



IQWiG-Berichte – Nr. 863

**Konzepte zur Generierung
versorgungsnaher Daten und
deren Auswertung zum
Zwecke der Nutzenbewertung
von Arzneimitteln nach § 35a
SGB V**

Rapid Report

Auftrag: A19-43

Version: 1.0

Stand: 10.01.2020

Impressum

Herausgeber:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema:

Konzepte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V

Auftraggeber:

Gemeinsamer Bundesausschuss

Datum des Auftrags:

02.05.2019

Interne Auftragsnummer:

A19-43

Anschrift des Herausgebers:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Im Mediapark 8

50670 Köln

Tel.: +49 221 35685-0

Fax: +49 221 35685-1

E-Mail: berichte@iqwig.de

Internet: www.iqwig.de

ISSN: 1864-2500

Mitarbeiterinnen und Mitarbeiter des IQWiG

- Sascha Abbas
- Ralf Bender
- Raphaela Gorris
- Elke Hausner
- Katharina Hirsch
- Thomas Kaiser
- Stefan Lange
- Jörg Lauterberg
- Guido Skipka
- Beate Wieseler
- Jürgen Windeler

Schlagwörter: Produktüberwachung nach Markteinführung, Register, Nutzenbewertung, Konzept

Keywords: Product Surveillance – Postmarketing, Registries, Benefit Assessment, Concept

Inhaltsverzeichnis

	Seite
Tabellenverzeichnis	v
Abbildungsverzeichnis	vi
Abkürzungsverzeichnis.....	vii
Kurzfassung	ix
1 Hintergrund	1
2 Fragestellung	2
3 Projektverlauf	3
4 Methoden	4
4.1 Informationsbeschaffung und -bewertung	4
4.2 Ermittlung von Qualitätskriterien für Patientenregister	5
5 Ergebnisse.....	9
5.1 Studiendesigns und Instrumente zur Datenerhebung zur Generierung versorgungsnaher Daten.....	9
5.2 Voraussetzungen für eine Nutzenbewertung.....	11
5.3 Versorgungsnahe Daten in der Nutzenbewertung.....	12
5.3.1 Formulierung der Fragestellung und Entscheidung für ein Studiendesign	14
5.3.2 Studienplanung.....	16
5.3.3 Datenerhebung.....	19
5.3.4 Analyse und Interpretation der Ergebnisse	21
5.4 Instrumente zur Datenerhebung	22
5.4.1 Studienindividuelle Datenerhebung	22
5.4.2 Studien auf Basis einer Datenerhebung in einem Register	23
5.4.3 Studien auf Basis einer Datenerhebung in elektronischen Patientenakten.....	23
5.4.4 Studien auf Basis einer Datenerhebung aus Abrechnungsdaten von Krankenkassen.....	30
5.5 Anforderungen an Daten und Auswertungen	38
5.5.1 Kriterien für Datenqualität	38
5.5.1.1 Kriterien für die Qualität von Registern gemäß nationaler und internationaler Standards	38
5.5.1.2 Ergebnisse der Gespräche mit Registerbetreibern.....	42
5.5.1.3 Kriterien für die Datenqualität bei der Generierung versorgungsnaher Daten zum Zwecke der Nutzenbewertung und Maßnahmen zur Sicherstellung der Datenqualität.....	51
5.5.2 Methodische Anforderungen an die Auswertung.....	54

5.5.3 Anforderungen an die Berichterstattung	61
5.6 Bewertung der Konzepte der Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung nach § 35a SGB V	63
5.7 Vorschläge für ein Vorgehen für die anwendungsbegleitende Datenerhebung..	66
5.7.1 Definition der Fragestellungen für die anwendungsbegleitende Datenerhebung.....	66
5.7.2 Evidenzlücken in bisherigen Nutzenbewertungen am Beispiel „Orphan Drugs“	67
5.7.3 Vorschläge für den Ablauf der anwendungsbegleitenden Datenerhebung im Verfahren der Nutzenbewertung nach § 35a SGB V	75
6 Diskussion.....	78
6.1 Studiendesigns für die Nutzung versorgungsnaher Daten.....	78
6.2 Allgemeine Handlungsempfehlungen auf Basis der Ergebnisse der Registergespräche	81
7 Fazit.....	86
8 Literatur	88
Anhang A – Fragebogen für die Interviews zu Patientenregistern	101
Anhang B – Diskussionsaspekte für den Expertenworkshop mit biometrischen Sachverständigen.....	109
Anhang C – Qualitätskriterien für Register und Registerstudien in der Literatur.....	112
Anhang D – Fragestellungen und Beschlüsse zu Orphan Drugs mit Marktzugang 2014 bis 2018	117

Tabellenverzeichnis

	Seite
Tabelle 1: Übersicht genereller und spezifischer Anforderungen für die einzelnen Schritte der Generierung versorgungsnaher Daten zum Zwecke der Nutzenbewertung in Abhängigkeit vom Studientyp.....	xiii
Tabelle 2: Kriterien für die Datenqualität und die Sicherstellung der Datenqualität für versorgungsnahe Datenerhebungen zum Zwecke der Nutzenbewertung von Arzneimitteln ..	xvi
Tabelle 3: Prozessschritte für eine anwendungsbegleitende Datenerhebung zum Zwecke der Nutzenbewertung nach § 35a	xxi
Tabelle 4: Interviews mit Registerbetreibern und biometrischen Sachverständigen	6
Tabelle 5: Komponenten der Nachbildung der Zielstudie aus einem nicht randomisierten Datensatz	18
Tabelle 6: Möglichkeiten und Einschränkungen von GKV-Routinedaten zur Beantwortung von Fragestellungen der Nutzenbewertungen	37
Tabelle 7: Zusammenstellung national und international verwendeter Qualitätskriterien für Patientenregister und registerbasierte Studien.....	40
Tabelle 8: Kriterien für die Datenqualität und die Sicherstellung der Datenqualität für versorgungsnahe Datenerhebungen zum Zwecke der Nutzenbewertung von Arzneimitteln ..	53
Tabelle 9: Definition der Fragestellung für eine anwendungsbegleitende Datenerhebung zum Zwecke der Nutzenbewertung nach § 35a SGB V	66
Tabelle 10: Größe der Zielpopulation und Datengrundlage für die Nutzenbewertung – G-BA-Beschlüsse zu Orphan Drugs mit Marktzugang 2014 bis 2018	70
Tabelle 11: Darstellung der Evidenzlücken je Endpunkt kategorie – Orphan Drugs mit Marktzugang 2014 bis 2018, Fragestellungen mit nicht quantifizierbarem Zusatznutzen	73
Tabelle 12: Anzahl der Endpunkt kategorien mit Evidenzlücke – Orphan Drugs mit Marktzugang 2014 bis 2018, Fragestellungen mit nicht quantifizierbarem Zusatznutzen	74
Tabelle 13: Prozessschritte für eine anwendungsbegleitende Datenerhebung zum Zwecke der Nutzenbewertung nach § 35a	76
Tabelle 14: Abgleich von Qualitätskriterien für Register und Registerstudien in der Literatur	112
Tabelle 15: Orphan Drug Bewertungen des G-BA mit Marktzugang 2014 bis 2018.....	117

Abbildungsverzeichnis

	Seite
Abbildung 1: Studiendesigns und Instrumente zur Datenerhebung zur Generierung versorgungsnaher Daten	xi
Abbildung 2: Studiendesigns und Instrumente zur Datenerhebung zur Generierung versorgungsnaher Daten	10
Abbildung 3: Ablauf von der Festlegung der Fragestellung der Nutzenbewertung bis zum Ergebnis der Untersuchung	13
Abbildung 4: Qualitätskriterien für Patientenregister und Registerstudien	39
Abbildung 5: Quantifizierung des Zusatznutzens und verfügbare Daten – Beschlüsse des G-BA zu Orphan Drugs bei Marktzugang 2014 bis 2018.....	69
Abbildung 6: Evidenzlücken in Fragestellungen mit nicht quantifizierbarem Zusatznutzen in der Übersicht - Beschlüsse des G-BA zu Orphan Drugs bei Marktzugang 2014 bis 2018 .	72

Abkürzungsverzeichnis

Abkürzung	Bedeutung
AHRQ	Agency for Healthcare Research and Quality
AMNOG	Arzneimittelmarktneuordnungsgesetz
AOK	Allgemeine Ortskrankenkasse
ATC	Anatomisch-Therapeutisch-Chemisch
CT	Computertomografie
CTT	Clinical Trial Transformation
DIMDI	Deutsches Institut für Medizinische Dokumentation und Information
DNVF	Deutsches Netzwerk Versorgungsforschung
DRG	Diagnosis Related Groups
EHR	Electronic health Record (elektronische Patientenakte)
EMA	European Medicines Agency (europäische Arzneimittel-Agentur)
EMR	Electronic medical Record (klinische Routinedokumentation in digitalen Akten)
EUnetHTA	European Network for Health Technology Assessment
FDA	Food and Drug Administration
G-BA	Gemeinsamer Bundesausschuss
GCP	Good clinical Practice
GKV	Gesetzliche Krankenversicherung
GSAV	Gesetz für mehr Sicherheit in der Arzneimittelversorgung
hdPS	Hoch-dimensionaler Propensity Score
HES	Hospital Episode Statistics (englische Krankenhausepisodenstatistik)
HTA	Health Technology Assessment
ICD	International Statistical Classification of Diseases and Related Health Problems (Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme)
ICH	International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
MedDRA	Medizinisches Wörterbuch für Aktivitäten im Rahmen der Arzneimittelzulassung
MDK	Medizinischer Dienst der Krankenkassen
NLP	Natural Language Processing

Abkürzung	Bedeutung
OECD	Organisation for Economic Co-operation and Development (Organisation für wirtschaftliche Zusammenarbeit und Entwicklung)
OPS	Operationen- und Prozedurenschlüssel
OTC	Over the Counter (rezeptfreies Arzneimittel)
PCORI	Patient-Centered Outcomes Research Institute
PRO	Patient reported Outcome (patientenberichteter Endpunkt)
PICO	Population, Intervention, Comparator; Outcomes
PKV	Private Krankenversicherung
RCT	Randomized controlled Trial (randomisierte kontrollierte Studie)
REQuEST	Registry Evaluation and Quality Standards Tool
SEER	Surveillance, Epidemiology and End Results
SGB	Sozialgesetzbuch
SVR	Sustained virologic Response (anhaltendes virologisches Ansprechen)
UE	Unerwünschtes Ereignis

Kurzfassung

Der Gemeinsame Bundesausschuss (G-BA) hat am 02.05.2019 das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) nach § 139 a, Abs. 3, SGB V mit einer wissenschaftlichen Ausarbeitung von Konzepten zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V beauftragt.

Fragestellung

Ziele der vorliegenden Ausarbeitung sind

- die Erstellung einer Übersicht möglicher Konzepte der Generierung und Auswertung versorgungsnaher Daten. Dabei sollen insbesondere auch Datenerhebungen berücksichtigt werden, die nicht der Kategorie „randomisierte kontrollierte Studie (RCT)“ zuzuordnen sind
- die Bewertung der identifizierten Konzepte der Datengenerierung und deren Auswertung hinsichtlich ihrer Eignung zur Beantwortung der Fragestellung der Nutzenbewertung nach § 35a SGB V insbesondere im Hinblick auf die Möglichkeit einer Quantifizierung des Zusatznutzens
- die Spezifizierung von Kriterien für die Qualität und methodische Anforderungen an die im Rahmen der jeweiligen Datengenerierung erhobenen Daten. Hierbei soll auch darauf eingegangen werden, welche Maßnahmen zur Sicherstellung der Datenqualität notwendig sind
- die Definition der Anforderungen an die Berichterstattung, die Aufbereitung und Struktur sowie die statistische Auswertung der Daten, die im Rahmen des jeweiligen Konzeptes der Datengenerierung erhoben werden.

Methoden

Informationsbeschaffung und -bewertung

Die Erarbeitung des Konzepts zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V wurde gemäß Projektskizze durch 3 Bausteine unterstützt:

- Empirische Information aus den Nutzenbewertungen von Arzneimitteln nach dem AMNOG (§ 35a SGB V)
- Orientierende Literaturrecherche für wissenschaftliche Fragestellungen, die sich als Teil der konzeptionellen Arbeit ergeben (z. B. zur Aussagekraft nicht randomisierter Studien abhängig von vorliegenden Datenkonstellationen)
- Interviews mit Expertinnen und Experten aus dem Bereich Register zu Kriterien für die Qualität und methodischen Anforderungen an die im Rahmen der jeweiligen Datengenerierung erhobenen Daten

Ermittlung von Qualitätskriterien für Patientenregister

Während der Bearbeitung des Projekts stellte sich heraus, dass als Instrumente zur Datenerhebung für die Generierung versorgungsnaher Daten zum Zwecke der Nutzenbewertung von Arzneimitteln neben einer studienindividuellen Datenerhebung primär Register infrage kommen. Die Spezifizierung von Kriterien für die Qualität der erhobenen Daten wurde deshalb auf Register beschränkt. Grundlage für die Beschreibung von Qualitätskriterien waren die oben genannten Interviews mit Expertinnen und Experten für Register und eine Zusammenstellung von Qualitätskriterien für Patientenregister aus nationalen und internationalen Empfehlungen.

Ergebnisse

Versorgungsnahe Daten zum Zwecke der Nutzenbewertung von Arzneimitteln sind wie folgt definiert:

- versorgungsnahe Daten werden bei den Patientenpopulationen erhoben, für die eine Indikation für das zu bewertende Arzneimittel im Rahmen der Zulassung besteht
- bei der Erhebung versorgungsnaher Daten erfolgt die Behandlung ohne spezifische Vorgaben

Da es bei der Bewertung von Arzneimitteln nach SGB V um die Versorgung von Patienten in Deutschland geht, müssen versorgungsnahe Daten die beiden oben genannten Kriterien so erfüllen, dass sie Rückschlüsse für die deutsche Versorgung zulassen.

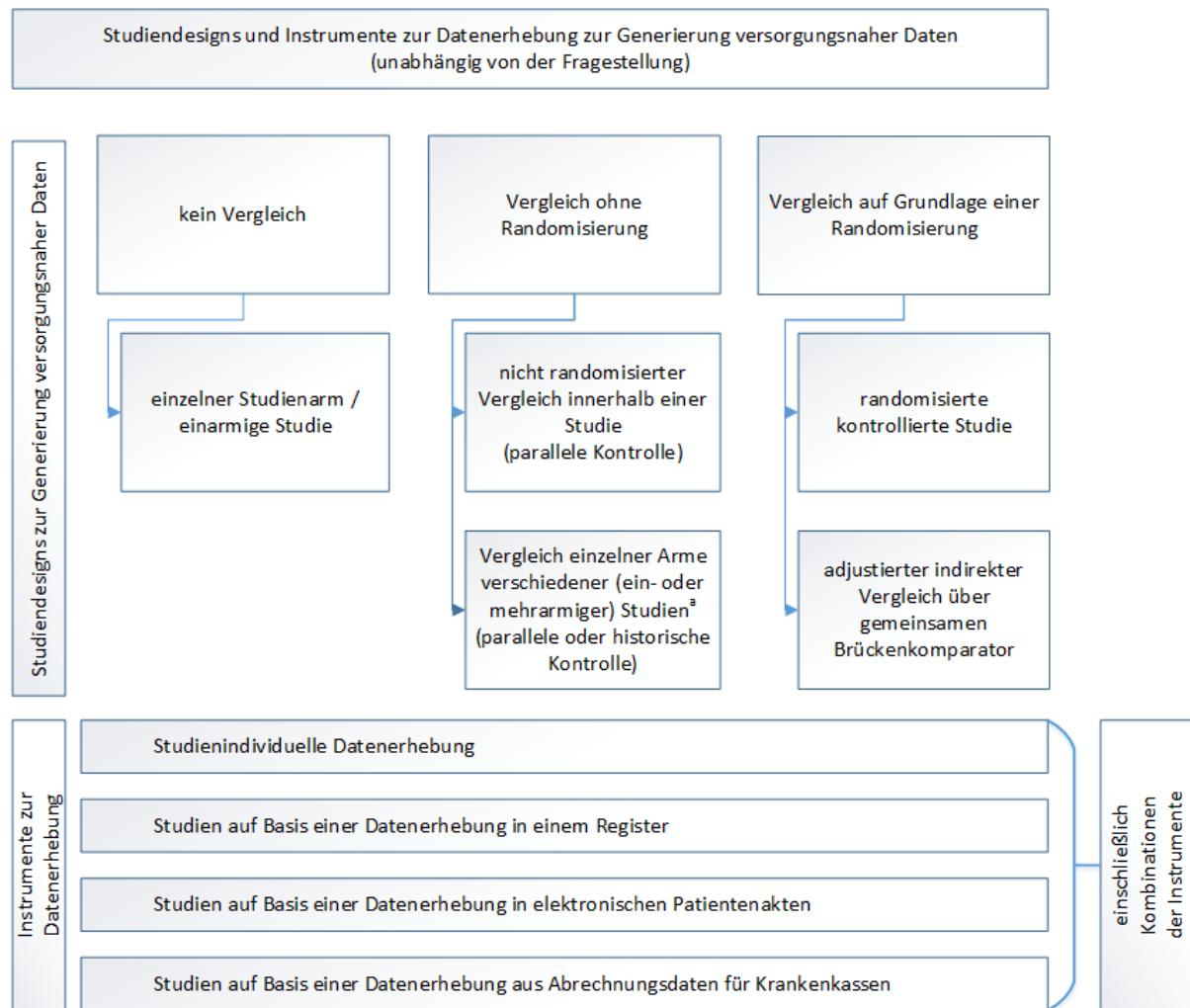
Die Definition versorgungsnaher Daten impliziert weder ein bestimmtes Studiendesign noch ein bestimmtes Instrument der Datenerhebung.

Das Ziel, Daten versorgungsnah zu erheben, verlangt nicht, dass die Datenerhebung sich auf solche Daten beschränkt, die im Versorgungsalltag per se erfasst werden. Eine derart falsch verstandene Beschränkung der Datenerhebung würde vielmehr das Ziel der Nutzenbewertung gefährden. Für die Nutzenbewertung sind regelhaft Daten notwendig, die im Behandlungsalltag nicht bei allen Patientinnen und Patienten dokumentiert werden (z. B. Daten zur gesundheitsbezogenen Lebensqualität, zu Symptomen oder Nebenwirkungen). Versorgungsnahe Daten müssen zudem ausreichend valide und strukturiert sein, um eine Nutzenbewertung zu ermöglichen.

Übersicht über Studiendesigns und Instrumente zur Datenerhebung mit dem Ziel der Generierung versorgungsnaher Daten

Die folgende Abbildung gibt eine Übersicht über Studiendesigns und Instrumente zur Datenerhebung, die zur Generierung versorgungsnaher Daten eingesetzt werden können. Der obere Teil der Abbildung beschreibt die grundsätzlich denkbaren Studiendesigns abhängig von der Art des möglichen Vergleichs von Interventionen. Der untere Teil benennt die Instrumente, die zur Erhebung versorgungsnaher Daten in Studien mit den verschiedenen Designs

herangezogen werden können. Dabei wird deutlich, dass die verschiedenen Instrumente der Datenerhebung generell für alle Studiendesigns eingesetzt werden können.



a: einschließlich Studien zum Spontanverlauf

Abbildung 1: Studiendesigns und Instrumente zur Datenerhebung zur Generierung versorgungsnaher Daten

Versorgungsnahe Daten in der Nutzenbewertung

Sollen versorgungsnahe Daten zum Zweck der Nutzenbewertung herangezogen werden, ist zu berücksichtigen, dass die Grundlage jeder Aussage über Effekte von Interventionen ein Vergleich ist. Nur auf Basis eines Vergleichs ist die für eine kausale Aussage notwendige Differenzierung zwischen „**nach** Intervention A“ und „**wegen** Intervention A“ möglich.

Aus diesen Überlegungen folgt, dass die alleinige Betrachtung einarmiger Studien oder einzelner Studienarme für die Nutzenbewertung nicht relevant ist. Damit entfällt für die weiteren Ausführungen der linke Strang der Übersicht von Studiendesigns in Abbildung 1, der

Designs ohne Vergleich darstellt. Für die Fragestellung der Nutzenbewertung sind ausschließlich vergleichende Studiendesigns relevant.

Abhängig vom vergleichenden Studiendesign, das für die Generierung versorgungsnaher Daten zum Zweck der Nutzenbewertung gewählt wird, entstehen unterschiedliche Anforderungen an die Durchführung und Auswertung der Studie. Tabelle 1 zeigt die Schritte von der Festlegung der Fragestellung der Nutzenbewertung bis zum Ergebnis der Untersuchung dieser Fragestellung und fasst die in diesem Ablauf bestehenden Anforderungen zusammen.

Während generelle wissenschaftliche Prinzipien wie die Formulierung der zu beantwortenden Fragestellung oder die Interpretation der Ergebnisse unter Berücksichtigung der erreichten Ergebnisqualität unabhängig vom gewählten Studiendesign vorgenommen werden, unterscheiden sich andere Schritte der Nutzenbewertung je nach Studiendesign. Denn abhängig vom Studiendesign kann ein fairer, kausal zu interpretierender Vergleich angenommen werden oder ist durch spezifische Schritte in der Studienplanung, der Datenerhebung und der Auswertung anzunähern.

Tabelle 1: Übersicht genereller und spezifischer Anforderungen für die einzelnen Schritte der Generierung versorgungsnaher Daten zum Zwecke der Nutzenbewertung in Abhängigkeit vom Studientyp

Prozessschritt	Generelle Anforderungen (für alle Studientypen)	Spezifische Anforderungen an vergleichende Studien ohne Randomisierung	Spezifische Anforderungen an vergleichende Studien mit Randomisierung	Spezifische Anforderungen an adjustierte indirekte Vergleiche über einen gemeinsamen Brückenkomparator
Formulierung Fragestellung und Entscheidung für ein Studiendesign	<ul style="list-style-type: none"> ▪ Identifizierung der Evidenzlücke ▪ Formulierung der Fragestellung (PICO) gemäß Evidenzlücke ▪ Berücksichtigung der Anforderungen an die Nutzenbewertung aus §35a 	<ul style="list-style-type: none"> ▪ keine Faktoren, die unwahrscheinlich machen, dass mit diesem Studiendesign ausreichend valide Ergebnisse erreicht werden können 	<ul style="list-style-type: none"> ▪ keine sehr großen (dramatischen) Effekte für entscheidungsleitende Endpunkte zu erwarten, interessierende Endpunkte auch unter Vergleichstherapie erreichbar 	<ul style="list-style-type: none"> ▪ prinzipielle Verfügbarkeit von Studien für einen solchen Vergleich (Vorabrecherche)
Studienplanung	<ul style="list-style-type: none"> ▪ explizites, vor Beginn der Datenerhebung finalisiertes Studienprotokoll ▪ präspezifizierter Analyseplan 	<ul style="list-style-type: none"> ▪ Nachbildung der Planung von vergleichenden Studien mit Randomisierung (emulation of target trial) ▪ Präspezifikation möglicher Confounder und deren Adjustierung in der Analyse 	<ul style="list-style-type: none"> ▪ Anpassung der Studienplanung an den Behandlungsaltag (pragmatische randomisierte Studie: Ein- und Ausschlusskriterien, Interventionen, Endpunkte, Visiten) 	<ul style="list-style-type: none"> ▪ Berücksichtigung pragmatischer randomisierter Studien in den Einschlusskriterien ▪ Planung eines systematischen Reviews inkl. Festlegung der Anforderungen an indirekte Vergleiche
Datenerhebung	<ul style="list-style-type: none"> ▪ gewähltes Instrument zur Datenerhebung muss Daten in der notwendigen Qualität zur Verfügung stellen können 	<ul style="list-style-type: none"> ▪ Sicherstellung der Verfügbarkeit der Daten zur Confounderkontrolle 	<ul style="list-style-type: none"> ▪ ggf. Nutzung bestehender Datenstrukturen, z. B. Register 	<ul style="list-style-type: none"> ▪ ggf. Re-Analyse vorhandener Studien, um Anforderungen an indirekten Vergleich zu erfüllen
Analyse und Interpretation	<ul style="list-style-type: none"> ▪ Berücksichtigung der Aussagekraft der unterschiedlichen Studiendesigns und der konkreten Datenqualität bei der Interpretation der Ergebnisse 	<ul style="list-style-type: none"> ▪ Annäherung an Strukturgleichheit der Behandlungsgruppen durch Adjustierung 	<ul style="list-style-type: none"> ▪ Analyse und Interpretation gemäß bestehender Standards 	<ul style="list-style-type: none"> ▪ Prüfung der Voraussetzungen für einen indirekten Vergleich (Ähnlichkeit, Homogenität, Konsistenz der Studien)

Wahl des Designs und Studienplanung

Die Entscheidung für ein Studiendesign sollte berücksichtigen, ob mit dem jeweils gewählten Design ausreichend valide Ergebnisse für eine Nutzenbewertung erreicht werden können.

Eine adäquate und auch in ihrem zeitlichen Ablauf nachvollziehbare Studienplanung ist für die Aussagekraft der Ergebnisse einer Studie von entscheidender Bedeutung. Vor Studienbeginn sind ein Studienprotokoll und ein statistischer Analyseplan zu erstellen, und die Studie sollte in einem Studienregister registriert werden.

Für die Planung von vergleichenden Studien ohne Randomisierung zum Zweck des Vergleichs von Therapieeffekten wird die explizite Nachbildung der Planung von vergleichenden Studien mit Randomisierung empfohlen (emulation of target trials).

Wesentlich für die Datenanalyse von Studien ohne Randomisierung sind Adjustierungen, die den Einfluss der Strukturungleichheit der Behandlungsgruppen ausgleichen sollen. Um eine ergebnisgesteuerte Analyse zu vermeiden, müssen die relevanten Confounder und der Ablauf der Adjustierung während der Analyse in der Studienplanung nachvollziehbar und in der notwendigen Detailtiefe präspezifiziert werden. Die relevanten Confounder müssen systematisch identifiziert (z. B. auf Basis wissenschaftlicher Literatur unter Einbindung von Fach-Experten) und in der Studienplanung präspezifiziert werden. Die Verfügbarkeit entsprechender Daten in der gewählten Datenquelle ist vor der Entscheidung für eine vergleichende Studie ohne Randomisierung sicherzustellen. Eine Adjustierung lediglich für die im Datensatz verfügbaren Confounder ist nicht ausreichend, wenn diese nicht die relevanten Confounder abdecken.

Bei der Durchführung einer vergleichenden Studie ohne Randomisierung besteht die Möglichkeit, die Daten retrospektiv oder prospektiv oder kombiniert teils retrospektiv und teils prospektiv zu erheben. Die retrospektive Datenerhebung ist nur sinnvoll, wenn der Datenkörper, auf dessen Grundlage die retrospektive Datenerhebung erfolgen soll, die notwendigen Daten in der erforderlichen Qualität enthält. Die Verfügbarkeit der relevanten Daten ist vor der Entscheidung für ein retrospektives Design sicherzustellen. Historische Kontrollen sind möglich, wenn die in der Vergangenheit beobachteten Patientenpopulationen hinreichend ähnlich zur aktuell behandelten Patientenpopulation sind und wenn in der Vergangenheit Daten in ausreichender Qualität für einen aussagekräftigen Vergleich gesammelt wurden. Darüber hinaus müssen aus dem historischen Datensatz auch die für die aktuelle Untersuchung relevanten spezifischen Daten verfügbar sein (z. B. patientenindividuelle Daten zu Confoundern), und die Daten müssen ausreichend ähnlich sein (z. B. Endpunkte und Confounder ausreichend ähnlich definiert und erhoben).

Sind die notwendigen Daten nicht in ausreichendem Umfang oder in ausreichender Qualität verfügbar, ist eine prospektive Datenerhebung notwendig. Dabei kann ggf. auf bestehende Datenquellen zurückgegriffen werden (z. B. indikationsspezifisches klinisches Register),

indem der Datensatz in der prospektiven Erhebung um die fehlenden Daten (z. B. einzelne Endpunkte) ergänzt wird.

Instrumente zur Datenerhebung

Die verschiedenen Instrumente zur Datenerhebung (studienindividuelle Datenerhebung, Datenerhebung in Registern, in elektronischen Patientenakten und aus Abrechnungsdaten von Krankenkassen) können grundsätzlich jeweils sowohl für vergleichende Studien ohne Randomisierung als auch für solche mit Randomisierung eingesetzt werden.

Praktisch erscheint die Erhebung von versorgungsnahen Daten zum Zwecke der Nutzenbewertung aus elektronischen Patientenakten und Abrechnungsdaten von Krankenkassen derzeit und absehbar nicht realistisch durchführbar zu sein. Die Hauptgründe liegen in der eingeschränkten Abbildung relevanter Daten für die Nutzenbewertung (Patientencharakteristika und Endpunkte) und in der eingeschränkten Datenqualität dieser Quellen.

Neben einer studienindividuellen Datenerhebung kommen für die versorgungsnahen Datenerhebung zum Zwecke der Nutzenbewertung insbesondere indikationsweite Patientenregister infrage. Denn von den nicht primär auf vergleichende Studien ausgerichteten Instrumenten zur Datenerhebung bieten solche Register am ehesten die Möglichkeit der Anpassung an die notwendige Datenerhebung für diese Studien. Dies betrifft sowohl die Spezifizierung der notwendigen Daten als auch die Datenqualität. In den vergangenen Jahren wurden Zielsetzung und Dokumentationsumfang von Registern erweitert. Insbesondere die zunehmende Dokumentation klinischer Informationen in Registern, die für die Beschreibung von Patientenpopulationen, Interventionen (einschließlich Komparatoren) und Endpunkten (PICO) für die Nutzenbewertung herangezogen werden können, ist in diesem Zusammenhang relevant. Bei prinzipieller Erweiterbarkeit eines Registers ist auch die Kombination mit einer ergänzenden studienindividuellen Datenerhebung für die jeweilige Registerstudie denkbar.

Studien auf Basis einer Datenerhebung in einem Register

Konzeptionell ist es wichtig, zwischen Registern (aktive, prospektive, standardisierte Dokumentation von Beobachtungseinheiten zu vorab festgelegten, aber im Zeitverlauf erweiterbaren Fragestellungen) und Studien in diesen Registern (Registerstudien) zu unterscheiden. Grundsätzlich sind in Registern nicht interventionelle und interventionelle vergleichende Studien möglich. In Registern können sowohl vergleichende Studien ohne Randomisierung als auch mit Randomisierung durchgeführt werden.

Anforderungen an die Datenqualität

Zur Beschreibung der Anforderungen an die Datenqualität in Registern liegen mehrere nationale und internationale Leitlinien, Übersichten und Positionspapiere vor. Diese stimmen in ihren Grundzügen in weiten Teilen überein. Letztlich ist jedoch nicht entscheidend und auch nicht erforderlich, dass alle dort genannten Maßnahmen vollständig umgesetzt sind, sondern dass die für die konkrete Fragestellung relevanten Daten in einer solchen Qualität vorliegen,

dass die Auswertungen im Rahmen einer Registerstudie ausreichend sicher interpretierbar sind. Um dies zu gewährleisten, lassen sich verschiedene Kategorien von Kriterien zur Datenqualität eines Registers unterscheiden (siehe Tabelle 2).

Tabelle 2: Kriterien für die Datenqualität und die Sicherstellung der Datenqualität für versorgungsnahe Datenerhebungen zum Zwecke der Nutzenbewertung von Arzneimitteln

Kategorie	Qualitätskriterien
Obligatorische Kriterien zur Sicherstellung der Datenqualität	<ul style="list-style-type: none"> ▪ Detaillierte Registerbeschreibung (Zielsetzung, Registerprotokoll) ▪ Exakte Definition / Operationalisierung von Expositionen, klinischen Ereignissen, Endpunkten und Confoundern ▪ Aktueller Datenplan / Kodierhandbuch ▪ Schulungen zur Datenerhebung und -erfassung ▪ Klar definierte Ein- und Ausschlusskriterien für Registerpatienten ▪ SOP-System zur Datenerhebung ▪ Maßnahmenpaket zur Sicherstellung der Richtigkeit der Daten und zur Information über Fehlerraten (z. B. source data verification, interne und externe Audits, IT-gestützte Prüfungen [z. B. Cross-Reference-Checks]) ▪ Documentation trail – Dokumentation der Prozess- und Definitionsänderungen im Register ▪ Wissenschaftliche Unabhängigkeit des Registers ▪ Nachhaltige Finanzierung
Allgemeine Kriterien, die regelhaft für Registerstudien zum Zwecke der Nutzenbewertung von Arzneimitteln relevant sind	<ul style="list-style-type: none"> ▪ Verwendung exakter Datumsangaben zum Patienten, zur Erkrankung und zu Ereignissen ▪ Detaillierte Informationen zur Arzneimitteltherapie (Wirkstoff, Dosis, Dosisänderung, inkl. Datumsangaben) ▪ Timeliness (Aktualität / schnelle Verfügbarkeit / Pünktlichkeit der benötigten Ergebnisse)
Allgemeine Kriterien, die je nach Fragestellung für Registerstudien zum Zwecke der Nutzenbewertung relevant sein können	<ul style="list-style-type: none"> ▪ Verwendung von Standard-Klassifikationen (z. B. ICD-10) und -Terminologien (z. B. MedDRA) ▪ Verwendung validier Standard-Erhebungsinstrumente (Fragebogen, Skalen, Tests) ▪ Flexibilität und Anpassungsfähigkeit (z. B. zur Einbettung von Studien, für weitere Datenerhebung, bei veränderter Versorgungssituation) ▪ Verknüpfbarkeit mit anderen Datenquellen
Kriterien, deren Erfüllungsgrad fragestellungsbezogen zu beurteilen ist ^a	<ul style="list-style-type: none"> ▪ Repräsentativität der Stichprobe / Selektion der Stichprobe ▪ Vollständigkeit der Daten je Erhebungszeitpunkt (loss-to-follow-up, Drop-outs) ▪ Vollständigkeit der Erhebungszeitpunkte ▪ Richtigkeit der Daten ▪ Erhebung aller für die Fragestellung relevanten Confounder ▪ Datenkonsistenz über die Zeit
<p>a: Die genannten Kriterien sind wichtige Kriterien der Datenqualität, jedoch nur fragestellungsbezogen zu beurteilen. Zum einen beziehen sich z. B. „Richtigkeit der Daten“ und „Datenkonsistenz über die Zeit“ nur auf die Daten, die für die jeweilige Fragestellung relevant sind. Zum anderen bezieht sich „Repräsentativität der Stichprobe“ nur auf die für die Fragestellung relevante Population, nicht jedoch auf die gesamte Registerpopulation.</p> <p>ICD: Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme; MedDRA: Medizinisches Wörterbuch für Aktivitäten im Rahmen der Arzneimittelzulassung; SOP: Standard Operating Procedure</p>	

Im Rahmen der Eignungsprüfung eines spezifischen Registers sollte mithilfe dieser Kriterienliste für die jeweilige konkrete Fragestellung evaluiert werden, ob und in welchem Ausmaß die einzelnen Kriterien erfüllt sind, welchen Einfluss eine mögliche Nichterfüllung auf die Ergebnisqualität vermutlich hat, und ob im Rahmen einer registerbasierten Studie eventuelle Defizite in einem für die Beantwortung der Fragestellung sinnvollen Rahmen und mit vertretbarem Aufwand korrigiert werden können.

Interviews mit Registerbetreibern

Aus den Interviews mit den Registerbetreibern hat sich ergeben, dass die Eignung des jeweiligen Registers für die Nutzenbewertung von Arzneimitteln in der Regel nicht pauschal zu beantworten ist, sondern von der konkreten Fragestellung abhängt. Die Register sind jedoch zumeist technisch und organisatorisch darauf vorbereitet, ggf. notwendige Erweiterungen des Datensatzes zu implementieren.

Aus den Interviews lassen sich zudem verschiedene förderliche und hinderliche Faktoren für einen Registerbetrieb ableiten. Daraus und allgemein aus den Ergebnissen der Interviews lassen sich Handlungsempfehlungen für Registerbetreiber, Verantwortliche für Registerstudien sowie Entscheidungsträger im Gesundheitswesen und in der Politik ableiten.

Anforderungen an die Auswertung

Bei Studien ohne Randomisierung ist die für einen fairen Vergleich notwendige Strukturgleichheit der zu vergleichenden Gruppen in der Regel nicht gegeben. Gruppenunterschiede in möglichen Confoundern, d. h. Faktoren, die sowohl mit der Behandlung als auch mit Endpunkten in Beziehung stehen und somit einen Behandlungseffekt verfälschen können, müssen daher bei der Schätzung von Effekten berücksichtigt werden. In einem detaillierten Studienprotokoll und Analyseplan soll daher unter anderem die systematische Identifikation der relevanten Confounder (z. B. mittels wissenschaftlicher Literatur unter Einbindung von Fach-Experten) beschrieben werden, da sich eine Confounderadjustierung daran orientieren muss, welche Confounder für die Fragestellung relevant sind und nicht, welche im Datensatz enthalten sind.

Für die Confounderadjustierung stehen verschiedene Ansätze zur Verfügung. Zum Zwecke der Nutzenbewertung von Arzneimitteln sind für die Confounderadjustierung regelhaft nur solche Verfahren sinnvoll, die unter Verwendung individueller Patientendaten durchgeführt werden. Eine häufig angewendete Methode zur Berücksichtigung von Confoundern in vergleichenden Studien ohne Randomisierung auf Basis von Registern stellt die Verwendung von Propensity Scores dar. Bei der Verwendung der Propensity-Score-Methode sind unter anderem Positivität, Überlappung und Balanciertheit wichtige Kriterien. Die diesbezügliche Entscheidungsstruktur ist im Analyseplan festzulegen. Diese Entscheidungsstruktur soll auch Vorgaben für Entscheidungen in Abhängigkeit von der konkreten Datensituation enthalten (z. B. Mindestmaß für die Überlappung und die Balanciertheit).

In der Praxis verbleibt auch bei striktem Befolgen der üblichen methodischen Guidelines, dass die Richtigkeit der Annahmen bezüglich Confounderadjustierung nicht vollumfänglich überprüfbar ist und ungemessene oder gänzlich unbekannte Confounder eine Rolle spielen können. Insgesamt liefern daher Ergebnisse aus vergleichenden Studien ohne Randomisierung regelhaft nur eine höchstens geringe qualitative Ergebnissicherheit. Auch wenn Studien ohne Randomisierung nur eine geringe qualitative Ergebnissicherheit aufweisen, können sie bei Zusammenführung mit anderen Daten die Ergebnissicherheit der Gesamtaussage zum Zusatznutzen erhöhen, z. B. wenn ergebnissichere Daten zu wichtigen Endpunkten durch die Studie ohne Randomisierung in anderen Endpunktkatoren ergänzt werden, oder bei gemeinsamer Betrachtung einer (kleinen) Studie mit Randomisierung mit einer (größeren) Studie ohne Randomisierung.

Auch bei sorgfältigster Analyse unter Erfüllung der genannten Qualitätsanforderungen sollte aufgrund potenziell unbekannter Confounder aus den in der Studie beobachteten Effekten erst ab einer bestimmten Effektstärke eine Aussage zum Nutzen oder Schaden einer Intervention abgeleitet werden. Eine (positive oder negative) Aussage zum Nutzen oder Schaden ergibt sich dann, wenn das Konfidenzintervall für den beobachteten Effekt oberhalb einer zu definierenden Schwelle liegt. Da die Erfüllung der genannten Qualitätsanforderungen Voraussetzung für die Betrachtung der Effekte ist, sollte dieser Schwellenwert deutlich unterhalb des Wertes für den „dramatischen Effekt“ (relatives Risiko von 5-10) liegen, z. B. in einem Bereich von 2-5 für das relative Risiko. Die konkrete Schwelle ergibt sich durch die Qualität der Daten im Einzelfall.

Bei Nutzenbewertungen von Arzneimitteln nach § 35a SGB V liegt, ausgehend von dieser Schwelle für eine Aussage zum Nutzen oder Schaden, bei bei Überschreiten der Schwelle mindestens ein geringer Zusatznutzen für den jeweiligen Endpunkt vor (Ausnahme: Endpunkte der Kategorie nicht schwerwiegende / nicht schwere Folgekomplikationen, weil für diese gemäß Arzneimittel-Nutzenbewertungsverordnung zusätzlich eine „nicht nur geringfügige Verbesserung“ gefordert wird). Die Einstufung in die Ausmaßkategorien „beträchtlich“ oder „erheblich“ bedarf für alle Endpunktkatoren höhere, d. h. oberhalb der oben genannten Schwelle liegende, nach Größe abgestufte Effektstärken.

Insbesondere für seltene Erkrankungen kann es sinnvoll und notwendig sein, dass Studien in internationaler Kooperation durchgeführt werden. Bei solchen Analysen ist zum einen eine standardisierte Datenharmonisierung erforderlich. Zum anderen ist bei Analysen, die außerhalb des interessierenden deutschen Versorgungskontextes generierte Daten heranziehen, zu begründen, dass die Daten versorgungsnah im Sinne der Versorgung in Deutschland sind bzw. Abweichungen davon für die Effektschätzung nicht relevant sind. Bei Analysen aus mehreren Registern kann es u. a. aus Aufwandsgründen sinnvoll sein, nicht einen gemeinsamen Datenpool zu bilden und diesen dann auszuwerten, sondern identisch geplante Studien in den einzelnen Registern zu planen und durchzuführen und diese Studien dann metaanalytisch zusammenzufassen.

Anforderungen an die Berichterstattung

Unabhängig vom Studientyp umfasst die vollständige Dokumentation einer Studie das Studienprotokoll (Planung der Methodik und Durchführung der Studie), den Analyseplan (Planung der Auswertung der Daten) und den Ergebnisbericht (Beschreibung der geplanten Methodik [inkl. Auswertung] und Durchführung der Studie, der Abweichungen von dieser Planung und der vollständigen Ergebnisse). Studienprotokoll und Analyseplan dienen neben der Beschreibung der Methodik und Durchführung der Studie bei prospektiv erhobenen Daten auch der Präspezifizierung der Studienplanung. Diese Präspezifizierung ist ein wesentliches Qualitätsmerkmal einer Studie mit prospektiver Datenerhebung. Für prospektive vergleichende Studien ohne Randomisierung sollte die Präspezifikation die Confounderkontrolle (Erhebung von Confoundern und Vorgehen bei der Adjustierung) so umfänglich wie möglich abdecken.

Die Möglichkeit einer ergebnisgesteuerten Analyse und Berichterstattung ist für retrospektive Studiendesigns ein ungelöstes Problem. Unabhängig davon sollte auch für Studien mit retrospektiver Datenerhebung aus vorhandenen Datenkörpern ein Studienprotokoll inkl. Analyseplan erstellt werden. Es wird empfohlen, diese Dokumente öffentlich verfügbar zu machen.

Optimierte Studien für die Entscheidungsfindung im Gesundheitswesen

Die Durchführung einer qualitativ hochwertigen vergleichenden Studie ohne Randomisierung ist mit einem hohen Aufwand verbunden. In diesem Zusammenhang ist die aktuelle Diskussion um Anpassungen bei der Durchführung von vergleichenden Studien mit Randomisierung relevant, die zum einen die Aussagekraft für breitere Populationen sicherstellen soll (pragmatische Studien mit Randomisierung) und zum anderen den notwendigen Aufwand verringern soll (large simple trials und registerbasierte Studien mit Randomisierung). Zusammenfassend kann es einfacher und zielführender sein, eine vergleichende Studie mit Randomisierung unter Berücksichtigung dieser Anpassungen durchzuführen, als zu versuchen, qualitativ hochwertige Ergebnisse aus einer vergleichenden Studie ohne Randomisierung zu generieren.

Vorschläge für ein Vorgehen für die anwendungsbegleitende Datenerhebung nach §35a Abs. 3b SGB V

Im vorliegenden Rapid Report wurden die Ergebnisse des Projekts auch im Zusammenhang mit der durch das GSAV eingeführten Möglichkeit der anwendungsbegleitenden Datenerhebung analysiert.

Definition der Fragestellung

Die Grundlage für die anwendungsbegleitende Datenerhebung ist die Definition der Fragestellung, die durch diese Datenerhebung beantwortet werden soll. Die Fragestellung beinhaltet mindestens die Komponenten des PICO-Schemas und die notwendige Dauer der Datenerhebung. Die genauen Spezifika der Fragestellung werden aus der Evidenzlücke abgeleitet, die sich in der Nutzenbewertung ergeben hat und die durch die Datenerhebung

geschlossen werden soll. Eine so definierte Fragestellung ist auch der Ausgangspunkt für die Beschreibung des notwendigen Umfangs der Datenerhebung (einschließlich Beobachtungsdauer und Fallzahlplanung).

Evidenzlücken in Nutzenbewertungen von Arzneimitteln (Orphan Drugs)

Um besser einschätzen zu können, welche Evidenzlücken bei der Nutzenbewertung bestehen und für die Festlegung des Ausmaßes des Zusatznutzens von Bedeutung sind, wurden Beschlüsse des G-BA zu Nutzenbewertungen von Orphan Drugs näher betrachtet. Untersucht wurden dabei Beschlüsse zu Orphan Drugs mit Marktzugang in den Jahren 2014 bis 2018 einschließlich Beschlüssen zu neuen Anwendungsgebiete von Orphan Drugs in diesem Zeitraum.

Die zusammenfassende Analyse zeigt einerseits, dass für Orphan-Drug-Bewertungen der Jahre 2014 bis 2018 zum Zeitpunkt des Marktzugangs für einen Großteil der Fragestellungen relevante Daten im zugehörigen Dossier zur Nutzenbewertung vorgelegt wurden (80 von 85 Fragestellungen, 94 %). In ca. 2 Dritteln der Fälle handelte es sich dabei um Studien mit Randomisierung, bei einem Drittel der Fälle um Studien ohne Randomisierung. Dennoch wurde nur für ca. ein Viertel aller Fragestellungen der Zusatznutzen quantifiziert. Dabei ließ sich nicht ableiten, dass der Zusatznutzen insbesondere bei kleinen Zielpopulationen nicht quantifizierbar wäre.

Die Analyse der Fragestellungen mit nicht quantifizierbarem Zusatznutzen zeigt, dass bei insgesamt 61 % (52 von 85) der in den Jahren 2014 bis 2018 vom G-BA bewerteten Fragestellungen Evidenzlücken benannt wurden, die für die fehlende Quantifizierbarkeit ausschlaggebend waren. Nahezu 2 Dritteln der Orphan-Drug-Bewertungen stellen demnach potenziell Kandidaten für eine anwendungsbegleitende Datenerhebung dar. In allen 52 Fällen fehlten auch Daten zur Kontrollgruppe, weshalb eine zielgerichtete anwendungsbegleitende Datenerhebung regelhaft vergleichend unter Hinzuziehung einer Kontrollgruppe (Vergleichstherapie) geplant und durchgeführt werden muss.

Evidenzlücken liegen oft in mehreren, wenn auch nicht regelhaft in allen Endpunktkatoren (Mortalität, Morbidität, gesundheitsbezogene Lebensqualität und Nebenwirkungen) vor. Da auch bei den Endpunktkatoren Morbidität und gesundheitsbezogene Lebensqualität häufig Evidenzlücken bestehen, werden für zielgerichtete anwendungsbegleitende Datenerhebungen oft Informationen zu patientenberichteten Endpunkten (PROs) erforderlich sein, da diese zur Adressierung der Evidenzlücken in diesen Endpunktkatoren essenziell sind.

Es ist daher insgesamt absehbar, dass als Zulassungsaufgabe geplante Datenerhebungen, die insbesondere der Identifikation seltener oder spät auftretender Nebenwirkungen des jeweiligen Orphan Drugs dienen sollen, in unveränderter Form häufig keine geeignete, zielgerichtete anwendungsbegleitende Datenerhebung zum Zwecke der Nutzenbewertung darstellen werden. Welche Änderung bzw. Erweiterung einer regulatorisch geforderten Datenerhebung für eine zielgerichtete anwendungsbegleitende Datenerhebung zum Zwecke der Nutzenbewertung

notwendig ist, ist jeweils im Einzelfall auf Basis der für die Quantifizierung des Zusatznutzens bestehenden Evidenzlücke zu prüfen.

Mögliche Prozessschritte der anwendungsbegleitenden Datenerhebung

Ausgehend von der Analyse der Orphan Drug Bewertungen der Jahre 2014 bis 2018 und auf Basis der Vorgaben des SGB V zeigt Tabelle 3 mögliche Prozessschritte der anwendungsbegleitenden Datenerhebung im Verfahren der Nutzenbewertung nach § 35a SGB V.

