

IQWiG-Berichte – Nr. 241

Bewertung und Auswertung von Studien bei seltenen Erkrankungen

Rapid Report

Auftrag: MB13-01
Version: 1.0
Stand: 05.09.2014

Impressum

Herausgeber:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema:

Bewertung und Auswertung von Studien bei seltenen Erkrankungen

Auftraggeber:

Bundesministerium für Gesundheit (BMG)

Datum des Auftrags:

10.12.2013

Interne Auftragsnummer:

MB13-01

Anschrift des Herausgebers:

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

Tel.: +49 (0)221 – 35685-0

Fax: +49 (0)221 – 35685-1

E-Mail: berichte@iqwig.de

Internet: www.iqwig.de

ISSN: 1864-2500

Externes Review des Rapid Reports

- Norbert Benda (Bundesinstitut für Arzneimittel und Medizinprodukte)

Dieser Rapid Report wurde einem externen Review unterzogen. Das IQWiG dankt dem externen Reviewer für seine Mitarbeit am Projekt. Der Reviewer war jedoch nicht in die Erstellung des Rapid Reports eingebunden. Daher geben einzelne Passagen und Schlussfolgerungen des Rapid Reports nicht notwendigerweise die Meinung des Reviewers wieder. Die Selbstangaben des externen Reviewers zu potenziellen Interessenkonflikten sind in Anhang B dargestellt. Es wurden keine Interessenkonflikte festgestellt, die die fachliche Unabhängigkeit im Hinblick auf eine Bearbeitung des vorliegenden Auftrags gefährden.

Für die Inhalte des Berichts ist allein das IQWiG verantwortlich.

Mitarbeiter des IQWiG¹:

- Ulrich Grouven
- Ralf Bender
- Ulrike Lampert
- Stefan Lange
- Stefan Sauerland
- Ulrich Siering
- Volker Vervölgyi

¹ Aufgrund gesetzlicher Datenschutzbestimmungen haben Mitarbeiter das Recht, ihrer Namensnennung nicht zuzustimmen.

Inhaltsverzeichnis

	Seite
Tabellenverzeichnis	v
Abbildungsverzeichnis	vi
Abkürzungsverzeichnis	vii
Kurzfassung	viii
1 Hintergrund	1
2 Ziele der Untersuchung	3
3 Projektbearbeitung	4
3.1 Zeitlicher Verlauf des Projekts	4
4 Methoden	5
4.1 Methodische Expertise	5
4.1.1 Gegenwärtige Rahmenbedingungen und Problemstellung	5
4.1.2 Methodische Aspekte bei der Bewertung und Auswertung von Studien.....	5
4.2 Empirische Untersuchung zur Studiengrundlage für die Zulassung von Orphan Drugs in Europa	5
4.2.1 Informationsbeschaffung.....	5
4.2.2 Informationssynthese.....	6
4.2.3 Informationsanalyse	7
5 Ergebnisse	8
5.1 Methodische Expertise	8
5.1.1 Grundlegende methodische Aspekte bei der Bewertung und Auswertung von Studien.....	8
5.1.2 Barrieren gegen randomisierte Studien	11
5.1.3 Situation bei seltenen Erkrankungen	13
5.1.4 Design-Vorschläge für randomisierte Therapiestudien bei seltenen Erkrankungen	14
5.1.4.1 Varianzreduzierende Verfahren und Designs	15
5.1.4.2 Sequenzielle und adaptive Designs	16
5.1.4.3 Bayes'sche Verfahren	17
5.1.4.4 Designs, bei denen alle Patienten (auch) die Prüfintervention erhalten	19
5.1.4.5 Sonstige Designs.....	21
5.1.5 Ergebnissicherheit nicht randomisierter Studien.....	22
5.1.6 Empfehlungen	24
5.1.6.1 Politische Vorgaben.....	25

5.1.6.2	Strukturen	25
5.1.6.3	Effizienz.....	26
5.1.6.4	Kompromisse bei der Aussagesicherheit.....	27
5.2	Ergebnisse der empirischen Untersuchung zur Studiengrundlage für die Zulassung von Orphan Drugs in Europa.....	28
5.2.1	Ergebnisse der Informationsbeschaffung und resultierender Publikationspool ...	28
5.2.2	Ergebnisse der Informationssynthese und Informationsanalyse	28
6	Diskussion.....	36
7	Fazit.....	40
8	Literatur	41
Anhang A – Extraktionstabellen zur Auswertung der Zulassungsstudien zu Orphan Drugs		48
Anhang B – Offenlegung potenzieller Interessenkonflikte des externen Reviewers.....		85

Tabellenverzeichnis

	Seite
Tabelle 1: Zusammenhang zwischen Einschränkung der Zulassung und Einbezug von RCT-Daten	29
Tabelle 2: Anzahl der Hauptstudien pro Arzneimittel	30
Tabelle 3: Studiencharakteristika (Arzneimittlebene).....	31
Tabelle 4: Studiencharakteristika der RCT (Studienebene).....	33
Tabelle 5: Verwendete statistische Methoden in den RCTs (Studienebene)	35
Tabelle 6: Orphan Drugs – Charakteristika.....	48
Tabelle 7: Methodische Aspekte der Zulassungsstudien zu Orphan Drugs (Studienebene)....	55
Tabelle 8: Übersicht der Zulassungsstudien zu Orphan Drugs (Arzneimittlebene).....	79

Abbildungsverzeichnis

	Seite
Abbildung 1: Anteil der Orphan Drugs, die auf Basis von RCTs zugelassen wurden, im zeitlichen Verlauf (N = 85)	29
Abbildung 2: Anteil der Zulassungen, bei denen patientenrelevante Endpunkte (primär oder sekundär) ausgewertet wurden, im zeitlichen Verlauf (N = 79)	32

Abkürzungsverzeichnis

Abkürzung	Bedeutung
ACHSE	Allianz Chronischer Seltener Erkrankungen
BfArM	Bundesinstitut für Arzneimittel und Medizinprodukte
BMBF	Bundesministerium für Bildung und Forschung
BMG	Bundesministerium für Gesundheit
EbM	evidenzbasierte Medizin
EMA	European Medicines Agency
EPAR	European Public Assessment Report (EMA)
EU	Europäische Union
FDA	Food and Drug Administration
G-BA	Gemeinsamer Bundesausschuss
INSERM	Institut national de la santé et de la recherche médicale
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	intention to treat
NAMSE	Nationales Aktionsbündnis für Menschen mit Seltene Erkrankungen
PICOS	Population, Intervention, Comparator, Outcome, Study design / Setting
RCT	randomized controlled trial (randomisierte kontrollierte Studie)
SGB	Sozialgesetzbuch

Kurzfassung

Das Bundesgesundheitsministerium (BMG) hat mit Schreiben vom 10.12.2013 das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) mit der Erstellung eines Rapid Reports zum Thema „Bewertung und Auswertung von Studien bei seltenen Erkrankungen“ beauftragt.

Fragestellung

Die Ziele des vorliegenden Rapid Reports sind

- die Erstellung einer Expertise zu methodischen Aspekten bei der Durchführung und Auswertung sowie zur Bewertung der Ergebnissicherheit von Studien bei seltenen Erkrankungen,
- die Beschreibung der Studiengrundlage für die Zulassung von Orphan Drugs in Europa.

Methoden

Methodische Expertise

In der methodischen Expertise wird die besondere Problematik bei der Durchführung und Bewertung von Studien zu seltenen Erkrankungen dargestellt. Dabei wird unterschieden zwischen seltenen ($\leq 5 / 10\,000$) und sehr seltenen Erkrankungen ($< 2 / 100\,000$ Einwohner). Es werden Kriterien der Ergebnis- und Aussagesicherheit von häufig bei seltenen Erkrankungen eingesetzten bzw. vorgeschlagenen Studiendesigns als Grundlage für Nutzenbewertungen dargestellt und diskutiert.

Empirische Untersuchung zur Studiengrundlage für die Zulassung von Orphan Drugs in Europa

Die Recherche für die empirische Untersuchung der Studiengrundlage für die Zulassung von Orphan Drugs in Europa erfolgte im Orphanet. Für die Extraktion der relevanten Daten zu den identifizierten Orphan Drugs wurden European Public Assessment Reports (EPARs) zu den Zulassungsstudien herangezogen. Relevante Charakteristika des Arzneimittels sowie der zugrunde liegenden Zulassungsstudien wurden extrahiert und mithilfe deskriptiver statistischer Verfahren ausgewertet.

Ergebnisse

Methodische Expertise

Grundlegende methodische Aspekte bei der Bewertung und Auswertung von Studien

Nach internationaler Übereinkunft ist die wissenschaftliche Grundlage der evidenzbasierten Medizin (EbM), im Hinblick auf eine klar definierte Fragestellung (gemäß PICOS [Patient, Intervention, Comparator, Outcome, Setting / Study design]-Schema) systematisch die zur Beantwortung dieser Fragestellung geeigneten klinischen Studien zu identifizieren, die Sicherheit der Ergebnisse der identifizierten Studien in nachvollziehbarer Weise zu bewerten und auf Basis der beobachteten Daten sowie unter Berücksichtigung der Ergebnissicherheit

eine zusammenfassende Bewertung abzugeben. Im Hinblick auf die Aussagesicherheit für die Beantwortung der zugrunde liegenden Fragestellung gemäß dem PICOS-Schema spielen 4 Komponenten eine Rolle. Die Sicherheit der Ergebnisse der identifizierten klinischen Studien wird dabei in erster Linie durch 3 Komponenten getragen:

- 1) eine qualitative Komponente, die durch das Verzerrungspotenzial der zu bewertenden Studien gekennzeichnet ist (interne Validität),
- 2) eine quantitative Komponente, die durch die Fallzahl(en), aber auch durch die Varianz der Beobachtungen bestimmt wird (Präzision der Ergebnisse),
- 3) die Größe beobachteter Unterschiede (Effektstärke).

Hinzu kommt noch eine weitere Komponente hinsichtlich der Aussagesicherheit:

- 4) die externe Validität (oder auch Anwendbarkeit), d. h. der Aspekt, inwieweit die Studienbedingungen ebendiese Fragestellung abbilden.

Für die Bewertung des Verzerrungspotenzials lassen sich 6 Kategorien von Verzerrungen (bias) abgrenzen, wobei es bei den ersten 4 Kategorien um Unterschiede zwischen den zu vergleichenden Interventionsgruppen geht: Selektionsbias (selection bias), Durchführungsbias (performance bias), Entdeckungsbias (detection bias), Abriebbias (attrition bias), Berichtsbias (reporting bias) und sonstige Verzerrungsmöglichkeiten, z. B. die Verwendung ungeeigneter statistischer Methoden.

Zur Vermeidung der ersten 4 genannten Verzerrungsmöglichkeiten lassen sich im Wesentlichen 3 Strategien einsetzen: randomisierte und verdeckte (concealed) Zuteilung von Patienten in die zu vergleichenden Interventionsgruppen, die Verblindung der Interventionen für Patienten und Studienpersonal (auch als Doppelblindheit bezeichnet) sowie die Auswertung aller in eine Studie eingeschlossenen Patienten gemäß der ihnen zugewiesenen Intervention (Intention-to-treat-Prinzip [ITT-Prinzip]).

Barrieren gegen randomisierte Studien

Alternativen zur Randomisierung, um insbesondere einen Selektionsbias zu vermeiden, sind derzeit nicht bekannt. Dennoch werden mitunter Argumente gegen randomisierte kontrollierte Studien vorgebracht, die die Durchführbarkeit, häufig aufgrund nicht näher spezifizierter ethischer Bedenken oder aber auch wegen logistischer Probleme, betreffen.

Ethisch zweifelhaft ist eine kontrollierte Studie in der Tat dann, wenn der Nutzen (oder Zusatznutzen im Vergleich zum Standard) einer Intervention in einer bestimmten Indikation mehr oder weniger belegt ist. Dann erübrigen sich aber Studien mit dem Ziel des Erkenntnisgewinnes im Hinblick auf die Frage nach dem Nutzen oder Zusatznutzen ohnehin, und das nicht nur bei seltenen Erkrankungen.

2 Szenarien könnten gegen eine Randomisierung sprechen: Die Durchführung einer adäquaten zentralen Randomisierung mit Wahrung des Concealment ist zu aufwendig oder eine bestimmte Intervention ist aus verschiedenen Gründen (z. B. mengenmäßig oder auch wegen mangelnder Expertise bei der Anwendung) für Studienzwecke nicht ausreichend verfügbar und kann auch nicht kurzfristig verfügbar gemacht werden.

Situation bei seltenen Erkrankungen

Es ist offensichtlich, dass es immer schwieriger wird, eine für die Beantwortung einer spezifischen Fragestellung ausreichende Fallzahl zu erreichen, je seltener diese Erkrankung ist. Somit wird die Ergebnissicherheit bei seltenen Erkrankungen vor allem im Hinblick auf die Präzision beeinträchtigt. Dies hat u. a. Auswirkungen auf den Fehler 2. Art (einen vorhandenen Unterschied nicht zu finden). Diese Fehlerwahrscheinlichkeit hängt von dem tatsächlichen Unterschied (von der „Wahrheit“) ab und kann somit nur bedingt über die durch Annahmen über diese „Wahrheit“ begründete Fallzahl kontrolliert werden. Demgegenüber kann der Fehler 1. Art (fälschlicherweise einen Unterschied anzunehmen, obwohl er gar nicht existiert) auch bei kleinen Fallzahlen über die Festlegung des Signifikanzniveaus zumindest theoretisch kontrolliert werden.

Die beiden anderen beeinflussbaren Komponenten der Ergebnis- bzw. Aussagesicherheit (interne und externe Validität) sind demgegenüber nicht quantifizierbar und können zu systematischen Fehlern (Abweichungen von der „Wahrheit“) führen. Diesbezügliche Schwächen im Design einer Studie lassen sich insbesondere im Hinblick auf das Verzerrungspotenzial in aller Regel nicht durch noch so ausgefeilte statistische Methoden ausgleichen. Deshalb ist es im Kontext der Bewertung medizinischer Interventionen bei seltenen Erkrankungen naheliegend, zunächst nach Möglichkeiten zu suchen, die Beschränkungen durch die ggf. erniedrigte Präzision durch optimierte statistische Verfahren auszugleichen.

Design-Vorschläge für randomisierte Therapiestudien bei seltenen Erkrankungen

Ausgangspunkt für vorgeschlagene Modifikationen ist das klassische randomisierte Parallelgruppen-Design, bei dem die Patienten auf Basis eines Zufallsmechanismus einer von 2 oder mehr unterschiedlichen Interventionen zugewiesen und über eine vorab definierte Zeit beobachtet werden. Die Verfahren und Designs lassen sich grob in 5 Gruppen unterteilen, wobei die Abgrenzung nicht scharf gezogen werden kann:

- Designs zur Reduktion der Varianz zwischen den Beobachtungen, z. B. Stratifizierung, regressionsanalytische Verfahren.
- Verfahren mit Anpassungen des Designs im Studienverlauf, z. B. sequenzielle Designs.
- Verfahren, bei denen Vorinformationen (außerhalb der Studie) in die statistische Auswertung mit einfließen (Bayes'sche Verfahren).
- Designs, bei denen alle Patienten auch die zu prüfende Therapie erhalten.
- Sonstige Designs, z. B. adaptive Randomisierung.

Ergebnissicherheit nicht randomisierter Studien

Nicht randomisierte Studien bergen inhärent erhebliche Einschränkungen bei der internen Validität (1. Komponente der Ergebnissicherheit). Gegenwärtig ist keine Methode bekannt, die in suffizienter Weise diese Einschränkungen aufheben könnte. Je nach Grund für einen eventuellen Verzicht auf eine Randomisierung ergeben sich unmittelbare Konsequenzen für ersatzweise nicht randomisierte Studien. Werden ethische Bedenken geltend gemacht, verbieten sich jegliche parallel vergleichenden Studien, und es verbleiben im Prinzip nur noch historische Kontrollen.

Historisch kontrollierte Studien stehen in den gängigen Evidenzhierarchien weit unten. Neben den üblichen Verzerrungsmöglichkeiten (z. B. Selektionsbias) kommt hier als weitere Störgröße der Faktor Zeit oder Chronologie dazu. Ergebnisse aus historisch kontrollierten Studien lassen Aussagen im Sinne eines Interventionseffektes nur dann zu, wenn (nahezu) eine Umkehr eines quasi deterministischen Verlaufs beobachtet wurde („dramatischer Effekt“) vorliegt. In der Literatur wird als Kriterium für das Vorliegen eines dramatischen Effektes ein um den Faktor 10 erhöhtes Risiko in der Interventionsgruppe im Vergleich zur Kontrollintervention in Verbindung mit einem (adäquaten) statistischen Test zum Irrtumsniveau 1 % vorgeschlagen.

Von prospektiv geplanten Interventionsstudien werden Beobachtungsstudien abgegrenzt, deren Sinn und Zweck entweder weniger der Frage nach den Effekten von Interventionen nachgehen oder deren Datengrundlage primär nicht der Gewinnung von medizinisch-klinischen Fragen dient. In den gängigen Evidenzhierarchien werden sie deshalb in ihrer Aussagekraft den randomisierten und nicht randomisierten Interventionsstudien nachgeordnet.

Empfehlungen

Es lässt sich keine wissenschaftliche Begründung für eine unterschiedliche Herangehensweise bei der Bewertung von medizinischen Interventionen für seltene und nicht seltene Erkrankungen ableiten. Studien mit geringer Präzision oder mit nicht ausreichenden Schutzmechanismen vor potenziell verzerrenden Faktoren oder mit Abweichungen der Studienbedingungen von der eigentlich interessierenden Fragestellung (z. B. Verwendung von Surrogatendpunkten anstelle von patientenrelevanten Endpunkten) haben die gleichen Auswirkungen auf die Aussagesicherheit bei seltenen und nicht seltenen Erkrankungen. Umgekehrt existieren keine spezifischen Designs und statistischen Methoden für seltene Erkrankungen, die nicht auch relevant für häufige(re) Erkrankungen sein könnten.

Im Kontext der Bewertung medizinischer Interventionen bei seltenen Erkrankungen wird es somit (besonders) notwendig sein, ein möglichst effizientes methodisches Vorgehen zu wählen und dabei auch effiziente Strukturen zu schaffen. Darüber hinaus kann es, insbesondere bei sehr seltenen Erkrankungen, erforderlich sein, Kompromisse bei der Aussagesicherheit einzugehen, die auch auf externen (politischen) Vorgaben beruhen können.

Im Rahmen des Arzneimittelmarktneuordnungsgesetzes (AMNOG) hat der Gesetzgeber in Deutschland politische Vorgaben gemacht, indem der Zusatznutzen von Arzneimitteln für die Behandlung seltener Erkrankungen (Orphan Drugs) qua Zulassung als belegt gilt, allerdings nur so lange, wie der Umsatz eines solchen Arzneimittels mit der gesetzlichen Krankenversicherung innerhalb eines Zeitraums von 12 Kalendermonaten einen Betrag von 50 Millionen € nicht übersteigt. Einerseits findet sich in der Begründung das inhaltliche Argument, dass regelmäßig davon auszugehen ist, dass es keine therapeutisch gleichwertige Behandlungsalternative gibt. Andererseits macht die Begründung auch die politische Vorgabe deutlich, indem der Zusatznutzen am Umsatz des Arzneimittels festgemacht wird. Wissenschaftlich ist es nicht zu begründen, warum inhaltliche Kriterien, die – wenn sie denn zutreffen – eine Bewertung (des [Zusatz-]Nutzens) obsolet machen, ab einem gewissen Umsatzvolumen nicht mehr zutreffen (sollen).

Bei seltenen Erkrankungen wird es mit abnehmender Häufigkeit der Erkrankung zunehmend schwieriger, der quantitativen Komponente der Ergebnissicherheit ausreichend Rechnung zu tragen. Eine diesbezüglich im Prinzip triviale Feststellung lautet daher, dass es für die klinische, patientenorientierte Erforschung seltener Erkrankungen besonders notwendig ist, in vernetzten, überregionalen und supranationalen Strukturen zu arbeiten. Daraus lässt sich ein starkes Plädoyer für Krankheitsregister mit klaren Qualitätskriterien bezüglich Vollzähligkeit und Vollständigkeit als Basis für hochwertige klinische, vor allem nicht randomisierte Studien ableiten.

Wenn es schon inhärent Beschränkungen bezüglich einer Komponente der Ergebnissicherheit gibt, scheint es wenig zielführend auch die anderen Komponenten zu kompromittieren. Von den zuvor beschriebenen Strategien wären somit diejenigen zu bevorzugen, die bei weitgehendem Erhalt von interner und externer Validität Fallzahleinsparungen ermöglichen, also sequenzielle Designs.

Bei sehr seltenen Erkrankungen könnte alternativ oder zusätzlich zum gängigen methodischen Vorgehen erwogen werden, ein größeres statistisches Irrtumsniveau für regulative Entscheidungen zuzulassen, z. B. das übliche zweiseitige Irrtumsniveau von den üblichen 5 % auf 10 % anzuheben. Der Vorteil einer solchen Vorgehensweise wäre, die Irrtumsmöglichkeit zumindest quantifizieren zu können. Ein derartiges Vorgehen kann auch als Annäherung an Bayes'sche Verfahren verstanden werden: Ob nämlich die Daten aus einer randomisierten Studie mit geringer Fallzahl mit einem erhöhten Irrtumsniveau zur Entscheidungsfindung herangezogen werden oder ob bei Verknüpfung mit z. B. optimistischem Vorwissen aus einer anderen Indikation (Prior) das übliche Irrtumsniveau beibehalten wird, dürfte auf vergleichbare Ergebnisse im Hinblick auf resultierende Entscheidungen hinauslaufen.

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. Auch der Einsatz von adaptiven Designs wird in der Regel mit Einschränkungen der externen Validität verbunden sein. Immerhin bliebe aber wenigstens die interne Validität erhalten.

Die Aufgabe der internen Validität durch Verzicht auf eine Randomisierung stünde in einem solchen hierarchischen Vorgehen aus logischen Gründen an letzter Stelle. Wesentliche Voraussetzung, um sie für (regulatorische) Entscheidungen nutzen zu können, wäre, dass die zugrunde liegenden Daten aus einem Krankheitsregister mit den o. g. Qualitätskriterien stammen oder dass ein derartig großer beobachteter Unterschied vorliegt, der nicht mehr allein durch Bias erklärt werden kann.

Ergebnisse der empirischen Untersuchung zur Studiengrundlage für die Zulassung von Orphan Drugs in Europa

Es wurden 85 Arzneimittel mit europäischer Orphan Drug Designation und europäischer Marktzulassung von 2001 bis 2013 identifiziert. Die Zulassung der 85 Arzneimittel stützte sich auf 125 Hauptstudien (ohne 6 Zulassungen, die auf Literatur-Reviews basierten); darunter waren 82 RCTs.

Auswertungen auf Arzneimittelebene

Alle Zulassungen

58 der identifizierten 85 Arzneimittel (68 %) dienen der Behandlung seltener, 27 Arzneimittel (32 %) der Behandlung sehr seltener Erkrankungen. Bei 59 Arzneimitteln (69 %) basierte die Zulassung auf RCTs (55 ausschließlich auf RCTs und 4 auf RCTs in Kombination mit Non-RCTs). Bei 20 Arzneimitteln (24 %) dienen Non-RCTs als Grundlage für die Zulassung.

