung an Kliniken	Zugang nur über Publikationen in wissenschaftlichen Zeitschriften (id)	ab 2009	Ia	A, B, C1, C2, C3, C4, F	http://www.aqmt.at/index.php?m=register&s=allregistries#5
Österreichisches Register für Medizinprodukte	BMG / Gesundheit Österreich GmbH (GÖG)	Dokumentation	produktbezogene elektronische Registrierung an der GÖG	nicht öffentlich; Auswertungen evt. über Anfragen beim BMG (id)	ab 2002	II	C1	http://www.goeg.at/de/Medizinprodukte-Register
IVF-Register	GÖG	Administrative Zwecke, Qualitätssicherung, Dokumentation	fallbezogene Online-Erhebung in den Zentren	aggregierte Auswertungen öffentlich über Jahresberichte (id)	an GÖG ab 2000	Ib	A, B, C1, C2, C3, C4, F	http://www.goeg.at/de/Berichte-Service/IVF-Register.html
Österreichisches Hämophilie-Register	Med. Univ. Wien	keine Information erhältlich	keine Information erhältlich	keine Information erhältlich	keine Info erhältlich	keine Klassifikation möglich	keine Aussage möglich	https://www.studienregister.at/web/guest/home http://www.meduniwien.ac.at/homepage/news-und-

Datenkategorie	Dateneigner*	Originalzweck	Datenerhebung	Zugang zu Daten/ Datenniveau ^α	Verfügbarkeit	Klassifikation#	Einsatz im HTA- Prozess##	Referenz
								topstories/?tx_ttnews%5Btt_news%5D=56&cHash=4de0f0dea7
Österreichisches Dialyse- und Transplantationsregister	Österreichische Gesellschaft für Nephrologie Arge Österreichisches Dialyse- und Transplantationsregister	Dokumentation, Qualitätssicherung	fallbezogene Erhebung der Daten von ARGE Dialyseregister + Austrotransplant	aggregierte Auswertungen öffentlich über Jahresberichte (id)	ab 2000	Ia	A, B, C1, C2, C3, C4, F	http://www.nephro.at/IB_all.htm
Austrian National CathLab Registry (ANCLAR)	Med. Univ. Innsbruck	Dokumentation, Qualitätssicherung	fallbezogene Fragebogenerhebung an den Zentren	aggregierte Auswertungen öffentlich über Jahresberichte (id)	ab 1992	Ia	A, B, C1, C2, C3, C4, F	http://iik.i-med.ac.at/index.en.php
Arzneispezialitätenregister	AGES PharmMed	Dokumentation	Veröffentlichung von AGES PharmMed	öffentlicher Online-Zugang (d)	jeweils aktuelle Einträge	II, IV	C1	http://pharmaweb.ages.at/index.jsf
Lebendspenderegister (derzeit im Aufbau)	GÖG	Verlaufskontrolle der Spender/innen, Transparenz und Situationsanalyse, Qualitätssicherung	fallbezogene elektronische Registrierung an der GÖG durch Dateneingabe des Transplantationszentrums, postalische Übermittlung der Nachuntersuchungsdaten von niedergelassenen Ärzten	noch offen	derzeit im Aufbau	I	A, B, C1	http://www.goeg.at/cxdata/media/download/berichte/TX_IB_2010.pdf
Biologica-Register für entzündlich-rheumatische Erkrankungen	Verein BioReg	Forschung, Behandlungsverbesserung	fallbezogene Weiterleitung routinemäßig erhobener Patientendaten durch Vereinsmitglieder	1/2-jährlicher Report der Fa. BIOCONSULT; derzeit nicht öffentlich	ab 2010	Ia	A, B, C1, C2, C3, C4, F	http://bioreg.at/
Sonstige								
Widerspruchsregister	GÖG	Dokumentation	fallbezogene Registrierung von individuellen Personen gemäß Antragsformular	anonymisierte Auswertungen in Transplant Jahresbericht (id)	ab 1995	n. r.	A	http://www.goeg.at/de/Widerspruchsregister http://www.goeg.at/cxdata/media/download/berichte/TX_IB_2010.pdf

Datenkategorie	Dateneigner*	Originalzweck	Datenerhebung	Zugang zu Daten/ Datenniveau ^α	Verfügbarkeit	Klassifikation#	Einsatz im HTA- Prozess##	Referenz
Critical Incident Reporting System (CIRS)	Österreichische Ärztekammer	Qualitätssicherung	fallbezogene Dokumentation von Angehörigen d. Gesundheitsberufe	öffentlicher Online-Zugang (d)	ab 2009	(Ib)	A	http://www.cirsmedical.at/
3) Surveys								
Österreichische Befragung zur Patientenzufriedenheit	GÖG	Politikgestaltung/Planung/Steuerung	Fragebogenerhebung bei Patienten nach Entlassung	Pilotphase (id)	ab 2011 (Pilot)	n. r.	A, C4	http://www.goeg.at/de/Bereich/Patientenzufriedenheit.html
Blutverbrauchs-erhebung	GÖG	Optimierung der Versorgung mit Blutprodukten, Qualitätssicherung	Online-Meldung einmal jährlich	aggregierte Auswertungen öffentlich, detaillierte Auswertungen nicht möglich (id)	ab 1999	II	A, B, C1, C4, F	http://www.goeg.at/de/Bereich/Blutverbrauchs-Erhebung.html
Gesundheitsbefragung	Statistik Austria	Politikgestaltung/Planung/Steuerung	Stichprobenerhebung mittels computeras-sistierter personal interviewing	Auswertungen öffentlich; anonymisierte Einzeldaten auf Anfrage kostenfrei erhältlich (d, id)±	ab 2006 als eigene Erhebung			
Mikrozensus Austria – z. B. Arbeitskräfte-erhebung	Statistik Austria	Politikgestaltung/Planung/Steuerung	Stichprobenerhebung	aggregierte Auswertung öffentlich, Sonderauswertung gegen Kostenersatz (d, id)±	ab 2004	IV, IIIc	C1, C3, C4	http://www.statistik.at/web_de/fragebogen/private_haushalte/mikrozensus/index.html
Konsumerhebung	Statistik Austria	Politikgestaltung/Planung/Steuerung	Stichprobenerhebung	aggregierte Auswertung öffentlich, Sonderauswertung gegen Kostenersatz (d, id)±	ab 1999	IV	C1, C3, C4	http://www.statistik.at/web_de/static/konsumerhebung_200405_016895.doc http://www.statistik.at/web_de/services/mikrodaten_fuer_forschung_und_lehre/datenangebot/standardisierte_datensaetze_sds/index.html#index4
WHO health survey	WHO	Politikgestaltung/Planung/Steuerung	Stichprobenerhebung	aggregierte Auswertung öffentlich (id)	ab 2002	IIIc	A, B, C1, C4	http://www.who.int/healthinfo/survey/en/

Datenkategorie	Dateneigner*	Originalzweck	Datenerhebung	Zugang zu Daten/ Datenniveau ^α	Verfügbarkeit	Klassifi- kation#	Einsatz im HTA- Prozess##	Referenz
EU-SILC: Einkommen, Armut und Lebensbedingungen ...	EUROSTAT/Statistik Austria	Politikgestaltung/Planung/Steuerung	Stichprobenerhebung	aggregierte Auswertung öffentlich, Sonderauswertung gegen Kostenersatz (d, id)±	ab 1995	IIIc	C1, C4	http://www.statistik.at/web_de/fragebogen/private_haushalte/eu_silc/index.html#index10 http://www.bmsk.gv.at/cms/site/dokument.html?doc=CMS1237196950667&channel=CH0020
HBSC-Survey (Health Behaviour in School aged children)	WHO, GÖG	Politikgestaltung/Planung/Steuerung	Stichprobenerhebung	aggregierte Auswertung öffentlich, Auswertungen gegen Kostenersatz (id)	ab 1985	IIIc	A, B, C1, C4	http://www.hbsc.org/
4) Amtliche nationale u. internationale Statistiken, zumeist auf Grundlage eines gesetzlichen Auftrags								
Gesundheitsstatistik Bevölkerungs- und Geburtenstatistik Todesursachen Krebserkrankungen Stationäre Aufenthalte Gesundheitszustand Gesundheitsdeterminanten Unfälle Gesundheitsversorgung Gesundheitsausgaben Gesundheitsindikatoren (Jahrbuch Gesundheitsstatistik)	Statistik Austria	Politikgestaltung/Planung/Steuerung	div. nationale Register, administrative Daten	aggregierte Auswertung öffentlich, Sonderauswertung gegen Kostenersatz (id)	je nach Art der Daten zwischen 1983 und 2006	IIIb IVb	A, B, C1, C4	http://www.statistik-austria.at/web_de/statistiken/gesundheit/index.html
Ärztelkostenstatistik	GÖG	Politikgestaltung/Planung/Steuerung	administrative Daten der Sozialversicherungsträger	Auswertungen gegen Kostenersatz (id)	ab 1991	IVa	A, B, C4, F	

Datenkategorie	Dateneigner*	Originalzweck	Datenerhebung	Zugang zu Daten/ Datenniveau ^α	Verfügbarkeit	Klassifikation#	Einsatz im HTA- Prozess##	Referenz
Daten aus der Zahnstaterhebung	GÖG	Politikgestaltung/Planung/Steuerung	Primärerhebung in Zufallsstichprobe	aggregierte Auswertung öffentlich, Auswertungen gegen Kostenersatz (id)	ab 2006	IIIc	A, B, C1, C4	
Pflegegeldstatistik	Statistik Austria, GÖG	Dokumentation/Administration	div. administrative Prozesse	aggregierte Auswertung öffentlich, Auswertungen gegen Kostenersatz (id)	ab 2006	IV	A, B, C3, C4	http://www.statistik.at/web_de/statistiken/soziales/sozialleistungen_auf_bundesebene/bundespflegegeld/index.html
Bevölkerungsstatistik	Statistik Austria	Politikgestaltung/Planung/Steuerung	Volkszählung/Registrierung + div. Statistiken	aggregierte Auswertung öffentlich, Sonderauswertung gegen Kostenersatz (d, id)±	ab 1952	n. r.	A, B, C4	http://www.statistik.at/web_de/statistiken/bevoelkerung/index.html
Familien- und Haushaltsstatistik (aus Mikrozensus)	Statistik Austria	Politikgestaltung/Planung/Steuerung	Registrierung	aggregierte Auswertung öffentlich, Sonderauswertung gegen Kostenersatz (d, id)±	ab 2004	n. r.	A, B, C1, C4	http://www.statistik.at/web_de/services/publikationen/3/index.html
Einkommensstatistik	Statistik Austria	Politikgestaltung/Planung/Steuerung	fallbezogene Lohnsteuerdaten, Verdienststrukturerhebung	aggregierte Auswertung öffentlich, Sonderauswertung gegen Kostenersatz (d, id)±	ab 1997	IV	C3, C4	http://www.statistik.at/web_de/statistiken/soziales/index.html
Umweltdaten	Umweltbundesamt / GÖG	Politikgestaltung/Planung/Steuerung	Luftgütemessung und Erhebung der Grundwasserqualität des Umweltbundesamts	Auswertung gegen Kostenersatz (id)	ab 1997	n.r.	A, B, C1, C4	http://www.umweltbundesamt.at/
Arbeitsmarktstatistik	Statistik Austria	Politikgestaltung/Planung/Steuerung	div. administrative Prozesse	tlw. öffentlich über Berichte (id)	ab 2002	IV	C3, C4	http://www.statistik.at/web_de/statistiken/arbeitsmarkt/index.html

Datenkategorie	Dateneigner*	Originalzweck	Datenerhebung	Zugang zu Daten/ Datenniveau ^α	Verfügbarkeit	Klassifikation#	Einsatz im HTA- Prozess##	Referenz
Pharma- Preisinformation	GÖG	Informati- on/Planung/Steueru ng	Recherche länderspe- zifischer Daten	Auswertungen möglich gegen Kostenersatz (id)	abhängig vom untersuchten Land	IVc	C3, C4, F	http://www.goeg.at/de/PPI
OECD family database	OECD	Politikgestal- tung/Planung/Steue rung	div. aggregierte OECD-Daten (z. B. social expenditure, benefits and wages, education)	aggregierte Auswertung öffentlich (id)	ab mid-2000s	IV	A, B, C1, C3, C4	http://www.oecd.org/document/4/0,3746,en_2649_34819_37836996_1_1_1_1,00.html
OECD health data	OECD	Politikgestal- tung/Planung/Steue rung	div. aggregierte nationale Gesund- heitsstatistiken	aggregierte Auswertung teilw. öffentlich, detaillierte Daten gegen Kostenersatz (id)	ab 1960 abhängig vom Indikator	III, IV	A, B, C1, C3, C4	http://www.oecd.org/document/30/0,3746,en_2649_33929_12968734_1_1_1_1,00.html
WHO Health statistics and health information systems	WHO	Politikgestal- tung/Planung/Steue rung	div. aggregierte nationale Gesund- heitsstatistiken	aggregierte Auswertung öffentlich über Berichte (id)	abhängig vom Indikator	III, IV	A, B, C1, C3, C4	http://www.who.int/healthinfo/statistics/en/
WHO Europe Database	WHO	Politikgestal- tung/Planung/Steue rung	div. aggregierte nationale Gesund- heitsstatistiken	aggregierte Auswertung tlw. öffentlich (id)	ab 1970 abhängig vom Indikator	III, IV	A, B, C1, C3, C4	http://www.euro.who.int/en/what-we-do/data-and-evidence/databases
The World Bank Health Indicators	World Bank	Politikgestal- tung/Planung/Steue rung	div. aggregierte Daten und Statisti- ken, Primärerhebun- gen	aggregierte Auswertung öffentlich (id)	ab 1960 abhängig vom Indikator	III, IV	A, B, C1, C3, C4	http://data.worldbank.org/topic/health http://data.worldbank.org/indicator/SH.DYN.MORT/countries/AT-SEEU-GB?display=graph
Freizeitunfallsta- tistik	Kuratorium für Verkehrssicherheit	Politikgestal- tung/Planung/Steue rung	div. Statistiken (Verkehrsunfall-, Spitals-, Todesursa- chenstatistik etc.)	aggregierte Auswertung öffentlich (id)	ab 2001	III	A, B, C1, C4	http://www.kfv.at/heim-freizeit-sport/freizeitunfallstatistik/
Straßenverkehrs- unfälle	Kuratorium für Verkehrssicherheit	Politikgestal- tung/Planung/Steue rung	div. Statistiken	aggregierte Auswertung öffentlich, detaillierte Daten gegen Kostenersatz (id)	ab 2000	III		http://www.kfv.at/unfallstatistik/index.php?id=57
Grunddaten des österreichischen Schulwesens	BM f. Unterricht, Kunst und Kultur	Politikgestal- tung/Planung/Steue rung	div. administrative Prozesse	aggregierte Auswertung öffentlich (id)	ab 1990 abhängig vom Indikator	n.r.	A, B, C1, C4	http://www.bmukk.gv.at/schulen/bw/ueberblick/grunddaten.xml

Datenkategorie	Dateneigner*	Originalzweck	Datenerhebung	Zugang zu Daten/ Datenniveau ^α	Verfügbarkeit	Klassifikation#	Einsatz im HTA- Prozess##	Referenz
Daten der Stellungs- untersuchung	Verteidigungsministerium (übermittelt an GÖG)	Dokumentation	Primärdatenerhebung bei Stellungsuntersuchung	Auswertungen möglich gegen Kostenersatz (id)	ab 1990 abhängig vom Indikator	III	A, B, C1	
Sozialstatistik Personeneinkommen Haushaltseinkommen Armut und soziale Eingliederung Tariflohnindex Arbeitskosten Verbrauchsausgaben Ausstattung privater Haushalte Gender-Statistik Zeitverwendung Freiwilligenarbeit Sozialschutz nach EU-Konzept Sozialleistungen auf Bundesebene Sozialleistungen auf Landesebene Kriminalität	Statistik Austria	Politikgestaltung/Planung/Steuerung	div. Methoden, fallbezogene und aggregierte Daten	aggregierte Auswertung öffentlich, Sonderauswertung gegen Kostenersatz (d, id) ±	ab 1997 abhängig vom Indikator	IV	A, B, C3, C4	http://www.statistik.at/web_de/statistiken/soziales/index.html
EUROSTAT-Daten (z. B. Arbeitslosigkeit)	EUROSTAT	Politikgestaltung/Planung/Steuerung	div. Methoden (z. B. Stichprobenerhebung)	aggregierte Auswertung öffentlich (id)	abhängig vom Indikator	IV	A, B, C3, C4	http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database http://epp.eurostat.ec.europa.eu/portal/page/portal/employment_unemployment_ifs/data/database
5) Daten zu Planungszwecken								
Erreichbarkeit und natürliche Einzugsgebiete	GÖG	Planung/Steuerung	Berechnung nach „Kürzeste-Weg-Matrix“ (Reisezeiten im motorisierten Individualverkehr	Auswertung gegen Kostenersatz (id)	Stand 2006	n. r.	A, B, C4	

Datenkategorie	Dateneigner*	Originalzweck	Datenerhebung	Zugang zu Daten/ Datenniveau ^α	Verfügbarkeit	Klassifikation#	Einsatz im HTA- Prozess##	Referenz
			zwischen Gemeinden)					
Österreichischer Strukturplan Gesundheit (ÖSG)	GÖG	Planung/Steuerung	div. fallbezogene und aggregierte Datenquellen	öffentlich (id)	ab 2006	(II)	A, B, C1, C4, F	http://www.goeg.at/de/BerichtDetail/Oesterreichischer-Strukturplan-Gesundheit-2010.html
Regionale Strukturpläne Gesundheit (RSG)	Länder	Planung/Steuerung	siehe ÖSG	tlw. öffentlich (id)	ab 2006 (abhängig von Bundesland)	(II)	A, B, C1, C4, F	
Rehabplan	HVB	Planung, Steuerung der Rehabilitationsleis- tungen	div. fallbezogene Datenquellen (Ist- Stand-Erhebung, Inanspruchnahmeda- ten, Antrags-, Erledigungs- u. Einweisungsdaten, LKF-Daten)	Bericht öffentlich (id)	ab 2004	II, IV	A, B, C1, C4, F	http://www.hauptverband.at/mediaDB/532621_Rehabilitationsplan%202009_Publikation%2024.2.2009.pdf
6) Datenbestände zu Versorgungsstrukturen								
Spitalskompass	GÖG	Information Leistungsangebot	aggregierte Daten der Leistungs- und Diagnosedokumenta- tion (BMG), Aktualisierung von den Kranken- anstalten	öffentlich (id)	jeweils aktuelles Angebot	II, IV	A, B, C1, C4	http://www.goeg.at/de/Kompass/Spitalskompass.html
Rehakompass	GÖG	Information Leistungsangebot	Aktualisierung durch Rehab-Einrichtungen bzw. -Träger	öffentlich (id)	jeweils aktuelles Angebot	II, IV	A, B, C1, C4	http://www.goeg.at/de/Kompass/Rehakompass.html
Suchthilfekompass	GÖG	Information Leistungsangebot	Aktualisierung durch Einrichtungen	öffentlich (id)	jeweils aktuelles Angebot	II, IV	C4	http://www.goeg.at/de/Kompass/Suchthilfekompass.html
Krankenanstalten in Zahlen	BMG	Information Leistungsangebot	Zusammenstellung administrativer Daten	öffentlich (id)	seit 2000	II, IV	C4	http://www.kaz.bmg.gv.at/index.htm

Datenkategorie	Dateneigner*	Originalzweck	Datenerhebung	Zugang zu Daten/ Datenniveau ^α	Verfügbarkeit	Klassifikation#	Einsatz im HTA- Prozess##	Referenz
Ärzte- und Zahnärzteliste	Ärzttekammer, (GÖG)	Politikgestaltung/Planung/Steuerung	administrative Daten der Ärztekammer	öffentlich, Sonderauswertungen an der GÖG gegen Kostenersatz (id)	ab 1999	IV	C4	
Daten zu medizinischen Berufsgruppen	Kammern und Berufsverbände (GÖG)	Dokumentation	div. administrative Daten	an der GÖG tief gegliedert nach Berufen, auch nicht-stationärer Bereich, gegen Kostenersatz (id)	2003	IV	C4	
Kapazitäten der Alten- und Langzeitversorgung	GÖG	Planung/Steuerung	unregelmäßige Erhebungen der GÖG im Auftrag von Sozial- u. Gesundheitsministerium für einzelne Jahre	Daten in GÖG-Datenbank nicht öffentlich; Ergebnisse werden in Berichtsform an Auftraggeber übermittelt (id)	2003	II, IV	A, B, C1, C3, C4	
7) Sonstige								
GAP-DRG (linkage zwischen stationären und ambulanten Daten)	HVB	Planung/Steuerung	Verknüpfung diverser fallbezogener administrativer Daten	projektabhängig (d, id)		II, IV	A, B, C1, C2, C3, C4, F	
ÖGIS (Österreichisches Gesundheitsinformationssystem)	GÖG	Planung/Steuerung	Verknüpfung diverser fallbezogener und aggregierter administrativer Daten und amtlicher Statistiken	Auswertungen gegen Kostenersatz (id)	unterschiedlich je nach Verfügbarkeit in verwendeten Primärdaten	I – IV	A, B, C1, C4	http://www.goeg.at/de/Berichte-Service/OeGIS1.html
REGIS (Regionales Gesundheitssystem)	GÖG	Planung/Steuerung	kartographische regional vergleichende Darstellung v. Eckdaten zu gesundheitsrelevanten Indikatoren	öffentlich (id)	siehe REGIS	I – IV	A, B, C1, C4	http://regis.goeg.at/

Datenkategorie	Dateneigner*	Originalzweck	Datenerhebung	Zugang zu Daten/ Datenniveau [±]	Verfügbarkeit	Klassifikation [#]	Einsatz im HTA- Prozess ^{##}	Referenz
Budgets der Gebietskörper- schaften	Jeweilige Gebiets- körperschaft		öffentliche Budgetdokumentati- on	öffentlich (id)	für jedes Jahr	IV	A, C4	

* Institutionen, die die Daten erheben, speichern und nützen sowie die rechtliche Verfügungsgewalt darüber besitzen; # I: Daten enthalten Informationen zu der zu evaluierende Technologie UND zu den relevanten Gesundheitszuständen; Ia: klinische bzw. Qualitätsregister (z. B. Chirurgie-Qualitätsregister); Ib: klinisch-administrative Datensets (z. B. Krankengeschichten); Ic: Bevölkerungsbasierte Datensets (z. B. Zervixkarzinom-Screening-Register in Großbritannien); Id: Sonstige; II: Daten enthalten Informationen zu der zu evaluierenden Technologie, aber nicht zu den relevanten Gesundheitszuständen (z. B. Medizinproduktregister); III: Daten enthalten Informationen zu den relevanten Gesundheitszuständen, aber nicht zu der zu evaluierenden Technologie; IIIa: adverse events reporting und vertrauliche Untersuchungen (z. B. confidential enquiry into stillbirths and deaths in infancy); IIIb: Krankheitsregister (z. B. Tumorregister); IIIc: Gesundheits-Surveys (z. B. HBSC-Survey); IV: Daten enthalten Informationen für Kostenberechnungen; IVa: Ressourcenverbrauch UND Preise (z. B. Arzneimittelverordnungen); IVb: Nur Ressourcenverbrauch; IVc: Nur Preise (z. B. Pharma-Preisinformationssystem); ##: A: Themenfindung/Priorisierung; B: Scoping; C: Durchführung des Assessments; C.1: Hintergrund (Statusbestimmung zu Krankheit und Technologie); C.2: Einschätzung der Wirksamkeit/Sicherheit; C.3: Wirtschaftlichkeit; C.4: Abschätzung der Tragweite (ökonomisch, organisatorisch, juristisch, ethisch, psychosozial); D: Dissemination der Ergebnisse; E: Umsetzung; F: Evaluierung des Impacts; k. A.: keine Angabe; ± Die Bundesanstalt Statistik Austria stellt ausgewählte Mikrodaten der amtlichen Statistik zur Nutzung in der wissenschaftlichen Forschung und Lehre bereit; [±] d: direkter Zugang zu (pseudonymisierten/anonymisierten) Originaldaten; id: indirekter Zugang über Aggregatstatistiken oder Sonderauswertungen.