Tabelle 3: Prozessschritte für eine anwendungsbegleitende Datenerhebung zum Zwecke der Nutzenbewertung nach § 35a

Prozessschritt	Anmerkung
Identifizierung einer Evidenzlücke im Beschluss des G-BA zur Nutzenbewertung nach §35a	<ul style="list-style-type: none"> ▪ Evidenzlücke: relevante Datenlücke zum Vergleich des neuen Arzneimittels mit der (zweckmäßigen) Vergleichstherapie bezüglich patientenrelevanter Endpunkte (insbesondere, wenn aufgrund der Evidenzlücke eine Quantifizierung des Zusatznutzens nicht möglich ist)
Beschreibung der Vorgaben für die anwendungsbegleitende Datenerhebung durch den G-BA und Übermittlung an den pharmazeutischen Unternehmer	<ul style="list-style-type: none"> ▪ Definition der Fragestellung ▪ Dauer, Art und Umfang der Datenerhebung (Dauer der Datenerhebung pro Patient, Fallzahl auf Basis einer Fallzahlabschätzung) ▪ Art und Umfang der Auswertung (abhängig vom verwendeten Studientyp) ▪ Festlegung der Zeitpunkte für die Überprüfung der gewonnenen Daten (mindestens alle achtzehn Monate) ▪ Festlegung der Vorgaben unter Berücksichtigung laufender und geplanter Datenerhebungen, insbesondere solcher, die sich aus Auflagen der Zulassungsbehörden (z. B. der EMA) ergeben
Überprüfung der erhobenen Daten und der Verpflichtung zur Datenerhebung	<ul style="list-style-type: none"> ▪ zum 1. Überprüfungszeitpunkt prüft der G-BA, ob ein (öffentlich verfügbares) Studienprotokoll inkl. Analyseplan vorliegt, das die anwendungsbegleitende Datenerhebung gemäß Anforderung abbildet ▪ zum 1. und jedem weiteren Überprüfungszeitpunkt prüft der G-BA die verfügbaren Daten und entscheidet, ob die Datenerhebung beendet werden kann oder fortgeführt werden soll

EMA: Europäische Arzneimittel-Agentur; G-BA: Gemeinsamer Bundesausschuss

Fazit

Studiendesign und Datenerhebung

- der Einsatz versorgungsnaher Daten zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V verlangt einen Vergleich zwischen dem neuen Arzneimittel und der vom G-BA bestimmten zweckmäßigen Vergleichstherapie; damit ist die Durchführung vergleichender Studien notwendig
- die Erhebung versorgungsnaher Daten zum Zwecke der Nutzenbewertung nach § 35a SGB V aus elektronischen Patientenakten und aus Abrechnungsdaten von Krankenkassen wird derzeit und absehbar als nicht realistisch eingeschätzt; es ist vielmehr eine studienindividuelle Datenerhebung oder eine Datenerhebung in Patientenregistern notwendig

Versorgungsnahe vergleichende Studien ohne Randomisierung

- sollen versorgungsnahen vergleichende Studien ohne Randomisierung für die Nutzenbewertung herangezogen werden, muss bereits mit der Studienplanung sichergestellt werden, dass der Ablauf der Studie und die erhobenen Daten die notwendige Qualität haben, um interpretierbare Ergebnisse zu generieren
- wesentliche Komponenten einer solchen Studienplanung sind ein Studienprotokoll inkl. Analyseplan, die Emulation einer Zielstudie, die die relevante Fragestellung behandelt, und die Sicherstellung der Erhebung ausreichender Daten für eine Confounderkontrolle
- ein zentraler Aspekt der Analyse einer vergleichenden Studie ohne Randomisierung ist die adäquate Adjustierung für Confounder; diese Adjustierung muss soweit möglich präspezifiziert werden, die enthaltenen Annahmen (z. B. die Definition der relevanten Confounder) sind dabei fundiert zu begründen
- aus versorgungsnahen vergleichenden Studien ohne Randomisierung können keine Effekte abgeleitet werden, wenn die Datenqualität in den verwendeten Datenquellen und die Qualität der Auswertung und Berichterstattung nicht hoch ist
- auch unter hohen Qualitätsanforderungen (für Daten, Auswertung und Berichterstattung) kann im Regelfall aus vergleichenden versorgungsnahen Studien ohne Randomisierung nicht mehr als ein Anhaltspunkt für einen Effekt abgeleitet werden
- wegen der inhärenten Unsicherheit der Ergebnisse aus vergleichenden Studien ohne Randomisierung sollte aufgrund potenziell unbekannter Confounder aus den in der Studie beobachteten Effekten erst ab einer bestimmten Effektstärke eine Aussage zum Nutzen oder Schaden einer Intervention abgeleitet werden; eine Quantifizierung eines Zusatznutzens gemäß der gesetzlich vorgegebenen Ausmaßkategorien erfordert entsprechende nach der Größe abgestufte Effektstärken
- die Möglichkeit, retrospektive Studiendesigns zu berücksichtigen, hängt davon ab, ob die verfügbaren Datenquellen die notwendigen Daten in der erforderlichen Qualität bereits enthalten; Vergleiche eines neuen Arzneimittels mit historischen Kontrollen erscheinen

nur dann realistisch, wenn für das neue Arzneimittel und die historische Kontrolle dieselbe Datenquelle (z. B. ein indikationsspezifisches klinisches Register) genutzt wird

Versorgungsnahe vergleichende Studien mit Randomisierung

- versorgungsnahe vergleichende Studien können auch mit Randomisierung durchgeführt werden (pragmatische klinische Studien)
- der Aufwand für eine versorgungsnahe vergleichende Studie mit Randomisierung wird bei vergleichbarer Datenqualität in der Regel geringer sein als der für eine Studie ohne Randomisierung, weil die Datenerhebung und Adjustierung für die Confounderkontrolle entfallen kann
- versorgungsnahe vergleichende Studien mit Randomisierung erreichen eine höhere Aussagekraft als solche ohne Randomisierung, eine Quantifizierung des Zusatznutzens ist zuverlässiger möglich
- insbesondere nach der Zulassung können versorgungsnahe vergleichende Studien mit Randomisierung abhängig von der bestehenden Fragestellung mit einer eingeschränkten Datenerhebung ([large] simple trials) durchgeführt werden; eine Durchführung in Registern hat ein zusätzliches Potenzial, die Studien zu beschleunigen und weniger aufwendig zu gestalten (registerbasierte vergleichende Studien mit Randomisierung)

Anwendungsbegleitende Erhebung versorgungsnaher Daten

- ob die verschiedenen Patientenregister bereits heute für die anwendungsbegleitende Datenerhebung nach § 35a SGB V geeignet sind, lässt sich nicht pauschal beantworten, da dies vom jeweiligen Register und den konkreten Fragestellungen abhängt
- auf Basis der Analysen und der Gespräche mit den Registerbetreibern lassen sich Handlungsfelder beschreiben, die dazu dienen, die einzelnen Register im Speziellen und die Registerlandschaft in Deutschland im Allgemeinen bei der anwendungsbegleitenden Datenerhebung zu unterstützen; diese Handlungsfelder sind im Bericht beschrieben
- die Erkenntnisse des vorliegenden Berichts können für die anwendungsbegleitende Datenerhebung zur Schließung von Evidenzlücken nach einer Nutzenbewertung eingesetzt werden; ein Vorschlag zum Ablauf dieser Datenerhebung wird mit dem Bericht zur Verfügung gestellt

1 Hintergrund

Der Gemeinsame Bundesausschuss (G-BA) hat am 02.05.2019 das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) nach § 139a, Abs. 3, SGB V mit einer wissenschaftlichen Ausarbeitung von Konzepten zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V beauftragt.

Die Abschnitte 5.1 bis 5.6 zeigen die Ergebnisse dieser Ausarbeitung. Ausgehend von diesen Ergebnissen werden in Abschnitt 5.7 Vorschläge zur anwendungsbegleitenden Datenerhebung nach dem GSAV [1] gemacht, da diese eine der möglichen Anwendungen einer versorgungsnahen Datenerhebung darstellt.

2 Fragestellung

Ziele der vorliegenden Ausarbeitung sind

- die Erstellung einer Übersicht möglicher Konzepte der Generierung und Auswertung versorgungsnaher Daten. Dabei sollen insbesondere auch Datenerhebungen berücksichtigt werden, die nicht der Kategorie „randomisierte kontrollierte Studie (RCT)“ zuzuordnen sind (siehe Abschnitte 5.1 bis 5.4).
- die Bewertung der identifizierten Konzepte der Datengenerierung und deren Auswertung hinsichtlich ihrer Eignung zur Beantwortung der Fragestellung der Nutzenbewertung nach § 35a SGB V insbesondere im Hinblick auf die Möglichkeit einer Quantifizierung des Zusatznutzens (siehe Abschnitt 5.6).
- die Spezifizierung von Kriterien für die Qualität und methodische Anforderungen an die im Rahmen der jeweiligen Datengenerierung erhobenen Daten. Hierbei soll auch darauf eingegangen werden, welche Maßnahmen zur Sicherstellung der Datenqualität notwendig sind (siehe Abschnitte 5.5.1 und 5.5.2).
- die Definition der Anforderungen an die Berichterstattung, die Aufbereitung und Struktur sowie die statistische Auswertung der Daten, die im Rahmen des jeweiligen Konzeptes der Datengenerierung erhoben werden (siehe Abschnitte 5.5.2 und 5.5.3).

3 Projektverlauf

Der Gemeinsame Bundesausschuss (G-BA) hat am 02.05.2019 das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) mit der wissenschaftlichen Ausarbeitung von Konzepten zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V beauftragt.

Die Bearbeitung der Fragestellung wurde auf Basis einer Projektskizze vorgenommen und in einem Rapid Report dokumentiert. Dieser Bericht wurde an den G-BA übermittelt und 2 Wochen später auf der Website des IQWiG veröffentlicht.

4 Methoden

Der vorliegende Auftrag umfasst die konzeptionelle Erarbeitung von Vorschlägen zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V. Die konzeptionelle Arbeit wurde durch eine Informationsbeschaffung und -bewertung unterstützt.

4.1 Informationsbeschaffung und -bewertung

Die Erarbeitung des Konzepts zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V wurde gemäß Projektskizze durch 3 Bausteine unterstützt:

Empirische Information aus den Nutzenbewertungen von Arzneimitteln nach dem AMNOG (§ 35a SGB V)

Die Evidenzgrundlage und die Ergebnisse der abgeschlossenen Nutzenbewertungen nach § 35a SGB V sollten systematisch dargestellt werden. Dabei sollten Konstellationen, die Informationen für die vorliegende Fragestellung enthalten können (insbesondere Bewertungen auf Basis nicht randomisierter Studien), charakterisiert und analysiert werden. Mögliche Formate waren Fallbeschreibungen und aggregierte Daten für Fallkonstellationen (ggf. einschließlich Effektgrößen inkl. Konfidenzintervallen in der gewählten Konstellation).

Die Analyse wurde für Bewertungen zu Orphan Drugs durchgeführt, da diese aufgrund des GSAV voraussichtlich im Fokus einer zukünftigen vom G-BA beauftragten Datenerhebung stehen.

Orientierende Literaturrecherche

Für wissenschaftliche Fragestellungen, die sich als Teil der konzeptionellen Arbeit ergeben (z. B. zur Aussagekraft nicht randomisierter Studien abhängig von vorliegenden Datenkonstellationen), wurde eine orientierende Literaturrecherche durchgeführt, um das zu erarbeitende Konzept durch den aktuellen Stand der wissenschaftlichen Erkenntnis zu unterstützen.

Interviews mit Expertinnen und Experten aus dem Bereich Register

Der Auftrag umfasst die Spezifizierung von Kriterien für die Qualität und methodischen Anforderungen an die im Rahmen der jeweiligen Datengenerierung erhobenen Daten. Hierbei soll auch darauf eingegangen werden, welche Maßnahmen zur Sicherstellung der Datenqualität notwendig sind.

Für die Datengenerierung werden Register eine relevante Rolle spielen. Um in die Anforderungen an die Register Informationen aus der Praxis einfließen lassen zu können, wurden fragebogengestützte Interviews mit Expertinnen und Experten für den Aufbau und die Nutzung von Registern durchgeführt (zur Durchführung siehe Abschnitt 4.2).

4.2 Ermittlung von Qualitätskriterien für Patientenregister

Interviews mit Expertinnen und Experten für Register

Zwischen Ende Juni und August 2019 wurden 6 Experteninterviews mit Vertreterinnen und Vertretern ausgewählter Patientenregister in Deutschland geführt. Ziele waren unter anderem, die Sicht der Praxis auf zentrale Qualitätsanforderungen für Patientenregister und darauf aufbauende Studien kennen zu lernen, sowie Einblicke in wichtige förderliche und hinderliche Faktoren beim Aufbau und Betrieb von Registern zu gewinnen. Die Auswahl wurde auf solche Register begrenzt, in denen die Arzneitherapie eine wichtige Rolle in der Therapie spielt und insofern auch in den Dokumentationen abgebildet ist. Tabelle 4 listet die beteiligten Register auf. Die Auswahl umfasste von der Anzahl eingeschlossener Patienten her unterschiedlich große Patientenregister, die sowohl seltene wie häufige Krankheitsbilder betreffen, onkologische und nicht onkologische Erkrankungen dokumentieren und regional oder bundesweit organisiert sind. Außerdem fand ein Expertenworkshop mit 3 biometrischen Sachverständigen statt, die in Auswertungsfragen von Patientenregisterdaten über fundierte Expertise verfügen (siehe ebenfalls Tabelle 4).

Tabelle 4: Interviews mit Registerbetreibern und biometrischen Sachverständigen

Registerinterviews	
Registername	Gesprächspartner (Affiliation)
Bayerisches Krebsregister des Bayerischen Landesamts für Gesundheit und Lebensmittelsicherheit und des Tumorzentrums Regensburg	<ul style="list-style-type: none"> ▪ Monika Klinkhammer-Schalke (DNVF; Tumorzentrum Regensburg) ▪ Brunhilde Steinger (Tumorzentrum Regensburg) ▪ Vinzenz Völkel (Tumorzentrum Regensburg)
Clinical Research platform Into molecular testing, treatment and outcome of (non-)Small cell lung carcinoma Patients (CRISP) der iOMEDICO	<ul style="list-style-type: none"> ▪ Frank Griesinger (Pius-Hospital Oldenburg; AIO e. V.) ▪ Martina Jänicke (iOMEDICO) ▪ Martin Sebastian (Universitätsklinikum Frankfurt; AIO e. V.)
Deutsches Mukoviszidose Register des Mukoviszidose e. V.	<ul style="list-style-type: none"> ▪ Manuel Burkhardt (Deutsches Mukoviszidose Register) ▪ Lutz Nährlich (Justus-Liebig-Universität Gießen; Deutsches Mukoviszidose Register) ▪ Miriam Schlangen (Mukoviszidose e. V.)
Deutsches MS Register der Deutschen MS Gesellschaft (DMSG)	<ul style="list-style-type: none"> ▪ David Ellenberger (MS Forschungs- und Projektentwicklungs-gGmbH, MS-Register der DMSG, Bundesverband e. V.) ▪ Alexander Stahmann (MS Forschungs- und Projektentwicklungs-gGmbH, MS-Register der DMSG, Bundesverband e. V.)
Patientenregister der European Society for Blood and Marrow Transplantation (EBMT)	<ul style="list-style-type: none"> ▪ Nicolaus Kröger (EBMT; Universitätsklinikum Eppendorf, Hamburg)
Rheumatoide Arthritis: Beobachtung der Biologika-Therapie (RABBIT), Register des Deutschen Rheuma-Forschungszentrums (DRFZ)	<ul style="list-style-type: none"> ▪ Anne Regierer (DRFZ) ▪ Anja Strangenfeld (DRFZ) ▪ Angela Zink (DRFZ)
Epidemiologisch-biometrische Gesprächspartner	
Name	Affiliation
Oliver Kuß	Institut für Biometrie und Epidemiologie, Deutsches Diabetes-Zentrum der Heinrich Heine-Universität Düsseldorf; Wissenschaftlicher Beirat des IQWiG
Rolf Lefering	Institut für Forschung in der operativen Medizin (IFOM), Universität Witten/Herdecke
Claudia Spix	Deutsches Kinderkrebsregister, Universitätsmedizin Mainz
AIO: Arbeitsgemeinschaft Internistische Onkologie in der Deutschen Krebsgesellschaft e. V.; DMSG: Deutsche Multiple Sklerose Gesellschaft Bundesverband e. V.; DNVF: Deutsches Netzwerk für Versorgungsforschung; DRFZ: Deutsches Rheuma-Forschungszentrum; EBMT: European Society for Blood and Marrow Transplantation; IFOM: Institut für Forschung in der operativen Medizin; iOMEDICO: International Organisation of Medical Oncology; MS: multiple Sklerose; RABBIT: Rheumatoide Arthritis: Beobachtung der Biologika-Therapie	

Interviews mit Registerbetreibern

Die Interviewpartner aus den ausgewählten Patientenregistern erhielten zur Vorbereitung auf die Gesprächsinhalte einen 3-fach gegliederten Fragebogen (siehe Anhang A). Der Fragebogen sollte in den ersten beiden Teilen vor dem Gesprächstermin ausgefüllt zurückgesendet werden. Die Antworten dienten dem IQWiG zur gezielten Gesprächsvorbereitung und wurden im Interview in Teilen vertieft. Der Fragebogen erhebt im 1. Teil zunächst wichtige Merkmale des betreffenden Patientenregisters. Im 2. Teil sollte eine Liste von 29 Qualitätskriterien von Patientenregistern nach den Aspekten Relevanz, schwierige und einfache Umsetzbarkeit sowie Aufwand eingeschätzt werden. Schließlich sollte für alle Kriterien der bislang erreichte Erfüllungsgrad im jeweiligen Register eingeschätzt werden.

Im dritten, nur zur Vorabinformation und Gesprächsvorbereitung aufgenommenen Teil des Fragebogens, sind 2 für die Arbeit des G-BA typische Bewertungsszenarien für Arzneimittel enthalten, die zum Thema einer möglichen versorgungsnahen Erhebung zusätzlicher Daten im Interview diskutiert werden sollten.

Der erste und dritte Teil des Fragebogens wurden auf Basis von Expertise bzw. nach Plausibilitätsgesichtspunkten gestaltet. Die Notwendigkeit, im knappen zeitlichen Rahmen der Berichterstellung die geplanten Interviews zügig vereinbaren und durchführen zu müssen, ließ für die Erarbeitung des 2. Teils des Expertenfragebogens (Qualitätskriterien) nur eine orientierende, nicht systematische Literaturrecherche zu. Das Memorandum des Deutschen Netzwerks Versorgungsforschung e. V. (DNVF) zu Patientenregistern [2] wurde zunächst zur Gliederung der Qualitätskriterienliste ausgewählt. Die Strukturierung der Publikation erschien hierfür sehr gut geeignet, denn sie berücksichtigt im Unterschied zu anderen Dokumenten zu Qualitätsstandards für Patientenregister in besonderer Weise die im Auftragsfokus stehenden Aspekte der Datenqualität und ihrer Sicherung.

Um einerseits häufig genannte Qualitätskriterien identifizieren zu können, und um andererseits auf für die Arzneimittelbewertung besonders relevante Kriterien (z. B. Datumsangaben zu Beginn und Ende von Therapien) Hinweise zu erhalten, wurde für die Vorauswahl auf 4 Quellen zurückgegriffen. Dies waren neben der erwähnten DNVF-Publikation [2,3] das AHRQ-Handbuch als etabliertes Standardwerk [4] (Kapitel 25), die aktuell von der EMA zur Diskussion gestellten Anforderungen an Patientenregister [5] (Kapitel 5 bis 6) und die von EUnetHTA im REQuEST-Tool formulierten Aspekte [6] (Kriterien 1 bis 19). Diese Auswahl von Publikationen spiegelt die nationale und europäische Expertise wider und ist zugleich aktuell sowie international repräsentativ für die bei Regulatoren und HTA-Organisationen geführte Diskussion um Patientenregisterqualität. Im Resultat wurde für die Interviews eine im Umfang zunächst begrenzte Liste für den 2. Teil des Fragebogens erstellt, die dann in der im nächsten Abschnitt beschriebenen Weise weiter ergänzt und qualitätsgesichert wurde.

Expertenworkshop mit biometrischen Sachverständigen

Für den Expertenworkshop mit biometrischen Sachverständigen wurden im Vorfeld Diskussionsaspekte zusammengestellt und an die Beteiligten verschickt (siehe Anhang B). Die

Diskussion wurde offen unter Berücksichtigung der genannten Diskussionsaspekte geführt. Die im Expertenworkshop diskutierten Aspekte wurden im vorliegenden Rapid Report berücksichtigt, insbesondere im Abschnitt 5.5 (Anforderungen an Daten und Auswertungen).

Zusammenstellung von Qualitätskriterien für Patientenregister

Um die im Interviewfragebogen mit den Registerbetreibern eingesetzte Liste von Qualitätskriterien zu ergänzen, wurden erstens weitere einschlägige Publikationen mit entsprechenden Empfehlungen zu dem Thema identifiziert und zweitens Hinweise der befragten Expertinnen und Experten auf in der Registerpraxis bedeutsame, aber noch nicht berücksichtigte Gütekriterien integriert. Hierzu gehören unter der Rubrik „Übergeordnete Kriterien“ gefasste Aspekte wie z. B. sichere Finanzierungsgrundlage und eine rasche und pünktliche Verfügbarkeit von benötigten Auswertungsergebnissen und Berichten („timeliness“).

Wie von Mandeville 2018 [7] nach einer systematischen Recherche von Publikationen zur Registerqualität im Rahmen eines aktuellen EUnetHTA-Projekts festgestellt, unterscheiden sich die Empfehlungsdokumente in mancher Hinsicht (z. B. Zweck, Inhalt, Gliederung, Detaillierungsgrad, Erstellungsmethodik, Format). Dennoch lassen sie sich überwiegend durch eine vergleichende Betrachtung für das Erkennen häufig oder übereinstimmend genannter Qualitätskriterien nutzen.

Für die Liste mit Qualitätskriterien für Patientenregister wurden für den vorliegenden Bericht die oben genannten Quellen [2,4-6] um eine 5. Publikation [8] ergänzt. Diese enthält die aus regulatorischer Perspektive formulierten Kriterien des Prüfverfahrens der US-amerikanischen CTT-Initiative (Clinical Trials Transformation Initiative) für die Eignung von Patientenregistern zur Durchführung eingebetteter klinischer Studien. Aus diesen 5 Quellen wurden die Qualitätskriterien für Register und Registerstudien extrahiert und miteinander abgeglichen. Tabelle 14 in Anhang C zeigt die Ergebnisse des Abgleichs. Der Abgleich der Quellen auf die Nennung der einzelnen Qualitätskriterien wurde von 2 wissenschaftlichen Mitarbeitern unabhängig durchgeführt. Im Falle der Nichtübereinstimmung zwischen den Bewertern wurden diese Fälle diskutiert und konsensual gelöst.

5 Ergebnisse

5.1 Studiendesigns und Instrumente zur Datenerhebung zur Generierung versorgungsnaher Daten

Dieses Kapitel skizziert Studiendesigns und Instrumente zur Datenerhebung zur Generierung versorgungsnaher Daten. Versorgungsnahe Daten können zur Beantwortung unterschiedlicher Fragestellungen herangezogen werden, z. B. zur Ermittlung der Häufigkeit verschiedener Erkrankungen in Deutschland, zur Beschreibung der Versorgung bestimmter Patientengruppen oder zur Bewertung des Nutzens von Arzneimitteln. Bei der Beurteilung der Eignung versorgungsnaher Daten zur Unterstützung von Entscheidungen im Gesundheitswesen muss die jeweils zu beantwortende Fragestellung berücksichtigt werden.

Was sind versorgungsnahe Daten zum Zwecke der Nutzenbewertung von Arzneimitteln?

Für den vorliegenden Bericht werden versorgungsnahe Daten wie folgt definiert:

- versorgungsnahe Daten werden bei den Patientenpopulationen erhoben, die ein zu bewertendes Arzneimittel im Versorgungsalltag (in der zugelassenen Indikation) erhalten können
- bei der Erhebung versorgungsnaher Daten erfolgt die Behandlung ohne spezifische Vorgaben

Da es bei der Bewertung von Arzneimitteln nach SGB V um die Versorgung von Patienten in Deutschland geht, müssen versorgungsnahe Daten die beiden oben genannten Kriterien so erfüllen, dass sie Rückschlüsse für die deutsche Versorgung zulassen.

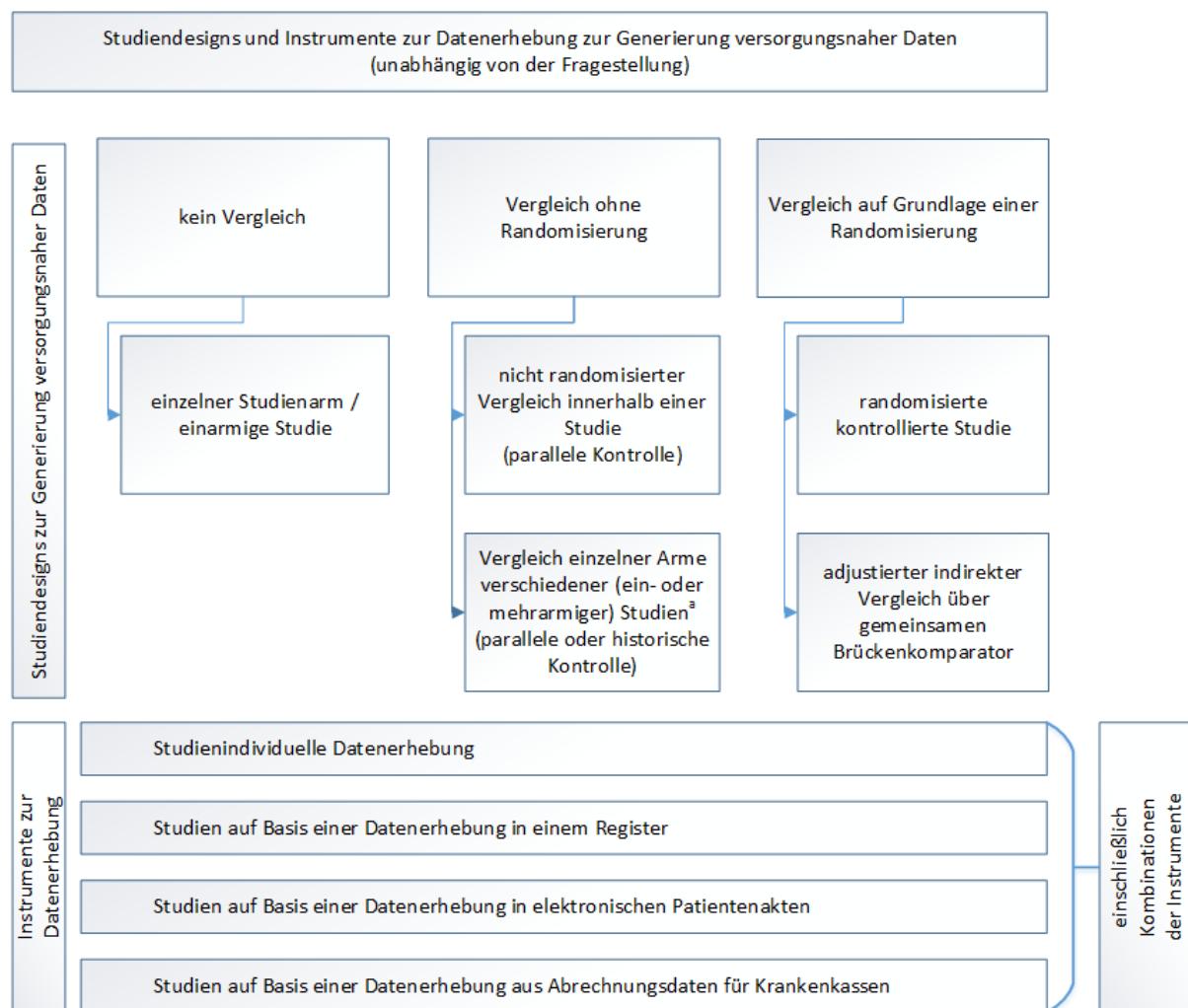
Die Definition versorgungsnaher Daten impliziert weder ein bestimmtes Studiendesign noch ein bestimmtes Instrument der Datenerhebung (siehe unten) [9].

Das Ziel, Daten versorgungsnah zu erheben, verlangt nicht, dass die Datenerhebung sich auf solche Daten beschränkt, die im Versorgungsalltag per se erfasst werden. Eine derart falsch verstandene Beschränkung der Datenerhebung würde vielmehr das Ziel der Nutzenbewertung gefährden. Für die Nutzenbewertung sind regelhaft Daten notwendig, die im Behandlungsalltag nicht bei allen Patientinnen und Patienten dokumentiert werden (z. B. Daten zur gesundheitsbezogenen Lebensqualität, zu Symptomen oder Nebenwirkungen). Versorgungsnahe Daten müssen zudem ausreichend valide und strukturiert sein, um eine Nutzenbewertung zu ermöglichen.

Übersicht über Studiendesigns und Instrumente zur Datenerhebung

Die folgende Abbildung gibt eine Übersicht über Studiendesigns und Instrumente zur Datenerhebung, die zur Generierung versorgungsnaher Daten eingesetzt werden können. Der obere Teil der Abbildung beschreibt die grundsätzlich denkbaren Studiendesigns abhängig von der Art des möglichen Vergleichs von Interventionen. Der untere Teil benennt die Instrumente,

die zur Erhebung versorgungsnaher Daten in Studien mit den verschiedenen Designs herangezogen werden können. Dabei wird deutlich, dass die verschiedenen Instrumente der Datenerhebung generell für alle Studiendesigns eingesetzt werden können.



a: einschließlich Studien zum Spontanverlauf

Abbildung 2: Studiendesigns und Instrumente zur Datenerhebung zur Generierung versorgungsnaher Daten

Zur Bewertung der Komponenten dieser Übersicht hinsichtlich ihrer Eignung zur Beantwortung der Fragestellung der Nutzenbewertung nach § 35a SGB V werden im folgenden Abschnitt zunächst die Voraussetzungen für eine Nutzenbewertung beschrieben. Anschließend werden ausgehend von der Fragestellung der Nutzenbewertung die Anforderungen an die Generierung und Auswertung von versorgungsnahen Daten innerhalb der verschiedenen Stränge der Übersicht diskutiert.

5.2 Voraussetzungen für eine Nutzenbewertung

Der Ausgangspunkt für Überlegungen zu einem sinnvollen Vorgehen bei Nutzenbewertungen kann prägnant mit folgendem Zitat beschrieben werden:

"The effect of any treatment for a given patient is the difference between what happened to the patient as a result of giving him the treatment and what would have happened had treatment been denied." (Der Effekt einer Behandlung für einen Patienten ist der Unterschied zwischen dem, was dem Patienten mit der Behandlung widerfahren ist, und dem, was passiert wäre, wenn dieser die Behandlung nicht erhalten hätte) [10].

Dieser Kernsatz macht das Prinzip und das aus ihm folgende Dilemma deutlich: Grundlage jeder Aussage über Effekte von Interventionen ist ein Vergleich. Dabei ist es gleichgültig, ob es sich um positive oder negative Effekte handelt und ob diese Effekte Wirkung, Wirksamkeit, Nutzen, Zusatznutzen oder Schaden genannt werden. Nur auf der Basis eines Vergleichs ist die für eine kausale Aussage notwendige Differenzierung zwischen „**nach** Intervention A“ und „**wegen** Intervention A“ möglich. Durch die identische Ausgangssituation wie im obigen Zitat beschrieben ergäbe sich dabei der notwendige „*faire*“ Vergleich.

Das Dilemma ergibt sich daraus, dass dieser Vergleich am einzelnen Patienten prinzipiell nicht möglich ist, da (siehe den Konjunktiv im Zitat) nicht 2 alternative Situationen gleichzeitig beobachtbar sind.

Ein Ausweg aus dem Dilemma bestünde darin, einen Patienten nacheinander einmal mit, einmal ohne die interessierende Behandlung zu beobachten. Abgesehen davon, dass dieser Weg nur bei reversiblen Endpunkten gangbar wäre, hat er 2 weitere Nachteile: Zum einen kann so die in dem Eingangssatz implizite Nebenbedingung, nämlich die identische Ausgangssituation, nicht gewährleistet werden (der Vergleich wäre nicht „*fair*“), zum anderen ist die so gewonnene Erkenntnis kaum allgemein, d. h. auch für andere Patienten brauchbar.

N-of-1-Trials oder Cross-over-Studien bieten hier weitergehende Möglichkeiten. Die kritischen Voraussetzungen und begrenzten Interpretationsmöglichkeiten führen jedoch dazu, dass die Anwendungsbereiche sehr eingeschränkt sind.

Das bestehende Dilemma lässt sich so auflösen, dass der notwendige Vergleich nicht innerhalb eines Patienten, sondern zwischen Patienten gemacht wird. Um eine gewisse Verallgemeinerbarkeit zu erreichen – die Ergebnisse sollen ja für zukünftige Patienten gelten – werden dabei größere Gruppen von Patienten betrachtet. Ein Teil der Patienten erhält die interessierende Behandlung, der andere Teil nicht, um so einen Vergleich der Ergebnisse zu ermöglichen. Um dem Ideal der identischen Ausgangssituation möglichst nahe zu kommen und einen fairen Vergleich zu ermöglichen, muss bestmöglich gewährleistet werden, dass sich zwischen den verglichenen Gruppen weder die Ausgangssituationen noch die Patienten in ihren Eigenschaften unterscheiden (Strukturgleichheit der Behandlungsgruppen).

So wie in anderen Wissenschaftsbereichen, aber auch im Alltag (z. B. bei Sportwettkämpfen), sind die Ansprüche an die Sicherstellung fairer Vergleiche umso höher, je geringer die erwarteten Unterschiede bzw. Effekte sind. In der modernen Medizin mit ihren heutzutage eher kleinen Fortschritten ist ein fairer Vergleich daher die zentrale Voraussetzung für Nutzenaussagen.

Aus diesen Überlegungen folgt, dass die alleinige Betrachtung einarmiger Studien oder einzelner Studienarme für die Nutzenbewertung nicht relevant ist. Damit entfällt für die weiteren Ausführungen in diesem Bericht der linke Strang der Übersicht von Studiendesigns in Abbildung 2, der Designs ohne Vergleich darstellt. Ausgehend von der Fragestellung der Nutzenbewertung sind ausschließlich vergleichende Studiendesigns geeignet.

5.3 Versorgungsnahe Daten in der Nutzenbewertung

Abhängig vom Studiendesign, das für die Generierung versorgungsnaher Daten zum Zwecke der Nutzenbewertung gewählt wird, entstehen unterschiedliche Anforderungen an die Durchführung und Auswertung der Studie. Abbildung 3 zeigt die Schritte von der Festlegung der Fragestellung der Nutzenbewertung bis zum Ergebnis der Untersuchung dieser Fragestellung.

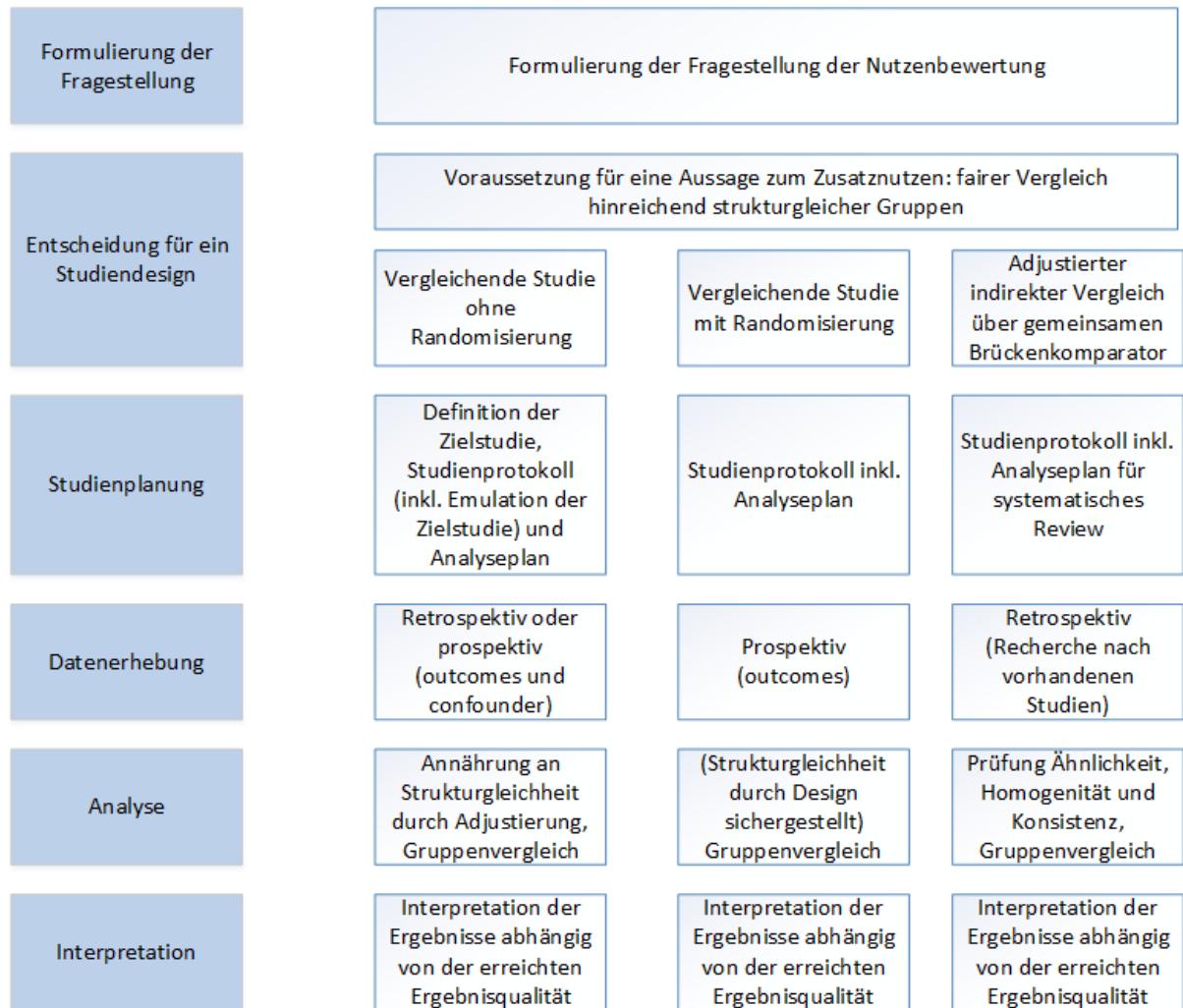


Abbildung 3: Ablauf von der Festlegung der Fragestellung der Nutzenbewertung bis zum Ergebnis der Untersuchung

Während generelle wissenschaftliche Prinzipien wie die Formulierung der zu beantwortenden Fragestellung oder die Interpretation der Ergebnisse unter Berücksichtigung der erreichten Ergebnisqualität unabhängig vom gewählten Studiendesign vorgenommen werden, unterscheiden sich andere Schritte der Nutzenbewertung je nach Studiendesign.

Die folgenden Kapitel beschreiben generelle und Studiendesign-spezifische methodische und prozedurale Anforderungen an die Studienplanung, Datenerhebung und -analyse sowie die Interpretation der Ergebnisse. Dabei werden u. a. folgende Empfehlungen berücksichtigt:

- Guidelines der International Conference on Harmonisation [11]
- Standards für die Berichterstattung verschiedener Studientypen des EQUATOR Netzwerks [12]
- Framework for FDA's Real-World Evidence Program [13]
- PCORI Methodology Standards [14]
- Cochrane Handbook for Systematic Reviews of Interventions [15]

5.3.1 Formulierung der Fragestellung und Entscheidung für ein Studiendesign

Generelle Anforderungen

Die Voraussetzung für die Generierung von versorgungsnahen Daten, die relevant zur Entscheidungsfindung des G-BA beitragen können, ist die Formulierung der Fragestellung, die durch die Datenerhebung beantwortet werden soll. Ausgangspunkt für die Fragestellung ist die Evidenzlücke, die geschlossen werden soll. Die Fragestellung muss die Elemente des PICO-Schemas abdecken. Das bedeutet, dass die relevante Patientenpopulation, die zu prüfende Intervention, der Komparator und die Endpunkte (Outcomes) definiert werden müssen.

Die spezifische Fragestellung ist gemäß den Anforderungen an eine Nutzenbewertung nach § 35a SGB V zu formulieren. Dabei müssen folgende Punkte berücksichtigt werden [16,17]:

- die Patientenpopulation entspricht der Zulassung des zu bewertenden Arzneimittels
- das zu bewertende Arzneimittel soll gemäß der Zulassung eingesetzt werden
- der Vergleichsarm der Studie bildet die (zweckmäßige) Vergleichstherapie ab
- bei den Endpunkten handelt es sich um patientenrelevante Endpunkte

Sollen die versorgungsnahen Daten im Anschluss an eine erste Nutzenbewertung erhoben werden, um Evidenzlücken zu schließen, sollte die Fragestellung diese Lücken explizit adressieren. Dabei können die Patientenpopulation, der Vergleichsarm oder die Endpunkte ggf. auf die noch nicht durch Evidenz abgedeckten Aspekte eingeschränkt werden.

Die Entscheidung für ein Studiendesign sollte berücksichtigen, ob mit dem jeweils gewählten Design ausreichend valide Ergebnisse für eine Nutzenbewertung erreicht werden können.

Besondere Anforderungen an vergleichende Studien ohne Randomisierung

Vergleichende Studien ohne Randomisierung sind unter bestimmten Bedingungen nicht durchführbar. So wird in der Literatur z. B. beschrieben, dass in Fällen, in denen der Krankheitsstatus oder das erwartete Ergebnis der Behandlung die Auswahl der Behandlung

stark beeinflussen, keine validen Effekte erwartet werden können [18]. Eine prospektive vergleichende Datenerhebung ist daher ggf. gar nicht möglich, wenn die Präferenz für eine Behandlungsoption (z. B. ein neues Arzneimittel) so stark ist, dass eine ausreichende Fallzahl für die Behandlung mit einer geeigneten Vergleichstherapie nicht in angemessener Zeit erreicht werden kann. Ein weiterer Faktor kann die erwartete Effektgröße sein. Je kleiner der erwartete Effekt ist, umso sinnvoller ist es, eine vergleichende Studie mit Randomisierung durchzuführen. Vor einer Entscheidung für eine vergleichende Studie ohne Randomisierung sollte daher geprüft werden, ob mit diesem Studiendesign ausreichend valide Ergebnisse erwartet werden können.

Im Fall der Entscheidung für eine vergleichende Studie ohne Randomisierung ist die Frage zu klären, ob mit einem historischen oder einem zeitlich parallelem Kontrollarm gearbeitet werden soll. Historische Kontrollen sind möglich, wenn die in der Vergangenheit beobachteten Patientenpopulationen hinreichend ähnlich zur aktuell behandelten Patientenpopulation sind und wenn in der Vergangenheit Daten in ausreichender Qualität für einen aussagekräftigen Vergleich gesammelt wurden. Darüber müssen aus dem historischen Datensatz auch die für die aktuelle Untersuchung relevanten spezifischen Daten verfügbar sein (z. B. patientenindividuelle Daten zu Confoundern), und die Daten müssen ausreichend ähnlich erhoben worden sein (z. B. ähnliche Endpunktdefinition). Diese Anforderung wird z. B. regelhaft zum Problem, wenn Populationen untersucht werden sollen, die durch bestimmte, in der Vergangenheit nicht erhobene Biomarker charakterisiert sind, oder wenn die Daten des historischen Datensatzes nicht für die notwendige Confounderadjustierung auf patientenindividueller Ebene zur Verfügung gestellt werden.

Sofern zu erwarten ist, dass das neue Arzneimittel nicht direkt nach Markteinführung parallel in allen Regionen oder Zentren Deutschlands zur Verfügung steht (z. B. aufgrund der Komplexität der Intervention und der damit verbundenen Anforderungen an eine qualitätsgesicherte Anwendung), kann ein prospektives Design mit zeitlich und / oder örtlich nicht parallelen Armen erwogen werden: Die Zuordnung zum neuen Arzneimittel wäre dann je nach Verfügbarkeit früher oder später in den jeweiligen Regionen bzw. Zentren möglich. Damit wäre die Zuordnung ggf. wesentlich durch die Verfügbarkeit des neuen Arzneimittels beeinflusst und nicht vorrangig durch Patientenkriterien.

Besondere Anforderungen an vergleichende Studien mit Randomisierung

Eine randomisierte Studie ist ggf. nicht notwendig, wenn sehr große (dramatische) Effekte erwartet werden oder es hinreichend sicher ist, dass der interessierende Endpunkt unter der zweckmäßigen Vergleichstherapie gar nicht erreicht wird (z. B. sustained virologic response [SVR] bei Hepatitis C, wenn keine aktiven Behandlungsoptionen zur Verfügung stehen, wie z. B. bei der Bewertung von Sofosbuvir bei vorbehandelten Jugendlichen mit Genotyp 2 oder 3 [19]).

Analog zur oben beschriebenen prospektiven Studie mit nicht parallelen Studienarmen ist ein solches Design auch randomisiert theoretisch denkbar (Stepped Wedge Design [20]).

Randomisiert wird dabei der Zeitpunkt, ab wann ein neues Arzneimittel an einem bestimmten Ort (Region, Zentrum) verordnet werden kann. Für die Nutzenbewertung von Arzneimitteln sind solche Studien aufgrund der grundsätzlichen Verordnungsfähigkeit neuer Arzneimittel bei Markteinführung jedoch weniger realistisch.

Besondere Anforderungen an adjustierte indirekte Vergleiche über einen gemeinsamen Brückenkomparator

Vor einer Entscheidung für einen adjustierten indirekten Vergleich über einen gemeinsamen Brückenkomparator sollte geprüft werden, ob prinzipiell Studien zur Verfügung stehen, die einen solchen Vergleich ermöglichen (vergleichende Studien mit Randomisierung). Da eine Bewertung auf Basis versorgungsnaher Daten angestrebt wird, sollten die einzuschließenden Studien die in Abschnitt 5.1 genannten Kriterien für versorgungsnaher Daten erfüllen. Ergibt die Recherche nach relevanten Studien für einen adjustierten indirekten Vergleich, dass die Studienlage unzureichend ist und dass die Lücke absehbar nicht geschlossen werden kann (siehe 5.3.3), muss ein anderes Design zur Bearbeitung der Fragestellung gewählt werden.

5.3.2 Studienplanung

Generelle Anforderungen

Unabhängig vom gewählten Studiendesign muss ein Studienprotokoll entwickelt und finalisiert werden. Die Sicherstellung einer adäquaten Studienplanung ohne Kenntnis der Daten ist bei prospektiven Studiendesigns einfacher sicherzustellen als bei retrospektiven. Bei prospektiven Studien ist generell die Anforderung der Finalisierung eines Studienprotokolls vor Einschluss der ersten Patientin oder des ersten Patienten in die Datenerhebung etabliert. Auch bei retrospektiven Datenerhebungen soll das Studienprotokoll vor Zugriff auf die Daten finalisiert werden. Ebenfalls unabhängig vom Studiendesign sind Änderungen der Studienplanung in Protokollamendments zu dokumentieren. Diese Änderungen sollen dabei entweder in Unkenntnis der Daten oder im Rahmen eines adaptiven Designs mit entsprechender statistischer Absicherung erfolgen.

Vor Beginn der Datenerhebung sollte die Studie in einem Studienregister registriert werden, eine solche Registrierung ist unabhängig vom Studiendesign z. B. bei ClinicalTrials.gov möglich [21].

Die Studienplanung umfasst auch die Planung der statistischen Analysen der Daten. Fallzahl und Beobachtungsdauer müssen so festgelegt werden, dass die Studie in der Lage ist, die Fragestellungen der Nutzenbewertung zu beantworten. Die Fertigstellung eines Analyseplans vor Kenntnis der Daten ist essenziell und daher nachvollziehbar zu dokumentieren.

Eine adäquate und auch in ihrem zeitlichen Ablauf nachvollziehbare Studienplanung ist für die Aussagekraft der Ergebnisse einer Studie von entscheidender Bedeutung. Das Protokoll inkl. der Amendments und des Analyseplans muss deshalb Teil des Berichts über die Studienergebnisse sein (siehe Abschnitt 5.5.3). Es wird empfohlen, die Studienplanung durch

Veröffentlichung des vollständigen Studienprotokolls und Analyseplans (z. B. in einem Studienregister) transparent zu machen.

Das Studienprotokoll (und der abschließende Studienbericht) soll beschreiben, wie sichergestellt wird, dass die für die Studie ausgewählten Patientinnen und Patienten den Anforderungen an die „Versorgungsnähe“ innerhalb der zu untersuchenden Fragestellung gerecht werden und wie die Studienteilnehmer identifiziert, ausgewählt, eingeschlossen und in der Studie gehalten werden, um Selektionsbias zu minimieren. Auch die Methoden mit denen eine unverzerrte und systematische Datenerhebung von allen Studienteilnehmern erreicht werden soll, sollen dargestellt werden.

Besondere Anforderungen an vergleichende Studien ohne Randomisierung

Für die Planung von vergleichenden Studien ohne Randomisierung zur Ermittlung von Therapieeffekten wird die explizite Nachbildung der Planung von vergleichenden Studien mit Randomisierung empfohlen (emulation of target trials [22,23]). Dabei wird zunächst ein Konzept für eine randomisierte Studie definiert, die für die Untersuchung der gegebenen Fragestellung geeignet wäre. In einem nächsten Schritt wird angestrebt, die Eigenschaften dieser randomisierten Studie aus dem Datensatz der Beobachtungsstudie nachzubilden.

Die folgende Tabelle stellt die Komponenten der Studienplanung für die Nachbildung einer Zielstudie dar.

Tabelle 5: Komponenten der Nachbildung der Zielstudie aus einem nicht randomisierten Datensatz

Protokollkomponente	Anforderung an den gewählten Datensatz
Einschlusskriterien für Patientinnen und Patienten	Die Überprüfung der verschiedenen Ein-/Ausschlusskriterien muss auf Basis der in der gewählten Datenquelle vorliegenden Daten möglich sein (sollen Patientinnen und Patienten z. B. in den 2 Jahren vor Einschluss in die Studie bestimmte Therapien nicht erhalten haben, müssen entsprechende Beobachtungsdauern und Informationen zu eingesetzten Therapien vorliegen).
Zu untersuchende Behandlungen	Patientinnen und Patienten, die die Ein-/Ausschlusskriterien erfüllen, werden der Intervention zugewiesen, die sie zu Beginn ihrer Behandlung der zu untersuchenden Erkrankung bzw. Indikation (z. B. Zweitlinientherapie) erhielten (new user design).
Behandlungszuweisung	Um sich der für den Vergleich notwendigen Strukturgleichheit anzunähern (die in diesem Design nicht durch Randomisierung erreicht wird) ist die Adjustierung für die relevanten Confounder notwendig. Falls der Datensatz der nicht randomisierten Studie unzureichende Informationen über die relevanten Confounder enthält, ist die Nachbildung einer randomisierten Behandlungszuweisung und damit die Annäherung an einen fairen Vergleich nicht möglich. In diesem Fall sind die Ergebnisse einer Studie für die Nutzenbewertung in der Regel nicht aussagekräftig.
Beobachtungsdauer (einschließlich eindeutiger Definition des Beobachtungsbeginns)	Zu Beginn der Beobachtung (in der Zielstudie wie in der nicht randomisierten Studie) sollen 3 Bedingungen erfüllt sein: die Ein- und Ausschlusskriterien sind erfüllt, die Behandlung beginnt und die Erfassung der Endpunkte beginnt. Diese Bedingungen müssen aus dem genutzten Datensatz ableitbar sein. Dazu ist z. B. notwendig, dass Datumsangaben für den Behandlungsbeginn und die Erfassung der Endpunkte vorliegen. Der Datensatz soll die Endpunkte über die geplante Beobachtungsdauer enthalten.
Endpunkte	Die relevanten Endpunkte sollen im Datensatz enthalten sein.
Interessierender Vergleich	Der interessierende Vergleich soll spezifisch präspezifiziert werden (z. B. Intention-to-treat Analyse der prädefinierten Endpunkte nach 2 Jahren).
Analyseplan	Ein Analyseplan soll präspezifiziert werden und die Abbildung der geplanten Analysen aus dem Datensatz soll möglich sein (z. B. Erfassung der Endpunkte unabhängig von Änderungen der Therapie während der Beobachtungsdauer, um eine Intention-to-treat Analyse zu ermöglichen).

Nach diesem Konzept muss der Datensatz, der für die vergleichende Studie ohne Randomisierung vorgesehen ist, die notwendigen Informationen enthalten, um die Zielstudie nachzubilden.

Wesentlich für die Datenanalyse von Studien ohne Randomisierung sind Adjustierungen, die den Einfluss der Struktungleichheit der Behandlungsgruppen ausgleichen sollen. Um eine ergebnisgesteuerte Analyse zu vermeiden, müssen die relevanten Confounder und der Ablauf der Adjustierung während der Analyse in der Studienplanung nachvollziehbar und in der notwendigen Detailtiefe präspezifiziert werden (siehe auch Abschnitt 5.5.2).

Besondere Anforderungen an vergleichende Studien mit Randomisierung

Die Anforderungen an die Planung der Durchfhrung und Analyse von randomisierten Studien sind etabliert (z. B. [11]). Diese Anforderungen sind auch fr die Erhebung versorgungsnaher Daten mit vergleichenden Studien mit Randomisierung umzusetzen. Um eine versorgungsnaher Datenerhebung sicherzustellen, sind insbesondere die Ein-/Ausschlusskriterien fr Patientinnen und Patienten, die Anwendung der Interventionen und die Durchfhrung von Visiten an den Behandlungsalltag anzupassen. Randomisierte Studien, die diese Anforderungen erfllen, werden als pragmatische randomisierte Studien bezeichnet [24-26]. Instrumente, die die Planung pragmatischer randomisierter Studien unterstützen, stehen zur Verfgung [27-29].

Besondere Anforderungen an adjustierte indirekte Vergleiche über einen gemeinsamen Brückenkomparator

Ein adjustierter indirekter Vergleich über einen gemeinsamen Brückenkomparator greift auf vorhandene randomisierte Studien zurück. Um die Bewertung des Arzneimittels auf Basis versorgungsnaher Daten zu ermöglichen, sollten die zugrundeliegenden Studien des indirekten Vergleichs darauf angelegt sein, versorgungsnaher Daten zu erheben.