Zulassungen ohne Literatur-Reviews

Die Zulassung der 79 Orphan Drugs ohne Literatur-Reviews beruhte auf jeweils 1 bis 5 Hauptstudien. Bei sehr seltenen Erkrankungen waren maximal 3 Studien Basis für die Zulassung. Der Anteil von Zulassungen basierend auf Daten aus RCTs betrug etwa 75 %. Die Patientenanzahl der Zulassungsstudien pro Arzneimittel betrug zwischen 27 und 2961 Patienten (Median 165). Etwa 70 % aller Patienten wurden in RCTs behandelt. Bei 66 Arzneimitteln (84 %) wurden patientenrelevante Endpunkte in den Studien berücksichtigt (bei 31 Orphan Drugs als primärer, bei 57 als sekundärer Endpunkt). Der Großteil der Studien wurde multizentrisch, multinational und multikontinental durchgeführt. In der Mehrzahl der Fälle zeigten sich keine auffälligen Unterschiede zwischen seltenen und sehr seltenen Erkrankungen in den dargestellten Studiencharakteristika.

Auswertungen auf Studienebene

Grundlage der deskriptiven Auswertungen zu Charakteristika und Methodik der Zulassungsstudien sind alle 82 RCTs (66 %) unter den 125 Hauptstudien (ohne Zulassungen basierend auf Literatur-Reviews) der zugelassenen Orphan Drugs (1 bis 3 RCTs pro Arzneimittel). In den RCTs wurden zwischen 8 und 769 Patienten (Median 160,5) eingeschlossen.

Der Anteil doppelblinder RCTs liegt bei 74 %. Bei 28 % der Studien wurde ein aktiver Komparator eingesetzt. Nichtunterlegenheits- bzw. Äquivalenzstudien kamen mit 6 % der

Studien nur in Ausnahmefällen zum Einsatz. Es zeigen sich keine auffälligen Unterschiede zwischen den Studien zu seltenen und sehr seltenen Erkrankungen.

In 52 % der RCTs kamen Methoden zur Kontrolle von Störgrößen zum Einsatz. Cross-over-Designs und sequenzielle Verfahren wurden in 5 % bzw. 12 % der Studien verwendet. Eine Erhöhung des Irrtumsniveaus wurde in 1 Studie vorgenommen. Eine adaptive Randomisierung, Bayes'sche Verfahren oder sonstige spezielle randomisierte Designs wurden in keiner der Zulassungsstudien eingesetzt. Insgesamt wurden in den Studien zu sehr seltenen Erkrankungen weniger häufig spezielle Auswertungsmethoden eingesetzt, sequenzielle Verfahren kamen gar nicht zur Anwendung.

Fazit

Eine Begründung für eine unterschiedliche Herangehensweise bei der Bewertung von medizinischen Interventionen für seltene gegenüber nicht seltenen Erkrankungen kann wissenschaftlich nicht abgeleitet werden. Umgekehrt existieren auch keine spezifischen Designs und statistischen Methoden für seltene Erkrankungen, die nicht auch relevant für häufige(re) Erkrankungen sein könnten. Dies gilt in gleicher Weise für medikamentöse wie für nichtmedikamentöse Interventionen.

Zulassungen und Zulassungsstudien für Orphan Drugs basieren zu einem großen Teil, auch bei sehr seltenen Erkrankungen, auf konventionellen (randomisierten) Designs, sodass die grundsätzliche Machbarkeit nicht infrage steht.

Im Kontext der Bewertung medizinischer Interventionen bei seltenen, insbesondere sehr seltenen Erkrankungen kann es dennoch notwendig oder auch politisch gewünscht sein, Kompromisse bei der Aussagesicherheit einzugehen. Solche Kompromisse sind grundsätzlich auf 3 Ebenen denkbar:

Das (statistische) Irrtumsniveau könnte über den üblichen Wert von (zweiseitig) 5 % angehoben werden (Kompromiss bei der geforderten Präzision).

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.

Die Aufgabe der internen Validität durch Verzicht auf eine Randomisierung stünde in einem hierarchischen Vorgehen aus logischen Gründen an letzter Stelle. Wesentliche Voraussetzung, um sie für (regulatorische) Entscheidungen nutzen zu können, wäre, dass die zugrunde liegenden Daten aus einem Krankheitsregister mit exzellenter Qualität im Hinblick auf Vollständigkeit und Vollzähligkeit stammen.

Schlagwörter: Seltene Krankheiten, Nutzenbewertung

Keywords: Rare Diseases, Benefit Assessment

1 Hintergrund

Die Kriterien zur Definition seltener Erkrankungen sind nicht einheitlich [1]. Die in Europa gebräuchlichste Definition spricht von einer seltenen Erkrankung bei einer Häufigkeit von 5 oder weniger Betroffenen pro 10 000 Einwohner (Prävalenz $\leq 0,0005$) [2]. Ca. 7000 bis 8000 der rund 30 000 bekannten Krankheiten werden als selten eingestuft. In Deutschland leiden bis zu 4 Millionen Menschen an einer seltenen Krankheit [3]. Eine Erkrankung mit einer Prävalenz von $< 2 / 100\ 000$ Einwohnern wird als sehr selten eingestuft [4,5].

Die aktuelle Relevanz des Themas zeigt sich auch darin, dass im Jahr 2010 das Bundesministerium für Gesundheit (BMG) gemeinsam mit dem Bundesministerium für Bildung und Forschung (BMBF) und der Allianz Chronischer Seltener Erkrankungen (ACHSE e. V.) das Nationale Aktionsbündnis für Menschen mit Seltene Erkrankungen (NAMSE) gegründet hat. Im Rahmen dieses Projekts wurde ein Maßnahmenkatalog erarbeitet mit dem Ziel, die Behandlung von Patienten mit seltenen Erkrankungen zu verbessern [6]. Der Maßnahmenkatalog beinhaltet auch den Aspekt der Entwicklung von Kriterien zur Bewertung und Auswertung von Studien mit geringer Teilnehmerzahl. Im Rahmen des 7. Forschungsrahmenprogramms der Europäischen Kommission fördert die Europäische Union (EU) aktuell 3 internationale, multidisziplinäre Forschungsprojekte zu methodischen Aspekten von klinischen Studien bei seltenen Erkrankungen, und zwar ASTERIX (Advances in Small Trials Design for Regulatory Innovation and Excellence) [7], IDEAL (Integrated Design and Analysis of Small Population Group Trials) [8] und InSPiRe (Innovative Methodology for Small Population Research) [9].

Die Durchführung aussagekräftiger Studien zur Bewertung von Interventionen bei seltenen Erkrankungen stellt eine besondere Herausforderung dar. Grundsätzlich ist auch hier die randomisierte klinische Studie (RCT) als Goldstandard anzusehen. Es wird häufig argumentiert, dass aufgrund der geringen Anzahl betroffener Patienten und somit potenzieller Studienteilnehmer die Durchführung qualitativ hochwertiger Studien erschwert oder gar praktisch nicht durchführbar sein kann [10]. In der Tat kann es bei sehr seltenen Erkrankungen problematisch sein, bei erwarteten geringen bis mäßigen Effektstärken ausreichend viele Probanden für eine Studie zu patientenrelevanten Endpunkten mit einer als adäquat angesehenen Power zu gewinnen. Dies gilt aber grundsätzlich nicht nur für RCTs, sondern auch für alle sonstigen Studien [1].

Das Ziel einer möglichst unverzerrten Schätzung des Behandlungseffekts gilt in kleinen wie in großen Studien gleichermaßen. Grundsätzlich gelten methodische Grundprinzipien und statistische Verfahren zur Planung und Auswertung klinischer Studien bei seltenen Erkrankungen ebenso wie bei Studien mit großen Fallzahlen [11]. Im Rahmen seiner Verordnung (EG) Nr. 141/2000 vom 16. Dezember 1999 stellen das Europäische Parlament und der Rat der Europäischen Union zudem fest, dass „Patienten mit solchen [seltenen] Leiden (...) denselben Anspruch auf Qualität, Unbedenklichkeit und Wirksamkeit von

Arzneimitteln [haben] wie andere Patienten. Arzneimittel für seltene Leiden sollten daher dem normalen Bewertungsverfahren unterliegen“ [2].

In der Literatur werden diverse methodische Ansätze speziell für klinische Studien zur Untersuchung seltener Erkrankungen vorgeschlagen und diskutiert [11,12]. Diese Ansätze zielen darauf ab, die vorhandenen eingeschränkten Informationen möglichst effizient zu nutzen (z. B. durch Verfahren zur Erhöhung der Präzision von Effektschätzern).

Es stellt sich die Frage nach der Eignung dieser Verfahren für die Durchführung klinischer Studien sowie nach Kriterien zur Bewertung der Ergebnis- und Aussagesicherheit von Studien bei seltenen Erkrankungen im Rahmen von Nutzenbewertungen.

2 Ziele der Untersuchung

Die Ziele des vorliegenden Rapid Reports sind

- die Erstellung einer Expertise zu methodischen Aspekten bei der Durchführung und Auswertung sowie zur Bewertung der Ergebnissicherheit von Studien bei seltenen Erkrankungen,
- die Beschreibung der Studiengrundlage für die Zulassung von Orphan Drugs in Europa.

3 Projektbearbeitung

3.1 Zeitlicher Verlauf des Projekts

Das Bundesgesundheitsministerium (BMG) hat mit Schreiben vom 10.12.2013 das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) mit der Erstellung eines Rapid Reports zum Thema „Bewertung und Auswertung von Studien bei seltenen Erkrankungen“ beauftragt.

Der vorliegende Rapid Report wurde innerhalb des IQWiG erstellt. Die Methodik zur Erstellung des Rapid Reports wurde intern in einer nicht veröffentlichten Projektskizze beschrieben (Version 1.0 vom 11.02.2014). Der vorläufige Rapid Report wurde extern begutachtet. Anschließend erfolgte die Übermittlung der endgültigen Fassung des Rapid Reports an das BMG und 4 Wochen später die Veröffentlichung auf der Internetseite des IQWiG.

4 Methoden

4.1 Methodische Expertise

4.1.1 Gegenwärtige Rahmenbedingungen und Problemstellung

In der methodischen Expertise wird die besondere Problematik bei der Durchführung und Bewertung von Studien zu seltenen Erkrankungen dargestellt. Dabei wird unterschieden zwischen seltenen ($\leq 5 / 10\,000$) und sehr seltenen Erkrankungen ($< 2 / 100\,000$ Einwohner [4,5]). Die Ausführungen gelten grundsätzlich für beliebige Interventionen; eine inhaltliche Einschränkung bzw. Unterscheidung hinsichtlich medikamentöser oder nichtmedikamentöser Interventionen erfolgte nicht.

4.1.2 Methodische Aspekte bei der Bewertung und Auswertung von Studien

Ebenso werden in der methodischen Expertise Ergebnis- und Aussagesicherheit von häufig bei seltenen Erkrankungen eingesetzten bzw. vorgeschlagenen Studiendesigns als Grundlage für Nutzenbewertungen unter Berücksichtigung relevanter methodischer Publikationen und Leitfäden einschlägiger Institutionen (z. B. der EMA [11]) dargestellt und diskutiert. Eine systematische Literaturrecherche zu der zugrunde liegenden Fragestellung erfolgte aufgrund mangelnder Praktikabilität und eingeschränkter Relevanz nicht. Sofern möglich, werden Empfehlungen zur Berücksichtigung der spezifischen Gegebenheiten bei seltenen und sehr seltenen Erkrankungen gemacht.

4.2 Empirische Untersuchung zur Studiengrundlage für die Zulassung von Orphan Drugs in Europa

4.2.1 Informationsbeschaffung

Die Recherche nach Orphan Drugs für die empirische Untersuchung der Studiengrundlage für die Zulassung von Orphan Drugs in Europa erfolgte im Orphanet „Verzeichnis der Arzneimittel für seltene Krankheiten in Europa“ [13]. Dieses Verzeichnis enthält eine Auflistung aller Arzneimittel mit europäischer Orphan-Drug-Designation und europäischer Marktzulassung von 2001 bis Januar 2014. Orphanet ist das Referenzportal für Informationen über seltene Krankheiten und Orphan Drugs, das von einem Konsortium aus rund 40 Ländern unter der Leitung von INSERM (Institut national de la santé et de la recherche médicale) geführt wird.

Gemäß Definition der EMA [14] erhält ein Arzneimittel eine Orphan-Drug-Designation, falls

- das Arzneimittel für die Diagnose, Verhütung oder Behandlung eines Leidens bestimmt ist, das lebensbedrohend ist oder eine chronische Invalidität nach sich zieht und von dem zum Zeitpunkt der Antragstellung in der Gemeinschaft nicht mehr als 5 von 10 000 Personen betroffen sind, oder
- das Arzneimittel eine Anwendung gemäß obigen Kriterien hat, es aber ohne (ökonomische) Anreize unwahrscheinlich ist, dass die erwarteten Verkaufszahlen die

notwendigen Entwicklungskosten decken, und das Arzneimittel für die Diagnose, Verhütung oder Behandlung eines lebensbedrohenden Leidens, eines zu schwerer Invalidität führenden oder eines schweren und chronischen Leidens in der Gemeinschaft bestimmt ist und dass das Inverkehrbringen des Arzneimittels in der Gemeinschaft ohne Anreize vermutlich nicht genügend Gewinn bringen würde, um die notwendigen Investitionen zu rechtfertigen, und

- in der Gemeinschaft noch keine zufriedenstellende Methode für die Diagnose, Verhütung oder Behandlung des betreffenden Leidens zugelassen wurde oder dass das betreffende Arzneimittel – sofern eine solche Methode besteht – für diejenigen, die von diesem Leiden betroffen sind, von erheblichem Nutzen sein wird.

Für die Extraktion der relevanten Daten zu den identifizierten Orphan Drugs wurden European Public Assessment Reports (EPARs) zu den Zulassungsstudien sowie das Orphanet [15,16] herangezogen.

4.2.2 Informationssynthese

Die folgenden Informationen wurden extrahiert:

Charakteristika des Produkts

- Name des Arzneimittels
- Name des Wirkstoffs
- Indikation(en)
- Häufigkeit der Erkrankung (Prävalenz)

Es wurden primär die Angaben aus dem Orphanet-Dokument „Prävalenz seltener Krankheiten: Bibliographische Angaben“ herangezogen [16]. Waren hier keine Angaben vorhanden, wurde alternativ im Orphanet bei den relevanten Indikationen, in den zugehörigen EPARs sowie in wissenschaftlichen Publikationen [17,18] nach Informationen zur Prävalenz der Erkrankung geschaut.

- Datum der Zulassung
- Bedingte Zulassung und Zulassung unter außergewöhnlichen Umständen

Eine bedingte Zulassung („conditional marketing authorisation“) bezeichnet eine zeitlich begrenzte Zulassung vorbehaltlich der Vorlage weiterer klinisch relevanter Daten. Eine Zulassung unter außergewöhnlichen Umständen („marketing authorisation under exceptional circumstances“) liegt vor, wenn umfassende Daten nicht verfügbar sind und auch in Zukunft nicht bereitgestellt werden können.

- ggf. Datum des Register-Austrags
- ggf. Datum der Markt-Rücknahme

Charakteristika der zugrunde liegenden Zulassungsstudien

- Anzahl zulassungsrelevanter Studien zur klinischen Wirksamkeit und Sicherheit

Es wurden die in den EPARs als Hauptstudien („main studies“) bezeichneten Studien zur klinischen Wirksamkeit („clinical efficacy“) für die Auswertungen im Rahmen dieses Rapid Reports herangezogen.

- Anzahl der Patienten pro Studie und in den Behandlungsgruppen
- Studiendesign (Studientyp, Kontrolle, Randomisierung, Verblindung)
- Betrachtete Endpunkte (primär / sekundär)

Endpunkte wurden als primär bzw. sekundär eingestuft, wenn sie explizit so benannt wurden oder eindeutig den genannten primären bzw. sekundären Studienzielen zuzuordnen waren. Als „exploratorisch“ bezeichnete Endpunkte wurden nicht berücksichtigt. Die primären und sekundären Endpunkte wurden cursorisch hinsichtlich ihrer Patientenrelevanz klassifiziert. Fälle mit unmittelbar erkennbarer Patientenrelevanz wurden mit „ja“, andernfalls mit „nein“ bewertet. Die anhand der Beschreibungen im EPAR nicht unmittelbar eindeutigen Fälle, die einer näheren Bewertung im Einzelfall bedürften, wurden als „unklar“ eingestuft. Insbesondere wurde progressionsfreies Überleben als „nein“ und krankheitsfreies Überleben als „unklar“ klassifiziert.

- Verwendete statistische Auswertungsmethoden

Die verwendeten statistischen Auswertungsmethoden wurden dargestellt. Insbesondere wurde erfasst, inwieweit die im methodischen Teil (Abschnitt 4.1.2) beschriebenen Verfahren zur Auswertung von Studien bei seltenen Erkrankungen in den Zulassungsstudien zur Anwendung gekommen sind.

Alle für die Bewertung relevanten Informationen wurden aus den genannten Unterlagen zu den eingeschlossenen Studien von einem Reviewer in standardisierte Tabellen extrahiert. Ein 2. Reviewer überprüfte die Extraktion. Bei auftretenden Diskrepanzen wurden diese durch Diskussion zwischen den beiden Reviewern aufgelöst.

4.2.3 Informationsanalyse

Die extrahierten Daten wurden mithilfe deskriptiver statistischer Verfahren ausgewertet. Die Analyse erfolgte zum einen auf Zulassungs- bzw. Arzneimittelebene und zum anderen auf Studienebene. Lagen mehrere Studien zu einem Arzneimittel vor, so erfolgte bei den Auswertungen auf Arzneimittelebene ein positiver Eintrag, wenn mindestens eine der Studien die entsprechende Eigenschaft aufwies (z. B. bei einem Arzneimittel mit 2 Zulassungsstudien, 1 RCT und ein 1 Non-RCT, erfolgte auf Arzneimittelebene ein positiver Eintrag für „RCT“). Zudem erfolgte eine abschließende zusammenfassende Bewertung der Informationen.

5 Ergebnisse

5.1 Methodische Expertise

5.1.1 Grundlegende methodische Aspekte bei der Bewertung und Auswertung von Studien

Nach internationaler Übereinkunft ist die wissenschaftliche Grundlage der EbM, im Hinblick auf eine klar definierte Fragestellung systematisch die zur Beantwortung dieser Fragestellung geeigneten klinischen Studien zu identifizieren, die Sicherheit der Ergebnisse der identifizierten Studien in nachvollziehbarer Weise zu bewerten und auf Basis der beobachteten Daten sowie unter Berücksichtigung der Ergebnissicherheit eine zusammenfassende Bewertung abzugeben. Im Zusammenhang mit der Bewertung von Nutzen und Schaden medizinischer Interventionen müssen für die klare Definition der Fragestellung die Indikation, die zu prüfende Intervention, die Vergleichsintervention sowie interessierende Endpunkte festgelegt werden. Weitere Festlegungen bezüglich geeigneter Designs sind in der Regel sinnvoll. Ausgehend von den entsprechenden englischen Begriffen wird hier auch von einem PICOS-Schema gesprochen: P für Population (im Sinn der Indikation), I für Intervention, C für Comparator, O für Outcome und S für Study design [19] oder auch Setting [20].

Komponenten der Ergebnis- und Aussagesicherheit

Im Hinblick auf die Aussagesicherheit für die Beantwortung der zugrunde liegenden Fragestellung gemäß dem PICOS-Schema spielen 4 Komponenten eine Rolle. Die Sicherheit der Ergebnisse der identifizierten klinischen Studien wird dabei in erster Linie durch 3 Komponenten getragen:

- 1) eine qualitative Komponente, die durch die sogenannte interne Validität der zu bewertenden Studien gekennzeichnet ist. Dafür wird das Risiko, dass die beobachteten Ergebnisse systematisch von der „Wahrheit“ abweichen, beschrieben. Dies wird auch als Verzerrungspotenzial bezeichnet. Die entsprechende Bewertung erfolgt sowohl auf Studienebene als auch auf Endpunktebene und basiert auf einer Analyse des Studiendesigns. Die nicht verblindete Erhebung von subjektiven Endpunkten birgt beispielsweise ein hohes Verzerrungspotenzial [21,22].
- 2) eine quantitative Komponente, die durch die Fallzahl(en), aber auch durch die Varianz der Beobachtungen bestimmt wird. Diese quantitative Komponente ist Ausdruck der Präzision der Ergebnisse und findet in der Breite von Konfidenzintervallen oder der Größe eines p-Wertes und damit in dem Überschreiten oder Nicht-Überschreiten der gewählten Signifikanzgrenze ihren Niederschlag.
- 3) die Größe beobachteter Unterschiede (Effektstärke). Beobachtete Unterschiede können trotz statistischer Signifikanz mitunter so klein sein, dass sie entweder keine ausreichende Relevanz besitzen [23] oder deren Bedeutung nicht ausreichend sicher interpretiert werden kann [24]. Umgekehrt können beobachtete Unterschiede so groß sein, dass davon

ausgegangen werden kann, dass sie nicht allein auf Verzerrung beruhen (können). Letzteres wird auch als „dramatischer Effekt“ bezeichnet [25] (siehe auch Abschnitt 5.1.5).

Hinzu kommt noch eine weitere Komponente hinsichtlich der Aussagesicherheit:

- 4) die externe Validität (oder auch Anwendbarkeit), d. h. der Aspekt, inwieweit die Studienbedingungen eben diese Fragestellung abbilden.

Diese 4. Komponente wird häufig der 1. Komponente, also dem Verzerrungspotenzial, gegenübergestellt. Bei einer solchen Gegenüberstellung darf allerdings nicht übersehen werden, dass Mängel bei der internen Validität nicht durch eine besonders hohe externe Validität aufgehoben werden können. Wenn eine Studie bzw. ein Studienergebnis wegen eines (zu) hohen Verzerrungspotenzials nicht ausreichend sicher interpretiert werden kann, ist auch eine Verallgemeinerbarkeit wertlos [10].

Zwischen den Komponenten kann es gewisse Interdependenzen geben. So können quasi definitionsgemäß die interne und die externe Validität sowohl die Präzision der Ergebnisse als auch die Größe beobachteter Unterschiede beeinflussen.

Kategorien von Verzerrungen

Da, wie zuvor ausgeführt, eine ungenügende interne Validität eine Interpretation grundlegend beeinträchtigt oder gar unmöglich macht, werden nachfolgend kurz die wesentlichen Verzerrungsmöglichkeiten erläutert. Hierbei lassen sich 6 Kategorien von Verzerrungen (bias) abgrenzen, wobei es bei den ersten 4 Kategorien um Unterschiede zwischen den zu vergleichenden Interventionsgruppen geht [26,27].