Quelle: LBI-eigene Darstellung 2011

15.5 Dateneigenschaften ausgewählter Datenbestände

15.5.1 Administrative Daten

15.5.1.1 Daten der Sozialversicherung

Originalzweck

Die Daten der Sozialversicherung entstammen verschiedenen Quellen (Vertragspartnerdaten, Versichertendaten, Krankenstandsmeldungen etc.) und entstehen durch die Erfüllung gesetzlicher Aufgaben. Sie dienen administrativen Zwecken und der Leistungsabrechnung.

Datenerhebung/Registrierung

Die Datenerhebung erfolgt durch unterschiedliche administrative Prozesse (z. B. Meldung der Beschäftigungsverhältnisse durch Dienstgeber, Vertragspartnerabrechnung) bei den einzelnen Sozialversicherungsträgern bzw. auf Ebene des Hauptverbandes. Datenauswertungen erfolgen durch Zusammenführung der verschiedenen „Vorsysteme“. Auf Basis von Richtlinien für die Zusammenarbeit zwischen Sozialversicherungsträgern werden zunehmend einheitliche EDV-Programme (sog. Standardprodukte) zur Datenerfassung verwendet (Hauptverband der österreichischen Sozialversicherungsträger 2011). Beispiele dafür sind die ärztliche Eigen- und Folgekostenanalyse (FOKO) der einzelnen Krankenversicherungsträger oder die zentrale Partnerverwaltung (ZPV), ein einheitliches zentrales Stammdatenverzeichnis für Dienstgeber, Versicherte und Leistungserbringer (zusammenfassend als „Partner“ bezeichnet).

Zugang zu Daten

Eine Reihe von aggregierten Daten ist elektronisch verfügbar und öffentlich zugänglich, wie z. B. die Daten des „Statistischen Handbuchs der Sozialversicherung“ (Hauptverband der Sozialversicherungsträger 2010). Datenauswertungen von Individualdaten sind möglich, wenn die entsprechende Rechtsgrundlage gegeben ist (nachvollziehbarer Rechtstitel wie z. B. universitäre Einrichtungen, Studierende) und/oder die Auswertung von öffentlichem Interesse ist. Daten können außerdem im Namen von Patienten/Patientinnen angefordert werden, sofern eine Zustimmung der Betroffenen vorliegt.

Dateneigenschaften/Validität

Die Gesamtheit der Daten kann unterteilt werden in

- » Personenstandsdaten / versichertenbezogene Rohdaten
 - » Daten zu Personen als Versicherte oder Mitversicherte
 - » ihre Stellung im Sozialversicherungssystem (erwerbstätig, in Mutterschutz, arbeitslos etc.) im Zeitverlauf (monatliche Tatbestände)
 - » Einkommen
- » Daten zu den Beziehungen zwischen Sozialversicherungsträger und Leistungserbringer
 - » Krankenanstalten
 - » Ärztliche Vertragspartner und diesen Gleichgestellte (z. B. Physiotherapeuten/Physiotherapeutinnen)
 - » Apotheker
 - » Sonstige
- » Leistungsdaten
 - » Abrechnungen von Sachleistungen
 - » Sonstige Leistungsdaten
 - » Geldleistungen (Wochengeld, Krankengeld, Unfallrenten, Pflegegeld)
 - » Kostenerstattungen (nach Wahlarztbesuchen)

Aus diesen Daten können – zusätzlich zu den periodisch durchgeführten Auswertungen – diverse Sonderauswertungen getätigt werden (z. B. zur Häufigkeit erbrachter Leistungen, zu Leistungstarifen, zur Anzahl von Vertragspartnern pro Bundesland). Dazu stehen systeminterne Datenbanken zur Verfügung (z. B. die Honorarordnungsverwaltung online und die Metahonorarordnung, die die unterschiedlichen Honorarordnungen in ein einheitliches Dokumentationssystem überführt).

Ein wesentliches Merkmal ist der hohe Grad an Vollständigkeit der Daten. Die Pflichtversicherung in der Krankenversicherung bedingt etwa, dass über die Daten fast hundert Prozent der Bevölkerung erfasst sind.

Die Daten haben jedoch bestimmte Eigenschaften, die bei der Interpretation zu beachten sind. Beispielweise enthält die Krankenversichertenstatistik, die im statistischen Handbuch ausgewiesen wird, nur jene Versicherten, für die ein Beitrag zur gesetzlichen Krankenversicherung entrichtet wird, während beitragsfrei mitversicherte Angehörige nicht erfasst sind. Gezählt werden außerdem nicht die krankenversicherten Personen, sondern die Versichertenverhältnisse. Dies gilt auch für die Daten der Pensions- und Unfallversicherung. Daten zu Leistungstarifen, die über die Online-Honorarordnungsverwaltung ermittelt werden, sind dadurch gekennzeichnet, dass die ausgewiesenen Tarife höher sein können als der in der Abrechnungsrealität gezahlte Tarif (siehe nachfolgendes Beispiel Metahonorarordnung). Bei der Verwendung der

Daten für wissenschaftliche Zwecke ist daher für die korrekte Interpretation eine Rücksprache mit den entsprechenden Datenexperten/Datenexpertinnen notwendig.

15.5.1.1.1 Beispiel Honorarordnungsverwaltung online und Metahonorarordnung (Wilbacher und Schröder 2006)

Originalzweck

Mit der Honorarordnungsdatenbank (seit Umstellung auf eine webbasierte Oberfläche „Honorarordnungsverwaltung online“) werden seit Mitte der 1990er Jahre die unterschiedlichen Honorarordnungen der Krankenversicherungsträger in ein einheitliches Dokumentationssystem übergeführt. Die Datenbank dient als Basis für die Metahonorarordnung, die ein Register für die Zusammenfassung aller Leistungspositionen der KV-Träger darstellt und einen raschen Überblick über vergleichbare Leistungen und deren Tarife bietet. Das Register beinhaltet rund 1.600 Positionen. Mit jeder Metahonorarordnungsposition sind die entsprechenden, vergleichbaren Positionen der Versicherungsträger verknüpft. Die Metahonorarordnungspositionen werden außerdem in den Katalog ambulanter Leistungen (KAL) übergeführt, in dem sie auf die entsprechenden KAL-Positionen „gemappt“ werden (siehe Katalog ambulanter Leistungen).

Datenerhebung/Registrierung

Die Kodierung der Positionen der Honorarordnungen und der Metahonorarordnung erfolgt numerisch nach folgender Systematik: Leistungsgruppe, Leistungsart, Leistungsposition. Aktualisierungen liegen in der Verantwortung der einzelnen KV-Träger.

Zugang zu Daten

Die Honorarordnungen der Krankenversicherungsträger sind online über den Hauptverband der österreichischen Sozialversicherung (www.avsv.at) oder bei einzelnen KV-Trägern zugänglich. Für die Metahonorarordnung werden SV-intern Zugriffsberechtigungen vergeben. Externe Institutionen haben keinen Zugriff. Externe Anfragen werden im Einzelfall mit der Generaldirektion des Hauptverbandes abgeklärt.

Dateneigenschaften/Validität

Die Metahonorarordnung umfasst ca. 1.600 Positionen, die nach Leistungsgruppen, -arten, und -positionen eingeteilt sind. Tarife zu Metahonorarordnungspositionen sind nur dann erfasst, wenn es sich um bundesweit gültige Gesamtverträge handelt (z. B. Mutter-Kind-Pass) bzw. wenn bei der Erarbeitung eines Musterkataloges Tarife kalkuliert wurden (Musterkataloge werden für Bereiche mit hohem Honoraraufwand wie z. B.

Laborleistungen von Experten/Expertinnen der KV-Träger und des Hauptverbandes unter medizinischer Beratung erarbeitet und dienen den KV-Trägern als Verhandlungsunterlage und Empfehlung). Für andere Leistungspositionen muss bei Fragen zu Tarifen ein Durchschnittstarif aus den Trägerpositionen gerechnet werden. Die Leistungs- und Tarifaktualisierungen liegen in der Verantwortung der Kassen und erfolgen derzeit meist erst nach Gültigkeitsbeginn (keine Tagesaktualität). Zukünftig wird eine Schnittstelle es ermöglichen, dass jede Änderung des Trägers gleichzeitig an die Honorarordnungsdatenbank gemeldet wird. Durch Regelungen bei der Vertragsgestaltung mit den Vertragspartnern (Limitierungen bzw. Degression von bestimmten Tarifen bei Mehrfacherbringung) kann der in der Honorarordnungsdatenbank ausgewiesene Tarif höher sein als der in der Abrechnungsrealität gezahlte Tarif.

15.5.1.1.2 Beispiel Vorsorgeuntersuchung

Originalzweck und Datenerhebung

Die Daten zu den in Anspruch genommenen Leistungen der Vorsorgeuntersuchung entstehen im Rahmen von administrativen Prozessen (Leistungsabrechnung).

Zugang zu Daten

Der Zugang (in Form von Sonderauswertungen) ist projektabhängig. Die Originaldaten werden außerdem an die Gesundheit Österreich GmbH (GÖG) übermittelt. Für Auswertungen ist – sofern diese möglich sind – ein Kostenersatz zu leisten.

Dateneigenschaften/Validität

Die Daten beinhalten alle über die Sozialversicherung verrechneten Vorsorgeuntersuchungen (für je ein Kalenderjahr) nach Geschlecht und Bundesland. Nicht enthalten sind privat bezahlte Untersuchungen oder solche von nicht versicherten Personen.

15.5.1.2 Klinische Dokumentation des Rehabilitationsverlaufes

Originalzweck

Der Rehabilitationsverlauf wird in unterschiedlicher Form und mit unterschiedlichem Ziel dokumentiert (z. B. für interne oder externe Qualitätssicherung, Institutionenvergleich, klinikinterne Kostenrechnung) (Abuzahra und Piso 2010).

Datenerhebung

Die Dokumentation der Daten erfolgt in unterschiedlicher Form (elektronisch, manuell), von verschiedenen Berufsgruppen und zu unterschiedlichen Zeitpunkten während eines Rehabilitationsprozesses bzw. –aufenthalts.

Zugang zu Daten

Die Zustimmung zum Datenzugang obliegt den Eigentümern der Rehabilitationszentren. Daten können ausschließlich in pseudonymisierter Form verwendet werden.

Dateneigenschaften/Validität

Es werden einerseits medizinische Parameter dokumentiert (z. B. Blutdruck, Gewicht, BMI, HDL-/LDL-Cholesterin), andererseits werden verschiedene Instrumente (z. B. NHPT, 10m-GZm VAS, HADS-D Scores für Angst/Depression) zur Dokumentation des Rehabilitationsverlaufs herangezogen (Abuzahra und Piso 2010, Gyimesi, Piso et al. 2010). Die Datenvalidität ist durch unvollständige oder uneinheitliche Dokumentation deutlich eingeschränkt. Statistische Auswertungen sind daher nur in begrenztem Ausmaß möglich und in ihrer Aussage oft eingeschränkt (Gyimesi, Piso et al. 2010). (Siehe dazu auch Kapitel 6.2.2.)

15.5.1.3 Dokumentation ambulanter Leistungen über den Katalog ambulanter Leistungen (KAL) (Bundesministerium für Gesundheit 2010, Geiersbach 2009, Geiersbach 2008)

Originalzweck

Ziel des KAL ist die einheitliche Dokumentation ambulanter Leistungen im spitalsambulantem und niedergelassenen Bereich, zu denen es – im Gegensatz zu Therapie- oder Diagnosedokumentationen – kein allgemein akzeptiertes oder verbreitetes Klassifikationssystem gibt. Die Leistungsdokumentation dient der Beobachtung von Leistungsentwicklungen, der Quantifizierung und Bewertung von Veränderung durch Reformpoolprojekte sowie der Planung im Gesundheitswesen. Die Dokumentation befindet sich derzeit in der Pilotierungsphase.

Datenerhebung/Registrierung

Die Leistungserfassung basiert auf einer multiaxialen Struktur (CCAM / Classification Commune des Actes Médicaux), die den medizinischen Zugang (z. B. über eine Körperöffnung), die Leistungsart (diagnostisch, therapeutisch etc.) und die Grob- bzw. Feinanatomie, der die Leistung zuzuordnen ist, beschreibt. Der Katalog umfasst derzeit 360 ambulante Zielleistungen im engeren Sinn und – im Sinne eines einheitlichen stationär-ambulanten Katalogs – insgesamt 760 dokumentierbare ambulant erbringbare Leistungen, die mit einem eindeutigen Code identifizierbar sind. Die Daten aus den einzelnen Bereichen werden mit Unterstützung des Programmes ADok gemeldet und sollen pseudonymisiert in einer gemeinsamen Datenbank, die sich im Aufbau befindet, zusammenfließen.

Zugang zu Daten

Im derzeitigen Stadium der Entwicklung sind die Daten noch nicht für planerische Zwecke oder zur Veröffentlichung vorgesehen.

Dateneigenschaften/Validität

Langfristig eignen sich die Daten zur groben Analyse von Leistungsbündeln im ambulanten Versorgungsbereich zum Zweck der Gesundheitsberichterstattung und Planung sowie – mit Einschränkungen – für Detailanalyse auf Bundesland- oder Regionsebene, etwa für Vergleiche zwischen spitalsambulanten und niedergelassenem Bereich. Für Detailanalysen auf Einzelleistungsebene und eine darauf aufbauende bundesweit ambulante Leistungsangebotsplanung wurden die Daten als nicht geeignet beschrieben (Geiersbach 2009).

15.5.1.4 Dokumentation spitalsambulanter Leistungen

Originalzweck

Spitalsambulante Leistungen sollen künftig über den einheitlichen Katalog ambulanter Leistungen (KAL; siehe ebendort) dokumentiert werden. Sie wurden aber auch bisher schon – jedoch in sehr unterschiedlicher Form und Ausführung – dokumentiert. Dies erfolgte zu administrativen Zwecken (z. B. zur Verrechnung von Kosten für ausländische Selbstzahler wie z. B. verletzte Skifahrer).

Datenerhebung/Registrierung

Die Registrierung erfolgt nach unterschiedlichen (elektronischen) Systemen und Dokumentationsstrukturen.

Zugang zu Daten

Es besteht kein genereller Zugang zu den Daten. In Einzelfällen (z. B. bei Forschungs-kooperationen) können pseudonymisierte Datensätze projektspezifisch zur Verfügung gestellt werden.

Dateneigenschaften/Validität

Die Daten beinhalten Art und Frequenz der erbrachten Leistungen. Die Eignung der Daten für wissenschaftliche Zwecke hängt von der Fragestellung ab. Schwierigkeiten ergeben sich etwa, wenn die Leistungsdokumentation nicht mit einer Diagnoseklassifikation verknüpft ist (Wild, Mad et al. 2007). Möglich sind Aussagen zur Häufigkeit von Einzelbesuchen, Leistungen oder Leistungsmuster, zur Häufigkeit von Frequenzen außerhalb der Normdienstzeit oder zu Patientenpfaden.

15.5.1.5 Minimal-Basic-Dataset-(MBDS-)Daten / Daten der leistungsorientierten Krankenanstellenfinanzierung (Bundesminis- terium für Gesundheit 2009)

Originalzweck

Das MBDS umfasst die Leistungs- und Diagnosedaten der österreichischen Fondskrankenanstalten, die den Ländern und den Landesgesundheitsfonds seit 1997 als Grundlage für die leistungsorientierte Krankenanstellenfinanzierung (LKF) vorzulegen sind (gesetzliche Grundlage: Bundesgesetz über die Dokumentation im Gesundheitswesen).

Datenerhebung/Registrierung

Die Dokumentation der Daten für den Diagnose- und Leistungsbericht erfolgt von den Krankenanstaltenträgern elektronisch mittels strukturiertem Datensatz. Die Diagnoseerfassung erfolgt nach dem Diagnoseschlüssel ICD-10, die Leistungserfassung nach dem Leistungskatalog des BMG.

Zugang zu Daten

Eine Reihe von aggregierten Daten ist elektronisch verfügbar und öffentlich zugänglich (z. B. über die Gesundheitsstatistik der Statistik Austria, Website des Ministeriums). Darüber hinausgehende Auswertungen der Originaldaten können projektbezogen auf konkrete Anfrage erfolgen, sofern sie die datenschutzrechtlichen Bestimmungen erfüllen und mit den gegebenen Ressourcen zu bewältigen sind.

Dateneigenschaften/Validität

Das Datenset ist pro Krankenhausaufenthalt anzulegen und enthält administrative Daten (aufenthaltsbezogene Daten wie Aufnahmeart, Aufnahmezeitpunkt etc. und patientenbezogene Daten wie Geburtsdatum, Geschlecht etc.), medizinische Daten (Haupt- und Nebendiagnosen, medizinische Einzelleistungen) und (je nach Fonds) LKF-Daten (z. B. Punkte LDF-Pauschale, Zusatzpunkte Intensivbetreuung).

Zur Sicherstellung einer bundeseinheitlichen Dokumentation dienen das „Handbuch Organisation und Datenverwaltung“ und das „Handbuch Medizinische Dokumentation“. Die Überprüfung auf Vollständigkeit und eine Plausibilitätsprüfung sind krankenhausern und auf Länderebene vorgesehen und im Handbuch beschrieben. Zusätzlich erfolgen händische Überprüfungen und automatisierte Prüfroutinen seitens des Bundesministeriums.

Während MBDS-Daten einen hohen Grad an Vollständigkeit aufweisen, ist die Validität hinsichtlich dokumentierter Diagnosen eingeschränkt, da die Dokumentation zu Abrechnungszwecken und nicht primär zu medizinischen oder wissenschaftlichen Zwecken erfolgt. Die administrierten Diagnosen können von den tatsächlichen Diagnosen prinzipiell oder in ihrer Gewichtung (Haupt- versus Nebendiagnose) abweichen. Jede Aufnahme (inklusive Wiederaufnahme) wird als neuer Fall registriert, daher können zwar Informationen zu den Krankenhausaufenthalten gewonnen werden, nicht aber personenbezogene Auswertungen erfolgen.

Die LKF-Daten enthalten Häufigkeiten einzelner Leistungspositionen und eine dazugehörige Punktezahl. Zur Ableitung von Kostengrößen müssen diese Punkte monetär bewertet werden. Dies geschieht mithilfe eines einheitlichen Durchschnittswertes pro Punkt. Ein klassischer Ansatz zur Ermittlung eines Bewertungsfaktors für LKF-Punkte ist die Division der stationären Endkosten durch die Summe der LKF-Punkte des jeweiligen Jahres. Limitierend bei dieser Methode ist, dass der tatsächlich ausgezahlte Punktwert von den unterschiedlichen Landesfonds festgelegt wird. Die angegebenen bewerteten LKF-Punkte sind daher nur eine grobe Schätzung der Ausgaben für stationäre Leistungen. Zudem existiert im Krankenanstaltenbereich eine Reihe von nicht über das LKF-System abgegoltenen Kosten, die hier ebenfalls nicht enthalten sind.