Der Studienpool für einen solchen Vergleich muss durch eine systematische Literaturbeschaffung zusammengestellt werden. Die Studienplanung muss der eines systematischen Reviews entsprechen, der Analyseplan die besonderen methodischen Anforderungen an indirekte Vergleiche erfüllen. Entsprechende Standards liegen vor [15,30-33].

5.3.3 Datenerhebung

Generelle Anforderungen

Bei der Auswahl des Instruments zur Datenerhebung ist zu berücksichtigen, ob das gewählte Instrument die für die Beantwortung der Fragestellung notwendigen Daten (Komponenten des PICO) in ausreichender Qualität zur Verfügung stellen kann.

Für Studien, die Daten aus verschiedenen Datenquellen nutzen (data linkage), ist eine detaillierte Beschreibung der genutzten Datenquellen, der notwendigen Informationen für das Gelingen der Zusammenführung der Daten sowie der verwendeten Algorithmen und ihrer Überprüfung notwendig [34]. Vor der Nutzung von Daten aus verschiedenen Datenquellen ist zu prüfen und nachvollziehbar zu dokumentieren, ob eine Zusammenführung überhaupt möglich ist oder ob z. B. Unterschiede in der Definition von Datenpunkten einer gemeinsamen Nutzung entgegenstehen.

Für jede Datenquelle muss anhand geeigneter Werkzeuge sichergestellt sein, dass die zu erhebenden Daten ausreichend valide (z. B. Plausibilitätsanalysen, automatisierte Überprüfungen, source data verification, Verwendung von Standardklassifikationen und -terminologien, Möglichkeit zu einem Audit) und integer (z. B. Verfügbarkeit von log-files) sind.

Besondere Anforderungen an vergleichende Studien ohne Randomisierung

Für die Analyse vergleichender Studien ohne Randomisierung sind zusätzlich zu den Komponenten des PICO Daten für die Confounderkontrolle notwendig. Die relevanten Confounder müssen systematisch identifiziert (z. B. auf Basis wissenschaftlicher Literatur unter Einbindung von Fach-Experten) und in der Studienplanung präspezifiziert werden (siehe Abschnitte 5.3.2 und 5.5.2). Die Verfügbarkeit entsprechender Daten in der gewählten Datenquelle ist vor der Entscheidung für eine nicht randomisierte Studie sicherzustellen. Eine Adjustierung lediglich für die im Datensatz verfügbaren Confounder ist nicht ausreichend, wenn diese nicht die relevanten Confounder abdecken.

Bei der Durchführung einer vergleichenden Studie ohne Randomisierung besteht die Möglichkeit, die Daten retrospektiv oder prospektiv oder kombiniert teils retrospektiv und teils prospektiv zu erheben. Die retrospektive Datenerhebung ist nur sinnvoll, wenn der Datenkörper, auf dessen Grundlage die retrospektive Datenerhebung erfolgen soll, die notwendigen Daten in der erforderlichen Qualität enthält. Die Verfügbarkeit der relevanten Daten ist vor der Entscheidung für ein retrospektives Design sicherzustellen.

Sind die notwendigen Daten nicht in ausreichendem Umfang oder in ausreichender Qualität verfügbar, ist eine prospektive Datenerhebung notwendig. Dabei kann ggf. auf bestehende Datenquellen zurückgegriffen werden (z. B. indikationsspezifisches klinisches Register), indem der Datensatz in der prospektiven Erhebung um die fehlenden Daten (z. B. einzelne Endpunkte) ergänzt wird.

Besondere Anforderungen an vergleichende Studien mit Randomisierung

In einer randomisierten Studie werden die Daten grundsätzlich prospektiv erhoben. Soll die randomisierte Studie in einer bestehenden Datenstruktur durchgeführt werden (z. B. in einem Register), ist zu prüfen, ob diese Datenstruktur die Erhebung der notwendigen Informationen vorsieht. Ist das nicht der Fall, kann die Erhebung in der Datenstruktur erweitert werden oder die fehlenden Informationen werden für die spezifische Studie parallel erhoben (studienindividuelle Datenerhebung).

Besondere Anforderungen an adjustierte indirekte Vergleiche über einen gemeinsamen Brückenkomparator

Ein adjustierter indirekter Vergleich über einen gemeinsamen Brückenkomparator greift auf vorhandene Studien zurück. Zeigt die Recherche nach Studien für einen solchen Vergleich, dass die Studienlage unzureichend ist, besteht ggf. die Möglichkeit, die Datenlücken innerhalb des indirekten Vergleichs durch zusätzliche noch durchzuführende Studien zu schließen. In der Regel ist in einem solchen Fall aber die Durchführung einer direkt vergleichenden Studie vorzuziehen, da diese eine höhere Aussagesicherheit erreichen kann.

Unter bestimmten Umständen kann die Evidenz für einen adjustierten indirekten Vergleich durch Re-Analyse vorhandener Studien erweitert werden, z. B. wenn zwar nicht die gesamte Studienpopulation, aber eine Subpopulation vorhandener Studien hinreichend ähnlich für den

indirekten Vergleich wäre (siehe z. B. [35]). Dabei sollte nicht nur auf eigene individuelle Patientendaten zurückgegriffen werden. Wenn notwendig sollten entsprechende Analysen auch von anderen Studiensponsoren angefordert werden. Diese Möglichkeit sollte geprüft werden, auch wenn das dazu führt, dass ggf. nur eine Teilfragestellung bearbeitet werden kann.

5.3.4 Analyse und Interpretation der Ergebnisse

Generelle Anforderungen

Die Präspezifikation der geplanten Analysen mit dem Ziel, verlässliche Antworten auf die bestehenden Fragestellungen zu bekommen und ergebnisgesteuerte Auswertungen zu verhindern, gehört zu den Standards hochwertiger Studien [11,14,15]. Das schließt auch die Beschreibung möglicher Anpassungen des Analyseplans ein.

Bei der Interpretation der Ergebnisse sind die Aussagekraft der verschiedenen Studiendesigns und die konkrete Datenqualität zu berücksichtigen. Innerhalb eines jeden Studiendesigns geht zusätzlich die Ergebnissicherheit der konkreten Studienergebnisse in die abschließende Aussagesicherheit der Ergebnisse zur Beantwortung der Fragestellung ein [36].

Die versorgungsnahen Daten müssen dazu geeignet sein, die Fragestellung der Nutzenbewertung zu beantworten. Sofern die konkreten Daten von einzelnen Aspekten der Fragestellung abweichen (z. B. andere Therapielinie, Dosierung oder Mutationsstatus), ist zu begründen, warum trotz dieser Abweichungen valide Aussagen zur Beantwortung der Fragestellung möglich sind. Versorgungsnahe Daten sind für die Nutzenbewertung geeignet, wenn Abweichungen im Hinblick auf die Eckpunkte der Fragestellung (PICO) für die Effektschätzung nicht relevant sind.

Besondere Anforderungen an vergleichende Studien ohne Randomisierung

Da in Studien ohne Randomisierung die Strukturgleichheit der Behandlungsgruppen wegen der fehlenden zufälligen Zuteilung nicht per se sichergestellt ist, wird durch Adjustierung versucht, sich einer Strukturgleichheit anzunähern.

Die Methoden zur Annäherung an die Strukturgleichheit der Daten in den Behandlungsgruppen gehen von bestimmten Annahmen aus. Diese Annahmen sind zu beschreiben und, wo möglich, zu überprüfen. Die Ergebnisse dieser Überprüfung sowie nicht überprüfbare Annahmen sind zu dokumentieren und in ihrer Auswirkung auf die Interpretation der Studienergebnisse einzuschätzen (siehe Abschnitt 5.5.2).

Da Analysen auf Basis retrospektiver Datenerhebungen besonders anfällig für ergebnisgesteuerte Auswertungen und Berichterstattung sind (siehe Abschnitt 5.5.3), bestehen besondere Anforderungen an die Dokumentation der Planung der Analysen dieser Daten. Dabei soll soweit möglich nachvollziehbar sichergestellt werden, dass die Planung der Analysen vor Sichtung der Daten erfolgt ist.

Besondere Anforderungen an vergleichende Studien mit Randomisierung

In Studien mit Randomisierung ist die Strukturgleichheit der Behandlungsgruppen als eine Voraussetzung für einen fairen Vergleich durch das Design sichergestellt. Ob spezifische Aspekte des Studiendesigns einen fairen Vergleich infrage stellen (z. B. unterschiedliche Therapieziele in den Studienarmen [37]), ist auch bei Studien mit Randomisierung zu prüfen.

Besondere Anforderungen an adjustierte indirekte Vergleiche über einen gemeinsamen Brückenkomparator

Die Voraussetzung für aussagekräftige adjustierte indirekte Vergleiche über einen gemeinsamen Brückenkomparator sind die Ähnlichkeit der eingeschlossenen Studien, die Homogenität der Ergebnisse der Einzelstudien, die in die Analyse eingehen, und die Konsistenz der geschätzten Effekte aus direkter und indirekter Evidenz [38,39]. Diese Annahmen sind zu überprüfen, Zutreffen oder Verletzung der Annahmen ist nachvollziehbar zu dokumentieren. Bestehen große Zweifel, dass eine oder mehrere der grundlegenden Annahmen in ausreichendem Maße erfüllt sind, so sollte von der Anwendung indirekter Vergleiche abgesehen werden [40]. Unsicherheiten bezüglich der Annahmen sollen mithilfe von Sensitivitätsanalysen untersucht werden. Die abschließende Aussagesicherheit adjustierter indirekter Vergleiche über einen gemeinsamen Brückenkomparator hängt vom Verzerrungspotenzial der eingeschlossenen Studien und der Erfüllung der oben genannten Annahmen ab [36].

5.4 Instrumente zur Datenerhebung

Die im Folgenden beschriebenen Instrumente zur Datenerhebung können jeweils für unterschiedliche Fragestellungen eingesetzt werden. Fokus des vorliegenden Berichts ist jedoch die Nutzenbewertung von Arzneimitteln. Die Diskussion der Instrumente zur Datenerhebung bezieht sich auf deren Eignung für die Fragestellung der Nutzenbewertung und adressiert deshalb den Einsatz in vergleichenden Studien.

5.4.1 Studienindividuelle Datenerhebung

Eine studienindividuelle Datenerhebung kann für versorgungsnahe vergleichende Studien mit und ohne Randomisierung herangezogen werden.

Der Vorteil einer studienindividuellen Datenerhebung ist, dass die zu erhebenden Daten auf den Bedarf der Studie zugeschnitten werden können. Standards für die studienindividuelle Datenerhebung sind etabliert. Diese Art der Datenerhebung wird deshalb im vorliegenden Bericht nicht weiter ausgeführt.

In der Literatur wird die Kombination studienindividueller Datenerhebung mit der Nutzung anderer Erhebungsinstrumente wie z. B. Registern oder elektronischen Patientenakten diskutiert [41]. Dabei sollen die Lücken, die in Registern oder elektronischen Patientenakten bestehen, durch eine zusätzliche studienindividuelle Datenerhebung geschlossen werden.

5.4.2 Studien auf Basis einer Datenerhebung in einem Register

Die Datenerhebung in einem Register kann für versorgungsnahe vergleichende Studien mit und ohne Randomisierung herangezogen werden.

In den vergangenen Jahren wurden Zielsetzung und Dokumentationsumfang von Registern erweitert [3]. Für den vorliegenden Bericht ist insbesondere die zunehmende Dokumentation klinischer Informationen in Registern zur Erfassung von Patientinnen und Patienten mit spezifischen Erkrankungen relevant, die für die Beschreibung von Patientenpopulationen, Interventionen (einschließlich Komparatoren) und Endpunkten (PICO) für die Nutzenbewertung herangezogen werden können. Register stellen damit voraussichtlich mehr relevante Informationen für die Nutzenbewertung zur Verfügung als elektronische Patientenakten oder GKV-Routinedaten (siehe Abschnitte 5.4.3, 5.4.4 und 5.5.1.2).

Konzeptionell ist es wichtig, zwischen Registern (aktive, prospektive, standardisierte Dokumentation von Beobachtungseinheiten zu vorab festgelegten, aber im Zeitverlauf erweiterbaren Fragestellungen, für die ein präziser Bezug zur Zielpopulation darstellbar ist [3]) und Studien in diesen Registern (Registerstudien) zu unterscheiden. Während das Register (im besten Fall) als dauerhaft verfügbare und fortgeführte Infrastruktur und Dokumentation besteht, werden Registerstudien anlässlich einer spezifischen Fragestellung, z. B. zur Nutzenbewertung eines neuen Arzneimittels, mithilfe des Registers durchgeführt.

Grundsätzlich sind in Registern nicht interventionelle und interventionelle vergleichende Studien möglich. In Registern können sowohl vergleichende Studien ohne Randomisierung als auch mit Randomisierung durchgeführt werden. Während aktuell die Mehrzahl der vergleichenden Registerstudien ohne Randomisierung vorgenommen wird, werden Registerstudien mit Randomisierung als wichtiges Instrument für die effiziente Erhebung verlässlicher Evidenz beschrieben [42-44]. Eine Reihe von Beispielen zeigen das Potenzial dieser Anwendung von Registern [44-48].

Da Registerstudien für die versorgungsnahe Datenerhebung mit dem Ziel Evidenz für die Nutzenbewertung von Arzneimitteln zu generieren von besondere Bedeutung sind, konzentriert sich die im Auftrag des G-BA enthaltene Spezifizierung von Qualitätskriterien für die versorgungsnahe zu erhebenden Daten auf Register (siehe Abschnitt 5.5.1).

5.4.3 Studien auf Basis einer Datenerhebung in elektronischen Patientenakten

Die Datenerhebung in elektronischen Patientenakten ist für versorgungsnahe vergleichende Studien mit und ohne Randomisierung denkbar. Eine systematische Nutzung dieser in der klinischen Routineversorgung digital dokumentierten Behandlungs- und Befunddaten für Forschungszwecke wird von manchen als Zukunftsmodell für ein „lernendes Gesundheitswesen“, speziell aber auch für die Arzneimittelentwicklung, Zulassung, Nutzenbewertung und Pharmakovigilanz gesehen [49]. Verglichen mit Studien mit primärer Datenerhebung erhofft man sich von Analysen auf Basis elektronischer Patientenakten eine

kostengünstigere und schnellere Durchführung von längsschnittigen Auswertungen an wenig selektierten Populationen [50]. Andererseits werden Daten aus elektronischen Patientenakten nicht zu Forschungszwecken, und damit weniger sorgfältig, standardisiert und vollständig dokumentiert. Auf ihnen basierende Analysen sind zudem in unterschiedlichem Ausmaß anfällig für diverse Formen von Bias und Confounding [51-53].

Die Möglichkeiten der Nutzung von Daten aus elektronischen Patientenakten werden auch im Rahmen von Zulassungsentscheidungen von Arzneimitteln und Medizinprodukten diskutiert. Es gibt deshalb inzwischen verschiedene regulatorische Empfehlungen dazu [54-56]. Sie adressieren diverse Fragen der Compliance mit den Standards der Good clinical Practice (GCP), die sich in teilweise neuartiger Weise durch die Verwendung elektronischer Patientenakten für die Transparenz und Qualität der Forschung sowie die Wahrung von Datenschutz und -sicherheit ergeben.

Definition

Unter einer elektronischen Patientenakte werden international immer noch unterschiedliche Dinge verstanden. Haas 2017 [57] spricht von einer „historischen Begriffsverwirrung“, die mit einer Vielfalt von Bezeichnungen und unterschiedlichen Definitionsansätzen einher geht. So wird z. B. entweder eine institutionelle digitale Akte einer Versorgungseinrichtung (z. B. Krankenhaus, Arztpraxis) oder eine einrichtungsübergreifende Akte zu einem Patienten darunter verstanden [58]. Auch unterschiedlich sind die Auffassungen darüber, welche Struktur, Inhalte und Funktionalitäten eine solche Patientenakte hat, welche Zugriffsmöglichkeiten bestehen sollen, wer die Akte füllt oder wer sie führt, z. B. der Patient bzw. die Patientin selbst, seine bzw. ihre medizinischen Betreuer und Versorgungseinrichtungen oder die Krankenversicherung. Zunehmend etablieren sich aber nationale Konzeptionen für übergreifende elektronische Patientenakten (engl. electronic health records, EHR), die aus klinischen Routinedokumentationen in digitalen Akten von einzelnen Gesundheitseinrichtungen (engl. electronic medical records, EMR) gespeist werden.

Einrichtungsübergreifende elektronische Patientenakten (EHR) haben als Datenbankanwendungen zum Ziel, kontinuierlich fall- und sektorübergreifend patientenbezogene Daten zu sammeln bzw. zu verknüpfen, und zwar zu Krankengeschichte, Behandlungen, Medikation, Befunden, speziellen Risiken und weiteren Gesundheitsmerkmalen. Sie können für große Patientenpopulationen in elektronischen EHR-Plattformen zu Analyse- und Forschungszwecken verknüpft oder aggregiert werden.

Elektronische Patientenakten variieren auf einem großen Spektrum von Eigenschaften. Es kann sich im einfachen Fall um ein auf eine Versorgungseinrichtung begrenztes digitales Archiv weniger Dokumente im PDF-Format, ergänzt um Röntgenbilder, eines Patienten handeln, das nur schlecht durchsuch- und auswertbar ist. Oder es besteht im Idealfall eine langfristige, gut verknüpf- und auswertbare Akte mit einer umfassenden, hoch formalisierten und strukturierten Sammlung validierter Daten, die nach Standardterminologien granular kodiert worden sind,

und die als integrales Element in einem Patientenaktensystem auch verschiedene Funktionalitäten für Patienten, medizinische Versorger und Forscher bietet.

Internationale und deutsche Situation

Eine im Auftrag der OECD erstellte Expertise [59] hat kürzlich gezeigt, dass vermutlich nur 10 von 28 weltweit analysierten Nationen (ohne Deutschland) mittels ihrer elektronischen Patientenaktenlösungen absehbar in der Lage sein werden, solide Grundlagen für Gesundheitsberichterstattung und medizinische Forschung zur Verfügung zu stellen. Dies liegt unter anderem an mangelnden organisatorisch-politischen und technischen Voraussetzungen, Datenschutzregelungen, zu geringer Flächendeckung, lückenhafter und zu wenig standardisierter Datenerhebung oder unzureichender Interoperabilität der eingesetzten IT-Lösungen. Untersuchungen von EMA-Mitarbeitern mit Fokus auf Zwecke der Arzneimittelzulassung [60] zeigen, dass auch auf elektronischen Patientenakten fußende Datenbestände in Europa sehr heterogen sind, sie aktuell nur selten den regulatorischen Anforderungen genügen, es an transparenter Information über sie mangelt und sie ungleich verteilt überwiegend in Zentral-, West- und Nordeuropa existieren. Damit schränkt sich für die Nutzenbewertung von Arzneimitteln die Zahl theoretisch zukünftig verfügbarer nationaler EHR-Plattformen deutlich ein.

Seit einigen Jahren organisieren sich international auch unterschiedliche Verbünde von Industrie, Forschung, Patientenorganisationen und Gesundheitsbehörden. Sie wollen speziell für Zwecke von Forschung und Arzneimittelentwicklung auf elektronischen Patientenakten basierenden Plattformen neue Möglichkeiten schaffen [61-63]. Zu den Zielen gehören unter anderem verbesserte Forschungsplanung, Patientenrekrutierung, Studiendurchführung und Generalisierbarkeit von Ergebnissen durch größere Nähe der Forschung zur klinischen Routineversorgung.

In Deutschland gibt es die Vorgabe für die gesetzlichen Krankenkassen, ihren Versicherten bis 2021 eine elektronische Patientenakte anzubieten, wofür seit geraumer Zeit Vorbereitungen und Abstimmungsprozesse laufen. Allerdings werden von Experten [57,64] zahlreiche Hindernisse für die Digitalisierung des deutschen Gesundheitswesens und relevante Hürden für die Realisierung des Projekts der nationalen elektronischen Patientenakte identifiziert. Dabei kommt auch der im europäischen Vergleich gesehen unterdurchschnittlichen Digitalisierung im Krankenhausbereich eine große Bedeutung zu [65].

Da die geplante nationale Lösung für die elektronische Patientenakte vorsieht, dass der Patient selbst festlegt, was an Daten bzw. welche Behandlungsepisoden in der Akte gespeichert werden sollen und wer außer ihm selbst noch Einsichtsrechte erhält, stellt sich mit Blick auf die Zukunft die Frage, wie viele Patientenakten a) vollständig genug für Forschungszwecke sein und b) generell von Versicherten dafür verfügbar gemacht werden? Gewiss scheint, dass die für Deutschland vorgesehene elektronische Patientenakte aktuell und vermutlich auch mittelfristig keine Optionen bieten wird, versorgungsnahe Datenerhebungen zur Bestimmung des Zusatznutzens von Arzneimitteln durchzuführen.

Auch in Deutschland existieren aktuelle Forschungsinitiativen zur patientenbezogenen Verknüpfung institutioneller elektronischer Patientenakten mit weiteren digitalen Datenbeständen z. B. aus Biobanken, von gesetzlichen Krankenkassen oder aus digitalen Gesundheitsanwendungen [66]. Ob und wann aus diesen Initiativen zuverlässige Möglichkeiten erwachsen, dass mit den gebildeten Datenbasen versorgungsnahe Datenerhebungen durchgeführt werden können, ist derzeit nicht absehbar.

Validität von Daten aus elektronischen Patientenakten

Es ist bekannt, dass vor dem Hintergrund der Vielgestaltigkeit von elektronischen Patientenakten auch deren Datenqualität sehr heterogen ist und diese von vielen Faktoren beeinflusst wird [67,68]. Daher wird für sekundäre Nutzungen solcher Daten wie die klinische Forschung geraten, in jedem Einzelfall die Datenqualität und Eignung für den jeweiligen Analysezweck („fit for purpose“) gründlich zu prüfen [41,68]. Weitverbreitet findet sich auch die Forderung, die Validität der Daten aus elektronischen Patientenakten wie auch anderer sogenannter „Real World Data“ systematisch in eigens aufgelegten Studien(-programmen) zu überprüfen und hierfür Standards zu entwickeln [60,69-72]. Allerdings wird die bisherige Transparenz über die Validität und Qualität der zum Beispiel in RCTs genutzten Daten aus elektronischen Patientenakten als völlig unzureichend betrachtet [73].

Eingaben in digitale Patientenakten werden im Alltag oft unter hoher Arbeitsbelastung von einer großen Zahl von Ärzten und Pflegekräften gemacht, die sich nicht primär Forschungsanliegen und den damit verbundenen Anforderungen bezüglich der Datenqualität verpflichtet fühlen. So kann eine Auswertung wie die von Brennan 2012 [74] zustande kommen, nach der die englische Krankenhausepisodenstatistik (HES), die auf elektronischen Klinikdokumentationen beruht, jährlich 1600 Fälle von Erwachsenen mit Inanspruchnahme von Diensten der Kinder- und Jugendpsychiatrie kodiert wurden, dazu 3000 Fälle von 0 bis 19-Jährigen mit Versorgung in ambulanten geriatrischen Einrichtungen und 17 000 Männer mit stationären Einweisungen zur Entbindung in gynäkologische Abteilungen. Green 2013 [75] untersucht in den USA aus einer nationalen Datenbank zu Krankenhausbehandlungen, die auf Angaben aus digitalen Patientenakten beruht, 875 Datensätze zu Notfallbehandlungen von bei Aufnahme intubierten Patienten. Er stellt in 27 % der Fälle Implausibilitäten fest, die nach Abgleich mit Textnotizen vermutlich zumeist auf Kodierfehlern beruhen, denn 81 intubierte Patientinnen und Patienten wurden nach den Daten sofort wieder nach Hause entlassen und 153 unmittelbar auf eine Normalstation verlegt.

Beim gemeinsamen Versuch von 4 Gesundheitsversorgern in den USA, mit klinischen Daten aus ihren elektronischen Patientenakten eine Vergleichsstudie zur Wirksamkeit verschiedener medikamentöser Bluthochdrucktherapien zu simulieren, konnten in einer qualitativen Analyse 5 Hauptprobleme mit den extrahierten Daten ermittelt werden [76]. Danach wäre die Validität der Studie durch fehlende, inkorrekte und nicht interpretierbare Daten ebenso beeinträchtigt worden wie durch nicht auswertbare wichtige Informationen in unstrukturierten Textdaten sowie Dateninkonsistenzen über die Zeit und zwischen beteiligten Einrichtungen.

Nach einer systematischen Recherche zu Methoden der Qualitätsbewertung von Daten aus elektronischen Patientenakten [68] existiert derzeit noch kein konsentierter Standard dafür, wie man die 5 aus dieser Literatursichtung abgeleiteten Qualitätsdimensionen „Vollständigkeit“, „Korrektheit“, „Konsistenz“, „Plausibilität“ und „zeitnahe Verfügbarkeit“ systematisch und mittels statistischer Methodik bestimmen soll. Infrage kommen dafür unter anderem Vergleiche mit einem Goldstandard oder anderen Datenquellen, datensatzinterne Validierungen, Vollständigkeits-, Verteilungs- und Plausibilitätsprüfungen sowie Log-Prüfungen zur Dateneingabe. Die Vereinbarung methodischer Standards für Qualitätsprüfungen sei den Autoren zufolge unbedingt notwendig, wenn man diese Daten verbreitet für klinische Forschungszwecke nutzen wollte. In den von ihnen identifizierten Studien seien hingegen intuitive und ad hoc-Prüfungen der Datenqualität verbreitete Praxis gewesen. Einen guten Ausgangspunkt für systematische Qualitätsprüfungen elektronischer Datenbasen stellt die Guideline der International Society of Pharmacoepidemiology dar [77].

Ein großes Problem der elektronischen Patientenakten ist, dass viele wichtige Informationen für Studien (z. B. Einschlusskriterien, Outcomes) nur in Textfeldern in nicht standardisierter Weise dokumentiert sind [78]. Manuelle Auswertungen werden als zu aufwendig betrachtet, nicht zuletzt, weil in den Befundnotizen der Mediziner eine große Zahl unterschiedlicher Begrifflichkeiten für dieselben Sachverhalte genutzt werden. Wassermann 2011 [79] berichtet zum Beispiel, dass an einem Tag im Kinderkrankenhaus von Philadelphia die Ärzte bei Kindern mit Otitis media in 465 Notizen in der Patientenakte 278 verschiedene Arten zur Angabe von Fieber (z. B. Fieber, Pyrexie, 39,2 Grad, erhöhte Temperatur) und 123 verschiedene Formulierungen zur Angabe von Ohrenschmerzen bei 213 Patienten genutzt haben. Solche Daten zur Kodierung in ausreichender Qualität mit Text-Mining-Anwendungen und Natural Language Processing (NLP) Werkzeugen automatisiert aus den Freitextnotizen zu extrahieren, bleibt sicher noch auf absehbare Zeit eine große Herausforderung. Dies zeigt auch eine aktuelle deutsche Publikation [80] für die automatisierte Auswertung von CT-Befunden im Rahmen einer epidemiologischen Studie. Im Fazit wird unter dem Aspekt der Datenqualität für künftige Studien ein gemeinsamer Analyse- und Extraktionsprozess von menschlichen Experten und Unterstützung durch ein algorithmusbasiertes Auswertungsprogramm empfohlen.

PICO

Die Fragestellungen im Rahmen der Nutzenbewertungen des G-BA gliedern sich regelhaft nach dem PICO-Schema, bei dem eine möglichst präzise Beschreibung der interessierenden Patientenpopulation, der zu bewertenden Intervention und ihrer Vergleichsintervention (Comparator) sowie der relevanten Ergebnismessgrößen (Outcomes) gefordert ist. Im Folgenden werden wichtige Aspekte für die präzise Abbildung des „PICO“ durch Daten aus elektronischen Patientenakten diskutiert.

Population

Für die Bestimmung der Patientenpopulation im Rahmen einer versorgungsnahen Datenerhebung sollten in einer elektronischen Patientenakte die notwendigen Angaben zu Ein-

und Ausschlusskriterien enthalten sein (emulation of target trial, siehe auch Abschnitt 5.3.2). Köpcke 2013 [81] untersuchten elektronische Patientenakten von 5 deutschen Universitätskliniken daraufhin, ob sie ausreichende Angaben zu den protokollgemäßen Auswahlkriterien aus 15 zufällig ausgewählten klinischen Studien enthielten. Im Mittel waren in den von Klinik zu Klinik unterschiedlichen Patientenakten Datenfelder zur Dokumentation von 55 % der Patientencharakteristika vorhanden, die zur Feststellung von Ein- und Ausschlusskriterien erforderlich gewesen wären. Betrachtet man zusätzlich die Datenvollständigkeit von 64 % in diesen Feldern, ergab sich ein in den digitalen Akten abgebildeter Anteil von nur 35 % der zur Auswahl notwendigen Patientencharakteristika. Zusätzliche umfangreiche Datenerhebungen, z. B. aus parallel geführten Dokumentationen (elektronisch oder in Papierform), wären für die Charakterisierung und Rekrutierung der Patientenpopulation im Rahmen von versorgungsnahen Datenerhebungen unvermeidlich, die sich auf im stationären Bereich geführte elektronische Patientenakten stützen wollen. Lau 2011 [82] stellen in einer Studie zum Datenabgleich elektronischer Patientenakten von 52 onkologischen Schwerpunkteinrichtungen in den USA mit anderen Datenbasen (u. a. SEER-Datenbank – Nationales Krebsinstitut, Medicare-Daten) fest, dass in den EHR des onkologischen Verbunds 70 % der Angaben zum Tumorstadium und 40 % zur Abstammung der Patienten fehlten. Damit zeigt sich auch hier die Unvollständigkeit wichtiger Daten als Problem für die Feststellung von Ein- und Ausschlusskriterien auf Basis elektronischer Patientenakten.

Intervention / Comparator

Die zuvor erwähnte Untersuchung in den 5 deutschen Universitätskliniken [81] zeigte auch, dass strukturierte Daten zu Therapien und hier insbesondere zur Medikation kaum in deren elektronischen Patientenakten existierten. Eine Ausnahme bildeten die über den OPS standardisiert verschlüsselten abrechnungsrelevanten Operationen und Prozeduren während des stationären Aufenthalts. Auch nach internationalen Erfahrungen stammen Daten zu Arzneiverordnungen in elektronischen Patientenakten zumeist aus dem ambulanten Versorgungssektor, was eine aktuelle Untersuchung der EMA zu 34 elektronischen Datenbasen aus Europa kürzlich bestätigte, von denen nur 2 (5,9 %) Verordnungen aus dem Krankenhaus enthielten [60]. Da aber neue und innovative Arzneitherapien oft im stationären Sektor begonnen werden, bedeutet der Mangel präziser Verordnungsdaten im Krankenhaus eine erhebliche Einschränkung der Möglichkeit zur Durchführung versorgungsnaher Datenerhebungen auf Basis elektronischer Patientenakten.

Outcomes

Elektronische Patientenakten bilden mit ihren Daten die klinische Routineversorgung ab. Daher enthalten sie für eine Nutzenbewertung von Arzneimitteln wichtige Daten wie die zu Symptomen und Nebenwirkungen unvollständig, und sie werden auch in der Regel nicht in standardisierter Weise kodiert (z. B. MedDRA-kodiert). Vielmehr sind solche Informationen zumeist in den unstrukturierten Textnotizen zu den Behandlungsverläufen enthalten und müssten von dort für Studienzwecke unter großem Aufwand extrahiert und sekundär kodiert

werden [76]. Gravierende unerwünschte Ereignisse finden sich eventuell als ICD-Codes abgebildet. Moore 2015 [83] arbeiten hierzu für 3 große Pharmakovigilanzinitiativen aus den USA und Europa heraus, dass sich sowohl in Krankenkassendaten wie auch elektronischen Patientenakten die Spezifität der über ICD-9-Kodierungen abgebildeten Daten zu Nebenwirkungen und unerwünschten Ereignissen in den bisherigen Validierungsstudien je nach Datenbasis als stark variabel und insgesamt unzureichend erwiesen hat. So zeigte sich zum Beispiel in einer Untersuchung zu den Datengrundlagen des von der FDA unterstützten Sentinel-Projekts, dass kodierte schwere akute Leberschäden im Abgleich mit den Krankenhausakten nur in 24,7 % der Fälle bestätigt werden konnten, und nur ein einziger von 56 Fällen in der Subgruppe mit kodiertem akutem Leberversagen. Zuverlässige Erkenntnisse über die Sensitivität, also wie viele der tatsächlichen unerwünschten Ereignisse in den EHR-basierten Datensätzen überhaupt abgebildet sind, liegen in Ermangelung entsprechender Validierungsstudien kaum vor.

Bestimmte Daten wie die zu gesundheitsbezogener Lebensqualität und PROs sind generell kein Bestandteil von elektronischen Patientenakten, oder darauf bezogene Informationen finden sich ebenfalls nur als unstrukturierte Angaben in den Textnotizen der Behandler. Damit sind wesentliche Outcomes nicht in den elektronischen Patientenakten enthalten.

Confounding und Effektmodifikation

Da die Inhalte und Datenfelder in elektronischen Patientenakten sehr heterogen sind, muss für jede Fragestellung einer vergleichenden Nutzenbewertung individuell geprüft werden, ob die entsprechend benötigten Daten zur Bildung von Confoundervariablen vorhanden sind und in welcher Qualität sie vorliegen. Dies gilt in gleicher Weise für mögliche Effektmodifikatoren.

Repräsentativität / Übertragbarkeit

Für die Frage der Übertragbarkeit von internationalen Studienergebnissen auf Basis von elektronischen Patientenakten auf GKV-Versicherte in Deutschland sind verschiedene Aspekte von Bedeutung, wozu u. a. die Vergleichbarkeit der Populationen gehört. Nach den Ergebnissen der bereits erwähnten Studie von Lau 2011 [82] spielen für Daten aus EHR-Beständen bereits auf nationaler Ebene Selektionseffekte eine wichtige Rolle. Der Vergleich von Daten aus über 160 000 elektronischen Patientenakten onkologischer Schwerpunkteinrichtungen mit registerbasierten Daten des nationalen Krebsinstituts (SEER) und Krankenkassendaten (Medicare-Daten und eine kommerzielle Datenbank) zeigte für 6 untersuchte Tumorentitäten systematische Verteilungsunterschiede von Tumorarten, Abstammung der Patienten und ambulanten onkologischen Therapien. Als Erklärung führen die Autoren neben geografischen Unterschieden der Datensammlungen auch an, dass bestimmte Gruppen Krebserkrankter seltener in onkologischen Spezialeinrichtungen, sondern nur von ihren Fachärzten behandelt werden. Beispielsweise trifft dies für ältere Prostatakrebspatienten zu, die sich bei ihren Urologen in größerer Zahl hormonell behandeln lassen und demnach z. B. stärker in den Daten von Medicare repräsentiert sind. Die Studie macht deutlich, dass sich die Frage der Generalisierbarkeit von Studienergebnissen auf Basis von EHR-Beständen grundsätzlich schon

auf nationaler Ebene ergibt, solange es sich nicht um Analysen aus einer national repräsentativen Datenbasis von einrichtungsübergreifenden elektronischen Patientenakten handelt. Zusätzlich muss dann bei internationalen EHR-Studien die Übertragbarkeit auf die jeweils betrachtete deutsche Patientengruppe aus der GKV-Population bewertet werden.

Fazit

Vor dem Hintergrund

- des Fehlens nationaler, für Forschungszwecke geeigneter einrichtungsübergreifender elektronischer Patientenakten,
- der sehr häufig fraglichen Qualität von aus elektronischen Patientenakten extrahierten Daten,
- den unter der Frage des PICO-Schemas erwähnten Einschränkungen von Daten aus elektronischen Patientenakten und
- der fraglichen Übertragbarkeit internationaler Studien auf Basis elektronischer Patientenakten,

wird derzeit eine Erhebung versorgungsnaher Daten zum Zwecke der Nutzenbewertung nach § 35a SGB V aus elektronischen Patientenakten als nicht realistisch durchführbar angesehen.

5.4.4 Studien auf Basis einer Datenerhebung aus Abrechnungsdaten von Krankenkassen

Grundsätzlich sind Studien auf Basis einer Datenerhebung aus Abrechnungsdaten von Krankenkassen sowohl für versorgungsnahe vergleichende Studien ohne Randomisierung als auch für solche mit Randomisierung denkbar.

Im vorliegenden Bericht stehen die im System der gesetzlichen Krankenversicherung in Deutschland zu Abrechnungszwecken erhobenen Daten im Fokus, die sogenannten GKV-Routinedaten. Einen guten Überblick über die vorhandenen Daten in den Abrechnungsdaten der Sozialversicherungsträger gibt das 2. Kapitel eines für das DIMDI (Deutsches Institut für medizinische Dokumentation und Information) erstellten Gutachtens [84]. Darin werden die Vor- und Nachteile der GKV-Routinedaten dargestellt. Im Kapitel 4 des Gutachtens wird auch beschrieben, wieso die Abrechnungsdaten der etwa 40 privaten Krankenversicherungen aufgrund eines anderen Abrechnungssystems und einer großen Vielfalt von Wahltarifen (z. B. mit Selbstbehalten und Leistungsausschlüssen) vermutlich vor allem bei den ambulanten Leistungen und Arzneimittelverordnungen lückenhaft sind. Daher sind sie für Forschungszwecke im Rahmen von Nutzenbewertungen als ungeeignet anzusehen. Im Folgenden soll auf zentrale Aspekte eingegangen werden, die für die Nutzenbewertung von Arzneimitteln mittels der GKV-Routinedaten besonders relevant sind.

Validität von GKV-Routinedaten

Im Kontext der Forschung mit GKV-Routinedaten kommt insbesondere der Validität der nach ICD-Klassifikation angegebenen Diagnosen eine hohe Bedeutung zu, da zu den abgerechneten Leistungen sonst keine klinischen Angaben dokumentiert werden. Eine Validierung dieser Diagnosedaten kann extern im Abgleich mit anderen Quellen (z. B. Krankenhausakte, Praxisdokumentation) oder intern erfolgen. Bei der internen Validierung in einem Datensatz wird beispielsweise überprüft, ob Diagnosen mit Arzneiverordnungen kongruent sind (z. B. Typ-1-Diabetiker mit Insulinverordnung), ob stationäre und ambulante Diagnosen übereinstimmen, ob Diagnosen irreversibler chronischer Erkrankungen im Zeitverlauf persistieren, oder ob bei Neuerkrankungen auch die in unmittelbarer zeitlicher Nähe zu erwartenden diagnostischen Abklärungsleistungen abgerechnet wurden [85].

Obwohl der externen Validierung im Sinne einer „source data verification“ eine größere Aussagekraft als einer internen Validierung zukommt, wurden systematische Abgleiche bislang für die GKV-Routinedaten nur selten durchgeführt bzw. publiziert.

Für den hausärztlichen Versorgungsbereich weisen sie auf ein bedeutsames Ausmaß an „Underreporting“, „Overreporting“ und Fehlkodierungen hin. So finden Erler 2009 [86] für 2158 betrachtete Behandlungsanlässe von 250 zufällig ausgewählten Patienten aus 10 Hausarztpraxen eine Übereinstimmung mit ICD-10 kodierten Abrechnungsdiagnosen von 40,1 %. In 29,7 % der Fälle mündeten Behandlungsanlässe wie vor allem Früherkennungsuntersuchungen, psychosoziale Beratungsleistungen und unspezifische Erkrankungen und Symptome nicht in Abrechnungsdiagnosen („Underreporting“), während sich in 19 % der Fälle Diagnosen zu Erkrankungen im Abrechnungsdatensatz befanden, die im betreffenden Zeitraum keinen Behandlungsanlass laut Patientenakte geboten hatten („Overreporting“). In 11,2 % der Fälle stimmten Behandlungsanlass und ICD-10-Kodierungen nicht überein („Fehlkodierung“). Für die 6 häufigsten Diagnosen im hausärztlichen Bereich lag aber zumindest auf grober Verschlüsselungsebene (3-stelliger Code) die Korrektheit der Routinedaten (Spezifität) zwischen 71 % und 93 %. Die Qualität der hausärztlichen Diagnosedokumentation im Bereich der Schilddrüsenerkrankungen war Gegenstand einer Studie von Münch 2016 [87] in sechs Hausarztpraxen. Bei 548 Patienten, denen dauerhaft Schilddrüsenmedikamente verordnet wurden, war die Diagnose bei 147 (26,8 %) lückenhaft und bei 100 (18 %) gar nicht dokumentiert worden.

Interne Validierungsstudien aus dem ambulanten Bereich weisen ebenfalls auf ein erhebliches Maß an Inkonsistenzen und damit verbundener Unsicherheit bezüglich der Diagnosevalidität selbst bei gut definierbaren Erkrankungen hin [88-92].

Gesetzliche Krankenkassen haben in der Vergangenheit vor dem Hintergrund des morbiditätsorientierten Risikostrukturausgleichs [93] zwischen den Kassen Gruppen von niedergelassenen Ärzten in speziellen Verträgen finanzielle Anreize zu „korrekten Diagnosekodierungen“ gesetzt. Dies ist in der Gesundheitspolitik seit Langem kontrovers diskutiert worden und hat aktuell zu verschiedenen gesetzlichen Initiativen geführt. Um die

diesbezügliche Sorge um eine Manipulationsanfälligkeit der ambulanten GKV-Routinedaten zu vermindern, wurde im Terminservice- und Versorgungsgesetz (TSVG) 2018 die verbindliche Einführung von Kodierrichtlinien für ambulante Diagnosen und Prozeduren mit Wirkung für Anfang 2022 festgelegt. Eine EDV-Integration in die zertifizierten Praxisverwaltungssysteme von Arztpraxen ist entsprechend vorgesehen, bundeseinheitliche Prüfmaßstäbe für die Kodierungen sollen geschaffen werden. Die Auswirkungen dieser gesetzlichen Maßnahmen auf die Validität der ambulanten Diagnosedaten sind derzeit noch nicht sicher vorhersagbar.

Im stationären Sektor hat sich seit der DRG-Einführung 2003 (Diagnosis Related Groups) durch ein kontinuierliches Überprüfungsverfahren seitens der Krankenkassen und Medizinischen Dienste (MDK) die Kodierqualität auch hinsichtlich abrechnungsrelevanter Diagnosen stetig verbessert [94]. Teile der amtlichen Krankenhausstatistik beruhen daher inzwischen auf Routinedaten wie zum Beispiel Entlassungsdiagnosen. Verbindliche DRG-Kodierrichtlinien werden ständig verfeinert und angepasst. Etwa 10 % der Krankenhausabrechnungen werden vom MDK geprüft, allerdings stellt der Zweifel an der Richtigkeit einer Haupt- oder Nebendiagnose vergleichsweise selten Anlass hierfür dar, sondern andere Fragen wie die einer möglichen Fehlbelegung stehen im Vordergrund [95]. Die Diagnosedaten aus dem Krankenhaus sind per se nicht vollständig in den Routinedaten abgebildet, denn laut Kodervorschriften dürfen nur solche Nebendiagnosen in den Abrechnungsdaten mitgeteilt werden, die den Behandlungsbedarf erhöhen. Maas 2015 [96] überprüften im Kontext einer Patientensicherheitsstudie 3000 Patientenakten deutscher Krankenhäuser im Abgleich mit den Abrechnungsdaten. Dort fanden sich nur 72,7 % der Fälle mit höhergradigem Dekubitus, 25 % der postoperativen tiefen Venenthrombosen oder 44,8 % der Wundinfektionen, sowie 20,8 % der Herzinfarkte und 23,3 % der Pneumonien während des stationären Aufenthaltes dokumentiert. Trotz niedriger Sensitivität war die Korrektheit der Angaben zu diesen unerwünschten Ereignissen in den Routinedaten im Sinne der Spezifität allerdings mit 99 % bis 100 % hoch.

Die Validität der GKV-Arzneiverordnungsdaten (quartalsweise Angaben zu verordneten Arzneimitteln nach ATC-Code, Verordnungs- und Abgabedatum, Packungsgröße) wird allgemein als relativ hoch und stabil angesehen. Für den Zeitraum 2000 bis 2006 wurde von Hoffmann 2008 [97] eine weitere Verbesserung der ohnehin hohen Datenqualität z. B. bei der Datumsdokumentation der Verordnung und Rezepteinlösung beschrieben.

Repräsentativität von GKV-Routinedaten

Zwar gibt es die GKV-Routinedaten für alle derzeit 73 Millionen gesetzlich Versicherten in Deutschland und damit die große Bevölkerungsmehrheit (Vollständigkeit), sie liegen aber nicht in einem Datenkörper vereint, sondern nur getrennt bei aktuell 109 gesetzlichen Krankenkassen vor [98]. Mehrere Studien [99-101] belegen unzweifelhaft, dass sich die Versichertenklientele von gesetzlichen Krankenkassen in Deutschland auch nach Adjustierung für Alter, Geschlecht und Komorbidität teils erheblich voneinander unterscheiden. Schäfer 2017 [101] zeigen beispielsweise im adjustierten Vergleich, dass Melanompatienten, die bei der AOK oder der

landwirtschaftlichen Krankenkasse versichert sind, bei Erstdiagnose vermehrt in deutlich fortgeschrittenem Stadium sind als Versicherte anderer gesetzlicher Kassen. Nur Analysen von Routinedatensätzen, die Versicherte aller gesetzlichen Kassenarten aus allen Regionen enthalten, würden generalisierbare Schlussfolgerungen erlauben. Mitunter vorgenommene Versuche von Hochrechnungen kassenspezifisch ermittelter Ergebnisse auf die gesamte GKV-Klientel sind wegen unklarer Verzerrungseffekte kritisch zu betrachten [100]. Der derzeit einzige kassenartenübergreifend aufgebaute Datenkörper nach § 303 SGB V beim DIMDI verfügt nicht über aktuelle Daten oder über die notwendige Vollständigkeit und Datentiefe für Nutzenbewertungen [102]. Zusammenfassend kann gesagt werden, dass es Analysen auf Basis kassenspezifischer GKV-Routinedaten für die Nutzenbewertung an der erforderlichen Repräsentativität für die gesamte Gruppe der gesetzlich Krankenversicherten mangelt.

Confounderkontrolle in GKV-Routinedaten

Für eine Analyse von GKV-Routinedaten im Rahmen von vergleichenden Studien ohne Randomisierung wäre eine Adjustierung des Vergleichs zwischen Intervention und Kontrolle notwendig. Wie zuvor schon erwähnt, fehlen in den GKV-Routinedaten wichtige Angaben für diese Adjustierung mit dem Ziel, Confounding zu kontrollieren. Dazu gehören klinische und genetische Daten, Befunde wie Laborwerte und andere Testergebnisse sowie Angaben zur Krankheitsschwere. Anthropometrische Daten wie Gewicht und Körpergröße sind nicht dokumentiert, sodass beispielsweise der Body-Mass-Index als möglicher Confounder nicht errechnet werden kann. Daten zum sozioökonomischen Status sind nur teilweise vorhanden, Angaben zu Lebensstilfaktoren (z. B. Rauchen) fehlen [84]. Somit ist eine Adjustierung von Vergleichen vielfach nicht oder nur unzureichend möglich.

Abbildung des PICO aus GKV-Routinedaten

Die Fragestellungen im Rahmen der Nutzenbewertungen des G-BA gliedern sich regelhaft nach dem PICO-Schema, bei dem eine möglichst präzise Beschreibung der interessierenden Patientenpopulation, der zu bewertenden Intervention und ihrer Vergleichsintervention (Comparator) sowie der relevanten Ergebnismessgrößen (Outcomes) gefordert ist. Im Folgenden wird für die GKV-Routinedaten beschrieben, inwieweit mit ihnen diese präzise Abbildung des „PICO“ möglich ist.

Population

Die Identifizierung von Patienten / Versicherten als einer bestimmten Population zugehörig erfolgt in GKV-Routinedaten über eine „Falldefinition“. Diese basiert primär auf soziodemografischen (Alter, Geschlecht) und ICD-Diagnosedaten, muss aber häufig zur Erhöhung der Validität durch weitere Kriterien wie die Dokumentation einer krankheitsspezifischen Medikation oder sonstigen Verordnungen über mehrere Quartale unterstützt werden. Denn es liegen in den GKV-Routinedaten keine klinischen Angaben vor, die zum Beispiel über eine eventuelle ICD-Kodierung hinaus Symptome oder Krankheitsschwere und -stadium abbilden. Exaktere Eingrenzungsmöglichkeiten einer Population sind mit GKV-Routinedaten auch deshalb erschwert, weil Labordaten oder

genetische Daten fehlen. Letztere spielen aber im Rahmen der Nutzenbewertungen von Arzneimitteln durch die „personalisierten“ Behandlungsansätze eine zunehmend größere Rolle. Oft werden in der Praxis Erkrankungen auf der Ebene der 4. Stelle des ICD nur mit „9“ kodiert, was eine mitunter größere Restklasse (Sonstige, nicht näher bezeichnet) bildet und eine spezifische Populationsdefinition in vielen Fällen unmöglich macht [84].

Wenn für die Spezifizierung einer Population der Zeitbezug der Erkrankung eine Rolle spielt, muss man wissen, dass im ambulanten Bereich Diagnosen ohne genaues Datum und nur im Quartalsbezug dokumentiert sind, während sie im stationären Bereich zumindest für neu aufgetretene Erkrankungen und akute Ereignisse über Aufnahme- und Entlassungsdatum abgebildet sind. Schwierigkeiten bereitet die klare Definition neu erkrankter (inzidenter) Patienten, weil die Abgrenzung zu prävalenten Fällen nur über eine längere retrospektive Betrachtung der Vorjahre (ohne entsprechende Diagnose) gelingt.

Da den Diagnosen eine herausragende Bedeutung in der Falldefinition in GKV-Routinedaten zukommt, schränken die im Abschnitt Validität beschriebenen Probleme die Möglichkeit, für die Nutzenbewertung hinreichend abgegrenzte und gleichzeitig aus dem Datensatz vollständige Patientenpopulationen zu bilden, vermutlich bedeutsam ein. Eine ausreichende Aussagesicherheit für Nutzenbewertungen kann auf Basis dieser Daten nach aktuellem Kenntnisstand nicht sichergestellt werden.

Für seltene Erkrankungen liegen nach Angaben des DIMDI im System des ICD-10 nur für etwa 355 von mehr als 8000 Erkrankungen (ca. 4,4 %) eigene spezifische Codes vor [103], was eine Identifikation dieser Patienten in GKV-Routinedaten über Diagnosedaten in den meisten Fällen unmöglich macht. Aktuell wird für die ICD-11 Version ein Doppelkodierungssystem diskutiert, das dieses Problem abmildern soll [103]. Es ist derzeit unklar, in welchem Umfang eine Doppelkodierung zur Lösung beitragen kann.

Zusammenfassend dürfte es vor dem Hintergrund der meist sehr spezifischen Fragestellungen des G-BA in der Arzneimittelbewertung mit GKV-Routinedaten nur selten gelingen, die Untersuchungspopulationen über z. B. genetische Merkmale, Erkrankungsschwere oder Behandlungsstadium mit der für die Nutzenbewertung notwendigen Präzision zu identifizieren.

Intervention / Comparator

Während bei der stationären Behandlung in den DRG-Abrechnungsdaten Codes enthalten sind, die durchgeführte Operationen und Prozeduren nach einem Standardschlüssel (OPS) genau beschreiben, fehlen – von ganz seltenen Ausnahmen hochpreisiger Spezialarzneien abgesehen – Angaben zur stationären Arzneimitteltherapie in den GKV-Routinedaten. Im Krankenhaus begonnene, fortgeführte oder geänderte Arzneitherapien sind also in den Daten nicht erfasst.