- Selektionsbias (selection bias), d. h. Unterschiede in den prognostisch bedeutsamen Patientencharakteristiken.
- Durchführungsbias (performance bias), d. h. Unterschiede in den Begleitumständen während der Studie.
- Entdeckungsbias (detection bias), d. h. Unterschiede bei der Endpunkterhebung.
- Abriebbias (attrition bias), d. h. Unterschiede im Anteil von Patienten, die vorzeitig aus der Studie ausscheiden und für die Endpunkte nicht erhoben werden (können).
- Berichtsbias (reporting bias), d. h. das selektive Nichtberichten von Ergebnissen, sei es von einzelnen Endpunkten oder von ganzen Studien (was zumeist als Publikationsbias bezeichnet wird).
- Sonstige Verzerrungsmöglichkeiten, z. B. die Veränderung von Endpunkt-Operationalisierungen in Kenntnis von bereits vorliegenden Ergebnissen [28], wobei dies auch unter Berichtsbias subsumiert werden kann, oder die Verwendung ungeeigneter Instrumente zur Endpunkterhebung oder statistischer Methoden. Auch der vorzeitige Abbruch von Studien vor dem eigentlich geplanten Ende kann zu Verzerrungen führen,

insbesondere wenn dafür keine geeigneten statistischen Schutzmechanismen in der Studienplanung vorgesehen waren. Und die Ermöglichung eines (vorzeitigen) Behandlungswechsels (treatment switching) auf Basis individueller Abbruchkriterien (z. B. Erreichen eines Surrogatendpunktes) kann die Interpretation von Interventionseffekten auf Endpunkte, die nachfolgen (z. B. Mortalität), erschweren oder gar unmöglich machen [29].

Strategien zur Vermeidung von Verzerrungen

Zur Vermeidung der ersten 4 genannten Verzerrungsmöglichkeiten und somit der Sicherstellung eines kausalen Zusammenhangs zwischen Behandlung und Ergebnis lassen sich im Wesentlichen 3 Strategien einsetzen:

- Die zufällige (randomisierte) und vor allem verdeckte (concealed) Zuteilung von Patienten in die zu vergleichenden Interventionsgruppen zur Vermeidung eines Selektionsbias.
- Die Maskierung (Verblindung) der Interventionen für Patienten und Studienpersonal (auch als Doppelblindheit bezeichnet) zur Vermeidung eines Durchführungs- und Entdeckungsbias. Sollte dies wegen offensichtlicher Unterschiede zwischen den Interventionen nicht möglich sein, so können aber dennoch häufig einzelne Endpunkte in Unkenntnis der Intervention erhoben werden (verblindete Endpunkterhebung). Beim Vergleich von chirurgischen Interventionen können identisch gestaltete Verbände mitunter zumindest eine Teilverblindung ermöglichen.
- Die Auswertung aller in eine Studie eingeschlossenen Patienten gemäß der ihnen zugewiesenen Intervention. Dies nennt man auch Intention-to-treat-Prinzip (ITT-Prinzip).

In diesem Zusammenhang ist es bedeutsam, dass der Verzicht auf eine Randomisierung bedeutet, dass auch die beiden anderen Strategien nicht oder allenfalls eingeschränkt eingesetzt werden (können). So ist eine Doppelblindheit ohne Randomisierung kaum vorstellbar, und das ITT-Prinzip erfordert die eindeutige Definition und Erfassung (oder zumindest zufällige Auswahl – hier wäre dann nicht zufällige Zuteilung gemeint) aller für die zu untersuchende Fragestellung geeigneter Patienten. Ohne eine vollständige Registrierung (z. B. im Rahmen eines Erkrankungs- bzw. Indikationsregisters) ist das nicht zu erreichen. Im Hinblick auf die interne Validität in einer randomisierten Studie muss keine repräsentative Auswahl der Patienten erfolgen. Die Randomisierung und dabei insbesondere das Concealment bieten aber die Gewähr dafür, eine eindeutige Studienpopulation festlegen zu können. Solange die Randomisierung nicht verletzt wird, gibt es keinen selektiven (Nicht-)Einschluss. Bei nicht randomisierten Studien existiert kein solcher Mechanismus. Deshalb kann ein selektiver (Nicht-)Einschluss nur durch eine vollständige Erfassung (und Auswertung) aller geeigneten Patienten vermieden werden. Darüber hinaus muss auch davon ausgegangen werden, dass ein Publikationsbias bei nicht randomisierten Studien, insbesondere bei Beobachtungsstudien, schwieriger zu entdecken sein wird.

Das Verzerrungspotenzial wird üblicherweise in 3 Stufen eingeteilt: niedrig, hoch oder unklar [27]. Letztlich hat aber die letzte Stufe (unklar) die gleichen Auswirkungen auf die Sicherheit einer Aussage wie die Kategorie „hoch“.

5.1.2 Barrieren gegen randomisierte Studien

Alternativen zur Randomisierung, um insbesondere einen Selektionsbias zu vermeiden, sind derzeit nicht bekannt. Dennoch werden mitunter Argumente gegen randomisierte kontrollierte Studien vorgebracht, die die Durchführbarkeit, häufig aufgrund nicht näher spezifizierter ethischer Bedenken oder aber auch wegen logistischer Probleme, betreffen [30]. Da auch und gerade im Zusammenhang mit seltenen Erkrankungen solche Argumente zu hören und zu lesen sind [31], ist es sinnvoll, zunächst diese beiden Argumente und ihre Konsequenzen zu beleuchten.

Ethische Aspekte

Ethisch zweifelhaft ist eine kontrollierte Studie in der Tat dann, wenn der Nutzen (oder Zusatznutzen im Vergleich zum Standard) einer Intervention in einer bestimmten Indikation mehr oder weniger belegt ist. Dann erübrigen sich aber Studien mit dem Ziel des Erkenntnisgewinnes im Hinblick auf die Frage nach dem Nutzen oder Zusatznutzen ohnehin, und das nicht nur bei seltenen Erkrankungen. Alternative Studiendesigns müssen nicht erwogen oder gar neu entwickelt werden.

Häufig existieren aber auch ohne einen entsprechenden Beleg hohe Erwartungen an den bzw. Überzeugungen zu dem Nutzen bzw. Zusatznutzen einer neuen Intervention, besonders in einer Situation, in der die verfügbaren Therapieoptionen nur äußerst begrenzt bis gar nicht vorhanden sind. Dies erschwert aber die Durchführung einer jeglichen prospektiven Interventionsstudie mit zeitlich paralleler Kontrollintervention unabhängig vom Zuteilungsmechanismus, ob also randomisiert oder nicht. Ähnlich ist die Situation bei starken Präferenzen für die eine oder andere Intervention. In einer derartigen Situation verbleiben somit – wenn überhaupt – nur noch Beobachtungsstudien, die a priori extrem selektionierte Gruppen (überzeugt vs. nicht überzeugt bzw. Präferenz für A vs. Präferenz für B) miteinander vergleichen (müssen). Man muss dann davon ausgehen, dass die zugrunde liegenden Selektionsmechanismen auch mit der Prognose assoziiert sind, und sie werden sich auch kaum durch statistische Bemühungen der Adjustierung oder des Matching ausgleichen lassen, weil die selektionsbildende Merkmalsausprägung quasi definitionsgemäß in den zu vergleichenden Gruppen (zumindest nahe) 100 % bzw. 0 % beträgt.

Für den Fall, dass zwar mehr oder weniger starke Präferenzen existieren, die aber nicht die zuvor skizzierte extreme Ausprägungsverteilung haben, sind Hybrid-Designs aus randomisierten und nicht randomisierten Vergleichen vorstellbar (z. B. [32]). Sollten beide Vergleiche zu ähnlichen Ergebnissen führen, würde das die Sicherheit erhöhen.

Logistische Aspekte

2 Szenarien könnten gegen eine Randomisierung sprechen:

- Die Durchführung einer adäquaten zentralen Randomisierung mit Wahrung des Concealment ist zu aufwendig. Eine solche Situation ist in westlichen Versorgungssystemen mit ihren Infrastrukturen nicht denkbar. Und im Prinzip lassen sich adäquate Randomisierungen auch durch einfachste Methoden erzielen, z. B. die Nutzung von blickdichten Briefumschlägen mit darin enthaltener Zuteilung. Dennoch mag es nicht ausgeschlossen sein, dass es unter schwierigsten Bedingungen nicht gelingt, eine solche Randomisierung durchzuführen. Dann kommt als nächstbeste Methode die alternierende Zuteilung (z. B. nach Anfangsbuchstaben des Namens, nach Geburtstag oder nach Wochentag) infrage. Auch solche Zuteilungsmethoden wurden früher als randomisiert bezeichnet, was aber definitiv falsch ist. Zum einen sind nämlich weder Namen noch Geburtstage zufällig verteilt (und man muss sich bei einer entsprechenden Studie genau mögliche Systematiken und damit verbundene Selektionsmechanismen überlegen, z. B. sind „Montagspatienten“ andere als „Dienstagspatienten“), zum anderen ist aber auch die Zuteilungsverdeckung aufgehoben, sodass ein selektionierter Nichteinschluss von Patienten mit resultierendem Selektionsbias erfolgen kann.
- Eine andere vorstellbare Situation ist, dass eine bestimmte Intervention aus verschiedenen Gründen (z. B. mengenmäßig oder auch wegen mangelnder Expertise bei der Anwendung) für Studienzwecke nicht ausreichend verfügbar ist und auch nicht kurzfristig verfügbar gemacht werden kann. Dann kann über einen Vergleich von Interventionsgruppen aus Regionen, in denen die Intervention verfügbar ist, mit solchen, in denen die Intervention nicht verfügbar ist (und die betroffenen Patienten eine entsprechende Kontrollintervention erhalten), nachgedacht werden. Wenngleich der regionale Unterschied und damit ein möglicher Selektionsbias letztlich nicht ausgeschlossen werden kann, müssen bei einer solchen Studie dennoch alle anderen Maßnahmen getroffen werden, um weitere Verzerrungsquellen zu minimieren. Beispielsweise müssen dann die Begleitmaßnahmen so standardisiert durchgeführt und deren standardisierte Durchführung überwacht werden, dass ein Durchführungsbias vermieden wird. Ähnliches gilt für die Erhebung von Endpunkten (Entdeckungsbias). Dem selektiven Einschluss bzw. Nichteinschluss von Patienten kann nur durch eine Registrierung (im Idealfall) aller infrage kommenden Patienten begegnet werden. Das bedeutet, dass dafür Erkrankungsregister (und nicht Prozedurenregister!) benötigt werden [10]. Bei seltenen, insbesondere bei sehr seltenen Erkrankungen sollte das einfacher zu realisieren sein als bei häufigeren Erkrankungen. In der Regel werden derartige Studien aber deutlich aufwendiger sein als einfache (!) randomisierte Studien. Auch in einer solchen Situation wäre ggf. ein Hybrid-Design aus randomisierten und nicht randomisierten Vergleichen denkbar.

5.1.3 Situation bei seltenen Erkrankungen

Es ist offensichtlich, dass es immer schwieriger wird, eine für die Beantwortung einer spezifischen Fragestellung ausreichende Fallzahl zu erreichen, je seltener diese Erkrankung ist. Somit wird die Ergebnissicherheit bei Studien zu seltenen Erkrankungen vor allem im Hinblick auf die Präzision (2. Komponente der Ergebnissicherheit) beeinträchtigt. Und dies hat u. a. Auswirkungen auf die Möglichkeiten einer Studie, einen Unterschied zwischen den Behandlungsgruppen nachzuweisen, wenn er tatsächlich existiert. Dies wird auch als die Macht oder Trennschärfe (Power) einer Studie bezeichnet. Einen Unterschied nicht zu finden, obwohl er existiert, kennzeichnet den sogenannten Fehler 2. Art, sozusagen ein „falsch-negatives“ Studienergebnis. Die entsprechende Fehlerwahrscheinlichkeit hängt von dem tatsächlichen Unterschied (von der „Wahrheit“) ab und kann somit nur bedingt über die durch Annahmen über diese „Wahrheit“ begründete Fallzahl kontrolliert werden. Der andere wesentliche Fehler, der begangen werden kann, nämlich fälschlicherweise einen Unterschied anzunehmen, obwohl er gar nicht existiert (Fehler 1. Art), kann demgegenüber auch bei kleinen Fallzahlen über die Festlegung des Signifikanzniveaus zumindest theoretisch kontrolliert werden. Für Studien ohne Verzerrungspotenzial – ebenfalls eine theoretische Annahme – sind diese beiden Fehlermöglichkeiten bei gegebener Fallzahl Gegenspieler, d. h., je geringer der Fehler 1. Art ist, desto größer ist der Fehler 2. Art (oder desto geringer ist auch die Power), und umgekehrt. Daneben hängen sie lediglich vom Zufall ab.

Die beiden anderen beeinflussbaren Komponenten der Ergebnis- bzw. Aussagesicherheit (interne und externe Validität) sind demgegenüber nicht quantifizierbar und können zu systematischen Fehlern (Abweichungen von der „Wahrheit“) führen. Diesbezügliche Schwächen im Design einer Studie lassen sich insbesondere im Hinblick auf das Verzerrungspotenzial in aller Regel nicht durch noch so ausgefeilte statistische Methoden ausgleichen. Insofern kommt den grundlegenden Schutzmechanismen im Studiendesign, um mögliche Verzerrungen einzudämmen oder gar auszuschalten, so eine große Bedeutung zu. Und deshalb ist es im Kontext der Bewertung medizinischer Interventionen bei seltenen Erkrankungen naheliegend, zunächst nach Möglichkeiten zu suchen, die Beschränkungen durch die ggf. erniedrigte Präzision durch optimierte statistische Verfahren auszugleichen. Hierfür existiert in der Literatur eine Reihe von Vorschlägen, die in zahlreichen Übersichtsarbeiten beschrieben sind [33-40]. Dabei ist zu betonen, dass in diesen Arbeiten überwiegend die Sorge vor dem fälschlichen Nichtentdecken eines tatsächlich existierenden Interventionseffektes, also eines Fehlers 2. Art, vorherrscht.

Ein wesentlicher Grund für die Schaffung gesetzlicher Anreize für die Entwicklung von Arzneimitteln zur Behandlung seltener Erkrankungen war die bis dato begrenzte bis gar nicht vorhandene Verfügbarkeit von insbesondere krankheitsmodifizierenden Therapien. In einer solchen Situation ist es vorstellbar, dass Patienten leichter zur Teilnahme an einer Studie motiviert werden können, wenn sie Aussicht auf die Gabe einer vielversprechenden Therapie haben. Insofern spielen im Zusammenhang mit seltenen Erkrankungen auch solche Designs

eine besondere Rolle, bei denen alle eingeschlossenen Patienten neben einer Kontrollbehandlung auch die neue Therapie erhalten.

5.1.4 Design-Vorschläge für randomisierte Therapiestudien bei seltenen Erkrankungen

Die nachfolgend beschriebenen Verfahren und Designs erhalten zumindest eine gewisse Art der randomisierten Zuteilung. Auf nicht randomisierte Studien wird in Abschnitt 5.1.5 eingegangen. Ausgangspunkt für vorgeschlagene Modifikationen ist das klassische randomisierte Parallelgruppen-Design, bei dem die Patienten auf Basis eines Zufallsmechanismus einer von 2 oder mehr unterschiedlichen Interventionen zugewiesen und über eine vorab definierte Zeit beobachtet werden. Bei einer sogenannten konfirmatorischen Studie ist es darüber hinaus erforderlich, vorab ein (Haupt-)Zielkriterium zu definieren, auf dessen Basis die Ergebnisse zwischen den Interventionsgruppen verglichen und für Entscheidungen (z. B. über die Zulassung von Arzneimitteln) herangezogen werden. Die Beschränkung auf ein (Haupt-)Zielkriterium vermeidet eine Quelle des Problems des multiplen Testens². Unter Einhaltung bestimmter Schutzmechanismen können jedoch (konfirmatorisch) auch mehrere Zielkriterien zur Entscheidungsfindung herangezogen werden. Die Verfahren und Designs lassen sich grob in 5 Gruppen unterteilen, wobei die Abgrenzung nicht scharf gezogen werden kann:

- 1) Solche, die in erster Linie auf eine Reduktion der Varianz zwischen den Beobachtungen abzielen. Die grundlegende statistische Methodik, um auffällige von durch Zufall verursachte Unterschiede zu differenzieren, besteht darin, Gruppenunterschiede im Hinblick auf Ergebnisse zu einem interessierenden Endpunkt ins Verhältnis zur Varianz der (Einzel-)Ergebnisse (zu diesem Endpunkt) zu setzen. Daraus folgt, dass ein Gruppenunterschied umso leichter als statistisch auffällig (= signifikant) aufgedeckt werden kann, je geringer die Varianz ist. Varianzreduktion führt zu einer erhöhten statistischen Power und damit zu einer Fallzahleinsparung.
- 2) Solche, bei denen frühe Ergebnisse in einer Studie zu Anpassungen des Designs führen können (adaptive Designs). Solche Anpassungen können, müssen aber nicht zu einer Fallzahleinsparung führen.
- 3) Solche, bei denen Vorinformationen (außerhalb der Studie) in die statistische Auswertung mit einfließen (Bayes'sche Verfahren). Für den Fall, dass derartige Vorinformationen oder Annahmen optimistisch sind, ist es offensichtlich, dass für die konkrete Studie weniger Patienten benötigt werden.
- 4) Solche, bei denen alle Patienten auch die zu prüfende Therapie erhalten.
- 5) Sonstige Designs

² Das Problem bei multiplen Tests ist, dass sich das Risiko für ein falsch-positives Ergebnis proportional zur Anzahl durchgeführter Tests erhöht.

5.1.4.1 Varianzreduzierende Verfahren und Designs

An erster Stelle können Verfahren genannt werden, um die durch prognostisch relevante Merkmale hervorgerufene Varianz zu verringern. Dies kann beispielsweise durch eine nach diesen Merkmalen stratifizierte Randomisierung erreicht werden, was den zusätzlichen Vorteil der zwangsläufigen Präspezifizierung bietet. Allerdings sind die Möglichkeiten für Stratifizierungen bei kleinen Fallzahlen schnell erschöpft.

Handelt es sich bei den entsprechenden Merkmalen um kontinuierliche Merkmale, für die sich keine Einteilung in wenige Kategorien anbietet (um eine Stratifizierung zu ermöglichen), dann können zumindest bei der statistischen Auswertung solche Merkmale in einem regressionsanalytischen Verfahren varianzmindernd berücksichtigt werden. Varianz kann darüber hinaus auch dadurch reduziert werden, dass durch Mehrfachmessungen (von kontinuierlichen Merkmalen) innerhalb einer Studie der Messfehler (bzw. die zufällige Schwankung) des interessierenden Endpunkts reduziert wird. Bei dem Einsatz solcher Verfahren ist in der Regel nicht mit einer Einschränkung der in Abschnitt 5.1.1 skizzierten Komponenten der Ergebnissicherheit zu rechnen. Bei einer nach prognostisch relevanten Merkmalen stratifizierten Rekrutierung ist jedoch zu bewerten, ob die Verteilung dieser Merkmale innerhalb der Studie der natürlichen Verteilung entspricht oder ob eine künstliche Verteilung durch Anreicherung von Patienten mit einem bestimmten Merkmal erzeugt wurde. Letzteres ist für die Anwendbarkeit der Ergebnisse bedeutsam, wenn das entsprechende Merkmal einen Effektmodifikator darstellt, d. h. Anzeichen dafür vorliegen, dass sich die Therapieeffekte von Patienten mit diesem Merkmal von Patienten ohne dieses Merkmal unterscheiden.

Eine andere Möglichkeit der Varianzminderung besteht darin, dass die Patienten als ihre eigene Kontrolle dienen. Dahinter steckt die Annahme, dass innerhalb einer Person im Hinblick auf die Auswirkungen der Anwendung einer Intervention weniger Varianz existiert als zwischen verschiedenen Personen. Dabei werden Endpunkte betrachtet, die entweder bei ein und derselben Person an verschiedenen Stellen des Körpers gleichzeitig (z. B. an verschiedenen Körperseiten) oder zu verschiedenen Zeitpunkten (also mehrfach) beobachtet werden können. Ersteres wird beispielsweise in der Dermatologie zur Testung von Externa oder auch in der Augenheilkunde verwendet (mit randomisierter Zuteilung der zu vergleichenden Interventionen bei einem Patienten zu den Körperseiten). Letzteres charakterisiert ein sogenanntes Cross-over-Design, bei dem ein Patient die zu vergleichenden Interventionen nacheinander in mindestens 2 Perioden in zufälliger Reihenfolge erhält. Zwischen die Perioden wird eine (mehr oder weniger kurze) Auswaschphase eingeschaltet, um Überhangeffekte (carry-over effect) zu vermeiden. Im Extremfall kann ein solches Cross-over-Design bei nur einer einzigen Person als „n-of-1 trial“ durchgeführt werden, bei dem diese Person in zahlreichen Perioden die zu vergleichenden Interventionen in zufälliger Reihenfolge erhält. Hier besteht das Ziel darin, für diese eine Person die für sie beste Alternative zu finden. Es gibt jedoch Vorschläge zu Meta-Analysen von n-of-1 trials, um generalisierbare Aussagen daraus ableiten zu können [41].

Cross-over-Studien haben die augenfällige Restriktion, dass nur Endpunkte betrachtet werden können, die vollständig reversibel sind. Damit sind ihre Einsatzmöglichkeiten erheblich begrenzt, und es ergeben sich Einschränkungen hinsichtlich der Patientenrelevanz der Endpunkte, insbesondere auch im Hinblick auf langfristige Effekte einer Intervention. Ein Ausfall von Patienten während der Studie führt darüber hinaus zu größeren Problemen bei der Auswertung.

Grundsätzlich führen komplexere statistische Modelle einerseits bei Gültigkeit der zugrundeliegenden Annahmen zu präziseren Ergebnissen, andererseits jedoch bei Abweichungen von diesen Annahmen zu einer eingeschränkten Robustheit und Validität [11].

5.1.4.2 Sequenzielle und adaptive Designs

Sequenzielle Designs erlauben eine (statistische) Mehrfachtestung der interessierenden Hypothese(n) innerhalb einer Studie. Im einfachsten Fall eines gruppensequenziellen Designs wird dabei zwar eine Studie in zunächst konventioneller Weise basierend auf einem angenommenen Interventionseffekt mit einer fixen Fallzahl geplant; allerdings werden im Verlauf der Studie vorab festgelegt eine oder mehrere Zwischenanalysen durchgeführt, die bei entsprechenden Ergebnissen (Nachweis eines Unterschieds in einem oder mehreren interessierenden Zielgrößen) einen vorzeitigen Abbruch der Studie ermöglichen, mit einer daraus gegenüber der ursprünglichen Planung resultierenden verminderten Fallzahl. Dabei wird der insgesamt einzuhaltende Fehler 1. Art in geeigneter Weise auf die verschiedenen (Zwischen-)Analysen verteilt. Ein solches Design kann dahin gehend erweitert werden, dass auch Abbruchregeln für den Fall festgelegt werden, dass keine Aussicht auf Erreichen des Studienziels (z. B. Nachweis eines Unterschiedes) mehr angenommen wird (futility). Bei Einsatz eines Sequenzialplans können Zwischenanalysen zu jedem beliebigen Zeitpunkt und in beliebiger Anzahl erfolgen, bei binären Endpunkten beispielsweise bei Eintritt eines jeden Ereignisses.