15.5.1.6 Kostenrechnungsdaten der Krankenanstalten (Bundesministerium für Gesundheit 2010, Bundesministerium für Gesundheit Familie und Jugend 2007)

Originalzweck

Die Krankenanstalten-Kostenrechnung dient krankenhausinternen Zwecken (Schaffung von Kostentransparenz, Stärkung des Kostenbewusstseins, Grundlage für Planung und Steuerung, Kontrolle der Wirtschaftlichkeit, Erfolgs-, Kennzahlen-, Vergleichsrechnungen etc.) sowie gesetzlich vorgegebenen Dokumentations- und Informationspflichten an Dritte. Seit 1978 existiert ein bundesweit einheitliches Kostenrechnungssystem, das seither beständig weiterentwickelt wurde und das Rechnungswesen und Berichtswesen umfasst. Das Rechnungswesen inkludiert das auf handelsrechtlichen Normen basierende pagatorische Rechnungswesen und das auf betriebswirtschaftlichen Grundlagen beruhende kalkulatorische Rechnungswesen. Zusätzlich kann es weitere nicht-kameralistische sowie kameralistische Teilbereiche des Rechnungswesens geben.

Datenerhebung/Registrierung

Erfassung und Prüfung der Daten sind in einem Handbuch geregelt (Bundesministerium für Gesundheit Familie und Jugend 2007), das die Kostenarten für die Kostenartenrechnung, die darauf aufbauende Kostenstellenrechnung (Zuordnung der Kosten(arten) zu einzelnen Kostenstellen) und die zu berücksichtigenden Kostenminderungen (Erlöse, Kostenersatzleistungen) beschreibt.

Zugang zu Daten

Es besteht kein genereller Zugang zu den Daten. In Einzelfällen (z. B. bei Forschungs-kooperationen) können aggregierte Daten projektspezifisch zur Verfügung gestellt werden.

Dateneigenschaften/Validität

Die Daten enthalten Informationen zu Kosten nach Kostenart und Kostenstellen.

15.5.1.7 Impfstatistik

Daten zu den durchgeführten und öffentlich finanzierten Impfungen werden prinzipiell auf Landesebene erfasst, da die Länder aufgrund der Teilung der Finanzierungsverantwortung für die Administration von Impfungen zuständig sind. Aus diesem Grund gibt es in Österreich bisher kein einheitliches System der Dokumentation, sondern eine jährliche Meldung der unterschiedlichen Länderdokumentationen an das Bundesministerium für Gesundheit. Im Rahmen des Projekts „Elektronische Gesundheitsakte“ wird an einer verbesserten Dokumentation gearbeitet. Daten zu den privat finanzierten Impfungen werden auf Landesebene nicht erfasst.

15.5.1.7.1 Impfstatistik zu öffentlich finanzierten Impfungen am Beispiel Kärnten

Originalzweck

Grundlage der Erfassung von Impfdaten ist das Impfkonzert 2006 (eine Vorläufervariante der Datenbank war seit 1998 im Einsatz); es handelt sich um einen Rahmenvertrag zwischen Kärntner Ärztekammer, Apothekerkammer und Land Kärnten, bei dem es um die Umsetzung des jeweils jährlichen Impfplanes des BMG geht. Das System besteht teilweise auch in anderen Bundesländern.

Datenerhebung/Registrierung

Sobald die Sozialversicherungsnummer eines Neugeborenen vergeben ist und weitere notwendige Daten bekannt sind, bekommt die Landessanitätsdirektion (LSD) die Anforderung zur Ausstellung eines Impfscheckheftes, das der Kindesmutter übermittelt wird. In diesem Scheckheft sind alle Impfungen, die im Mutter-Kind-Pass angeführt sind (bis zum 6. Lebensjahr), inkludiert. Die Daten werden von den niedergelassenen Ärzten/Ärztinnen überwiegend in eine Web-Applikation eingetragen, die den Ärzten/Ärztinnen mit registriertem Benutzernamen und Passwort zur Verfügung gestellt wird (teilweise werden noch Impflisten geführt, die von Mitarbeitern/Mitarbeiterinnen an den Bezirksverwaltungsbehörden in die Impfdatenbank eingegeben werden). Die Web-Applikation ist so ausgelegt, dass per Barcodeleser der jeweilige Impfgutschein gescannt und somit eindeutig identifiziert erfasst wird. Manuell eingetragen werden nur die Chargennummer und das Impfdatum, da diese Parameter im Vorhinein nicht bekannt sind. Die niedergelassene Ärzteschaft (Fachärzte/Fachärztinnen, Ärzte für allg. Med.) erhält eine Impfgebühr vom Land Kärnten mit der Auflage, die getätigten Impfungen (in der Impfdatenbank) zu dokumentieren; Amtsärzte/Amtsärztinnen und Jugendfürsorgeärzte/-ärztinnen erhalten ein vermindertes Impfhonorar. Die Abrechnung der Impfkosten erfolgt zentral in der LSD.

Zugang zu Daten

Zugang zum System haben ausschließlich Mitarbeiter/innen der LSD und die impfenden Ärzte/Ärztinnen per Authentifizierung. Datenauswertungen für wissenschaftliche Zwecke waren bisher nicht vorgesehen, sind aber zukünftig nicht ausgeschlossen und datenschutzrechtlich möglich.

Dateneigenschaften/Validität

Derzeit sind ca. 900.000 Datensätze in der Datenbank erfasst (alle Bezirksverwaltungsbehörden + Mag. Villach), und es sind ca. 140 Ärzte/Ärztinnen am Impfkonzept beteiligt. Die Sozialversicherungsnummer dient als eindeutige Identifizierung. Die Administration der Datenbank (d. h. zentrale Anforderung der Impfgutscheinhefte, Neuanlegen von Chargen, Änderungen der Stammdaten wie Wohnsitz, Korrekturen etc.) erfolgt zentral in der LSD. Dort werden auch qualitätssichernde Maßnahmen (Identifizierung doppelter Datensätze, fehlerhafter Stammdaten, Bereinigungen/Datenpflege etc.) vorgenommen.

Ein Problemfeld bez. Validität stellt schon die Frage nach Durchimpfungsraten pro Bezirk dar, denn es gibt bisher keinen Abgleich mit dem zentralen Melderegister, d. h., natürliche Migrationsbewegungen verhindern valide Aussagen über Bezirke und Durchimpfungsraten.

15.5.1.8 Daten der Stellungsuntersuchung

Originalzweck

Die Daten werden im Rahmen der Stellungsuntersuchung zur Prüfung der Wehrtauglichkeit erhoben.

Datenerhebung

Elektronische Dokumentation medizinischer Parameter im Zuge des Stellungsverfahrens.

Zugang zu Daten

Das Verteidigungsministerium übermittelt die Daten an die GÖG. Es existieren keine periodischen Auswertungen oder Publikationen. Sonderauswertungen bei externen

Anfragen sind möglich, aber in der Regel mit einer finanziellen Abgeltung des Zeitaufwandes verbunden.

Dateneigenschaften/Validität:

Es werden Erkrankungen, biometrische Größen (Länge, Gewicht, Blut- und Harnwerte ...), Angaben zum Gesundheitsverhalten (Zigarettenkonsum) und der ausgeübte Beruf dokumentiert. Die Daten werden von allen männlichen Österreichern in dem Jahr erhoben, in dem sie das 18. Lebensjahr erreichen.

15.5.2 Register

Register sind nach einer allgemeinen Definition organisierte Systeme, die beobachtende Studienmethoden für die Sammlung von einheitlichen (klinischen, aber auch anderen) Daten einsetzen, um ein definiertes Ergebnis für eine Population zu evaluieren, die durch eine bestimmte Erkrankung, einen Umstand oder eine Einwirkung gekennzeichnet sind (Gliklich und Dreyer 2007). Im Kontext der Epidemiologie sind Register „the file of data concerning all cases of a particular disease or other health-related condition in a defined population such that the cases can be related to a population base“(Last 2001).

Bei den Arten von Registern zu unterscheiden sind demnach epidemiologische Register (die populationsbezogene Daten zu Prävalenz und Inzidenz von Erkrankungen beinhalten), Qualitätsregister (die Information zur Behandlungsqualität auf individualmedizinischer oder institutioneller Ebene liefern) und risiko-, krankheits- oder interventionsbezogene Register (die Fälle von Individuen mit bestimmten gesundheitlichem Risiko, definierten Erkrankungen oder Fälle von Interventionen registrieren) (Mathis und Wild 2008).

15.5.2.1 Epidemiologische Register

15.5.2.1.1 Epidemiologische Krebsregister

Mit epidemiologischen Krebsregistern wird das Krebsgeschehen beobachtet, also wie häufig bestimmte Tumorerkrankungen in einer Region auftreten. In Österreich existieren sowohl ein nationales, von der Statistik Austria geführtes Krebsregister als auch regionale Register in Salzburg, Tirol, Kärnten und Vorarlberg. Die Register unterscheiden sich primär durch den Grad an Vollständigkeit (z. B. % an „Death certificate only“-Einträgen, Daten ausschließlich aus dem stationären oder auch aus dem niedergelassenen Bereich) und durch die Anzahl der berechneten Indikatoren (Standardindikatoren

sind Inzidenz, Mortalität und Prävalenz; Überlebensraten werden nur teilweise berechnet). Unten stehend werden das Tumorregister Österreich und – als Beispiel für ein regionales Register – das Tumorregister Tirol beschrieben.

15.5.2.1.1.1 Tumorregister Österreich (Statistik Austria 2010)

Originalzweck

Das österreichische Krebsregister wird von der Statistik Austria auf Basis des Krebsstatistikgesetzes (1969) und der Krebsstatistikverordnung (1978) geführt. Es dient der Dokumentation aller Krebsfälle der Wohnbevölkerung, der Auswertung dieser Daten und dem Monitoring von Umfang, Struktur und Folgen der Krebslast.

Datenerhebung/Registrierung

Jede Krebserkrankung und jeder Sterbefall aufgrund einer solchen sind von den Krankenanstalten an die Statistik Austria zu melden (gesetzliche Verpflichtung). In den Bundesländern Kärnten, Salzburg, Tirol und Vorarlberg, wo eigene regionale Krebsregister bestehen, übernehmen diese als Dienstleister für die Krankenanstalten die Meldung.

Zugang zu Daten

Die aggregierten Ergebnisse werden auf der Internetseite der Statistik Austria (http://www.statistik.at/web_de/statistiken/gesundheit/krebserkrankungen/index.html) sowie im Jahrbuch der Gesundheitsstatistik veröffentlicht. Diese sind kostenfrei zugänglich (http://www.statistik.at/web_de/services/publikationen/4/index.html). Eine kostenpflichtige CD-ROM enthält tiefer gegliederte Daten im Excel-Format. Es besteht außerdem die Möglichkeit eines Online-Zugriffs auf eine Datenbank (nach Anmeldung), wobei Abfragen zum Teil kostenpflichtig sind. Darüber hinausgehende Datenbedarfe können gegen Gebühr erstellt werden.

Dateneigenschaften/Validität

Es werden Daten zur Krebslokalisierung und Histologie (ICD-O-3), zum Tumorstadium und zu Geschlecht, Alter und Wohnort erhoben. Daraus werden die Indikatoren Inzidenz, Mortalität, Überleben und Prävalenz, nach verschiedenen Merkmalen gegliedert, berechnet. Da Meldepflicht nur für Krankenanstalten besteht, werden die Daten zu Krebserkrankungen, die ausschließlich im niedergelassenen Bereich diagnostiziert und behandelt werden, zwar erfasst, wegen potenzieller Untererfassung aber nicht publiziert (z. B. bösartige Neubildungen der Haut, ausgenommen bösartige Melanome).

Zur Beurteilung der Datenqualität werden Empfehlungen der Internationalen Vereinigung der Krebsregister (IACR) herangezogen. Ein zentrales Qualitätskriterium ist, eine Rate an „Death certificate only“-Meldungen unter zehn Prozent zu erreichen, was seit 1997 gegeben ist. Die Qualität der Daten ist seit der Veröffentlichung in der Publikation *Cancer Incidence in Five Continents (Vol. IX)* von der IARC anerkannt. Das Register verfolgt das Konzept einer lebenden Datenbank und zieht für jede Veröffentlichung die jeweils aktuellen Gesamtdaten (für alle verfügbaren Berichtsjahre) heran. Bei der Interpretation der Daten und Verwendung der Daten für wissenschaftliche Zwecke ist die für die Altersstandardisierung verwendete Standardbevölkerung zu beachten (z. B. Welt-Standardbevölkerung der WHO, SEGI-Weltbevölkerung oder Europäische Standardbevölkerung).

15.5.2.1.1.2 Regionale Register am Beispiel Tumorregister Tirol (Institut für klinische Epidemiologie der TILAK 2008)

Originalzweck

Das Tumorregister Tirol wurde 1986 zum Zweck der Dokumentation aller Krebsfälle in Tirol, zur Berechnung von Überlebenszeiten und Überlebensraten, zur Berichterstattung über Krebsinzidenz und -mortalität in Tirol und zum Zweck der Durchführung epidemiologischer Studien gegründet.

Datenerhebung/Registrierung

Für die Basisdokumentation erhält das Register von den Tiroler Krankenhäusern und Sanatorien Daten über neu auftretende Tumorfälle (von einigen Abteilungen auf elektronischem Weg, von ca. der Hälfte der Abteilungen auf Papierfragebogen; bei den restlichen Abteilungen wird von Dokumentaren/Dokumentarinnen des Tumorregisters direkt aus den Krankengeschichten dokumentiert). Tumorfälle, die ausschließlich von Praktikern behandelt werden, werden über Pathologiebefunde der Pathologieinstitute registriert, die das Tumorregister erhält und die über ein Record-Linkage-Verfahren mit den Registerdaten abgeglichen werden. In einigen Kliniken wird zudem ein Spezialregister mit detaillierten Dokumentationen geführt.

Zur Überprüfung der Vollständigkeit werden Daten aus der Datenbank mit Krankenhausabgangsdiagnosen, Pathologiedaten und Abteilungsdaten abgeglichen und gegebenenfalls fehlende Daten nachrecherchiert. Für Tumortodesfälle, für die keine Meldung eines Krankenhauses vorliegt, wird in Krankenhäusern und bei niedergelassenen Ärzten/Ärztinnen recherchiert. Falls keine Information gefunden wird, wird der Tumorfall als „Death certificate only“-Fall ausgewertet. Todesdaten werden von der Statistik Austria (codiert und ausgewertet) zur Verfügung gestellt.

Zugang zu Daten

Es werden jährliche Berichte zu Standardauswertungen vorgelegt, die öffentlich über die IET-Website zugänglich sind. Per Anfrage können Datenauswertungen, die die Datenschutzbestimmungen erfüllen, durchgeführt werden. Anfragen können sowohl von Ärzten/Ärztinnen als auch von anderen Personen gestellt werden (z. B. Verdacht auf erhöhte Inzidenzrate in einer Gemeinde) und werden beantwortet, ohne Kosten zu verrechnen. Ergebnisse für einzelne Abteilungen werden nur mit schriftlicher Zustimmung des jeweiligen Abteilungsvorstandes weitergegeben.

Dateneigenschaften/Validität

Die Basisdokumentation erfolgt auf Basis der Vorgaben des österreichischen Krebsstatistikgesetzes (Krebsmeldeblatt). Für Melanome, Tumoren des Zentral- und peripheren Nervensystems und für Struma maligna erfolgt eine erweiterte Spezialdokumentation.

Es werden alle bösartigen Neubildungen und Carcinoma in situ erfasst. Jeder Tumor wird nach ICD-O V-3 codiert (Lokalisation und Histologie des Tumors). Das Tumorstadium wird nach Möglichkeit nach der TNM-Klassifikation dokumentiert, für gynäkologische Tumoren wird wahlweise das TNM- oder FIGO-Stadium verwendet, für hämatologische Systemerkrankungen werden die jeweiligen Systeme (Durie-Salmon, Rai etc.) verwendet. Der Wohnort der Patienten/Patientinnen wird nach dem Gemeindeschlüssel der Statistik Austria codiert. Dadurch ist eine getrennte Auswertung nach Gemeinden und Bezirken möglich.

Die Qualitätsprüfung der Daten erfolgt mit selbst entwickelten Überprüfungsprogrammen und Programmen der IARC.

Die Daten dienen zur Berechnung von Standardmaßzahlen (z. B. DCO-Prozentsatz, Verhältnis Mortalität zu Inzidenz, standardisierte Mortalitäts- bzw. Inzidenzrate, relative Überlebensraten). Die verwendeten Maßzahlen, die dazugehörigen Formeln und die Interpretation sind in der Methodenbeschreibung dokumentiert (Institut für klinische Epidemiologie der TILAK 2008).

15.5.2.1.2 Geburtenregister Österreich (Institut für klinische Epidemiologie der TILAK 2011, Oberaigner und Leitner 2008)

Originalzweck

Das Hauptziel des Geburtenregisters ist es, einen Beitrag zur Senkung der perinatalen Mortalität und Morbidität zu leisten. Daneben soll das Register zur Transparenz über

die geburtshilflich-perinatologische Qualität, zur Entscheidungsunterstützung, für Forschungszwecke und zur Gewinnung basisepidemiologischer Daten dienen.

Datenerhebung/Registrierung

Die Daten werden in den geburtshilflichen Abteilungen mit unterschiedlicher Software (z. B. GebDat) erhoben und pro Quartal vom IET bei den Abteilungen eingeholt, in einen gemeinsamen Datensatz eingespielt und auf Vollständigkeit, fehlerhafte Werte, Plausibilität und perinatale Mortalität (Vergleich mit Statistik Austria) überprüft. Hausgeburten sind nicht Teil des Geburtenregisters und werden seit 2006 vom österreichischen Hebammengremium dokumentiert. Seit 2005 nehmen Abteilungen aller Bundesländer teil und seit 2011 alle Krankenhäuser mit einer geburtshilflichen Abteilung (inkl. Privatkrankenhäuser). Die Datenerhebung orientiert sich an der deutschen Perinatalerhebung, auf Basis derer ein verbindlicher Datensatz festgelegt wurde.

Zugang zu Daten

Aggregierte Daten für alle österreichischen Abteilungen werden über die regelmäßige Auswertung des Geburtenregister-Betreibers zugänglich gemacht (Kernauswertung vierteljährlich, Qualitätsindikatoren jährlich). Abteilungsspezifische Ergebnisse werden nur an die jeweilige Abteilung weitergegeben. Regelmäßig erscheinende öffentlich zugängliche Jahresberichte werden in Tirol seit 1999 erstellt, für die Steiermark bzw. KAGES (KAGES 2008) seit 2004 und für Oberösterreich (GESPAG) seit 2008. Spezielle Auswertungen können auf Anfrage per Formular nach Genehmigung durch den Fachbeirat des Geburtenregisters Österreich nach freien Ressourcen durchgeführt werden. Bisher wurden keine Kosten verrechnet.

Dateneigenschaften/Validität

Die Daten entstammen einem perinatologischen Datensatz, der österreichweit zum Einsatz kommt (auf die Struktur des Datensatzes kann zugegriffen werden). Der Datensatz des Geburtenregisters enthält Daten zum Ort der Geburt, zu den Charakteristika der Mutter (Alter, Berufstätigkeit, Raucherstatus), zur Schwangerschaft (Teilnahme an Vorsorgeuntersuchungen, stationäre Aufenthalte, Risikoschwangerschaften, medizinische Maßnahmen), zur Entbindung (Anzahl, Ablauf, Art, medizinische Interventionen, Komplikationen), zum Kind (Anzahl, Gesundheitszustand, Mortalität), zum Wochenbett (Komplikationen) und zum Entlassungszeitpunkt. Aus den Daten kann eine Reihe von Indikatoren berechnet werden. Ein Beispiel für die Bandbreite an Auswertungen ist bei Leitner (2009) angeführt. 2010 wurde eine strukturierte Qualitätsverbesserung vereinbart. Die Daten werden sowohl auf den Abteilungen durch die Abteilungssoftware ausführlich auf Plausibilität getestet als auch im Geburtenregister Österreich.

Die Vollzähligkeit der Meldungen über Geburten wird mit jedem Haus über die Geburtenbücher überprüft, die Vollständigkeit der Meldungen von kindlichen Todesfällen wird mit Daten der Statistik Austria geprüft.

15.5.2.1.3 Register der anzeigepflichtigen Erkrankungen / epidemiologisches Meldesystem (EMS) (Bundeskanzleramt 2008)

Originalzweck

Das Meldesystem wurde im Rahmen von Bundesgesetzen zur Überwachung und Kontrolle von Infektionskrankheiten und aufgrund von EU-Regelungen geschaffen und soll in drei Ausbaustufen (Beginn 2008) etabliert werden.

Datenerhebung/Registrierung

Im Vollausbau soll ein elektronischer Zusammenschluss der Gesundheitsbehörden auf allen administrativen Ebenen bestehen (Bundesministerium, Bezirksverwaltungsbehörden, Landessanitätsdirektionen). In der derzeitigen Ausbaustufe erfolgen die Meldungen der Amtsärzte/Amtsärztinnen und teilweise der Labors elektronisch im EMS; Meldungen aller weiteren meldepflichtigen Personen und der restlichen Labors erfolgen derzeit noch mit dem herkömmlichen Meldesystem (Erstmeldeformular, Erregerfeststellung) an die zuständige Bezirksverwaltungsbehörde, die die elektronische Eintragung im EMS übernimmt. Eine Ausnahme sind die HIV-/Aids-Fälle, die direkt an das Bundesministerium für Gesundheit gemeldet werden. In den weiteren Ausbaustufen erfolgen schrittweise alle Meldungen direkt im EMS.

Zugang zu Daten

Der Zugang ist nur definierten Personen mit Authentifizierung möglich. Aggregierte monatliche Auswertungen zu den Krankheitshäufigkeiten und Todesfällen pro Bundesland werden online veröffentlicht. Weitere öffentliche Datenauswertungsmöglichkeiten sind in Planung (in Anlehnung an das SurvNet am Robert Koch Institut in Deutschland), aber derzeit noch nicht möglich. Die Möglichkeit der Auswertungen individueller Anfragen wird fallbezogen entschieden.