Im ambulanten Bereich sind in den Abrechnungen mit der GKV quartalsweise Angaben zu verordneten Arzneimitteln (nach ATC-Code), zu Verordnungsdatum, Abgabedatum in der Apotheke und Packungsgröße enthalten. Diese Daten sind zwar wie zuvor erwähnt an sich

valide, lassen aber keine hinreichend präzisen Rückschlüsse auf die Exposition zu, z. B. ob es zu Dosis- oder Kombinationsänderungen oder zu Therapiewechseln oder -abbrüchen kam. Hierzu werden bei Routinedatenanalysen typischerweise auf Basis des Verordnungsmusters und bestimmter Annahmen Berechnungen zur Abschätzung der Exposition durchgeführt [104]. Nicht rezeptpflichtige (OTC-Präparate) und über Privatrezepte erhaltene Arzneimittel sind in den GKV-Routinedaten nicht erfasst.

Nicht medikamentöse aktive Vergleichstherapien (z. B. operative Verfahren, Heilmittel, Psychotherapie etc.) sind in den GKV-Routinedaten nach den Abrechnungsvorgaben für den jeweiligen Leistungsbereich unterschiedlich detailliert dokumentiert.

Zusammenfassend kann man sagen, dass für eine Datenerhebung zur exakten Beschreibung von Intervention und Vergleichstherapien im Zuge einer vergleichenden Nutzenbewertung von Pharmakotherapien GKV-Routinedaten allgemeine und insbesondere sektorspezifische Lücken aufweisen, die je nach Fragestellung und Anspruch an die Aussagensicherheit solche Analysen kritisch erscheinen lassen.

Outcomes

Mortalität als eindeutiger Endpunkt ist in den GKV-Routinedaten in den Versichertenstammdaten als Grund für die Beendigung des Versichertenverhältnisses zuverlässig erfasst.

Direkte Angaben zur Morbidität sind wegen des Fehlens von klinischen Daten, Labordaten, Daten aus diagnostischer Bildgebung sowie sonstigen Testbefunden nur sehr eingeschränkt mit GKV-Routinedaten möglich. Da die Nutzenbewertung von Arzneimitteln typischerweise aber wesentlich auf der präzisen Erfassung von Symptomen und Komplikationen der Erkrankung oder der Behandlung beruht, stellt das Fehlen von klinischen Angaben zu Symptomen (Übelkeit, Kopfschmerzen, Diarrhoe usw.) einen wesentlichen Mangel der GKV-Routinedaten für diesen Zweck dar. Unerwünschte Ereignisse werden nicht wie in klinischen Studien vollständig erfasst und auch nicht mithilfe von Standardterminologien wie z. B. MedDRA kodiert. Stattdessen ist die Erfassung nur sehr unvollständig und darüber hinaus ungenau über die ICD-Kodierung im ambulanten (nur Quartalsangabe ohne Datum) und stationären Bereich abgebildet. Wie im Abschnitt zur Validität dargestellt, bilden sich die Diagnosen selbst gravierender unerwünschter Ereignisse in den GKV-Routinedaten zum Beispiel im Krankenhaus nicht in genügend sensibler Weise ab [96]. In einer Studie von Kuklik 2017 [105] wurde in 4 deutschen Krankenhäusern ein Abgleich von Krankenakten mit Abrechnungsdaten vorgenommen, bei dem Sensitivität und Spezifität von ICD-10-GM Codes zur Kennzeichnung von im Krankenhaus entstandenen unerwünschten Ereignissen (UE) im Zusammenhang mit Arzneimitteltherapien (z. B. L27.0 – Generalisierte Hauteruption durch Drogen oder Arzneimittel) im Fokus standen. Von 807 begutachteten Fällen mit spezifischen Codes in den Abrechnungsdaten wurden letztlich 65,1 % nach Studium der Krankenakte als im Krankenhaus erworbenes arzneitherapiebedingtes unerwünschtes Ereignis bestätigt (Spezifität). Für die Bestimmung der Sensitivität wurden 1510 zufällig ausgewählte Patientenakten ausgewertet und

dabei in 358 Fällen insgesamt 495 im Krankenhaus erworbene UE identifiziert. Davon fanden sich 246 Ereignisse in den Abrechnungsdaten wieder, was insgesamt einer Sensitivität von 49,7 % entspricht. Diese Ergebnisse wären für eine Nutzenbewertung unzureichend.

Daten zur gesundheitsbezogenen Lebensqualität der Patienten oder zu anderen patientenberichteten Endpunkten (PROs), die für eine Nutzenbewertung ebenfalls von großer Bedeutung sind, sind nicht in den GKV-Routinedaten vorhanden.

Mit GKV-Routinedaten könnten über längere Zeiträume über einzelne Behandlungsepisoden hinaus gröbere Indikatoren betrachtet werden, mit denen man indirekt etwas über den Zustand und das Schicksal des Patienten aussagen kann [106]. Dazu gehören die Mortalität, erneute stationäre Behandlungen und deren Anlässe, Behandlungsdauern und -intensitäten, Eintritt von oder Veränderungen im Grad der Pflegebedürftigkeit, Inanspruchnahme medizinischer Leistungen usw. Solche Indikatoren, die vorwiegend in der vergleichenden Qualitätssicherung der stationären Versorgung zum Einsatz kommen, entsprechen, auch aufgrund ihrer Operationalisierung, mit Ausnahme der Mortalität in der Regel nicht den aktuell für eine Nutzenbewertung von Arzneimitteln verwendeten Endpunkten.

Zusammenfassend gibt Tabelle 6 Möglichkeiten und Einschränkungen von GKV-Routinedaten zur Beantwortung von Fragestellungen zu Arzneimitteltherapien nach dem PICO-Schema wieder:

Tabelle 6: Möglichkeiten und Einschränkungen von GKV-Routinedaten zur Beantwortung von Fragestellungen der Nutzenbewertungen

Population	Intervention - Comparator	Outcomes
<ul style="list-style-type: none"> ▪ eingeschränkte Repräsentativität bei kassenspezifischen Datensätzen ▪ Eingrenzung über Alter, Geschlecht und weitere „Hilfsmerkmale“ sowie fraglich valide, oft unspezifizierte ICD-Diagnosedaten, je nach Erkrankung nur eingeschränkt möglich ▪ Keine klinischen und genetischen Merkmale, Labordaten, Schweregradangaben ▪ Seltene Erkrankungen im ICD-10 nur in ca. 4,4 % spezifisch abgebildet ▪ Inzidente Fälle nur über längere retrospektive Analysen identifizierbar 	<ul style="list-style-type: none"> ▪ Medikament, Verordnungs- und Abgabedatum sowie Packungsgröße im ambulanten Bereich sauber dokumentiert ▪ stationäre Arzneimitteltherapien nicht dokumentiert ▪ Genaue Einnahme (z. B. Dauer, Dosis) unklar ▪ Therapieabbrüche, -unterbrechungen und -wechsel mangelhaft abgebildet ▪ nicht medikamentöse Zusatz- oder Vergleichstherapien sind in GKV-Routinedaten abgebildet 	<ul style="list-style-type: none"> ▪ Mortalität abgebildet ▪ Morbidität wegen des Fehlens von klinischen Daten nur sehr eingeschränkt und mittelbar über ICD-10 Codes abgebildet ▪ Keine Abbildung von Symptomen ▪ Nebenwirkungen (unerwünschte Ereignisse) unvollständig erfasst und nicht spezifisch kodiert (z. B. MedDRA), sondern nur mangelhaft sensitiv und spezifisch über ICD-Codes abgebildet ▪ genauer Zeitbezug für Endpunkte wie z. B. Nebenwirkungen schwierig (ambulant > stationär) ▪ keine Daten zur gesundheitsbezogenen Lebensqualität ▪ keine Daten zu PROs ▪ meist eher unspezifische Indikatoren für langfristige Behandlungsergebnisse können gebildet werden

GKV: gesetzliche Krankenversicherung; ICD: Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme; MedDRA: Medizinisches Wörterbuch für Aktivitäten im Rahmen der Arzneimittelzulassung; PRO: patientenberichteter Endpunkt

Internationale Abrechnungsdaten („claims data“)

Aus den auch auf Abrechnungsdatenbestände bezogenen Analysen von Moore 2015 [83] und Pacurariu 2018 [60] geht mittelbar hervor, dass die dort identifizierten Herausforderungen in der Nutzung von Abrechnungsdaten zu Forschungszwecken international von ihrer Art her deckungsgleich mit denen sind, die sich bei GKV-Routinedaten zu Zwecken der Nutzenbewertung zeigen. Dies bezieht sich zum Beispiel auf verschiedene Aspekte der Datenqualität ebenso wie auf das weitgehende Fehlen wichtiger klinischer Angaben und möglicher Confounder. Externe Validierungsstudien mit Krankenaktenabgleich [107] kommen zu ernüchternden Befunden, insbesondere was die Sensitivität der ICD-kodierten Diagnosen in den Abrechnungsdaten betrifft, und die mit der Nutzung von Krankenkassendaten assoziierten methodischen Probleme werden nicht nur aktuell [108], sondern wurden schon vor mehr als zwei Jahrzehnten in der internationalen Literatur ausführlich diskutiert [109]. In Staaten mit Krankenversicherungssystemen wie dem der USA kommt als Problem hinzu, dass für noch

Berufstätige und deren Angehörige mit dem Arbeitgeberwechsel oft ein Wechsel des Krankenversicherers verbunden ist und damit die für Forschung erforderliche Informationskontinuität fehlen kann.

Fazit

Aus den in Tabelle 6 zusammengefassten Einschränkungen von GKV-Routinedaten und gleich gelagerten Einschränkungen und Qualitätsproblemen bei Abrechnungsdaten aus anderen Ländern ergibt sich deren mangelnde Eignung für Zwecke einer belastbaren Nutzenbewertung von Arzneimitteln, wie sie der G-BA nach den dafür bestimmten Vorgaben vornimmt.

5.5 Anforderungen an Daten und Auswertungen

5.5.1 Kriterien für Datenqualität

Das vorliegende Kapitel beschreibt Kriterien für die Datenqualität von Registern. Der Grund für die Fokussierung auf dieses Instrument zur Datenerhebung liegt darin, dass Register für die Fragestellung des Berichts von besonderer Bedeutung sind. Versorgungsnahe Daten aus elektronischen Patientenakten und aus Abrechnungsdaten der Krankenkassen spielen für die Fragestellungen der Nutzenbewertung derzeit und absehbar eine untergeordnete Rolle (siehe 5.4.3 und 5.4.4). Die Kriterien für die Qualität von Daten aus studienindividuellen Datenerhebungen sind als Standards weitgehend etabliert.

5.5.1.1 Kriterien für die Qualität von Registern gemäß nationaler und internationaler Standards

Qualitätskriterien für Patientenregister und Registerstudien

Unter einem Patientenregister soll wie in Gliklich 2014 [4] definiert „ein organisiertes System“ verstanden werden, „das mit der Methodik einer Beobachtungsstudie einheitliche Daten (klinische und andere) sammelt, die bei definierten Populationen von Erkrankten, Exponierten oder Merkmalsträgern der Evaluation festgelegter Endpunkte dienen, und das vorab bestimmte wissenschaftliche, klinische oder programmatiche Zwecke verfolgt“.

Solche patientenbezogenen Datensammlungen sind von unterschiedlichster Art und Zweckbestimmung. Sie können über die routinehaften Basisauswertungen hinaus zur Grundlage zusätzlich geplanter beobachtender oder interventioneller Studien werden. Dann spricht man von einer Registerstudie, bei der über grundlegende Qualitätsanforderungen für ein Patientenregister hinaus häufig weitere Kriterien zu erfüllen sind, z. B. bezüglich eines Studienprotokolls, eines erweiterten Patienteneinverständnisses, der Erhebung zusätzlicher Daten, einer intensiveren Sicherung der Datenqualität oder der Auswertungsmethodik (siehe Abbildung 4).

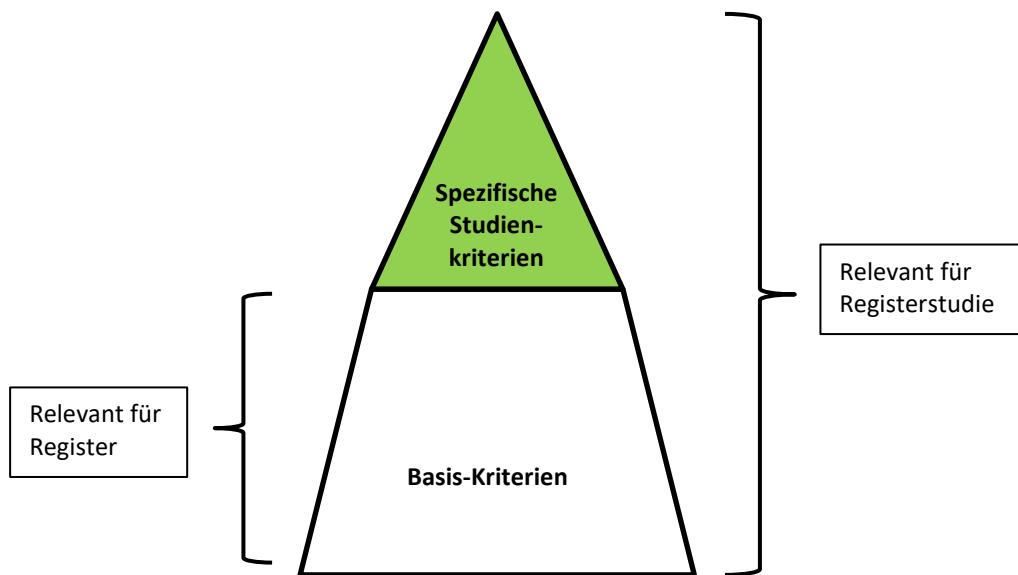


Abbildung 4: Qualitätskriterien für Patientenregister und Registerstudien

In Publikationen zur Qualität von Patientenregistern wird regelhaft ein Mix aus Struktur-, Prozess- und Ergebnisqualitätskriterien empfohlen, der eine gute Registerpraxis auszeichnet. Dabei wird im vorliegenden Rapid Report analog zu Donabedians Qualitätsmodell davon ausgegangen, dass Strukturen und Prozesse dem Ziel einer möglichst hohen Ergebnisqualität dienen, die damit „die ultimative Beurteilungsebene“ [110, S. 694] für Qualität darstellt.

Übertragen auf Patientenregister wäre z. B. ein Strukturqualitätsmerkmal das Vorhandensein eines Steuerungsgremiums oder eines umfassenden, aktuellen Registerprotokolls. Prozessqualitäts-Merkmale sind u. a. regelmäßige Audits in den lokalen Meldezentren oder automatische Überprüfungen bei der Dateneingabe. Letztendlich zielen Strukturen und Prozesse in einem Patientenregister auf eine hohe Ergebnisqualität, die sich vor allem in zentralen Merkmalen der Datenqualität (Vollzähligkeit, Vollständigkeit, Richtigkeit) ausdrückt. Ein weiterer Aspekt von Ergebnisqualität kann übergeordnet in effektivem Datenschutz oder schneller Verfügbarkeit von Auswertungsergebnissen gesehen werden.

Im Abschnitt 4.2 wurde beschrieben, wie die in Tabelle 7 enthaltene Liste von Qualitätskriterien für Patientenregister für den vorliegenden Bericht erarbeitet wurde. Auch diese Liste stellt eine Mischung aus Struktur-, Prozess- und Ergebnisqualitätsmerkmalen dar und enthält wie von anderen publizierte Kriterienlisten [4,6,8] auch Basiskriterien für ein Register und solche, die besondere oder sogar spezifische Relevanz für registerbasierte Studien haben.

Tabelle 7: Zusammenstellung national und international verwendeter Qualitätskriterien für Patientenregister und registerbasierte Studien

Nr.	Qualitätskriterium
Systematik	
1	Detaillierte Registerbeschreibung (Protokoll)
Standardisierung	
2	Exakte Definition / Operationalisierung von Expositionen, klinischen Ereignissen, Endpunkten und Confoundern
3	Aktueller Datenplan / Kodierhandbuch
4	Verwendung von Standard-Klassifikationen (z. B. ICD-10) und -Terminologien (z. B. MedDRA)
5	Verwendung von validierten Standard-Erhebungsinstrumenten (Fragebogen, Skalen, Tests)
6	Schulungen zur Datenerhebung und -erfassung
7	Umsetzung eines konsentierten krankheitsspezifischen Kerndatensatzes („core data set“)
8	Verwendung exakter Datumsangaben zum Patienten (z. B. Geburt, Tod, Schwangerschaft)
9	Verwendung exakter Datumsangaben zur Erkrankung (z. B. definitive Diagnose, klinisch relevante Ereignisse)
10	Verwendung exakter Datumsangaben zu wichtigen Untersuchungen
11	Verwendung exakter Datumsangaben zu Behandlungen / Interventionen (z. B. bei Arzneimitteln Start- / Stop-Datum, Dosis, Dosisänderungen)
Erreichung des Rekrutierungsziels / Stichprobengewinnung	
12	Klar definierte Ein- und Ausschlusskriterien für Registerpatienten
13	Vollzähligkeit der Registerpatienten (Vollerhebung oder repräsentative Stichprobe)
14	Strategien zur Vermeidung ungewollter Selektionen beim Patienteneinschluss, um Repräsentativität zu erreichen
Validität der Datenerhebung	
15	Vollständigkeit der Daten je Erhebungszeitpunkt (loss-to-follow-up, Drop-outs)
16	Vollständigkeit der Erhebungszeitpunkte
17	Richtigkeit der Daten (accuracy)
18	Datenkonsistenz über die Zeit
19	Source data verification (z. B. für 10 % zufällig ausgewählter Patienten je Erhebungszentrum)
20	Registermonitoring durch interne audits
21	Registermonitoring durch externe audits
22	QM-System (gfs. mit regelmäßiger Erhebung von Qualitätsindikatoren)
23	SOPs (Standard Operating Procedures) zur Datenerhebung

(Fortsetzung)

Tabelle 7: Zusammenstellung national und international verwendeter Qualitätskriterien für Patientenregister und registerbasierte Studien (Fortsetzung)

Nr.	Qualitätskriterium
Übergeordnete Qualitätskriterien	
24	Transparenz eines Registers (u. a. Finanzierung, Entscheidungswege, Interessenkonflikte)
25	Wissenschaftliche Unabhängigkeit
26	Sichere Finanzierung (für geplanten Erhebungszeitraum)
27	Steering committee, Leitungsgremium
28	Aktualität der Registerdokumente (z. B. Protokoll, Datenplan, SAP, Einwilligungserklärung etc.)
29	Wahrung von Patientenrechten und Datenschutz, Berücksichtigung ethischer Aspekte
30	Timeliness (Aktualität / schnelle Verfügbarkeit / Pünktlichkeit der benötigten Ergebnisse)
31	Flexibilität und Anpassungsfähigkeit (z. B. zur Einbettung von Studien, für weitere Datenerhebung, bei veränderter Versorgungssituation)
32	Documentation trail - Dokumentation aller Prozess- und Definitionsänderungen im Register
33	Audit trail - Dokumentation und Attribuierbarkeit aller Datentransaktionen
34	Verknüpfbarkeit mit anderen Datenquellen
Validität der statistischen Analysen und Berichte zu Registerstudien	
35	Öffentliche Registrierung der geplanten Registerstudie
36	Erstellung eines Studienprotokolls und eines statistischen Analyseplans zur geplanten Registerstudie
37	Präspezifizierung der Auswertungsmethodik im statistischen Analyseplan
38	Darlegung des Umgangs mit fehlenden Werten
39	Adjudikations-Komitee für Schlüssel-Outcomes
40	Adjustierung von Vergleichsergebnissen für potenziell konfundierende Variablen und Berücksichtigung von Effekt modifizierenden Variablen
41	Sensitivitätsanalysen (z. B. für unterschiedliche Falldefinitionen oder Confounder-Berücksichtigung)
42	Auswertung / Kontrolle von Meldeeinrichtungseffekten
43	Bericht über Maßnahmen zur Vermeidung sonstiger Verzerrungen (z. B. Selektionsbias)
44	Vollständiger Bericht der Ergebnisse
45	Publikation des Ergebnisberichts einschließlich Studienprotokoll und Analyseplan
Weitere mögliche Kriterien aus regulatorischer Sicht	
46	Erhebung und Handling unerwünschter Ereignisse (UEs) gemäß regulativer Anforderungen
ICD: Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme; MedDRA: Medizinisches Wörterbuch für Aktivitäten im Rahmen der Arzneimittelzulassung; QM: Qualitätsmanagement; SAP: statistischer Analyseplan; SOP: Standard Operating Procedure; UE: unerwünschtes Ereignis	

Diese Kriterienliste kann Ausgangspunkt für eine Übersicht und Einschätzung der Struktur-, Prozess- und Ergebnisqualität eines Patientenregisters sein.

Zur Beurteilung der Güte eines Patientenregisters ist vor allem die glaubwürdig dargelegte Ergebnisqualität hinsichtlich Repräsentativität, Vollständigkeit und Richtigkeit der Daten relevant. Es ist daher im Rahmen einer Prüfung auf Eignung eines Registers für die

Durchführung von Studien nicht angemessen, alle einzelnen Struktur- oder Prozessqualitätsmerkmale der Liste als obligat zu betrachten. Der Zusammenhang zwischen einzelnen Struktur- und Prozessqualitätsmerkmalen und der erreichten Ergebnisqualität ist auch bei Patientenregistern nicht linear. Es ist eher davon auszugehen, dass umso bessere Ergebnisse zu erwarten sind, je mehr dieser Kriterien in hohem Maße erfüllt sind. Gleichzeitig muss aber auch ein geringer Erfüllungsgrad bei einzelnen Kriterien nicht zwangsläufig dazu führen, dass ein Register grundsätzlich für die versorgungsnahe Datenerhebung zum Zwecke der Nutzenbewertung ungeeignet ist.

Im Rahmen der Eignungsprüfung eines spezifischen Registers sollte daher mithilfe einer auf den vorliegenden Anwendungsfall, der Nutzenbewertung von Arzneimitteln, ausgerichteten Kriterienliste für die jeweilige konkrete Forschungsfragestellung evaluiert werden, ob und in welchem Ausmaß die einzelnen Kriterien erfüllt sind, welchen Einfluss eine mögliche Nichterfüllung auf die Ergebnisqualität vermutlich hat, und ob im Rahmen einer registerbasierten Studie eventuelle Defizite in zeitlich zumutbarem Rahmen und mit vertretbarem Aufwand korrigiert werden können.

Eine solche auf den vorliegenden Anwendungsfall, die Nutzenbewertung von Arzneimitteln, ausgerichtete Kriterienliste findet sich in Abschnitt 5.5.1.3. Die Liste wurde ausgehend von der obigen Gesamtliste auf Basis der Literaturanalyse zu national und international verwendeten Qualitätskriterien für Register, den Gesprächen mit Registerbetreibern (siehe Abschnitt 5.5.1.2) sowie den Erfahrungen aus über 300 Nutzenbewertungen von Arzneimitteln und den für die Bewertungen relevanten Anforderungen an Studiendesign und -auswertung gebildet. Kriterien, die sich auf die Erstellung eines Studienprotokolls und eines Analyseplans für eine Registerstudie sowie auf konkrete Auswertungen zu einer Registerstudie beziehen (z. B. Sensitivitätsanalysen) werden dabei nicht aufgeführt, da diese im Abschnitt 5.5.2 (Methodische Anforderungen an die Auswertung) separat adressiert werden.

5.5.1.2 Ergebnisse der Gespräche mit Registerbetreibern

Wie in Abschnitt 4.2 beschrieben, wurde zur Vorbereitung der Gespräche mit den Registerbetreibern ein 3-teiliger Fragebogen verschickt. Nicht allen Registerbetreibern war es möglich, die ersten beiden Teile des Fragebogens (Teil 1: Informationen zum Register; Teil 2: Allgemeine Qualitätskriterien für Register) vor dem Gesprächstermin auszufüllen. Auch in diesen Fällen dienten die im Fragebogen aufgeführten Fragen (Teil 1) bzw. Kriterien (Teil 2) als Gesprächsleitfaden.

Neben dem Ziel, ein allgemeines Verständnis von Aufbau, Organisation und quantitativem sowie qualitativem Datenbestand zu erhalten sowie eine Einschätzung zur Bedeutung und zum Erfüllungsgrad der verschiedenen Qualitätskriterien zu gewinnen, sollten mit Hilfe der praktischen Erfahrungen der Registerbetreiber besonders förderliche sowie besonders hinderliche Faktoren für den Betrieb eines bestehenden oder für den Aufbau eines neuen Registers identifiziert werden. Diese Erfahrungen sollten auch dazu dienen, den Zeitbedarf für unterschiedliche Ausgangssituationen (neues Register vs. bestehendes Register) und

unterschiedliche Herangehensweisen (prospektive Datenerhebung im Register vs. retrospektive Auswertung auf Basis eines bereits vorliegenden Datenbestands) abzuschätzen.

Nachfolgend werden zunächst die Rückmeldungen der Registerbetreiber zu den in Teil 2 des Fragebogens aufgeführten Qualitätskriterien zusammenfassend beschrieben, insbesondere im Hinblick auf den jeweiligen Erfüllungsgrad (Selbsteinschätzung der Registerbetreiber) und die Bedeutung der genannten Kriterien. Im Anschluss werden die in den Gesprächen hervorgehobenen besonders förderlichen und besonders hinderlichen Faktoren für den Registerbetrieb aufgeführt.

Danach findet sich eine allgemeine Einschätzung zur Eignung der betrachteten Register für die versorgungsnahe Datenerhebung zum Zwecke der Nutzenbewertung. Abschließend wird beschrieben, welcher Zeitraum für versorgungsnahe Datenerhebungen zum Zwecke der Nutzenbewertung voraussichtlich in der Regel erforderlich ist.

Erfüllung der allgemeinen Qualitätskriterien aus Sicht der Registerbetreiber

Die im Fragebogen genannten Qualitätskriterien waren den Registerbetreibern durchweg bekannt und wurden im Grundsatz auch sämtlich als sinnvoll angesehen, wenn auch teilweise in modifizierter Form (siehe z. B. nachfolgend zum Thema source data verification).

Nach Selbsteinschätzung der Registerbetreiber war der Erfüllungsgrad der Kriterien, die der Standardisierung der Daten im Register dienen, in der Regel hoch. Eine Ausnahme bildeten bei manchen Registern hierbei zum einen Klassifikationssysteme, die nicht regelhaft in der klinischen Praxis verwendet werden (z. B. das MedDRA-System zur Kodierung unerwünschter Ereignisse / Nebenwirkungen), zum anderen exakte Datumsangaben, wobei sich dies je nach Register auf Datumsangaben zu den Patienten selbst, zur Erkrankung (Diagnosestellung und Untersuchungen im Krankheitsverlauf) oder zur Arzneimitteltherapie bezog. Insbesondere bei letzterem wurde in Bezug auf Angaben zur Initialdosis oder zu Dosisänderungen im Verlauf teilweise ein geringerer Erfüllungsgrad konstatiert.

Bei den Kriterien zur Validität der Stichprobengewinnung wurde insbesondere bei der Gewährleistung der Vollzähligkeit und bei Strategien zur Vermeidung einer Patientenselektion teilweise ein geringerer Erfüllungsgrad beschrieben (Ausnahme insbesondere: Register mit gesetzlicher Grundlage zur Erfassung der Daten für alle in Frage kommenden Patientinnen und Patienten). In diesem Zusammenhang wurde von einigen Registerbetreibern angegeben, dass mangels valider externer Datenquellen die Repräsentativität der Registerpopulation nur unzureichend beurteilt werden könne.

Für die Kriterien zur Validität der Datenerhebung (Vollständigkeit und Richtigkeit der Daten) wurde häufiger der Erfüllungsgrad als nicht vollständig angegeben. Begründet wurde dies zumeist mit dem damit verbundenen sehr hohen Aufwand. Mit Blick auf die source data verification gaben die meisten Registerbetreiber an, diese zwar durchzuführen, die z. B. von der EMA genannte Zielgröße eines zu verifizierenden Anteils von 10 % der Daten aufgrund

des hohen finanziellen und organisatorischen Aufwands derzeit jedoch nicht zu erreichen. Aus Sicht einiger Registerbetreiber sei eine solche feste Zielmarke allerdings auch nicht sinnvoll. Denn zum einen kämen auch andere Verfahren der Sicherstellung der Datenqualität zum Einsatz (z. B. Kodierhilfen und Schulungen, Plausibilitätsprüfungen, Cross-Reference-Prüfungen [bei Eingabe von Daten zu einem Patienten durch mehrere Personen, z. B. bei Kontakt zu unterschiedlichen Versorgungsebenen]). Zum anderen wäre aus Sicht einiger Registerbetreiber eine adaptierte Quelldatenprüfung, z. B. mittels Prüfung eines geringen Datenanteils in jedem Zentrum mit Erweiterung der Prüfung bei Auffälligkeiten, sinnvoller. Externe Audits (externes Monitoring) wurden in den betrachteten Registern nach Angabe der Registerbetreiber nicht regelhaft durchgeführt.

Die Einschätzung zum Erfüllungsgrad der Kriterien zur Validität der statistischen Analyse und Berichte war zwischen den Registerbetreibern heterogen. Teilweise wurden hier keine Einschränkungen konstatiert, teilweise wurde in der Darlegung des Umgangs mit fehlenden Werten und bei der Confounder-Adjustierung ein unvollständiger Erfüllungsgrad festgestellt. Bei letzterem waren insbesondere die eingeschränkte Erfassung potenziell relevanter Confounder ausschlaggebend für diese Einschätzung.

Schließlich war die Selbsteinschätzung zur Registertransparenz sowie zur wissenschaftlichen Unabhängigkeit durchweg uneingeschränkt positiv. Die Betrachtung der öffentlich zugänglichen Informationen zu den Registern (Webseiten, Ergebnisberichte und wissenschaftliche Publikationen) erbrachte keine gegenteilige Einschätzung. Eine Anbindung an die Spontanmeldesysteme bei vermuteten unerwünschten Arzneimittelwirkungen war hingegen bei keinem der betrachteten Register vollständig gegeben. Auch hier wurde insbesondere der hohe Aufwand als Grund genannt.

Besonders förderliche oder hinderliche Faktoren für den Registerbetrieb

Bei den sich aus den Registergesprächen ergebenden besonders förderlichen und hinderlichen Faktoren für den Registerbetrieb zeigte sich eine auffällige Konsistenz zwischen den verschiedenen Registern bzw. Registerbetreibern. Abweichende Einschätzungen ergaben sich insbesondere dann, wenn durch besondere Rahmenbedingungen wie die gesetzliche Verankerung eines Registers und der zu erfassenden Daten der Registerbetrieb gestützt wird.

Als besonders förderliche Faktoren wurden insbesondere genannt:

- Eine gewachsene Gemeinschaft klinisch tätiger Ärztinnen und Ärzte und weiteren medizinischen Personals mit intrinsischer Motivation zur Unterstützung des Registerbetriebs
- Lokales Feedback für die beteiligten Zentren innerhalb der Registersoftware (Zugriff auf die Daten zu den Zentrumspatientinnen und -patienten einschließlich der Möglichkeit zur Darstellung zeitlicher Verläufe, ggf. auch Benchmarking mittels Vergleich mit anderen beteiligten Zentren)

- Zur Motivation der Zentren: Verwendbarkeit der Daten für Forschungsaufgaben / Beteiligung der Zentren an wissenschaftlichen Publikationen, wissenschaftliche Unabhängigkeit von Sponsoren
- Einfache Kontaktmöglichkeit für die Zentren zu den Registerbetreibern mit zeitnaher Unterstützung bei Problemen
- Im Hinblick auf die Datenqualität: Ausführliche Schulung der beteiligten Zentren, professionelle Unterstützung der Zentren bei der Dateneingabe, z. B. durch geschulte Dokumentationskräfte
- Für den Aufbau eines neuen Registers: eine vorhandene technische Infrastruktur (Hard- und Software), z. B. aufgrund des etablierten Betriebs eines Registers zu einer anderen (verwandten) Erkrankung

Als besonders hinderliche Faktoren wurden insbesondere genannt:

- Fehlende langfristige Finanzierungsmöglichkeiten
- Heterogene Datenschutzanforderungen (auf verschiedenen Ebenen, z. B. zwischen Bundesländern, zwischen verschiedenen Universitäten, zwischen verschiedenen Kliniken auch innerhalb eines Bundeslandes)
- Fehlende Standardisierung der IT-Infrastruktur in den Zentren (insbesondere Heterogenität der Krankenhausinformationssysteme [KIS], teilweise veraltete Hardware)
- Fehlender oder erschwerter Zugriff auf andere Datenquellen (z. B. Sterberegister)
- Teilweise mangelnde Bereitschaft zur source data verification in den Zentren (wird als Kontrolle empfunden, Störung des Betriebsablaufs)

Nachfolgend wird auf einige der genannten Punkte näher eingegangen, da diese in den Gesprächen mit den Registerbetreibern häufig und besonders betont wurden.

Finanzierung und Unabhängigkeit

Die meisten der befragten Registerbetreiber haben auf die Notwendigkeit einer nachhaltigen Finanzierung des Registers hingewiesen. Für viele Register sei dies ein Dauerthema, da eine Finanzierung häufig aus Eigenmitteln der betreibenden Organisation mit oder ohne finanzielle Unterstützung der (zumeist pharmazeutischen) Industrie, ggf. ergänzt durch zeitlich befristete Drittmittelfinanzierung von Forschungsprojekten, gewährleistet werden müsse. Die Industriefinanzierung wurde von einigen Registerbetreibern kritisch gesehen, da dies aus ihrer Sicht die Unabhängigkeit der Register gefährde. Unabhängig von der konkreten Finanzierung des eigenen Registers waren sich die Registerbetreiber dahingehend einig, dass alle im Register erfassten Daten zur Auswertung für die Registerbetreiber (oder von ihnen beauftragte Institutionen oder andere Forschergruppen) uneingeschränkt zur Verfügung stehen und eine Publikation dieser Auswertungen ergebnisunabhängig möglich sein müssen.

Die teilweise stark begrenzten Mittel verhinderten nach Angabe der Registerbetreiber bei manchen Registern wünschenswerte Erweiterungen der Datenerfassung, teilweise sei auch eine Reduktion der derzeitigen Datenerfassung angedacht. Dies betreffe nicht nur Datenfelder und Datenzeitpunkte, sondern auch die professionelle Unterstützung der Datenerfassung vor Ort. Insgesamt sei aus Sicht der meisten Registerbetreiber eine der Bedeutung von Registern und dem damit verbundenen Aufwand angemessene und nachhaltige (d. h. für mehrere Jahre gesicherte) Grundfinanzierung durch eine von der Industrie und der GKV / PKV unabhängige Stelle erforderlich. Als Option wurden in diesem Zusammenhang die Erhebung eines Systemzuschlags für die Finanzierung von Registern genannt.

Datenschutz

Für alle Registerbetreiber stellte der Datenschutz ein wesentliches Qualitätsmerkmal der Register zur Gewährleistung der Vertrauenswürdigkeit dar. Dem entsprechend wurde nicht der Datenschutz als solcher, sondern das damit verbundene lokale oder regionale Vorgehen als wesentlicher hinderlicher Faktor beschrieben. Genannt wurden in diesem Zusammenhang insbesondere 3 Punkte:

- Unterschiedliche Anforderungen und Einschätzungen zum Datenschutz durch die zuständigen Stellen in den Bundesländern, einschließlich der benannten Datenschutzbeauftragten: identische Sachverhalte (Vorgehensweisen, Einverständniserklärungen etc.) würden in den verschiedenen Ländern teilweise unterschiedlich bewertet
- Teilweise spezifische Anforderungen auf Zentrumsebene: Selbst bei grundsätzlicher Freigabe durch die zuständigen Stellen (z. B. der Bundesländer) gebe es auf Zentrumsebene zum Teil langwierige Verhandlungen wegen spezifischer weitergehender Anforderungen, die im Registerbetrieb nicht ohne Weiteres umzusetzen sind
- Fehlende praxisnahe Hilfestellungen (Unterlagen, Ablaufbeschreibungen) für den Fall, dass ein Register neu aufgesetzt wurde; die von der Technologie- und Methodenplattform für die vernetzte medizinische Forschung (TMF e. V.) hierzu bereitgestellten Informationen und Unterlagen wurden von den Registerbetreibern mehrheitlich als zu praxisfern bezeichnet

Fehlende Standardisierung der IT-Infrastruktur in den Zentren

In den Gesprächen wurde wiederholt die fehlende Standardisierung der IT-Infrastruktur sowie die teilweise veraltete Hardware in den beteiligten Zentren als hinderlicher Faktor genannt. Dabei wurde anschaulich beschrieben, warum die fehlende Standardisierung der KIS je nach Zentrum die Dateneingabe vor Ort unnötig aufwändig und fehleranfällig machen kann. Denn viele Informationen, die auch für das Register erforderlich sind (z. B. demografische Daten der Patientinnen und Patienten, Laborwerte, Untersuchungsbefunde etc.) müssen auch in das KIS eingegeben werden bzw. sind in diesem patientenbezogenen enthalten. 2 Punkte wurden in diesem Zusammenhang besonders hervorgehoben:

- Zum einen fehlen je nach KIS Schnittstellen, die eine Übernahme dieser bereits digital vorliegenden Informationen in das Register ermöglichen. Dies verursacht nicht nur unnötigen Mehraufwand durch Doppeleingabe (was auch die Rekrutierung neuer Zentren behindert), sondern stellt auch eine potenzielle Fehlerquelle bei händischer Übertragung dieser Informationen in die Registersoftware dar. So wurde beschrieben, dass Dokumentationskräfte, die vor Ort für die Dateneingabe in die Registersoftware zuständig sind, parallel mit 2 IT-Systemen arbeiten, weil die notwendigen Informationen vom Bildschirm des einen Systems abgelesen werden (KIS) und dann in das andere System (Registersoftware) übertragen werden.
- Die je nach Datenfeld fehlende semantische Interoperabilität zwischen den verschiedenen KIS erlaubt darüber hinaus nicht, dass selbst aus KIS mit vorhandener Schnittstelle die Daten mittels einer einheitlichen Schnittstelle in das Register übernommen werden können. Dieses Hindernis wird nach Angaben einiger Registerbetreiber durch individuelle Anpassungen der KIS in den einzelnen Zentren noch verschärft.

Anbindung an andere Datenquellen zur Verwendung anderweitig erfasster Informationen

Aus Sicht der Registerbetreiber wäre es zu begrüßen, wenn bereits anderweitig in hoher Qualität erfasste Daten für die Register zur Verfügung gestellt werden könnten, um die patientenbezogenen Informationen zu vervollständigen. Beispielhaft genannt wurde hier der Zugriff auf das Sterberegister: Aus Gründen des Datenschutzes oder des Fehlens einer einheitlichen patientenbezogenen Identifikationsnummer sei der Zugriff hierauf nach Darstellung einiger Registerbetreiber kaum möglich. Stattdessen würde für Registerstudien, die u. a. Aussagen zum Überleben der Patientinnen und Patienten treffen wollen, mit hohem Aufwand durch gezielten Zentrumskontakt versucht, die entsprechenden Informationen zu erlangen. Unklar war den Registerbetreibern auch, welche potenziellen Datenquellen es überhaupt gebe, und welche Möglichkeiten und Grenzen die jeweiligen Datenquellen bieten. Hierzu fehle eine zentrale Informationsplattform, die idealerweise auch die niederschwellige Möglichkeit zum Austausch zwischen den Nutzern der Datenquellen ermöglichen sollte.

Eignung der Daten für die Zwecke der Nutzenbewertung von Arzneimitteln

Aus den Befragungen wurde deutlich, dass auf der einen Seite die bestehenden Register in der Regel nicht primär darauf ausgerichtet sind, Daten zum Zwecke der Nutzenbewertung von Arzneimitteln zu generieren. Auf der anderen Seite ist aufgrund der bestehenden technischen und organisatorischen Infrastruktur, teilweise auch aufgrund der bereits bestehenden Datenstruktur, davon auszugehen, dass solche Daten prinzipiell in den betrachteten Registern zumindest prospektiv generiert werden können. Ob und mit welchem zusätzlichen Aufwand dies verbunden wäre, hängt stark von der jeweiligen Fragestellung ab, was nachfolgend an verschiedenen Punkten beispielhaft erläutert wird.

- Die meisten der betrachteten Register erfassen auch patientenberichtete Endpunkte (PROs), wenn auch teilweise erst seit kurzer Zeit mit daraus folgend aktuell geringem Datenbestand. Ob die verwendeten Instrumente (Symptomfragebogen, Fragebogen zur

generischen oder krankheitsspezifischen Lebensqualität) und die Erfassungsintervalle für die Nutzenbewertung geeignet sind, hängt von der konkreten Fragestellung ab.

- In den betrachteten Registern werden Nebenwirkungen, im Vergleich zu klinischen Studien, nur in begrenztem Umfang erfasst. Zumeist sind bekannte Nebenwirkungen ausgewählter Arzneimittel das Aufgreifkriterium für deren Erfassung. In manchen Registern werden zumindest die erfassten Nebenwirkungen analog zu klinischen Studien nach dem MedDRA-System kodiert, sodass prinzipiell eine Studien-analoge Erfassung möglich scheint. Bei einigen Registern wird dies durch die Unterscheidung von Zentren mit begrenzter (Kerndatensatz) und erweiterter Datenerfassung unterstützt. Zentren mit erweiterter Datenerfassung sind in einigen Registern prinzipiell auf eine Studien-analoge Datenerfassung ausgerichtet. Dabei wird von einigen Registerbetreibern dies auch im Rahmen von registerbasierten RCTs angestrebt. Relevante Hürden für die Durchführung interventioneller Registerstudien stellen aus Sicht mancher Registerbetreiber die damit verbundenen gesetzlichen Anforderungen (insbesondere gemäß Arzneimittelgesetz [AMG] und SGB V) dar.
- Bei einigen der betrachteten Register sind nicht alle Versorgungsebenen (z. B. hausärztliche Versorgung, Versorgung in Zentren, Rehabilitation) an der Datenerfassung direkt beteiligt. Ob Informationen, die durch Kontakt mit einer nicht involvierten Versorgungsebene entstehen, in den Registerdatensatz einfließen, hängt in diesen Fällen vom Informationsfluss zwischen den Versorgungsebenen ab. Dies beeinflusst potenziell nicht nur die Vollständigkeit der Daten (Werden diese Informationen im Register erfasst?), sondern auch die Datenqualität (ggf. Überführung einer bereits verarbeiteten und in unstrukturierter Form vorliegenden Information [Diagnose im Arztbrief] durch eine dritte Person in das Register ohne Möglichkeit einer source data verification). Ob diese Informationen für die Datengenerierung zum Zwecke der Nutzenbewertung relevant sind und daher mit entsprechendem Zusatzaufwand in ausreichender Qualität erfasst werden sollten, hängt wiederum von der konkreten Fragestellung ab: Auf der einen Seite gibt es je nach Erkrankung ggf. kaum Arztkontakte außerhalb der an dem Register teilnehmenden Zentren, oder die in anderen Versorgungsebenen entstehenden Informationen sind für die Nutzenbewertung ggf. nicht entscheidungsleitend (z. B. Diagnose von Bagatellerkrankungen). Auf der anderen Seite könnten im Rahmen eines (nicht direkt erfassten) Notfalls erforderliche Therapien und Komplikationen von zentraler Bedeutung für die Nutzenbewertung sein.

Die genannten Punkte beschreiben die Möglichkeiten und Grenzen, die sich bei der Durchführung einer (retrospektiven oder prospektiven) vergleichenden Studie innerhalb derselben Datenquelle, d. h. desselben Registers, ergeben. Aus der Gesamtschau der Befragungen wurde darüber hinaus deutlich, dass eine vergleichende Studie, die unter Verwendung verschiedener Datenquellen für den zu bewertenden Wirkstoff bzw. die Vergleichstherapie durchgeführt wird, vermutlich nur in Ausnahmefällen für Zwecke der Nutzenbewertung geeignet sein wird. Bei solchen Analysen ist potenziell eine Confounder-

Adjustierung allein nicht ausreichend, um systematische Verzerrungen zu vermeiden, da vermeintlich gleiche Endpunkte in den verschiedenen Datenquellen potenziell deutlich unterschiedlich definiert sind und selbst bei ähnlicher Definition die Art der Erhebung zu relevanten Unterschieden führt. Beispiele hierfür sind:

- Unerwünschte Ereignisse: Werden die Ereignisse systematisch bei dem Patientenkontakt abgefragt? Werden Ereignisse erfasst, wenn der Patient bzw. die Patientin auf diese hinweist? Werden die Ereignisse nach Schweregrad eingestuft? Wenn ja, erfolgt die Einstufung des Schweregrades durch die Patientinnen bzw. Patienten, die Ärztinnen bzw. Ärzte, nach einem standardisierten System?
- Patientenberichtete Endpunkte: Wo werden die Daten erhoben, zu Hause oder vor Ort im Zentrum? Erfolgt die Erhebung in zeitlichem Zusammenhang mit dem Arztkontakt? Wenn ja, vor oder nach der Visite? Wird das Ausfüllen der Fragebögen unterstützt, z. B. durch Personal oder durch eine IT-Anwendung?
- Erfassung von Diagnosen, Folgekomplikationen, Ereignissen generell: Nach welchen Vorgaben werden Diagnosen im Register kodiert, und in welchem Detailgrad? Wie detailliert werden Ereignisse dokumentiert (z. B. stummer Herzinfarkt)? Wie werden unstrukturierte Informationen (z. B. Arztbriefe) in strukturierte Datenfelder überführt?

Zu den beispielhaft genannten Punkten werden bei den befragten Registern durch verschiedene Maßnahmen Anstrengungen unternommen, Konsistenz zwischen den Zentren und damit innerhalb des Registers zu gewährleisten. Eine Konsistenz mit anderen Registern oder mit klinischen Studien steht dabei nicht im Fokus und kann auch nicht das Ziel sein, da die verschiedenen externen Datenquellen und klinischen Studien wiederum selbst diesbezüglich nicht konsistent sind. Bei einem Vergleich eines neuen Arzneimittels mit einer Vergleichstherapie unter Verwendung unterschiedlicher Datenquellen (z. B. einarmige Studie zum neuen Wirkstoff vs. Kontrolle zur Vergleichstherapie aus einem Register) ist daher je nach interessierendem Endpunkt eine Confounder-Adjustierung zur Adressierung der potenziellen Verzerrung alleine nicht ausreichend. Lediglich für Endpunkte bei denen die Erfassung sich zwischen den Quellen nicht potenziell wesentlich unterscheidet, erscheint ein Vergleich unter Verwendung unterschiedlicher Datenquellen für eine zielgerichtete Datenerhebung sinnvoll. Dazu gehören Mortalität und ggf. als Surrogat validierte, standardisierte Labor- bzw. sonstige Messwerte (z. B. die SVR bei Hepatitis C). Für sonstige Endpunkte wäre dies im Einzelfall zu prüfen und nachzuweisen. Unabhängig davon wäre eine vergleichbare Verfügbarkeit und Operationalisierung der relevanten Confounder in den verschiedenen Datenquellen notwendig.

Aus den zuvor genannten Ausführungen ergibt sich auch, dass für die versorgungsnahe Datenerhebung zum Zwecke der Nutzenbewertung vorrangig solche Register infrage kommen, bei denen Patientenmerkmale das Einschlusskriterium darstellen (z. B. Indikationsregister). Register, bei denen die konkrete Therapie (z. B. Behandlung mit einem neuen Wirkstoff) das Einschlusskriterium darstellt, sind voraussichtlich nur in Ausnahmefällen für Zwecke der Nutzenbewertung geeignet. Dies gilt insbesondere dann, wenn es sich um Produkt- oder

Interventionsregister handelt, bei denen ein Großteil vorhandener Therapieoptionen (z. B. etablierte Standardtherapien) nicht im Register betrachtet wird, da dann voraussichtlich regelhaft keine Vergleiche zum Zwecke der Nutzenbewertung innerhalb dieser Datenquelle durchgeführt werden können.

Abschätzung des notwendigen Zeitrahmens für Registerstudien

Die nachfolgende Abschätzung des notwendigen Zeitrahmens für Registerstudien stützt sich vorrangig auf die in den Gesprächen dargelegten, konkreten praktischen Erfahrungen der Registerbetreiber. Die Abschätzung beschränkt sich auf Registerstudien, die allein innerhalb des jeweiligen Registers durchgeführt werden, also ohne Anbindung an eine weitere Datenquelle. Dabei werden etwaige Verzögerungen wegen fehlender Finanzierung der Studien nicht betrachtet, da für die vorliegende Anwendungssituation davon ausgegangen wird, dass die Finanzierung mindestens durch die Auftraggeber der geplanten Studie gesichert wird.

Die Abschätzung erfolgt nachfolgend für die prospektive Datenerhebung in einem neu zu errichtenden Register sowie die prospektive Datenerhebung in einem bestehenden (und grundsätzlich geeigneten) Register. Zusätzlich wird der Zeitrahmen für eine rein retrospektive Datenerhebung in einem bestehenden (und grundsätzlich geeigneten) Register dargestellt, auch wenn aufgrund der in Abschnitt 5.3 beschriebenen Anforderungen und Limitationen dies vermutlich keinen häufigen Anwendungsfall darstellen wird. Für den wahrscheinlicheren Anwendungsfall einer kombinierten prospektiven Datenerhebung (für das zu bewertende Arzneimittel) und retrospektiven Auswertung (für die Vergleichstherapie) innerhalb derselben Datenquelle kann die Abschätzung des Zeitrahmens für retrospektive Auswertungen davon unabhängig hilfreich sein.

1) Prospektive Datenerhebung in einem neuen Register

- Notwendiger Zeitraum zum Aufsetzen eines neuen Registers: ca. 1 bis 4 Jahre, abhängig insbesondere von
 - Bereits vorhandener (ggf. vom Registerbetreiber bereits für ein anderes Register verwendeter) Software und sonstiger technischer und organisatorischer Infrastruktur für den Registerbetrieb
 - Bereitschaft zur Teilnahme in der Fläche (z. B. durch bereits bestehendes Register zu einer anderen Erkrankung im selben Fachgebiet, einen bestehenden Verbund von Fachärztinnen und -ärzte [z. B. Kompetenznetze, zertifizierte Zentren mit etablierten gemeinsamen Fachtagungen oder Versorgungszielen], anderweitig organisierte Arztgruppen [z. B. langjährige Kooperation mit Patientenvertretung])
 - Anzahl der notwendigen Versorgungsebenen zur Datenerfassung einschließlich der damit verbundenen Notwendigkeit zur dezentralen Abstimmung im Bereich Datenschutz
 - Vorliegen eines etablierten Datensatzes vs. Notwendigkeit der Definition eines Großteils der zu erhebenden Daten

- Zusätzlicher Zeitrahmen für die Registerstudie siehe unter 2), wobei die notwendigen Vorbereitungen teilweise im Zuge des Aufbaus des Registers getroffen werden
- 2) Prospektive Datenerhebung in einem bestehenden und grundsätzlich geeigneten Register
- Notwendiger Zeitraum z. B. für eine Registerstudie mit einjähriger Beobachtungsdauer: ca. 2 bis 4 Jahre. Dieser Zeitraum setzt sich zusammen aus 6 bis 24 Monaten Vorbereitungszeit, der unter anderem von Fallzahl und Rekrutierungsmöglichkeiten abhängigen Rekrutierung- und Beobachtungsdauer sowie einem Zeitraum von 3 bis 6 Monaten für die Auswertung und die Erstellung des Ergebnisberichts
 - Erläuterung des notwendigen Zeitraums für die Vorbereitung: Der Zeitbedarf hängt stark von der konkreten Fragestellung und dem damit verbundenen Zeitraum für die Protokoll- und Analyseplanerstellung plus Register-internen und -externen Genehmigungsprozessen für die Registerstudie ab. Einen teilweise großen und pauschal schwer abschätzbarer Zeitfaktor stellt zudem die ggf. notwendige Erweiterung des Datensatzes dar, weniger aus technischer, sondern primär aus organisatorischer Sicht: ggf. ist eine gesonderte Rekrutierung von Zentren erforderlich, die bereit sind, an der erweiterten Datenerfassung teilzunehmen, plus Schulung der beteiligten Zentren plus Vorlaufzeit zur Gewährleistung einer qualitativ hochwertigen Datenerfassung für die Erfassung im Rahmen der konkreten Registerstudie; ggf. ist eine Erweiterung der Einverständniserklärung mit dezentralen Abstimmungen zum Datenschutz erforderlich.
- 3) Retrospektive Auswertung in einem bestehenden und grundsätzlich geeigneten Register
- Notwendiger Zeitraum: 6 bis 18 Monate, stark abhängig von konkreter Fragestellung und dem damit verbundenen Zeitraum für die Protokoll- und Analyseplanerstellung sowie Register-internen und -externen Genehmigungsprozessen für die Registerstudie.