Letztlich kann bei sequenziellen Verfahren gänzlich auf eine Fallzahlplanung verzichtet werden, was aber in gewisser Weise das eigentliche Ziel, nämlich die Fallzahleinsparung, konterkarieren kann, da solche Studien ohne Konvergenz der Abbruchkriterien theoretisch bis ins Unendliche fortgeführt werden können. Darüber hinaus führen sequenzielle Designs auch nur dann zu einer (relevanten) Fallzahleinsparung, wenn die für die jeweiligen Analysen heranzuziehenden Endpunkte frühzeitig im Verhältnis zur Rekrutierungsdauer erhoben werden (können). Anderenfalls können sie lediglich die notwendige Studiendauer reduzieren. Simulationen haben gezeigt, dass – wenn die genannten Bedingungen gegeben sind – die durch den Einsatz sequenzieller Verfahren im Durchschnitt erzielbaren Einsparungen gegenüber einer fixen Fallzahlfestlegung bis zu 50 % betragen können [42].

Es sollte ergänzt werden, dass auf Basis empirischer Beobachtungen gemutmaßt wurde, dass Studien, die aufgrund eines positiven Ergebnisses vorzeitig abgebrochen wurden („stopping early for benefit“), zu einer systematischen Verzerrung führen [43,44]. Wenngleich dies sicherlich für Studien zutrifft, bei denen ein solcher Abbruch nicht durch adäquate statistische

Methoden kontrolliert wird, ist das Ausmaß der möglichen Verzerrung bei einem Vorgehen *lege artis* aus statistisch-methodischer Sicht vernachlässigbar [45-47]. Dennoch bleiben Aspekte, die durchaus für ein Verzerrungspotenzial sprechen, insbesondere der Umstand, dass ein besonders großer beobachteter Unterschied in einer Studie – und je größer der beobachtete Unterschied, desto frühzeitiger der Abbruch – eine nachfolgende Replikation erschwert, wenn nicht gar unmöglich macht [45,48]. Somit würden also mit einer gewissen Systematik Studien verhindert, die einen geringeren Effekt zeigen. Darüber hinaus kann ein frühzeitiger Abbruch dazu führen, dass Effekte nicht mehr so präzise geschätzt werden können, dass verlässliche Aussagen zur Größenordnung eines Effekts möglich sind (3. Komponente der Ergebnissicherheit). Eine weitere Quelle möglicher Verzerrungen besteht in einem Bekanntwerden von Ergebnissen von Interim-Analysen, die zu Modifikationen im weiteren Studienverlauf z. B. durch eine veränderte Rekrutierung führen können [49].

Während sich die „Adaptation“ bei sequenziellen Designs auf die Fallzahl beschränkt (und dies auch nur bei einem frühzeitigen Abbruch der Studie), ermöglichen adaptive Designs (auch flexible Designs genannt) im eigentlichen Sinne eine weitere Flexibilisierung, indem bestimmte Planungsannahmen (z. B. Annahmen über die zu erwartende Varianz) im Studienverlauf sowohl auf Basis interner als auch externer Informationen³ überprüft werden und das Studiendesign eben nicht nur hinsichtlich der Fallzahl, sondern auch nahezu aller anderen Aspekte (z. B. Zielpopulation, Endpunkte, Auswahl von Vergleichsinterventionen) entsprechend angepasst werden kann [38,50,51]. Dabei werden in der Regel die Phasen der Studie (vor und nach der Anpassung) zunächst getrennt ausgewertet, dann aber wieder in geeigneter Weise zusammengefasst [52]. Je gravierender die Änderung im Studiendesign ist, desto schwieriger kann hinterher die Interpretation und Anwendbarkeit der Ergebnisse werden, z. B. bei einem Wechsel von Endpunkten (4. Komponente der Aussagesicherheit). Solche adaptiven bzw. flexiblen Designs werden vorzugsweise in den frühen Phasen der Entwicklung einer Intervention eingesetzt. Die zuvor beschriebenen sequenziellen Designs können als ein Spezialfall von adaptiven Designs angesehen werden. Die Bezeichnung „Design“ impliziert, dass die Möglichkeit der Anpassungen und das genaue Vorgehen, wie man dazu kommt (z. B. Art und Zahl von Zwischenanalysen), vorab geplant und im Studienprotokoll beschrieben sind. Dies dient dem Zweck, die Integrität der Studie und die Validität der (End-)Ergebnisse zu wahren.

5.1.4.3 Bayes'sche Verfahren

Schon recht früh wurde die Anwendung Bayes'scher Verfahren als möglicher Königsweg („a way out of a conundrum“) aus dem Dilemma der kleinen Fallzahlen bei seltenen Erkrankungen vorgeschlagen [53]. Bayes'sche Verfahren und die (insbesondere für regulatorische Fragen) üblichen statistischen (frequentistischen) Methoden (klassische statistische Signifikanztests, Konfidenzintervalle) unterscheiden sich in ihrer grundlegenden Philosophie [54,55]:

³ Die meisten Definitionen für adaptive Designs beinhalten allerdings lediglich interne Informationen (z. B. [50]).

- Bei den frequentistischen Methoden wird davon ausgegangen, dass „die Wahrheit“ fix ist und in Gestalt einer eindeutigen Hypothese formuliert werden kann (z. B.: Intervention A hat gegenüber Intervention B keinen positiven Effekt). Nur die Beobachtungen z. B. im Rahmen eines Experiments unterliegen einem Zufallsmechanismus, der durch eine Wahrscheinlichkeitsverteilung beschrieben wird. Sind die Abweichungen der Beobachtungen von der ursprünglichen Hypothese (auch Nullhypothese genannt) unter Berücksichtigung dieser Wahrscheinlichkeitsverteilung zu extrem, wird die Hypothese verworfen und die Alternative angenommen (Intervention A hat gegenüber Intervention B einen positiven Effekt), allerdings mit dem Vorbehalt einer Irrtumswahrscheinlichkeit (Signifikanzniveau). Die Wahrscheinlichkeitsaussage bezieht sich also nicht auf die Richtigkeit der Hypothese, sondern auf die Abweichung der Beobachtungen von der Nullhypothese. Bei diesen Überlegungen spielt es keine Rolle, ob statistische Tests oder Konfidenzintervalle als frequentistische Methode eingesetzt werden.
- Demgegenüber wird bei der Anwendung von Bayes'schen Verfahren die Ausgangshypothese selbst mit einer Wahrscheinlichkeit (und einer entsprechenden Verteilung) belegt (z. B.: Die Wahrscheinlichkeit, dass Intervention A gegenüber Intervention B keinen positiven Effekt hat, beträgt im Mittel 30 % und folgt einer bestimmten Wahrscheinlichkeitsverteilung). Diese wahrscheinlichkeitsbasierte Ausgangshypothese wird a priori Wahrscheinlichkeit genannt (kurz: Prior). Die Beobachtungen z. B. im Rahmen eines Experiments dienen dann dazu, je nach Abweichungen von dieser Ausgangshypothese, diese anzupassen und in eine A-posteriori-Wahrscheinlichkeitsaussage zu überführen.

Der zentrale Unterschied zwischen frequentistischen und Bayes'schen Verfahren besteht also darin, dass bei den frequentistischen Methoden die Bewertung der beobachteten Daten ohne Bezug zu irgendeinem Vorwissen (über die Richtigkeit der Ausgangshypothese) erfolgt und die Daten für sich allein stehen (müssen). Bei Bayes'schen Verfahren kann in gewisser Weise aus dem Vorwissen „gelernt“ werden. Deshalb sind sie auch bedeutsam im Zusammenhang mit adaptiven Designs. Allerdings lassen sich Bayes'sche Verfahren auch anwenden, wenn das Vorwissen quasi auf null gesetzt wird: Man spricht dann von einem nicht informativen Prior. Bei der Verwendung von einem nicht informativen Prior führen Bayes'sche und frequentistische Verfahren bei gleicher Datengrundlage zu nahezu identischen Ergebnissen, jedenfalls was Entscheidungsgrenzen angeht [56].

Wenn jedoch der eigentliche Gedanke der Bayes'schen Verfahren aufgegriffen und ein informativer Prior verwendet wird, dann können diese selbstverständlich präzisere Ergebnisse liefern, da das Vorwissen genutzt werden kann und sich Aussagen und daraus abgeleitete Entscheidungen nicht mehr allein auf die beobachteten Daten stützen (müssen). Die Krux dabei ist, dass die Festlegung der informativen Prior subjektiv und letztlich auch spekulativ ist. In die resultierenden Ergebnisse fließen also Annahmen über die tatsächlichen Effekte (bzw. deren Verteilung) ein. Und gleiche Daten führen bei unterschiedlichem Prior zu unterschiedlichen Ergebnissen. Das ist insbesondere für administrative Entscheidungen eine

unbefriedigende Situation, weshalb die Anwendung von Bayes'schen Verfahren z. B. im Rahmen der Zulassung von Arzneimitteln von den zuständigen Behörden zumindest in der Vergangenheit recht reserviert gesehen wurde [11,57]. Je besser die Datenbasis für die Prior ist (z. B. Informationen aus einer systematischen Übersicht von RCTs hoher Ergebnissicherheit), umso überzeugender werden die Ergebnisse einer Bayes'schen Analyse mit informativen Prior aufgenommen werden können. Darin wird aber im Zusammenhang mit seltenen Erkrankungen gerade das Problem liegen, weil hier ja typischerweise solche hochwertige Evidenz nicht vorliegt und die Übertragbarkeit von Studienergebnissen aus einer anderen (häufigeren) Indikation wieder ein subjektives und spekulatives Moment bietet.

5.1.4.4 Designs, bei denen alle Patienten (auch) die Prüflintervention erhalten

Den nachfolgend aufgeführten Designs, die teilweise unter anderen Namen kursieren, ist gemein, dass sie weniger auf eine Fallzahleinsparung abzielen als vielmehr darauf, alle eingeschlossenen Patienten im Verlauf einer entsprechenden Studie auch der experimentellen Intervention zuzuführen bzw. umgekehrt die Zeit unter der Kontrollintervention (häufig Placebo) möglichst kurz zu halten. Das markiert dann auch gleichzeitig den Nachteil dieser Designs, dass nämlich die Phase, in der in einem parallel kontrollierten (randomisierten) Vergleich Aussagen zu Effekten möglich sind, vergleichsweise kurz ist und dass z. B. Fragen der Langzeitsicherheit oder zu später sich manifestierender Morbidität oder Mortalität nicht hinreichend sicher beantwortet werden können. Damit ist dann auch die Anwendbarkeit der Ergebnisse eingeschränkt.

Bei einem delayed start bzw. randomized placebo-phase design werden die Patienten zunächst randomisiert auf die experimentelle oder die Kontrollintervention (diese muss nicht zwangsläufig Placebo sein) zugeteilt. Nach einem einzigen fixen Zeitpunkt wechseln die Patienten im Kontrollarm auf die experimentelle Intervention. Eine ursprüngliche Rationale für dieses Design war es, bei chronischen, sich progredient verschlechternden Erkrankungen krankheitsmodifizierende von rein symptomatisch wirkenden sowie gemischt krankheitsmodifizierend und symptomatisch wirkenden Therapien unterscheiden zu können [58]. Bei Zulassungsstudien von Arzneimitteln wird häufig der ursprünglich randomisierten Studie eine sogenannte Open-Label-Extensionsphase für alle Patienten (oder auch nur für diejenigen, die das wünschen) mit dem Prüfpräparat angeschlossen. Das ist letztendlich nichts wesentlich anderes.

Bei einem early escape design erhalten die initial der Kontrollintervention zugewiesenen Patienten nicht zu einem fixen Zeitpunkt, sondern bei Erreichen eines ungünstigen Status oder Ereignisses die experimentelle Intervention. In der Onkologie wird ein solches Design bzw. eine Variante bevorzugt bei fortgeschrittenen Erkrankungen eingesetzt: Bei Feststellung einer Progression können die Patienten auf die experimentelle Intervention wechseln. Dies wird auch häufig als „Cross-over“ bezeichnet, darf aber nicht mit den eigentlichen, zu Beginn dieses Abschnitts erwähnten Cross-over-Studien verwechselt werden. Eine günstigere Bezeichnung ist deshalb „Treatment switching“. Ein solcher Behandlungswechsel ist

grundsätzlich in beide Richtungen möglich, also sowohl von Kontroll- auf die experimentelle Intervention als auch von experimenteller auf die Kontrollintervention. Letzteres passiert aber vergleichsweise selten.

Bei einem stepped wedge design erfolgt die (ggf. randomisierte) Zuteilung auf mehrere verschiedene Startzeitpunkte mit der experimentellen Intervention (z. B. sofort, nach 3-monatiger Kontrollbehandlung und nach 6-monatiger Kontrollbehandlung). Ein solches Design wird insbesondere für die Evaluation von bevölkerungsbezogenen Maßnahmen empfohlen, bei denen die Implementation ohnehin nicht flächendeckend gleichzeitig erfolgen kann [59].

Bei einem randomized withdrawal design schließlich erhalten – umgekehrt – zunächst alle Patienten die experimentelle Intervention. Nach einem bestimmten Zeitpunkt erfolgt die Randomisierung auf Fortführung der experimentellen oder Wechsel auf die Kontrollintervention. Teilweise wird die randomisierte Phase auf diejenigen beschränkt, die auf die experimentelle Therapie im Hinblick auf ein bestimmtes Kriterium angesprochen haben (Responder) [60]. In diesem Fall muss allerdings die beschränkte Übertragbarkeit des eigentlich interessierenden Vergleichs (zwischen experimenteller und Kontrollintervention) beachtet werden, da hier eben nur Patienten betrachtet werden, die auf die experimentelle Intervention angesprochen haben. Dies entspricht jedoch nicht der Entscheidungssituation, in der das Ansprechen in aller Regel nicht vorhergesagt werden kann. Ein weiteres Problem eines randomized withdrawal designs besteht darin, dass der interessierende Endpunkt, nämlich die Verschlechterung nach der Randomisierung in der Kontrollgruppe, eher Ausdruck einer Toleranzentwicklung gegenüber der in der ersten Phase gegebenen experimentellen Intervention ist und wenig bis nichts mit deren tatsächlicher Effektivität zu tun hat.

Weiterhin ist eine Kombination dieser Designs vorstellbar, so wurde 2001 ein Three-stage Design vorgeschlagen, das eine Kombination aus einem early escape und einem randomized withdrawal design darstellt [61]: Nach einer initialen Phase des randomisierten Vergleichs der experimentellen mit der Kontrollintervention treten die Responder aus dem experimentellen Arm in eine zweite Phase, in der randomisiert auf Fortführung oder Wechsel auf die Kontrollintervention randomisiert wird (randomized withdrawal). Die Non-Responder aus dem Kontrollarm der ersten Phase erhalten alle die experimentelle Intervention. Und diejenigen, die dann ansprechen, treten in der dritten Phase wieder in ein randomized withdrawal design ein. In der ersten Phase bildet der Response, in den beiden folgenden Phasen der Relapse das Zielkriterium. Die Ergebnisse der 3 Phasen werden schließlich statistisch adäquat kombiniert. Da hier bestimmte Patienten doppelt in Vergleichen betrachtet werden (können), ist das Design effizient und kann – je nach der tatsächlichen Effektivität der experimentellen Intervention – zu moderaten (25 % bis 30 %) bis erheblichen (60 %) Fallzahleinsparungen führen.

5.1.4.5 Sonstige Designs

Einen weiteren Spezialfall stellen Studien mit einer adaptiven Randomisierung dar, bei denen das Zuteilungsverhältnis, das (bei 2 zu vergleichenden Interventionen) zunächst häufig 1:1 beträgt, auf Basis der Beobachtungen bei zuvor eingeschlossenen Patienten modifiziert wird [39]. Ziel der adaptiven Randomisierung ist es, die Zahl der Patienten, die (allein) die unterlegene Therapie erhalten, möglichst gering zu halten. Dieses Ziel wird umso besser erreicht, je effektiver die zu prüfende Intervention ist.

Bei einem „play-the-winner“-Design wird die Zuteilungswahrscheinlichkeit für diejenige Intervention, für die zuvor häufiger positive Ereignisse (Response) beobachtet wurden, erhöht – oder umgekehrt: für diejenige Intervention, für die zuvor häufiger negative Ereignisse (z. B. Todesfälle) beobachtet wurden, erniedrigt –, bis hin zu einer nahezu deterministischen Zuteilung. Eine solche Studie wird beendet, wenn eine zuvor festgelegte Anzahl von Patienten hintereinander in einer Gruppe ein positives oder auch ein negatives Resultat gezeigt haben. Die statistische Analyse von Daten aus einer Studie mit adaptiver Randomisierung basiert nicht auf Standardverfahren und kann sehr komplex werden [11].

Bei der ersten Anwendung eines solchen Designs mit dem Endpunkt Mortalität wurden 11 Patienten der experimentellen Intervention und nur ein Patient der Kontrollintervention zugewiesen [62]. Der eine Patient im Kontrollarm war verstorben, während dies bei keinem Patienten der experimentellen Therapie beobachtet wurde. Dies Ergebnis zeigt die Grenzen eines solchen Designs: Die Studie hatte nunmehr praktisch den Status einer Fallserie, und ihr Ergebnis wurde in der Fachwelt umstritten aufgenommen. Eine 11 Jahre später veröffentlichte Studie im konventionellen Parallelgruppen-Design bestätigte die Überlegenheit der experimentellen Therapie, allerdings nun mit einem deutlich geringeren Unterschied, dafür aber besser interpretierbaren Ergebnis [63]. Das Risiko für einen solchen Extremfall kann allerdings verringert werden, indem dafür Sorge getragen wird, dass die Zuteilungswahrscheinlichkeiten zu Beginn nicht zu stark divergieren.

Der Einsatz einer adaptiven Randomisierung wurde in der Vergangenheit und insbesondere in dem oben dargestellten Beispiel auch wegen ethischer Überlegungen propagiert: Bei Vorliegen von Indizien oder auch nur Überzeugungen einer besonders (quasi „dramatisch“) ausgeprägten Überlegenheit sollte nur möglichst wenigen Patienten die mutmaßlich unterlegene Therapie vorenthalten werden. Eine solche Argumentation ist allerdings nicht überzeugend, da das ethische Dilemma letztlich nicht über die Anzahl der Betroffenen gelöst wird. Das mag ein wesentlicher Grund – neben anderen – dafür sein, dass Studien mit einer adaptiven Randomisierung bis heutzutage Raritäten sind.

Zuteilungsregeln, die auf dem Ergebnis zuvor beobachteter Patienten basieren, werden auch als responseadaptive Methoden bezeichnet und lassen sich auf nicht randomisierte Zuteilungen erweitern. Solche werden vorwiegend in Studien (Phase 1) zur Findung einer maximal tolerablen Dosis eingesetzt.

5.1.5 Ergebnissicherheit nicht randomisierter Studien

Verzerrungen

Nicht randomisierte Studien bergen inhärent erhebliche Einschränkungen bei der internen Validität (1. Komponente der Ergebnissicherheit). Gegenwärtig ist keine Methode bekannt, die in suffizienter Weise diese Einschränkungen aufheben könnte. Noch so ausgefeilte statistische Verfahren können keine Imbalancen zwischen den zu vergleichenden Gruppen aufheben, die auf nicht gemessenen oder unbekanntem Störgrößen beruhen. Es verbleibt somit immer eine gewisse Restunsicherheit. Das heißt aber nicht, dass in einer nicht randomisierten Studie nicht dennoch alle Anstrengungen unternommen werden sollten, um möglichen Verzerrungen aufgrund der in Abschnitt 5.1.1 beschriebenen Mechanismen zu begegnen. Daraus folgt, dass nicht randomisierte Studien in der Regel mit einem deutlich höheren Aufwand verbunden sind als RCTs, da die Randomisierung in Verbindung mit dem Einsatz anderer adäquater Designinstrumente auf einfache Weise eine Reihe von Verzerrungsmöglichkeiten ausschaltet.

Je nach Grund für einen eventuellen Verzicht auf eine Randomisierung ergeben sich unmittelbare Konsequenzen für ersatzweise nicht randomisierte Studien (siehe Abschnitt 5.1.2). Werden ethische Bedenken geltend gemacht, verbieten sich jegliche parallel vergleichenden Studien, und es verbleiben im Prinzip nur noch historische Kontrollen.

Historische Kontrollen

Historisch kontrollierte Studien stehen in den gängigen Evidenzhierarchien weit unten. Das sind einarmige Studien oder Fallserien, bei denen die Patienten die experimentelle Intervention erhalten und über einen gewissen Zeitraum beobachtet werden. Dabei erfolgt der Vergleich mit der Kontrollintervention nicht zeitlich parallel, sondern implizit oder explizit mit den Erfahrungen aus der Vergangenheit. Neben den üblichen Verzerrungsmöglichkeiten (z. B. Selektionsbias) kommt hier als weitere Störgröße der Faktor Zeit oder Chronologie dazu, z. B. geänderte Indikationsstellungen, Stadienverschiebung oder Begleitbehandlungen. Ähnlich wie bei den o. g. (starken) Überzeugungen oder Präferenzen lässt sich dies kaum durch statistische oder andere Verfahren ausgleichen. Ergebnisse aus historisch kontrollierten Studien lassen Aussagen im Sinne eines Interventionseffektes nur dann zu, wenn (nahezu) eine Umkehr eines quasideterministischen Verlaufs beobachtet wurde, wofür wiederum die genaue Kenntnis dieses natürlichen Verlaufs erforderlich ist [11]. Hierfür wurde (auch) der Begriff eines „dramatischen Effektes“ geprägt. Dies ist terminologisch nicht ganz sauber, da die Bezeichnung „Effekt“ besser dem kausal begründeten Unterschied, der „Wahrheit“, vorbehalten bleiben und besser von „dramatischem Unterschied“ gesprochen werden sollte. Eine etwas andere (ggf. schwächere) Operationalisierung liegt dem „Alles-oder-nichts-Kriterium“ zugrunde: Während in der Vergangenheit alle Patienten mit einer bestimmten Erkrankung z. B. verstarben und nach Anwendung einer (neuen) Intervention nun einige überleben, könne auch das als kausal begründet akzeptiert werden [10]. Ein gewisser Nachteil dieses Kriteriums besteht darin, dass die dafür notwendige Vorbedingung, nämlich die hundertprozentige Eintrittswahrscheinlichkeit, nicht wirklich nachgewiesen werden kann. Es

muss eben eine gewisse Relativierung des Deterministischen akzeptiert werden, was letztlich eher wieder in die Richtung eines „dramatischen Effekts“ führt.