Dateneigenschaften/Validität

Die Daten umfassen Informationen zu allen in Österreich meldepflichtigen Erkrankungen gemäß Epidemie-, Tuberkulose-, AIDS- und Geschlechtskrankheitengesetz. Für Krankheiten, die in der EU und in Österreich meldepflichtig sind, erfolgt die Dokumentation gemäß EU-Falldefinition, für solche, die nur in Österreich meldepflichtig sind,

wurden – sofern vorhanden – die Falldefinitionen des Robert Koch Instituts angewendet (Bundesministerium für Gesundheit Familie und Jugend 2008).

Zur Überprüfung der Datenqualität erfolgt eine Kennzeichnung unvollständiger oder ungenauer Informationen beim entsprechenden Dateneintrag und eine Recherche der fehlenden Information. Problematisch hinsichtlich Validität ist die mangelnde Meldetätigkeit (underreporting) und die mangelnde Vollständigkeit bei den Variablen pro Fall.

15.5.2.2 Qualitätsregister

15.5.2.2.1 Diabetesregister Tirol (Institut für klinische Epidemiologie der TILAK 2011)

Originalzweck

Das Diabetesregister wurde 2006 zum Zweck der Qualitätsverbesserung eingerichtet. In der ersten Phase wurde der Großteil der Diabetesambulanzen der Tiroler Krankenhäuser eingebunden, in der zweiten Phase (seit 2008) erfolgt eine Erweiterung auf andere Stationen der Krankenhäuser. Ein weiteres Ziel ist das Gewinnen von epidemiologischen Fakten über Diabetes-Patienten und -Patientinnen in Tirol wie Inzidenz und Prävalenz von Diabetes mellitus und Häufigkeit von diabetischen Spätkomplikationen. Die erhobenen Informationen sollen schlussendlich auch eine Abschätzung der zu erwartenden Kosten in der Behandlung von Diabetes mellitus und der Spätkomplikationen ermöglichen.

Datenerhebung/Registrierung

Die Daten werden elektronisch von den teilnehmenden Institutionen dokumentiert und nach Pseudonymisierung elektronisch an das IET übermittelt.

Zugang zu Daten

Es werden jährlich öffentlich zugängliche Berichte über aggregierte Daten vorgelegt. Anfragen können sowohl von Ärzten/Ärztinnen als auch von anderen Personen/Institutionen gestellt werden und werden bisher beantwortet, ohne Kosten zu verrechnen.

Dateneigenschaften/Validität

Es werden Daten zum Migrationshintergrund, anamnestische Daten, Daten zu Spät-komplikationen, Laborwerte (HbA1c), Blutdruckwerte, Daten zu Gewicht, Bauchumfang und körperlichen Aktivitäten und zur Insulintherapie erhoben (Fragebogen ist über die Website des IET zugreifbar). Die Abteilungen erhalten pro Quartal eine Übersicht über die Daten und können die Ausprägungen der eigenen Abteilung mit den anderen Abteilungen vergleichen. Hauptqualitätsparameter ist der Laborparameter „HbA1c“. Weiters werden Anteile über durchgeführte Untersuchungen (z. B. Fußinspektion), über Begleiterkrankungen (z. B. Bluthochdruck) sowie über Spätkomplikationen ausgewertet.

15.5.2.2.2 Erwachsenen–Herzchirurgie–Register

Originalzweck

Das Register – ein gemeinsames Projekt des Bundesinstituts für Qualität im Gesundheitswesen (BIQG) und der Gesellschaft für Thorax- und Herzchirurgie – wurde eingerichtet, um eine vergleichende Auswertung der Behandlungsergebnisse der neun herzchirurgischen Zentren in Österreich zu ermöglichen.

Datenerhebung/Registrierung

Die Daten werden dezentral in den einzelnen Abteilungen erfasst und nach einer Überprüfung an das Qualitätsinstitut übermittelt. Es nehmen alle neun herzchirurgischen Zentren freiwillig an dem Register teil und übermitteln einmal jährlich die Daten aus dem vergangenen Jahr.

Zugang zu Daten

Die teilnehmenden Zentren können selbstständig die eigenen Daten auswerten. Einmal jährlich wird ein nicht öffentlicher Ergebnisbericht mit allen österreichischen Gesamtdaten erstellt. Für jedes Zentrum wird zusätzlich ein eigener Bericht erstellt, in dem das eigene Zentrum sichtbar ist und die anderen Zentren pseudonymisiert dargestellt werden. Aggregierte österreichweite Gesamtdaten können gegebenenfalls auf Anfrage gegen Kostenersatz weitergegeben werden, wenn sie den vertraglichen Regelungen im Hinblick auf die Datenverwendung entsprechen.

Dateneigenschaften/Validität

Zur Dokumentation wird das international anerkannte Indikatorenset der Society of Thoracic Surgeons (STS) angewandt. Basis der Dokumentation bilden alle herzchirurgischen Operationen, die in einem der neun österreichischen herzchirurgischen Zentren bei einer erwachsenen Patientin / einem erwachsenen Patienten durchgeführt wurden. Es besteht daher die Möglichkeit, dass eine Patientin / ein Patient, bei der/dem mehrere herzchirurgische Eingriffe stattgefunden haben, mehrmals in das Register aufgenommen wurde. Durch die Definition von Pflichtfeldern können unvollständige Datensätze vermieden werden.

15.5.2.2.3 Herzschrittmacher- und Loop-Recorder-Register

Originalzweck

Das Herzschrittmacher-Register wurde von 1980 bis 2006 von der Österreichischen Gesellschaft für Kardiologie betrieben. Seit 2007 ist die GÖG/BIQG beauftragt, das Herzschrittmacher-Register zu führen. Eine Internetapplikation wurde 2011 entwickelt, damit die Datenübertragung an das nun vom BIQG geführte Register online und plausibilitätsgeprüft erfolgen kann. Das Register wird ab 2012 um implantierbare Defibrillatoren (ICD) und Loop-Recorder erweitert. Das Register wird zum Zweck der Medizinprodukte-Vigilanz und der Information über die Produktperformance/Lebensdauer und für einen Marktüberblick betrieben.

Datenerhebung/Registrierung

Zukünftig erfolgt die Datenübermittlung online über das Internet mittels eines Registers. Krankenhäuser mit eigenen Datenerfassungssystemen für die jeweiligen Datensätze können die Daten dezentral erfassen, speichern und periodisch mithilfe des Webservice an die GÖG/BIQG übermitteln.

Zugang zu Daten

Die Daten werden einmal jährlich ausgewertet und dem Bundesministerium für Gesundheit für statistische Zwecke zur Verfügung gestellt.

Dateneigenschaften/Validität

Es werden Stammdaten, Schrittmacherdaten und Sondendaten erhoben.

15.5.2.2.4 Hüftendoprothesenregister (derzeit Pilotbetrieb mit ausgewählten Abteilungen)

Originalzweck

Das Register für Hüftendoprothetik wird mit den Fachgesellschaften für Orthopädie und orthopädische Chirurgie und für Unfallchirurgie betrieben. Es wurde zum Zweck der Medizinprodukte-Vigilanz (Frühwarnsystem für fehlerhafte Produkte) und zur Datengewinnung über weitere Aspekte der Behandlungsqualität (Behandlungserfolge, Re-Implantationen ...) eingerichtet.

Datenerhebung/Registrierung

Die Daten werden von den teilnehmenden Krankenanstalten direkt in das Register eingegeben. Zur Erfassung der Implantatdaten werden Kabelscanner verwendet, mit deren Hilfe die Barcodes der Artikel erfasst und die Datenfelder der Interneteingabemaske automatisch mit den hinter den Artikelnummern stehenden Produktinformationen gespeist werden. Um laufend die aktuellen Produkteigenschaften erfassen zu können, ist eine enge Kooperation mit den Herstellern von Implantaten notwendig.

Zugang zu Daten

Die teilnehmenden Abteilungen können tagesaktuell alle eingegebenen Daten downloaden, um sie für interne Zwecke zu verwenden. Aggregierte Auswertungen werden jährlich durch die GÖG/BIQG erstellt und an alle teilnehmenden Abteilungen versandt. Alle Daten stehen ausschließlich für die interne Qualitätsarbeit zur Verfügung und dürfen nicht veröffentlicht werden.

Dateneigenschaften/Validität

Der Datensatz umfasst zwei Teile, einen Teil für die Erstoperation und einen für die Revisionsoperation(en). In beiden Teilen werden Daten zum Zeitpunkt der Aufnahme (verpflichtende Eingabe), zur Operation (verpflichtende Eingabe), zu den verwendeten Implantaten (verpflichtende Eingabe) und zum Zeitpunkt der Entlassung (freiwillige Eingabe) erhoben. Im Falle einer Revisionsoperation werden zusätzlich noch die Gründe für die Revisionsoperation erfasst, wobei differenziert wird zwischen einer Hauptindikation zur Revision bzw. weiteren Befunden (Mehrfachnennungen möglich).

15.5.2.2.5 Prothesenregister Tirol (Institut für klinische Epidemiologie der TILAK 2011, Oberaigner und Labek 2007)

Originalzweck

Das Prothesenregister Tirol (Gründung 2003) hat als Hauptziel die Verbesserung der Qualität der Behandlung im Bereich der Endoprothesen. Sowohl die Ergebnisqualität der behandelnden Abteilungen als auch die Qualität der Implantate (im Rahmen internationaler Kooperationen) sollen verbessert werden.

Datenerhebung/Registrierung

Die Daten werden einerseits von Patienten/Patientinnen (präoperative WOMAC-Fragebogen, postoperative WOMAC-Fragebögen) erhoben und andererseits von operierenden Ärzten/Ärztinnen dokumentiert. Der präoperative WOMAC-Fragebogen wird von den Stationen ausgegeben, wieder eingesammelt und an das Institut für klinische Epidemiologie der TILAK (IET) gesendet. Die Operationsdaten werden von den Ärzten dokumentiert, entweder über eine Eingabemaske oder über einen Fragebogen. Der Versand der postoperativen WOMAC-Fragebögen wird vom IET durchgeführt, die Fragebögen werden von den Patienten/Patientinnen wiederum direkt an das IET zurückgeschickt. Die Daten werden in einer Oracle-Datenbank gespeichert.

Zugang zu Daten

Seit 2005 wurden jährliche Berichte erstellt (mit einer Unterbrechung im Jahr 2009), die öffentlich zugänglich sind. Anfragen können sowohl von Ärzten/Ärztinnen als auch von anderen Personen gestellt werden (z. B. Verdacht auf erhöhte Inzidenzrate in einer Region oder Verdacht auf erhöhte Revisionsrate in einem Krankenhaus) und werden bei gegebenen Ressourcen beantwortet, ohne Kosten zu verrechnen. Alle teilnehmenden Abteilungen werden über ihre Qualitätsparameter informiert und erhalten Informationen, mit denen sie die Qualität der eigenen Abteilung mit allen anderen Abteilungen in einer Gruppe zusammengefasst vergleichen können. Es werden keine abteilungsspezifischen Ergebnisse weitergegeben, außer es liegt ein schriftlicher Auftrag des jeweiligen Klinikvorstands/Abteilungsleiters oder der Abteilungsleiterin vor.

Dateneigenschaften/Validität

Der WOMAC-Fragebogen ist ein international validierter Fragebogen, mit dem Daten zu den Themen Schmerzen, Steifigkeit, Schwierigkeiten und - in der postoperativen Version - zusätzlich zu Thrombose, Nervenlähmung und Re-Operation erhoben

werden. Die Einstufungen der Patienten/Patientinnen erfolgen auf einer Skala von 1 bis 11.

Der Umfang der dokumentierten Daten über die Implantation orientiert sich an den internationalen Empfehlungen und wird vom Fachbeirat beschlossen. Ab dem Jahr 2011 wurden auch die Vorgaben des Hüftprothesenregisters der GÖG eingehalten.

Es erfolgen Auswertungen zu Patientencharakteristika, zur Differenz zwischen prä- und postoperativer Einschätzung der Patienten/Patientinnen (statische Signifikanz und klinische Relevanz), zu den Erstimplantationen und zu den Revisionen (Revisionsrate, Revisionslast). Bezüglich Qualität der Daten ist eine hohe Vollständigkeit (mind. 90 %) und hohe Qualität der Informationen gegeben (Oberaigner und Labek 2007).

15.5.2.2.6 Kinderkardiologieregister (noch nicht im Regelbetrieb)

Originalzweck

Das Register wurde mit Vertretern der vier kinder-kardiologischen Zentren in Österreich aufgebaut und erfasst die Ergebnisqualität der Herzkatheter-Untersuchungen und der Interventionen bei angeborenen sowie bei erworbenen Herzfehlern im Kindes- und Jugendalter (0-18 Jahre). Die nachfolgenden Angaben beziehen sich auf den geplanten Regelbetrieb:

Datenerhebung/Registrierung

Im Register werden alle Herzkatheteruntersuchungen und -interventionen bei angeborenen Herzfehlern sowie erworbenen Herzfehlern im Kindes- und Jugendalter (0-18 Jahre) erfasst. Die Eingabe erfolgt kontinuierlich.

Zugang zu Daten

Die vier kinder-kardiologischen Zentren in Österreich können tagesaktuell alle eingegebenen Daten beziehen, um sie für interne Zwecke zu verwenden. Auswertungen sind nur für das hausinterne Qualitätsmanagement bestimmt und nicht öffentlich zugänglich. Jährlich wird ein Ergebnisbericht durch die GÖG/BIQG mit den österreichweiten Gesamtdaten erstellt.

Dateneigenschaften/Validität

Der Datensatz gliedert sich nach Patientendaten, Untersuchungsdaten, Katheteruntersuchung/Intervention, Transfer, Ergebnis und Nachbeobachtung. Die Daten unterliegen

bei der Eingabe einer Plausibilitätsprüfung. Durch die Definition von Pflichtfeldern können unvollständige Datensätze vermieden werden.

15.5.2.2.7 Chirurgie-Qualitätsregister

Originalzweck

Das Register wird in Zusammenarbeit mit der Österreichischen Gesellschaft für Chirurgie und dem Berufsverband Österreichischer Chirurgen geführt. Das Projekt wird auch von der Arbeitsgemeinschaft Chirurgische Onkologie unterstützt. Ziel des Registers ist, mit minimalem Dokumentations- und Zeitaufwand die Qualitätssicherung in der Allgemeinchirurgie zu unterstützen.

Datenerhebung/Registrierung

Von den teilnehmenden Abteilungen sind alle ungeplanten Re-Eingriffe bis zum 14. postoperativen Tag nach allgemeinchirurgischen Eingriffen ins Register einzutragen. Neben patientenbezogenen Parametern werden insbesondere Daten zum Ersteingriff sowie zum Re-Eingriff erfasst. Die Dateneingabe erfolgt kontinuierlich. Jährlich im Nachhinein erfolgt eine Auswertung in Hinblick auf Hauptkomplikationen, Re-Operations- und Mortalitätsraten. Zur Berechnung dieser Raten erfolgt ein Abgleich mit der Diagnosen- und Leistungsdokumentation (DLD).

Zugang zu Daten

Die teilnehmenden Abteilungen können tagesaktuell alle eingegebenen Daten downloaden, um sie für interne Zwecke (z. B. im Rahmen von Fehler- oder Komplikationskonferenzen) zu verwenden. Auswertungen werden jährlich durch die GÖG/BIQG erstellt und an alle teilnehmenden Abteilungen versandt. Alle Daten stehen ausschließlich für die interne Qualitätsarbeit zur Verfügung und dürfen nicht veröffentlicht werden.

Dateneigenschaften/Validität

Der Datensatz ist, um den Dokumentationsaufwand möglichst gering zu halten, relativ kurz gehalten. Er umfasst Patientendaten, wesentliche Informationen zur Primär-Operation sowie die Erfassung der Hauptkomplikation, die zur Re-Operation geführt hat, und das Datum dieser Re-Operation. Im Rahmen der Berechnung der Re-Operations- und Mortalitätsraten erfolgt ein Abgleich mit den in der DLD abgerechneten Fällen. Auf diese Weise ist auch die Beurteilung der Vollständigkeit der eingegebenen Fälle möglich. Auswertungen erfolgen für verschiedene Leistungsbereiche.

15.5.2.2.8 Stroke-Unit-Register

Originalzweck

Das österreichische Stroke-Unit-Register wurde im Jahr 2003 von der Gesundheit Österreich GmbH (GÖG) und der Österreichischen Gesellschaft für Schlaganfallforschung (ÖGSF) ins Leben gerufen. Mit dem Register werden unter anderem folgende Ziele verfolgt: Qualitätssicherung und Evaluierung im Behandlungsprozess, Evaluierung des Zustransportmanagements und des Outcomes, anonymisiertes Benchmarking mit Feedback-Schleifen und Einleitung von Rückkoppelungsprozessen und Qualitätsverbesserungsmaßnahmen, standardisierte statistische Auswertungen, Schaffung einer umfassenden Datenbasis für wissenschaftliche Auswertungen und Etablierung einer Service-Plattform für teilnehmende Zentren.

Datenerhebung/Registrierung

Insgesamt nehmen derzeit 35 Stroke-Units am Register teil. Die Daten werden laufend von den teilnehmenden Stroke-Units über eine internetbasierte Eingabemaske in das Register eingetragen.

Zugang zu Daten

Die teilnehmenden Stroke-Units können tagesaktuell auf die eigenen Daten sowie auf vordefinierte Auswertungen (inkl. Benchmark-Auswertungen) zugreifen. Zudem können die eigenen Daten jederzeit mittels Download bezogen werden und weiter analysiert werden. Zumindest einmal im Jahr wird zudem eine Auswertung durch die GÖG/BIQG an alle Stroke-Units versandt. Seit 2009 wird zudem einmal im Jahr eine Benchmarking-Veranstaltung durch die GÖG/BIQG organisiert, die den Teilnehmern die Gelegenheit zum Austausch („Lernen vom Besten“) bietet. Interessierte Teilnehmer haben die Möglichkeit, vertiefende (wissenschaftliche) Auswertungen aus den Registerdaten anzufragen. Einige österreichweite Ergebnisse werden möglicherweise zukünftig veröffentlicht.

Dateneigenschaften/Validität

Der Datensatz ist im Vergleich zu den anderen von der GÖG/BIQG geführten Registern relativ umfangreich. Er gliedert sich in drei große Teile: Informationen zur Aufnahme (inkl. Patientendaten), Aufenthalts-/Entlassungsdaten (inkl. Therapie) und Drei-Monats-Follow-up. Plausibilitätsprüfungen finden direkt im Rahmen der Dateneingabe statt. Durch die Definition von Pflichtfeldern können unvollständige Datensätze vermieden werden.

15.5.2.2.9 Implantatregister für Silikonimplantate

Originalzweck

Das an der Medizinischen Universität Wien angesiedelte Register ist ausschließlich für Ärzte/Ärztinnen, die Silikonimplantate verwenden, zum Zweck der Qualitätssicherung vorgesehen.

Datenerhebung/Registrierung

Die Datenerhebung erfolgt mittels Implantatpass.

Zugang zu Daten

Der Zugang ist auf Teilnehmer/innen des Registers beschränkt.

Dateneigenschaften/Validität

Zu den Dateneigenschaften liegen keine Informationen vor.

15.5.2.2.10 Pharmakovigilanz und Qualitätsmangelmeldung Arzneimittel

Originalzweck

Die Meldung von Nebenwirkungen und Qualitätsmängeln ist im Arzneimittelgesetz und in der Pharmakovigilanzverordnung gesetzlich verpflichtend zum Zweck der Qualitätssicherung vorgesehen.

Datenerhebung/Registrierung

Die Meldung erfolgt verpflichtend von Angehörigen diverser Gesundheitsberufe (Ärzte/Ärztinnen, Hebammen), von Apothekern/Apothekerinnen, Gewerbetreibenden und Zulassungsinhabern per bei der AGES (online) erhältlichem Formular elektronisch oder per Post an die zuständige Behörde (AGES PharmMed). Eine nationale Datenbank, in der die Nebenwirkungsmeldungen für weitere Nutzung aufbereitet werden, befindet sich derzeit im Aufbau. Die Daten werden außerdem an die Europäische Arzneimittelbehörde (EMA) weitergeleitet, wo ebenfalls eine Datenbank für die weitere Nutzung der Daten in Aufbau ist. Meldungen zu Qualitätsmängeln werden fallbezogen auf der Webseite der AGES PharmMed veröffentlicht, sie sind jedoch nicht retrospektiv einsehbar.

Zugang zu Daten

Die Datenbank zu gemeldeten Nebenwirkungen soll ab 2012 öffentlich zugänglich sein. Die Qualitätsmängelmeldungen sind auf der Website in chronologischer Reihenfolge ersichtlich.

Dateneigenschaften/Validität

Die Nebenwirkungsmeldung umfasst personenbezogene Daten zu den betroffenen Patienten/Patientinnen und deren Erkrankung(en), zu Art und Zeitpunkt der aufgetretenen Nebenwirkung, deren Behandlung und Verlauf und zum verwendeten Arzneimittel. Welche Daten davon in welcher Form in der vorgesehenen Datenbank aufbereitet werden, ist zum jetzigen Zeitpunkt noch nicht bekannt.

Trotz gesetzlicher Verpflichtung zur Meldung ist von einem „underreporting“ auszugehen. Da wichtige Informationen fehlen (z. B. zur Frage, wie häufig das Arzneimittel insgesamt verschrieben wurde), ist die Interpretation (zur Häufigkeit auftretender Nebenwirkungen) eingeschränkt. Seltene Nebenwirkungen sind auf internationaler Ebene aussagekräftiger erfasst als in nationalen Datenbanken.

15.5.2.3 Risiko-, krankheits- oder interventionsbezogene Register

15.5.2.3.1 Klinische Krebsregister

Klinische Krebsregister sammeln Daten zur Erkrankung und Therapie bestimmter Krebsarten, um die Behandlung zu verbessern.