5.5.1.3 Kriterien für die Datenqualität bei der Generierung versorgungsnaher Daten zum Zwecke der Nutzenbewertung und Maßnahmen zur Sicherstellung der Datenqualität

Voraussetzung dafür, dass versorgungsnahe Daten für Zwecke der Nutzenbewertung von Arzneimitteln geeignet sind, ist u. a. eine angemessene Datenqualität, die durch entsprechende Maßnahmen sichergestellt werden muss. Welche Maßnahmen hierfür gemäß internationaler Standards infrage kommen und als Qualitätskriterien für Register genannt werden, ist in Abschnitt 5.5.1.1 beschrieben.

Letztlich ist jedoch nicht entscheidend und auch nicht erforderlich, dass alle dort genannten Maßnahmen vollständig umgesetzt sind, sondern dass die für die konkrete Fragestellung relevanten Daten in einer solchen Qualität vorliegen, dass die Auswertungen im Rahmen einer Registerstudie ausreichend sicher interpretierbar sind. Um dies zu gewährleisten lassen sich folgende Kriterien zur Datenqualität eines Registers unterscheiden:

- 1) Obligatorische Kriterien zur Sicherstellung der Datenqualität
- 2) Kriterien der Datenqualität, darunter
 - a) Allgemeine Kriterien, die regelhaft für Registerstudien zum Zwecke der Nutzenbewertung von Arzneimitteln relevant sind
 - b) Allgemeine Kriterien, die je nach Fragestellung für Registerstudien zum Zwecke der Nutzenbewertung relevant sein können
 - c) Kriterien, deren Erfüllungsgrad fragestellungsbezogen zu beurteilen ist

Ausgehend von den international und national definierten Qualitätskriterien, der Berücksichtigung der Gespräche mit den Registerbetreibern (siehe Abschnitt 5.5.1.2) sowie der Kenntnis über die für die Nutzenbewertung von Arzneimitteln relevanten Informationen zu Patientencharakteristika, Endpunkten, Studienablauf, Einschlusskriterien sind in der nachfolgenden Tabelle 8 diese relevanten Kriterien aufgeführt. Kriterien, die sich auf die Erstellung eines Studienprotokolls und eines Analyseplans für eine Registerstudie sowie auf konkrete Auswertungen zu einer Registerstudie beziehen (z. B. Sensitivitätsanalysen) werden dabei nicht aufgeführt, da diese im Abschnitt 5.5.2 (Methodische Anforderungen an die Auswertung) separat adressiert werden.

Tabelle 8: Kriterien für die Datenqualität und die Sicherstellung der Datenqualität für versorgungsnahe Datenerhebungen zum Zwecke der Nutzenbewertung von Arzneimitteln

Kategorie	Qualitätskriterien
Obligatorische Kriterien zur Sicherstellung der Datenqualität	<ul style="list-style-type: none"> ▪ Detaillierte Registerbeschreibung (Zielsetzung, Registerprotokoll) ▪ Exakte Definition / Operationalisierung von Expositionen, klinischen Ereignissen, Endpunkten und Confoundern ▪ Aktueller Datenplan / Kodierhandbuch ▪ Schulungen zur Datenerhebung und -erfassung ▪ Klar definierte Ein- und Ausschlusskriterien für Registerpatienten ▪ SOP-System zur Datenerhebung ▪ Maßnahmenpaket zur Sicherstellung der Richtigkeit der Daten und zur Information über Fehlerraten (z. B. source data verification, interne und externe Audits, IT-gestützte Prüfungen [z. B. Cross-Reference-Checks]) ▪ Documentation trail - Dokumentation der Prozess- und Definitionsänderungen im Register ▪ Wissenschaftliche Unabhängigkeit des Registers ▪ Nachhaltige Finanzierung
Allgemeine Kriterien, die regelhaft für Registerstudien zum Zwecke der Nutzenbewertung von Arzneimitteln relevant sind	<ul style="list-style-type: none"> ▪ Verwendung exakter Datumsangaben zum Patienten, zur Erkrankung und zu Ereignissen ▪ Detaillierte Informationen zur Arzneimitteltherapie (Wirkstoff, Dosis, Dosisänderung, inkl. Datumsangaben) ▪ Timeliness (Aktualität / schnelle Verfügbarkeit / Pünktlichkeit der benötigten Ergebnisse)
Allgemeine Kriterien, die je nach Fragestellung für Registerstudien zum Zwecke der Nutzenbewertung relevant sein können	<ul style="list-style-type: none"> ▪ Verwendung von Standard-Klassifikationen (z. B. ICD-10) und -Terminologien (z. B. MedDRA) ▪ Verwendung von validen Standard-Erhebungsinstrumenten (Fragebögen, Skalen, Tests) ▪ Flexibilität und Anpassungsfähigkeit (z. B. zur Einbettung von Studien, für weitere Datenerhebung, bei veränderter Versorgungssituation) ▪ Verknüpfbarkeit mit anderen Datenquellen
Kriterien, deren Erfüllungsgrad fragestellungsbezogen zu beurteilen ist ^a	<ul style="list-style-type: none"> ▪ Repräsentativität der Stichprobe / Selektion der Stichprobe ▪ Vollständigkeit der Daten je Erhebungszeitpunkt (loss-to-follow-up, Drop-outs) ▪ Vollständigkeit der Erhebungszeitpunkte ▪ Richtigkeit der Daten ▪ Erhebung aller für die Fragestellung relevanten Confounder ▪ Datenkonsistenz über die Zeit
<p>a: Die genannten Kriterien sind wichtige Kriterien der Datenqualität, jedoch nur fragestellungsbezogen zu beurteilen. Zum einen beziehen sich z. B. „Richtigkeit der Daten“ und „Datenkonsistenz über die Zeit“ nur auf die Daten, die für die jeweilige Fragestellung relevant sind. Zum anderen bezieht sich „Repräsentativität der Stichprobe“ nur auf die für die Fragestellung relevante Population, nicht jedoch auf die gesamte Registerpopulation.</p> <p>ICD: Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme; MedDRA: Medizinisches Wörterbuch für Aktivitäten im Rahmen der Arzneimittelzulassung; SOP: Standard Operating Procedure</p>	

Im Rahmen der Eignungsprüfung eines spezifischen Registers sollte mithilfe dieser Kriterienliste für die jeweilige konkrete Fragestellung evaluiert werden, ob und in welchem

Ausmaß die einzelnen Kriterien erfüllt sind, welchen Einfluss eine mögliche Nichterfüllung auf die Ergebnisqualität vermutlich hat, und ob im Rahmen einer registerbasierten Studie eventuelle Defizite in einem für die Beantwortung der Fragestellung zumutbaren Rahmen und mit vertretbarem Aufwand korrigiert werden können.

5.5.2 Methodische Anforderungen an die Auswertung

Die methodischen Anforderungen an Auswertung von vergleichenden Studien mit Randomisierung und von adjustierten indirekten Vergleichen über einen gemeinsamen Brückenkomparator sind etabliert [15,36]. Das vorliegende Kapitel beschreibt daher im Wesentlichen die methodischen Anforderungen an die Auswertung vergleichender Studien ohne Randomisierung. Grundlage hierfür sind zum einen publizierte wissenschaftliche Literatur, zum anderen die Diskussion mit biometrischen Sachverständigen, die in Auswertungsfragen von Patientenregisterdaten über fundierte Expertise verfügen (siehe auch Abschnitt 4.2).

Ergänzend wird am Ende des Kapitels kurz auf besondere Anforderungen an die Auswertung von vergleichenden Studien mit Randomisierung sowie auf die Zusammenführung von Daten aus unterschiedlichen Datenquellen eingegangen.

Besondere Anforderungen an die Auswertung von vergleichenden Studien ohne Randomisierung

Statistischer Analyseplan

Vergleichende Studien ohne Randomisierung zum Zweck des Vergleichs medizinischer Interventionen sollen bei der Planung den Anforderungen von vergleichenden Studien mit Randomisierung entsprechen [111] (siehe auch Abschnitt 5.3.2). Dazu gehört auch ein detaillierter Analyseplan, in dem vorab festgelegt wurde,

- welche statistischen Methoden und Modelle zur Anwendung kommen,
- welche Verfahren und Kriterien bei der Modellauswahl und -anpassung verwendet werden,
- in welchem Umfang und aus welchen Gründen fehlende Daten zu erwarten sind,
- welche Maßnahmen zur Vermeidung fehlender Daten getroffen werden,
- welche Auswertungsstrategien zum Umgang mit fehlenden Daten gewählt werden,
- wie mit unplausiblen Daten und Ausreißern umgegangen wird, und
- mit welchen Sensitivitätsanalysen die Robustheit der Ergebnisse überprüft werden.

Adjustierung für Störgrößen (Confounder)

Bei Studien ohne Randomisierung ist die für einen fairen Vergleich notwendige Strukturgleichheit der zu vergleichenden Gruppen in der Regel nicht gegeben. Gruppenunterschiede in möglichen Confoundern, d. h. Faktoren, die sowohl mit der

Behandlung als auch mit Endpunkten in Beziehung stehen und somit einen Behandlungseffekt verfälschen können, müssen daher bei der Schätzung von Effekten berücksichtigt werden. Einen zentralen Aspekt bei vergleichenden Studien ohne Randomisierung stellt daher die adäquate Adjustierung für Confounder dar, um – trotz Verzerrungspotenzial durch Confounding – interpretierbare Schätzungen des interessierenden Effekts zu erhalten. Hierzu liegen gute Forschungspraktiken der International Society for Pharmacoeconomics and Outcomes Research (ISPOR) vor, nach denen vorgegangen werden sollte [112-114]. Um eine adäquate Kontrolle für Confounder zu erreichen, ist es, unabhängig von der verwendeten Methodik, insbesondere notwendig

- vorab alle wichtigen Confounder (inklusive wichtiger Wechselwirkungen) zu identifizieren und in angemessener Form im Modell zu berücksichtigen,
- diese wichtigen Confounder in der Studie vollständig zu erheben,
- die Studie mit einem ausreichenden Stichprobenumfang zu planen, um auch alle Confounder modellmäßig verarbeiten zu können,
- das kausale Modell exakt zu beschreiben, z. B. mithilfe kausaler Grafiken,
- die Annahmen des kausalen Modells darzustellen, und
- fundiert z. B. anhand wissenschaftliche Literatur, zu begründen, warum diese Annahmen im konkreten Anwendungsfall zu rechtfertigen sind.

Falls einer oder mehrere dieser wichtigen Confounder im Datenbestand nicht enthalten sind, sind Auswertungsergebnisse auf Basis dieses Datenbestands nicht für eine Nutzenbewertung geeignet. Zudem müssen bei der Bildung des Modells zur Datenauswertung auch alle weiteren wichtigen Eigenschaften der vorliegenden Datensituation adäquat berücksichtigt werden, wie z. B. eine vorhandene hierarchische Struktur der Daten (z. B. Region, Klinik, Patient).

Methodische Ansätze zur Adjustierung für Confounder

Im Expertenworkshop mit den biometrischen Sachverständigen wurde herausgearbeitet, dass zum Zwecke der Nutzenbewertung von Arzneimitteln für die Confounderadjustierung regelhaft nur solche Verfahren sinnvoll sind, die unter Verwendung der individuellen Patientendaten durchgeführt werden. Für die Confounderadjustierung mit individuellen Patientendaten stehen verschiedene Ansätze zur Verfügung. Im Allgemeinen lassen sich insbesondere die folgenden 3 Ansätze unterscheiden [115-117]:

- Die einfachste Art ist die direkte Adjustierung nach Confoundern in einer multiplen Regressionsanalyse. Bei dieser Art der Adjustierung werden im Prinzip Strata gebildet, die sich bezüglich der Confounder (z. B. Alter, Geschlecht und Schweregrad der Erkrankung) ähneln. Für diese in sich homogenen Strata wird der Behandlungseffekt jeweils separat geschätzt, und diese Schätzungen werden anschließend in einer gemeinsamen Schätzung zusammengefasst. Dieses Verfahren ist seit langem etabliert. Es hat aber den Nachteil, dass diese Analysen je nach Fallzahl bzw. Ereigniszahl nur eine

begrenzte Anzahl an Confoundern aufnehmen können, da sie sonst unzuverlässige Resultate liefern oder mathematisch gar nicht funktionieren. Zudem kann nur nach den beobachteten Confoundern adjustiert werden.

- Eine mittlerweile ebenfalls etablierte und verbreitete Klasse an Methoden beruht auf Propensity Scores. Hier wird in einem 1. Schritt für jede Person in Abhängigkeit der beobachteten Confounder die Wahrscheinlichkeit dafür berechnet, eine der beiden zu vergleichenden Behandlungen (z. B. den zu bewertenden Wirkstoff) zu erhalten. In diesem 1. Schritt wird also der Einfluss einer größeren Menge potenzieller Confounder je Patientin bzw. Patient zu einer einzelnen Maßzahl zwischen 0 und 1, dem Propensity Score, verdichtet. Dieser Propensity Score wird in einem 2. Schritt zur Adjustierung genutzt. Hierfür existieren verschiedene Verfahren, z. B. die (direkte) Adjustierung in einer Regressionsanalyse (siehe oben), die Bildung von Personenpaaren, deren Propensity Scores sehr ähnlich sind (ein sogenanntes Matching), oder die Gewichtung anhand des Propensity Scores. Der große Vorteil der Verfahren, die auf Propensity Scores beruhen, ist, dass im Vergleich zur direkten Adjustierung deutlich mehr mögliche Confounder einbezogen werden können. Zudem können mit Hilfe von Propensity Scores Patientinnen bzw. Patienten identifiziert und ggf. ausgeschlossen werden, für die nicht beide der zu vergleichenden Therapieoptionen theoretische infrage kommen (siehe hierzu auch unten: Propensity Scores als Methode zur Berücksichtigung von Confoundern). Allerdings können sie ebenfalls nur nach den beobachteten Confoundern adjustieren.
- Eine weitere Möglichkeit der Confounderkontrolle ist die Verwendung von Instrumentenvariablen. In der Regel wird in der medizinischen Versorgung eine konkrete Therapie in Abhängigkeit prognostischer Faktoren (z. B. Schweregrad der Erkrankung) gewählt. Dies führt bei Studien ohne Randomisierung zu Inbalancen der zu vergleichenden Gruppen bezüglich dieser Faktoren. Die Idee der Methode der Instrumentenvariablen ist, dass diese die Auswahl einer Therapie beschreibt, aber nicht mit prognostischen Faktoren assoziiert ist. Z. B. könnte dies bei bestimmten medizinischen Schulen der Fall sein (Art einer Operation), wenn die Anwendung der jeweiligen Operationsmethode lediglich davon abhängt, in welchem Zentrum der Patient bzw. die Patientin (zufällig) behandelt wird, nicht jedoch von der Schwere der Erkrankung. Eine Instrumentenvariable muss daher mit der Wahrscheinlichkeit, eine bestimmte Therapie zu erhalten, hoch korrelieren, darf aber weder direkt noch über eine Assoziation mit Confoundern die Endpunkte beeinflussen. Im oben genannten Beispiel wäre es dann möglich, die verschiedenen Operationsmethoden unter der Annahme einer ausreichenden Strukturgleichheit durch Adjustierung mit der passenden Instrumentenvariable zu vergleichen. Der (theoretische) Vorteil dieses Verfahrens besteht darin, dass mit einer passenden Instrumentenvariablen sowohl nach beobachteten als auch nach nicht beobachteten Confoundern adjustiert werden kann. Praktisch sind die mit der Auswahl der Instrumentenvariablen verbundenen strengen Annahmen oft jedoch kaum überprüfbar.

Im Expertenworkshop mit den biometrischen Sachverständigen wurde herausgearbeitet, dass die Verwendung von Instrumentenvariablen, u. a. aufgrund der damit verbundenen kaum überprüfbaren Annahmen, für Nutzenbewertungen von Arzneimitteln von untergeordneter Relevanz ist. Eine häufig angewendete Methode zur Berücksichtigung von Confoundern in vergleichenden Studien ohne Randomisierung auf Basis von Registern stellen dem gegenüber Propensity Scores dar, auf die daher nachfolgend detaillierter eingegangen wird.

Propensity Scores als Methode zur Berücksichtigung von Confoundern

Wie oben beschrieben wird bei der Verwendung von Propensity Scores für jede Patientin bzw. jeden Patienten ein Propensity Score gebildet, der den Einfluss einer größeren Menge an Confoundern beschreiben soll. Bei der Auswahl des Modells zur Bestimmung des Propensity Scores, der Schätzung der Therapieeffekte mithilfe des Propensity Scores und der Interpretation der Daten sind verschiedene Punkte zu beachten. Für das allgemeine Verständnis der Verwendung von Propensity Scores zum Zwecke der Nutzenbewertung von Arzneimitteln wesentliche Punkte werden nachfolgend im Detail beschrieben.

Positivität, Überlappung und Balanciertheit

Bei der Verwendung von Propensity Scores sind 3 Begriffe von wesentlicher Bedeutung: Positivität, Überlappung und Balanciertheit.

Voraussetzung für einen Vergleich zweier Patientengruppen mithilfe von Propensity Scores ist zunächst, dass die Patientinnen und Patienten beider Gruppen theoretisch für beide interessierenden Therapien (neuer Wirkstoff bzw. Vergleichstherapie) infrage kommen. Dies bedeutet zum einen, dass die Patientengruppen durch die Fragestellung der Nutzenbewertung vorgegeben werden und z. B. Patientinnen und Patienten mit einer Kontraindikation gegen eine der untersuchten Therapien nicht in der Auswertung berücksichtigt werden dürfen (Positivität). Zum anderen bedeutet dies, dass in der zur Verfügung stehenden Datenbasis eine ausreichende Überlappung (gemessen am Propensity Score) zwischen der Population, die die eine Therapie erhalten hat (z. B. neuer Wirkstoff), und der Population, die die andere Therapie erhalten hat (Vergleichstherapie), bestehen muss. Ausreichende Überlappung bedeutet, dass sich die Verteilung der Patientinnen und Patienten auf die verschiedenen Propensity Scores ähneln muss. Wenn beispielsweise bei einem Großteil der Patientinnen und Patienten, die den neuen Wirkstoff erhalten, der Propensity Score unter 0,3 liegt, jedoch bei denen, die die Vergleichstherapie erhalten, über 0,7, dann besteht keine ausreichende Überlappung der Populationen.

Neben einer ausreichenden Überlappung ist es erforderlich, dass die Populationen in den Behandlungsgruppen ausreichend balanciert sind. Damit ist gemeint, dass sich die Behandlungsgruppen bezüglich wichtiger Confounder (z. B. Alter, Schweregrad der Erkrankung) nicht relevant unterscheiden. Dies ist durch starke Überlappung der Populationen gemessen am Propensity Score allein nicht gewährleistet, da sich die Überlappung durch Betrachtung des Propensity Scores als Gesamtmaß ergibt, nicht jedoch durch Ähnlichkeit der einzelnen Confounder im Detail. So könnten sich in der einen Patientengruppe deutlich mehr

ältere Patientinnen und Patienten befinden, während in der anderen Gruppe der Schweregrad im Mittel deutlich höher ausfällt, ohne dass dies zu bedeutsamen Unterschieden im Propensity Score führt.

Wie stark die Überlappung und die Balanciertheit der Gruppen sind, hängt zunächst von dem gewählten Modell zur Bildung des Propensity Scores ab. Sie kann aber auch durch „Trimmen“ (Ausschluss von Patientinnen und Patienten in sich nicht überlappenden Bereichen des Propensity Scores) und die Methodik zur Adjustierung beeinflusst werden. Die sich ausreichend überlappende und ausreichend balancierte Patientenpopulation ist letztlich diejenige Population, für die die geschätzten Effekte unter Anwendung des Propensity Scores gelten. Daher ist diese Population genau zu beschreiben, und es ist zu untersuchen, ob diese die für die ursprüngliche Fragestellung gewählte Population ausreichend abbildet. Ist das nicht der Fall, so gelten die geschätzten Effekte ggf. nur für eine eingeschränkte Population. Zum einen geht dann allerdings potenziell der postulierte Vorteil einer versorgungsnahen Datenerhebung (keine wesentliche Einschränkung der untersuchten Population) verloren. Zum anderen sind die Effekte möglicherweise gar nicht sinnvoll interpretierbar, z. B. wenn die aufgrund der Überlappung zustande gekommene künstliche Population im deutschen Versorgungskontext nicht abgrenzbar ist. Die notwendige Confounder-Adjustierung kann daher im konkreten Anwendungsfall zu einer deutlichen Einschränkung der Übertragbarkeit und inhaltlichen Interpretierbarkeit der Ergebnisse führen.

Unterschiedliche Methoden der Effektschätzung mittels Propensity Scores

Für die Schätzung des Therapieeffekts mithilfe von Propensity Scores gibt es unterschiedliche Methoden [118]. Bei Anwendung von Propensity Scores sind die vorhandenen Anleitungen zur Datenanalyse mithilfe dieser Methode zu beachten [118-120]. Welche Methode im jeweiligen Anwendungsfall die am besten geeignete ist, lässt sich teilweise erst anhand der konkreten Datensituation entscheiden, da unterschiedliche Methoden z. B. zu unterschiedlich starker Überlappung oder Balanciertheit führen können [121]. Allerdings kann und sollte im Analyseplan die Entscheidungsstruktur für die Auswahl des Verfahrens beschrieben werden. Dazu gehören z. B. das notwendige Mindestmaß der Überlappung und der Balanciertheit. Darüber hinaus sollten Sensitivitätsanalysen mit unterschiedlichen Propensity-Score-Methoden durchgeführt werden, sofern diese ebenfalls das notwendige Mindestmaß der Überlappung und Balanciertheit erfüllen.

Propensity Scores ersetzen nicht die Notwendigkeit der Messung aller relevanten Confounder

Auch bei der Anwendung von Propensity Scores muss man mit ungemessenem Confounding rechnen, so dass es auch bei dieser Methodik erforderlich ist, wichtige Confounder festzulegen, zu erheben und in der Auswertung zu berücksichtigen [122]. Als Ausweg hieraus wurde der hoch-dimensionale Propensity Score (hdPS) vorgeschlagen, der über eine automatische Suche eine sehr hohe Zahl an Confoundern berücksichtigt. Dies soll gewährleisten, dass wichtige ungemessene Confounder über gemessene Confounder (Proxy-Variablen) berücksichtigt werden [123,124]. Die automatische Suche widerspricht allerdings dem Prinzip, die wichtigen

Confounder systematisch auf Basis vorhandener wissenschaftlicher Literatur und Expertenwissen zu identifizieren, und es ist kaum möglich, die angewandte Methode transparent darzustellen [125]. Davon unabhängig verbleibt auch bei Anwendung des hdPS das Risiko von ungemessenem Confounding und dadurch ein Verzerrungspotenzial der entsprechenden Ergebnisse [125]. Denn die hdPS-Methode kann nur die im Datensatz erhobenen Variablen als mögliche Proxy-Variablen berücksichtigen, und es ist für den jeweiligen Anwendungsfall unklar, ob über diese alle bekannten oder unbekannten, ungemessenen Confounder berücksichtigt werden können. Die hdPS-Methode kann daher die oben beschriebene Forderung, die relevanten Confounder vollständig zu erheben und im Modell zu berücksichtigen, nicht ersetzen.

Qualitative Ergebnissicherheit und Replikation der Ergebnisse

In der Praxis verbleibt – auch bei striktem Befolgen der oben genannten Guidelines –, dass die Richtigkeit der Annahmen bezüglich Confounderadjustierung nicht vollumfänglich überprüfbar ist und ungemessene oder gänzlich unbekannte Confounder eine Rolle spielen können. Bei vergleichenden Studien ohne Randomisierung ist die Validität der Ergebnisse daher immer abhängig von den getroffenen Annahmen [126]. Für diese besteht daher grundsätzlich eine besondere, durch wissenschaftliche Literatur gestützte Begründungspflicht.

Insgesamt liefern daher Ergebnisse aus vergleichenden Studien ohne Randomisierung gemäß den Allgemeinen Methoden des Instituts regelhaft nur eine höchstens geringe qualitative Ergebnissicherheit, was bei einer Metaanalyse von mindestens 2 solcher Studien mit statistisch signifikantem Resultat maximal zu einem Anhaltspunkt für das Vorliegen eines Effekts führt [36]. Die Replikation der Ergebnisse kann auch innerhalb desselben Datensatzes z. B. durch Auswertung nach Regionen bzw. innerhalb derselben Datenquelle durch eine 2. unabhängige Stichprobe angestrebt werden.

Auch wenn Studien ohne Randomisierung nur eine geringe qualitative Ergebnissicherheit aufweisen, können sie bei Zusammenführung mit anderen Daten die Ergebnissicherheit der Gesamtaussage zum Zusatznutzen erhöhen. Dies ist jeweils im Einzelfall zu beurteilen. Beispielhaft genannt seien hier ergebnissichere Daten zu wichtigen Endpunkten, die durch die Studie ohne Randomisierung in anderen Endpunkt-kategorien ergänzt werden, oder die gemeinsame Betrachtung einer (kleinen) Studie mit Randomisierung mit einer (größeren) Studie ohne Randomisierung.

Quantifizierung / Größenordnung des Effekts

Auch bei sorgfältigster Analyse unter Erfüllung der genannten Qualitätsanforderungen sollte aufgrund potenziell unbekannter Confounder aus den in der Studie beobachteten Effekten erst ab einer bestimmten Effektstärke eine Aussage zum Nutzen oder Schaden einer Intervention abgeleitet werden. Eine (positive oder negative) Aussage zum Nutzen oder Schaden ergibt sich dann, wenn das Konfidenzintervall für den beobachteten Effekt ober- bzw. unterhalb einer zu definierenden Schwelle liegt. Da die Erfüllung der genannten Qualitätsanforderungen Voraussetzung für die Betrachtung der Effekte ist, sollte dieser Schwellenwert deutlich

unterhalb des Wertes für den „dramatischen Effekt“ (relatives Risiko von 5-10 [36]) liegen, z. B. in einem Bereich von 2-5 für das relative Risiko. Die konkrete Schwelle ergibt sich durch die Qualität der Daten im Einzelfall, u. a. durch das Wissen über relevante Confounder. Je nach Datenlage ist eine solche Schwelle auch endpunktspezifisch, z. B. aufgrund fehlender Verblindung der Therapien oder unterschiedliche Verzerrungsrichtung für positive bzw. negative Effekte, anzuwenden.

Bei Nutzenbewertungen von Arzneimitteln gemäß § 35a SGB V ist das Ausmaß des Zusatznutzens zu quantifizieren [17]. Für die Einstufung in die Ausmaßkategorien gering, beträchtlich und erheblich sind dabei unterschiedliche Effektstärken erforderlich. Für diese Einstufung kann die derzeit bei Dossierbewertungen des IQWiG verwendete Methode nicht unverändert übernommen werden, da diese von Daten mit höherer Ergebnissicherheit ohne Notwendigkeit eines Schwellenwerts für Aussagen zum Nutzen oder Schaden ausgeht [127].

Ausgehend von der oben genannten Schwelle für eine Aussage zum Nutzen oder Schaden liegt bei Überschreiten dieser Schwelle mindestens ein geringer Zusatznutzen für den jeweiligen Endpunkt vor (Ausnahme: Endpunkte der Kategorie nicht schwerwiegende / nicht schwere Folgekomplikationen, weil für diese gemäß Arzneimittel-Nutzenbewertungsverordnung zusätzlich eine „nicht nur geringfügige Verbesserung“ gefordert wird [17]). Die Einstufung in die Ausmaßkategorien „beträchtlich“ oder „erheblich“ bedarf für alle Endpunktkatoren höhere, d. h. oberhalb der oben genannten Schwelle liegende, nach Größe abgestufte Effektstärken.

Besondere Anforderungen an die Auswertung von vergleichenden Studien mit Randomisierung

Die Anforderungen an die Auswertung von vergleichenden Studien mit Randomisierung sind in zahlreichen Guidelines der Arzneimittelzulassungsbehörden beschrieben und stehen nicht im Fokus dieses Berichts. Beispielhaft sei hier verwiesen auf die Guidelines des International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) [11], hierbei insbesondere auf die Guideline „Statistical Principles for Clinical Trials“ [128], sowie die Biostatistik-Guidelines der European Medicines Agency (EMA) [129].

Zu beachten ist, dass im Vergleich mit Studien ohne Randomisierung der Aufwand der umfänglichen Confounder-Adjustierung entfällt, da die Confounderkontrolle durch die Randomisierung gewährleistet ist. Ebenso entfällt die Notwendigkeit, den Stichprobenumfang der Studie so zu erhöhen, dass eine umfängliche Confounder-Adjustierung auch modellmäßig möglich ist.

Zusammenführung von Analysen aus mehreren Datenquellen

Insbesondere für seltene Erkrankungen kann es sinnvoll und notwendig sein, dass Studien in internationaler Kooperation durchgeführt werden. In diesem Fall sind insbesondere folgende Grundsätze zu beachten:

- Sofern für die Analysen mehrere Datenquellen genutzt werden (z. B. Register aus den beteiligten Staaten), ist eine Harmonisierung der Daten auf inhaltlicher und struktureller Ebene erforderlich. Dies erfordert ein standardisiertes Vorgehen, z. B. gemäß Fortier 2016 (Maelstrom Research Guidelines [130]). Das dort für die retrospektive Datenharmonisierung beschriebene Vorgehen kann und sollte im Grundsatz auch für prospektive Studien vor Studienbeginn angewendet werden.
- Bei Analysen aus mehreren Registern kann es sinnvoll sein, für das gemeinsame Projekt nicht einen gemeinsamen Datenpool zu bilden und diesen dann auszuwerten, sondern identisch geplante Studien in den einzelnen Registern zu planen und durchzuführen und diese Studien dann metaanalytisch zusammenzufassen. Dies verringert zum einen Anforderungen an den Datenschutz, da die individuellen Patientendaten nur den ohnehin Zugriffsberechtigten zur Verfügung stehen müssen. Zum anderen können etwaige Effektunterschiede zwischen den landesspezifischen Studien im Rahmen der Metaanalyse explizit gemacht und adressiert werden. Diese Vorteile sind mit möglichen Nachteilen im Vergleich zu einer statistischen Analyse eines einzelnen Datenpools abzuwägen [131].
- Bei Analysen, die außerhalb des interessierenden deutschen Versorgungskontextes generierte Daten heranziehen, ist zu begründen, dass die Daten versorgungsnah im Sinne der Versorgung in Deutschland sind bzw. Abweichungen davon für die Effektschätzung nicht relevant sind. In diesem Zusammenhang ist auf ggf. unterschiedliche Versorgungspfade, Begleittherapien, Resistenzlagen (bei Antibiotikatherapie) etc. zu achten. Mittels Heterogenitätstests bei der metaanalytischen Zusammenfassung der Studien aus unterschiedlichen Ländern (siehe vorhergehenden Punkt) kann diese Begründung unterstützt werden.

5.5.3 Anforderungen an die Berichterstattung

Eine adäquate Berichterstattung über die Methoden und Ergebnisse von Studien ist integraler Bestandteil qualitativ hochwertiger Forschung [12]. Die Qualität der Dokumentation einer Studie bestimmt ganz wesentlich die Aussagekraft der erzielten Ergebnisse. Die vollständige Dokumentation umfasst dabei das Studienprotokoll (Planung der Methodik und Durchführung der Studie), den Analyseplan (Planung der Auswertung der Daten) und den Ergebnisbericht (Beschreibung der geplanten Methodik [inkl. Auswertung] und Durchführung der Studie, der Abweichungen von dieser Planung und der vollständigen Ergebnisse). Darüber hinaus wird, auch im regulatorischen Umfeld, diskutiert, inwieweit die kontrollierte Verfügbarkeit der anonymisierten individuellen Patientendaten Bestandteil der Dokumentation einer Studie sein sollte [132-134].

Das Studienprotokoll und der Analyseplan dienen dabei neben der Beschreibung der Methodik und Durchführung der Studie auch der Präspezifizierung der Studienplanung. Das bedeutet, dass das Vorgehen in der Studie vor Erhebung der Daten und damit ohne Kenntnis der Daten bestimmt wird. Dadurch wird verhindert, durch Wahl der Methoden unter Kenntnis der Daten

(bewusst oder unbewusst) Einfluss auf die erzielten Ergebnisse zu nehmen (siehe auch Abschnitt 5.3.2).

Während in den vorangegangenen Abschnitten die Anforderungen davon geprägt waren, ob eine vergleichende Studie mit oder ohne Randomisierung durchgeführt werden sollte, ist bezüglich der Möglichkeiten der Präspezifizierung der Studienplanung als wesentlichem Charakteristikum aussagekräftiger Studienergebnisse primär von Bedeutung, ob eine vergleichende Studie mit prospektiver oder retrospektiver Datenerhebung vorliegt.

Prospektive vergleichende Studien

Die Präspezifizierung der Methodik prospektiver vergleichender Studien mit oder ohne Randomisierung kann einfach erreicht und überprüft werden. Hier ist lediglich notwendig, das Studienprotokoll und den Analyseplan vor Beginn der Datenerhebung nachvollziehbar dokumentiert zu finalisieren. Für kontrollierte Änderungen im Studienverlauf ist die Dokumentation durch Studienprotokoll-Amendments etabliert.

Für eine nachvollziehbare Dokumentation wird empfohlen, das Studienprotokoll und den Analyseplan in einem Studienregister vor Beginn der Datenerhebung zu publizieren. Mit dem Studienregister ClinicalTrials.gov eine geeignete Datenbank zur Verfügung, die eine Registrierung mit Ablage von Studienprotokollen und Analyseplänen für alle Studientypen erlaubt [21].

Retrospektive vergleichende Studien (einschließlich Studien mit retrospektiven Vergleichsarmen)

Die Transparenz der Studienplanung und der Analysemethoden vor der Durchführung von retrospektiven Studien ist weniger leicht zu gewährleisten. Da die Daten bereits vorliegen, kann nicht abschließend sichergestellt werden, dass die Studienplanung ohne Kenntnis der Daten erfolgt ist.

Auch die FDA weist in diesem Zusammenhang auf Probleme insbesondere bei retrospektiven Beobachtungsstudien hin. So beschreibt die FDA in ihrem Papier zu Real-World Evidence [13]: „Das mögliche Fehlen der Transparenz der Vorab-Planung der Studien (insbesondere bei retrospektiven Beobachtungsstudien) verbunden mit der Tatsache, dass retrospektive Analysen in elektronischen Datensätzen relativ kostengünstig mit variierenden Studiendesignelementen wiederholt werden können, macht es möglich, zahllose retrospektive Studien durchzuführen, bis das gewünschte Ergebnis erreicht wird, und dann ausschließlich die positiven Ergebnisse einzureichen, als ob sie das Ergebnis einer einzelnen Studie mit einer präspezifizierten Studienplanung seien“ (eigene Übersetzung). Aktuell hat auch die FDA keine Lösung für dieses Problem, Anforderungen an die Berichterstattung für solche Studien werden noch erarbeitet [13].

Unabhängig von diesem grundsätzlichen Problem wird empfohlen, auch für retrospektive vergleichende Studien ein Studienprotokoll und ein Analyseplan in einem Studienregister zu

publizieren. Auch für diesen Studientyp ist die Registrierung und Veröffentlichung von Studienprotokoll und Analyseplan in ClinicalTrials.gov möglich.

Besondere Anforderungen an die Berichterstattung vergleichender Studien ohne Randomisierung

Methodisch soll bei der Planung von vergleichenden Studien ohne Randomisierung zur Ermittlung von Therapieeffekten eine explizite Nachbildung der Planung einer vergleichenden Studie mit Randomisierung erfolgen (siehe Abschnitt 5.3.2). Diese Emulation einer Zielstudie soll im Studienprotokoll explizit beschrieben werden. Der Studienbericht soll darüber hinaus beschreiben, in welchem Ausmaß die Nachbildung gelungen ist (z. B. bez. der Abbildung der Ein-/Ausschlusskriterien).

Wie in den Abschnitten 5.3.2 und 5.3.4 dargelegt, ist für vergleichende Studien ohne Randomisierung insbesondere die Annäherung an die Strukturgleichheit der zu vergleichenden Behandlungsgruppen durch Adjustierung für relevante Confounder notwendig. Für die Berichterstattung bedeutet das, dass die relevanten Confounder in der Studienplanung benannt werden müssen und dass die Durchführung der Confounderkontrolle im Analyseplan soweit wie möglich präspezifiziert werden muss. Der Ergebnisbericht der Studie muss Angaben zur Verfügbarkeit der Daten zu relevanten Confoundern enthalten. Im Ergebnisbericht ist zudem eine transparente Darstellung des Vorgehens in der Analyse und der nicht adjustierten und adjustierten Vergleiche der Behandlungsgruppen notwendig [135-137].

5.6 Bewertung der Konzepte der Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung nach § 35a SGB V

Mögliche Studiendesigns für die Generierung versorgungsnaher Daten zum Zwecke der Nutzenbewertung

Der vorliegende Bericht untersucht die Generierung versorgungsnaher Daten zum Zwecke der Nutzenbewertung. Aus dieser Zielsetzung ergibt sich unmittelbar, dass Daten erhoben werden müssen, die einen Vergleich zwischen Patientengruppen ermöglichen, die mit unterschiedlichen Interventionen behandelt werden (siehe Abschnitt 5.2).

Die Generierung vergleichender versorgungsnaher Daten ist nicht an ein bestimmtes Studiendesign gebunden, sondern kann auf Basis unterschiedlicher Studiendesigns durchgeführt werden. Insbesondere können versorgungsnahe Daten sowohl in vergleichenden Studien ohne Randomisierung als auch in solchen mit Randomisierung erhoben werden [9,27,138-140].

Art und Umfang der Datenerhebung

Ziel der Erhebung versorgungsnaher Daten für die Nutzenbewertung von Arzneimitteln ist es, hinreichend sichere Aussagen zum Nutzen und Schaden der zu bewertenden Arzneimittel sowohl unter den Bedingungen der Versorgung als auch unter Berücksichtigung der Fragestellung der Nutzenbewertung machen zu können.

Dieses Ziel verlangt nicht, dass die Datenerhebung sich auf solche Daten beschränkt, die im Versorgungsalltag auch ohne den Zweck, Informationen für eine Nutzenbewertung zu generieren, erfasst werden. Eine derart falsch verstandene Beschränkung der Datenerhebung würde vielmehr das Ziel der Nutzenbewertung gefährden. Für die Nutzenbewertung sind regelhaft Daten notwendig, die im Behandlungsalltag nicht bei allen Patientinnen und Patienten dokumentiert werden (z. B. Daten zur gesundheitsbezogenen Lebensqualität, zu Symptomen oder Nebenwirkungen, siehe auch Abschnitte 5.4.3 und 5.4.4). Eine Analyse der vom G-BA benannten Evidenzlücken in Bewertungen von Orphan Drugs unterstreicht die Notwendigkeit, Daten zu allen Endpunktekategorien der Nutzenbewertung zu erheben (siehe Abschnitt 5.7.2).

Der Umfang der ergänzenden Datenerhebung hängt u. a. vom gewählten Studiendesign ab. Wie in den Abschnitten 5.3.2 und 5.5.2 ausgeführt, kann insbesondere für vergleichende Studien ohne Randomisierung eine ergänzende Datenerhebung notwendig werden, da außer Daten zu Endpunkten auch Informationen, die eine Confounderkontrolle ermöglichen, gesammelt werden müssen.

Die Erhebung der notwendigen Daten kann dabei in den Behandlungsalltag integriert werden. Die Bewertung der möglichen Instrumente zur Datenerhebung hat ergeben, dass Register, ggf. ergänzt durch eine studienindividuelle Erhebung der nicht im Register verfügbaren Daten, aktuell und für absehbare Zeit die realistischste Möglichkeit darstellen (siehe Abschnitt 5.4). Register bieten am ehesten die Möglichkeit der Anpassung an die notwendige Datenerhebung. Das betrifft sowohl die Spezifizierung der notwendigen Daten als auch die Datenqualität.

Auswertung der versorgungsnah erhobenen Daten

Die Anforderungen an die Auswertung versorgungsnaher Daten werden maßgeblich durch das verwendete Studiendesign bestimmt. Dabei stellt die Annäherung an eine hinreichende Strukturgleichheit der Behandlungsgruppen durch Adjustierung für Confounder bei vergleichenden Studien ohne Randomisierung die primäre Herausforderung dar.

Aus der Notwendigkeit für die Adjustierung für Confounder in der Auswertung ergeben sich eine Reihe von Anforderungen an den zu erhebenden Datensatz. Dazu gehört z. B. die Prä-Spezifizierung und Erhebung der relevanten Confounder und eine ausreichende Studiengröße, um die Adjustierung adäquat durchführen zu können. Insbesondere die ausreichende Studiengröße für die Adjustierung von Confoundern fehlt aktuell in der Diskussion von vergleichenden Studien ohne Randomisierung, dabei ist gerade diese Anforderung für Studien in kleinen Populationen, für die dieser Studientyp besonders häufig vorgeschlagen wird, von Bedeutung. So kann in der Regel bei einer Studiengröße von weniger als 100 Patientinnen und Patienten keine adäquate Confounderkontrolle durchgeführt werden. Gerade für Studien in kleinen Populationen sind effiziente Studiendesigns notwendig.

Aktuell ist die Planung vergleichender Studien ohne Randomisierung und die damit verbundene Datenerhebung in vielen Fällen unzureichend, um hochwertige Auswertungen durchführen zu können [141]. Die Qualitätsmängel sind u. a. darauf zurückzuführen, dass die gewählten

Datenquellen die notwendigen Informationen z. B. zu Confoundern oder Endpunkten nicht enthalten [142]. In Abschnitt 6.2 sind daher Handlungsempfehlungen für Registerbetreiber und Verantwortliche für Registerstudien aufgeführt, die der Vorbereitung auf zukünftige Datenerhebungen und Auswertungen zum Zwecke der Nutzenbewertung von Arzneimitteln dienen.

Aussagekraft der versorgungsnah erhobenen Daten

Die Aussagekraft der erhobenen Daten ergibt sich nach den Standards der evidenzbasierten Medizin auch bei einer versorgungsnahen Datenerhebung aus der Kombination von qualitativer und quantitativer Ergebnissicherheit.

Bezüglich der qualitativen Ergebnissicherheit ist wissenschaftlicher Konsens, dass eine vergleichende Studie ohne Randomisierung bei höchstens moderaten Effekten nicht die Aussagekraft einer vergleichenden Studie mit Randomisierung erreichen kann. Deswegen ist eine anwendungsbegleitende Datenerhebung auf Basis von Studien ohne Randomisierung nicht sinnvoll, wenn auf Basis der vorhandenen Evidenz zu erwarten ist, dass zwischen neuem Wirkstoff und Vergleichstherapie keine relevanten Unterschiede bestehen.

Es stellt sich vielmehr die Frage, in welchem Ausmaß die inhärente Unsicherheit der vergleichenden Studie ohne Randomisierung durch Maßnahmen zur Annäherung an die Strukturgleichheit der Behandlungsgruppen verringert werden kann.

Wenn in bestimmten Fällen bei der Nutzenbewertung die inhärent erhöhte Unsicherheit vergleichender Studien ohne Randomisierung in Kauf genommen werden soll, so ergeben sich aus der Bewertung der Konzepte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung folgende Eckpunkte:

- aus vergleichenden Studien ohne Randomisierung können keine Effekte abgeleitet werden, wenn die Datenqualität in den verwendeten Datenquellen und die Qualität der Auswertung und Berichterstattung nicht hoch ist (Ausnahmefälle sind unter bestimmten Umständen Effektstärken, die so groß sind, dass sie durch Störgrößen nicht mehr plausibel allein erklärt werden können)
- auch unter hohen Qualitätsanforderungen (für Daten, Auswertung und Berichterstattung) kann im Regelfall aus vergleichenden Studien ohne Randomisierung nicht mehr als ein Anhaltspunkt für einen Effekt abgeleitet werden
- wegen der inhärenten Unsicherheit der Ergebnisse aus vergleichenden Studien ohne Randomisierung durch verbleibende unbekannte Confounder kann selbst bei hohen Qualitätsanforderungen aus den Effekten erst ab einer bestimmten Effektstärke eine Aussage zum Nutzen oder Schaden einer Intervention abgeleitet werden; eine Quantifizierung eines Zusatznutzens gemäß der gesetzlich vorgegebenen Ausmaßkategorien erfordert entsprechende nach der Größe abgestufte Effektstärken

5.7 Vorschläge für ein Vorgehen für die anwendungsbegleitende Datenerhebung

Die vorangehenden Kapitel dieses Berichts beschreiben Konzepte zur versorgungsnahen Datenerhebung zum Zweck der Nutzenbewertung nach § 35a SGB V. Im Folgenden werden die Ergebnisse des Projekts im Zusammenhang mit der durch das GSAV eingeführten Möglichkeit der anwendungsbegleitenden Datenerhebung diskutiert.

5.7.1 Definition der Fragestellungen für die anwendungsbegleitende Datenerhebung

Die Grundlage für die anwendungsbegleitende Datenerhebung ist die Definition der Fragestellung, die durch diese Datenerhebung beantwortet werden soll. Die Fragestellung beinhaltet mindestens die Komponenten des PICO-Schemas und die notwendige Dauer der Datenerhebung. Die genauen Spezifika der Fragestellung werden aus der Evidenzlücke abgeleitet, die sich in der Nutzenbewertung ergeben hat und die durch die Datenerhebung geschlossen werden soll. Eine so definierte Fragestellung ist auch der Ausgangspunkt für die Beschreibung des notwendigen Umfangs der Datenerhebung (einschließlich Beobachtungsdauer und Fallzahlplanung).

Tabelle 9 beschreibt die Komponenten der Fragestellung.

Tabelle 9: Definition der Fragestellung für eine anwendungsbegleitende Datenerhebung zum Zwecke der Nutzenbewertung nach § 35a SGB V

Bestandteil der Fragestellung	Ausprägung der Komponenten abhängig von der in der Nutzenbewertung identifizierten Evidenzlücke
Patientenpopulation (P)	<ul style="list-style-type: none"> ▪ Patientenpopulation gemäß dem zugelassenen Anwendungsgebiet oder ▪ eingeschränkte Patientenpopulation, für die die in der Nutzenbewertung identifizierte Evidenzlücke besteht
Neues Arzneimittel (I)	<ul style="list-style-type: none"> ▪ zu bewertendes Arzneimittel
Zweckmäßige Vergleichstherapie (C)	<ul style="list-style-type: none"> ▪ zweckmäßige Vergleichstherapie
Patientenrelevante Endpunkte (O)	<ul style="list-style-type: none"> ▪ patientenrelevante Endpunkte zur Mortalität, Morbidität und gesundheitsbezogenen Lebensqualität oder ▪ spezifische patientenrelevante Endpunkte, die die in der Nutzenbewertung identifizierte Evidenzlücke schließen können
Dauer der Datenerhebung	<ul style="list-style-type: none"> ▪ Dauer der Datenerhebung abhängig von der Therapiesituation (z. B. Dauertherapie vs. akute Therapie) für die eine Evidenzlücke geschlossen werden soll
SGB: Sozialgesetzbuch	

Beispiele für Evidenzlücken wären:

- **Patientenpopulation:** für das Arzneimittel besteht ein besonderer Bedarf bei Patientinnen und Patienten mit einer schweren Ausprägung der von der Indikation umfassten Erkrankung, für die Nutzenbewertung standen aber ausschließlich Daten für Patientinnen

und Patienten mit einer leichten Ausprägung der Erkrankung zur Verfügung, Daten für die Patientenpopulation mit schwerer Erkrankung fehlen

- **neues Arzneimittel:** ein neues Arzneimittel wurden in den Studien, die für die Nutzenbewertung zur Verfügung standen, in einer anderen Darreichungsform als in der zugelassenen Anwendung eingesetzt, es fehlen damit Ergebnisse zum Vergleich der zugelassenen Darreichungsform mit der zweckmäßigen Vergleichstherapie
- **zweckmäßige Vergleichstherapie:** für die Nutzenbewertung lagen keine aussagekräftigen Daten für einen Vergleich des neuen Arzneimittels mit der zweckmäßigen Vergleichstherapie vor
- **patientenrelevante Endpunkte:** es fehlen Daten zur gesundheitsbezogenen Lebensqualität, obwohl die vorliegende Therapiesituation (z. B. palliative Versorgung) oder die Ergebnisse der Nutzenbewertung (z. B. Überlebensvorteil bei gleichzeitig verstärktem Auftreten schwerer Nebenwirkungen) unterstreichen, dass Ergebnisse zur gesundheitsbezogenen Lebensqualität relevant für Therapieentscheidungen sind
- **Dauer der Datenerhebung:** für ein Arzneimittel für eine chronische Erkrankung, dass sehr lange oder dauerhaft eingenommen werden muss, liegen für die Nutzenbewertung ausschließlich Daten für einen sehr begrenzten Behandlungszeitraum vor, Nutzen und Schaden bei längerer Behandlung bleiben unklar

5.7.2 Evidenzlücken in bisherigen Nutzenbewertungen am Beispiel „Orphan Drugs“

Anwendungsbegleitende Datenerhebungen sollen zum Zweck der Nutzenbewertung durchgeführt werden und darauf ausgerichtet sein, den Zusatznutzen zu quantifizieren [1]. Eine zielgerichtete anwendungsbegleitende Datenerhebung muss deshalb solche Evidenzlücken adressieren, die für die Nutzenbewertung und die damit verbundene Festlegung des Ausmaßes des Zusatznutzens von Bedeutung sind.