„Dramatischer Effekt“

Unabhängig von der Terminologie bleibt die Frage, wann ein beobachteter Unterschied als so groß angesehen werden kann, dass er nicht mehr allein auf Bias zurückgeführt werden kann. Auf Basis ein paar weniger Simulationsuntersuchungen und eigener Überlegungen wurde dafür von Glasziou et al. [25] der Faktor 10 vorgeschlagen: Das heißt, wenn z. B. die Erfolgsrate einer neuen Intervention 10-mal höher ist als die Erfolgsrate der Kontrollintervention, dann sei nicht mehr zu erwarten, dass dieser Unterschied allein auf Bias zurückgeführt werden könne. Darüber hinaus sei ein solches Ergebnis mit einem (adäquaten) statistischen Test (mit üblicher Nullhypothese, also zum Nachweis irgendeines Unterschieds > 0) zum Irrtumsniveau 1 % abzusichern [25]. Bemerkenswerterweise unterschieden die Autoren bei dieser Festlegung nicht zwischen kontrollierten und unkontrollierten (bzw. „historisch“ kontrollierten) Studien, sondern bezogen sich allein auf das (Nicht-)Erfordernis einer Randomisierung. Für eine differenzierte Festlegung von entsprechenden „Grenzwerten“ entlang der Evidenzhierarchie (jenseits von RCT) gibt es somit bisher keine Vorschläge.

Sprechen logistische Gründe gegen eine Randomisierung, können alternative Zuteilungsregeln erwogen werden (siehe Abschnitt 5.1.2); immerhin verbleibt aber das Charakteristikum einer prospektiv geplanten Interventionsstudie mit zeitlich parallelen Kontrollen.

Beobachtungsstudien

Von prospektiv geplanten Interventionsstudien werden Beobachtungsstudien abgegrenzt, deren Sinn und Zweck entweder weniger der Frage nach den Effekten von Interventionen nachgehen, sondern vielmehr nach der Verteilung und den Ursachen von Krankheiten (epidemiologische Kohorten- oder Fall-Kontroll-Studien), oder deren Datengrundlage primär nicht der Gewinnung von medizinisch-klinischen Fragen dient, sondern allenfalls der Dokumentation (Routinedaten oder Sekundärdaten). In den gängigen Evidenzhierarchien werden sie deshalb in ihrer Aussagekraft den randomisierten und nicht randomisierten Interventionsstudien nachgeordnet. Nicht immer erfolgt allerdings eine solche scharfe Trennung zwischen Interventions- und Beobachtungsstudien.

Aus der divergierenden Zweckbestimmung folgen die inhärenten Nachteile derartiger Studien: Ein Schutz vor den einen Interventionseffekt möglicherweise verzerrenden Mechanismen ist oft kaum gegeben. Die arztgesteuerte oder Selbstselektion beruht zumeist auf prognostisch bedeutsamen Merkmalen, die nicht erfasst wurden oder erst gar nicht bekannt sind. Ein Durchführungsbias ist nicht vermeidbar und eine ITT-Auswertung nur bei eindeutiger Definition der Zielpopulation (Vollzähligkeit) und zumeist unter erheblichem Aufwand durchführbar. Routine- oder Sekundärdaten leiden notorisch unter fälschlichen (Fehl-) und fehlenden Dokumentationen (Unvollständigkeit).

Als zentrales Qualitätskriterium solcher Studien bleibt dann nur noch die adäquate Befassung mit möglichen Störgrößen (Confounder) im Rahmen der statistischen Auswertung im Sinne einer Adjustierung oder als Paarbildung (Matching), um zumindest ansatzweise die Problematik bei der internen Validität zu adressieren. Dies sollte idealerweise in Unkenntnis der erhobenen bzw. zu erhebenden Daten vorab geplant werden, inkl. des zu verwendenden statistischen Modells und eventuellen Interaktionen. Bei Studien mit kleinen Fallzahlen stößt aber die Adjustierung (bzw. das Matching) schnell an Grenzen, da nicht beliebig viele prognostische Merkmale in ein entsprechendes statistisches Modell aufgenommen werden können. Näherungsweise kann man sich eine Adjustierung wie eine Schichtung nach den in das Modell aufzunehmenden Merkmalen vorstellen; die Interventionseffekte werden dabei zunächst innerhalb der einzelnen Schichten geschätzt und dann wieder geeignet zusammengefasst. Wenn lediglich binäre prognostische Merkmale betrachtet werden, resultieren bereits 2^n Strata, wobei n die Anzahl Merkmale bezeichnen soll. Es muss also gerade bei Studien mit kleinen Fallzahlen bezweifelt werden, dass eine suffiziente statistische Berücksichtigung der möglichen Störgrößen erfolgen kann.

Beobachtungsstudien werden zumeist in prospektiv und retrospektiv eingeteilt. Dabei wird den prospektiven Studien eine höhere Aussagekraft zugesprochen als den retrospektiven; diese Unterscheidung gelingt aber nicht immer, es existieren fließende Unterschiede mit Studien, die sowohl retro- als auch prospektive Anteile haben. Ein entscheidenderes Gütekriterium ist eher, inwieweit der Studie ein tatsächlicher Plan mit vorformulierten Hypothesen und Strategien der statistischen Auswertung und des Vermeidens bzw. Minimierens von Verzerrungsquellen zugrunde liegt.

5.1.6 Empfehlungen

Orientiert man sich an den in Abschnitt 5.1.1 genannten Komponenten der Ergebnissicherheit, kann daraus keine wissenschaftliche Begründung für eine unterschiedliche Herangehensweise bei der Bewertung von medizinischen Interventionen für seltene und nicht seltene Erkrankungen abgeleitet werden. Studien mit geringer Präzision oder mit nicht ausreichenden Schutzmechanismen vor potenziell verzerrenden Faktoren oder mit Abweichungen der Studienbedingungen von der eigentlich interessierenden Fragestellung (z. B. Verwendung von Surrogatendpunkten anstelle von patientenrelevanten Endpunkten) haben die gleichen Auswirkungen auf die Aussagesicherheit bei seltenen und nicht seltenen Erkrankungen. Dies wird auch mit dem Fokus auf den Stellenwert randomisierter Studien durchgängig in der einschlägigen Literatur so gesehen [1,10,11,31,33,35]. Umgekehrt existieren keine spezifischen Designs und statistischen Methoden für seltene Erkrankungen, die nicht auch relevant für häufige(re) Erkrankungen sein könnten [11].

Im Kontext der Bewertung medizinischer Interventionen bei seltenen Erkrankungen wird es somit (besonders) notwendig sein, ein möglichst effizientes methodisches Vorgehen zu wählen und dabei auch effiziente Strukturen zu schaffen. Darüber hinaus kann es,

insbesondere bei sehr seltenen Erkrankungen, erforderlich sein, Kompromisse bei der Aussagesicherheit einzugehen, die auch auf externen (politischen) Vorgaben beruhen können.

5.1.6.1 Politische Vorgaben

Im Rahmen des Arzneimittelmarktneuordnungsgesetzes (AMNOG) hat der Gesetzgeber in Deutschland politische Vorgaben gemacht, indem der Zusatznutzen von Arzneimitteln für die Behandlung seltener Erkrankungen (Orphan Drugs, gemäß Verordnung (EG) Nr. 141/2000 des Europäischen Parlaments und des Rates vom 16.12.1999) qua Zulassung als belegt gilt, allerdings nur so lange, wie der Umsatz eines solchen Arzneimittels mit der gesetzlichen Krankenversicherung (zu Apothekenverkaufspreisen einschließlich Umsatzsteuer) innerhalb eines Zeitraums von 12 Kalendermonaten einen Betrag von 50 Millionen € nicht übersteigt (§ 35a Abs. 1 Satz 10 SGB V).

In der Begründung für diese Ausnahmeregelung wird einerseits (auch) ein inhaltliches Argument vorgebracht: „Es ist regelmäßig davon auszugehen, dass es für die Behandlung dieser Erkrankung keine therapeutisch gleichwertige Alternative gibt. Der pharmazeutische Unternehmer erbringt daher keine Nachweise zum Nutzen und zum Zusatznutzen im Dossier“ [64]. Dies könnte als Annäherung an einen „dramatischen Effekt“ verstanden werden; immerhin beinhalten die Anforderungen der o. g. EG-Verordnung an die Ausweisung als „orphan drug“ für den Fall, dass es sich um eine nicht seltene Erkrankung handelt (sic! – d. h. eine Erkrankung, von der mehr als 5 von zehntausend Personen betroffen sind), den Passus: „... wenn der Investor nachweisen kann, dass ... b) in der Gemeinschaft noch keine zufriedenstellende Methode für die Diagnose, Verhütung oder Behandlung des betreffenden Leidens zugelassen wurde oder dass das betreffende Arzneimittel — sofern eine solche Methode besteht — für diejenigen, die von diesem Leiden betroffen sind, von erheblichem Nutzen sein wird“ [2].

Andererseits macht die Begründung auch den politischen Kompromiss deutlich: „Diese Erleichterung ist für Arzneimittel angemessen, die aufgrund ihrer Zulassung für seltene Erkrankungen einen geringen Umsatz haben. Erreicht der pharmazeutische Unternehmer mit dem Arzneimittel in der gesetzlichen Krankenversicherung einen Umsatz von mehr als 50 Millionen € in den letzten 12 Kalendermonaten, ist ihm zuzumuten, den Nachweis des Zusatznutzens zu erbringen und hierfür ein vollständiges Dossier vorzulegen“ [64]. Es ist wissenschaftlich nicht zu begründen, warum inhaltliche Kriterien, die – wenn sie denn zutreffen – eine Bewertung (des [Zusatz-]Nutzens) obsolet machen, ab einem gewissen Umsatzvolumen nicht mehr zutreffen (sollen).

5.1.6.2 Strukturen

Bei seltenen Erkrankungen wird es mit abnehmender Häufigkeit der Erkrankung zunehmend schwieriger, der quantitativen Komponente der Ergebnissicherheit ausreichend Rechnung zu tragen. Geht man von der derzeitigen Definition einer seltenen Erkrankung in der Europäischen Union aus, dann wären am Rand dieser Definition (bei einer Häufigkeit von

genau 5 pro 10 000 Personen) in Deutschland etwa 40 000 Menschen, in der gesamten Europäischen Union etwa 250 000 Menschen von einer solchen Erkrankung betroffen. Nähme man weiter an, dass, ausgehend von den Erfahrungen bei der Durchführung klinischer Studien, vielleicht nur 1 % der Betroffenen (auf Basis von spezifischen Ein- und Ausschlusskriterien) geeignet wären und sich dazu bewegen ließen, an einer randomisierten Studie teilzunehmen, lassen sich die Begrenztheiten erkennen, insbesondere dann, wenn zu regional geschaut wird.

Eine diesbezüglich im Prinzip triviale Feststellung lautet daher, dass es für die klinische, patientenorientierte Erforschung seltener Erkrankungen unabdingbar ist, in vernetzten überregionalen und supranationalen Strukturen zu arbeiten [35,65-67]. Daraus lässt sich ein starkes Plädoyer für Krankheitsregister als Unterstützung solcher Strukturen ableiten, die aussagefähige Studien erst ermöglichen. Solche Krankheitsregister müssen – um als Basis für hochwertige klinische, vor allem nicht randomisierte, Studien dienen zu können – klaren Qualitätskriterien bezüglich Vollzähligkeit und Vollständigkeit genügen [10]. Vollzähligkeit meint dabei die Erfassung eines möglich großen Anteils der Betroffenen, um einem Selektionsbias begegnen zu können. Vollständigkeit bedeutet die Erfassung aller relevanten Merkmale von einem möglichst großen Anteil der Patienten im Verlauf, um mögliche Unterschiede in den Begleitumständen erfassen und ggf. bei Interventionsvergleichen berücksichtigen zu können, und natürlich um als (patienten-)relevante Outcome-Kriterien zu dienen. Solche Qualitätsaspekte werden sich in der Regel am ehesten in Zentren umsetzen und gewährleisten lassen. Die Zentrenbildung hätte darüber hinaus den Vorteil, Patienten besser erreichen zu können. Krankheitsregister sollten also in erster Linie dazu genutzt werden, um (auch) die quantitativen Voraussetzungen für aussagefähige klinische Studien zu schaffen. Es gibt Beispiele für eine erfolgreiche Netzwerkbildung, z. B. auf dem Gebiet der zystischen Fibrose [68-70]. Jüngst hat die Deutsche Krankenhausgesellschaft im Rahmen eines Positionspapiers zur Qualitätssicherung interessante Vorschläge zum Thema Register gemacht, beispielsweise bei kleinen Fallzahlen (hier sollen offenbar seltene oder sehr seltene Erkrankungen gemeint sein) eine Verpflichtung aller Leistungserbringer zur Teilnahme [71].

5.1.6.3 Effizienz

Wenn es schon inhärent Beschränkungen bezüglich einer Komponente der Ergebnissicherheit gibt, scheint es wenig zielführend, auch die anderen Komponenten zu kompromittieren. Von den zuvor beschriebenen Strategien wären somit diejenigen zu bevorzugen, die bei weitgehendem Erhalt von interner und externer Validität Fallzahleinsparungen ermöglichen, also sequenzielle Designs. Dabei muss man sich bewusst machen, dass die Einsparungen nur beim Vorliegen großer beobachteter Unterschiede (frühzeitiger Abbruch wegen Überlegenheit) und durch den Verzicht auf das Entdecken kleiner Effekte (frühzeitiger Abbruch wegen Futility) erzielt werden können.

5.1.6.4 Kompromisse bei der Aussagesicherheit

Irrtumsniveau

Bei sehr seltenen Erkrankungen könnte alternativ oder zusätzlich zum gängigen methodischen Vorgehen erwogen werden, ein größeres statistisches Irrtumsniveau für regulative Entscheidungen zuzulassen. Üblicherweise gilt ein Ergebnis dann als „statistisch signifikant“, wenn der Fehler 1. Art zweiseitig (d. h., es sollen Abweichungen in beide Richtungen von der Nullhypothese belegt werden können) 5 % beträgt (da die meisten Fragestellungen dem Grunde nach einseitig sind, beträgt das diesbezügliche Irrtumsniveau nur 2,5 %). Die Ausgangssituation bei seltenen Erkrankungen ist meist ein hoher Bedarf einer neuen Therapie („medical need“) in Kombination mit geringeren bzw. beschränkten Mitteln bzw. Fallzahlen. Ein höherer Bedarf bedeutet, dass der Fehler, eine wirksame Therapie nicht als solche zu erkennen, wichtiger wird im Verhältnis zu dem Fehler, eine unwirksame Therapie als wirksam zu erklären. Dies spiegelt bei gegebener Fallzahl das Abwägen zwischen den Fehlern 1. und 2. Art wider. Daraus kann sich der Kompromiss einer erhöhten Irrtumswahrscheinlichkeit ergeben. Bei einer Anhebung des zweiseitigen Irrtumsniveaus auf 10 % und unter Annahme eines einfachen Modells (t-Test) könnte bei einer Power von 80 % die Fallzahl um gut 20 %, bei einem zweiseitigen Irrtumsniveau von 20 % gar um gut 40 % reduziert werden. Der Vorteil einer solchen Vorgehensweise wäre, die Irrtumsmöglichkeit zumindest quantifizieren zu können. Ein derartiges Vorgehen kann auch als Annäherung an Bayes'sche Verfahren verstanden werden: Ob nämlich die Daten aus einer randomisierten Studie mit geringer Fallzahl mit einem erhöhten Irrtumsniveau zur Entscheidungsfindung herangezogen werden oder ob bei Verknüpfung mit z. B. optimistischem Vorwissen aus einer anderen Indikation (Prior) das übliche Irrtumsniveau beibehalten wird, dürfte auf vergleichbare Ergebnisse im Hinblick auf resultierende Entscheidungen hinauslaufen.

Externe Validität

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. Bei seltenen Erkrankungen wird die Validität von Surrogatendpunkten in der Regel unklar sein, da für die Validierung Meta-Analysen von aussagefähigen Studien erforderlich sind, eine Situation also, die gerade nicht angenommen werden kann. Auch der Einsatz von adaptiven Designs wird in der Regel mit Einschränkungen der externen Validität verbunden sein. Immerhin bliebe aber wenigstens die interne Validität erhalten.

Interne Validität

Die Aufgabe der internen Validität durch Verzicht auf eine Randomisierung stünde in einem solchen hierarchischen Vorgehen aus logischen Gründen an letzter Stelle. Wesentliche Voraussetzung, um sie für (regulatorische) Entscheidungen nutzen zu können, wäre, dass die zugrunde liegenden Daten aus einem Krankheitsregister mit den in Abschnitt 5.1.6.2 genannten Qualitätskriterien stammen oder dass ein derartig großer beobachteter Unterschied vorliegt, der nicht mehr allein durch Bias erklärt werden kann.

5.2 Ergebnisse der empirischen Untersuchung zur Studiengrundlage für die Zulassung von Orphan Drugs in Europa

5.2.1 Ergebnisse der Informationsbeschaffung und resultierender Publikationspool

Die Recherche nach zugelassenen Orphan Drugs in Europa im Orphanet-„Verzeichnis der Arzneimittel für seltene Krankheiten in Europa“ [13] identifizierte 85 Arzneimittel mit europäischer Orphan Drug Designation und europäischer Marktzulassung von 2001 bis 2013. Die Zulassung der 85 Arzneimittel stützte sich auf 125 Hauptstudien (ohne 6 Zulassungen, die auf Literatur-Reviews basierten); darunter waren 82 RCTs. Detaillierte Informationen zu den identifizierten Orphan Drugs werden in Tabelle 6 bis Tabelle 8 (Anhang A) dargestellt. Tabelle 6 zeigt Charakteristika der Zulassung der einzelnen Arzneimittel. In Tabelle 7 werden die zugrunde liegenden Hauptstudien (siehe Abschnitt 4.2.2) und die dort verwendete Methodik charakterisiert. Tabelle 8 fasst die relevanten Informationen auf Arzneimittelzebene zusammen.

5.2.2 Ergebnisse der Informationssynthese und Informationsanalyse

Auswertungen auf Arzneimittelzebene

Alle Zulassungen

Insgesamt erhielten im Zeitraum von 2001 bis 2013 85 Orphan Drugs die europäische Marktzulassung. Bei 14 Arzneimitteln wurde später die Orphan Drug Designation zurückgezogen. 4 Arzneimittel wurden wieder vom Markt genommen. 7 Arzneimittel erhielten eine bedingte Zulassung, bei 24 Arzneimitteln wurde eine Zulassung unter außergewöhnlichen Umständen ausgesprochen.

58 der 85 Arzneimittel (68 %) dienen der Behandlung seltener, 27 Arzneimittel (32 %) der Behandlung sehr seltener Erkrankungen.

Bei 59 Arzneimitteln (69 %) basierte die Zulassung auf RCTs (55 ausschließlich auf RCTs und 4 auf RCTs in Kombination mit Non-RCTs). Bei 20 Arzneimitteln (24 %) dienten Non-RCTs als Grundlage für die Zulassung. Bei 6 Orphan Drugs (7 %) basierte die Zulassung auf Ergebnissen von Studien, die nicht explizit für die Zulassung konzipiert und durchgeführt wurden, sondern im Rahmen eines Literatur-Reviews identifiziert wurden. Bei 2 der 6 Literatur-Reviews waren RCTs enthalten.

Den Zusammenhang zwischen Art der Zulassung und dem Einbezug von RCT-Daten zeigt Tabelle 1. Der Anteil von RCTs ist bei eingeschränkter Zulassung (bedingt bzw. unter außergewöhnlichen Umständen) geringer als bei Zulassungen ohne Einschränkung.

Tabelle 1: Zusammenhang zwischen Einschränkung der Zulassung und Einbezug von RCT-Daten

Berücksichtigung von RCT-Daten	Zulassung [n (%)]		
	ohne Einschränkung	bedingt	außergewöhnliche Umstände
ja	41 (75,9)	4 (57,1)	16 (66,7)
nein	13 (24,1)	3 (42,9)	8 (33,3)

n: absolute Häufigkeit

Abbildung 1 zeigt den Anteil der Orphan Drugs, die unter Berücksichtigung von Daten aus RCTs zugelassen wurden. Es zeigt sich kein deutlicher Trend im Zeitverlauf. Es ist allerdings zu beachten, dass die errechneten Prozentzahlen pro Jahr nur auf wenigen Studien basieren und mit dementsprechender Unsicherheit behaftet sind.

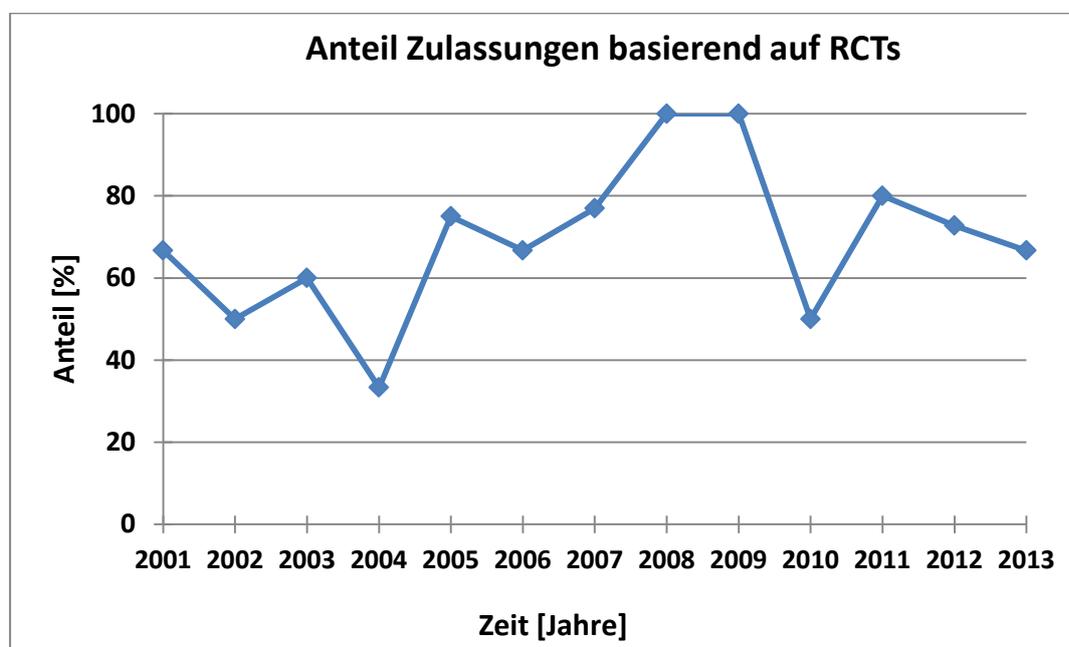


Abbildung 1: Anteil der Orphan Drugs, die auf Basis von RCTs zugelassen wurden, im zeitlichen Verlauf (N = 85)

Zulassungen ohne Literatur-Reviews

Die folgenden deskriptiven Auswertungen basieren auf den 79 Orphan Drugs, deren Zulassung nicht auf Grundlage von Literatur-Reviews erfolgte. Eine detaillierte Darstellung der Ergebnisse zeigt Tabelle 3.

Die Zulassung der Orphan Drugs beruhte auf jeweils 1 bis 5 Hauptstudien. Bei sehr seltenen Erkrankungen waren maximal 3 Studien Basis für die Zulassung (Tabelle 2).

Der Anteil von Zulassungen basierend auf Daten aus RCTs betrug etwa 75 %. Die Patientenzahl der Zulassungsstudien pro Arzneimittel betrug zwischen 27 und 2961 Patienten (Median 165). Etwa 70 % aller Patienten wurden in RCTs behandelt.