Ein Beispiel für klinische Krebsregister sind die Register der Arbeitsgemeinschaft für medikamentöse Tumorthherapie (Arbeitsgemeinschaft Medikamentöse Tumorthherapie (AGMT) 2011) (Kopf-Halstumor Register, Registry and Biobank for the collection of clinical data and biomaterials from adult ALL patients, Register zu „Surgical Resection of Solitary Metastases in Patients with Metastatic Breast Cancer“, Register zu „Screening for HER2-positive patients with inoperable locally advanced or metastatic gastric or gastro-esophageal junction cancer“, Register zu „Biomarkers in Metastatic Colorectal Cancer“, Register zu „VIDAZ® in patients with MDS, CMML or AML“).

Es handelt sich um eine freiwillige prospektive und/oder retrospektive elektronische Datenerhebung von österreichischen Kliniken zur Untersuchung von Tumorfrequenzen, durchgeführten Behandlungen, Outcomes (Lebensqualität, Überleben) und Nebenwirkungen.

Ein öffentlicher Zugang zu den Daten und den Datenerhebungsblättern besteht nicht. Anfragen können an die jeweiligen Verantwortlichen der Register gestellt werden. Es ist geplant, dass Ergebnisse in wissenschaftlichen Zeitschriften und in Form von Berichten veröffentlicht werden.

15.5.2.3.2 Österreichisches Register für Medizinprodukte

Originalzweck

Gemäß Medizinproduktegesetz und den europäischen Richtlinien für Medizinprodukte und In-vitro-Diagnostika ist in Österreich ein Register für Medizinprodukte zu führen. Seit 2002 müssen all jene Personen bzw. Firmen mit Sitz in Österreich registriert werden, die für das erstmalige Inverkehrbringen von Medizinprodukten im europäischen Wirtschaftsraum verantwortlich sind oder die eine Prüf- und Überwachungsstelle mit Sitz in Österreich betreiben. Händler und Vertreiber von Medizinprodukten und In-vitro-Diagnostika können sich seit 2005 freiwillig registrieren.

Datenerhebung/Datenregistrierung

Die Daten werden internetbasiert nach Zuteilung eines Benutzernamens und Kennwortes an der GÖG registriert.

Zugang zu Daten

Das Register ist nicht öffentlich zugänglich. Anfragen sind an das BMG (Dateneigner) zu richten.

Dateneigenschaften/Validität

Die Daten beinhalten die Medizinprodukte-Kategorie nach international anerkannten Codes (GMDN-/UMDNS-/EDMS-Bezeichnung), Bezeichnung des Gewerbes, gewerbescheinausstellende Behörde, Ausstellungsdatum und Geschäftszahl des Gewerbescheins, den Handels- bzw. Produktnamen oder den generischen Namen des Medizinprodukts. Es erfolgt keine spezifische Validitätsprüfung.

15.5.2.3.3 IVF-Register

Originalzweck

Das IVF-Register wurde auf der Grundlage des IVF-Fonds-Gesetzes 1999 (Bundeskanzleramt 2010) mit dem Zweck eingeführt, abrechnungsrelevante Informationen für den IVF-Fonds bereitstellen und Analysen durchführen zu können. Es ist die Aufgabe des Registers, für jede Vertragskrankenanstalt die Anzahl der Schwangerschaftsversuche und die dabei erreichten Schwangerschaften zu dokumentieren.

Datenerhebung/Registrierung

Das IVF-Register, in dem die anonymisierten Daten der Behandlungsdokumentation aller Patienten/Patientinnen gespeichert werden, wird von der GÖG/ÖBIG ca. seit dem Jahr 2000 geführt. Die Daten für das IVF-Register werden von den IVF-Zentren selbst kontinuierlich (da damit auch die Abrechnung verbunden ist) in die Online-Eingabemaske eingegeben bzw. elektronisch an das Register übermittelt. Dies gilt für alle Fälle mit Kostenbeteiligung durch den IVF-Fonds. Zuvor wurden die Daten per Erhebungsblatt an die GÖG/ÖBIG gesendet und eingegeben.

Zugang zu Daten

Gemäß IVF-Fondsgesetz (Bundeskanzleramt 2010) sind die Daten bis auf Name und Anschrift der Vertragskrankenanstalten nicht öffentlich zugänglich. Die Auswertung der Daten erfolgt durch die GÖG/ÖBIG und wird in einem jährlichen und öffentlichen Bericht dem Bundesministerium für Gesundheit zur Verfügung gestellt (http://www.bmg.gv.at/home/Service/Foerderungen/Der_IVF_Fonds_Hilfe_bei_unerfuelltem_Kinderwunsch).

Dateneigenschaften/Validität

Die Daten beinhalten u. a. Name und Anschrift des IVF-Zentrums, das den Versuch durchgeführt hat, Versuche pro Person, Mitteilung eines erfolglosen Versuchs oder einer frühzeitigen Beendigung der Schwangerschaft (inkl. Grund), Mitteilung einer Schwangerschaft und Geburt sowie Anzahl der Kinder.

15.5.2.3.4 Österreichisches Hämophilie-Register (Österreichische Hämophilie Gesellschaft 2011)

Originalzweck

Ziel des Registers ist es, Daten zur Hämophilietherapie in Österreich zu erheben und eine bessere Planung der Versorgung mit Gerinnungsfaktor-Präparaten sowie eine frühere Erfassung von Nebenwirkungen, wie z. B. Auftreten eines Hemmkörpers oder von bestimmten Infektionen, zu ermöglichen. Zudem dient das Register der Erfassung fundierter wissenschaftlicher Daten.

Datenerhebung/Registrierung

Die Daten werden über eine webbasierte Eingabesoftware erhoben. Nähere Informationen waren dazu nicht erhältlich.

Zugang zu Daten

Keine Information erhältlich

Dateneigenschaften-/Validität

Keine Information erhältlich

15.5.2.3.5 Österreichisches Dialyse- und Transplantationsregister (ÖDTR)

Originalzweck

Das ÖDTR der Österreichischen Gesellschaft für Nephrologie beruht auf einer freiwilligen Kooperation aller österreichischen Behandlungszentren, die Nierenersatztherapien anbieten. Ziel ist die Erfassung der Versorgungsstruktur, der betroffenen Patienten/Patientinnen, epidemiologischer und klinischer Parameter und die wissenschaftliche Erkenntnisgewinnung.

Datenerhebung/Registrierung

Die Daten werden von der ARGE „Dialyseregister“ der Österreichischen Gesellschaft für Nephrologie in Zusammenarbeit mit Austrotransplant gesammelt, überprüft und validiert.

Zugang zu Daten

Es besteht kein direkter Zugang zu den Daten. Daten werden periodisch in Jahresberichten veröffentlicht.

Dateneigenschaften/Validität

Daten werden via Zentrums- und Patientenfragebogen erhoben. Der Zentrumsfragebogen erhebt Daten zu Struktur, Personal und zu den durchgeführten Behandlungen. Der Patientenfragebogen enthält Daten zur Patientenidentifikation (Name, Geburtsdatum, Versicherungsnummer, Wohnort etc.), zu Komorbiditäten, Behandlungsarten und zum behandelnden Zentrum.

Die Qualitätsprüfung fokussiert auf die Identifizierung von Mehrfachmeldungen und das Auffinden nicht gemeldeter Patienten/Patientinnen. Datenvergleiche mit anderen Datenbanken (z. B. Eurotransplant) werden durchgeführt.

15.5.2.3.6 Lebendspenderegister (Koordinationsbüro für das Transplantationswesen 2010)

Originalzweck

Das Register wird derzeit eingerichtet und dient der Dokumentation bzw. Qualitätssicherung sowie der medizinischen Nachsorge von Lebendspendern/Lebendspenderinnen.

Datenerhebung/Registrierung

Nach derzeitigem Stand ist eine Online-Registrierung an den jeweiligen Transplantationszentren vorgesehen, die an das Register der GÖG übermittelt wird. Über das Register erfolgt eine Einladung zur medizinischen Nachsorge der Spender/innen. Die medizinischen Daten werden von den niedergelassenen Ärzten/Ärztinnen per Post an die GÖG gesendet.

Zugang zu Daten

Noch zu definieren

Dateneigenschaften/Validität

Noch unbekannt

15.5.2.3.7 Austrian National CathLab Registry (ANCLAR) (Arbeitsgruppe Interventionelle Kardiologie 2011)

Originalzweck

Seit 1992 erfolgt in Österreich eine Erhebung von koronaren Interventionen (PCI), diagnostischen Herzkathetereingriffen und diagnostischen Koronarangiographien (CA). Es handelt sich um ein universitäres Projekt ohne externe Finanzierung, an dem sich alle österreichischen Zentren beteiligen. Das Register wurde zum Zweck des Qualitätsmanagements eingerichtet, insbesondere zur Verbesserung der Ergebnisqualität.

Datenerhebung/Registrierung

Die Daten werden per Fragebogen erhoben und von den Zentren übermittelt.

Zugang zu Daten

Es erfolgen jährliche Auswertungen, die online über eine frei zugängliche Publikation erhältlich sind.

Dateneigenschaften/Validität

Es werden Leistungszahlen, Komplikationen (Mortalität, periphere Gefäßkomplikationen, Kontrastmittelreaktionen etc.) und die Struktur der Zentren abgefragt. Zur Qualitätssicherung der Daten erfolgen Audits in den jeweiligen Zentren sowie intensive Kommunikation zwischen Zentren und Betreibern des Registers, die auch zur Erhöhung der Vollständigkeit der Daten beitragen soll. Limitierend wird das Underreporting der Komplikationen beschrieben und die fehlende Langzeitbeobachtung der Patienten/Patientinnen.

15.5.2.3.8 Arzneispezialitätenregister

Originalzweck

Gemäß Arzneimittelgesetz ist am Bundesamt für Sicherheit im Gesundheitswesen ein Arzneispezialitätenregister zu führen.

Datenerhebung/Registrierung

Alle Zulassungen, Registrierungen und Änderungen sind spätestens zwei Monate nach ihrer Rechtswirksamkeit vom Bundesamt für Sicherheit im Gesundheitswesen zu veröffentlichen.

Zugang zu Daten

Das Register ist elektronisch und öffentlich zugänglich.

Dateneigenschaften/Validität

Das Register enthält Daten zu zugelassenen, homöopathischen, apothekeneigenen und zu traditionellen pflanzlichen Arzneispezialitäten. Registriert werden Zulassungsnummer, -inhaber und -datum, Verwendung, Packung, Wirkstoff, ATC-Codes und diverse Dokumente (z. B. Gebrauchsinformation).

15.5.2.4 Sonstige

15.5.2.4.1 Critical Incident Reporting System (CIRS)

Originalzweck

Das Critical Incident Reporting System (CIRS) ist ein Berichtssystem zur anonymen Meldung von Fehlern oder Beinahe-Fehlern in Einrichtungen des Gesundheitssystems. Damit soll die Fehlerkultur verbessert werden und die Qualität der medizinischen Versorgung sowie die Patientensicherheit erhöht werden. Es wurde von der österreichischen Ärztekammer initiiert und wird in Zusammenarbeit mit dem BMG, der GÖG/BIQG und anderen Institutionen (Krankenpflegeverband etc.) seit 2009 umgesetzt. Operative Betreiberin ist die zur Österreichischen Ärztekammer gehörende Österreichische Gesellschaft für Qualitätssicherung und Qualitätsmanagement in der Medizin (ÖQMed).

Datenerhebung/Registrierung

Berichtende füllen anonym ein Online-Formular über den Vorfall aus, der auch Lösungsvorschläge für ein Verhindern eines erneuten Ereignisses beinhalten kann. Experten/Expertinnen geben anschließend ihre Vorschläge ab.

Zugang zu Daten

Die Berichte sind online öffentlich zugänglich.

Dateneigenschaften/Validität

Die Berichterstattung erfolgt nach einem vorgegebenen Template (enthält z. B. Alter, Geschlecht der betroffenen Patienten / Patientinnen, betroffener Bereich, Fallbeschreibung).

15.5.2.4.2 Widerspruchsregister (Koordinationsbüro für das Transplantationswesen 2010)

Originalzweck

In Österreich darf einem potenziellen Spender ein Organ, Organteil oder Gewebe nur dann entnommen werden, wenn zu Lebzeiten kein Widerspruch abgegeben wurde. Zur wirksamen Dokumentation eines Widerspruchs wurde 1995 das Widerspruchsregister eingerichtet.

Datenerhebung/Datenregistrierung

Die Daten werden auf Basis des Antragsformulars der GÖG, das von Personen ab 16 Jahren und für nicht geschäftsfähige Personen bzw. für Kinder von den gesetzlichen Vertretern/Vertreterinnen an die GÖG gesendet werden kann, im Register eingetragen.

Zugang zu Daten

Der Zugang zu den Daten ist über das Krankenanstalten- und Kuranstaltengesetz § 62 geregelt. Personenbezogene Anfragen sind nur von den Transplantationszentren möglich. Die Auskunft erfolgt über die Vergiftungsinformationszentrale, die im Rahmen ihres 24-Stunden-Service das Widerspruchsregister im Bedarfsfall aufruft. Anonymisierte aggregierte Auswertungen werden von ÖBIG-Transplant an der GÖG durchgeführt, weitere Auswertungen sind nicht vorgesehen (<http://www.goeg.at/de/Transplant>)

Dateneigenschaften/Validität

Die Daten beinhalten Name, Sozialversicherungsnummer, Geburtsdatum, Geschlecht, Adresse und E-Mail-Adresse der Antragsteller/innen, gegebenenfalls des gesetzlichen Vertreters und das Datum der Antragsstellung. Die Validität ist durch die Unterschrift der Antragsteller/innen bestätigt.

15.5.3 Surveys

15.5.3.1 Befragung zur Patientenzufriedenheit

Originalzweck

Die derzeit in Kooperation mit 49 Krankenanstalten durchgeführte systematische Befragung von Patienten und Patientinnen hat zum Ziel, über aufgezeigte Defizite bezüglich der Naht- und Schnittstellen im Gesundheitssystem und bezüglich institutioneninterner Prozesse (z. B. Entlassungsmanagement) eine wirkungsvolle integrierte Versorgung zu schaffen. Die Befragung soll langfristig auf weitere Bereiche der Gesundheitsversorgung ausgeweitet werden.

Datenerhebung

Im Befragungszeitraum von November 2010 bis März 2011 wurden rund 99.000 Fragebögen und Antwortkuverts an Patientinnen und Patienten durch Krankenanstalten im Rahmen der Austrittsgespräche ausgeteilt. Die Patientinnen und Patienten wurden gebeten, den Fragebogen mittels des frankierten Antwortkuverts an die GÖG/BIQG innerhalb von zwei Wochen zu übermitteln.

Zugang zu Daten

Die teilnehmenden Krankenanstalten erhalten Auswertungen der eigenen Daten sowie (anonymisierte) Referenzwerte der anderen Spitäler (Benchmarking) und können daraus Maßnahmen zur Verbesserung ableiten (best practice). Damit sind sie in der Lage, Probleme bei den Prozessen im Haus und an den Nahtstellen zu erkennen und an deren Behebung zu arbeiten. Aus Gründen des Datenschutzes erfolgt eine Veröffentlichung der Ergebnisse nur in aggregierter bzw. anonymisierter Form in Form eines österreichweiten Ergebnisberichtes. Den teilnehmenden Häusern wurde vertraglich Vertraulichkeit zugesichert.

Dateneigenschaften/Validität

Die personenbezogenen Daten beinhalten Angaben zur Dauer des Krankenhausaufenthalts, zu Geburtsjahr und Geschlecht, zur höchsten abgeschlossenen Schulbildung, zum Schweregrad der Erkrankung, zur Existenz einer Privatversicherung, zur Muttersprache und zur Selbstständigkeit beim Ausfüllen des Fragebogens. Die systembezogenen Daten beinhalten Informationen zum Gesundheitszustand, zum Aufnahmeprozess in eine Krankenanstalt (Überweisung, Vorbereitung, Untersuchungen, Wartezeiten auf Aufnahme und Bett), zur Betreuung während des Aufenthalts (Informationsweitergabe, Zeit des Personals für Betreuung und Information, Zusammenarbeit des Personals), zum Entlassungsmanagement (Info über Entlassung, nötige Maßnahmen und Angebote nach der Entlassung, Hilfe bei Nachbetreuung und Unterstützung), zur wahrgenommenen Koordination der Dienstleister und zur Gesamtzufriedenheit. Der Fragebogen wurde mittels kognitiver Interviews und einer Pilotbefragung (1.400 Fragebögen) validiert und getestet. Die im Rahmen der Hauptbefragung eingelangten Daten wurden mittels definierter Kriterien bereinigt und einer Plausibilitätsprüfung unterzogen.

15.5.3.2 Blutverbrauchserhebung (Gesundheit Österreich GmbH 2011)

Originalzweck

Zweck der seit 1999 jährlichen Erhebung ist es, den Einsatz von Blutprodukten zu optimieren und ihren Verwurf zu minimieren, qualitätssichernde Maßnahmen zu implementieren und die Anzahl an in österreichischen Krankenanstalten transfundierten und verworfenen Konserven transparent zu machen.

Datenerhebung

Alle Krankenanstalten in Österreich melden einmal pro Jahr per Online-Eingabe die Daten des Blutverbrauchs.

Zugang zu Daten

Jährlich erfolgt eine aggregierte Standardauswertung, die online abrufbar ist. Weitere öffentliche Auswertungen sind aufgrund des Datenschutzes nicht möglich.

Dateneigenschaften/Validität

Es werden Daten zu den transfundierten Produkten je 1.000 Einwohner erhoben und die Einheiten der transfundierten Produkte im Jahresvergleich sowie (für ausgewählte Produkte) im Ländervergleich dargestellt. Seit dem Jahr 2000 konnte jährlich ein hundertprozentiger Rücklauf erzielt werden.

15.5.3.3 Weitere Surveys

Für eine ausführliche Beschreibung der weiteren in Tabelle 9.1 angeführten Surveys siehe Referenzen ebendort.

15.5.4 Amtliche Statistiken

Siehe Tabelle 15.1.; ausführliche Beschreibung der Statistiken in den dort angeführten Referenzen.

15.5.5 Daten zu Planungszwecken

Siehe Tabelle 15.1.; Beschreibung der Daten in den dort angeführten Referenzen.

15.5.6 Daten zu Versorgungsstrukturen

Siehe Tabelle 15.1.; Beschreibung der Daten in den dort angeführten Referenzen.

15.5.7 Sonstige

15.5.7.1 Österreichisches Gesundheitsinformationssystem (ÖGIS)

Originalzweck

Das Österreichische Gesundheitsinformationssystem (ÖGIS) ist als geografisches Informationssystem (GIS) konzipiert. Es bietet Karten- und Zeitreihendarstellungen sowie einfache statistische Analysemethoden zu allen epidemiologischen Kernthemen an (z. B. Lebenserwartung, Sterblichkeit, Krebsinzidenz, Krankenhaushäufigkeit, subjektiver Gesundheitszustand). Außerdem sind über das ÖGIS detaillierte Informationen zu nahezu allen Teilsektoren des Gesundheitssystems verfügbar (z. B. Akutkrankeanstalten, Rehabilitationszentren, Alten- und Pflegeheime, niedergelassene Ärzte, mobile Dienste, Rettungsstellen).

Datenerhebung

Die GÖG holt die Informationen zu den einzelnen gesundheitsrelevanten Themenbereichen von verschiedenen Quellen ein und führt diese Sekundärdaten in der dem ÖGIS zugrunde liegenden Datenbank QGIS zusammen.

Zugang zu Daten

ÖGIS-Auswertungen können auf Wunsch unter Berücksichtigung der Datenschutzerfordernisse gegen Bezahlung durchgeführt werden (Kosten auf Anfrage).

Dateneigenschaften/Validität

Die Daten werden je nach Verfügbarkeit mindestens einmal jährlich aktualisiert, sodass für viele Themenbereiche Zeitreihendarstellungen möglich sind. Die im ÖGIS verfügbaren Basisdaten werden umfassend validiert und ebenso wie die Methoden bzw. die daraus resultierenden Indikatoren mit den Publikationen anderer einschlägig befasster Institutionen im Inland (v. a. BMG und ST.AT) wie im Ausland (v. a. EU/Eurostat, OECD und WHO/Europe) abgestimmt.

15.5.7.2 Regionales Gesundheitsinformationssystem (REGIS)

Originalzweck

Das Regionale Gesundheitsinformationssystem REGIS ist eine Dienstleistung der GÖG/ÖBIG und Teil des Österreichischen Gesundheitsinformationssystems ÖGIS. In REGIS werden Eckdaten zu gesundheitsrelevanten Indikatoren regional vergleichbar kartografisch dargestellt. Es wendet sich an sämtliche Personen und Institutionen, die im Gesundheitsbereich tätig bzw. an diesem Bereich interessiert sind.

Datenerhebung

Siehe ÖGIS

Zugang zu Daten

REGIS steht kostenfrei auf der Homepage <http://regis.goeg.at/> zur Verfügung.

Dateneigenschaften/Validität

REGIS enthält fertige Auswertungen zu den Themenbereichen Demographie, Lebenserwartung, Sterblichkeit, Spitalsaufnahmen, Lifestyle und Verkehrsunfälle.

15.5.7.3 Pharma-Preisinformation (PPI)

Originalzweck

Das Pharma-Preisinformationssystem wurde eingerichtet, um zuverlässige und unabhängige Information über Arzneimittelpreise in allen EU-Mitgliedsländern, in der Schweiz, in Kroatien und Norwegen zu erhalten.

Datenerhebung

Grundlage sind Daten aus den einzelnen Ländern, die in das PPI-System integriert werden.