Um besser einschätzen zu können, welche Evidenzlücken bei der Nutzenbewertung bestehen und für die Festlegung des Ausmaßes des Zusatznutzens von Bedeutung sind, wurden Beschlüsse des G-BA zu Nutzenbewertungen von Orphan Drugs näher betrachtet. Untersucht wurden dabei Beschlüsse zu Orphan Drugs mit Marktzugang in den Jahren 2014 bis 2018 einschließlich Beschlüssen zu neuen Anwendungsgebiete von Orphan Drugs in diesem Zeitraum. Nicht betrachtet wurden erneute Bewertungen desselben Anwendungsgebiets, also z. B. nach Ablauf einer etwaigen Befristung oder nach Überschreiten einer Jahresumsatzgrenze von 50 Mio. €

Im nachfolgenden Abschnitt werden die Beschlüsse des G-BA zunächst zusammenfassend analysiert. Ziel ist es, Kategorien von Evidenzlücken zu identifizieren sowie etwaige Zusammenhänge zwischen Evidenzlücken und den zum Zeitpunkt des Marktzugangs verfügbaren Daten und der Prävalenz der Erkrankung zu beschreiben.

Zusammenfassende Analyse der Beschlüsse zu Orphan Drugs mit Marktzugang 2014 bis 2018

Für die zusammenfassende Analyse waren insgesamt 67 Beschlüsse zu Orphan Drugs in dem genannten Zeitraum relevant. In diesen 67 Beschlüssen wurden für 85 verschiedene Fragestellungen Aussagen zum Ausmaß des Zusatznutzens getroffen. Eine Aufstellung der Beschlüsse und der zugehörigen Fragestellungen findet sich in Anhang D.

Gemäß § 35a Abs. 1 Satz 11 SGB V gilt der Zusatznutzen für Orphan Drugs bei Marktzugang als belegt. Dem entsprechend können für das Ausmaß des Zusatznutzens bei Orphan Drugs die in der Arzneimittel-Nutzenbewertungsverordnung aufgeführten Kategorien „erheblich“, „beträchtlich“ oder „gering“ (quantifizierter Zusatznutzen) oder „nicht quantifizierbar“ verwendet werden [17]. Abbildung 5 zeigt die Aufteilung der Fragestellungen in solche, in denen der Zusatznutzen quantifiziert wurde, und solche, bei denen der Zusatznutzen als „nicht quantifizierbar“ eingestuft wurde. Ergänzt wurde dabei jeweils, wie häufig unterschiedliche Studiendesigns gemäß Beschluss des G-BA Grundlage dieser Einstufung waren.

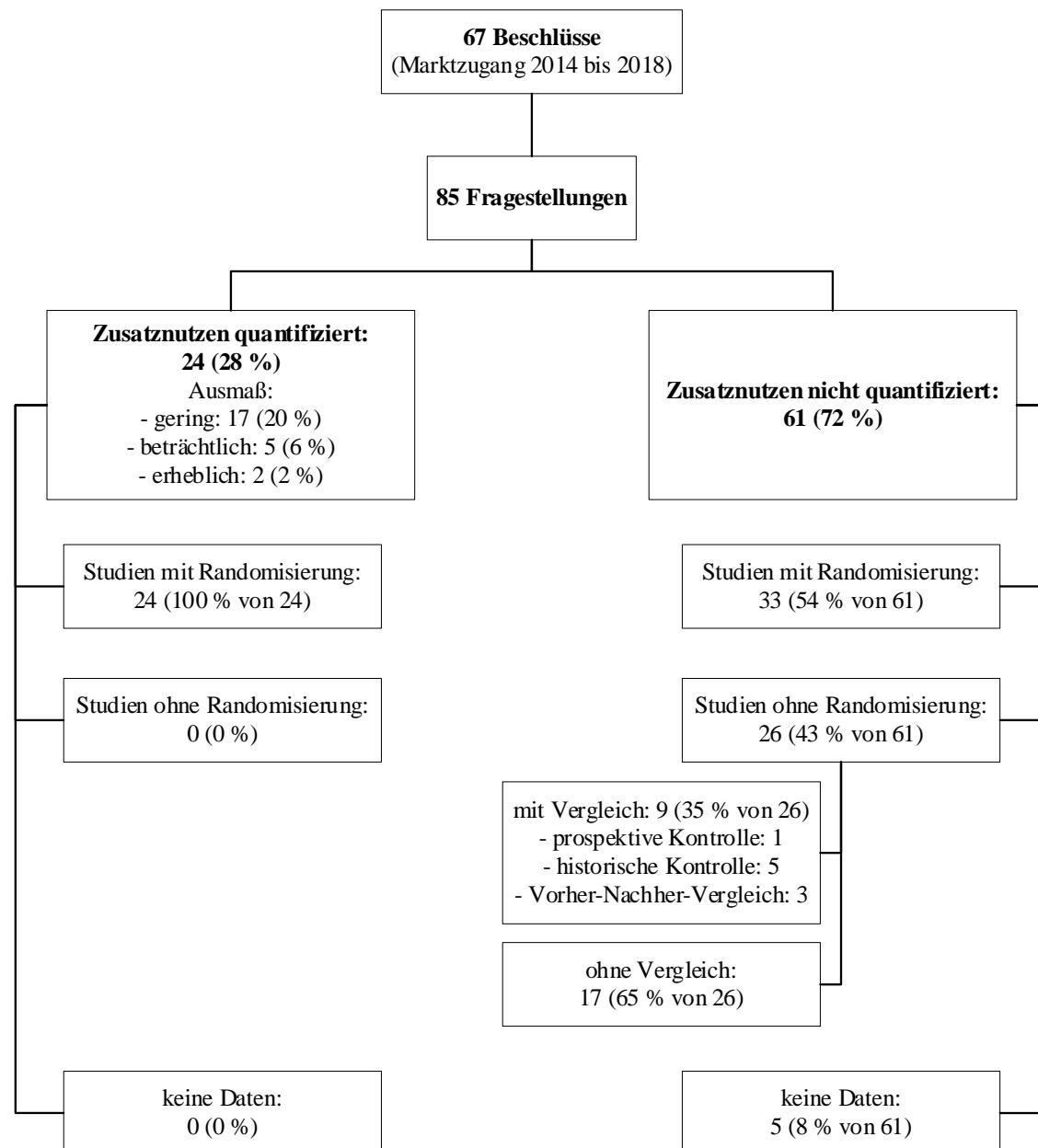


Abbildung 5: Quantifizierung des Zusatznutzens und verfügbare Daten – Beschlüsse des G-BA zu Orphan Drugs bei Marktzugang 2014 bis 2018

Nur bei ca. einem Viertel der Fragestellungen wurde der Zusatznutzen quantifiziert (24 von 85 Fragestellungen, 28 %). In diesen Fällen stellten ausschließlich Studien mit Randomisierung die Datengrundlage dar. Bei keiner Fragestellung wurde der Zusatznutzen auf Basis vergleichender Studien ohne Randomisierung quantifiziert.

Bei ca. 3 Viertel der betrachteten Fragestellungen wurde der Zusatznutzen nicht quantifiziert (61 von 85 Fragestellungen, 72 %). Für die meisten dieser Fragestellungen (92 %) lagen Daten vor, wobei es sich mehrheitlich um Studien mit Randomisierung handelte (33 der 61 Fragestellungen, 54 %). Für 26 dieser 61 Fragestellungen (43 %) lagen Daten aus Studien ohne

Randomisierung vor, in 3 Fällen zusätzlich zu einer Studie mit Randomisierung. Dabei handelte es sich häufig um Daten allein zum Orphan Drug (einarmige Studien) und damit nicht um vergleichende Daten. In einigen Fällen lagen auch Daten für eine Vergleichsgruppe vor, zumeist als historische Kontrolle. Für 5 der 61 Fragestellungen mit nicht quantifizierbarem Zusatznutzen (8 %) lagen gar keine Daten vor.

In der nachfolgenden Tabelle 10 werden die Größe der Zielpopulation gemäß G-BA-Beschluss und die dem G-BA-Beschluss zugrunde liegenden Daten (Studientypen) gegenübergestellt, getrennt nach Fragestellungen mit Quantifizierung des Zusatznutzens und Fragestellungen mit nicht quantifizierbarem Zusatznutzen.

Tabelle 10: Größe der Zielpopulation und Datengrundlage für die Nutzenbewertung – G-BA-Beschlüsse zu Orphan Drugs mit Marktzugang 2014 bis 2018

Studientyp	Größe der Zielpopulation ^a		
	Fragestellungen mit quantifiziertem Zusatznutzen N Median [IQR]	Fragestellungen mit nicht quantifizierbarem Zusatznutzen N Median [IQR]	Jeglicher Zusatznutzen N Median [IQR]
Studien mit Randomisierung	24 Fragestellungen 428 [144; 3255]	33 Fragestellungen 815 [350; 2300]	57 Fragestellungen 433 [255; 2450]
Studien ohne Randomisierung	0 Fragestellungen	26 Fragestellungen 165 [65; 500]	26 Fragestellungen 165 [65; 500]
Keine Daten	0 Fragestellungen	5 Fragestellungen 250 [26; 455]	5 Fragestellungen 250 [26; 455]
Gesamt (jeglicher Studientyp)	24 Fragestellungen 428 [144; 3255]	61 Fragestellungen 433 [120; 865]	85 Fragestellungen 433 [120; 1400]

a: Angabe gemäß G-BA-Beschluss; bei Angabe einer Spanne im Beschluss wurde der jeweilige Mittelwert herangezogen; bei einer für mehrere Fragestellungen gültigen Angabe wurde diese zwischen den Fragestellungen gleichmäßig aufgeteilt.
IQR: Interquartilsbereich.

Aus der Gegenüberstellung lässt sich einerseits ableiten, dass Studien ohne Randomisierung häufiger bei Fragestellungen mit kleiner Zielpopulation vorgelegt wurden, dass aber andererseits eine kleine Zielpopulation nicht per se ursächlich für eine fehlende Quantifizierbarkeit des Zusatznutzens war.

Charakterisierung der Evidenzlücken für Fragestellungen mit nicht quantifizierbarem Zusatznutzen

Für die Charakterisierung von Evidenzlücken werden nachfolgend die 61 Fragestellungen näher betrachtet, bei denen der G-BA einen nicht quantifizierbaren Zusatznutzen festgestellt hat. Hintergrund ist, dass in diesen Fällen die wissenschaftliche Datengrundlage nicht ausreichend war, um den Zusatznutzen quantifizieren zu können [17]. Die anwendungsbegleitende Datenerhebung soll gemäß § 130b Abs. 3 SGB V darauf ausgerichtet

sein, den Zusatznutzen zu quantifizieren, denn wenn sich anhand der gewonnenen Daten keine Quantifizierung des Zusatznutzens belegen lässt, ist ein geringerer Erstattungsbetrag für ein Orphan Drug zu vereinbaren [1].

Die Charakterisierung der Evidenzlücken wurde anhand einer Textanalyse der Tragenden Gründe zum jeweiligen Beschluss des G-BA vorgenommen, indem die dort vorrangig für die fehlende Quantifizierbarkeit des Zusatznutzens genannten Gründe extrahiert und kategorisiert wurden. Folgende Aspekte wurden in diesem Zusammenhang betrachtet:

- Bezieht sich die vom G-BA beschriebene Evidenzlücke sowohl auf den zu bewertenden Wirkstoff (Orphan Drug) als auch die Kontrollgruppe (z. B. bei Evidenzlücken in Studien mit Randomisierung) oder nur auf die Kontrollgruppe (z. B. wenn Daten zur Kontrollgruppe vollständig fehlten)? Welches Ausmaß hatte die Evidenzlücke (Daten fehlen vollständig, Daten fehlen teilweise)?
- In welcher Endpunkt-kategorie besteht die Evidenzlücke (Mortalität, Morbidität, gesundheitsbezogene Lebensqualität, Nebenwirkungen)?
- Welcher Grund wird für die Evidenzlücke genannt (Daten lagen nicht vor [z. B. weil sie gar nicht erhoben wurden], Daten wurden in unzureichender Qualität oder Menge erhoben, Datenauswertung war mangelhaft)?
- Für Fragestellungen, bei denen in den Tragenden Gründen keine wesentlichen Datenlücken benannt wurden: Welche sonstigen Gründe gab es für die Feststellung eines nicht quantifizierbaren Zusatznutzens?

Abbildung 6 zeigt, wie häufig Evidenzlücken für den neuen Wirkstoff (Orphan Drug) bzw. die Kontrollgruppe festgestellt wurden, und welches Ausmaß diese jeweils hatten.

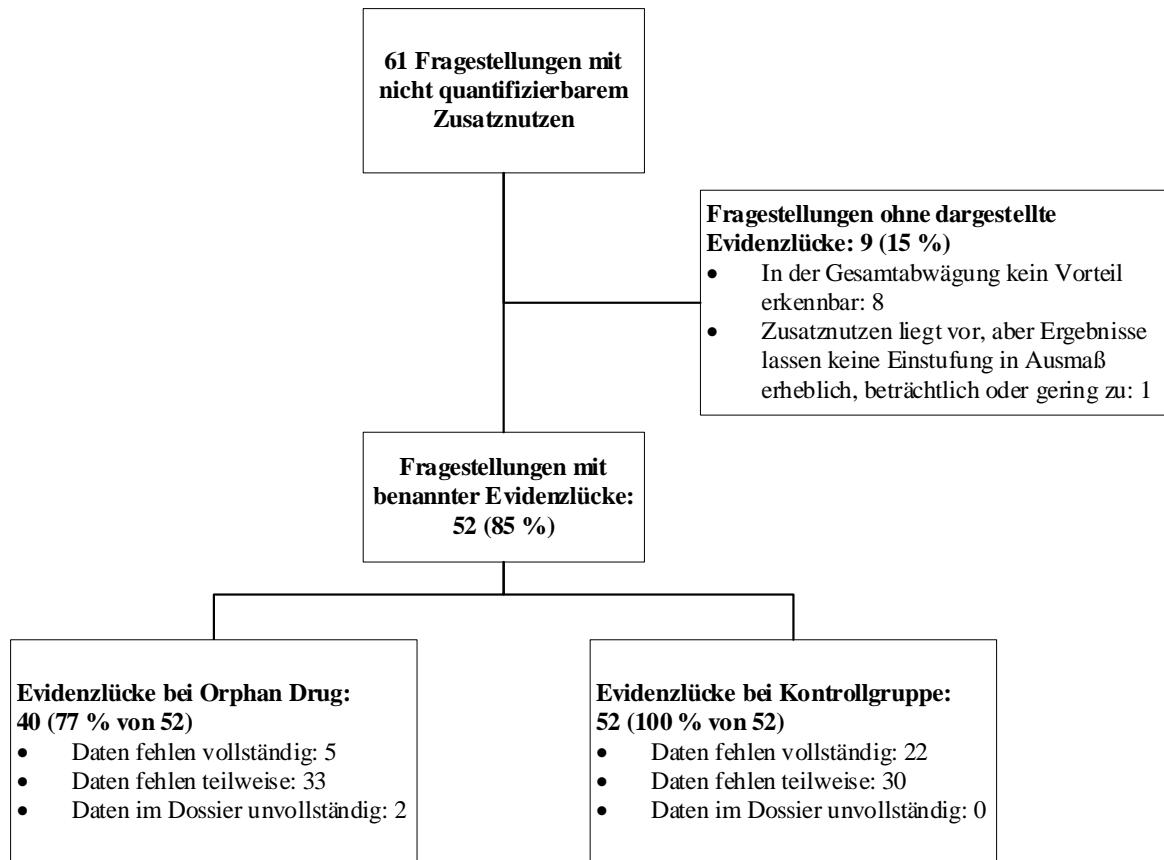


Abbildung 6: Evidenzlücken in Fragestellungen mit nicht quantifizierbarem Zusatznutzen in der Übersicht - Beschlüsse des G-BA zu Orphan Drugs bei Marktzugang 2014 bis 2018

Die Analyse zeigt, dass seitens des G-BA nur in wenigen Fällen (9 der 61 Fragestellungen, 15 %) der Zusatznutzen nicht quantifiziert wurde, ohne dass eine Evidenzlücke benannt wurde. Bei 8 dieser 9 Fragestellungen war in der Gesamtabwägung der positiven und negativen Effekte kein Vorteil für das Orphan Drug im Vergleich zur Kontrollgruppe erkennbar. Die Feststellung eines „nicht quantifizierbaren Zusatznutzens“ folgte in diesen Fällen offenbar der gesetzlichen Fiktion eines Zusatznutzens für Orphan Drugs bei Marktzugang (§ 35a Abs. 1 Satz 11 SGB V). In einem Fall waren die Daten nach Einschätzung des G-BA zwar ausreichend für die Feststellung eines Zusatznutzens, ließen aber keine Einstufung in eine der Kategorien gering, beträchtlich oder erheblich zu, ohne dass hierfür Evidenzlücken konkret benannt wurden.

Bei der weitaus überwiegenden Zahl der Fragestellungen ohne Quantifizierung des Zusatznutzens lagen Evidenzlücken vor (52 von 61 Fragestellungen, 85 %). Das Ausmaß der Evidenzlücken war dabei für die Kontrollgruppen größer als für die Orphan Drugs selbst, was im Wesentlichen durch das vollständige Fehlen von Daten für die Kontrollgruppe bedingt ist. Auch für Orphan Drugs wurden in der Mehrzahl der Fälle Evidenzlücken festgestellt. Zusammenfassend zeigt sich folgendes Bild:

- Bei allen 52 Fragestellungen mit Evidenzlücken (100 %) fehlten Daten für die Kontrollgruppe, für das jeweilige Orphan Drug war dies bei 40 Fragestellungen der Fall (77%).
- Auch das Ausmaß der festgestellten Evidenzlücken war für die Kontrollgruppe höher als für das Orphan Drug selbst: in 22 der 52 Fälle fehlten Daten zur Kontrollgruppe vollständig (42 %), während dies bei Orphan Drugs nur in 5 der 40 Fragestellungen mit Evidenzlücken der Fall war (13 %).

Die nachfolgende Tabelle 11 zeigt, für welche Endpunktakategorien gemäß der Darstellung des G-BA Evidenzlücken vorlagen, und aus welchen Gründen dies jeweils der Fall war.

Tabelle 11: Darstellung der Evidenzlücken je Endpunktakategorie – Orphan Drugs mit Marktzugang 2014 bis 2018, Fragestellungen mit nicht quantifizierbarem Zusatznutzen

Gruppe Begründung für Evidenzlücke	Mortalität n (% von N)	Morbidität n (% von N)	Lebensqualität n (% von N)	Nebenwirkungen n (% von N)
Orphan Drugs (N = 40)				
<i>Daten lagen nicht vor</i>	20 (50 %)	27 (68 %)	26 (65 %)	18 (45 %)
<i>Mängel in der Erhebung^a</i>	7 (18 %)	8 (20 %)	16 (40 %)	7 (18 %)
<i>Mängel bei der Auswertung</i>	11 (28 %)	12 (30 %)	5 (13 %)	5 (13 %)
	2 (5 %)	7 (18 %)	5 (13 %)	6 (15 %)
Kontrollgruppen (N = 52)				
<i>Daten lagen nicht vor</i>	33 (63 %)	40 (77 %)	35 (67 %)	34 (65 %)
<i>Mängel in der Erhebung^a</i>	22 (42 %)	27 (52 %)	30 (58 %)	27 (52 %)
<i>Mängel bei der Auswertung</i>	11 (21 %)	8 (15 %)	2 (4 %)	4 (8 %)
	0 (0 %)	5 (10 %)	3 (6 %)	3 (6 %)

a: Einschließlich zu kurzer Beobachtungsdauer oder zu geringer Fallzahl.

Die nachfolgende Tabelle 12 zeigt, bei wie vielen Fragestellungen Evidenzlücken in 1, 2, 3 oder allen 4 Endpunktakategorien auftraten.

Tabelle 12: Anzahl der Endpunkt-kategorien mit Evidenzlücke – Orphan Drugs mit Marktzugang 2014 bis 2018, Fragestellungen mit nicht quantifizierbarem Zusatznutzen

Anzahl der Endpunkt-kategorien mit Evidenzlücken	Orphan Drugs n (% von 40)	Kontrollgruppen n (% von 52)
1 Endpunkt-kategorie	12 (30 %)	12 (23 %)
2 Endpunkt-kategorien	12 (30 %)	12 (23 %)
3 Endpunkt-kategorien	9 (23 %)	6 (12 %)
4 Endpunkt-kategorien	7 (18 %)	22 (42 %)

Aus der detaillierten Darstellung der Evidenzlücken ergeben sich insbesondere folgende Schlussfolgerungen:

- Evidenzlücken lagen in den 4 Endpunkt-kategorien Mortalität, Morbidität, gesundheitsbezogene Lebensqualität und Nebenwirkungen häufig und nahezu gleichmäßig verteilt vor. Dies gilt für die Orphan Drugs selbst wie für die jeweiligen Kontrollgruppen gleichermaßen.
- Für die Kontrollgruppen lag über alle Endpunkt-kategorien hinweg das Ausmaß der Evidenzlücken höher als bei den Orphan Drugs.
- Während für die Kontrollgruppen in der deutlichen Mehrzahl die Evidenzlücke auf das vollständige Fehlen von Daten zurückzuführen war, waren bei den Orphan Drugs zusätzlich Mängel in der Erhebung in den im Dossier vorgelegten Studien ein häufiger Grund.
- Für Orphan Drugs und insbesondere für die Kontrollgruppen waren die Evidenzlücken nur in wenigen Fällen auf Mängel bei der Auswertung der erhobenen und im Dossier vorgelegten Daten zurückzuführen.
- Bei den meisten Fragestellungen lagen Evidenzlücken in 2 oder mehr der 4 Endpunkt-kategorien vor.

Fazit der zusammenfassenden Analyse der Orphan-Drug-Bewertungen und der bestehenden Evidenzlücken

Die zusammenfassende Analyse zeigt einerseits, dass für Orphan-Drug-Bewertungen der Jahre 2014 bis 2018 zum Zeitpunkt des Marktzugangs für einen Großteil der Fragestellungen relevante Daten im zugehörigen Dossier zur Nutzenbewertung vorgelegt wurden (80 von 85 Fragestellungen, 94 %). In ca. 2 Dritteln der Fälle handelte es sich dabei um Studien mit Randomisierung, bei einem Drittel der Fälle um Studien ohne Randomisierung. Andererseits zeigt die Analyse auch, dass trotz dieser hohen Rate an Fragestellungen mit relevanten Daten der Zusatznutzen nur für ca. ein Viertel der Fragestellungen quantifiziert werden konnte. Aus den vorliegenden Informationen ließ sich dabei nicht ableiten, dass der Zusatznutzen insbesondere bei kleinen Zielpopulationen nicht quantifizierbar wäre.

Die Analyse der Fragestellungen mit nicht quantifizierbarem Zusatznutzen zeigt, dass bei insgesamt 52 der 85 in den Jahren 2014 bis 2018 vom G-BA bewerteten Fragestellungen Evidenzlücken benannt wurden, die für die fehlende Quantifizierbarkeit ausschlaggebend waren (61 %). Nahezu 2 Drittel der Orphan-Drug-Bewertungen stellen demnach potenziell Kandidaten für eine anwendungsbegleitende Datenerhebung dar, wenn diese der Quantifizierung des Zusatznutzens dienen soll. Da die Evidenzlücken bei ca. 80 % der Fragestellungen für das jeweilige Orphan Drug selbst vorliegen und zu 100 % bei der Kontrollgruppe, auch weil häufig Daten zur Kontrollgruppe vollständig fehlten, muss eine zielgerichtete anwendungsbegleitende Datenerhebung dabei regelhaft vergleichend unter Hinzuziehung einer aus Sicht des G-BA relevanten Kontrollgruppe (Vergleichstherapie) geplant und durchgeführt werden.

Evidenzlücken liegen oft in mehreren, wenn auch nicht regelhaft in allen Endpunktkatoren (Mortalität, Morbidität, gesundheitsbezogene Lebensqualität und Nebenwirkungen) vor. Da auch bei den Endpunktkatoren Morbidität und gesundheitsbezogene Lebensqualität häufig Evidenzlücken bestehen, werden für zielgerichtete anwendungsbegleitende Datenerhebungen oft Informationen zu patientenberichteten Endpunkten (PROs) erforderlich sein, da diese zur Adressierung der Evidenzlücken in diesen Endpunktkatoren essenziell sind.

Es ist daher insgesamt absehbar, dass als Zulassungsaufgabe geplante Datenerhebungen, die insbesondere der Identifikation seltener oder spät auftretender Nebenwirkungen des jeweiligen Orphan Drugs dienen sollen, in unveränderter Form häufig keine geeignete, zielgerichtete anwendungsbegleitende Datenerhebung zum Zwecke der Nutzenbewertung darstellen werden. Denn Evidenzlücken liegen zumeist in mehreren und damit auch in anderen Endpunktkatoren vor, z. B. der Mortalität und der gesundheitsbezogenen Lebensqualität, und sie bestehen regelhaft auch für die Kontrollgruppe. Welche Änderung bzw. Erweiterung einer regulatorisch geforderten Datenerhebung für eine zielgerichtete anwendungsbegleitende Datenerhebung zum Zwecke der Nutzenbewertung notwendig ist (insbesondere Endpunkte [z. B. PROs], Kontrollgruppe, Beobachtungsdauer und -intervalle), ist jeweils im Einzelfall auf Basis der für die Quantifizierung des Zusatznutzens bestehenden Evidenzlücke zu prüfen.

5.7.3 Vorschläge für den Ablauf der anwendungsbegleitenden Datenerhebung im Verfahren der Nutzenbewertung nach § 35a SGB V

Ausgehend von der Analyse der Orphan Drug Bewertungen der Jahre 2014 bis 2018 und auf Basis der Vorgaben des SGB V zeigt Tabelle 13 mögliche Prozessschritte der anwendungsbegleitenden Datenerhebung im Verfahren der Nutzenbewertung nach § 35a SGB V.

Tabelle 13: Prozessschritte für eine anwendungsbegleitende Datenerhebung zum Zwecke der Nutzenbewertung nach § 35a

Prozessschritt	Anmerkung
Identifizierung einer Evidenzlücke im Beschluss des G-BA zur Nutzenbewertung nach § 35a SGB V	<ul style="list-style-type: none"> Evidenzlücke: relevante Datenlücke zum Vergleich des neuen Arzneimittels mit der (zweckmäßigen) Vergleichstherapie bezüglich patientenrelevanter Endpunkte (insbesondere, wenn aufgrund der Evidenzlücke eine Quantifizierung des Zusatznutzens nicht möglich ist)
Beschreibung der Vorgaben für die anwendungsbegleitende Datenerhebung durch den G-BA und Übermittlung an den pharmazeutischen Unternehmer	<ul style="list-style-type: none"> Definition der Fragestellung (siehe Abschnitte 5.3.1 und 5.7.1) Dauer, Art und Umfang der Datenerhebung (Dauer der Datenerhebung pro Patient, Fallzahl auf Basis einer Fallzahlschätzung) Art und Umfang der Auswertung (abhängig vom verwendeten Studientyp, siehe Abschnitte 5.3.4 und 5.5.2) Festlegung der Zeitpunkte für die Überprüfung der gewonnenen Daten (mindestens alle achtzehn Monate) Festlegung der Vorgaben unter Berücksichtigung laufender und geplanter Datenerhebungen, insbesondere solcher, die sich aus Auflagen der Zulassungsbehörden (z. B. der EMA) ergeben
Überprüfung der erhobenen Daten und der Verpflichtung zur Datenerhebung	<ul style="list-style-type: none"> zum 1. Überprüfungszeitpunkt prüft der G-BA, ob ein (öffentlich verfügbares) Studienprotokoll inkl. Analyseplan vorliegt, das die anwendungsbegleitende Datenerhebung gemäß Anforderung abbildet zum 1. und jedem weiteren Überprüfungszeitpunkt prüft der G-BA die verfügbaren Daten und entscheidet, ob die Datenerhebung beendet werden kann oder fortgeführt werden soll

EMA: europäische Arzneimittel-Agentur; G-BA: Gemeinsamer Bundesausschuss; SGB: Sozialgesetzbuch

Mit dem Beschluss zur Nutzenbewertung nach § 35a SGB V stellt der G-BA fest, ob und wenn ja, welche Evidenzlücke für die Bewertung eines Arzneimittels besteht. Diese Feststellung bildet die Grundlage für die Beschreibung der Vorgaben für die anwendungsbegleitende Datenerhebung. Die Vorgaben für den pharmazeutischen Unternehmer umfassen die in Tabelle 13 genannten Komponenten. Bei der Festlegung der Vorgaben berücksichtigt der G-BA laufende und geplante Datenerhebungen, insbesondere solche, die sich aus Auflagen der Zulassungsbehörden (z. B. der EMA) ergeben. In diesem Schritt wird ebenfalls festgelegt, zu welchem Zeitpunkt die Überprüfung der erhobenen Daten erfolgen soll.

Es ist zu diskutieren, in welchem Umfang der G-BA oder das IQWiG in die Entwicklung von Studienprotokollen für die notwendige Datenerhebung eingebunden werden sollten. Ziel einer solchen Einbindung wäre es, sicherzustellen, dass die geplante Datenerhebung grundsätzlich zum Zwecke der Nutzenbewertung verwendbar ist.

Zum Zeitpunkt der 1. Überprüfung der anwendungsbegleitend erhobenen Daten prüft der G-BA, ob die durchzuführende Studie in einem Studienregister registriert ist und ob mindestens ein Studienprotokoll (inkl. Analyseplan), das die Datenerhebung gemäß den Vorgaben abbildet, öffentlich verfügbar vorliegt. Dieses Protokoll ist eine Voraussetzung dafür, dass die Datenerhebung durchgeführt werden wird.

Zu den festgelegten Überprüfungszeitpunkten werden außerdem die verfügbaren Daten bewertet. Dabei kann in einem 1. Schritt geprüft werden, ob sich die Daten grundsätzlich für die Nutzenbewertung eignen. In einem 2. Schritt kann dann die erneute Nutzenbewertung vorgenommen werden.

Abhängig von den vorliegenden Daten kann der G-BA beschließen, ob die Datenerhebung fortgeführt werden soll oder ob sie abgeschlossen werden kann, weil entweder die bestehende Evidenzlücke geschlossen wurde oder die Datenerhebung keine hinreichenden Belege zur Neubewertung des Zusatznutzens erbringen kann.

Potenzielle Anforderung einer anwendungsbegleitenden Datenerhebung im Vorfeld der Nutzenbewertung

In bestimmten Fällen kann es sinnvoll sein, mit dem Marktzugang und damit bereits zu Beginn der Nutzenbewertung eine anwendungsbegleitende Datenerhebung zu beginnen. Das wäre z. B. der Fall, wenn zu erwarten ist, dass ein großer Teil der Patientinnen und Patienten, für die ein neues Arzneimittel infrage kommt, direkt nach Marktzugang behandelt wird oder wenn ein Arzneimittel einmalig eingesetzt wird. Eine unmittelbar zu Marktzugang beginnende Datenerhebung könnte dazu beitragen, die Zeitraum, in dem nur unzureichende Evidenz zum Zusatznutzen des Arzneimittels verfügbar ist, relevant zu verkürzen.

Um ein anwendungsbegleitende Datenerhebung bereits im Vorfeld der Nutzenbewertung zu spezifizieren und dann ab dem Zeitpunkt des Markteintritts anzufordern, kann der G-BA auf Informationen aus dem Zulassungsverfahren oder auf eigene Recherchen zurückgreifen.

6 Diskussion

6.1 Studiendesigns für die Nutzung versorgungsnaher Daten

Eine sinnvolle Diskussion von geeigneten Studiendesigns für die Nutzung versorgungsnaher Daten muss zwingend die Fragestellung, die durch die Datenerhebung beantwortet werden soll, berücksichtigen. Bereits die Entscheidung, ob eine Studie mit einem einzelnen Studienarm oder als vergleichende Studie mit mindestens 2 Armen durchgeführt werden muss, ist durch die Fragestellung bestimmt. So können z. B. Daten zur Beschreibung des Versorgungsgeschehens innerhalb einer Erkrankung oder zur Charakterisierung der Patientenpopulation, die bestimmte Interventionen erhält, in einarmigen Studien erhoben werden. Sollen dagegen Effekte einer Intervention zum Zwecke einer Nutzenbewertung untersucht werden, sind vergleichende Studien notwendig (siehe Abschnitt 5.2). Im Folgenden werden deshalb ausschließlich vergleichende Studien diskutiert.

Auch unter Berücksichtigung der Argumente der aktuellen intensiven Diskussion um die Nutzung von Daten aus der Versorgung zur Unterstützung von Entscheidungen im Gesundheitswesen, muss festgestellt werden, dass Aussagen zum Nutzen und Schaden von medizinischen Interventionen auf Basis vergleichender Studien ohne Randomisierung regelhaft eine größere Unsicherheit aufweisen als Studien mit Randomisierung. Ausgehend vom Konzept der Emulation eines Target Trials (siehe Abschnitt 5.3.2) sind vergleichende Studien ohne Randomisierung darüber hinaus bei gleichen Ansprüchen an die Datenqualität regelhaft mit einem deutlich höheren Aufwand verbunden als solche mit Randomisierung. Wie relevant der Unterschied in der Unsicherheit ist, hängt vom Entscheidungskontext ab. Die größere Unsicherheit ist ggf. von untergeordneter Bedeutung, wenn Interventionen ein geringes Risiko für einen Schaden für die einzelnen Patientinnen und Patienten oder für das Gesundheitssystem haben. Unsicherheit kann auch akzeptabel sein, wenn die beobachteten Effekte einer Intervention so groß sind, dass sie durch den Einfluss von Störgrößen nicht mehr plausibel erklärbar sind. In diesen Fällen kann davon ausgegangen werden kann, dass ein Effekt erhalten bleibt [143], auch wenn er dann nicht quantifiziert werden kann.

Die erhöhte Unsicherheit von Ergebnissen vergleichender Studien ohne Randomisierung ist Anlass für eine intensive Methodendiskussion [22,23,140,144]. Das grundsätzliche Problem in vergleichenden Studien ohne Randomisierung ist die potenziell fehlende Strukturgleichheit der Behandlungsgruppen, die dazu führt, dass der kausale Zusammenhang zwischen Intervention und Effekt infrage stellt werden muss. Die methodischen Ansätze zur Nutzung von Daten aus diesen Studien haben deshalb primär eine Annäherung an die Strukturgleichheit der Behandlungsgruppen zum Ziel, z. B. durch eine Emulation einer randomisierten Zielstudie und durch Adjustierung für Confounder (siehe Abschnitte 5.3.2 und 5.5.2). Diese methodischen Ansätze erfordern eine detaillierte Planung der Studie, um die notwendige Datenerhebung sicherzustellen, und sind nur nach entsprechend umfangreicher und spezifischer Datenerhebung und bei ausreichend großen Stichproben möglich. Die Durchführung einer qualitativ hochwertigen vergleichenden Studie ohne Randomisierung ist damit mit einem hohen Aufwand verbunden. Abschließend können die Methoden der Confounderkontrolle eine Verzerrung

jedoch nicht grundsätzlich ausschließen. Insbesondere bleibt für die einzelne Studie unklar, in welchem Ausmaß eine Adjustierung für Confounder gelingt, da auch bei Berücksichtigung der bekannten relevanten Confounder zusätzlich unbekannte relevante Confounder vorliegen können.

Eine Randomisierung stellt die Strukturgleichheit dagegen mit geringem Aufwand sicher und ermöglicht so abhängig von den geplanten Endpunkten Studien mit geringerer Fallzahl und weniger umfangreicher Datenerhebung. Dieser Vorteil ist insbesondere für Studien in kleinen Patientenpopulationen relevant.

Optimierte Studien für die Entscheidungsfindung im Gesundheitswesen

Trotz der im vorangegangenen Abschnitt beschriebenen Zusammenhänge zwischen Randomisierung und Effizienz des Studiendesigns besteht die Wahrnehmung, dass vergleichende Studien ohne Randomisierung leichter, schneller und ggf. mit relevanteren Ergebnissen durchgeführt werden könnten als vergleichende Studien mit Randomisierung. Die gängigen Argumente beschreiben vergleichende Studien mit Randomisierung als Studien, die eng umrissene Patientenpopulationen einschließen, die die im Versorgungsalltag behandelten Patientinnen und Patienten nicht abbilden. Darüber hinaus seien diese Studien nicht durchführbar in kleinen Patientenpopulationen (z. B. bei seltenen Erkrankungen) oder zur Untersuchung von seltenen oder späten Ereignissen. Nicht zuletzt sei die Machbarkeit durch den hohen Aufwand bzw. die hohen Kosten eingeschränkt.

Patientenpopulationen mit Relevanz für den Versorgungsalltag

Es ist richtig, dass aktuell viele vergleichende Studien mit Randomisierung über ihre Ein- und Ausschlusskriterien eng definierte Patientenpopulationen einschließen. Eine Randomisierung ist aber nicht zwingend mit einer derartigen Einschränkung der Patientenpopulation verbunden. Eine vergleichende Studie mit Randomisierung kann auch eine Population einschließen, die die Patienten umfasst, die im Versorgungsalltag mit den zu untersuchenden Interventionen behandelt werden. Diese sogenannten pragmatischen randomisierten Studien werden bereits durchgeführt [145,146], es wurden auch bereits Empfehlungen für das Design solcher Studien entwickelt [27,29].

Studien bei seltenen und späten Ereignissen

Die Untersuchung seltener Ereignisse verlangt hohe Fallzahlen, die Beobachtung spät auftretender Ereignisse lange Beobachtungszeiten. Gerade für diese Fallkonstellationen werden häufig vergleichende (nicht interventionelle) Studien ohne Randomisierung anstelle von vergleichenden Studien mit Randomisierung vorgeschlagen.

Als Lösung für diese Situationen werden auf der anderen Seite sogenannte *large simple trials* diskutiert, das sind vergleichende Studien mit Randomisierung, die einen begrenzten, spezifisch auf die bestehende Fragestellung zugeschnittenen Datensatz erheben und häufig im Versorgungsalltag durchgeführt werden [29]. Damit bestehen Überschneidungen zu den oben beschriebenen pragmatischen randomisierten Studien. Für diese Studien wird auch über eine

Datenerhebung in Registern aber auch in elektronischen Patientenakten oder mithilfe von Abrechnungsdaten von Krankenkassen diskutiert.

Effizientes Design für aussagekräftige Ergebnisse in kleinen Populationen

Insbesondere bei Interventionen, die in kleinen Patientenpopulationen wie z. B. bei seltenen Erkrankungen eingesetzt werden sollen, wird die Machbarkeit von vergleichenden Studien mit Randomisierung infrage gestellt. Empirische Untersuchungen von Studien bei seltenen Erkrankungen entkräften dieses Argument. So hat z. B. eine systematische Untersuchung von Studien, die in dem Studienregister ClinicalTrials.gov registriert sind, ergeben, dass auch bei seltenen und sehr seltenen Erkrankungen in großem Umfang vergleichende Studien mit Randomisierung durchgeführt werden [147]. Zum gleichen Ergebnis kommt ein Bericht des IQWiG, der die zugrunde liegenden Studien für Orphan Drug Zulassungen der Jahre 2001 bis 2013 durch die EMA untersucht [148]. Die Überprüfung der zugrundliegenden Studien in Beschlüssen des G-BA zur frühen Nutzenbewertung von Orphan Drugs im vorliegenden Bericht ergab ebenfalls einen hohen Anteil von vergleichenden Studien mit Randomisierung (siehe Abschnitt 5.7.2). Diese Ergebnisse zeigen, dass die Machbarkeit von vergleichenden Studien mit Randomisierung in kleinen Populationen nicht infrage steht.

Gerade in kleinen Populationen erscheint die Durchführung vergleichender Studien ohne Randomisierung fragwürdig, weil die verfügbaren Fallzahlen in der Regel keine adäquate Adjustierung für die relevanten Confounder ermöglichen. In kleinen Populationen kann die notwendige Methodik für diese Studien ohne Randomisierung also ggf. gar nicht angewendet werden. Eine Entscheidungsfindung auf Basis solcher Studien ist für die betroffenen Patientinnen und Patienten deshalb mit einer hohen Unsicherheit verbunden.

In der Situation sehr begrenzter Patientenzahlen erscheint es deshalb sinnvoller, auf eine kontrollierte Erhöhung der Unsicherheit zurückzugreifen. So kann z. B. das statistische Irrtumsniveau über den üblichen Wert von (2-seitig) 5 % auf 10 % angehoben werden [148,149]. In absteigender Priorität könnten auch Einschränkungen der externen Validität hingenommen werden, z. B. durch den Einbezug von Daten aus ähnlichen Indikationsgebieten oder durch den Einsatz von etablierten Surrogatendpunkten innerhalb kombinierter Endpunkte [148].

Diese Kompromisse bei der geforderten Präzision der Ergebnisse oder der Definition der Endpunkte führen eher zu interpretierbaren Ergebnissen als die Durchführung einer nicht ausreichend für Confounder adjustierten vergleichenden Studie ohne Randomisierung.

Minimierung von Aufwand und Kosten

Als weiterer Grund für Schwierigkeiten vergleichende Studien mit Randomisierung durchzuführen, werden die Kosten, der Aufwand und die Dauer von vergleichenden Studien mit Randomisierung genannt. Dabei ist zu berücksichtigen, dass Kosten und Aufwand nicht primär durch die Randomisierung, sondern durch die für diese Studien geforderten Maßnahmen

zur Sicherstellung der Datenqualität zustande kommen. Die Sicherstellung der Datenqualität ist für Studien ohne Randomisierung jedoch gleichermaßen erforderlich.

Auch hier erscheint die Lösung eher in der Veränderung der Durchführung dieser Studien als im Verzicht auf dieses aussagekräftige Studiendesign zu liegen. So sollten neue Instrumente, wie z. B. registerbasierte RCTs eingesetzt werden, um Studien innerhalb bestehender Datenstrukturen durchzuführen und damit die Rekrutierung zu verbessern, Kosten zu senken und eine langfristige Datenerhebung zu ermöglichen [13,42,44]. Insbesondere in Kombination mit den oben beschriebenen Ansätzen der pragmatischen RCT und der *large simple trials* mit einer Beschränkung auf die Erhebung der wesentlichen Endpunkte können registerbasierte RCTs eine Möglichkeit sein, den Aufwand für vergleichende Studien mit Randomisierung zu verringern.

Zusammenfassung

Zusammenfassend kann es einfacher und zielführender sein, eine vergleichende Studie mit Randomisierung unter Berücksichtigung der oben beschriebenen Anpassungen durchzuführen, als zu versuchen, qualitativ hochwertige Ergebnisse aus einer vergleichenden Studie ohne Randomisierung zu generieren.

6.2 Allgemeine Handlungsempfehlungen auf Basis der Ergebnisse der Registergespräche

Die Analyse der möglichen Konzepte zur Generierung versorgungsnaher Daten zum Zwecke der Nutzenbewertung hat gezeigt, dass für den deutschen Versorgungskontext auf absehbare Zeit Register voraussichtlich das wichtigste Instrument für eine versorgungsnahe Datenerhebung darstellen werden. Dies gilt unabhängig davon, ob es sich bei den registerbasierten Studien um vergleichende Studien ohne oder mit Randomisierung handelt.

Ob und inwieweit die verschiedenen Patientenregister bereits zum jetzigen Zeitpunkt dafür geeignet sind, die zukünftigen Fragestellungen bei der Nutzenbewertung von Arzneimitteln zu beantworten, lässt sich nicht pauschal beantworten, da dies vom jeweiligen Register und der Art der Registerstudie, aber insbesondere auch von den konkreten Fragestellungen abhängt. Es lassen sich jedoch auf Basis der Analysen und der Gespräche mit den Registerbetreibern verschiedene Handlungsfelder beschreiben, die dazu dienen, die einzelnen Register im Speziellen und die Registerlandschaft in Deutschland im Allgemeinen bei der Aufgabe der versorgungsnahen Datenerhebung zum Zwecke der Nutzenbewertung von Arzneimitteln zu unterstützen. Dabei lassen sich 3 (teilweise überlappende) Zielgruppen unterscheiden: Registerbetreiber, Auftraggeber bzw. Verantwortliche für Registerstudien sowie Entscheidungsträger im Gesundheitswesen und in der Politik. Nachfolgend werden für diese 3 Zielgruppen Handlungsempfehlungen aufgeführt mit dem Ziel, die versorgungsnahe Datenerhebung zum Zwecke der Nutzenbewertung von Arzneimitteln zu unterstützen.

Registerbetreiber

Für die einzelnen Register könnte insbesondere durch folgende 3 Schritte geprüft werden, inwieweit sie eine geeignete Datenquelle mit Blick auf die versorgungsnahe Datenerhebung zum Zwecke der Nutzenbewertung von Arzneimitteln darstellen:

- Prüfung des Registers im Hinblick auf die in Abschnitt 5.5.1.3 genannten obligatorischen Kriterien zur Sicherstellung der Datenqualität sowie der allgemeinen Kriterien, die regelhaft für Registerstudien zum Zwecke der Nutzenbewertung von Arzneimitteln relevant sind.
- Emulation eines oder mehrerer Target Trials (siehe Abschnitt 5.3.2) im Register auf Basis tatsächlich vorhandener Evidenzlücken in abgeschlossenen Nutzenbewertungen zu Arzneimitteln. Hierfür können die Beschlüsse und die Tragenden Gründe des G-BA sowie die zugehörigen Nutzenbewertungen des IQWiG bzw. des G-BA (bei Orphan Drugs) herangezogen werden. Bei Registern zu seltenen Erkrankungen bietet es sich an, Target Trials zu Fragestellungen bei Orphan Drugs mit nicht quantifiziertem Zusatznutzen zu emulieren (siehe Anhang D). Für die übrigen Erkrankungen bieten sich primär befristete Beschlüsse oder Beschlüsse zu Fragestellungen, für die der Zusatznutzen nicht belegt wurde, an.
- Identifikation der relevanten Confounder für diese Target Trials bzw. die damit untersuchten Fragestellungen. Dies sollte möglichst systematisch, z. B. auf Basis wissenschaftlicher Literatur in Abstimmung mit Fachexperten, erfolgen.
- Prüfung, ob in dem Register prinzipiell Studien mit Randomisierung durchgeführt werden können bzw. welche (organisatorischen) Hürden diesbezüglich bestehen und wie diese beseitigt werden können.

Mit diesen Schritten kann Register-individuell die Eignung für die versorgungsnahe Datenerhebung zum Zwecke der Nutzenbewertung von Arzneimitteln abgeschätzt und etwaiger Anpassungsbedarf beschrieben werden. Zugleich kann hierdurch das konkrete Vorgehen der Emulation eines Target Trials und der systematischen Zusammenstellung der relevanten Confounder pilotiert werden, sofern Erfahrungen hierzu nicht bereits anderweitig vorliegen. Zudem können die Register auf die Durchführung registerbasierter Studien mit Randomisierung vorbereitet werden.

Das vorgeschlagene Vorgehen bedeutet nicht, dass damit eine (möglicherweise sehr kostenträchtige und an organisatorische Grenzen stoßende) Erweiterung jedes Registers mit dem Ziel, jede absehbare Fragestellung zukünftig beantworten zu können, angestoßen werden soll. Eine solche Bestandsaufnahme kann vielmehr der gezielten Anpassung in einzelnen Bereichen dienen, oder auch der bewussten Entscheidung gegen eine solche Anpassung, falls diese anderen Registerzielen widersprechen würde.

Es bietet sich an, das konkrete Vorgehen und die entsprechenden Ergebnisse zu publizieren, zum einen, um G-BA, IQWiG und potenzielle Auftraggeber für Registerstudien über die

Einschätzung zur Eignung des jeweiligen Registers zu informieren, zum anderen zur gegenseitigen Unterstützung der Registerbetreiber bei der Eignungsprüfung.

Auftraggeber bzw. Verantwortliche für Registerstudien

Für die Durchführung von Registerstudien zum Zwecke der Nutzenbewertung von Arzneimitteln im Kontext des § 35a SGB V sollten Standardprotokolle und -analysepläne entwickelt werden. Es ist zwar davon auszugehen, dass hierfür in großen Teilen auf bereits vorhandene Standarddokumente, z. B. aus dem regulatorischen Umfeld, zurückgegriffen werden kann. Insbesondere die nachfolgenden Punkte sollten allerdings spezifisch implementiert werden:

- Entwicklung einer generischen Methodik zur systematischen Identifikation aller für die jeweilige Fragestellung relevanten Confounder; Beschreibung dieser Methodik im Standard-Studienprotokoll
- Target Trial – Emulation mit Beispielen für typische PICO-Fragestellungen zur Nutzenbewertung von Arzneimitteln
- Generische Beschreibung, wie die fragestellungsbezogene Datenqualität des Registers darzulegen ist (inkl. Angaben zur Vollständigkeit und Richtigkeit für die einzelnen relevanten Datenfelder)
- Generische Beschreibung der Auswertung gemäß Abschnitt 5.5.2 einschließlich prospektiver Festlegung der Entscheidungsstruktur für von den konkreten Daten abhängige Vorgehensweisen zur Confounderadjustierung

Diese Standarddokumente sollten unter Einbindung von Registerbetreibern entwickelt werden, da bei konkreter Ausgestaltung für eine spezifische Fragestellung einige Elemente detailliertes Wissen über die jeweilige Datenquelle (das jeweilige Register) erfordern. Darüber hinaus sollte eine Berichtsstruktur für einen Standardbericht von Registerstudien zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V entwickelt werden.

Entscheidungsträger im Gesundheitswesen und in der Politik

Die versorgungsnahe Datenerhebung zum Zwecke der Nutzenbewertung von Arzneimitteln ist politischer Wille in Deutschland, was sich in der Gesetzgebung zur anwendungsbegleitenden Datenerhebung niedergeschlagen hat [1]. Die Gespräche mit den Registerbetreibern haben verschiedene hinderliche Faktoren zum Aufbau und Betrieb eines Registers und damit zur Erreichung des oben genannten Ziels, der Verwendung versorgungsnaher Datenerhebungen zum Zwecke der Nutzenbewertung von Arzneimitteln, identifiziert. Einige dieser Faktoren sind auf allgemeine Rahmenbedingungen zurückzuführen, die nur Register-übergreifend von Entscheidungsträgern im Gesundheitswesen und in der Politik adressiert werden können. Zu den 3 Themen Datenschutz, IT-Landschaft und Finanzierung werden nachfolgend die wesentlichen Hindernisse und daraus abgeleitet Handlungsoptionen beschrieben.

Datenschutz

Heterogene Datenschutzanforderungen auf verschiedenen Ebenen, z. B. zwischen Bundesländern, zwischen verschiedenen Universitäten, zwischen verschiedenen Kliniken auch innerhalb eines Bundeslandes, behindern den Aufbau und die potenziell für die versorgungsnahe Datenerhebung zum Zwecke der Nutzenbewertung von Arzneimitteln notwendige Erweiterung eines Registers. Es ist nicht nachvollziehbar, warum für Patientinnen und Patienten in Klinik A in einem nördlichen Bundesland andere Datenschutzbestimmungen gelten sollten als in Klinik B im Süden oder Universität C im Osten.

Es empfiehlt sich, für Zwecke der Patientenregister einheitliche und für alle Beteiligten (insbesondere Registerbetreiber, Zentren, Patientinnen und Patienten, Datenschützer) verbindliche Rahmenbedingungen zu schaffen.

IT-Landschaft

Die heterogene IT-Landschaft in den Zentren und dabei insbesondere die unterschiedlichen Krankenhausinformationssysteme (KIS), verbunden mit teilweise veralteter Hardware, stellen nach Angabe der Registerbetreiber ein wesentliches Hindernis beim Registerbetrieb dar.

Als Handlungsfelder ergeben sich hieraus zum einen ggf. (Sonder-)Investitionen in die technische Infrastruktur auf Basis zu definierender Mindeststandards. Zum anderen könnte erwogen werden, für KIS Anforderungen analog derjenigen für die Praxissoftware im ambulanten Bereich (§73 Abs. 9 SGB V) gesetzlich zu verankern mit dem Ziel, einheitliche Schnittstellen und einheitliche Kerndatensätze unter Berücksichtigung semantischer Interoperabilität zu schaffen.