Bei 66 Arzneimitteln (84 %) wurden patientenrelevante Endpunkte in den Studien berücksichtigt (bei 31 Orphan Drugs als primärer, bei 57 Arzneimitteln als sekundärer Endpunkt). Der Großteil der Studien wurde multizentrisch, multinational und multikontinental durchgeführt.

In der Mehrzahl der Fälle zeigen sich keine auffälligen Unterschiede zwischen seltenen und sehr seltenen Erkrankungen in den dargestellten Studiencharakteristika. Insbesondere ist der Anteil von Zulassungen basierend auf RCTs bei sehr seltenen Erkrankungen und seltenen Erkrankungen vergleichbar. Die mittlere Patientenzahl in Studien zu sehr seltenen Erkrankungen (Median 96) ist erwartungsgemäß deutlich geringer als in Studien zu seltenen Erkrankungen (Median 213,5). Der Anteil Patienten, die in RCTs behandelt wurden, ist mit 84 % bei sehr seltenen Erkrankungen höher als bei seltenen Erkrankungen mit 67 %. Bei der Verwendung patientenrelevanter Endpunkte zeigen sich tendenzielle Unterschiede. Bei sehr seltenen Erkrankungen werden häufiger Endpunkte zur Morbidität und seltener zur Mortalität ausgewertet als bei seltenen Erkrankungen (Tabelle 3).

Tabelle 2: Anzahl der Hauptstudien pro Arzneimittel

Anzahl Hauptstudien	Erkrankung		
	gesamt (N = 79) n (%)	selten (N = 56) n (%)	sehr selten (N = 23) n (%)
1	44 (55,7)	30 (53,6)	14 (60,9)
2	28 (35,4)	21 (37,5)	7 (30,4)
3	4 (5,1)	2 (3,6)	2 (8,8)
4	2 (2,5)	2 (3,6)	0 (0)
5	1 (1,3)	1 (1,8)	0 (0)

N: Anzahl Zulassungen mit vorhandenen Angaben; n: absolute Häufigkeit

Tabelle 3: Studiencharakteristika (Arzneimittlebene)

Charakteristikum	gesamt (N = 79)			Erkrankung					
				selten (N = 56)			sehr selten (N = 23)		
	Median (unteres; oberes Quartil)			Median (unteres; oberes Quartil)			Median (unteres; oberes Quartil)		
Patientenanzahl	165 (76; 449)			213,5 (117,5; 509)			96 (43; 207)		
Patientenanzahl in RCTs	131 (0; 393)			161 (4; 429,5)			58 (0; 203)		
	n (%)			n (%)			n (%)		
	ja	nein	unklar	ja	nein	unklar	ja	nein	unklar
RCT	59 (74,7)	20 (25,3)	0 (0)	42 (75,0)	14 (25,0)	0 (0)	17 (73,9)	6 (26,1)	0 (0)
multizentrisch	74 (93,7)	4 (5,1)	1 (1,3)	52 (93,9)	3 (5,4)	1 (1,8)	22 (95,7)	1 (4,4)	0 (0)
multinational	59 (74,7)	14 (17,7)	6 (7,6)	41 (73,2)	10 (17,9)	5 (8,9)	18 (78,3)	4 (17,4)	1 (4,4)
multikontinental	49 (62,0)	19 (24,1)	11 (13,9)	35 (62,5)	14 (25,0)	7 (12,5)	14 (60,9)	5 (21,7)	4 (17,4)
Patientenrelevanter EP									
primär oder sekundär									
gesamt	66 (83,5)	11 (13,9)	2 (2,5)	46 (82,1)	8 (14,3)	2 (3,6)	20 (87,0)	3 (13,0)	0 (0)
Mortalität	28 (35,4)	51 (64,6)	0 (0)	23 (41,1)	33 (58,9)	0 (0)	5 (21,7)	18 (78,3)	0 (0)
Morbidität	47 (59,5)	25 (31,7)	7 (8,9)	29 (51,8)	20 (35,7)	7 (12,5)	18 (78,3)	5 (12,7)	0 (0)
Lebensqualität	13 (16,5)	66 (83,5)	0 (0)	11 (19,6)	45 (80,4)	0 (0)	2 (8,7)	21 (91,3)	0 (0)
primär									
gesamt	31 (39,2)	40 (50,6)	8 (10,1)	16 (28,6)	33 (58,9)	7 (12,5)	15 (65,2)	7 (30,4)	1 (4,4)
Mortalität	7 (8,9)	72 (91,1)	0 (0)	5 (8,9)	51 (91,1)	0 (0)	2 (8,7)	21 (91,3)	0 (0)
Morbidität	26 (32,9)	45 (57,0)	8 (10,1)	11 (19,6)	38 (67,9)	7 (12,5)	15 (65,2)	7 (30,4)	1 (4,4)
Lebensqualität	1 (1,3)	78 (98,7)	0 (0)	0 (0)	56 (100)	0 (0)	1 (4,4)	22 (95,6)	0 (0,0)
sekundär									
gesamt	57 (72,2)	20 (25,3)	2 (2,5)	40 (71,4)	14 (25,0)	2 (3,6)	17 (73,9)	6 (26,1)	0 (0)
Mortalität	21 (26,6)	58 (73,4)	0 (0)	18 (32,1)	38 (67,9)	0 (0)	3 (13,0)	20 (87,0)	0 (0)
Morbidität	42 (53,2)	31 (39,2)	6 (7,6)	27 (48,2)	23 (41,1)	6 (10,7)	15 (65,2)	8 (34,8)	0 (0)
Lebensqualität	12 (15,2)	67 (84,8)	0 (0)	11 (19,6)	45 (80,4)	0 (0)	1 (4,4)	22 (95,6)	0 (0)
EP: Endpunkt; N: Anzahl Studien; n: absolute Häufigkeit									

Abbildung 2 zeigt den Anteil der Zulassungen mit Berücksichtigung patientenrelevanter Endpunkte (primär oder sekundär). Es zeigt sich kein deutlicher Trend im Zeitverlauf. Es ist allerdings zu beachten, dass die errechneten Prozentzahlen pro Jahr nur auf wenigen Studien basieren und mit dementsprechender Unsicherheit behaftet sind.

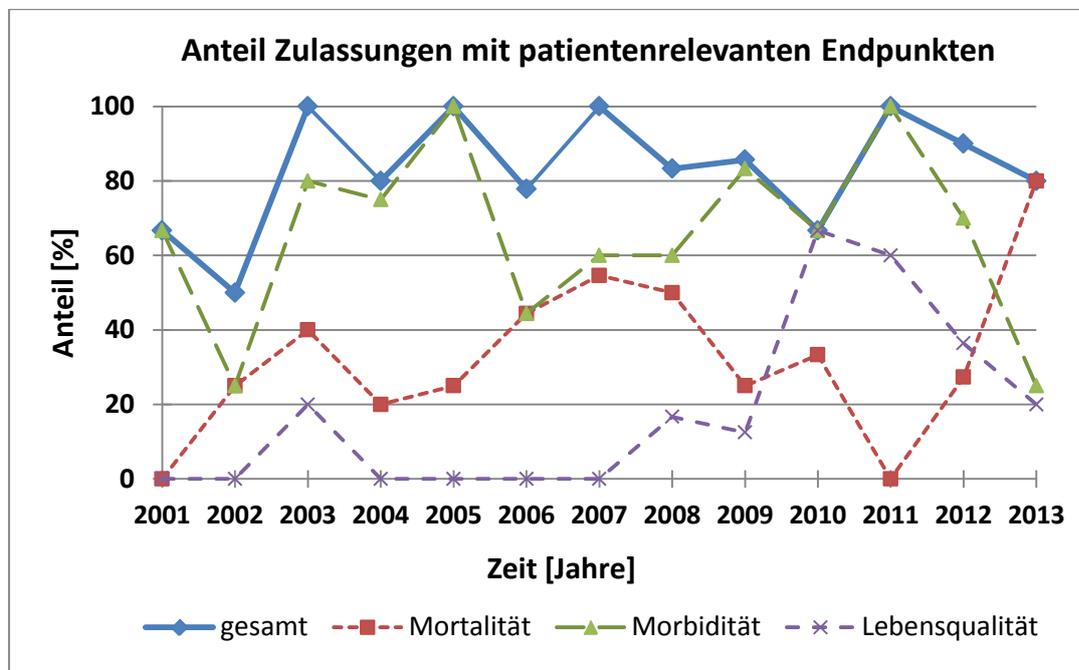


Abbildung 2: Anteil der Zulassungen, bei denen patientenrelevante Endpunkte (primär oder sekundär) ausgewertet wurden, im zeitlichen Verlauf (N = 79)

Auswertungen auf Studienebene

Grundlage der folgenden deskriptiven Auswertungen zu Charakteristika und Methodik der Zulassungsstudien sind alle 82 RCTs (66 %) unter den 125 Hauptstudien (ohne Zulassungen basierend auf Literatur-Reviews) der zugelassenen Orphan Drugs (1 bis 3 RCTs pro Arzneimittel). In den RCTs wurden zwischen 8 und 769 Patienten (Median 160,5) eingeschlossen.

Tabelle 4 zeigt eine detaillierte Darstellung der Studiencharakteristika insgesamt und getrennt für seltene und sehr seltene Erkrankungen. Methodische Aspekte werden in Tabelle 5 dargestellt.

Erwartungsgemäß ist die Patientenzahl der RCTs zu sehr seltenen Erkrankungen deutlich geringer. Angaben zur Studiendauer und zur Rekrutierungsdauer waren nur sehr lückenhaft verfügbar. Die meisten RCTs wurden multizentrisch, multinational und multikontinental durchgeführt; dabei zeigen sich keine wesentlichen Unterschiede zwischen den Studien zu seltenen und sehr seltenen Erkrankungen. In den RCTs zu sehr seltenen Erkrankungen werden häufiger Endpunkte zur Morbidität und seltener zur Mortalität betrachtet als bei seltenen Erkrankungen.

Tabelle 4: Studiencharakteristika der RCT (Studienebene)

Charakteristikum	Erkrankung								
	gesamt (N = 82)			selten (N = 59)			sehr selten (N = 23)		
	Median (unteres; oberes Quartil)			Median (unteres; oberes Quartil)			Median (unteres; oberes Quartil)		
Patientenanzahl	160,5 (63; 312)			197 (86; 344)			87 (42; 201)		
Studiendauer [Monate]	24 (17; 36) (N* = 35)			24 (19; 38) (N* = 31)			17 (13,5; 23) (N* = 4)		
Rekrutierungsdauer [Monate]	23 (20; 27) (N* = 11)			26 (23; 30) (N* = 6)			22 (20; 23) (N* = 5)		
Studienarme (1 ^a / 2 / 3 / 4) [n (%)]	4 (4,9) / 60 (73,2) / 12 (14,6) / 6 (7,3)			3 (5,1) / 45 (76,3) / 7 (11,9) / 4 (6,8)			1 (4,4) / 15 (65,2) / 5 (21,7) / 2 (8,7)		
	n (%)			n (%)			n (%)		
	ja	nein	unklar	ja	nein	unklar	ja	nein	unklar
multizentrisch	78 (95,1)	3 (3,7)	1 (1,2)	57 (96,6)	1 (1,2)	1 (1,2)	21 (91,3)	2 (8,7)	0 (0)
multinational	65 (79,3)	12 (14,6)	5 (6,1)	48 (81,4)	7 (11,9)	4 (6,8)	17 (73,9)	5 (21,7)	1 (4,4)
multikontinental	59 (72,0)	16 (19,5)	7 (8,5)	43 (72,9)	11 (18,6)	5 (8,5)	16 (69,6)	5 (21,7)	2 (8,7)
Patientenrelevanter EP									
primär oder sekundär									
gesamt	67 (81,7)	11 (13,4)	4 (4,9)	47 (79,7)	8 (13,6)	4 (6,8)	20 (87,0)	3 (13,0)	0 (0)
Mortalität	20 (24,4)	62 (75,6)	0 (0)	18 (30,5)	41 (69,5)	0 (0)	2 (8,7)	21 (91,3)	0 (0)
Morbidity	55 (67,1)	19 (23,2)	8 (9,8)	36 (61,0)	15 (25,4)	8 (13,6)	19 (82,6)	4 (17,4)	0 (0)
Lebensqualität	16 (19,5)	66 (80,5)	0 (0)	13 (22,0)	46 (78,0)	0 (0)	3 (13,0)	20 (87,0)	0 (0)
primär									
gesamt	38 (46,3)	34 (41,5)	10 (12,2)	21 (35,6)	29 (49,2)	9 (15,3)	17 (73,9)	5 (21,7)	1 (4,4)
Mortalität	5 (6,1)	77 (93,9)	0 (0)	5 (8,5)	54 (91,5)	0 (0)	0 (0)	23 (100)	0 (0)
Morbidity	33 (40,2)	39 (47,6)	10 (12,2)	16 (27,1)	34 (57,6)	9 (15,3)	17 (73,9)	5 (21,7)	1 (4,4)
Lebensqualität	1 (1,2)	81 (98,8)	0 (0)	0 (0)	59 (100)	0 (0)	1 (4,4)	22 (95,7)	0 (0)
sekundär									
gesamt	59 (72,0)	19 (23,2)	4 (4,9)	41 (69,5)	14 (23,7)	4 (6,8)	18 (78,3)	5 (21,7)	0 (0)
Mortalität	15 (18,3)	67 (81,7)	0 (0)	13 (22,0)	46 (78,0)	0 (0)	2 (8,7)	21 (91,3)	0 (0)
Morbidity	50 (61,0)	25 (30,5)	7 (8,5)	34 (57,6)	18 (30,5)	7 (11,9)	16 (69,6)	7 (30,4)	0 (0)
Lebensqualität	15 (18,3)	67 (81,7)	0 (0)	13 (22,0)	46 (78,0)	0 (0)	2 (8,7)	21 (91,3)	0 (0)
a: Cross-over-Studie									
EP: Endpunkt; N: Anzahl RCTs; N*: Anzahl verfügbarer Datensätze; n: absolute Häufigkeit									

Der Anteil doppelblinder RCTs liegt bei 74 %. Bei 28 % der Studien wurde ein aktiver Komparator eingesetzt. Nichtunterlegenheits- bzw. Äquivalenzstudien kamen mit 6 % der Studien nur in Ausnahmefällen zum Einsatz. Es zeigen sich keine auffälligen Unterschiede zwischen den Studien zu seltenen und sehr seltenen Erkrankungen.

Zur Klassifizierung und Charakterisierung der eingesetzten statistischen Methoden in den RCTs wurde – basierend auf den methodischen Ausführungen in Abschnitt 5.1 – eine Einteilung der Studien in folgende grobe Kategorien vorgenommen:

- M1: Berücksichtigung von Störgrößen (Regressionsmodell / Stratifizierung mit Berücksichtigung von Messungen im Zeitverlauf und / oder prognostischer Faktoren und Baseline-Werten)
- M2: Cross-over-Studie
- M3: Adaptive Verfahren / sequenzielle Verfahren / Interim-Analysen
- M4: Adaptive Randomisierung
- M5: Bayes'sche Verfahren
- M6: Sonstige spezielle randomisierte Designs
- M7: Erhöhung des Irrtumsniveaus
- M8: Sonstige Verfahren (nicht M1-M7)

In der Kategorie Sonstige Verfahren (M8) werden z. B. eine rein deskriptive Darstellung der Ergebnisse sowie einfache Testverfahren wie t-Test oder Chi-Quadrat-Test subsumiert.

Tabelle 5: Verwendete statistische Methoden in den RCTs (Studienebene)

Methodik	gesamt (N = 82)	Erkrankung	
	n (%)	selten (N = 59) n (%)	sehr selten (N = 23) n (%)
doppelblind	61 (74,4)	41 (69,5)	20 (87,0)
aktive Kontrolle	23 (28,1)	17 (28,8)	6 (26,1)
Nichtunterlegenheit / Äquivalenz	5 (6,1)	2 (3,4)	3 (13,0)
Auswertung			
M1	43 (52,4)	33 (55,9)	10 (43,5)
M2	4 (4,9)	3 (5,1)	1 (4,4)
M3	10 (12,2)	10 (17,0)	0 (0)
M4	0 (0)	0 (0)	0 (0)
M5	0 (0)	0 (0)	0 (0)
M6	0 (0)	0 (0)	0 (0)
M7	1 (1,2)	1 (1,7)	0 (0)
M8	27 (32,9)	15 (25,4)	12 (52,2)

M1: Berücksichtigung von Störgrößen (Regressionsmodell / Stratifizierung mit Berücksichtigung von Messungen im Zeitverlauf und / oder prognostischer Faktoren und Baseline-Werten)
M2: Cross-over-Studie
M3: Adaptive Verfahren / sequenzielle Verfahren / Interim-Analysen
M4: Adaptive Randomisierung
M5: Bayes'sche Verfahren
M6: Sonstige spezielle randomisierte Designs
M7: Erhöhung des Irrtumsniveaus
M8: Sonstige Verfahren (nicht M1–M7)
N: Anzahl RCT mit vorhandenen Angaben; n: absolute Häufigkeit

In 67 % der RCTs kam mindestens eines der definierten statistischen Verfahren M1 bis M7 zum Einsatz. In 52 % der Fälle handelte es sich um Methoden zur Berücksichtigung von Störgrößen (M1). Cross-over-Designs (M2) und sequenzielle Verfahren (M3) wurden in 5 % bzw. 12 % der Studien eingesetzt. Eine Erhöhung des Irrtumsniveaus (M7) wurde in 1 Studie vorgenommen. Eine adaptive Randomisierung (M4), Bayes'sche Verfahren (M5) oder sonstige spezielle randomisierte Designs (M6) wurden in keiner der Zulassungsstudien eingesetzt.

Insgesamt wurden in den Studien zu sehr seltenen Erkrankungen weniger häufig spezielle Auswertungsmethoden (M1–M7) eingesetzt, sequenzielle Verfahren kamen gar nicht zur Anwendung.

6 Diskussion

Methodische Ansätze für Studien bei seltenen Erkrankungen

Die Durchführung aussagekräftiger Studien zur Bewertung von Interventionen bei seltenen Erkrankungen ist mit besonderen Problemen verbunden. Trivialerweise können wenige verfügbare Daten auch nur eingeschränkte Informationen liefern. Dies gilt jedoch für alle Studientypen und ist kein Argument gegen die Durchführung von RCTs [1]. Kriterien der Ergebnissicherheit und Aussagekraft von Studiendesigns und -ergebnissen gelten gleichermaßen für Studien bei seltenen und nicht seltenen Erkrankungen [11,34]. Häufig angeführte ethische Aspekte werden kontrovers diskutiert und liefern auch keine klar überzeugenden Argumente gegen die Durchführung von RCTs [31,37].

Die in Abschnitt 5.1 beschriebenen statistischen Methoden sind geeignet, die Aussagekraft der Ergebnisse von RCTs zu erhöhen. Cornu et al. [33] beschreiben einen Algorithmus zur Auswahl des am besten geeigneten Studiendesigns für eine konkret vorliegende Fragestellung. Es sei noch einmal darauf hingewiesen, dass diese methodischen Überlegungen grundsätzlich für alle RCTs Gültigkeit haben, bei Studien zu seltenen Erkrankungen aufgrund der inhärent kleinen Fallzahlen jedoch von besonderer Wichtigkeit sind, um zu belastbaren Aussagen zu gelangen.

In einem Rapid Report [5] untersuchte das IQWiG die Manuale relevanter HTA-Agenturen im Hinblick auf den Umgang mit Evidenz bei seltenen Erkrankungen. Hinweise und Anleitungen zur Methodik und zum Umgang mit Evidenz bei seltenen Erkrankungen fanden sich in 8 von 24 identifizierten HTA-Manualen, allerdings nur in sehr eingeschränkter Form. Konkretere methodische Aspekte fanden sich lediglich im Methodenpapier des IQWiG [72].

Anwendung in Zulassungsstudien

Die Auswertung der zugelassenen Orphan Drugs in der EU zeigt, dass ein Großteil der zugrunde liegenden Zulassungsstudien RCTs war. Dies gilt auch für sehr seltene Erkrankungen, wo die verfügbaren Fallzahlen naturgemäß besonders gering sind.

Bei etwa zwei Drittel der RCTs unter den Zulassungsstudien kamen in Abschnitt 5.1 beschriebene methodische Verfahren zum Einsatz, die auf eine Erhöhung der Präzision und damit verbunden eine Verbesserung der Aussagefähigkeit der Ergebnisse abzielen. Mit Abstand am häufigsten wurden Methoden zur Berücksichtigung von Störgrößen eingesetzt, die seit Jahrzehnten zum methodischen Standard bei der Auswertung klinischer Studien – und nicht nur bei geringen Fallzahlen – gehören [73]. Andere Verfahren wie z. B. Interimsanalysen oder Cross-over-Studien spielten eine eher untergeordnete Rolle. Ein erhöhtes Irrtumsniveau wurde lediglich in einer Studie beobachtet (Arzneimittel Tracleer). Es bleibt jedoch unklar, inwieweit Überlegungen hinsichtlich eines möglichen Kompromisses bei der Aussagesicherheit – wie in Abschnitt 5.1 diskutiert – hierbei eine Rolle gespielt haben. Auffälligerweise war insgesamt der Anteil dieser Verfahren in Studien zu sehr seltenen Erkrankungen vergleichsweise geringer.

Der hohe Anteil von RCTs unter den Zulassungsstudien zeigt, dass die Durchführung von RCTs – auch bei sehr seltenen Erkrankungen – durchaus möglich ist. Auch weitere für die Aussagekraft einer klinischen Studie relevante Aspekte wie die multizentrische Durchführung und die Verwendung patientenrelevanter Endpunkte fanden bei Studien zu sehr seltenen Erkrankungen nicht weniger Berücksichtigung als bei seltenen Erkrankungen.

Picavet et al. [66] führten ebenfalls eine Auswertung der Zulassungsstudien für Orphan Drugs durch, allerdings basierend auf den zugelassenen Orphan Drugs bis lediglich Mitte 2012 und mit einem anderen Schwerpunkt als im vorliegenden Rapid Report. Die Autoren betrachteten primär die allgemeine Qualität der Zulassungsstudien und deren Berichtsqualität sowie die Verwendung von Surrogatendpunkten und Endpunkten zur Lebensqualität. Insgesamt sehen die Autoren methodische Mängel in den Zulassungsstudien und fordern eine bessere Evidenzgrundlage, u. a. auch durch den Einsatz neuer adaptiver Studiendesigns. Zur Auswertung der Zulassungsstudien verwendeten die Autoren ein von ihnen selbst entwickeltes und validiertes Instrument zur Qualitätsbewertung der Evidenz zu Orphan Drugs (COMPASS-Tool) [74]. Das COMPASS-Tool ist ein strukturierter dreiteiliger Fragebogen zur Erfassung der relevanten allgemeinen Informationen und zur Bewertung der methodischen Qualität und der Berichtsqualität. Auch Joppi et al. [18,75,76] führten im Laufe der Jahre verschiedene Auswertungen der Orphan-Drug-Zulassungen der EMA mit ähnlichen Schwerpunkten wie Picavet et al. [66] durch. Die Autoren sahen zwar leichte Verbesserungen in den letzten Jahren, bemängelten aber die nach wie vor eingeschränkte Qualität und Evidenz der Daten.