Zugang zu Daten

Eine Datenauswertung kann auf schriftliche Anfrage mit Angaben zu Wirkstoffbezeichnung oder Produktnamen, zum ATC-Code (sofern vorhanden) und zu den gewünschten Ländern erfolgen. Das Ergebnis wird in einem Excel-File mit allen Preisstufen (Fabriksabgabepreis, Apothekenverkaufspreis inkl./exkl. Umsatzsteuer) zur Verfügung gestellt. Die Kosten belaufen sich auf 70 Euro pro Produkt (zwei Stärken, eine Darreichungsform, alle Packungsgrößen). Es können auch Spezialauswertungen (z. B. zu Preisen aller Arzneimittel, die einen bestimmten Wirkstoff enthalten) durchgeführt werden.

Dateneigenschaften/Validität

Die Daten enthalten aktuelle, europaweit vergleichbare Preisinformationen.

15.5.8 Kompass

15.5.8.1 Spitalskompass

Originalzweck

Der österreichische Spitalskompass hat zum Ziel, die interessierte Öffentlichkeit über das Leistungsangebot österreichischer Spitäler zu informieren.

Datenerhebung

Die Daten der Rubrik „Diagnosen und Leistungen“ stammen aus der Diagnosen- und Leistungsdokumentation der österreichischen Krankenanstalten (jährliche Aktualisierung meist im November). Die meisten anderen Informationen (z. B. Telefonnummer, Abteilungsinhalte etc.) werden von den Spitälern über entsprechende Eingabemasken selbst aktualisiert.

Zugang zu Daten

Eine Online-Suchmaske ermöglicht die Suche nach Spitälern in Österreich, klassifiziert nach Standort, vorhandenen medizinischen Einrichtungen (z. B. Geräte, Ambulanzen), Fachabteilungen, Häufigkeit erbrachter medizinischer Leistungen und Ausstattung des

Krankenzimmers. Zum Thema Geburt liegen besonders ausführliche Informationen zu den spezifischen Angeboten vor.

Dateneigenschaften/Validität

Es werden Informationen zu Anzahl und Art der Spitäler in österreichischen Bundesländern, zu den vorhandenen Abteilungen, zur Häufigkeit durchgeführter bestimmter medizinischer Leistungen und zur Ausstattung der Spitäler bzw. Krankenzimmer zur Verfügung gestellt.

15.5.8.2 Rehakompass

Originalzweck

Der Rehakompass ist ein elektronisches Verzeichnis, das Patienten/Patientinnen, Angehörigen und der interessierten Öffentlichkeit Informationen bietet zur Organisation, zum medizinischen Angebot und zur Ausstattung der stationären Rehabilitationseinrichtungen in Österreich.

Datenerhebung

Die Eingabe und Aktualisierung der Daten erfolgt durch die Rehabilitationseinrichtungen bzw. -träger im Eingabemodus des österreichischen Rehakompasses.

Zugang zu Daten

Eine frei zugängliche Online-Suchmaske ermöglicht die Suche nach Rehabilitationseinrichtungen nach Bundesland, Indikation und Sozialversicherungsträger. Zusätzlich sind über das Portal allgemeine Informationen zur Rehabilitation zugänglich (Grundsätze, Zuständigkeiten, Antragsformalitäten etc.).

Dateneigenschaften/Validität

Es erfolgt eine jährliche Aktualisierung und eine Kontrolle/Freigabe der eingegebenen Daten durch die GÖG, gegebenenfalls nach Rücksprache mit dem HVB bzw. den einzelnen Sozialversicherungsträgern.

15.5.8.3 Suchthilfekompass

Originalzweck

Der Suchthilfekompass ist ein elektronisches Verzeichnis und bietet derzeit Informationen zu jenen Drogenhilfeeinrichtungen, die nach § 15 Suchtmittelgesetz vom Bundesministerium für Gesundheit kundgemacht sind. Dabei handelt es sich um ambulante oder stationäre Einrichtungen, die vorgegebene Qualitätskriterien erfüllen. Einrichtungsprofile verschaffen einen umfassenden und differenzierten Überblick über die zur Verfügung stehenden Betreuungs- und Behandlungsmöglichkeiten.

Datenerhebung

Die Daten werden von den Einrichtungen eingegeben.

Zugang zu Daten

Die Daten sind online frei zugänglich. Es besteht die Möglichkeit, über Volltextsuche oder das Setzen von Filtern die gewünschten Anforderungen anzugeben. Das Ergebnis dieser Suche ist eine nach Bundesländern sortierte Liste jener Einrichtungen, die über das gesuchte Angebot verfügen. Eine nach Zielgruppen, Angeboten und begleitenden Maßnahmen aufgeschlüsselte Suchfunktion trägt dazu bei, geeignete Hilfseinrichtungen zu finden.

Dateneigenschaften/Validität

Der Suchthilfekompass beinhaltet Informationen über die Basisdaten der Einrichtungen (z. B. Adresse, Öffnungszeiten, Schwerpunkt der Tätigkeit etc.), Zielgruppen nach Geschlecht, Alter, Betreuungsproblematik (z. B. Alkohol, Medikamente etc.) und Angebote im Zusammenhang mit gesundheitsbezogenen Maßnahmen (§ 11 SMG). Weiters werden Tätigkeiten der Einrichtungen u. a. in den Bereichen aufsuchende und nachgehende Drogenhilfe, suchtbegleitende Schadensminimierung und Soforthilfe, Alltags- und Freizeitgestaltung etc. dargestellt. Auch Auskunft über die in der Einrichtung beschäftigten Berufsgruppen wird zur Verfügung gestellt.

Die Aktualisierung erfolgt jährlich.

15.5.9 Verlinkung unterschiedlicher administrativer Daten am Beispiel des GAP-DRG-Projekts

Originalzweck

Das GAP-DRG-Projekt (Grundlagenforschung für ambulante patientenbezogene diagnosis related groups / DRGs) dient der Erstellung einer Datenbank, die unterschiedliche personenbezogene (aber pseudonymisierte) Datenbestände des österreichischen Gesundheitssystems verlinkt.

Datenerhebung

Die bisher integrierten Datenquellen sind die Daten der ärztlichen Eigen- und Folgekostenanalyse (FOKO), die Minimal-Basic-DataSet-(MBDS-)Daten der Leistungsdokumentation in Krankenanstalten inklusive der Daten des privaten Krankenanstaltenfinanzierungsfonds (PRIKRAF), Daten der „zentralen Partnerverwaltung“ (ZPV) des HVB und der Metahonorarordnung. Zusätzlich wurden zahlreiche Metadaten (z. B. Details über Bettenanzahl pro Spitalsabteilung) und Zusatzinformationen (z. B. Schlüssel für die Übersetzung von ATC-Codes in ICD-Codes) eingespeist.

Zugang zu Daten

Die Datenbank wird kontinuierlich weiterentwickelt. Ob Auswertungen möglich sind, ist abhängig von der Fragestellung und der Rolle des Antragstellers.

Dateneigenschaften/Validität

Durch die Verlinkung von Leistungs- und Diagnosedaten aus dem stationären und ambulanten (niedergelassenen) Bereich mit Patientenstammdaten können mit der Datenbank Analysen zum Leistungsgeschehen über traditionelle Systemgrenzen hinweg erfolgen. Eine Reihe von generellen Qualitätssicherungs- und Validierungsmaßnahmen wurden durchgeführt, diese ersetzen aber nicht eine projektspezifische Prüfung bei jedem einzelnen Projekt, das auf die Datenbank zurückgreift.

15.6 Checklisten zur Qualitätsbewertung der Studien

An den Beginn jeder der im Folgenden dargestellten Checklisten gehören folgende Angaben:

Ref ID: _____

Autor und Jahr der Publikation: _____

Initialen der Reviewerin / des Reviewers: _____

15.6.1 Checkliste zur Beurteilung von randomisierten kontrollierten Studien (RCT)

Kriterien zur Beurteilung von RCTs	Ja	Nein	Unklar
SELEKTION			
Wurde eine adäquate Methode der Randomisierung verwendet, um die Studienteilnehmer/innen unterschiedlichen Behandlungsgruppen zuzuteilen?			
War die Geheimhaltung der Randomisierungssequenz gewährleistet? (allocation concealment)			
VERGLEICHBARKEIT			
Waren die Behandlungsgruppen nach der Randomisierung ähnlich in Bezug auf wesentliche prognostische Merkmale oder Confounder?			
Waren die Studienteilnehmer/innen verblindet?			
Waren jene Personen, die die Intervention verabreichten, verblindet?			
Waren jene Personen, die die Endpunkte erhoben, verblindet?			
Erhielten alle Studiengruppen außer der zu untersuchenden Intervention identische Behandlungen?			
ENDPUNKTE			
Wurden die Endpunkte in allen Behandlungsgruppen zum selben Zeitpunkt erhoben?			
War die allgemeine Drop-out-Rate geringer als 20 %?			
War die differenzielle Drop-out-Rate zwischen den Studiengruppen geringer als 15 Prozentpunkte?			
Wurde eine Intention-to-treat-(ITT-)Analyse durchgeführt und war diese korrekt?			
Kann angenommen werden, dass alle erhobenen Endpunkte auch berichtet wurden?			
Beurteilung des Bias-Risikos	Gering	Unklar	Hoch
Kommentare			

15.6.2 Checkliste zur Beurteilung von Kohortenstudien

Kriterien zur Beurteilung von Kohortenstudien	Ja	Nein	Unklar
SELEKTION			
Wurden die Studiengruppen aus derselben Population und während derselben Zeitperiode rekrutiert?			
Wurde durch die Autoren ausgeschlossen, dass ein definierter Endpunkt bereits zu Studienbeginn vorhanden war?			
Wurden Interventionen in allen Gruppen auf gleiche Art und Weise beurteilt?			
VERGLEICHBARKEIT			
Ist die Verteilung der prognostischen Faktoren zwischen den Gruppen ausreichend beschrieben?			
Ist die Verteilung der prognostischen Faktoren zwischen den Gruppen ähnlich?			
ENDPUNKTE			
Wurden Endpunkte auf dieselbe Art und verblindet beurteilt?			
Wurden potenzielle Confounder in der statistischen Auswertung berücksichtigt?			
War die Studienlaufzeit adäquat und für alle Gruppen identisch?			
War die allgemeine Drop-out-Rate geringer als 20 %?			
War die differenzielle Drop-out-Rate zwischen den Studiengruppen geringer als 15 Prozentpunkte?			
Beurteilung des Bias-Risikos	Gering	Unklar	Hoch
Kommentare			

15.6.3 Checkliste zur Beurteilung von Fall-Kontroll-Studien

Kriterien zur Beurteilung von Fall-Kontroll-Studien	Ja	Nein	Unklar
SELEKTION			
Wurde ein „Fall“ exakt definiert?			
Wurden die Studiengruppen aus derselben Population rekrutiert?			
Hatten Personen in der Fall- und in der Kontrollgruppe dasselbe Risiko, der Exposition ausgesetzt zu werden?			
Ist die Verteilung prognostischer Faktoren zwischen den Gruppen ausreichend beschrieben?			
Ist die Verteilung der prognostischen Faktoren zwischen den Gruppen ähnlich? Ist „Over-matching“ möglich?			
EXPOSITION			
Wurde die Exposition in der Fall- und Kontrollgruppe auf gleiche Art und im besten Fall verblindet beurteilt?			
War die Rate der Personen, die eine Beteiligung an der Studie abgelehnt haben, in beiden Gruppen ähnlich?			
Wurde eine adäquate statistische Analyse durchgeführt?			
Beurteilung des Bias-Risikos	Gering	Unklar	Hoch
Kommentare			

15.6.4 Checkliste zur Beurteilung von systematischen Reviews und Metaanalysen

Kriterien zur Beurteilung von systematischen Reviews und Metaanalysen	Ja	Nein	Unklar
Basiert der Review auf einer klar definierten Forschungsfrage(PICO)?			
Wurden Selektionskriterien für Studien klar definiert?			
Wurde eine umfangreiche systematische Literatursuche durchgeführt?			
Wurde der Ein- oder Ausschluss von Studien von zumindest zwei Personen beurteilt?			
Wurde die methodische Qualität der eingeschlossenen Studien von zumindest zwei Personen beurteilt?			
Wurde die methodische Qualität der Studien bei der Evidenzsynthese berücksichtigt?			
METAANALYSEN			
Wurde Publikationsbias beurteilt?			
Wurde Heterogenität statistisch beurteilt?			
Wurden die Ursachen für vorhandene Heterogenität adäquat analysiert?			
War die Auswahl des statistischen Modells adäquat?			
Beurteilung des Bias-Risikos	Gering	Unklar	Hoch
Kommentare			

15.6.5 Checkliste zur Beurteilung von diagnostischen Studien

Kriterien zur Beurteilung von diagnostischen Studien	Ja	Nein	Unklar
SELEKTION			
Repräsentiert die Studie jene Patientinnen und Patienten, die den Test in der Praxis erhalten werden?			
Sind die Auswahlkriterien für Patientinnen und Patienten klar definiert?			
VERGLEICHBARKEIT			
Wurde ein Referenztest verwendet, der als „Goldstandard“ angesehen werden kann?			
Wurden Indextest und Referenztest zum selben Zeitpunkt durchgeführt?			
Wurde der Referenztest an der gesamten Studienpopulation ausgeführt?			
Wurde der Referenztest unabhängig vom Resultat des Indextests durchgeführt?			
Wurde ausgeschlossen, dass der Indextest bereits Bestandteil des gewählten Referenztests ist?			
Wurde die Durchführung der Tests so genau beschrieben, dass eine Replikation möglich wäre?			
ENDPUNKTE			
Wurden die Ergebnisse des Indextests ohne Kenntnis der Ergebnisse des Referenzverfahrens interpretiert?			
Wurden die Ergebnisse des Referenzverfahrens ohne Kenntnis der Ergebnisse des Indextests interpretiert?			
Waren bei der Interpretation der Testergebnisse dieselben klinischen Daten verfügbar, die auch vorliegen würden, wenn der Test in der klinischen Praxis durchgeführt wird?			
Wurde der Anteil an nicht-interpretierbaren Resultaten berichtet?			
Wurde die Höhe der Drop-out-Rate während der Tests genannt?			
Beurteilung der internen Validität der Studie: Bias-Risiko	Gering	Unklar	Hoch
Kommentare			

15.7 Checklisten und Tabellen zum Kapitel ökonomische Übersichtsarbeit

Tabelle 15.2:

Beispiel für Datenextraktionsformular für gesundheitsökonomische Evaluationen

Allgemeine Information
Datum der Datenextraktion
Studienautor/autorin, Jahr
Identifizierung des Reviewers
Anmerkungen
Spezifische Information
Allgemeine Studiencharakteristika
Verifizierung der erfüllten Einschlusskriterien
Forschungsfrage (Ziel der gesundheitsökonomischen Evaluation)
Studienpopulation und Setting
Intervention und Vergleichsintervention(en)
Studientyp (CEA, CUA, CBA)
Land/Währung und Jahr auf das sich Währung bezieht
Perspektive (Gesundheitssystem, Kostenträger, Gesellschaft etc.)
Methodische Qualität
Anmerkungen
Spezifische Studiencharakteristika: Berechnung von Effekten und Kosten
Datenquellen
» Datenquellen für Effektivitätsdaten (Einzelstudie, Review etc.)
» Datenquellen für Kostendaten (Sekundär- oder Primärdaten)
Methode zur Berechnung von Effekten und Kosten
» Berechnung und Bewertung der Gesundheitseffekte (z. B. QALYs auf Basis von standard gamble)
» Berechnung der Kosten
» Modellierung, wenn angewendet (Modelltyp, Beschreibung der wichtigsten Komponenten)
Ergebnisse
Gesundheitseffekte
Kosten
Synthese von Effekten und Kosten (z.B. inkrementelles Kosten-Effektivitätsverhältnis)
Ergebnisse der Sensitivitätsanalyse
Gesamtsynthese der Ergebnisse
Größendimension und Richtung der Ergebnisse
Auswirkungen für die Praxis
Anmerkungen

Quelle: Khan, Riet ter 2001

Tabelle 15.3:
BMJ-Checkliste

Referees' checklist (also to be used, implicitly, by authors)				
Item	Yes	No	Not clear	Not appropriate
Study design				
(1) The research question is stated	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
(2) The economic importance of the research question is stated	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
(3) The viewpoint(s) of the analysis are clearly stated and justified	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
(4) The rationale for choosing the alternative programmes or interventions compared is stated	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
(5) The alternatives being compared are clearly described	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
(6) The form of economic evaluation used is stated	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
(7) The choice of form of economic evaluation is justified in relation to the questions addressed	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Data collection				
(8) The source(s) of effectiveness estimates used are stated	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
(9) Details of the design and results of effectiveness study are given (if based on a single study)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(10) Details of the method of synthesis or meta-analysis of estimates are given (if based on an overview of a number of effectiveness studies)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(11) The primary outcome measure(s) for the economic evaluation are clearly stated	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
(12) Methods to value health states and other benefits are stated	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(13) Details of the subjects from whom valuations were obtained are given	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(14) Productivity changes (if included) are reported separately	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(15) The relevance of productivity changes to the study question is discussed	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(16) Quantities of resources are reported separately from their unit costs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
(17) Methods for the estimation of quantities and unit costs are described	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
(18) Currency and price data are recorded	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
(19) Details of currency of price adjustments for inflation or currency conversion are given	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
(20) Details of any model used are given	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(21) The choice of model used and the key parameters on which it is based are justified	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Analysis and interpretation of results				
(22) Time horizon of costs and benefits is stated	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
(23) The discount rate(s) is stated	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(24) The choice of rate(s) is justified	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(25) An explanation is given if costs or benefits are not discounted	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(26) Details of statistical tests and confidence intervals are given for stochastic data	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(27) The approach to sensitivity analysis is given	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(28) The choice of variables for sensitivity analysis is justified	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(29) The ranges over which the variables are varied are stated	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(30) Relevant alternatives are compared	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
(31) Incremental analysis is reported	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(32) Major outcomes are presented in a disaggregated as well as aggregated form	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
(33) The answer to the study question is given	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
(34) Conclusions follow from the data reported	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
(35) Conclusions are accompanied by the appropriate caveats	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

Quelle: Drummond und Jefferson 1996

Tabelle 15.4:
CHEC-Checkliste

	Item	Yes	No
1.	Is the study population clearly described?	<input type="checkbox"/>	<input type="checkbox"/>
2.	Are competing alternatives clearly described?	<input type="checkbox"/>	<input type="checkbox"/>
3.	Is a well-defined research question posed in answerable form?	<input type="checkbox"/>	<input type="checkbox"/>
4.	Is the economic study design appropriate to the stated objective?	<input type="checkbox"/>	<input type="checkbox"/>
5.	Is the chosen time horizon appropriate to include relevant costs and consequences?	<input type="checkbox"/>	<input type="checkbox"/>
6.	Is the actual perspective chosen appropriate?	<input type="checkbox"/>	<input type="checkbox"/>
7.	Are all important and relevant costs for each alternative identified?	<input type="checkbox"/>	<input type="checkbox"/>
8.	Are all costs measured appropriately in physical units?	<input type="checkbox"/>	<input type="checkbox"/>
9.	Are costs valued appropriately?	<input type="checkbox"/>	<input type="checkbox"/>
10.	Are all important and relevant outcomes for each alternative identified?	<input type="checkbox"/>	<input type="checkbox"/>
11.	Are all outcomes measured appropriately?	<input type="checkbox"/>	<input type="checkbox"/>
12.	Are outcomes valued appropriately?	<input type="checkbox"/>	<input type="checkbox"/>
13.	Is an incremental analysis of costs and outcomes of alternatives performed?	<input type="checkbox"/>	<input type="checkbox"/>
14.	Are all future costs and outcomes discounted appropriately?	<input type="checkbox"/>	<input type="checkbox"/>
15.	Are all important variables, whose values are uncertain, appropriately subjected to sensitivity analysis?	<input type="checkbox"/>	<input type="checkbox"/>
16.	Do the conclusions follow from the data reported?	<input type="checkbox"/>	<input type="checkbox"/>
17.	Does the study discuss the generalizability of the results to other settings and patient/ client groups?	<input type="checkbox"/>	<input type="checkbox"/>
18.	Does the article indicate that there is no potential conflict of interest of study researcher(s) and funder(s)?	<input type="checkbox"/>	<input type="checkbox"/>
19.	Are ethical and distributional issues discussed appropriately?	<input type="checkbox"/>	<input type="checkbox"/>

Quelle: Shemilt, Mugford 2009, Evers, Goossens et al. 2005

Tabelle 15.5:

Checkliste zur Beurteilung der methodischen Qualität gesundheitsökonomischer Studien (entwickelt im Konsensusverfahren von den gesundheitsökonomischen Projektgruppen München, Hannover, Ulm)