Finanzierung des Registerbetriebs

Für Patientenregister gibt es, mit gesetzlich begründeten Ausnahmen (z. B. Krebsregister), keine nachhaltige öffentliche Finanzierung, was aus Sicht der Registerbetreiber ein Hindernis für den Aufbau und Betrieb eines Registers, insbesondere aber für die Gewährleistung einer für die Nutzenbewertung von Arzneimitteln hohen Datenqualität darstellen kann.

Zwar kann argumentiert werden, dass Registerstudien zu Arzneimitteln, so wie klinische Studien auch, durch die Auftraggeber der Studien finanziert werden sollten. Im Gegensatz zu klinischen Studien wird bei Registern jedoch erwartet, dass dauerhaft eine technische und organisatorische Infrastruktur aufrechterhalten wird, auf die kurzfristig für die Planung und Durchführung von Registerstudien zurückgegriffen werden kann. Unabhängig von der Frage, ob die Finanzierung von Registerstudien durch die pharmazeutische Industrie die Unabhängigkeit eines Registers gefährdet, kann sie daher jedenfalls nicht den Dauerbetrieb von Registern sichern.

Aus Sicht einiger Registerbetreiber ist es daher kaum machbar, das Register auf die zukünftige Durchführung von Registerstudien zum Zwecke der Nutzenbewertung von Arzneimitteln auszurichten, sofern dies mit hohem und nicht gegenfinanziertem Aufwand verbunden ist. Denn

dadurch könnten potenziell andere Registerziele gefährdet werden, weil für diese die notwendigen Mittel dann nicht mehr zur Verfügung stehen.

Angesichts des offensichtlich hohen öffentlichen Interesses an Registerstudien zum Zwecke der Nutzenbewertung von Arzneimitteln erscheint es daher sinnvoll, über eine öffentliche Finanzierung von Registern unter bestimmten Voraussetzungen nachzudenken. Dies könnte z. B. (Zusatz-)Investitionen zur Erfüllung der obligatorischen Kriterien zur Sicherstellung der Datenqualität umfassen. Inwieweit hierfür aufgrund der Zielrichtung des GSAV, der Nutzenbewertung nach § 35a SGB V, eine Finanzierung über Mittel der Selbstverwaltung mit Beteiligung der PKV infrage kommt, wäre zu diskutieren.

Finanzierung interventioneller Registerstudien

Die Durchführung registerbasierter interventioneller Studien ist nachvollziehbarer Weise nur innerhalb des vorgegebenen rechtlichen Rahmens gemäß AMG möglich. Eine Finanzierung solcher Studien durch die Registerbetreiber als Studiensponsor ist aufgrund der häufig hohen Arzneimittelkosten derzeit allerdings zumeist nicht realistisch, wodurch die Durchführung unabhängiger interventioneller Registerstudien erschwert wird. Es könnte daher sinnvoll sein, eine Kostenübernahme für versorgungsnahe interventionelle Studien im Rahmen des SGB V zu regeln. Denkbar wäre z. B. ein Modell einer Antragsregelung analog § 35c Abs. 2 SGB V zur Kostenübernahme für Studien im Off-Label-Bereich.

7 Fazit

Studiendesign und Datenerhebung

- der Einsatz versorgungsnaher Daten zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V verlangt einen Vergleich zwischen dem neuen Arzneimittel und der vom G-BA bestimmten zweckmäßigen Vergleichstherapie; damit ist die Durchführung vergleichender Studien notwendig
- die Erhebung versorgungsnaher Daten zum Zwecke der Nutzenbewertung nach § 35a SGB V aus elektronischen Patientenakten und aus Abrechnungsdaten von Krankenkassen wird derzeit und absehbar als nicht realistisch eingeschätzt; es ist vielmehr eine studienindividuelle Datenerhebung oder eine Datenerhebung in Patientenregistern notwendig

Versorgungsnahe vergleichende Studien ohne Randomisierung

- sollen versorgungsnahen vergleichende Studien ohne Randomisierung für die Nutzenbewertung herangezogen werden, muss bereits mit der Studienplanung sichergestellt werden, dass der Ablauf der Studie und die erhobenen Daten die notwendige Qualität haben, um interpretierbare Ergebnisse zu generieren
- wesentliche Komponenten einer solchen Studienplanung sind ein Studienprotokoll inkl. Analyseplan, die Emulation einer Zielstudie, die die relevante Fragestellung behandelt, und die Sicherstellung der Erhebung ausreichender Daten für eine Confounderkontrolle
- ein zentraler Aspekt der Analyse einer vergleichenden Studie ohne Randomisierung ist die adäquate Adjustierung für Confounder; diese Adjustierung muss soweit möglich präspezifiziert werden, die enthaltenen Annahmen (z. B. die Definition der relevanten Confounder) sind dabei fundiert zu begründen
- aus versorgungsnahen vergleichenden Studien ohne Randomisierung können keine Effekte abgeleitet werden, wenn die Datenqualität in den verwendeten Datenquellen und die Qualität der Auswertung und Berichterstattung nicht hoch ist
- auch unter hohen Qualitätsanforderungen (für Daten, Auswertung und Berichterstattung) kann im Regelfall aus vergleichenden versorgungsnahen Studien ohne Randomisierung nicht mehr als ein Anhaltspunkt für einen Effekt abgeleitet werden
- wegen der inhärenten Unsicherheit der Ergebnisse aus vergleichenden Studien ohne Randomisierung sollte aufgrund potenziell unbekannter Confounder aus den in der Studie beobachteten Effekten erst ab einer bestimmten Effektstärke eine Aussage zum Nutzen oder Schaden einer Intervention abgeleitet werden; eine Quantifizierung eines Zusatznutzens gemäß der gesetzlich vorgegebenen Ausmaßkategorien erfordert entsprechende nach der Größe abgestufte Effektstärken
- die Möglichkeit, retrospektive Studiendesigns zu berücksichtigen, hängt davon ab, ob die verfügbaren Datenquellen die notwendigen Daten in der erforderlichen Qualität bereits

enthalten; Vergleiche eines neuen Arzneimittels mit historischen Kontrollen erscheinen nur dann realistisch, wenn für das neue Arzneimittel und die historische Kontrolle dieselbe Datenquelle (z. B. ein indikationsspezifisches klinisches Register) genutzt wird

Versorgungsnahe vergleichende Studien mit Randomisierung

- versorgungsnahe vergleichende Studien können auch mit Randomisierung durchgeführt werden (pragmatische klinische Studien)
- der Aufwand für eine versorgungsnahe vergleichende Studie mit Randomisierung wird bei vergleichbarer Datenqualität in der Regel geringer sein als der für eine Studie ohne Randomisierung, weil die Datenerhebung und Adjustierung für die Confounderkontrolle entfallen kann
- versorgungsnahe vergleichende Studien mit Randomisierung erreichen eine höhere Aussagekraft als solche ohne Randomisierung, eine Quantifizierung des Zusatznutzens ist zuverlässiger möglich
- insbesondere nach der Zulassung können versorgungsnahe vergleichende Studien mit Randomisierung abhängig von der bestehenden Fragestellung mit einer eingeschränkten Datenerhebung ([large] simple trials) durchgeführt werden; eine Durchführung in Registern hat ein zusätzliches Potenzial, die Studien zu beschleunigen und weniger aufwendig zu gestalten (registerbasierte vergleichende Studien mit Randomisierung)

Anwendungsbegleitende Erhebung versorgungsnaher Daten

- ob die verschiedenen Patientenregister bereits heute für die anwendungsbegleitende Datenerhebung nach § 35a SGB V geeignet sind, lässt sich nicht pauschal beantworten, da dies vom jeweiligen Register und den konkreten Fragestellungen abhängt
- auf Basis der Analysen und der Gespräche mit den Registerbetreibern lassen sich Handlungsfelder beschreiben, die dazu dienen, die einzelnen Register im Speziellen und die Registerlandschaft in Deutschland im Allgemeinen bei der anwendungsbegleitenden Datenerhebung zu unterstützen; diese Handlungsfelder sind im Bericht beschrieben
- die Erkenntnisse des vorliegenden Berichts können für die anwendungsbegleitende Datenerhebung zur Schließung von Evidenzlücken nach einer Nutzenbewertung eingesetzt werden; ein Vorschlag zum Ablauf dieser Datenerhebung wird mit dem Bericht zur Verfügung gestellt

8 Literatur

1. Bundesministerium für Gesundheit. Gesetz für mehr Sicherheit in der Arzneimittelversorgung (GSAV). Bundesgesetzblatt 2019; Teil 1(30): 1202-1220.
2. Müller D, Augustin M, Banik N, Baumann W, Bestehorn K, Kieschke J et al. Memorandum Register für die Versorgungsforschung. Gesundheitswesen 2010; 72(11): 824-839.
3. Stausberg J, Maier B, Bestehorn K, Gothe H, Gröne O, Jacke C et al. Memorandum Register für die Versorgungsforschung: Update 2019 [angenommenes Manuskript]. Gesundheitswesen.
4. Gliklich R, Dreyer N, Leavy M. Registries for evaluating patient outcomes: a user's guide; 3rd edition; AHRQ pub. no. 13(14)-EHC111. Rockville: Agency for Healthcare Research and Quality; 2014. URL: https://effectivehealthcare.ahrq.gov/sites/default/files/related_files/registries-guide-3rd-edition-vol-2-140430.pdf
5. European Medicines Agency. Discussion paper: use of patient disease registries for regulatory purposes; methodological and operational considerations [online]. 05.11.2018 [Zugriff: 23.05.2019]. URL: https://www.ema.europa.eu/documents/other/discussion-paper-use-patient-disease-registries-regulatory-purposes-methodological-operational_en.docx.
6. European Network for Health Technology Assessment. REQuEST (Registry Evaluation and Quality Standards Tool) [online]. 30.09.2019 [Zugriff: 18.12.2019]. URL: <https://eunetha.eu/wp-content/uploads/2019/10/Registry-Evaluation-and-Quality-Standards-Tool-REQuEST-1.xlsm>.
7. Mandeville KL, Valentic M, Ivankovic D, Pristas I, Long J, Patrick HE. Quality assurance of registries for Health Technology Assessment. Int J Technol Assess Health Care 2018; 34(4): 360-367.
8. Clinical Trials Transformation Initiative. CTTI recommendations: registry trials [online]. [Zugriff: 18.12.2019]. URL: <https://www.ctti-clinicaltrials.org/files/recommendations/registrytrials-recs.pdf>.
9. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL et al. Real-world evidence: what is it and what can it tell us? N Engl J Med 2016; 375(23): 2293-2297.
10. Senn S. Statistical issues in drug development. Chichester: Wiley; 1997.
11. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. ICH guidelines [online]. [Zugriff: 18.12.2019]. URL: <https://www.ich.org/page/ich-guidelines>.
12. EQUATOR Network. Webauftritt [online]. [Zugriff: 28.11.2019]. URL: <https://www.equator-network.org>.

13. Food and Drug Administration. Framework for FDA's Real-World Evidence Program [online]. 12.2018 [Zugriff: 16.07.2019]. URL: <https://www.fda.gov/media/120060/download>.
14. Patient-Centered Outcomes Research Institute. PCORI methodology standards [online]. 26.02.2019 [Zugriff: 18.12.2019]. URL: <https://www.pcori.org/sites/default/files/PCORI-Methodology-Standards.pdf>.
15. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ et al (Ed). Cochrane handbook for systematic reviews of interventions, version 6.0. Hoboken: Wiley-Blackwell; 2019.
16. Bundesministerium für Gesundheit. Gesetz zur Neuordnung des Arzneimittelmarktes in der gesetzlichen Krankenversicherung (Arzneimittelmarktneuordnungsgesetz - AMNOG). Bundesgesetzblatt Teil 1 2010; (67): 2262-2277.
17. Bundesministerium für Gesundheit. Verordnung über die Nutzenbewertung von Arzneimitteln nach § 35a Absatz 1 SGB V für Erstattungsvereinbarungen nach § 130b SGB V (Arzneimittel-Nutzenbewertungsverordnung - AM-NutzenV) [online]. 09.08.2019 [Zugriff: 18.12.2019]. URL: <http://www.gesetze-im-internet.de/am-nutzenv/index.html>.
18. Schneeweiss S. Real-world evidence of treatment effects: the useful and the misleading. Clin Pharmacol Ther 2019; 106(1): 43-44.
19. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Sofosbuvir (chronische Hepatitis C bei Jugendlichen): Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag A17-55 [online]. 11.01.2018 [Zugriff: 22.01.2018]. (IQWiG-Berichte; Band 582). URL: https://www.iqwig.de/download/A17-55_Sofosbuvir_Nutzenbewertung-35a-SGB-V_V1-0.pdf.
20. Wellek S, Donner-Banzhoff N, König J, Mildenberger P, Blettner M. Studienplanung und -auswertung nach dem Stepped-Wedge-Design
Dtsch Arztebl 2019; 116(26): 453-458.
21. U.S. National Library of Medicine. ClinicalTrials.gov [online]. 2019 [Zugriff: 28.11.2019]. URL: <https://clinicaltrials.gov>.
22. Hernan MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol 2016; 183(8): 758-764.
23. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016; 355: i4919.
24. Peto R, Collins R, Gray R. Large-scale randomized evidence: large, simple trials and overviews of trials. J Clin Epidemiol 1995; 48: 23-40.
25. Eapen ZJ, Lauer MS, Temple RJ. The imperative of overcoming barriers to the conduct of large, simple trials. JAMA 2014; 311(14): 1397-1398.
26. Ford I, Norrie J. Pragmatic trials. N Engl J Med 2016; 375(5): 454-463.

27. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ* 2015; 350: h2147.
28. Patient-Centered Outcomes Research Institute. Guidance on the design and conduct of trials in real-world settings: factors to consider in pragmatic patient centered outcomes research [online]. [Zugriff: 18.12.2019]. URL: <https://www.pcori.org/sites/default/files/PCORI-Guidance-Design-and-Conduct-of-Trials-Real-World-Settings-Factors-to-Consider-Pragmatic-PCOR.pdf>.
29. Zuidgeest MGP, Goetz I, Groenwold RHH, Irving E, Van Thiel G, Grobbee DE. Series: pragmatic trials and real world evidence; paper 1: introduction. *J Clin Epidemiol* 2017; 88: 7-13.
30. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews* 2015; 4: 1.
31. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004; 23(20): 3105-3124.
32. Rücker G. Network meta-analysis, electrical networks and graph theory. *Res Syn Meth* 2012; 3(4): 312-324.
33. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 1997; 50(6): 683-691.
34. March S, Andrich S, Drepper J, Horenkamp-Sonntag D, Icks A, Ihle P et al. Gute Praxis Datenlinkage (GPD). *Gesundheitswesen* 2019; 81(8-9): 636-650.
35. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Biotechnologisch hergestellte Wirkstoffe bei rheumatoider Arthritis: Abschlussbericht; Auftrag A16-70 [online]. 23.07.2019 [Zugriff: 30.10.2019]. URL: https://www.iqwig.de/download/A16-70_Biotechnologisch-hergestellte-Wirkstoffe-bei-rheumatoider-Arthritis_Abschlussbericht_V1-0.pdf.
36. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden: Version 5.0. Köln: IQWiG; 2017. URL: https://www.iqwig.de/download/Allgemeine-Methoden_Version-5-0.pdf.
37. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Linagliptin: erneute Nutzenbewertung gemäß § 35a Absatz 5b SGB V; Dossierbewertung; Auftrag A12-11 [online]. 29.11.2012 [Zugriff: 11.03.2013]. (IQWiG-Berichte; Band 144). URL: https://www.iqwig.de/download/A12-11_Linagliptin_Erneute_Nutzenbewertung_35a_Absatz_5b_SGB_V.pdf.
38. Song F, Loke YK, Walsh T, Glenny AM, Eastwood AJ, Altman DG. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. *BMJ* 2009; 338: b1147.

39. Donegan S, Williamson P, Gamble C, Tudur-Smith C. Indirect comparisons: a review of reporting and methodological quality. *PLoS One* 2010; 5(11): e11054.
40. Kiefer C, Sturtz S, Bender R. Indirekte Vergleiche und Netzwerk-Metaanalysen: Schätzung von Effekten bei fehlenden direkten Vergleichsstudien; Teil 22 der Serie zur Bewertung wissenschaftlicher Publikationen. *Dtsch Arztebl* 2015; 112(47): 803-808.
41. Meinecke AK, Welsing P, Kafatos G, Burke D, Trelle S, Kubin M et al. Series: pragmatic trials and real world evidence; paper 8: data collection and management. *J Clin Epidemiol* 2017; 91: 13-22.
42. Lauer MS, D'Agostino RB. The Randomized Registry Trial: the next disruptive technology in clinical research? *N Engl J Med* 2013; 369(17): 1579-1581.
43. Li G, Sajobi TT, Menon BK, Korngut L, Lowerison M, James M et al. Registry-based randomized controlled trials: what are the advantages, challenges, and areas for future research? *J Clin Epidemiol* 2016; 80: 16-24.
44. Foroughi S, Wong HI, Gately L, Lee M, Simons K, Tie J et al. Re-inventing the randomized controlled trial in medical oncology: the registry-based trial. *Asia Pac J Clin Oncol* 2018; 14(6): 365-373.
45. Lagerqvist B, Fröbert O, Olivecrona GK, Gudnason T, Maeng M, Alström P et al. Outcomes 1 year after thrombus aspiration for myocardial infarction. *N Engl J Med* 2014; 371(12): 1111-1120.
46. Glinatsi D, Heiberg MS, Rudin A, Nordstrom D, Haavardsholm EA, Gudbjornsson B et al. Head-to-head comparison of aggressive conventional therapy and three biological treatments and comparison of two de-escalation strategies in patients who respond to treatment: study protocol for a multicenter, randomized, open-label, blinded-assessor, phase 4 study. *Trials* 2017; 18(1): 161.
47. Mathes T, Buehn S, Prengel P, Pieper D. Registry-based randomized controlled trials merged the strength of randomized controlled trials and observational studies and give rise to more pragmatic trials. *J Clin Epidemiol* 2018; 93: 120-127.
48. Yndigegn T, Hofmann R, Jernberg T, Gale CP. Registry-based randomised clinical trial: efficient evaluation of generic pharmacotherapies in the contemporary era. *Heart* 2018; 104(19): 1562.
49. Eichler HG, Bloechl-Daum B, Broich K, Kyrle PA, Oderkirk J, Rasi G et al. Data rich, information poor: can we use electronic health records to create a learning healthcare system for pharmaceuticals? *Clin Pharmacol Ther* 2019; 105(4): 912-922.
50. Casey JA, Schwartz BS, Stewart WF, Adler NE. Using electronic health records for population health research: a review of methods and applications. *Annu Rev Public Health* 2016; 37: 61-81.

51. Cave A, Kurz X, Arlett P. Real-world data for regulatory decision making: challenges and possible solutions for Europe. *Clin Pharmacol Ther* 2019; 106(1): 36-39.
52. Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JPA. Routinely collected data and comparative effectiveness evidence: promises and limitations. *CMAJ* 2016; 188(8): E158-E164.
53. Horenkamp-Sonntag D, Linder R, Engel S, Verheyen F. Brachytherapie beim Prostata-Ca: Möglichkeiten und Limitationen von GKV-Routinedaten für eine Nutzenbewertung. *Gesundheitswesen* 2016; 78(5): 298-305.
54. Medicines & Healthcare products Regulatory Agency. MHRA position statement and guidance: electronic health records [online]. 16.09.2015 [Zugriff: 18.12.2019]. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/470228/Electronic%20Health%20Records%20MHRA%20Position%20Statement.pdf.
55. Food and Drug Administration. Use of electronic health record data in clinical investigations: guidance for industry [online]. 07.2018 [Zugriff: 21.11.2019]. URL: <https://www.fda.gov/media/97567/download>.
56. European Medicines Agency. Reflection paper on expectations for electronic source data and data transcribed to electronic data collection tools in clinical trials [online]. 09.06.2010 [Zugriff: 21.11.2019]. URL: https://www.ema.europa.eu/documents/regulatory-procedural-guideline/reflection-paper-expectations-electronic-source-data-data-transcribed-electronic-data-collection_en.pdf.
57. Bertelsmann Stiftung. Elektronische Patientenakten: Einrichtungsübergreifende Elektronische Patientenakten als Basis für integrierte patientenzentrierte Behandlungsmanagement-Plattformen [online]. 04.2017 [Zugriff: 18.12.2019]. URL: https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/VV_eEPA_Expertise_final.pdf.
58. Garets D, Davis M. Electronic medical records vs. electronic health records: yes, there is a difference; a HIMSS Analytics white paper. Chicago: HIMSS Analytics; 2006.
59. Organisation for Economic Co-operation and Development. Using routinely collected data to inform pharmaceutical policies: analytical report for OECD and EU countries [online]. 2019 [Zugriff: 24.05.2019]. URL: <https://www.oecd.org/health/health-systems/Using-Routinely-Collected-Data-to-Inform-Pharmaceutical-Policies-Analytical-Report-2019.pdf>.
60. Pacurariu A, Plueschke K, McGettigan P, Morales DR, Slattery J, Vogl D et al. Electronic healthcare databases in Europe: descriptive analysis of characteristics and potential for use in medicines regulation. *BMJ Open* 2018; 8(9): e023090.
61. De Moor G, Sundgren M, Kalra D, Schmidt A, Dugas M, Claerhout B et al. Using electronic health records for clinical research: the case of the EHR4CR project. *J Biomed Inform* 2015; 53: 162-173.

62. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. *J Am Med Inform Assoc* 2014; 21(4): 578-582.
63. Hernandez AF, Fleurence RL, Rothman RL. The ADAPTABLE Trial and PCORnet: shining light on a new research paradigm. *Ann Intern Med* 2015; 163(8): 635-636.
64. Caumanns J. Zur Diskussion: Stand der Digitalisierung im deutschen Gesundheitswesen. *Z Evid Fortbild Qual Gesundhwes* 2019; 143: 22-29.
65. Stephani V, Busse R, Geissler A. Benchmarking in der Krankenhaus-IT: Deutschland im internationalen Vergleich. In: Klauber J, Geraedts M, Friedrich J, Wasem J (Ed). *Krankenhaus-Report 2019*. Berlin: Springer; 2019. URL: https://link.springer.com/content/pdf/10.1007%2F978-3-662-58225-1_2.pdf.
66. Krüger-Brand HE. Medizininformatik-Initiative: Impulse für die digitale Medizin. *Dtsch Arztbl* 2019; 116(42): A1883-A1888.
67. Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PR, Bernstam EV et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. *Med Care* 2013; 51(8 Suppl 3): S30-S37.
68. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc* 2013; 20(1): 144-151.
69. Cowie MR, Blomster JI, Curtis LH, Duclaux S, Ford I, Fritz F et al. Electronic health records to facilitate clinical research. *Clin Res Cardiol* 2017; 106(1): 1-9.
70. Ehrenstein V, Petersen I, Smeeth L, Jick SS, Benchimol EI, Ludvigsson JF et al. Helping everyone do better: a call for validation studies of routinely recorded health data. *Clin Epidemiol* 2016; 8: 49-51.
71. Mc Cord KA, Al-Shahi Salman R, Treweek S, Gardner H, Strech D, Whiteley W et al. Routinely collected data for randomized trials: promises, barriers, and implications. *Trials* 2018; 19(1): 29.
72. Najafzadeh M, Schneeweiss S. From trial to target populations: calibrating real-world data. *N Engl J Med* 2017; 376(13): 1203-1205.
73. Mc Cord KA, Ewald H, Ladanie A, Briel M, Speich B, Bucher HC et al. Current use and costs of electronic health records for clinical trial research: a descriptive study. *CMAJ Open* 2019; 7(1): E23-E32.
74. Brennan L, Watson M, Klaber R, Charles T. The importance of knowing context of hospital episode statistics when reconfiguring the NHS. *BMJ* 2012; 344: e2432.
75. Green SM. Congruence of disposition after emergency department intubation in the National Hospital Ambulatory Medical Care Survey. *Ann Emerg Med* 2013; 61(4): 423-426.e8.

76. Bayley KB, Belnap T, Savitz L, Masica AL, Shah N, Fleming NS. Challenges in using electronic health record data for CER: experience of 4 learning organizations and solutions applied. *Med Care* 2013; 51(8 Suppl 3): S80-S86.
77. Hall GC, Sauer B, Bourke A, Brown JS, Reynolds MW, LoCasale R. Guidelines for good database selection and use in pharmacoepidemiology research. *Pharmacoepidemiol Drug Saf* 2012; 21(1): 1-10.
78. Guinn D, Wilhelm EE, Lieberman G, Khozin S. Assessing function of electronic health records for real-world data generation. *BMJ Evid Based Med* 2019; 24(3): 95-98.
79. Wasserman RC. Electronic medical records (EMRs), epidemiology, and epistemology: reflections on EMRs and future pediatric clinical research. *Acad Pediatr* 2011; 11(4): 280-287.
80. Pokora RM, Cornet LL, Daumke P, Mildenberger P, Zeeb H, Blettner M. Validierung von semantischen Analysen von unstrukturierten medizinischen Daten für Forschungszwecke. *Gesundheitswesen* 09.10.2019 [Epub ahead of print].
81. Köpcke F, Trinczek B, Majeed RW, Schreiweis B, Wenk J, Leusch T et al. Evaluation of data completeness in the electronic health record for the purpose of patient recruitment into clinical trials: a retrospective analysis of element presence. *BMC Med Inform Decis Mak* 2013; 13: 37.
82. Lau EC, Mowat FS, Kelsh MA, Legg JC, Engel-Nitz NM, Watson HN et al. Use of electronic medical records (EMR) for oncology outcomes research: assessing the comparability of EMR information to patient registry and health claims data. *Clin Epidemiol* 2011; 3: 259-272.
83. Moore TJ, Furberg CD. Electronic health data for postmarket surveillance: a vision not realized. *Drug Saf* 2015; 38(7): 601-610.
84. PMV Forschungsgruppe. Gutachten: Daten für die Versorgungsforschung; Zugang und Nutzungsmöglichkeiten; Datengutachten für das Deutsche Institut für Medizinische Dokumentation und Information (DIMDI) [online]. 08.10.2014 [Zugriff: 18.12.2019]. URL: <https://www.dimdi.de/static/.downloads/deutsch/dimdi-sekundaerdaten-expertise.pdf>.
85. Horenkamp-Sonntag D, Linder R, Wenzel F, Gerste B, Ihle P. Prüfung der Datenqualität und Validität von GKV-Routinedaten. In: Swart E, Ihle P, Gothe H, Matusiewicz D (Ed). *Routinedaten im Gesundheitswesen*. Bern: Huber; 2014. S. 314-330.
86. Erler A, Beyer M, Muth C, Gerlach FM, Brennecke R. Garbage in - Garbage out? Validität von Abrechnungsdiagnosen in hausärztlichen Praxen. *Gesundheitswesen* 2009; 71(12): 823-831.
87. Munch C, Gottschall M, Hübsch G, Köberlein-Neu J, Schubel J, Bergmann A et al. Qualität der hausärztlichen Diagnosedokumentation in Patientenakten: eine Analyse am Beispiel von Schilddrüsenerkrankungen. *Z Evid Fortbild Qual Gesundhwes* 2016; 115-116: 56-62.

88. Czwikla J, Domhoff D, Giersiepen K. ICD-Codierqualität ambulanter Krebsdiagnosen in GKV-Routinedaten. *Z Evid Fortbild Qual Gesundhwes* 2016; 118-119: 48-55.
89. Gerste B, Gutschmidt S. Datenqualität von Diagnosedaten aus dem ambulanten Bereich: kritische Anmerkungen am Beispiel Diabetes. *Gesundheits- und Sozialpolitik* 2006; (3-4): 10-24.
90. Schubert I, Ihle P, Koster I. Interne Validierung von Diagnosen in GKV-Routinedaten: Konzeption mit Beispielen und Falldefinition. *Gesundheitswesen* 2010; 72(6): 316-322.
91. Walker J, Buchwald V. Plausibilität der Diagnosekodierung in der ambulanten Versorgung: Analyse auf Basis eines Ärztepanels. In: Repschläger U, Schulte C, Osterkamp N (Ed). *Gesundheitswesen aktuell 2011*. Barmer GEK; 2011. S. 128-151.
92. AG medizinische Grouperanpassung. Analysen der Diagnoseeinträge in der Datengrundlage des Bewertungsausschusses für die Ermittlung der diagnosebezogenen Veränderungsraten für das Jahr 2016 (Geburtstagsstichprobe Version 7.0) [online]. 14.06.2017 [Zugriff: 08.01.2020]. URL: https://www.gkv-spitzenverband.de/media/dokumente/krankenversicherung_1/aerztliche_versorgung/verguetung_und_leistungen/klassifikationsverfahren/7_Analysen_der_Diagnoseeintraege_in_der_Daten_grundlage_des_BA_fuer_die_Ermittlung_der_diagnosebezogenen_Veraenderungsraten_fuer_das_Jahr_2016_Stand_14.06.2017.pdf.
93. Drösler S, Garbe E, Hasford J, Schubert I, Ulrich V, Van de Ven W et al. Sondergutachten zu den Wirkungen des morbiditätsorientierten Risikostrukturausgleichs: erstellt durch den Wissenschaftlichen Beirat zur Weiterentwicklung des Risikostrukturausgleichs im Auftrag des Bundesministeriums für Gesundheit [online]. 27.11.2017 [Zugriff: 08.01.2020]. URL: https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/5_Publikationen/Gesundheit/Berichte/Sondergutachten_Wirkung_RSA_2017.pdf.
94. Gerste B, Günster C. Erkrankungshäufigkeiten und Inanspruchnahme von Gesundheitsleistungen. In: Günster C, Klose J, Schmacke N (Ed). *Versorgungsreport 2012*. Stuttgart: Schattauer; 2012. S. 315-384.
95. Fiori W, Brüning K, Bunzemeier H, Roeder N. Unberechtigte Vorwürfe. *Dtsch Arztebl* 2009; 106(33): A1596-A1598.
96. Maass C, Kuske S, Lessing C, Schrappe M. Are administrative data valid when measuring patient safety in hospitals? A comparison of data collection methods using a chart review and administrative data. *Int J Qual Health Care* 2015; 27(4): 305-313.
97. Hoffmann F, Andersohn F, Giersiepen K, Scharnetzky E, Garbe E. Validierung von Sekundärdaten: Grenzen und Möglichkeiten. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 2008; 51(10): 1118-1126.

98. GKV-Spitzenverband. Anzahl der Krankenkassen im Zeitverlauf: Konzentrationsprozess durch Fusionen (Angaben zum Stichtag 1.1.) [online]. [Zugriff: 18.12.2019]. URL: https://www.gkv-spitzenverband.de/media/grafiken/krankenkassen/Grafik_Krankenkassenanzahl_Konzentrationsprozess_300dpi_2019-03-29.jpg.
99. Hoffmann F, Icks A. Unterschiede in der Versichertenstruktur von Krankenkassen und deren Auswirkungen für die Versorgungsforschung: Ergebnisse des Bertelsmann-Gesundheitsmonitors. *Gesundheitswesen* 2012; 74(5): 291-297.
100. Hoffmann F, Koller D. Verschiedene Regionen, verschiedene Versichertenpopulationen? Soziodemografische und gesundheitsbezogene Unterschiede zwischen Krankenkassen. *Gesundheitswesen* 2017; 79(1): e1-e9.
101. Schafer I, Reusch M, Siebert J, Hilbring C, Augustin M. Assoziation von Krankenversicherung und soziodemografischen Faktoren mit der Versorgung maligner Melanome. *Gesundheitswesen* 2017; 79(1): 21-27.
102. Ohlmeier C. Versorgungsepidemiologische Analysen kardiovaskulärer Erkrankungen auf Basis von Routinedaten der gesetzlichen Krankenkassen [Dissertation] Bremen: Universität Bremen; 2015.
103. Marx MM, Dulas FM, Schumacher KM. Verbesserung der Sichtbarkeit seltener Erkrankungen in Gesundheitssystemen durch spezifische Routinekodierung. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 2017; 60(5): 532-536.
104. Schröder H. Arzneimittelverordnungen. In: Swart E, Ihle P, Gothe H, Matusiewicz D (Ed). Routinedaten im Gesundheitswesen. Bern: Huber; 2014. S. 74-87.
105. Kuklik N, Stausberg J, Jockel KH. Adverse drug events in German hospital routine data: a validation of International Classification of Diseases, 10th revision (ICD-10) diagnostic codes. *PLoS One* 2017; 12(11): e0187510.
106. Schellschmidt H, Mansky T, Heller G, Robra B. Indikatoren und Messinstrumente zur Bewertung der medizinischen Ergebnisqualität im Krankenhaus. In: Stock J, Szecsenyi J (Ed). Stichwort: Qualitätsindikatoren. Bonn: Kompart Verlagsgesellschaft; 2007. S. 173-200.
107. Wilchesky M, Tamblyn RM, Huang A. Validation of diagnostic codes within medical services claims. *J Clin Epidemiol* 2004; 57(2): 131-141.
108. Nabhan C, Klink A, Prasad V. Real-world evidence: what does it really mean? *JAMA Oncol* 2019; 5(6): 781-783.
109. Motheral BR, Fairman KA. The use of claims databases for outcomes research: rationale, challenges, and strategies. *Clin Ther* 1997; 19(2): 346-366.
110. Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q* 1966; 44(3 Suppl): 166-206.

111. Rubin DB. The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Stat Med* 2007; 26(1): 20-36.
112. Berger ML, Mamdani M, Atkins D, Johnson ML. Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources; the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report; part I. *Value Health* 2009; 12(8): 1044-1052.
113. 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 Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report; part II. *Value Health* 2009; 12(8): 1053-1061.
114. 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 Health* 2009; 12(8): 1062-1073.
115. Agoritsas T, Merglen A, Shah ND, O'Donnell M, Guyatt GH. Adjusted analyses in studies addressing therapy and harm: users' guides to the medical literature. *JAMA* 2017; 317(7): 748-759.
116. D' Agostino RB Jr, D' Agostino RB Sr. Estimating treatment effects using observational data. *J Am Med Assoc* 2007; 297(3): 314-316.
117. Stukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. *J Am Med Assoc* 2007; 297(3): 278-285.
118. Kuss O, Blettner M, Borgermann J. Propensity Score: an alternative method of analyzing treatment effects. *Dtsch Arztebl Int* 2016; 113(35-36): 597-603.
119. Austin PC. The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments. *Stat Med* 2014; 33(7): 1242-1258.
120. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 1998; 17(19): 2265-2281.
121. Desai RJ, Franklin JM. Alternative approaches for confounding adjustment in observational studies using weighting based on the propensity score: a primer for practitioners. *BMJ* 2019; 367: l5657.

122. Austin PC, Mamdani MM, Stukel TA, Anderson GM, Tu JV. The use of the propensity score for estimating treatment effects: administrative versus clinical data. *Stat Med* 2005; 24(10): 1563-1578.
123. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology* 2009; 20(4): 512-522.
124. Wyss R, Fireman B, Rassen JA, Schneeweiss S. Erratum: high-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology* 2018; 29(6): e63-e64.
125. Austin PC, Wu CF, Lee DS, Tu JV. Comparing the high-dimensional propensity score for use with administrative data with propensity scores derived from high-quality clinical data. *Stat Methods Med Res* 11.04.2019 [Epub ahead of print].
126. Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. *Stat Sci* 1999; 14(1): 29-46.
127. Skipka G, Wieseler B, Kaiser T, Thomas S, Bender R, Windeler J et al. Methodological approach to determine minor, considerable, and major treatment effects in the early benefit assessment of new drugs. *Biom J* 2016; 58(1): 43-58.
128. International Conference on Harmonisation E9 Expert Working Group. ICH harmonised tripartite guideline: statistical principles for clinical trials. *Stat Med* 1999; 18(15): 1905-1942.
129. European Medicines Agency. Clinical efficacy and safety guidelines: biostatistics [online]. 2019 [Zugriff: 28.11.2019]. URL: <https://www.ema.europa.eu/en/human-regulatory/research-development/scientific-guidelines/clinical-efficacy-safety/biostatistics>.
130. Fortier I, Raina P, Van den Heuvel ER, Griffith LE, Craig C, Saliba M et al. Maelstrom Research guidelines for rigorous retrospective data harmonization. *Int J Epidemiol* 2017; 46(1): 103-115.
131. Mathew T, Nordström K. Comparison of one-step and two-step meta-analysis models using individual patient data. *Biom J* 2010; 52(2): 271-287.
132. Taichman DB, Backus J, Baethge C, Bauchner H, De Leeuw PW, Drazen JM et al. Sharing clinical trial data: a proposal from the International Committee of Medical Journal Editors. *PLoS Med* 2016; 13(1): e1001950.
133. Biner BE, Li R, Barnes M, Sim I. A global, neutral platform for sharing trial data. *N Engl J Med* 2016; 374(25): 2411-2413.
134. Agency EM. European Medicines Agency policy on publication of clinical data for medicinal products for human use (Policy/0070) [online]. [Zugriff: 08.01.2020]. URL: https://www.ema.europa.eu/documents/other/european-medicines-agency-policy-publication-clinical-data-medicinal-products-human-use_en.pdf.

135. Elm Ev, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandebroucke JP. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 2007; 335(7624): 806.
136. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I et al. The Reporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med* 2015; 12(10): e1001885.
137. Langan SM, Schmidt SA, Wing K, Ehrenstein V, Nicholls SG, Filion KB et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). *BMJ* 2018; 363: k3532.
138. Freemantle N, Blonde L, Bolinder B, Gerber RA, Hobbs FD, Martinez L et al. Real-world trials to answer real-world questions. *Pharmacoeconomics* 2005; 23(8): 747-754.
139. Freemantle N, Strack T. Real-world effectiveness of new medicines should be evaluated by appropriately designed clinical trials. *J Clin Epidemiol* 2010; 63(10): 1053-1058.
140. Eichler HG, Koenig F, Arlett P, Enzmann H, Humphreys A, Petavy F et al. Are novel, non-randomised analytic methods fit for decision-making? The need for prospective, controlled and transparent validation. *Clin Pharmacol Ther* 01.10.2019 [Epub ahead of print].
141. Dhiman P, Lee H, Kirtley S, Collins GS. More consideration was needed when conducting non-randomised studies of interventions. *J Clin Epidemiol* 2019; 117: 99-108.
142. Bartlett VL, Dhruva SS, Shah ND, Ryan P, Ross JS. Feasibility of using real-world data to replicate clinical trial evidence. *JAMA Netw Open* 2019; 2(10): e1912869.
143. Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. *BMJ* 2007; 334(7589): 349-351.
144. Booth CM, Karim S, Mackillop WJ. Real-world data: towards achieving the achievable in cancer care. *Nat Rev Clin Oncol* 2019; 16(5): 312-325.
145. Corrigan-Curay J, Sacks L, Woodcock J. Real-world evidence and real-world data for evaluating drug safety and effectiveness. *JAMA* 2018; 320(9): 867-868.
146. Woodcock A, Bakerly ND, New JP, Gibson JM, Wu W, Vestbo J et al. The Salford Lung Study protocol: a pragmatic, randomised phase III real-world effectiveness trial in asthma. *BMC Pulm Med* 2015; 15(1): 160.
147. Hee SW, Willis A, Tudur Smith C, Day S, Miller F, Madan J et al. Does the low prevalence affect the sample size of interventional clinical trials of rare diseases? An analysis of data from the aggregate analysis of clinicaltrials.gov. *Orphanet J Rare Dis* 2017; 12(1): 44.
148. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Bewertung und Auswertung von Studien bei seltenen Erkrankungen: Rapid Report; Auftrag MB13-01 [online]. 05.09.2014 [Zugriff: 14.10.2014]. (IQWiG-Berichte; Band 241). URL: https://www.iqwig.de/download/MB13-01_Rapid-Report_Studien-bei-seltenen-Erkrankungen.pdf.

149. Renfro LA, Ji L, Piao J, Onar-Thomas A, Kairalla JA, Alonzo TA. Trial design challenges and approaches for precision oncology in rare tumors: experiences of the Children's Oncology Group. *JCO Precision Oncology* 2019; 3: 1-13.

Anhang A – Fragebogen für die Interviews zu Patientenregistern

Hintergrund

Das IQWiG ist vom Gemeinsamen Bundesausschuss (G-BA) mit einer wissenschaftlichen Ausarbeitung von Konzepten zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln beauftragt worden. Zu Ihrer Information finden Sie hier den Auftrag und seine Konkretisierung durch den G-BA:

- https://www.g-ba.de/downloads/39-261-3773/2019-05-02_IQWIG-Beauftragung_Konzepte-Daten-Nutzenbewertung.pdf
- https://www.g-ba.de/downloads/40-268-5720/2019-05-02_IQWIG-Beauftragung_Konzepte-Daten-Nutzenbewertung_Auftragskonkretisierung.pdf

Ein Element des Konzepts wird die Erhebung von Daten in Registern sein. Der Auftrag umfasst u. a. die Spezifizierung von Kriterien für die Qualität und methodische Anforderungen an die im Rahmen der jeweiligen Datengenerierung erhobenen Daten, sowie die Beschreibung von Maßnahmen zur Sicherstellung der Datenqualität. Für diese Aspekte würden wir gern auch von Ihren Erfahrungen aus der Praxis von Registern profitieren. Dabei beschränkt sich der Auftrag und damit auch unser Interview auf die Datenerhebung und -auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln.

Inhalt des Fragebogens

Der vorliegende Fragebogen besteht aus 3 Teilen:

- In Teil I stellen wir Ihnen einige Fragen zu Ihrem eigenen Register, um dieses näher kennenzulernen. Hier bitten wir Sie um stichpunktartige Antworten auf die Freitextfragen bzw. um konkrete Angaben zu Mengen, Zeitpunkten etc.
- Teil II stellt eine tabellarische Auflistung allgemeiner Qualitätskriterien dar. Hier bitten wir Sie um Ihre Einschätzung zur Relevanz und Umsetzbarkeit der genannten Qualitätskriterien. Außerdem bitten wir Sie um eine Einschätzung zum Erfüllungsgrad der einzelnen Qualitätskriterien für Ihr Register.
- Teil III benennt 2 Fragestellungen zur Nutzenbewertung (Szenarien). Anhand dieser Szenarien möchten wir mit Ihnen im Interview besprechen, ob und wie Sie sich die Beantwortung entsprechender Fragestellungen in Ihrem Register vorstellen könnten. Eine schriftliche Rückmeldung zu diesen Szenarien vorab ist nicht erforderlich.

Der Fragebogen dient der Vorbereitung unseres Interviews. Wir bitten Sie daher, uns den ausgefüllten Fragebogen 2 Arbeitstage vor dem Interviewtermin zu schicken. Sollten es Ihnen aus zeitlichen Gründen nicht möglich sein, beide Teile I und II vor dem Interview auszufüllen und uns zu übermitteln, dann bearbeiten Sie bitte vorrangig Teil II des Fragebogens

(Qualitätskriterien). Sie können dann gerne Teil I erst nach dem Interview ausfüllen und uns schicken.

Ablauf des Interviews

Im ersten Teil des Interviews werden wir anhand Ihrer Rückmeldung zu Teil II des Fragebogens konkrete Qualitätskriterien und Ihre Einschätzung zur Relevanz und zum Aufwand ansprechen. Sollten Sie Teil I des Fragebogens vorab übermittelt haben, werden wir ggf. einzelne dort genannte Punkte ansprechen

Im zweiten Teil des Interviews werden wir mit Ihnen die anhand der genannten Szenarien zur Nutzenbewertung Ihre Einschätzung zur Eignung des Registers für die Beantwortung dieser Fragestellungen sowie das konkrete Vorgehen diesbezüglich besprechen.

Ein Protokoll des Interviews (Verlaufs- oder Ergebnisprotokoll) wird nicht erstellt. Zur Gewährleistung der Nachvollziehbarkeit des Gesprächsinhalts nehmen wir das Interview allerdings auf. Diese Aufnahme wird nur für interne Zwecke benötigt. Eine Veröffentlichung erfolgt nicht, und die Aufnahme wird auch nicht außerhalb der Auftragsbearbeitung verwendet.

Abbildung der Inhalte des Fragebogens und des Interviews im Bericht

In unserem Bericht werden wir die betrachteten Register und die Diskussionspartner benennen. Informationen zu den einzelnen Registern, die nicht öffentlich zugänglich sind, sowie wörtliche Zitate aus den Interviews werden wir nicht in den Bericht aufnehmen.

Die Rückmeldungen zu Teil II des Fragebogens einschließlich des zugehörigen Austauschs im Interview werden wir in aggregierter Form im Bericht abbilden (z. B. Angaben dazu, welche Qualitätskriterien von den Befragten am häufigsten als besonders relevant angesehen wurden, welche Qualitätskriterien nach Angabe der Befragten mit hohem Aufwand verbunden ist und warum etc.).

Teil I: Informationen zu Ihrem Register

- 1) Welche Patientengruppe (-n) ist / sind in Ihrem Register eingeschlossen?
- 2) Seit welchem Jahr existiert das Patientenregister?
- 3) Was sind die Hauptziele bzw. Fragestellungen Ihres Patientenregisters?
- 4) Wer ist der Träger des Patientenregisters?
- 5) Wie finanziert sich das Patientenregister?
- 6) Welche medizinischen Versorgungseinrichtungen (z. B. Akutkliniken, REHA-Kliniken, Praxen von niedergelassenen Ärzten) sind in das Patientenregister für die Datenerhebung eingebunden?
- 7) Wie hoch war initial die Anzahl beteiligter Erhebungszentren und wie ist sie aktuell?
- 8) Wie erfolgt die Meldung der Daten von den Erhebungszentren an das Patientenregister?
- 9) Gibt es eine Rückmeldung an die beteiligten Erhebungszentren und in welcher Form?
- 10) Über welche personelle Infrastruktur verfügt das Patientenregister (Vollzeitkräfte, Qualifikation)?
- 11) Welcher Art sind die im Patientenregister erhobenen Daten (Mehrfachnennung möglich)?
 - soziodemografische Daten
 - anthropometrische Daten
 - klinische Daten
 - genetische Daten
 - Patient Reported Outcomes (PROs) / gesundheitsbezogene Lebensqualität
 - Versorgungsdaten
 - Laborparameter
- 12) Wie viele Variablen erheben Sie ungefähr pro Patient und Erhebungszeitpunkt?
- 13) Wie oft bzw. bei welchen Anlässen erheben Sie die Daten von den Registerpatienten?
- 14) Welche Hauptmaßnahmen nutzen Sie zur Qualitätssicherung der erhobenen Daten in Ihrem Patientenregister?
- 15) Existiert eine schriftliche Registerbeschreibung bzw. ein entsprechendes Protokoll für das Patientenregister?
- 16) Haben Sie bereits Erfahrungen mit Registerstudien zur Bewertung medizinischer Interventionen gemacht?
 - a) Falls ja, waren darunter auch vergleichende Bewertungen von Interventionen bzw. Outcome-Studien?

- b) Falls ja, waren darunter auch randomisierte vergleichende Bewertungen von Interventionen?
 - c) Falls ja, haben Sie solche Studien auch in nationaler oder internationaler Kooperation mit anderen Registerbetreibern durchgeführt?
 - d) Falls ja, benennen Sie bitte die aus Ihrer Sicht wichtigsten Projekte unter Angabe zugehöriger Publikationen (sofern vorhanden)
- 17) Wir würden uns freuen, wenn Sie uns Hinweise auf schriftliche, öffentlich verfügbare Informationen zu Ihrem Patientenregister geben könnten, z. B. auf der Webseite des Registers oder in der Fachliteratur.

Teil II: Allgemeine Qualitätskriterien für Register

Bitte benennen Sie durch Ankreuzen [x] von den in der Tabelle aufgelisteten Qualitätskriterien jeweils aus Ihrer fachlichen Sicht die Kriterien, auf die die in den Spalten a. bis d. genannte Eigenschaft (hohe Relevanz, Umsetzungsschwierigkeiten, einfache Umsetzbarkeit, hoher Aufwand) zutrifft. Benennen Sie dabei bitte je Spalte a. bis d. nur jeweils 5 (ggf. unterschiedliche) Kriterien.

Zusätzlich möchten wir Sie im Hinblick auf das eigene Patientenregister um eine grobe Einschätzung zum Erfüllungsgrad aller 29 Kriterien bitten, und zwar auf einer vierstufigen Skala von 1 = nicht erfüllt bis 4 = voll erfüllt (letzte Spalte).

Nr.	Qualitätskriterium	Hohe Relevanz	Umsetzungsschwierigkeiten in der Praxis	Einfache Umsetzbarkeit in der Praxis	Hoher Aufwand (Zeit, Ressourcen)	Erfüllungsgrad im eigenen Register 1 = nicht erfüllt bis 4 = voll erfüllt
Systematik						
1	Detaillierte Registerbeschreibung (Protokoll)					
Standardisierung						
2	Exakte Definition / Operationalisierung von Expositionen, klin. Ereignissen, Endpunkten und Confoundern					
3	Aktueller Datenplan / Kodierhandbuch					
4	Verwendung von Standard-Klassifikationen (z. B. ICD-10) und -Terminologien (z. B. MedDRA)					
5	Verwendung von validierten Standard-Erhebungsinstrumenten (Fragebogen, Skalen)					
6	Schulungen zur Datenerhebung und -erfassung					
7	Umsetzung eines konsentierten krankheitsspezifischen Kerndatensatzes („core data set“)					

Nr.	Qualitätskriterium	Hohe Relevanz	Umsetzungsschwierigkeiten in der Praxis	Einfache Umsetzbarkeit in der Praxis	Hoher Aufwand (Zeit, Ressourcen)	Erfüllungsgrad im eigenen Register 1 = nicht erfüllt bis 4 = voll erfüllt
8	Verwendung exakter Zeitdaten zum Patienten (z. B. Geburt, Tod, Schwangerschaft)					
9	Verwendung exakter Zeitdaten zur Erkrankung (z. B. definitive Diagnose, klinisch relevante Ereignisse)					
10	Verwendung exakter Zeitdaten zu wichtigen Untersuchungen					
11	Verwendung exakter Zeitdaten zu Behandlungen / Interventionen (z. B. bei Arzneimitteln Start-/Stop-Datum, Dosis, Dosisänderungen)					
Validität der Stichprobengewinnung						
12	Klar definierte Ein- und Ausschlusskriterien für Registerpatienten					
13	Vollzähligkeit der Registerpatienten					
14	Strategien zur Vermeidung ungewollter Selektionen beim Patienteneinschluss					
Validität der Datenerhebung						
15	Vollständigkeit der Daten					
16	Vollständigkeit der Erhebungszeitpunkte (drop-outs)					
17	Registermonitoring durch interne audits					
18	Registermonitoring durch externe audits					
19	QM-System mit regelmäßiger Erhebung von Qualitätsindikatoren					
20	SOPs (Standard Operating Procedures) zur Datenerhebung					

Nr.	Qualitätskriterium	Hohe Relevanz	Umsetzungsschwierigkeiten in der Praxis	Einfache Umsetzbarkeit in der Praxis	Hoher Aufwand (Zeit, Ressourcen)	Erfüllungsgrad im eigenen Register 1 = nicht erfüllt bis 4 = voll erfüllt
21	Source data verification (z. B. für 10 % zufällig ausgewählter Patienten je Erhebungszentrum)					
22	Erhebung und Handling sicherheitsrelevanter Daten (Adverse events) wie für PASS (Post-Authorization Safety Studies)					
Validität der statistischen Analysen und Berichte						
23	Prä-Spezifikation der Auswertungsmethodik					
24	Darlegung des Umgangs mit fehlenden Werten					
25	Vollständiger Bericht der Ergebnisse zu allen erhobenen Variablen					
26	Adjustierung der Ergebnisse für potenziell konfundierende oder Effekt modifizierende Variablen					
Übergeordnete Qualitätskriterien						
27	Transparenz eines Registers (Finanzierung, Entscheidungswege, Interessenkonflikte)					
28	Wissenschaftliche Unabhängigkeit					
29	Nutzung der vorhandenen Spontanmeldesysteme bei vermuteten UEs / UAWs					
MedDRA: Medizinisches Wörterbuch für Aktivitäten im Rahmen der Arzneimittelzulassung; ICD: Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme; PASS: Post-Authorization Safety Study; QM: Qualitätsmanagement; SOP: Standard Operation Procedure; UAW: unerwünschte Arzneimittelwirkung; UE: unerwünschtes Ereignis						

Sofern wichtige Qualitätskriterien in der tabellarischen Auflistung fehlen, benennen Sie diese bitte im Interview.