Insgesamt zeigen sich bei allen genannten Untersuchungen methodische Mängel und eine eingeschränkte Evidenz der Zulassungsstudien. Die teilweise abweichenden Ergebnisse der unterschiedlichen Auswertungen sind durch eine unterschiedliche Datengrundlage aufgrund des Auswertungszeitpunktes, abweichende Definitionen einzelner Studiencharakteristika, abweichende Interpretation unklarer Informationen und unterschiedliche Betrachtungsebenen (Studienebene oder Arzneimittelzebene) zu erklären.

Downing et al. [77] analysierten die zugrunde liegenden pivotalen Studien für die Zulassung neuer Medikamente durch die US Food and Drug Administration (FDA) in den Jahren 2005 bis 2012. Unter den neu zugelassenen 188 Medikamenten befanden sich 31 Medikamente (56 Studien) mit Orphan Status. Etwa 54 % der Studien waren RCTs, davon 70 % doppelblind. Diese Ergebnisse der FDA haben eine vergleichbare Größenordnung wie bei den Daten der EMA.

Insgesamt sind bei der Interpretation aller Ergebnisse mögliche Selektionseffekte zu beachten. In die Auswertungen gingen nur Studien ein, die zu einer erfolgreichen Zulassung führten. Nicht betrachtet wurden Arzneimittel zur Behandlung seltener Erkrankungen, bei denen die Industrie aber dann an der Durchführung geplanter Studien scheiterte oder die Zulassungsbehörde die Zulassung wegen mangelhafter Studienqualität oder -größe ablehnte. Weitergehende Informationen zu diesen Arzneimitteln wären jedoch auch – abgesehen von

den wenigen bei der EMA eingereichten aber abgelehnten Markteinreichungen⁴ – kaum in systematischer Form zu erhalten.

Kriterien der Zulassung von Orphan Drugs

Simeons et al. [79] diskutieren kritisch die Argumente und Kriterien für derzeit gültige gesetzliche Bestimmungen bei der Zulassung von Orphan Drugs. Die Autoren kommen zu dem Schluss, dass es an der Zeit ist, die speziellen Anreize und Regelungen für den Marktzugang von Orphan Drugs neu zu überdenken. Hughes-Wilson et al. [80] diskutieren die vergleichsweise hohen Kosten von Orphan Drugs und schlagen die Entwicklung eines neuen Bewertungssystems zur Erstattung von Orphan Drugs basierend auf einer Reihe relevanter Kriterien vor wie z. B. dem Schweregrad der behandelten Krankheit und dem Forschungsaufwand. Ein weiteres Kriterium betrifft methodische Aspekte (Studiendesign, untersuchte Endpunkte, Komparator, Effektgröße) und die Evidenz der zugrunde liegenden Daten.

Medizinprodukte

Die methodischen Überlegungen in Abschnitt 5.1 haben grundsätzlich Gültigkeit bei beliebigen Interventionen. Neben der Zulassung von Arzneimitteln (Abschnitt 5.2) ist es daher auch von Interesse, einen Blick auf den Bereich der Medizinprodukte bei seltenen Erkrankungen zu werfen. Hierzu wurden systematisch die öffentlich verfügbaren Dokumente, die die Zulassung von Medizinprodukten in Europa und in den USA regulieren, gesichtet. Ferner erfolgte am 21. Mai 2014 eine PubMed-Suche über „orphan AND device*[Ti]“, die 13 Treffer ergab.

Im Bereich der europäischen Medizinproduktezulassung gibt es keine spezifischen regulatorischen Vorschriften, die auf seltene Erkrankungen Bezug nehmen. Die regulatorischen Dokumente [81], die gegenwärtig für die CE-Kennzeichnung von Medizinprodukten gelten (insbesondere die Richtlinie MEDDEV 2.7.1 Rev.3 [82]), enthalten keine spezifischen Aussagen zu seltenen Erkrankungen. Bei der Durchsicht der Dokumente zur geplanten Revision der europäischen Regulierung (Stand März 2014) lässt sich erkennen, dass auch weiterhin keine spezifischen Vorgaben für Medizinprodukte zur Behandlung seltener Erkrankungen (sogenannte „orphan devices“) geplant sind [83].

In den USA dagegen existiert für Medizinprodukte ein regulatorischer Sonderweg, wenn das Medizinprodukt in einer Indikation angewendet wird, die maximal 4000 US-Bürger pro Jahr betrifft [84]. In diesen Fällen kann aus humanitären Gründen („Humanitarian Use Device“) die Zulassung erteilt werden. Über diesen Weg wurden seit 1997 mehr als 50 Medizinprodukte von der FDA zugelassen [85]. Die für die einzelnen Medizinprodukte vorgelegten klinischen Daten reichen von kleinen Fallserien bis zu RCTs. Auch Medizinprodukte der höchsten Risikoklasse (z. B. die Melody Transkatheter-Pulmonalklappe, Medtronic Inc.) erhielten auf der Basis von Studien ohne Kontrollgruppe so eine Zulassung.

⁴ Es handelt sich um 17 Arzneimittel mit „negative opinion“ im Zeitraum von 2002 bis 2014 [78]).

Die große Mehrzahl der auf diesem Wege in den USA zugelassenen Medizinprodukte verfügt auch über eine CE-Kennzeichnung.

Ein aktuelles Beispiel zeigt klar die Risiken der FDA-Regulierung auf: Auch der Wingspan-Stent (Boston Scientific Corp.) kam 2005 über den Sonderweg der humanitären Ausnahme in den USA auf den Markt [50,86]. Die Zulassung für die Indikation der intrakraniellen Stenose erfolgte auf der Basis einer Fallserie an 45 Patienten. Ein RCT (n = 451) zeigte dann 2011 keine Vor-, sondern deutliche Nachteile, die mit der Stentimplantation verbunden sind [87]. Fraglich ist ferner, ob Stenosen intrakranieller Gefäße überhaupt als selten gelten können. Da allein 2011 in Deutschland 1368 Patienten einen Stent erhalten haben [88], dürfte in den USA aufgrund der etwa dreifach größeren Bevölkerungszahl die selbstgesetzte Schwelle von 4000 Patienten mit einer möglichen Stent-Indikation überschritten worden sein. Auch epidemiologische Studien, die eine Prävalenz symptomatischer intrakranieller Stenosen von etwa 10 auf 100 000 Einwohner fanden, führen zu einem ähnlichen rechnerischen Schätzwert [89]. Erneut zeigt sich, dass der Verzicht auf eine scheinbar kaum durchführbare RCT erhebliche Schäden verursachen kann.

7 Fazit

Eine Begründung für eine unterschiedliche Herangehensweise bei der Bewertung von medizinischen Interventionen für seltene gegenüber nicht seltenen Erkrankungen kann wissenschaftlich nicht abgeleitet werden. Umgekehrt existieren auch keine spezifischen Designs und statistischen Methoden für seltene Erkrankungen, die nicht auch relevant für häufige(re) Erkrankungen sein könnten. Dies gilt in gleicher Weise für medikamentöse wie für nichtmedikamentöse Interventionen.

Zulassungen und Zulassungsstudien für Orphan Drugs basieren zu einem großen Teil, auch bei sehr seltenen Erkrankungen, auf konventionellen (randomisierten) Designs, sodass die grundsätzliche Machbarkeit nicht infrage steht.

Im Kontext der Bewertung medizinischer Interventionen bei seltenen, insbesondere sehr seltenen Erkrankungen kann es dennoch notwendig oder auch politisch gewünscht sein, Kompromisse bei der Aussagesicherheit einzugehen. Solche Kompromisse sind grundsätzlich auf 3 Ebenen denkbar:

Das (statistische) Irrtumsniveau könnte über den üblichen Wert von (zweiseitig) 5 % angehoben werden (Kompromiss bei der geforderten Präzision).

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.

Die Aufgabe der internen Validität durch Verzicht auf eine Randomisierung stünde in einem hierarchischen Vorgehen aus logischen Gründen an letzter Stelle. Wesentliche Voraussetzung, um sie für (regulatorische) Entscheidungen nutzen zu können, wäre, dass die zugrunde liegenden Daten aus einem Krankheitsregister mit exzellenter Qualität im Hinblick auf Vollständigkeit und Vollzähligkeit stammen.

8 Literatur

1. Windeler J, Lange S. Nutzenbewertung in besonderen Situationen: seltene Erkrankungen. Z Evid Fortbild Qual Gesundheitswes 2008; 102(1): 25-30.
2. Europäisches Parlament, Rat der Europäischen Union. Verordnung (EG) Nr. 141/2000 des Europäischen Parlaments und des Rates vom 16. Dezember 1999 über Arzneimittel für seltene Leiden. Amtsblatt der Europäischen Gemeinschaften 2000; 43(L18): 1-5.
3. Eidt D, Frank M, Reimann A, Wagner TOF, Mittendorf T, Graf von der Schulenburg JM. Maßnahmen zur Verbesserung der gesundheitlichen Situation von Menschen mit Seltenen Erkrankungen in Deutschland [online]. 09.06.2009 [Zugriff: 03.02.2014]. URL: http://www.gesundheitsforschung-bmbf.de/media/Autorentexte_Massnahmen-seltene-Krankheiten_200908.pdf.
4. Rat der Europäischen Union. Empfehlung des Rates vom 8. Juni 2009 für eine Maßnahme im Bereich seltener Krankheiten. Amtsblatt der Europäischen Gemeinschaften 2009; 52(C151): 7-10.
5. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Welche Evidenz wird für die Erstellung von Leitlinien für seltene Erkrankungen derzeit herangezogen? Rapid Report; Auftrag V10-01 [online]. 23.03.2011 [Zugriff: 28.01.2014]. (IQWiG-Berichte; Band 83). URL: https://www.iqwig.de/download/V10-01_Rapid-Report_Evidenz_fuer_Leitlinien_zu_seltenen_Erkrankungen.pdf.
6. Nationales Aktionsbündnis für Menschen mit Seltenen Erkrankungen. Nationaler Aktionsplan für Menschen mit Seltenen Erkrankungen: Handlungsfelder, Empfehlungen und Maßnahmenvorschläge. Bonn: NAMSE; 2013. URL: http://www.namse.de/images/stories/Dokumente/nationaler_aktionsplan.pdf.
7. Advances in Small Trials dESign for Regulatory Innovation and eXcellence. Webauftritt [online]. [Zugriff: 20.08.2014]. URL: <http://www.asterix-fp7.eu/>.
8. Integrated DEsign and AnaLysis of small population group trials. Webauftritt [online]. [Zugriff: 20.08.2014]. URL: <http://www.ideal.rwth-aachen.de/>.
9. University of Warwick. InSPiRE: innovative methodology for small populations research [online]. 04.08.2014 [Zugriff: 20.08.2014]. URL: <http://www2.warwick.ac.uk/fac/med/research/hscience/stats/currentprojects/inspire/>.
10. Behera M, Kumar A, Soares HP, Sokol L, Djulbegovic B. Evidenz-basiertes Medizin für seltene Krankheiten: Implikationen für die Dateninterpretation und das klinische Studien-Design. Cancer Control 2007; 14(2): 160-166.
11. European Medicines Agency. Guideline on clinical trials in small populations [online]. 27.07.2006 [Zugriff: 28.01.2014]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003615.pdf.

12. European Medicines Agency. Concept paper on extrapolation of efficacy and safety in medicine development [online]. 19.03.2013 [Zugriff: 25.08.2014]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/04/WC500142358.pdf.
13. Verzeichnis der Arzneimittel für seltene Krankheiten in Europa [online]. 10.2013 [Zugriff: 29.01.2014]. (Orphanet Berichtsreihe: Orphan Drugs Datenerhebung; Band Oktober 2013). URL: http://www.orpha.net/orphacom/cahiers/docs/DE/Verzeichnis_der_in_Europa_zugelassenen_Orphan_Drugs.pdf.
14. European Medicines Agency. Orphan drugs and rare diseases at a glance [online]. 03.07.2007 [Zugriff: 21.03.2014]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Other/2010/01/WC500069805.pdf.
15. Verzeichnis der Arzneimittel für seltene Krankheiten in Europa [online]. 01.2014 [Zugriff: 19.02.2014]. (Orphanet Berichtsreihe: Orphan Drugs Datenerhebung; Band Januar 2014). URL: http://www.orpha.net/orphacom/cahiers/docs/DE/Verzeichnis_der_in_Europa_zugelassenen_Orphan_Drugs.pdf.
16. Prävalenz seltener Krankheiten: bibliographische Angaben; alphabetische Liste der Krankheiten oder Krankheitsgruppen [online]. 11.2013 [Zugriff: 30.04.2014]. (Orphanet Berichtsreihe: Seltene Krankheiten Datenerhebung; Band November 2013, Nummer 1). URL: http://www.orpha.net/orphacom/cahiers/docs/DE/Pravalenzen_seltener_Krankheiten_Alphabetische_Liste.pdf.
17. Committee for Orphan Medicinal Products, European Medicines Agency Scientific Secretariat. European regulation on orphan medicinal products: 10 years of experience and future perspectives. *Nature Reviews Drug Discovery* 2011; 10(5): 341-349.
18. Joppi R, Bertele V, Garattini S. Orphan drugs, orphan diseases: the first decade of orphan drug legislation in the EU. *Eur J Clin Pharmacol* 2013; 69(4): 1009-1024.
19. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med* 2009; 6(7): e1000100.
20. Atkins D, Chang SM, Gartlehner G, Buckley DI, Whitlock EP, Berliner E et al. Assessing applicability when comparing medical interventions: AHRQ and the Effective Health Care Program. *J Clin Epidemiol* 2011; 64(11): 1198-1207.
21. Hróbjartsson A, Thomsen AS, Emanuelsson F, Tendal B, Hilden J, Boutron I et al. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. *BMJ* 2012; 344: e1119.

22. Wood L, Egger M, Gluud LL, Schulz KF, Jüni P, Altman DG et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008; 336(7644): 601-605.
23. Bundesministerium für Gesundheit. Verordnung über die Nutzenbewertung von Arzneimitteln nach § 35a Absatz 1 SGB V für Erstattungsvereinbarungen nach § 130b SGB V (Arzneimittel-Nutzenbewertungsverordnung – AM-NutzenV). *Bundesgesetzblatt Teil 1* 2010; (68): 2324-2328.
24. Schünemann HJ, Akl EA, Guyatt GH. Interpreting the results of patient reported outcome measures in clinical trials: the clinician's perspective. *Health Qual Life Outcomes* 2006; 4: 62.
25. Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. *BMJ* 2007; 334(7589): 349-351.
26. Higgins JPT, Green S (Ed). *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley-Blackwell; 2008.
27. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011; 343: d5928.
28. Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. *JAMA* 2009; 302(9): 977-984.
29. Morden JP, Lambert PC, Latimer N, Abrams KR, Wailoo AJ. Assessing methods for dealing with treatment switching in randomised controlled trials: a simulation study. *BMC Med Res Methodol* 2011; 11: 4.
30. Jakobsen JC, Gluud C. The necessity of randomized clinical trials. *Br J Med Med Res* 2013; 3(4): 1453-1468.
31. Day S. Evidence-based medicine and rare diseases. In: Posada de la Paz M, Groft SC (Ed). *Rare diseases epidemiology*. Dordrecht: Springer; 2010. S. 41-53. (*Advances in Experimental Medicine and Biology*; Band 686).
32. Schmoor C, Caputo A, Schumacher M. Evidence from nonrandomized studies: a case study on the estimation of causal effects. *Am J Epidemiol* 2008; 167(9): 1120-1129.
33. Cornu C, Kassai B, Fisch R, Chiron C, Alberti C, Guerrini R et al. Experimental designs for small randomised clinical trials: an algorithm for choice. *Orphanet J Rare Dis* 2013; 8: 48.
34. Gerß JWO, Köpcke W. Clinical trials and rare diseases. In: Posada de la Paz M, Groft SC (Ed). *Rare diseases epidemiology*. Springer: Dordrecht; 2010. S. 173-190. (*Advances in Experimental Medicine and Biology*; Band 686).
35. Griggs RC, Batshaw M, Dunkle M, Gopal-Srivastava R, Kaye E, Krischer J et al. Clinical research for rare disease: opportunities, challenges, and solutions. *Mol Genet Metab* 2009; 96(1): 20-26.
36. Gupta S, Faughnan ME, Tomlinson GA, Bayoumi AM. A framework for applying unfamiliar trial designs in studies of rare diseases. *J Clin Epidemiol* 2011; 64(10): 1085-1094.

37. Kianifard F, Islam MZ. A guide to the design and analysis of small clinical studies. *Pharm Stat* 2011; 10(4): 363-368.
38. van der Lee JH, Wesseling J, Tanck MW, Offringa M. Efficient ways exist to obtain the optimal sample size in clinical trials in rare diseases. *J Clin Epidemiol* 2008; 61(4): 324-330.
39. Tudur Smith C, Williamson PR, Beresford MW. Methodology of clinical trials for rare diseases. *Best Pract Res Clin Rheumatol* 2014; 28(2): 247-262.
40. Korn EL, McShane LM, Freidlin B. Statistical challenges in the evaluation of treatments for small patient populations. *Sci Transl Med* 2013; 5(178): 178sr3.
41. Zucker DR, Ruthazer R, Schmid CH. Individual (N-of-1) trials can be combined to give population comparative treatment effect estimates: methodologic considerations. *J Clin Epidemiol* 2010; 63(12): 1312-1323.
42. Sebillé V, Bellissant E. Comparison of four sequential methods allowing for early stopping of comparative clinical trials. *Clin Sci* 2000; 98(5): 569-578.
43. Bassler D, Briel M, Montori VM, Lane M, Glasziou P, Zhou Q et al. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. *JAMA* 2010; 303(12): 1180-1187.
44. Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, Briel M et al. Randomized trials stopped early for benefit: a systematic review. *JAMA* 2005; 294(17): 2203-2209.
45. Bassler D, Montori VM, Briel M, Glasziou P, Walter SD, Ramsay T et al. Reflections on meta-analyses involving trials stopped early for benefit: is there a problem and if so, what is it? *Stat Methods Med Res* 2013; 22(2): 159-168.
46. Goodman S, Berry D, Wittes J. Bias and trials stopped early for benefit. *JAMA* 2010; 304(2): 157.
47. Goodman SN. Systematic reviews are not biased by results from trials stopped early for benefit. *J Clin Epidemiol* 2008; 61(1): 95-96.
48. Guyatt GH, Briel M, Glasziou P, Bassler D, Montori VM. Problems of stopping trials early. *BMJ* 2012; 344: e3863.
49. European Medicines Agency. Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design [online]. 18.10.2007 [Zugriff: 18.08.2014]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003616.pdf.
50. Food and Drug Administration. Guidance for industry: adaptive design clinical trials for drugs and biologics; draft guidance [online]. 02.2010 [Zugriff: 29.04.2014]. URL: <http://www.fda.gov/downloads/Drugs/.../Guidances/ucm201790.pdf>.

51. Chow SC, Corey R. Benefits, challenges and obstacles of adaptive clinical trial designs. *Orphanet J Rare Dis* 2011; 6: 79.
52. Bauer P, Brannath W. The advantages and disadvantages of adaptive designs for clinical trials. *Drug Discov Today* 2004; 9(8): 351-357.
53. Lilford RJ, Thornton JG, Braunholtz D. Clinical trials and rare diseases: a way out of a conundrum. *BMJ* 1995; 311(7020): 1621-1625.
54. Goodman SN. Introduction to Bayesian methods I: measuring the strength of evidence. *Clin Trials* 2005; 2(4): 282-290.
55. Louis TA. Introduction to Bayesian methods II: fundamental concepts. *Clin Trials* 2005; 2(4): 291-294.
56. Wellek S. Das Bayessche Prinzip der Inferenz und seine Bedeutung für die Medizinische Statistik. *Inform Biom Epidemiol Med Biol* 1994; 25(3): 157-171.
57. Lee JJ, Chu CT. Bayesian clinical trials in action. *Stat Med* 2012; 31(25): 2955-2972.
58. D'Agostino RB. The delayed-start study design. *N Engl J Med* 2009; 361(13): 1304-1306.
59. Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. *BMC Med Res Methodol* 2006; 6: 54.
60. Rosner GL, Stadler W, Ratain MJ. Randomized discontinuation design: application to cytostatic antineoplastic agents. *J Clin Oncol* 2002; 20(22): 4478-4484.
61. Honkanen VE, Siegel AF, Szalai JP, Berger V, Feldman BM, Siegel JN. A three-stage clinical trial design for rare disorders. *Stat Med* 2001; 20(20): 3009-3021.
62. Bartlett RH, Roloff DW, Cornell RG, Andrews AF, Dillon PW, Zwischenberger JB. Extracorporeal circulation in neonatal respiratory failure: a prospective randomized study. *Pediatrics* 1985; 76(4): 479-487.
63. UK Collaborative ECMO Trial Group. UK collaborative randomised trial of neonatal extracorporeal membrane oxygenation. *Lancet* 1996; 348(9020): 75-82.
64. Deutscher Bundestag. Beschlussempfehlung und Bericht des Ausschusses für Gesundheit (14. Ausschuss): Bundestag Drucksache 17/3698 [online]. 10.11.2010 [Zugriff: 25.04.2014]. URL: <http://dipbt.bundestag.de/dip21/btd/17/036/1703698.pdf>.
65. Luisetti M, Campo I, Scabini R, Zorzetto M, Kadija Z, Mariani F et al. The problems of clinical trials and registries in rare diseases. *Respir Med* 2010; 104(Suppl 1): S42-S44.
66. Picavet E, Cassiman D, Hollak CE, Maertens JA, Simoens S. Clinical evidence for orphan medicinal products: a cause for concern? *Orphanet J Rare Dis* 2013; 8(1): 164.
67. Krischer JP, Gopal-Srivastava R, Groft SC, Eckstein DJ. The Rare Diseases Clinical Research Network's Organization and Approach to Observational Research and Health Outcomes Research. *J Gen Intern Med* 2014; 29(3 Suppl): 739-744.