Checkliste methodische Qualität		
Autoren, Titel und Publikationsorgan:	1 = Kriterium erfüllt 1/2 = Kriterium teilweise erfüllt 0 = Kriterium nicht erfüllt nr = nicht relevant	1, ½, 0, nr
Fragestellung		
1. Wurde die Fragestellung präzise formuliert?		
2. Wurde der medizinische und ökonomische Problemkontext ausreichend dargestellt?		
Evaluationsrahmen		
3. Wurden alle in die Studie einbezogenen Technologien hinreichend detailliert beschrieben?		
4. Wurden alle im Rahmen der Fragestellung relevanten Technologien verglichen?		
5. Wurde die Auswahl der Vergleichstechnologien schlüssig begründet?		
6. Wurde die Zielpopulation klar beschrieben?		
7. Wurde ein für die Fragestellung angemessener Zeithorizont für Kosten und Gesundheitseffekte gewählt und angegeben?		
8. Wurde der Typ der gesundheitsökonomischen Evaluation explizit genannt?		
9. Wurden sowohl Kosten als auch Gesundheitseffekte untersucht?		
10. Wurde die Perspektive der Untersuchung eindeutig gewählt und explizit genannt?		
Analysemethoden und Modellierung		
11. Wurden adäquate statistische Tests/Modelle zur Analyse der Daten gewählt und hinreichend gründlich beschrieben?		
12. Wurden in entscheidungsanalytischen Modellen die Modellstruktur und alle Parameter vollständig und nachvollziehbar dokumentiert (in der Publikation bzw. einem technischen Report)?		
13. Wurden die relevanten Annahmen explizit formuliert?		
14. Wurden in entscheidungsanalytischen Modellen adäquate Datenquellen für die Pfadwahrscheinlichkeiten gewählt und eindeutig genannt?		
Gesundheitseffekte		
15. Wurden alle für die gewählte Perspektive und den gewählten Zeithorizont relevanten Gesundheitszustände berücksichtigt und explizit aufgeführt?		
16. Wurden adäquate Quellen für die Gesundheitseffektdata gewählt und eindeutig genannt?		
17. Wurden das epidemiologische Studiendesign und die Auswertungsmethoden adäquat gewählt und beschrieben und wurden die Ergebnisse detailliert dargestellt? (falls auf einer einzelnen Studie basierend)		
18. Wurden angemessene Methoden zur Identifikation, Extraktion und Synthese der Effektparameter verwendet und wurden sie detailliert beschrieben? (falls auf einer Informationssynthese basierend)		
19. Wurden die verschiedenen Gesundheitszustände mit Präferenzen bewertet und dafür geeignete Methoden und Messinstrumente gewählt und angegeben?		
20. Wurden adäquate Quellen der Bewertungsdaten für die Gesundheitszustände gewählt und eindeutig genannt?		
21. Wurde die Evidenz der Gesundheitseffekte ausreichend belegt? (s. ggf. entsprechende Kontextdokumente)		

Checkliste methodische Qualität	
Kosten	
22. Wurden die den Kosten zugrunde liegenden Mengengerüste hinreichend gründlich dargestellt? 23. Wurden adäquate Quellen und Methoden zur Ermittlung der Mengengerüste gewählt und eindeutig genannt? 24. Wurden die den Kosten zugrunde liegenden Preisgerüste hinreichend gründlich beschrieben? 25. Wurden adäquate Quellen und Methoden zur Ermittlung der Preise gewählt und eindeutig genannt? 26. Wurden die einbezogenen Kosten anhand der gewählten Perspektive und des gewählten Zeithorizontes schlüssig begründet und wurden alle relevanten Kosten berücksichtigt? 27. Wurden Daten zu Produktivitätsausfallkosten (falls berücksichtigt) getrennt aufgeführt und methodisch korrekt in die Analyse einbezogen? 28. Wurde die Währung genannt? 29. Wurden Währungskonversionen adäquat durchgeführt? 30. Wurden Preisanpassungen bei Inflation oder Deflation adäquat durchgeführt?	
Diskontierung	
31. Wurden zukünftige Gesundheitseffekte <u>und</u> Kosten adäquat diskontiert? 32. Wurde das Referenzjahr für die Diskontierung angegeben bzw. bei fehlender Diskontierung das Referenzjahr für die Kosten? 33. Wurden die Diskontraten angegeben? 34. Wurde die Wahl der Diskontraten bzw. der Verzicht auf eine Diskontierung plausibel begründet?	
Ergebnispräsentation	
35. Wurden Maßnahmen zur Modellvalidierung ergriffen und beschrieben? 36. Wurden absolute Gesundheitseffekte und absolute Kosten jeweils pro Kopf bestimmt und dargestellt? 37. Wurden inkrementelle Gesundheitseffekte und inkrementelle Kosten jeweils pro Kopf bestimmt und dargestellt? 38. Wurde eine für den Typ der gesundheitsökonomischen Evaluation sinnvolle Maßzahl für die Relation zwischen Kosten und Gesundheitseffekt angegeben? 39. Wurden reine (nicht lebensqualitätsadjustierte) klinische Effekte berichtet? 40. Wurden die relevanten Ergebnisse in disaggregierter Form dargestellt? 41. Wurden populationsaggregierte Kosten und Gesundheitseffekte dargestellt?	
Behandlung von Unsicherheiten	
42. Wurden univariate Sensitivitätsanalysen für die relevanten Parameter durchgeführt? 43. Wurden multivariate Sensitivitätsanalysen für die relevanten Parameter durchgeführt? 44. Wurde Sensitivitätsanalysen für die relevanten strukturellen Elemente durchgeführt? 45. Wurden in den Sensitivitätsanalysen realistische Werte oder Wertebereiche bzw. Strukturvarianten berücksichtigt und angegeben? 46. Wurden die Ergebnisse der Sensitivitätsanalysen hinreichend dokumentiert? 47. Wurden adäquate statistische Inferenzmethoden (statistische Tests, Konfidenzintervalle) für stochastische Daten eingesetzt und die Ergebnisse berichtet?	
Diskussion	
48. Wurde die Datenqualität kritisch beurteilt? 49. Wurden Richtung und Größe des Einflusses unsicherer oder verzerrter Parameterschätzung auf das Ergebnis konsistent diskutiert? 50. Wurde Richtung und Größe des Einflusses struktureller Modellannahmen auf das Ergebnis konsistent diskutiert?	

Checkliste methodische Qualität		
51. Wurden die wesentlichen Einschränkungen und Schwächen der Studie diskutiert?		
52. Wurden plausible Angaben zur Generalisierbarkeit der Ergebnisse gemacht?		
53. Wurden wichtige ethische und Verteilungsfragen diskutiert?		
54. Wurde das Ergebnis sinnvoll im Kontext mit unabhängigen Gesundheitsprogrammen diskutiert?		
Schlussfolgerungen		
55. Wurden in konsistenter Weise Schlussfolgerungen aus den berichteten Daten/Ergebnissen abgeleitet?		
56. Wurde eine auf Wissensstand und Studienergebnissen basierende Antwort auf die Fragestellung gegeben?		

Quelle: Siebert, Behrend 1999

Tabelle 15.6:
Checkliste (Philips, Ginnelly 2004)

Beurteilung der Qualität von entscheidungsanalytischen gesundheitsökonomischen Modellen	Ja	Nein	Nicht enthalten
Modellstruktur			
1. Beschreibung des Entscheidungsproblems			
1.1. Wird das Entscheidungsproblem klar beschrieben? 1.2. Sind das Ziel der Evaluation und das beschriebene Modell mit dem Entscheidungsproblem konsistent? 1.3. Wird der primäre Entscheidungsträger genannt?			
2. Beschreibung der Perspektive			
2.1. Wird die Perspektive der Untersuchung eindeutig gewählt und explizit genannt? 2.2. Sind die Input-Parameter des Modells konsistent mit der genannten Perspektive? 2.3. Wird der Gültigkeitsbereich des Modells angegeben und begründet? 2.4. Sind die Modellergebnisse mit Perspektive, Gültigkeitsbereich und Gesamtziel der Modellierung konsistent?			
3. Logik der Modellstruktur			
3.1. Ist die Struktur des Modells mit einer kohärenten Theorie zum evaluierten Gesundheitszustand konsistent? 3.2. Werden die Daten(quellen) zur Konstruktion der Modellstruktur beschrieben und schlüssig begründet? 3.3. Werden die im Modell verwendeten kausalen Zusammenhänge beschrieben und schlüssig begründet?			
4. Strukturelle Annahmen			
4.1. Werden die strukturellen Annahmen transparent beschrieben und begründet? 4.2. Sind die strukturellen Annahmen vor dem Hintergrund von Evaluierungsziel, Perspektive und Gültigkeit zweckvoll?			
5. Strategien/Vergleichstechnologien			
5.1. Werden alle im Rahmen der Fragestellung relevanten Technologien hinreichend detailliert beschrieben? 5.2. Werden alle im Rahmen der Fragestellung relevanten Technologien verglichen? 5.3. Wird der Ausschluss einer plausiblen Alternative ausreichend begründet?			
6. Modelltyp Wurde ein dem Entscheidungsproblem angemessener Modelltyp gewählt?			
7. Zeithorizont			
7.1. Ist der Zeithorizont des Modells ausreichend lang, um alle Unterschiede zwischen den gewählten Alternativen aufzuzeigen?			

Beurteilung der Qualität von entscheidungsanalytischen gesundheitsökonomischen Modellen	Ja	Nein	Nicht enthalten
7.2. Werden der gewählte Zeithorizont des Modells sowie die Dauer von Behandlung und Behandlungseffekt beschrieben und schlüssig begründet?			
8. Krankheitszustände/Pfade			
Spiegeln die Krankheitszustände oder die Pfade (Entscheidungsbaum) den tatsächlichen biologischen Prozess der untersuchten Erkrankung und der Auswirkung der Intervention wider?			
9. Zykluslänge			
Wird die gewählte Zykluslänge beschrieben und auf Basis des natürlichen Krankheitsverlaufs ausreichend begründet?			
Daten			
1. Datenidentifikation			
1.1. Werden die Methoden zur Identifikation der Daten transparent dargestellt und sind sie für das Evaluationsziel geeignet? 1.2. Wird die Entscheidung zur Datenauswahl ausreichend begründet? 1.3. Wird der Identifikation von Daten für zentrale Modellparameter entsprechende Aufmerksamkeit gewidmet? 1.4. Wird die Datenqualität angemessen evaluiert? 1.5. Wird bei der Nutzung von ExpertInnenmeinungen die Methode der Datengenerierung beschrieben und begründet?			
2. Modellierung			
2.1. Basiert die Modellierungsmethode auf begründeten und anerkannten statistischen und epidemiologischen Methoden?			
2a. Basisfalldaten			
2a.1. Wird die Datenauswahl für die Basisfallanalyse ausreichend beschrieben und begründet? 2a.2. Werden die Übergangswahrscheinlichkeiten richtig kalkuliert? 2a.3. Werden ‚half-cycle‘ Korrekturen für Kosten und Gesundheitseffekte durchgeführt oder die Unterlassung begründet?			
2b. Behandlungseffekte			
2b.1. Wird die Datensynthese bei der Verwendung relativer Effekte aus klinischen Studien methodisch korrekt und transparent durchgeführt? 2b.2. Werden die Methoden und Annahmen zur Extrapolation von kurzfristigen auf langfristige Gesundheitseffekte beschrieben und begründet? 2b.3. Werden Annahmen bezüglich des Behandlungseffektes nach Behandlungsstop beschrieben und begründet? 2b.4. Werden Sensitivitätsanalysen zu alternativen Annahmen durchgeführt? 2b.5. Werden die Diskontierungsraten für Gesundheitseffekte beschrieben und begründet?			
2c. Kosten			

Beurteilung der Qualität von entscheidungsanalytischen gesundheitsökonomischen Modellen	Ja	Nein	Nicht enthalten
2c.1. Werden alle relevanten Kosten berücksichtigt? 2c.2. Werden alle Quellen für die Kostendaten beschrieben? 2c.3. Werden geeignete Methoden zur Ermittlung der Mengen- und Preisgerüste angewendet? 2c.4. Werden die Diskontierungsraten für Kosten beschrieben und begründet?			
2d. Lebensqualitätsbewertung			
2d.1. Werden Nutzwerte korrekt in das Modell integriert? 2d.2. Wird die Quelle für Nutzwerte beschrieben? 2d.3. Ist die zur Gewinnung der Nutzwerte verwendete Methode geeignet und begründet?			
3. Datensynthese			
3.1. Werden alle Quellen für die verwendeten Daten ausreichend beschrieben? 3.2. Werden die Datenauswahl und damit verbundene Annahmen ausreichend begründet? 3.3. Ist der Prozess der Datensynthese transparent? 3.4. Wird bei einer Verwendung von Wahrscheinlichkeitsverteilungen die gewählte Verteilung begründet? 3.5. Wird bei einer Verwendung von Wahrscheinlichkeitsverteilungen die Unsicherheit zweiten Ranges berücksichtigt?			
4. Behandlung von Unsicherheit			
4.1. Werden die vier zentralen Unsicherheitstypen mittels Sensitivitätsanalysen behandelt? 4.2. Wenn nicht, wird die Entscheidung begründet?			
4a. Methodologische Unsicherheit			
Wird die Modellierung mit mehreren methodologischen Annahmen zur Behandlung methodologischer Unsicherheit durchgeführt?			
4b. Strukturelle Unsicherheit			
Wird strukturelle Unsicherheit in Sensitivitätsanalysen behandelt?			
4c. Heterogenität			
Wird das Modell mit unterschiedlichen Subgruppen gerechnet, um Heterogenität zu berücksichtigen?			
4d. Parameterunsicherheit			
4d.1. Werden geeignete Methoden zur Analyse von Parameterunsicherheit angewandt? 4d.2. Werden bei der Verwendung von Punktschätzern die für die Sensitivitätsanalyse herangezogenen Intervalle klar beschrieben und begründet?			
Validierung			
1. Interne Validierung			

Beurteilung der Qualität von entscheidungsanalytischen gesundheitsökonomischen Modellen	Ja	Nein	Nicht enthalten
Wurde die mathematische Logik des Modells getestet?			
2. Externe Validierung			
2.1. Werden kontra-intuitive Ergebnisse erklärt und begründet? 2.2. Werden bei einer Validierung mit externen Daten auftretende Unterschiede erklärt und begründet?			
2.3. Werden die Modellergebnisse mit früheren Modellen verglichen und Abweichungen erklärt?			

Quelle: (Philips, Ginnelly 2004); GÖG/BIQG-eigene Übersetzung, leichte Modifikation und Darstellung

Tabelle 15.7:

Evidenzsynthese zu Effektivität und Kosteneffektivität von HIV-Präventionsmaßnahmen bei unterschiedlichen Subgruppen

Interventions	FSW	MSM	IDU	SDC	Preg	PI	HCW	Young	G pop
I. Interventions that affect knowledge, attitude and beliefs and influence psychological and social correlates of risk									
Abstinence-only programmes	X	X	X	X	X	X	X	No	X
Abstinence-plus programmes	X	X	X	X	X	X	X	■	X
Community-based education	■	■	No	-	X	-	X	■	■
Mass media campaigns	X	X	X	X	X	X	X	X	■
School-based sex education programs (combined with life skills)	X	X	X	X	X	X	X	■	X
Routine (provider-initiated) voluntary HIV screening at health care settings	-	-	-	-	■	-	-	■	■
Voluntary HIV counselling and testing (VCT) (+ STI clinic and condom distribution)	■	■	■	■	■	■	-	■	■
Workplace-based education (+ condom distribution/free STI clinic)	■	X	X	X	X	X	-	X	■
II. Harm reduction interventions that lower the risk of a behaviour, but do not eliminate the behaviour									
Condom use (availability and accessibility)	■	■	■	■	X	-	X	■	■
Introduction of female condoms	■	-	-	-	X	-	X	-	-
Needle and syringe exchange	X	X	■	X	X	X	X	X	X
Needle social marketing	X	X	■	X	X	X	X	X	X
Street outreach	X	X	■	X	X	X	X	X	X
III. Biological/biomedical interventions that strive to reduce HIV infection and transmission risk									
HIV vaccine	-	-	No	-	X	-	-	-	-
Improved STI treatment services	■	■	■	■	X	-	X	■	■
Mass or community treatment of sexually transmitted infections	X	X	X	X	X	X	X	X	No
Male circumcision	X	-	-	-	X	-	X	-	■
Microbicides	No	X	-	-	X	-	X	-	-
Postexposure prophylaxis	-	-	No	-	X	-	No	X	X

Fortsetzung nächste Seite

Prevention of mother-to-child transmission of HIV	X	X	X	X	■	X	X	X	X
Screening blood products and donated organs for HIV	X	X	X	X	X	X	X	X	■
Substitution treatment	X	X	■	X	X	X	X	X	X
Nucleic acid test screening (NAT) of volunteer blood donations	X	X	X	X	X	X	X	X	■
IV. Mitigation of barriers to prevention and negative social outcomes of HIV infection									
Increased alcohol tax	X	X	X	X	X	X	X	X	■
Microfinance	X	X	X	X	X	X	X	X	No
Microfinance (combined with education)	X	X	X	X	X	X	X	X	■
FSW, female sex worker; MSM, men who have sex with men; IDU, injecting drug user; SDC, serodiscordant couples; Preg, pregnant women; PI, prison inmate; HCW, health care worker; Young, people aged 10–24 years old; G pop, general population.									
Legend to shading in the table									
Colours	Effectiveness	Cost-effectiveness	Description						
■	Yes	Yes	The intervention proven to be effective and cost-effective						
■	Yes	Data not available	The intervention proven to be effective but no evidence regarding cost-effectiveness						
■	Yes	No	The intervention proven to be effective but not cost-effective						
No	No	No, data not available	The intervention proven to be neither effective nor cost-effective						
–	Data not available	Data not available	No evidence concerning effectiveness or cost-effectiveness of the intervention						
X			The intervention is not relevant or not used for particular target subpopulation						

Quelle: Anderson und Shemilt 2010

Tabelle 15.8:

GDP-PPP (1980–2008), Nationale Wahrung in US \$ (Teil 1)

	Australien	sterreich	Belgien	Kanada	Tschechische Republik	Danemark	Finnland	Frankreich	Deutschland	Griechenland	Ungarn	Island	Irland
1980	0,99	0,96	0,90	1,15		7,66	0,77	0,85	1,11	0,08		5,82	0,62
1981	1,00	0,94	0,86	1,17		7,73	0,78	0,86	1,06	0,09		7,93	0,66
1982	1,05	0,93	0,87	1,20		8,02	0,81	0,91	1,04	0,11		11,49	0,72
1983	1,10	0,93	0,89	1,21		8,29	0,84	0,96	1,03	0,13		19,60	0,76
1984	1,12	0,94	0,90	1,21		8,47	0,88	0,99	1,01	0,15		23,97	0,78
1985	1,14	0,94	0,91	1,21		8,57	0,90	1,02	1,01	0,18		30,71	0,80
1986	1,19	0,94	0,92	1,22		8,60	0,92	1,05	1,01	0,21		37,43	0,83
1987	1,25	0,94	0,91	1,24		8,77	0,93	1,05	1,00	0,23		43,69	0,83
1988	1,32	0,92	0,90	1,25		8,82	0,97	1,04	0,98	0,26		52,01	0,83
1989	1,37	0,92	0,91	1,26		8,92	1,00	1,04	0,97	0,29		61,57	0,84
1990	1,38	0,91	0,90	1,25	5,39	8,84	1,01	1,03	0,97	0,34		68,27	0,80
1991	1,36	0,91	0,89	1,25	7,09	8,77	1,00	1,02	0,96	0,39	30,09	71,48	0,79
1992	1,35	0,92	0,90	1,23	7,79	8,71	0,98	1,01	0,99	0,44	35,73	72,28	0,79
1993	1,34	0,93	0,92	1,22	9,22	8,58	0,98	1,01	1,00	0,49	42,36	71,95	0,82
1994	1,32	0,93	0,92	1,21	10,24	8,53	0,97	1,00	1,01	0,53	49,58	72,30	0,81
1995	1,32	0,93	0,91	1,21	11,06	8,46	1,00	0,99	1,00	0,57	61,57	72,98	0,82
1996	1,32	0,93	0,91	1,21	11,94	8,44	1,00	0,99	0,99	0,60	73,05	74,88	0,83
1997	1,32	0,92	0,91	1,21	12,70	8,43	1,00	0,97	0,99	0,63	84,97	74,41	0,85
1998	1,31	0,92	0,92	1,19	13,89	8,39	1,00	0,97	0,99	0,66	94,14	77,21	0,88
1999	1,30	0,92	0,92	1,19	14,14	8,47	1,00	0,96	0,97	0,68	101,07	79,68	0,93
2000	1,31	0,90	0,89	1,23	14,23	8,42	1,00	0,94	0,97	0,68	108,03	84,42	0,96
2001	1,33	0,92	0,89	1,22	14,22	8,47	1,01	0,92	0,96	0,67	110,65	88,92	0,99
2002	1,34	0,90	0,87	1,23	14,32	8,30	1,00	0,90	0,94	0,66	114,88	91,34	1,00
2003	1,35	0,88	0,88	1,23	14,03	8,54	1,01	0,94	0,92	0,69	120,52	94,49	1,01
2004	1,37	0,87	0,90	1,23	14,28	8,40	0,97	0,94	0,90	0,70	126,21	94,18	1,01
2005	1,39	0,89	0,90	1,21	14,32	8,59	0,98	0,92	0,87	0,71	128,59	99,08	1,01
2006	1,41	0,88	0,90	1,21	14,26	8,59	0,97	0,92	0,86	0,72	130,98	109,33	1,00
2007	1,42	0,88	0,89	1,21	14,23	8,60	0,98	0,91	0,86	0,72	134,78	115,10	0,97
2008	1,47	0,88	0,89	1,22	14,25	8,70	0,99	0,91	0,85	0,73	139,58	123,09	0,94

Quelle: OECD (OECD 2010)

Tabelle 15.9:

GDP-PPP (1980–2008), Nationale Wahrung in US \$ (Teil 2)