Teil III: Szenarien zur Nutzenbewertung von Arzneimitteln

Szenario 1

- Erkrankung: maligne Hautveränderungen ohne Spontanremission
- Intervention: neue Immuntherapie
- Zweckmäßige Vergleichstherapie des G-BA: bislang keine Therapiemöglichkeiten außer Best Supportive Care (BSC)
- Zulassung u.a. mit einarmiger Studie mit folgenden Hauptergebnissen:
 - Response bei 30 % der Patienten, Remission bei 5 % der Patienten; gemessen mit bildgebenden Verfahren, keine Beurteilung der Relevanz aus Patientensicht
 - Schwere unerwünschte Ereignisse bei 30 % der Patienten
- Ergebnis Nutzenbewertung: nicht quantifizierbarer Zusatznutzen, Begründung für „nicht quantifizierbar“:
 - Patientenrelevanz der Response unklar
 - Abwägung schwerer unerwünschter Ereignisse im Vergleich zur Vergleichstherapie (BSC) nicht möglich, da keine Daten für BSC vorliegen und die beobachteten schweren unerwünschten Ereignisse auch durch die Erkrankung selbst (und nicht durch die Therapie) bedingt sein können

Szenario 2

- Erkrankung: chronische Lungenerkrankung
- Intervention: inhalative Zweifachkombination mit einem neuen und einem bekannten Wirkstoff
- Vergleichstherapie: inhalative Zweifachkombination mit bekannten Wirkstoffen
- Zulassung u.a. mit direkt vergleichender RCT mit folgenden Hauptergebnissen:
 - Symptomverbesserung bei 30 % mit neuem Arzneimittel vs. 20 % mit Vergleichstherapie
 - Daten zur gesundheitsbezogenen Lebensqualität wegen großer Menge fehlender Werte nicht interpretierbar
 - Kein relevanter Unterschied im Bereich unerwünschter Ereignisse
- Ergebnis Nutzenbewertung: nicht quantifizierbarer Zusatznutzen, Begründung für „nicht quantifizierbar“:
 - Eskalation der Vergleichstherapie vor Studienbeginn nicht gemäß deutscher Leitlinien (Ort der Studiendurchführung: Nordamerika und Asien), Vergleichstherapie in der durchgeführten RCT unterdosiert; Effektstärke zur Symptomverbesserung und zu unerwünschten Ereignissen daher nicht abschließend einschätzbar
 - Daten zur gesundheitsbezogenen Lebensqualität fehlen

Anhang B – Diskussionsaspekte für den Expertenworkshop mit biometrischen Sachverständigen

Hintergrund

Das IQWiG ist vom Gemeinsamen Bundesausschuss (G-BA) mit einer wissenschaftlichen Ausarbeitung von Konzepten zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln beauftragt worden. Zu Ihrer Information finden Sie hier den Auftrag und seine Konkretisierung durch den G-BA:

- https://www.g-ba.de/downloads/39-261-3773/2019-05-02_IQWIG-Beauftragung_Konzepte-Daten-Nutzenbewertung.pdf
- https://www.g-ba.de/downloads/40-268-5720/2019-05-02_IQWIG-Beauftragung_Konzepte-Daten-Nutzenbewertung_Auftragskonkretisierung.pdf

Ein Element des Konzepts wird die Erhebung von Daten in Registern sein. Der Auftrag umfasst u. a. die Spezifizierung von Kriterien für die Qualität und methodische Anforderungen an die Berichterstattung, die Aufbereitung und Struktur sowie die statistische Auswertung der Daten, die im Rahmen des jeweiligen Konzepts der Datengenerierung erhoben werden. Für diese Aspekte würden wir gern auch von Ihren Erfahrungen aus der Praxis von Registern profitieren. Dabei beschränkt sich der Auftrag und damit auch unsere Expertendiskussion auf die Datenerhebung und -auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln.

Diskussionsaspekte für die Biometrische Expertenrunde

Folgende Punkte möchten wir gerne mit Ihnen am 19.07.2019 diskutieren:

Generelles:

- Welche Art von Registern in Deutschland erscheinen Ihnen prinzipiell für den Zweck der Nutzenbewertung von Arzneimitteln geeignet und welche nicht?
- Welche Anforderungen gelten für das Datenmanagement (Datenimport/-export, Zugang, Berechtigungen ...)?
- Welche Anforderungen gelten bezüglich Datenkorrekturen (Initiierung, Überwachung, Protokollierung)?
- Was sind die Anforderungen an die Dokumentation der Datenstruktur (Kodierung, Kategorien, Skala ...), auch in Hinblick auf Anpassungen im Zeitverlauf?
- Welche besonderen Herausforderungen sind an die multizentrische Organisation von Registern zu stellen?

Fehlende Daten:

- Wie sieht es mit der Vollzähligkeit und Vollständigkeit der grundsätzlich geeigneten Register aus?

- Gibt es neben dem Aspekt der Vollständigkeit weitere Probleme mit fehlenden Daten (z. B. zu einzelnen Zeitpunkten)?
- Welche Anforderungen gelten zum Umgang mit fehlenden Daten bei der statistischen Datenanalyse?
- Können diese Anforderungen in der Praxis eingehalten werden?

Statistische Aspekte:

- Welche Anforderungen gelten zum Umgang mit Confounding bei der statistischen Datenanalyse?
- Welche Auswirkungen haben die Anwendung der verschiedenen Methoden zur Adjustierung für Confounder (z. B. Propensity Scores) auf die Population, für die man eine Aussage machen kann?
- Werden Maßnahmen benötigt zur Berücksichtigung von multiplem Testen?
- Besteht eine Verpflichtung zur Fallzahlplanung bei statistischen Auswertungen von Registerdaten?
- Welche Anforderungen gelten zur Berücksichtigung von Clustereffekten (z. B. wenn die Exposition einrichtungsbezogen gesteuert wird)?

Sonstiges:

- Welche Anforderungen gelten bei der Durchführung register-basierter RCTs?
- Wie würden Sie ein neues Register planen, das für den Zweck der Nutzenbewertung von Arzneimitteln geeignet ist?

Diese Aspekte stellen nur ein Grundgerüst für die Diskussion am 19.07.2019 dar. Gerne können Sie weitere wichtige Aspekte benennen, die diskutiert werden sollen.

Zum Abschluss der Biometrischen Expertenrunde möchten wir gerne mit Ihnen ein Fallbeispiel durchgehen, um die Datenerhebung und -auswertung auf Basis von Registern zum Zwecke der Nutzenbewertung von Arzneimitteln etwas konkreter zu machen.

Fallbeispiel zur Diskussion in der Biometrischen Expertenrunde

- Erkrankung: maligne Hautveränderungen ohne Spontanremission
- Intervention: neue Immuntherapie
- Zweckmäßige Vergleichstherapie des G-BA: bislang keine Therapiemöglichkeiten außer Best Supportive Care (BSC)
- Zulassung u.a. mit einarmiger Studie mit folgenden Hauptergebnissen:
 - Response bei 30 % der Patienten, Remission bei 5 % der Patienten; gemessen mit bildgebenden Verfahren, keine Beurteilung der Relevanz aus Patientensicht

- Schwerwiegende unerwünschte Ereignisse bei 30 % der Patienten
- Ergebnis Nutzenbewertung: nicht quantifizierbarer Zusatznutzen, Begründung für „nicht quantifizierbar“:
 - Patientenrelevanz der Response unklar
 - Abwägung schwerwiegender unerwünschter Ereignisse im Vergleich zur Vergleichstherapie (BSC) nicht möglich, da keine Daten für BSC vorliegen und die beobachteten schweren unerwünschten Ereignisse auch durch die Erkrankung selbst (und nicht durch die Therapie) bedingt sein können
 - Keine Daten zur Lebensqualität erhoben, keine Daten zur Lebensqualität unter BSC vorgelegt

Ablauf der Expertenrunde

Ein Protokoll der Biometrischen Expertenrunde (Verlaufs- oder Ergebnisprotokoll) wird nicht erstellt. Zur Gewährleistung der Nachvollziehbarkeit des Gesprächsinhalts nehmen wir die Diskussion der Expertenrunde allerdings auf. Diese Aufnahme wird nur für interne Zwecke benötigt. Eine Veröffentlichung erfolgt nicht, und die Aufnahme wird auch nicht außerhalb der Auftragsbearbeitung verwendet.

Abbildung der Inhalte der Biometrischen Expertenrunde im Bericht

In unserem Bericht werden wir die diskutierten Aspekte zu den methodischen Anforderungen an die statistische Datenauswertung geeignet zusammenfassen (z.B. Angaben dazu, welche methodischen Anforderungen von den Experten als besonders relevant angesehen wurden, welche methodischen Anforderungen in der Praxis eingehalten werden können und welche nicht etc.). Wörtliche Zitate der einzelnen Experten werden wir nicht in den Bericht aufnehmen.

Anhang C – Qualitätskriterien für Register und Registerstudien in der Literatur

Tabelle 14: Abgleich von Qualitätskriterien für Register und Registerstudien in der Literatur (mehrseitige Tabelle)

Nr.	Qualitätskriterium	DNVF e. V. 2010/19 ^a	EMA Discussion Paper 2018	AHRQ- Handbuch 2014	REQuEST EUnetHTA 2019 ^b	CTTI Recommendations 2017
Systematik						
1	Detaillierte Registerbeschreibung (Protokoll)	X	X	X	X	X
Standardisierung						
2	Exakte Definition / Operationalisierung von Expositionen, klin. Ereignissen, Endpunkten und Confoundern	X	X	X	X	X
3	Aktueller Datenplan / Kodierhandbuch	X		X	X	X
4	Verwendung von Standard-Klassifikationen (z. B. ICD-10) und – Terminologien (z. B. MedDRA)	X	X	X	X	X
5	Verwendung von validierten Standard-Erhebungsinstrumenten (Fragebögen, Skalen, Tests)	X		X		
6	Schulungen zur Datenerhebung und – erfassung	X	X	X	X	X
7	Umsetzung eines konsentierten krankheitsspezifischen Kerndatensatzes („core data set“)	X	X		X	
8	Verwendung exakter Datumsangaben zum Patienten (z. B. Geburt, Tod, Schwangerschaft)		X			
9	Verwendung exakter Datumsangaben zur Erkrankung (z. B. definitive Diagnose, klinisch relevante Ereignisse)		X			X
10	Verwendung exakter Datumsangaben zu wichtigen Untersuchungen		X			X

Tabelle 14: Abgleich von Qualitätskriterien für Register und Registerstudien in der Literatur (mehrseitige Tabelle)

Nr.	Qualitätskriterium	DNVF e. V. 2010/19 ^a	EMA Discussion Paper 2018	AHRQ- Handbuch 2014	REQuEST EUnetHTA 2019 ^b	CTTI Recommendations 2017
11	Verwendung exakter Datumsangaben zu Behandlungen / Interventionen (z. B. bei Arzneimitteln Start-/Stop-Datum, Dosis, Dosisänderungen)		X	X		X
Erreichung des Rekrutierungsziels / Stichprobengewinnung						
12	Klar definierte Ein- und Ausschlusskriterien für Registerpatienten	X	X	X	X	X
13	Vollzähligkeit der Registerpatienten (Vollerhebung oder repräsentative Stichprobe)	X	X	X	X	X
14	Strategien zur Vermeidung ungewollter Selektionen beim Patienteneinschluss, um Repräsentativität zu erreichen	X	X	X	X	X
Validität der Datenerhebung						
15	Vollständigkeit der Daten je Erhebungszeitpunkt	X	X	X	X	X
16	Vollständigkeit der Erhebungszeitpunkte (loss-to-follow-up, drop-outs)	X	X	X	X	X
17	Richtigkeit der Daten (accuracy)	X	X	X	X	X
18	Datenkonsistenz über die Zeit		X			X
19	Source data verification (z. B. für 10 % zufällig ausgewählter Patienten je Erhebungszentrum)		X	X		X
20	Registermonitoring durch interne audits	X	X	X	X	X
21	Registermonitoring durch externe audits	X	X	X	X	X
22	QM-System (gfs. mit regelmäßiger Erhebung von Qualitätsindikatoren)	X	X	X	X	X

Tabelle 14: Abgleich von Qualitätskriterien für Register und Registerstudien in der Literatur (mehrseitige Tabelle)

Nr.	Qualitätskriterium	DNVF e. V. 2010/19 ^a	EMA Discussion Paper 2018	AHRQ- Handbuch 2014	REQuEST EUnetHTA 2019 ^b	CTTI Recommendations 2017
23	SOPs (Standard Operating Procedures) zur Datenerhebung	X	X	X		X
Übergeordnete Qualitätskriterien						
24	Transparenz eines Registers (u. a. Finanzierung, Entscheidungswege, Interessenkonflikte)	X	X	X	X	X
25	Wissenschaftliche Unabhängigkeit	X	X			
26	Sichere Finanzierung (für geplanten Erhebungszeitraum)		X		X	
27	Steering committee, Leitungsgremium	X	X	X	X	
28	Aktualität der Registerdokumente (z. B. Protokoll, Datenplan, SAP, Einwilligungserklärung etc.)				X	
29	Wahrung von Patientenrechten und Datenschutz, Berücksichtigung ethischer Aspekte	X	X	X	X	X
30	Timeliness (Aktualität / schnelle Verfügbarkeit / Pünktlichkeit der benötigten Ergebnisse)	X	X			X
31	Flexibilität und Anpassungsfähigkeit (z. B. zur Einbettung von Studien, für weitere Datenerhebung, bei veränderter Versorgungssituation)	X	X			X
32	Documentation trail - Dokumentation aller Prozess- und Definitionsänderungen im Register	X		X		
33	Audit trail - Dokumentation und Attribuierbarkeit aller Datentransaktionen			X		X
34	Verknüpfbarkeit mit anderen Datenquellen	X	X	X		X

Tabelle 14: Abgleich von Qualitätskriterien für Register und Registerstudien in der Literatur (mehrseitige Tabelle)

Nr.	Qualitätskriterium	DNVF e. V. 2010/19 ^a	EMA Discussion Paper 2018	AHRQ- Handbuch 2014	REQuEST EUnetHTA 2019 ^b	CTTI Recommendations 2017
Validität der statistischen Analysen und Berichte zu Registerstudien						
35	Öffentliche Registrierung der geplanten Registerstudie					
36	Erstellung eines Studienprotokolls und eines statistischen Analyseplans zur geplanten Registerstudie		X			
37	Präspezifizierung der Auswertungsmethodik im SAP	X	X	X		X
38	Darlegung des Umgangs mit fehlenden Werten	X	X	X	X	X
39	Adjudikations-Komitee für Schlüssel- Outcomes			X		X
40	Adjustierung von Vergleichs-ergebnissen für potenziell konfundierende Variablen und Berücksichtigung von Effekt modifizierenden Variablen	X	X	X	X	
41	Sensitivitätsanalysen (z. B. für unterschiedliche Falldefinitionen oder Confounder-Berücksichtigung)		X	X		
42	Auswertung / Kontrolle von Meldeeinrichtungseffekten	X				
43	Bericht über Maßnahmen zur Vermeidung von Verzerrung, detaillierte Diskussion möglicher Verzerrungen (z. B. Selektionsbias, „unmeasured confounding“)	X	X	X	X	X
44	Vollständiger Bericht der Ergebnisse zu allen erhobenen Variablen	X		X		

Tabelle 14: Abgleich von Qualitätskriterien für Register und Registerstudien in der Literatur (mehrseitige Tabelle)

Nr.	Qualitätskriterium	DNVF e. V. 2010/19 ^a	EMA Discussion Paper 2018	AHRQ- Handbuch 2014	REQuEST EUnetHTA 2019 ^b	CTTI Recommendations 2017
45	Publikation des Ergebnisberichts einschließlich Studienprotokoll und Analyseplan					
Weitere mögliche Kriterien aus regulatorischer Sicht						
46	Erhebung und Handling unerwünschter Ereignisse (UEs) gemäß regulativer Anforderungen		X	X		
<p>a: Die dieser Auswertung initial zugrundeliegende Version des Memorandums war die Publikation aus 2010, das zur Publikation aktuell eingereichte Update 2019 lag dank freundlicher Kooperation des DNVF e. V. ebenfalls vor Fertigstellung des Rapid Reports vor. Es haben sich für diese Vergleichsbetrachtung keine wesentlichen Änderungen zur Vorversion ergeben.</p> <p>b: Die dieser Auswertung zugrundeliegende Version der REQuEST Tools war die im Mai 2019 noch in einem Stellungnahmeverfahren von EUnetHTA befindliche Fassung. Es haben sich zur aktuell publizierten finalen Version im Hinblick auf diese Vergleichsbetrachtung keine wesentlichen Veränderungen ergeben.</p> <p>AHRQ: Agency for Healthcare Research and Quality; CTTI: Clinical Trial Transformation; DNVF: Deutsches Netzwerk Versorgungsforschung; EMA: Europäische Arzneimittel-Agentur; EUnetHTA: European Network for Health Technology Assessment; ICD: Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme; MedDRA: Medizinisches Wörterbuch für Aktivitäten im Rahmen der Arzneimittelzulassung; QM: Qualitätsmanagement; REQuEST: Registry Evaluation and Quality Standards Tool; SAP: statistischer Analyseplan; SOP: Standard Operating Procedure; UE: unerwünschtes Ereignis</p>						

Anhang D – Fragestellungen und Beschlüsse zu Orphan Drugs mit Marktzugang 2014 bis 2018

Tabelle 15: Orphan Drug Bewertungen des G-BA mit Marktzugang 2014 bis 2018 (mehrseitige Tabelle)

Wirkstoff	Fragestellung	Verfahrensstart	Beschlussdatum	Zusatznutzen quantifiziert	Beschluss des G-BA (URL)	Tragende Gründe zum Beschluss des G-BA (URL)
Macitentan	pulmonal arterielle Hypertonie (PAH)	01.02.2014	17.07.2014	ja	https://www.g-ba.de/downloads/39-261-2030/2014-07-17_AM-RL-XII_Macitentan_2014-02-01-D-096_BAnz.pdf	https://www.g-ba.de/downloads/40-268-2888/2014-07-17_AM-RL-XII_Macitentan_2014-02-01-D-096_TrG.pdf
Riociguat	chronisch thromboembolische pulmonale Hypertonie (CTEPH)	01.05.2014	16.10.2014	ja	https://www.g-ba.de/downloads/39-261-2076/2014-10-16_AM-RL-XII_Riociguat_2014-05-01-D-103_BAnz.pdf	https://www.g-ba.de/downloads/40-268-2978/2014-10-16_AM-RL-XII_Riociguat_2014-05-01-D-103_TrG.pdf
Riociguat	pulmonal arterielle Hypertonie (PAH)	01.05.2014	16.10.2014	ja	https://www.g-ba.de/downloads/39-261-2076/2014-10-16_AM-RL-XII_Riociguat_2014-05-01-D-103_BAnz.pdf	https://www.g-ba.de/downloads/40-268-2978/2014-10-16_AM-RL-XII_Riociguat_2014-05-01-D-103_TrG.pdf
Cholsäure	Angeborene Störungen der primären Gallensäuresynthese	15.05.2014	06.11.2014	nein	https://www.g-ba.de/downloads/39-261-2088/2014-11-06_AM-RL-XII_Cholsaeure_2014-05-15-D-105_BAnz.pdf	https://www.g-ba.de/downloads/40-268-2995/2014-11-06_AM-RL-XII_Cholsaeure_2014-05-15-D-105_TrG.pdf
Elosulfase alfa	Mucopolysaccharidose vom Typ IVA	01.06.2014	20.11.2014	ja	https://www.g-ba.de/downloads/39-261-2100/2014-11-20_AM-RL-XII_Elosulfase%20alfa_2014-06-01-D-114_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3008/2014-11-20_AM-RL-XII_Elosulfase%20alfa_2014-06-01-D-114_TrG.pdf
Siltuximab	Multizentrische Castleman-Krankheit	15.06.2014	04.12.2014	nein	https://www.g-ba.de/downloads/39-261-2118/2014-12-04_AM-RL-XII_Siltuximab_2014-06-15-D-119_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3021/2014-12-04_AM-RL-XII_Siltuximab_2014-06-15-D-119_TrG.pdf
Cabozantinib	Schilddrüsennoplasien	01.08.2014	22.01.2015	ja	https://www.g-ba.de/downloads/39-261-2147/2015-01-22_AM-RL-XII_Cabozantinib_2014-08-01-D-121_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3084/2015-01-22_AM-RL-XII_Cabozantinib_2014-08-01-D-121_TrG.pdf

Tabelle 15: Orphan Drug Bewertungen des G-BA mit Marktzugang 2014 bis 2018 (mehrseitige Tabelle)

Wirkstoff	Fragestellung	Verfahrens-start	Beschluss-datum	Zusatznutzen quantifiziert	Beschluss des G-BA (URL)	Tragende Gründe zum Beschluss des G-BA (URL)
Obinutuzumab	CLL	15.08.2014	05.02.2015	nein	https://www.g-ba.de/downloads/39-261-2167/2015-02-05_AM-RL-XII_Obinutuzumab_2014-08-15-D-120_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3098/2015-02-05_AM-RL-XII_Obinutuzumab_2014-08-15-D-120_TrG.pdf
Ivacaftor	zystische Fibrose (Gating-Mutationen im CFTR-Gen)	01.09.2014	19.02.2015	ja	https://www.g-ba.de/downloads/39-261-2178/2015-02-19_AM-RL-XII_Ivacaftor-nAWG_2014-09-01-D-133_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3114/2015-02-19_AM-RL-XII_Ivacaftor-nAWG_2014-09-01-D-133_TrG.pdf
Teduglutid	Malabsorptions-syndrome	01.09.2014	19.02.2015	ja	https://www.g-ba.de/downloads/39-261-2182/2015-02-19_AM-RL-XII_Teduglutid_2014-09-01-D-130_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3118/2015-02-19_AM-RL-XII_Teduglutid_2014-09-01-D-130_TrG.pdf
Alipogentiparvovec	Hyperlipoproteinämie Typ I	01.11.2014	21.05.2015	nein	https://www.g-ba.de/downloads/39-261-2250/2015-05-21_AM-RL-XII_Alipogentiparvovec_2014-11-01-D-138_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3224/2015-05-21_AM-RL-XII_Alipogentiparvovec_2014-11-01-D-138_TrG.pdf
Ibrutinib	CLL (mindestens 1 vorangehende Therapie)	01.11.2014	16.04.2015	nein	https://www.g-ba.de/downloads/39-261-2229/2015-04-16_AM-RL-XII_Ibrutinib_2014-11-01-D-141_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3187/2015-04-16_AM-RL-XII_Ibrutinib_2014-11-01-D-141_TrG.pdf
Ibrutinib	CLL (17p-Mutation oder TP53-Mutation, Erstlinie)	01.11.2014	16.04.2015	nein	https://www.g-ba.de/downloads/39-261-2229/2015-04-16_AM-RL-XII_Ibrutinib_2014-11-01-D-141_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3187/2015-04-16_AM-RL-XII_Ibrutinib_2014-11-01-D-141_TrG.pdf
Ibrutinib	Mantelzell-Lymphom	01.11.2014	16.04.2015	nein	https://www.g-ba.de/downloads/39-261-2229/2015-04-16_AM-RL-XII_Ibrutinib_2014-11-01-D-141_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3187/2015-04-16_AM-RL-XII_Ibrutinib_2014-11-01-D-141_TrG.pdf
Ataluren	Duchenne-Muskeldystrophie	01.12.2014	21.05.2015	ja	https://www.g-ba.de/downloads/39-261-2252/2015-05-21_AM-RL-XII_Ataluren_2012-01-D-149_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3226/2015-05-21_AM-RL-XII_Ataluren_2012-01-D-149_TrG.pdf

Tabelle 15: Orphan Drug Bewertungen des G-BA mit Marktzugang 2014 bis 2018 (mehrseitige Tabelle)

Wirkstoff	Fragestellung	Verfahrens-start	Beschluss-datum	Zusatznutzen quantifiziert	Beschluss des G-BA (URL)	Tragende Gründe zum Beschluss des G-BA (URL)
Pasireotid	Akromegalie	01.01.2015	18.06.2015	ja	https://www.g-ba.de/downloads/39-261-2263/2015-06-18_AM-RL-XII_Pasireotid_nAWG_2015-01-01-D-148_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3250/2015-06-18_AM-RL-XII_Pasireotid_nAWG_2015-01-01-D-148_TrG.pdf
Ramucirumab	Magenkrebs (Kombination mit Paclitaxel)	01.02.2015	16.07.2015	ja	https://www.g-ba.de/downloads/39-261-2292/2015-07-16_AM-RL-XII_Ramucirumab_2015-02-01-D-150_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3275/2015-07-16_AM-RL-XII_Ramucirumab_2015-02-01-D-150_TrG.pdf
Ramucirumab	Magenkrebs (Monotherapie)	01.02.2015	16.07.2015	ja	https://www.g-ba.de/downloads/39-261-2292/2015-07-16_AM-RL-XII_Ramucirumab_2015-02-01-D-150_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3275/2015-07-16_AM-RL-XII_Ramucirumab_2015-02-01-D-150_TrG.pdf
Nintedanib	idiopathische pulmonale Fibrose	15.03.2015	03.09.2015	ja	https://www.g-ba.de/downloads/39-261-2322/2015-09-03_AM-RL-XII_Nintedanib_nAWG_2015-03-15-D-156_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3339/2015-09-03_AM-RL-XII_Nintedanib_nAWG_2015-03-15-D-156_TrG.pdf
Eliglustat	Morbus Gaucher Typ 1	01.04.2015	01.10.2015	nein	https://www.g-ba.de/downloads/39-261-2350/2015-10-01_AM-RL-XII_Eliglustat_2015-04-01-D-159_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3381/2015-10-01_AM-RL-XII_Eliglustat_2015-04-01-D-159_TrG.pdf
Olaparib	Ovarialneoplasie	01.06.2015	27.11.2015	nein	https://www.g-ba.de/downloads/39-261-2383/2015-11-27_AM-RL-XII_Olaparib_2015-06-01-D-166_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3451/2015-11-27_AM-RL-XII_Olaparib_2015-06-01-D-166_TrG.pdf
Lenvatinib	Schilddrüsene-neoplasien	01.07.2015	17.12.2015	nein	https://www.g-ba.de/downloads/39-261-2413/2015-12-17_AM-RL-XII_Lenvatinib_2015-07-01-D-164_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3495/2015-12-17_AM-RL-XII_Lenvatinib_2015-07-01-D-164_TrG.pdf
Asfotase alfa	Hypophosphatasie (Patienten ≤ 5 Jahre)	01.10.2015	17.03.2016	nein	https://www.g-ba.de/downloads/39-261-2526/2016-03-17_AM-RL-XII_Asfotase-alfa_D-188_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3662/2016-03-17_AM-RL-XII_Asfotase-alfa_D-188_TrG.pdf

Tabelle 15: Orphan Drug Bewertungen des G-BA mit Marktzugang 2014 bis 2018 (mehrseitige Tabelle)

Wirkstoff	Fragestellung	Verfahrens-start	Beschluss-datum	Zusatznutzen quantifiziert	Beschluss des G-BA (URL)	Tragende Gründe zum Beschluss des G-BA (URL)
Asfotase alfa	Hypophosphatasie (Patienten > 5 Jahre)	01.10.2015	17.03.2016	nein	https://www.g-ba.de/downloads/39-261-2526/2016-03-17_AM-RL-XII_Asfotase-alfa_D-188_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3662/2016-03-17_AM-RL-XII_Asfotase-alfa_D-188_TrG.pdf
Idebenon	Lebersche hereditäre Optikusneuropathie	01.10.2015	17.03.2016	nein	https://www.g-ba.de/downloads/39-261-2527/2016-03-17_AM-RL-XII_Idebenon_D-191_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3663/2016-03-17_AM-RL-XII_Idebenon_D-191_TrG.pdf
Panobinostat	multiples Myelom	01.10.2015	17.03.2016	nein	https://www.g-ba.de/downloads/39-261-2529/2016-03-17_AM-RL-XII_Panobinostat_D-180_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3666/2016-03-17_AM-RL-XII_Panobinostat_D-180_TrG.pdf
Sebelipase alfa	LAL-Mangel (nicht bereits im Säuglingsalter rasch fortschreitend)	01.10.2015	17.03.2016	nein	https://www.g-ba.de/downloads/39-261-2528/2016-03-17_AM-RL-XII_Sebelipase-alfa_D-187_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3664/2016-03-17_AM-RL-XII_Sebelipase-alfa_D-187_TrG.pdf
Sebelipase alfa	LAL-Mangel (bereits im Säuglingsalter rasch fortschreitend)	01.10.2015	17.03.2016	nein	https://www.g-ba.de/downloads/39-261-2528/2016-03-17_AM-RL-XII_Sebelipase-alfa_D-187_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3664/2016-03-17_AM-RL-XII_Sebelipase-alfa_D-187_TrG.pdf
Isavuconazol	Aspergillose	15.11.2015	04.05.2016	nein	https://www.g-ba.de/downloads/39-261-2574/2016-05-04_AM-RL-XII_Isavuconazol_2015-11-15_D-192_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3754/2016-05-04_AM-RL-XII_Isavuconazol_2015-11-15_D-192_TrG.pdf
Isavuconazol	Mukormyrkose	15.11.2015	04.05.2016	nein	https://www.g-ba.de/downloads/39-261-2574/2016-05-04_AM-RL-XII_Isavuconazol_2015-11-15_D-192_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3754/2016-05-04_AM-RL-XII_Isavuconazol_2015-11-15_D-192_TrG.pdf
Blinatumomab	ALL	15.12.2015	02.06.2016	nein	https://www.g-ba.de/downloads/39-261-2605/2016-06-02_AM-RL-XII_Blinatumomab_D-201_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3801/2016-06-02_AM-RL-XII_Blinatumomab_D-201_TrG.pdf

Tabelle 15: Orphan Drug Bewertungen des G-BA mit Marktzugang 2014 bis 2018 (mehrseitige Tabelle)

Wirkstoff	Fragestellung	Verfahrens-start	Beschluss-datum	Zusatznutzen quantifiziert	Beschluss des G-BA (URL)	Tragende Gründe zum Beschluss des G-BA (URL)
Carfilzomib	multiples Myelom	15.12.2015	02.06.2016	nein	https://www.g-ba.de/downloads/39-261-2606/2016-06-02_AM-RL-XII_Carfilzomib_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3802/2016-06-02_AM-RL-XII_Carfilzomib_TrG.pdf
Ivacaftor	zystische Fibrose (Patienten ab 18 Jahren mit R117H-Mutation)	15.12.2015	02.06.2016	ja	https://www.g-ba.de/downloads/39-261-2601/2016-06-02_AM-RL-XII_Ivacaftor_nAWG_2015-12-15-D-200_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3797/2016-06-02_AM-RL-XII_Ivacaftor_nAWG_2015-12-15-D-200_TrG.pdf
Ivacaftor	zystische Fibrose (Kinder 2 bis 5 Jahre mit Gating-Mutation)	15.12.2015	02.06.2016	nein	https://www.g-ba.de/downloads/39-261-2601/2016-06-02_AM-RL-XII_Ivacaftor_nAWG_2015-12-15-D-200_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3797/2016-06-02_AM-RL-XII_Ivacaftor_nAWG_2015-12-15-D-200_TrG.pdf
Afamelanotid	erythropoetische Protoporphyrrie	15.02.2016	04.08.2016	nein	https://www.g-ba.de/downloads/39-261-2674/2016-08-04_AM-RL-XII_Afamelanotid_D-218_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3923/2016-08-04_AM-RL-XII_Afamelanotid_D-218_TrG.pdf
Albutrepenonacog alfa	Hämophilie B	01.06.2016	01.12.2016	nein	https://www.g-ba.de/downloads/39-261-2776/2016-12-01_AM-RL-XII_Albutrepenonacog_D-227_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4075/2016-12-01_AM-RL-XII_Albutrepenonacog_D-227_TrG.pdf
Daratumumab	multiples Myelom	01.06.2016	01.12.2016	nein	https://www.g-ba.de/downloads/39-261-2772/2016-12-01_AM-RL-XII_Daratumumab_D-238_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4071/2016-12-01_AM-RL-XII_Daratumumab_D-238_TrG.pdf
Migalastat	Morbus Fabry	01.06.2016	01.12.2016	nein	https://www.g-ba.de/downloads/39-261-2777/2016-12-01_AM-RL-XII_Migalastat_D-225_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4076/2016-12-01_AM-RL-XII_Migalastat_D-225_TrG.pdf
Eftrenocog alfa	Hämophilie B	15.06.2016	15.12.2016	nein	https://www.g-ba.de/downloads/39-261-2804/2016-12-15_AM-RL-XII_Eftrenonacog-alfa_D-233_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4108/2016-12-15_AM-RL-XII_Eftrenonacog-alfa_D-233_TrG.pdf

Tabelle 15: Orphan Drug Bewertungen des G-BA mit Marktzugang 2014 bis 2018 (mehrseitige Tabelle)

Wirkstoff	Fragestellung	Verfahrens-start	Beschluss-datum	Zusatznutzen quantifiziert	Beschluss des G-BA (URL)	Tragende Gründe zum Beschluss des G-BA (URL)
Obinutuzumab	follikuläres non-Hodgkin Lymphom	01.07.2016	15.12.2016	nein	https://www.g-ba.de/downloads/39-261-2801/2016-12-15_AM-RL-XII_Obinutuzumab_D-229_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4105/2016-12-15_AM-RL-XII_Obinutuzumab_D-229_TrG.pdf
Brentuximab Vedotin	Hodgkin-Lymphom	01.08.2016	19.01.2017	nein	https://www.g-ba.de/downloads/39-261-2835/2017-01-19_AM-RL-XII_Brentuximab-Vedotin_D-253_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4156/2017-01-19_AM-RL-XII_Brentuximab-Vedotin_D-253_TrG.pdf
Carfilzomib	multiples Myelom	01.08.2016	19.01.2017	ja	https://www.g-ba.de/downloads/39-261-2836/2017-01-19_AM-RL-XII_Carfilzomib_nAWG-Kombi-Dexamethason_D-255_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4157/2017-01-19_AM-RL-XII_Carfilzomib_nAWG-Kombi-Dexamethason_D-255_TrG.pdf
Pitolisant	Narkolepsie	01.08.2016	19.01.2017	nein	https://www.g-ba.de/downloads/39-261-2838/2017-01-19_AM-RL-XII_Pitolisant_D-250_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4159/2017-01-19_AM-RL-XII_Pitolisant_D-250_TrG.pdf
Tasimelton	Schlafstörungen	01.08.2016	19.01.2017	nein	https://www.g-ba.de/downloads/39-261-2837/2017-01-19_AM-RL-XII_Tasimelton_D-242_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4158/2017-01-19_AM-RL-XII_Tasimelton_D-242_TrG.pdf
Teduglutid	Malabsorptions-syndrom	01.08.2016	19.01.2017	nein	https://www.g-ba.de/downloads/39-261-2839/2017-01-19_AM-RL-XII_Teduglutid_D-254_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4160/2017-01-19_AM-RL-XII_Teduglutid_D-254_TrG.pdf
Olaratumab	Sarkom	01.12.2016	18.05.2017	ja	https://www.g-ba.de/downloads/39-261-2949/2017-05-18_AM-RL-XII_Olaratumab_D-265_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4391/2017-05-18_AM-RL-XII_Olaratumab_D-265_TrG.pdf
Venetoclax	CLL (17p-Deletion oder TP53-Mutation)	01.01.2017	15.06.2017	nein	https://www.g-ba.de/downloads/39-261-2970/2017-06-15_AM-RL-XII_Venetoclax_D-266_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4425/2017-06-15_AM-RL-XII_Venetoclax_D-266_TrG.pdf
Venetoclax	CLL (ohne 17p-Deletion oder TP53-Mutation)	01.01.2017	15.06.2017	nein	https://www.g-ba.de/downloads/39-261-2970/2017-06-15_AM-RL-XII_Venetoclax_D-266_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4425/2017-06-15_AM-RL-XII_Venetoclax_D-266_TrG.pdf

Tabelle 15: Orphan Drug Bewertungen des G-BA mit Marktzugang 2014 bis 2018 (mehrseitige Tabelle)

Wirkstoff	Fragestellung	Verfahrens-start	Beschluss-datum	Zusatznutzen quantifiziert	Beschluss des G-BA (URL)	Tragende Gründe zum Beschluss des G-BA (URL)
Ixazomib	multiples Myelom	15.01.2017	06.07.2017	nein	https://www.g-ba.de/downloads/39-261-2994/2017-07-06_AM-RL-XII_Ixazomib_D-272_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4454/2017-07-06_AM-RL-XII_Ixazomib_D-272_TrG.pdf
Obeticholsäure	biliäre Leberzirrhose	15.01.2017	06.07.2017	nein	https://www.g-ba.de/downloads/39-261-2995/2017-07-06_AM-RL-XII_Obeticholsaeure_D-269_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4455/2017-07-06_AM-RL-XII_Obeticholsaeure_D-269_TrG.pdf
Cerliponase-alfa	Neuronale Ceroid-Lipofuszinose	01.07.2017	21.12.2017	nein	https://www.g-ba.de/downloads/39-261-3168/2017-12-21_AM-RL-XII_Cerliponase-alfa_D-298_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4721/2017-12-21_AM-RL-XII_Cerliponase-alfa_D-298_TrG.pdf
Nusinersen	spinale Muskelatrophie Typ 2	01.07.2017	21.12.2017	ja	https://www.g-ba.de/downloads/39-261-3169/2017-12-21_AM-RL-XII_Nusinersen_D-294_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4722/2017-12-21_AM-RL-XII_Nusinersen_D-294_TrG.pdf
Nusinersen	spinale Muskelatrophie Typ 1	01.07.2017	21.12.2017	ja	https://www.g-ba.de/downloads/39-261-3169/2017-12-21_AM-RL-XII_Nusinersen_D-294_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4722/2017-12-21_AM-RL-XII_Nusinersen_D-294_TrG.pdf
Nusinersen	spinale Muskelatrophie Typ 3	01.07.2017	21.12.2017	nein	https://www.g-ba.de/downloads/39-261-3169/2017-12-21_AM-RL-XII_Nusinersen_D-294_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4722/2017-12-21_AM-RL-XII_Nusinersen_D-294_TrG.pdf
Nusinersen	spinale Muskelatrophie Typ 4	01.07.2017	21.12.2017	nein	https://www.g-ba.de/downloads/39-261-3169/2017-12-21_AM-RL-XII_Nusinersen_D-294_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4722/2017-12-21_AM-RL-XII_Nusinersen_D-294_TrG.pdf
Inotuzumab Ozogamicin	ALL	15.07.2017	18.01.2018	ja	https://www.g-ba.de/downloads/39-261-3192/2018-01-18_AM-RL-XII_Inotuzumab-Ozogamicin_D-297_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4763/2018-01-18_AM-RL-XII_Inotuzumab-Ozogamicin_D-297_TrG.pdf
Avelumab	Merkelzellkarzinom (nicht vorbehandelt)	01.10.2017	16.03.2018	nein	https://www.g-ba.de/downloads/39-261-3252/2018-03-16_AM-RL-XII_Avelumab_D-308_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4873/2018-03-16_AM-RL-XII_Avelumab_D-308_TrG.pdf

Tabelle 15: Orphan Drug Bewertungen des G-BA mit Marktzugang 2014 bis 2018 (mehrseitige Tabelle)

Wirkstoff	Fragestellung	Verfahrens-start	Beschluss-datum	Zusatznutzen quantifiziert	Beschluss des G-BA (URL)	Tragende Gründe zum Beschluss des G-BA (URL)
Avelumab	Merkelzellkarzinom (vorbehandelt)	01.10.2017	16.03.2018	nein	https://www.g-ba.de/downloads/39-261-3252/2018-03-16_AM-RL-XII_Avelumab_D-308_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4873/2018-03-16_AM-RL-XII_Avelumab_D-308_TrG.pdf
Midostaurin	AML	15.10.2017	05.04.2018	ja	https://www.g-ba.de/downloads/39-261-3277/2018-04-05_AM-RL-XII_Midostaurin_AML_D-319_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4916/2018-04-05_AM-RL-XII_Midostaurin_AML_D-319_TrG.pdf
Midostaurin	Mastozytose	15.10.2017	05.04.2018	nein	https://www.g-ba.de/downloads/39-261-3279/2018-04-05_AM-RL-XII_Midostaurin_ASM_D-319_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4918/2018-04-05_AM-RL-XII_Midostaurin_ASM_D-319_TrG.pdf
Obinutuzumab	Follikuläres Lymphom	15.10.2017	05.04.2018	nein	https://www.g-ba.de/downloads/39-261-3275/2018-04-05_AM-RL-XII_Obinutuzumab_nAWG_D-305_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4914/2018-04-05_AM-RL-XII_Obinutuzumab_nAWG_D-305_TrG.pdf
Telotristat-ethyl	Karzinoid-Syndrom	15.10.2017	05.04.2018	nein	https://www.g-ba.de/downloads/39-261-3276/2018-04-05_AM-RL-XII_Telotristatethyl_D-318_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4915/2018-04-05_AM-RL-XII_Telotristatethyl_D-318_TrG.pdf
Cenegermin	neurotrophe Keratitis	15.11.2017	03.05.2018	nein	https://www.g-ba.de/downloads/39-261-3305/2018-05-03_AM-RL-XII_Cenegermin-D-329_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4967/2018-05-03_AM-RL-XII_Cenegermin-D-329_TrG.pdf
Niraparib	Ovarialkarzinom	15.12.2017	07.06.2018	nein	https://www.g-ba.de/downloads/39-261-3361/2018-06-07_AM-RL-XII_Niraparib_D-331_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5050/2018-06-07_AM-RL-XII_Niraparib_D-331_TrG.pdf
Allogene, genetisch modifizierte T-Zellen	Graft-versus-Host Disease	15.01.2018	05.07.2018	nein	https://www.g-ba.de/downloads/39-261-3396/2018-07-05_AM-RL-XII_allogene_genetisch_modifizierte_T-Zellen_D-333_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5102/2018-07-05_AM-RL-XII_allogene_genetisch_modifizierte_T-Zellen_D-333_TrG.pdf

Tabelle 15: Orphan Drug Bewertungen des G-BA mit Marktzugang 2014 bis 2018 (mehrseitige Tabelle)

Wirkstoff	Fragestellung	Verfahrens-start	Beschluss-datum	Zusatznutzen quantifiziert	Beschluss des G-BA (URL)	Tragende Gründe zum Beschluss des G-BA (URL)
Brentuximab Vedotin	T-Zell-Lymphom	15.01.2018	05.07.2018	ja	https://www.g-ba.de/downloads/39-261-3397/2018-07-05_AM-RL-XII_Brentuximab-Vedotin_D-340_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5103/2018-07-05_AM-RL-XII_Brentuximab-Vedotin_D-340_TrG.pdf
Letermovir	Prophylaxe einer CMV-Reaktivierung	15.02.2018	02.08.2018	nein	https://www.g-ba.de/downloads/39-261-3431/2018-08-02_AM-RL-XII_Letermovir_D-342_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5164/2018-08-02_AM-RL-XII_Letermovir_D-342_TrG.pdf
Glycerolphenylbutyrat	Harnstoffzyklusstörung	01.03.2018	16.08.2018	nein	https://www.g-ba.de/downloads/39-261-3449/2018-08-16_AM-RL-XII_Glycerolphenylbutyrat_D-303_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5184/2018-08-16_AM-RL-XII_Glycerolphenylbutyrat_D-303_TrG.pdf
Burosumab	Hypophosphatämie	15.04.2018	04.10.2018	nein	https://www.g-ba.de/downloads/39-261-3515/2018-10-04_AM-RL-XII_Burosumab_D-349_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5295/2018-10-04_AM-RL-XII_Burosumab_D-349_TrG.pdf
Darvadstrocel	Morbus Crohn	01.06.2018	22.11.2018	nein	https://www.g-ba.de/downloads/39-261-3571/2018-11-22_AM-RL-XII_Darvadstrocel_D-366_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5411/2018-11-22_AM-RL-XII_Darvadstrocel_D-366_TrG.pdf
Velmanase alfa	α-Mannosidose	01.07.2018	20.12.2018	nein	https://www.g-ba.de/downloads/39-261-3617/2018-12-20_AM-RL-XII_Velmanase-alfa_D-365_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5500/2018-12-20_AM-RL-XII_Velmanase-alfa_D-365_TrG.pdf
Gemtuzumab Ozogamicin	AML	01.09.2018	21.02.2019	nein	https://www.g-ba.de/downloads/39-261-3679/2019-02-21_AM-RL-XII_Gemtuzumab-Ozogamicin_D-380_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5585/2019-02-21_AM-RL-XII_Gemtuzumab-Ozogamicin_D-380_TrG.pdf
Tisagenlecleucel	diffus großzelliges B-Zell-Lymphom	15.09.2018	07.03.2019	nein	https://www.g-ba.de/downloads/39-261-3700/2019-03-07_AM-RL-XII_Tisagenlecleucel-DLBCL_D-375_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5622/2019-03-07_AM-RL-XII_Tisagenlecleucel-DLBCL_D-375_TrG.pdf

Tabelle 15: Orphan Drug Bewertungen des G-BA mit Marktzugang 2014 bis 2018 (mehrseitige Tabelle)

Wirkstoff	Fragestellung	Verfahrens-start	Beschluss-datum	Zusatznutzen quantifiziert	Beschluss des G-BA (URL)	Tragende Gründe zum Beschluss des G-BA (URL)
Tisagenlecleucel	akute B-Zell-Leukämie	15.09.2018	07.03.2019	nein	https://www.g-ba.de/downloads/39-261-3701/2019-03-07_AM-RL-XII_Tisagenlecleucel-ALL_D-376_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5623/2019-03-07_AM-RL-XII_Tisagenlecleucel-ALL_D-376_TrG.pdf
Caplacizumab	thrombotisch-thrombozytopenische Purpura	01.10.2018	22.03.2019	nein	https://www.g-ba.de/downloads/39-261-3726/2019-03-22_AM-RL-XII_Caplacizumab_D-387_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5655/2019-03-22_AM-RL-XII_Caplacizumab_D-387_TrG.pdf
Daunorubicin / Cytarabin	AML	01.10.2018	22.03.2019	ja	https://www.g-ba.de/downloads/39-261-3716/2019-03-22_AM-RL-XII_Daunorubicin-Cytarabin_D-382_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5645/2019-03-22_AM-RL-XII_Daunorubicin-Cytarabin_D-382_TrG.pdf
Inotersen	Amyloidose	01.10.2018	22.03.2019	nein	https://www.g-ba.de/downloads/39-261-3717/2019-03-22_AM-RL-XII_Inotersen_D-381_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5646/2019-03-22_AM-RL-XII_Inotersen_D-381_TrG.pdf
Metreleptin	Generalisierte Lipodystrophie	01.10.2018	22.03.2019	nein	https://www.g-ba.de/downloads/39-261-3712/2019-03-22_AM-RL-XII_Metreleptin_D-385_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5641/2019-03-22_AM-RL-XII_Metreleptin_D-385_TrG.pdf
Metreleptin	Partielle Lipodystrophie	01.10.2018	22.03.2019	nein	https://www.g-ba.de/downloads/39-261-3712/2019-03-22_AM-RL-XII_Metreleptin_D-385_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5641/2019-03-22_AM-RL-XII_Metreleptin_D-385_TrG.pdf
Patisiran	Amyloidose	01.10.2018	22.03.2019	ja	https://www.g-ba.de/downloads/39-261-3718/2019-03-22_AM-RL-XII_Patisiran_D-391_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5647/2019-03-22_AM-RL-XII_Patisiran_D-391_TrG.pdf
Vestronidase alfa	Sly-Syndrom	01.10.2018	22.03.2019	nein	https://www.g-ba.de/downloads/39-261-3723/2019-03-22_AM-RL-XII_Vestronidase%20alfa_D-392_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5730/2019-03-22_AM-RL-XII_Vestronidase%20alfa_D-392_TrG.pdf

Tabelle 15: Orphan Drug Bewertungen des G-BA mit Marktzugang 2014 bis 2018 (mehrseitige Tabelle)

Wirkstoff	Fragestellung	Verfahrens-start	Beschluss-datum	Zusatznutzen quantifiziert	Beschluss des G-BA (URL)	Tragende Gründe zum Beschluss des G-BA (URL)
Axicabtagen-Ciloleucel	diffus großzelliges B-Zell-Lymphom	01.11.2018	02.05.2019	nein	https://www.g-ba.de/downloads/39-261-3771/2019-05-02_AM-RL-XII_Axicabtagen-Ciloleucel_D-406_D-416_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5741/2019-05-02_AM-RL-XII_Axicabtagen-Ciloleucel_D-406_D-416_TrG.pdf
Axicabtagen-Ciloleucel	primär mediastinale B-Zell-Lymphom	01.11.2018	02.05.2019	nein	https://www.g-ba.de/downloads/39-261-3772/2019-05-02_AM-RL-XII_Axicabtagen-Ciloleucel_D-406_D-416_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5742/2019-05-02_AM-RL-XII_Axicabtagen-Ciloleucel_D-406_D-416_TrG.pdf
Tezacaftor / Ivacaftor	zystische Fibrose (ab 12 Jahre, F508del-homozygot)	01.12.2018	16.05.2019	ja	https://www.g-ba.de/downloads/39-261-3785/2019-05-16_AM-RL-XII_Tezacaftor-Ivacaftor_D-408_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5745/2019-05-16_AM-RL-XII_Tezacaftor-Ivacaftor_D-408_TrG.pdf
Tezacaftor / Ivacaftor	zystische Fibrose (ab 12 Jahre, F508del-heterozygot, weitere Mutationen im CFTR-Gen)	01.12.2018	16.05.2019	ja	https://www.g-ba.de/downloads/39-261-3785/2019-05-16_AM-RL-XII_Tezacaftor-Ivacaftor_D-408_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5745/2019-05-16_AM-RL-XII_Tezacaftor-Ivacaftor_D-408_TrG.pdf
ALL: akute lymphoblastische Leukämie; AML: akute myeloische Leukämie; CFTR: Cystic Fibrosis Transmembrane Conductance Regulator; CLL: chronisch lymphatische Leukämie; CMV: Cytomegalovirus; CTEPH: chronisch thromboembolische pulmonale Hypertonie; G-BA: Gemeinsamer Bundesausschuss; LAL: lysosomale saure Lipase; PAH: pulmonal arterielle Hypertonie; URL: Uniform Ressource Locator						