	Italien	Japan	Korea	Luxemburg	Mexiko	Niederlande	Neuseeland	Norwegen	Polen	Portugal	Slowakische Republik	Spanien	Schweden
1980	0,39	228,80	384,77	0,91	0,02	1,18	0,86	8,02		0,15		0,38	6,33
1981	0,42	218,43	415,89	0,89	0,02	1,13	0,93	8,28		0,16		0,39	6,34
1982	0,47	210,29	418,15	0,93	0,03	1,12	0,95	8,62		0,19		0,42	6,46
1983	0,52	207,07	427,23	0,96	0,05	1,10	0,99	8,87		0,22		0,45	6,85
1984	0,55	205,95	436,49	0,96	0,08	1,07	1,03	9,08		0,27		0,48	7,10
1985	0,59	204,44	443,68	0,96	0,12	1,06	1,14	9,26		0,32		0,50	7,34
1986	0,62	203,32	457,91	0,94	0,21	1,03	1,31	9,00		0,37		0,55	7,65
1987	0,64	198,42	470,79	0,92	0,49	0,99	1,43	9,39		0,40		0,56	7,80
1988	0,66	193,30	489,80	0,91	0,95	0,97	1,52	9,52		0,43		0,58	8,02
1989	0,67	190,53	498,96	0,91	1,16	0,94	1,53	9,69		0,46		0,60	8,34
1990	0,70	187,85	530,91	0,90	1,43	0,92	1,52	9,68	0,27	0,50		0,62	8,73
1991	0,73	186,84	567,69	0,89	1,71	0,92	1,48	9,56	0,40	0,53		0,64	9,20
1992	0,74	185,61	597,20	0,90	1,91	0,92	1,47	9,28	0,55	0,58	9,65	0,66	9,08
1993	0,76	182,40	620,86	0,93	2,04	0,91	1,47	9,28	0,70	0,61	10,89	0,68	9,16
1994	0,77	178,80	655,67	0,94	2,17	0,91	1,46	9,07	0,94	0,64	12,10	0,69	9,21
1995	0,79	174,34	690,04	0,95	2,93	0,91	1,46	9,16	1,18	0,65	13,03	0,71	9,36
1996	0,81	170,12	711,81	0,95	3,76	0,91	1,47	9,04	1,36	0,66	13,34	0,72	9,24
1997	0,82	168,32	732,44	0,96	4,35	0,91	1,45	9,08	1,52	0,67	13,70	0,72	9,30
1998	0,81	166,52	766,56	0,95	4,96	0,91	1,45	9,38	1,66	0,69	14,16	0,72	9,37
1999	0,82	162,04	754,89	0,94	5,63	0,91	1,43	9,33	1,74	0,70	15,08	0,73	9,29
2000	0,82	154,94	748,82	0,94	6,11	0,89	1,45	9,14	1,84	0,70	15,86	0,74	9,15
2001	0,81	149,44	757,08	0,95	6,31	0,91	1,47	9,18	1,86	0,71	15,72	0,74	9,35
2002	0,85	143,77	769,77	0,93	6,55	0,90	1,47	9,11	1,83	0,71	15,90	0,73	9,35
2003	0,85	139,68	795,67	0,94	6,82	0,93	1,50	9,11	1,84	0,71	16,72	0,75	9,34
2004	0,87	134,32	794,27	0,92	7,22	0,91	1,51	8,98	1,86	0,72	17,24	0,76	9,10
2005	0,87	129,55	788,92	0,95	7,13	0,90	1,54	8,90	1,87	0,68	17,05	0,76	9,38
2006	0,86	124,33	760,67	0,95	7,38	0,89	1,52	8,89	1,87	0,68	17,08	0,76	9,27
2007	0,85	120,14	749,88	0,95	7,53	0,88	1,55	9,05	1,92	0,67	17,10	0,74	9,14
2008	0,87	116,32	759,72	0,94	7,86	0,88	1,57	9,40	1,94	0,67	17,55	0,75	9,27

Quelle: OECD (OECD 2010)

Tabelle 15.10:
GDP-PPP (1980–2008), Nationale Währung in US \$ (Teil 3)

	Schweiz	Türkei	Großbritannien	Vereinigte Staaten
1980	2,12	5,47	0,49	1,00
1981	2,05	7,20	0,50	1,00
1982	2,06	8,71	0,51	1,00
1983	2,03		0,51	1,00
1984	2,02		0,52	1,00
1985	2,01		0,53	1,00
1986	2,02		0,54	1,00
1987	2,01		0,55	1,00
1988	2,00		0,57	1,00
1989	1,98		0,59	1,00
1990	1,99		0,61	1,00
1991	2,02		0,63	1,00
1992	2,02		0,64	1,00
1993	2,02	0,01	0,64	1,00
1994	2,00	0,01	0,64	1,00
1995	1,98	0,02	0,64	1,00
1996	1,94	0,04	0,64	1,00
1997	1,89	0,08	0,63	1,00
1998	1,88	0,13	0,64	1,00
1999	1,87	0,20	0,65	1,00
2000	1,85	0,28	0,64	1,00
2001	1,84	0,43	0,63	1,00
2002	1,77	0,61	0,63	1,00
2003	1,78	0,77	0,64	1,00
2004	1,75	0,81	0,63	1,00
2005	1,74	0,83	0,64	1,00
2006	1,69	0,86	0,64	1,00
2007	1,66	0,89	0,65	1,00
2008	1,67	0,98	0,65	1,00

Quelle: OECD (OECD 2010)

Tabelle 15.11:

Bruttoinlandsprodukt – Preis Index (2000 = 100); Teil 1

	Australien	Österreich	Belgien	Kanada	Tschechische Republik	Dänemark	Finnland	Frankreich	Deutschland	Griechenland	Ungarn	Island
1980	41,2	58,7	54,4	51,4		48,1	41,4	47,6	64,2	6,8		3,9
1981	46,0	62,6	57,2	56,9		53,2	46,2	53,1	66,9	8,2		5,9
1982	50,7	65,9	61,5	61,8		58,5	50,4	59,4	70,0	10,5		9,0
1983	54,7	68,2	65,0	65,1		62,9	54,6	65,1	71,9	12,6		16,0
1984	57,4	71,6	68,5	67,2		66,6	59,2	69,8	73,4	15,4		20,3
1985	60,6	73,7	71,7	69,3		69,5	62,4	73,7	74,9	18,4		26,8
1986	64,8	75,9	73,7	71,4		71,3	65,4	77,6	77,2	21,8		33,4
1987	70,4	77,7	74,9	74,7		74,8	68,3	79,7	78,2	25,1		40,1
1988	77,0	79,2	76,6	78,1		77,7	73,6	82,2	79,5	29,3		49,4
1989	81,3	81,7	80,2	81,6		81,5	78,2	84,9	81,8	33,6		60,7
1990	84,2	84,2	82,5	84,2	31,2	84,0	82,7	87,1	84,6	40,5		69,9
1991	85,8	87,4	84,9	86,7	42,4	86,2	84,2	89,4	87,2	48,6	23,3	75,7
1992	86,9	90,6	87,8	87,9	47,7	87,6	84,8	91,3	91,5	55,7	28,3	78,3
1993	87,7	93,1	91,3	89,1	57,7	88,2	86,5	92,7	94,9	63,8	34,3	79,8
1994	89,0	95,6	93,2	90,1	65,4	89,6	87,6	93,9	97,2	70,9	40,9	81,8
1995	91,0	97,5	94,3	92,2	72,1	90,7	91,8	95,2	99,0	77,9	51,9	84,3
1996	92,2	98,4	94,8	93,7	79,6	92,5	91,6	96,7	99,5	83,6	62,9	86,4
1997	93,4	98,1	95,9	94,8	86,2	94,4	93,5	97,7	99,8	89,2	74,5	88,9
1998	93,4	98,5	97,8	94,4	95,8	95,5	96,7	98,6	100,3	93,9	83,9	93,4
1999	95,4	98,9	98,2	96,0	98,5	97,1	97,5	98,6	100,7	96,7	91,0	96,5
2000	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
2001	102,8	101,9	102,0	101,1	104,9	102,5	103,0	102,0	101,2	103,1	108,4	108,6
2002	105,9	103,2	103,9	102,2	107,8	104,9	104,3	104,4	102,6	106,6	116,9	114,8
2003	109,6	104,5	105,6	105,6	108,8	106,6	103,8	106,4	103,9	110,5	123,7	115,5
2004	113,7	106,2	108,1	109,0	113,8	109,1	104,5	108,1	104,9	114,2	129,1	118,4
2005	119,0	108,4	110,8	112,6	113,4	112,2	105,0	110,3	105,6	118,1	132,0	121,7
2006	124,5	110,4	113,3	115,4	114,5	114,4	106,4	113,1	106,1	121,9	137,1	132,6
2007	130,0	112,7	116,0	119,0	118,6	116,7	109,8	115,8	108,1	125,4	145,0	139,9
2008		115,4	118,0				112,8	118,4	109,7	129,7	150,2	

Quelle: OECD (OECD 2010)

Tabelle 15.12:

Bruttoinlandsprodukt – Preis Index (2000 = 100); Teil 2

	Irland	Italien	Japan	Korea	Luxemburg	Mexiko	Niederlande	Neuseeland	Norwegen	Polen	Portugal	Slowakische Republik
1980	35,2	25,4	79,3	27,9	52,1	0,2	67,1	32,4	39,2		11,8	
1981	41,3	30,1	82,8	33,0	55,9	0,2	70,7	38,2	44,3		13,8	
1982	47,6	35,4	84,6	35,2	61,9	0,3	74,4	41,3	48,9		16,7	
1983	52,7	40,7	86,6	37,4	66,1	0,6	75,8	44,8	52,3		20,8	
1984	56,0	45,1	89,3	39,6	69,0	0,9	76,6	48,2	55,5		25,9	
1985	9,0	49,3	91,4	41,5	71,2	1,4	77,9	55,1	58,4		31,6	
1986	62,8	53,0	92,9	43,8	71,1	2,4	77,6	65,0	58,0		38,0	
1987	64,2	56,1	93,2	46,3	71,2	5,9	76,7	72,8	62,2		41,9	
1988	66,3	59,9	93,9	49,8	73,1	11,7	77,4	78,3	65,2		46,5	
1989	69,9	63,6	96,1	52,7	76,1	14,8	78,4	82,4	68,9		51,4	
1990	69,4	68,9	98,4	58,2	78,0	18,9	79,6	84,4	71,5	11,9	58,2	
1991	70,7	74,1	101,3	64,4	79,4	23,3	82,1	85,4	73,1	18,5	64,1	
1992	72,7	77,4	102,9	69,3	82,4	26,7	84,2	87,1	72,6	25,7	71,4	50,9
1993	76,4	80,4	103,5	73,7	87,3	29,2	85,5	88,6	74,2	33,6	76,7	58,7
1994	77,7	83,3	103,6	79,5	90,4	31,7	87,3	90,2	74,1	46,1	82,2	66,6
1995	80,1	87,4	103,0	85,4	92,5	43,6	89,1	91,9	76,3	59,0	85,1	73,2
1996	81,9	91,6	102,5	89,8	95,2	57,0	90,2	93,2	79,5	69,5	87,3	76,3
1997	85,0	93,9	103,1	93,9	93,5	67,1	92,6	95,0	81,7	79,2	90,6	81,0
1998	90,6	96,4	103,1	99,4	93,1	77,5	94,4	96,2	81,1	88,0	94,0	85,1
1999	94,3	98,1	101,8	99,3	98,0	89,2	96,0	96,8	86,5	93,2	97,1	91,4
2000	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
2001	105,5	103,0	98,8	103,5	100,1	105,8	105,1	103,8	101,7	103,5	103,7	105,0
2002	110,3	106,3	97,2	106,5	102,2	113,2	109,1	104,0	99,9	105,8	107,8	109,1
2003	113,0	109,6	95,7	109,4	108,4	122,9	111,5	106,5	102,9	106,2	111,2	114,9
2004	115,3	112,5	94,7	112,3	110,4	133,8	112,3	110,0	108,3	110,6	113,9	121,6
2005	118,0	114,8	93,5	112,1	115,3	139,4	115,0	112,4	117,7	113,5	116,8	124,5
2006	122,1	117,0	92,7	111,6	121,5	149,0	117,0	116,0	127,7	115,2	120,1	128,2
2007	123,8	119,8	92,0	112,9	123,6	156,0	118,8	120,4	130,6	119,7	123,6	129,6
2008		123,2					122,2		142,7	123,1		133,4

Quelle: OECD (OECD 2010)

Tabelle 15.13:

Bruttoinlandsprodukt – Preis Index (2000 = 100); Teil 3

	Spanien	Schweden	Schweiz	Türkei	Großbritannien	Vereinigte Staaten
1980	27,2	37,6	61,6	0,0	40,1	54,0
1981	30,6	41,2	65,0	0,0	44,7	59,1
1982	34,7	44,5	69,3	0,0	48,0	62,7
1983	38,8	49,1	71,1	0,0	50,6	65,1
1984	43,0	52,8	73,6	0,0	52,9	67,6
1985	46,7	56,2	75,2	0,1	56,0	69,7
1986	51,8	59,9	77,5	0,1	57,9	71,2
1987	54,9	62,8	79,2	0,1	61,0	73,2
1988	58,2	66,8	81,3	0,2	64,9	75,7
1989	62,2	72,1	83,7	0,3	69,6	78,6
1990	66,7	78,4	87,2	0,5	75,0	81,6
1991	71,4	85,4	91,9	0,7	79,8	84,5
1992	76,2	86,3	93,8	1,2	82,9	86,4
1993	79,6	89,1	96,0	2,0	85,2	88,4
1994	82,7	91,4	97,2	4,1	86,6	90,3
1995	86,8	94,8	97,9	7,7	88,9	92,1
1996	89,8	95,6	98,1	13,6	92,1	93,9
1997	91,9	97,1	98,0	24,7	94,7	95,4
1998	94,2	97,7	98,3	43,5	96,8	96,5
1999	96,7	98,6	98,9	67,0	98,8	97,9
2000	100,0	100,0	100,0	100,0	100,0	100,0
2001	104,2	102,3	100,8	152,9	102,1	102,4
2002	108,7	104,0	101,3	210,1	105,3	104,2
2003	113,2	106,0	102,3	258,9	108,5	106,4
2004	117,7	106,2	102,9	291,0	111,2	109,5
2005	122,8	107,2	103,0	311,7	113,7	113,1
2006	127,8	109,0	104,7	340,7	116,7	116,7
2007	131,8	112,3	106,6	368,4	120,1	119,8
2008	135,8	116,0				

Quelle: OECD (OECD 2010)

15.8 Beispiel für eine Budgetauswirkungsanalyse

Berechnung der Budgetauswirkung bei Einführung einer neuen Impfung in das **Impfprogramm zur Virusimmunisierung bei** Kindern bis 5 Jahre aus Sicht der Gesellschaft, des Gesundheitswesens† (Staat + Patient) sowie des öffentlichen Gesundheitssystems‡ (Gesundheitsministerium und pharmazeutische Zulassungsstelle)

† Gesundheitsministerium, pharmazeutische Zulassungsstelle, und direkte Kosten der Betreuungspersonen inklusive Krankenhausaufenthalte, Inanspruchnahme der Notaufnahme, Konsultation Allgemeinmediziner/innen und Zuschüsse für Behandlung der virusbedingten Erkrankung

‡ Gesundheitsministerium und pharmazeutische Zulassungsstelle

Tabelle 15.14:
Beispiel zur Berechnung der Budgetauswirkung (Angabe)

Kosten	Parameter
Anschaffungskosten pro Impfdosis	\$ 50
Krankenhaus	
Krankenhausaufenthalt	\$ 1.889
Krankenhausaufenthalt (DRG basierte Kosten)	1.512
Inanspruchnahme der Notaufnahme	585
Konsultation Allgemeinmediziner/in (Kosten für Staat)	39,70
Tageseinkommen der Betreuungsperson	59,36
Zuschüsse für med. Behandlung von virusbedingter Erkrankung	\$ 2,86
Erwerbsausfall der Betreuungsperson	
Krankenhausaufenthalt	\$ 245
Inanspruchnahme der Notaufnahme	\$ 105
Allgemeinmediziner/in	\$ 101
Transport	
Krankenhausaufnahme	\$ 8,00
Inanspruchnahme der Notaufnahme	\$ 4,00
Wahrscheinlichkeiten, Anteile, Verhältnisse, Proportionen	
Effizienz der Impfung (drei Dosen)	
Krankenhausaufenthalt	0,958
Inanspruchnahme der Notaufnahme	0,937
Konsultation Allgemeinmediziner/in	0,860
Einzeldosis (Krankenhausaufenthalte, Notaufnahme)	0,83
Zwei Dosen (Krankenhausaufenthalte, Notaufnahme)	0,81
Einzeldosis (Allgemeinmediziner/in und Arbeitsausfall)	0,75
Zwei Dosen (Allgemeinmediziner/in und Arbeitsausfall)	0,73
Durchimpfungsrate‡	
Einzeldosis	0,910
Zwei Dosen	0,890
Drei Dosen	0,846
Inzidenzrate (Aufnahme) siehe Tabelle 2	
Letalität pro 1.000 Aufnahmen	
Alter < 12 Monate	1,37
Alter 12–23 Monate	0,48
Alter 24–35 Monate	0,56
Alter >35 Monate	0
Med. Behandlung von virusbedingter Erkrankung (Proportion PatientInnen der Notaufnahme zu PatientInnen der AllgemeinmedizinerInnen)	0,5
Verhältnis von Notaufnahme zu Krankenhausaufenthalte bei virusbedingter Erkrankung (siehe Tabelle 2)	2,05
Verhältnis von Konsultation Allgemeinmediziner/in zu Krankenhausaufenthalte (siehe Tabelle 2)	7,8

Kosten	Parameter
Proportion von Krankenhausfällen (überwiesen von Allgemeinmediziner/in)	0,7
Rückgang des Impfschutzes pro Jahr	
Jahre 3-5	0
Proportion der Fälle von virusbedingter Erkrankung verursacht durch Virus; Alter 3-4 Jahre	0,45
Andere Parameter	
Geburtskohorten 2009-2020	64.140
Zeithorizont	10 Jahre
Negativnutzen pro Virus Fall	
Kind	0,0022
Betreuungsperson	0,0018
Diskontierungsrate (für Lebensjahre)	0,035
Arbeitsausfall Betreuungsperson	
Krankenhausaufenthalt	3,6
Inanspruchnahme der Notaufnahme	1,6
Allgemeinmediziner/in	1,5

Quelle und Darstellung: GÖG/BIQG

Tabelle 15.15:

Voraussichtliche jährliche Fälle im Jahr 2009 und Inzidenzraten virusbedingter Erkrankungen bei Kindern unter 5 Jahren (Berechnung)

Alter (Jahre)	Fälle pro Jahr				Jährliche Inzidenzrate pro 100.000			
	Krankenhausaufenthalte	Notaufnahme	Allgemeinmediziner/in	Gesamt	Krankenhausaufenthalte	Notaufnahme	Allgemeinmediziner/in	Gesamt
< 1	466	956	3,135	4.557	727	1.491	4.887	7.105
1-2	549	1.125	3,690	5.364	861	1.765	5.788	8.414
2-3	253	519	1,703	2.476	398	815	2.673	3.885
3-4	143	292	958	1.393	224	459	1.504	2.186
4-5	94	193	634	922	148	304	996	1.448
> 5 Jahre	1.506	3.086	10.120	14.712	476	977	3.202	4.655
5 Jahresrisiko	1:43	1:21	1:7	1:5				

Quelle und Darstellung: GÖG/BIQG

Tabelle 15.16:

Voraussichtliche Kosten für die Gesellschaft im Jahr 2009 für die Behandlung virusbedingter Erkrankungen bei Kindern unter 5 Jahren ohne Einführung eines Impfprogramms zur Virusimmunisierung (Berechnung)

	Kosten für Land (in Millionen \$)	Kosten pro Kind unter 5 Jahre (in \$)	Kostenverteilung (in %)
Krankenhausaufenthalte	2,89	9,07	40,9%
Inanspruchnahme der Notaufnahme, kein stationärer Krankenhaus- aufenthalt	1,81	5,66	25,5%
Einkommen der Betreuungsperson	1,80	5,64	25,4%
Arztbesuche, staatliche Vergütung*	0,53	1,66	7,5%
Transport zum Krankenhaus	0,02	0,08	0,3%
Zuschüsse für Behandlung virusbedingte Erkrankung	0,02	0,06	0,3%
Gesamt	7,07	22,17	100,0%

Quelle und Darstellung: GÖG/BIQG

Tabelle 15.17:

Voraussichtliche Fälle in fünf Jahren bei Einführung eines Impfprogrammes zur Virusimmunisierung im Vergleich zur Nichteinführung* (Berechnung)

	Keine Impfung	Impfung	Verhinderte Krankheitsfälle	NNV (number needing to be vaccinated)
Krankenhausaufenthalte	1.506	314	1.191	49
Inanspruchnahme der Notaufnahme	3.086	644	2.442	24
Konsultation Allgemeinmediziner/in†	13.335	3.543	9.792	6
Todesfälle	1,0	0,2	0,8	73.357

* keine Diskontierung

† inklusive 2.543 Fälle, die von Allgemeinmediziner/innen in ein Krankenhaus eingewiesen wurden

NNV = number of children needing to be vaccinated each year to prevent a single health-care encounter of the type specified / Anzahl an Kindern, die geimpft werden müssen um einen Krankenhausaufenthalt/die Inanspruchnahme der Notaufnahme/die Konsultation Allgemeinmediziner/in oder einen Todesfall zu verhindern.

Quelle und Darstellung: GÖG/BIQG

Tabelle 15.18:

Voraussichtliche Budgetauswirkung in fünf Jahren bei Einführung eines Impfprogramms zur Virusimmunisierung* (Berechnung)

	Keine Impfung (in Millionen \$)	Impfung (in Millionen \$)	Inkrementelle Kosten (in Millionen \$)	Break-even Impfung Preis pro Dosis (in \$)
Kosten für Gesellschaft	7,07	10,06	2,99	32,39
Kosten für Gesundheitswesen †	5,27	9,62	4,35	24,40
Kosten für das öffentliche Gesund- heitssystem ‡	4,72	9,50	4,78	21,82

*alle Kosten wurden nicht diskontiert, da sie im gleichen Jahr angefallen sind

† Gesundheitsministerium, pharmazeutische Zulassungsstelle, und direkte Kosten der Betreuungspersonen inklusive Krankenhausaufenthalte, Inanspruchnahme der Notaufnahme, Konsultation Allgemeinmediziner/innen und Zuschüsse für Behandlung der virusbedingten Erkrankung

‡ Gesundheitsministerium und pharmazeutische Zulassungsstelle

Quelle und Darstellung: GÖG/BIQG