68. De Boeck K, Bulteel V, Tiddens H, Wagner T, Fajac I, Conway S et al. Guideline on the design and conduct of cystic fibrosis clinical trials: the European Cystic Fibrosis Society-Clinical Trials Network (ECFS-CTN). *J Cyst Fibros* 2011; 10(Suppl 2): S67-S74.
69. Kruer MC, Steiner RD. The role of evidence-based medicine and clinical trials in rare genetic disorders. *Clin Genet* 2008; 74(3): 197-207.
70. Goss CH, Mayer-Hamblett N, Kronmal RA, Ramsey BW. The cystic fibrosis therapeutics development network (CF TDN): a paradigm of a clinical trials network for genetic and orphan diseases. *Adv Drug Deliv Rev* 2002; 54(11): 1505-1528.
71. Deutsche Krankenhausgesellschaft. Positionen der Deutschen Krankenhausgesellschaft zur Weiterentwicklung der Qualitätssicherung und der Patientensicherheit [online]. 05.2014 [Zugriff: 13.05.2014]. URL: http://www.dkgev.de/media/file/16586.2014-05-09_Anlage_DKG-Positionen-Qualitaets-sicherung.pdf.
72. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden: Version 4.1. Köln: IQWiG; 2013. URL: https://www.iqwig.de/download/IQWiG_Methoden_Version_4-1.pdf.
73. Saquib N, Saquib J, Ioannidis JP. Practices and impact of primary outcome adjustment in randomized controlled trials: meta-epidemiologic study. *Br Med J* 2013; 347: f4313.
74. Picavet E, Cassiman D, Aertgeerts B, Simoens S. Development and validation of COMPASS: clinical evidence of orphan medicinal products: an assessment tool. *Orphanet J Rare Dis* 2013; 8: 157.
75. Joppi R, Bertele V, Garattini S. Orphan drug development is progressing too slowly. *Br J Clin Pharmacol* 2006; 61(3): 355-360.
76. Joppi R, Bertele V, Garattini S. Orphan drug development is not taking off. *Br J Clin Pharmacol* 2009; 67(5): 494-502.
77. Downing NS, Aminawung JA, Shah ND, Krumholz HM, Ross JS. Clinical trial evidence supporting FDA approval of novel therapeutic agents, 2005-2012. *JAMA* 2014; 311(4): 368-377.
78. European Medicines Agency. Rare disease (orphan) designations [online]. [Zugriff: 04.09.2014]. URL: http://www.ema.europa.eu/ema/index.jsp?curl=pages%2Fmedicines%2Flanding%2Forphan_search.jsp&mid=WC0b01ac058001d12b&searchkwByEnter=false&alreadyLoaded=true&isNewQuery=true&status=Negative&startLetter=View+all&keyword=Enter+keywords&searchType=Active+Substance.
79. Simoens S, Cassiman D, Dooms M, Picavet E. Orphan drugs for rare diseases: is it time to revisit their special market access status? *Drugs* 2012; 72(11): 1437-1443.

80. Hughes-Wilson W, Palma A, Schuurman A, Simoens S. Paying for the orphan drug system: break or bend? Is it time for a new evaluation system for payers in Europe to take account of new rare disease treatments? *Orphanet J Rare Dis* 2012; 7: 74.
81. European Commission. Clinical investigation and evaluation [online]. [Zugriff: 16.04.2014]. URL: http://ec.europa.eu/health/medical-devices/dialogue-parties/working-groups/cie_wg_index_en.htm.
82. European Commission. Clinical evaluation: a guide for manufacturers and notified bodies [online]. 12.2009 [Zugriff: 16.04.2014]. (Guidelines on Medical Devices; Band MEDDEV. 2.7.1 Rev.3). URL: http://ec.europa.eu/health/medical-devices/files/meddev/2_7_1rev_3_en.pdf.
83. European Commission. Revision of the medical device directives [online]. [Zugriff: 16.04.2014]. URL: http://ec.europa.eu/health/medical-devices/documents/revision/index_en.htm.
84. Food and Drug Administration. Developing products for rare diseases & conditions [online]. 15.04.2014 [Zugriff: 16.04.2014]. URL: <http://www.fda.gov/forindustry/developingproductsforrareconditions/default.htm>.
85. Food and Drug Administration. Listing of CDRH humanitarian device exemptions [online]. 20.11.2013 [Zugriff: 11.03.2014]. URL: <http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/DeviceApprovalsandClearances/HDEApprovals/ucm161827.htm>.
86. Food and Drug Administration. Wingspan Stent System with Gateway PTA Balloon Catheter: H050001 [online]. 09.08.2005 [Zugriff: 16.04.2014]. URL: <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cftopic/pma/pma.cfm?num=H050001>.
87. Chimowitz MI, Lynn MJ, Derdeyn CP, Turan TN, Fiorella D, Lane BF et al. Stenting versus aggressive medical therapy for intracranial arterial stenosis. *N Engl J Med* 2011; 365(11): 993-1003.
88. GKV Spitzenverband. Antrag auf Bewertung der Methode "Einsatz von Stents zur Behandlung von intrakraniellen arteriellen Stenosen" [online]. 27.03.2013 [Zugriff: 29.04.2014]. URL: http://www.g-ba.de/downloads/40-268-2430/2013-07-18_Antragsannahme_137c_Stents_intrakranielle-arterielle-Stenosen_Antrag.pdf.
89. Suri MF, Johnston SC. Epidemiology of intracranial stenosis. *J Neuroimaging* 2009; 19(Suppl 1): 11S-16S.

Anhang A – Extraktionstabellen zur Auswertung der Zulassungsstudien zu Orphan Drugs

Tabelle 6: Orphan Drugs – Charakteristika

Arzneimittel	Wirkstoff	Indikation	Geschätzte Prävalenz [Fälle / 100 000] ^a	Zulassung	
				Jahr	bedingt / außer- gewöhnliche Umstände
Adcetris	Brentuximab vedotin	Hodgkin-Lymphom	10	2012	ja / nein
		Anaplastisch großzelliges Lymphom	2		
Afinitor	Everolimus	Nicht familiäres Nierenzellkarzinom	< 42	2009 ^b	nein / nein
		Familiäres Nierenzellkarzinom	< 10		
Aldurazyme	Laronidase	Mukopolysaccharidose Typ 1	1 ^c	2003 ^d	nein / ja
Arzerra	Ofatumumab	Chronische lymphatische B-Zell-Leukämie	27	2010	ja / nein
Atriance	Nelarabine	T-Vorläuferzell-Leukämie / akutes lymphoblastisches Lymphom	11	2007	nein / ja
Bosulif	Bosutinib	Philadelphia-Chromosom positive chronische myeloische Leukämie	6	2013	ja / nein
Bronchitol	Mannitol	Zystische Fibrose	12,6	2012	nein / nein
Busilvex	Busulfan	Hämatopoetische Stammzell-Transplantation	< 1	2003 ^d	nein / nein
Carbaglu	Carglumic acid	N-Acetylglutamat-Synthetase-Mangel	< 0,1	2003	nein / nein
		Isovalerianazidämie	1		
		Methylmalonazidämie ohne Homocystinurie	< 10		
		Propionazidämie	0,002		
Cayston	Aztreonam	Zystische Fibrose	12,6	2009	ja / nein
Ceplene	Histamine dihydrochloride	Akute myeloische Leukämie	11 ^c	2008	nein / ja

(Fortsetzung)

Tabelle 6: Orphan Drugs – Charakteristika (Fortsetzung)

Arzneimittel	Wirkstoff	Indikation	Geschätzte Prävalenz [Fälle / 100 000] ^a	Zulassung	
				Jahr	bedingt / außer-gewöhnliche Umstände
Cystadane	Betaine anhydrous	Klassische Homocystinurie	1,65	2007	nein / nein
		Homocystinurie durch Methylen-Tetrahydrofolat-Reduktase-Mangel	–		
		Homocystinurie ohne Methylmalonazidurie	37 ^e		
		Methylmalonazidämie mit Homocystinurie	> 500 ^e		
Dacogen	Decitabine	Akute myeloische Leukämie	11 ^c	2012	nein / nein
Defitelio	Defibrotide	Lebervenen-Verschlusskrankheit	11	2013	nein / ja
Diacomit	Stiripentol	Myoklonusepilepsie des Kindesalters	106 ^e	2007	ja / nein
Elaprase	Idursulfase	Mukopolysaccharidose Typ 2	0,6 ^c	2007	nein / ja
Esbriet	Pirfenidone	Idiopathische Lungenfibrose	11,5	2011	nein / nein
Evoltra	Clofarabine	Akute lymphoblastische Leukämie	< 10	2006	nein / ja
Exjade	Deferasirox	Beta-Thalassämie	0,5	2006	nein / nein
Fabrazyme	Agalsidase Beta	Fabry-Syndrom	0,22 ^e	2001 ^b	nein / nein
Firazyr	Icatibant acetate	Hereditäres Angioödem Typ 1 und 2	1 ^f	2008	nein / nein
Firdapse	Amifampridine	Lambert-Eaton-Myasthenisches Syndrom	1	2009	nein / ja
Gliolan	5-aminole-vulinic acid hydrochloride	Glioblastom	1	2007	nein / nein
Glivec	Imatinib mesilate	Chronische myeloische Leukämie ⁱ	6	2001 ^b	nein / ja
Glybera	Alipogene tiparvovec	Familiärer Lipoprotein-Lipase-Mangel	0,2	2012	nein / ja
Iclusig	Ponatinib	Chronische myeloische Leukämie	6	2013	nein / nein
		Akute lymphoblastische Leukämie	< 10		
Ilaris	Canakinumab	Idiopathische juvenile systemische Arthritis	5	2009 ^b	nein / ja
		Cryopyrin-assoziiertes periodisches Syndrom	< 1		

(Fortsetzung)

Tabelle 6: Orphan Drugs – Charakteristika (Fortsetzung)

Arzneimittel	Wirkstoff	Indikation	Geschätzte Prävalenz [Fälle / 100 000] ^a	Zulassung	
				Jahr	bedingt / außer-gewöhnliche Umstände
Imnovid	Pomalidomide	Multipl. Myelom	11,5	2013	nein / nein
Increlex	Mecasermin	Wachstumsverzögerung durch IGF-1-Mangel	4 ^e	2007	nein / ja
Inovelon	Rufinamide	Lennox-Gastaut-Syndrom	15	2007	nein / nein
Jakavi	Ruxolitinib	Essenzielle Thrombozythämie	24	2012	nein / nein
		Myelofibrose mit myeloischer Metaplasie	1		
		Polycythämia vera	30		
Kalydeco	Ivacaftor	Zystische Fibrose	12,6	2012	nein / nein
Kuvan	Sapropterin dihydrochloride	Dihydropteridinreduktase-Mangel	> 150 ^e	2008	nein / nein
Litak	Cladribine	Haarzell-Leukämie (HCL)	3,12 ^e	2004	nein / nein
Lysodren	Mitotane	Nebennierenrinden-Karzinom	1	2004	nein / nein
Mepact	Mifamurtide	Osteosarkom	5	2009	nein / nein
Mozobil	Plerixafor	Lymphom	–	2009	nein / nein
		Multipl. Myelom	11,5		
Myozyme	Alglucosidase alpha	Glykogenose Typ 2	0,8 ^c	2006	nein / nein
Naglazyme	Galsulfase	Mukopolysaccharidose Typ 6	0,16 ^c	2006	nein / ja
Nexavar	Sorafenib tosylate	Nicht familiäres Nierenzellkarzinom	< 42	2006	nein / nein
		Familiäres Nierenzellkarzinom	< 10		
		Hepatozelluläres Karzinom	–		
Nexobrid	Concentrate of proteolytic enzymes enriched in bromelain	Thermische Verletzungen / Verbrennungswunden des Grades IIB und III	< 60	2012	nein / nein
NovoThirteen	Catridecacog	Kongenitaler Faktor XIII-Mangel	0,05	2012 ^g	nein / nein

(Fortsetzung)

Tabelle 6: Orphan Drugs – Charakteristika (Fortsetzung)

Arzneimittel	Wirkstoff	Indikation	Geschätzte Prävalenz [Fälle / 100 000] ^a	Zulassung	
				Jahr	bedingt / außer-gewöhnliche Umstände
Nplate	Romiplostim	Autoimmune thrombozytopenische Purpura	25	2009	nein / nein
Onsenal	Celecoxib	Adenomatöse familiäre Polyposis	6	2003 ^b	nein / ja
Orfadin	Nitisinone	Tyrosinämie Typ 1	0,05	2005	nein / nein
Orphacol	Cholic acid	Kongenitaler Gallensäuresynthesedefekt, Typ 1 und 2	< 1	2013	nein / ja
Pedea	Ibuprofen	Persistierender Ductus arteriosus	50	2004	nein / nein
Peyona	Caffeine citrate	Frühgeborenen-Apnoe	8,5	2009	nein / nein
Photobarr	Porfimer sodium	Barrett Oesophagus	> 100	2004 ^h	nein / nein
Plenadren	Hydrocortison	Primäre Nebenniereninsuffizienz	45	2011	nein / nein
Prialt	Ziconotide	Chronischer Schmerz, der eine intraspinale Analgesie erfordert	< 51	2005	nein / ja
Procysbi	Mercaptamine bitartrate	Nephropathische Cystinose	0,5 ^c	2013	nein / nein
Replagal	Agalsidase alfa	Fabry-Syndrom	0,22 ^c	2001 ^b	nein / ja
Revatio	Sildenafil citrate	Idiopathische und/oder familiäre pulmonale arterielle Hypertonie	1,5	2005	nein / nein
Revestive	Teduglutide	Sekundäres Kurzdarm-Syndrom	3,4	2012	nein / nein
Revlimid	Lenalidomide	Multipl. Melanom Myelodysplastische Syndrome	11,5 5	2007	nein / nein
Revolade	Eltrombopag	Autoimmune thrombozytopenische Purpura	25	2010 ^g	nein / nein
Rilonacept Regeneron	Rilonacept	CINCA-Syndrom Cryopyrin-assoziiertes periodisches Syndrom Familiäre Kälte-Urtikaria Muckle-Wells-Syndrom	100 < 0,1 0,1 –	2009 ^h	nein / ja

(Fortsetzung)

Tabelle 6: Orphan Drugs – Charakteristika (Fortsetzung)

Arzneimittel	Wirkstoff	Indikation	Geschätzte Prävalenz [Fälle / 100 000] ^a	Zulassung	
				Jahr	bedingt / außer-gewöhnliche Umstände
Savene	Dexrazoxane	Anthracyclin-induzierte Extravasation	< 0,1	2006	nein / nein
Signifor	Pasireotide	Cushing-Krankheit	4	2012	nein / nein
Siklos	Hydroxycarbamide	Drepanozytose (Sichelzellanämie)	15	2007	nein / nein
Soliris	Eculizumab	Atypisches hämolytisch-urämisches Syndrom	1	2007	nein / nein
		Nächtliche paroxysmale Hämoglobinurie	0,3		
Somavert	Pegvisomant	Akromegalie	6	2002 ^g	nein / nein
Sprycel	Dasatinib	Akute lymphoblastische Leukämie	< 10	2006	nein / nein
		Chronische myeloische Leukämie	6		
Sutent	Sunitinib malate	Gastrointestinaler Stroma-Tumor	13	2006 ^j	ja / nein
		Nicht familiäres Nierenzellkarzinom	< 42		
		Familiäres Nierenzellkarzinom	< 10		
Tasigna	Nilotinib	Chronische myeloische Leukämie	6	2007	nein / nein
Tepadina	Thiotepa	Hämatopoetische Stammzell-Transplantation	< 1	2010	nein / nein
Thalidomide Celgene	Thalidomide	Multipl. Myelom	11,5	2008	nein / nein
Thelin	Sitaxentan sodium	Idiopathische pulmonale arterielle Hypertonie	1,5 ^k	2006 ^l	nein / nein
		Pulmonale arterielle Hypertonie mit assoziierter Bindegewebskrankheit			
Tobi Podhaler	Tobramycin	Zystische Fibrose	12,6	2011	nein / nein
Torisel	Temsirolimus	Nicht familiäres Nierenzellkarzinom	< 42	2007	nein / nein
		Familiäres Nierenzellkarzinom	< 10		
		Mantelzell-Lymphom	4		

(Fortsetzung)

Tabelle 6: Orphan Drugs – Charakteristika (Fortsetzung)

Arzneimittel	Wirkstoff	Indikation	Geschätzte Prävalenz [Fälle / 100 000] ^a	Zulassung	
				Jahr	bedingt / außer-gewöhnliche Umstände
Tracleer	Bosentan monohydrate	Idiopathische und/oder familiäre pulmonale arterielle Hypertonie	1,5	2002	nein / ja
		Pulmonale arterielle Hypertonie mit assoziierter kongenitaler Herzkrankheit	–		
		Pulmonale arterielle Hypertonie mit assoziierter Bindegewebskrankheit	–		
		Systemische Sklerodermie	25		
Trisenox	Arsenic trioxide	Akute Promyelozytenleukämie	8	2002 ^g	nein / ja
Ventavis	Iloprost	Idiopathische pulmonale arterielle Hypertonie	< 1	2003 ^d	nein / ja
Vidaza	Azacitidine	Akute myeloische Leukämie mit multilineärer Dysplasie	–	2008	nein / nein
		Chronische myelomonozytäre Leukämie	–		
		Myelodysplastische Syndrome	5		
Volibris	Ambrisentan	Idiopathische pulmonale arterielle Hypertonie	1,5 ^k	2008	nein / nein
		Pulmonale arterielle Hypertonie mit assoziierter Bindegewebskrankheit			
Votubia	Everolimus	Tuberöse Sklerose	8,8	2011	ja / nein
Vpriv	Velaglucerase alfa	Gaucher-Krankheit Typ 1	1	2010	nein / nein
Vyndaqel	Tafamidis	Familiäre Amyloidneuropathie	< 1	2011	nein / ja
Wilzin	Zinc acetate dihydrate	Wilson-Krankheit	6	2004	nein / nein
Xagrid	Anagrelide hydrochloride	Essenzielle Thrombozythämie	24	2004	nein / ja
Xaluprine	Mercaptopurine	Akute lymphoblastische Leukämie	< 10	2012	nein / nein
Xyrem	Sodium oxybate	Gelineau-Krankheit	25	2005 ^m	nein / nein

(Fortsetzung)

Tabelle 6: Orphan Drugs – Charakteristika (Fortsetzung)

Arzneimittel	Wirkstoff	Indikation	Geschätzte Prävalenz [Fälle / 100 000] ^a	Zulassung	
				Jahr	bedingt / außer-gewöhnliche Umstände
Yondelis	Trabectedin	Ovarielles Adenokarzinom	–	2007	nein / ja
		Weichteilsarkom	23,7		
Zavesca	Miglustat	Gaucher-Krankheit Typ 1	1	2002	nein / ja
		Niemann-Pick-Krankheit Typ C	1		
<p>–: keine oder unklare Angaben a: in der Europäischen Union (EU), Quelle: Orphanet [15] b: Registeraustrag 2011 c: Prävalenz bei der Geburt d: Registeraustrag 2013 e: Gesamtzahl Fälle f: für „hereditäres Angioödem“ g: Registeraustrag 2012 h: Marktrücknahme 2012 i: Indikationserweiterungen: Akute lymphoblastische Leukämie, chronische eosinophile Leukämie, Dermatofibrosarcoma protuberans, gastrointestinaler Stromatumor, idiopathisches hypereosinophiles Syndrom, myelodysplastische / myeloproliferative Erkrankung j: Registeraustrag 2008 k: für „pulmonale arterielle Hypertonie“ l: Marktrücknahme 2011 m: Registeraustrag 2010</p>					

Tabelle 7: Methodische Aspekte der Zulassungsstudien zu Orphan Drugs (Studienebene)

Arzneimittel	Studien ^a	Studiendesign	Patienten [gesamt (IG / KG)]	Anzahl Zentren / Länder / Kontinente ^b	Rekrutierungs- / Studienzeit raum ^c [Monate]	Primärer Endpunkt	Patientenrelevante sekundäre Endpunkte (ohne UE)	Statistische Methodik (primäre Analyse)
Adcetris	2	unkontrolliert, multizentrisch, multinational, Phase II	102	≥ 2 / ≥ 2 / –	– / 19	objektive Ansprechrate	Gesamtüberleben	Häufigkeiten mit exakten 95 %-KI
		unkontrolliert, multizentrisch, multinational, Phase II	58	3 / 3 / 2	– / 15	objektive Ansprechrate	Gesamtüberleben	Häufigkeiten mit exakten 95 %-KI
Afinitor	1	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Phase III)	410 (272 / 138)	88 / 10 / 4	– / 15	progressions- freies Überleben	Gesamtüberleben; Symptome (FKSI- DRS); Lebensqualität	Überlebenszeitanalyse (Log-Rank-Test, stratifiziert; Cox-Modell, stratifiziert); Interim- Analysen (Gruppen- sequenzielle Verfahren)
Aldurazyme	1	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Phase III)	45 (22 / 23)	5 / ≤ 4 / 2	– / –	prozentuale Veränderung des erwarteten FVC; 6-Minuten- Gehtest	Symptome (CHAQ/HAQ Disability Index, AHI); Bluttransfusionen	Wilcoxon-Rangsummentest
Arzerra	1	unkontrolliert, multizentrisch, multinational, Phase II	154	41 / 10 / –	– / 24 ^d	Ansprechrate	Gesamtüberleben; Symptome;	Häufigkeiten mit 99 %-KI; Interim-Analysen
Atriance	2	unkontrolliert, multizentrisch, multinational, Phase II	70	78 / 3 / 2	– / –	Ansprechrate	Gesamtüberleben	Häufigkeiten mit 95 %-KI; Interim-Analysen
		unkontrolliert, multizentrisch, Phase II	40	≥ 2 / – / –	– / –	Ansprechrate	Gesamtüberleben	Häufigkeiten mit exakten 95 %-KI

(Fortsetzung)

Tabelle 7: Methodische Aspekte der Zulassungsstudien zu Orphan Drugs (Studienebene) (Fortsetzung)

Arzneimittel	Studien ^a	Studiendesign	Patienten [gesamt (IG / KG)]	Anzahl Zentren / Länder / Kontinente ^b	Rekrutierungs- / Studienzeit- raum ^c [Monate]	Primärer Endpunkt	Patientenrelevante sekundäre Endpunkte (ohne UE)	Statistische Methodik (primäre Analyse)
Bosulif	2	unkontrolliert, multizentrisch, multinational, Phase I/II (ongoing)	571	≥ 26 / 26 / ≥ 2	- / -	vollständiges oder partielles zytogene- tisches Ansprechen	Gesamtüberleben	Häufigkeiten mit 95 %-KI
		RCT (unverblindet, aktive Kontrolle, multizentrisch, multinational, Phase III)	502 (250 / 252)	149 / 29 / 5	- / -	vollständiges zytogeneti- sches Ansprechen	-	CMH-Test
Bronchitol	2	RCT (doppelblind, aktive Kontrolle, multizentrisch, multinational, Phase III)	324 (192 / 132)	40 / 4 / 2	- / 25	Veränderung FEV ₁	pulmonale Exazerbationen; Hospitalisierung; Lebensqualität (CFQ-R)	MMRM
		RCT (doppelblind, aktive Kontrolle, multizentrisch, multinational, Phase III)	318 (192 / 126)	53 / 7 / 3	- / 19	Veränderung FEV ₁	pulmonale Exazerbationen; Lebensqualität (CFQ-R)	MMRM
Busilvex	2	unkontrolliert, multizentrisch, multinational, Phase II	42	5 / ≥ 2 / -	- / -	-	Gesamtüberleben	deskriptiv
		unkontrolliert, multizentrisch, multinational, Phase II	62	7 / ≥ 2 / -	- / -	-	Gesamtüberleben	deskriptiv
Carbaglu	1	Beobachtungsstudie (retrospektiv, multizentrisch, multinational, Phase IIIb)	57	21 / 7 / 1	- / 190	biologisches Ansprechen (Ammonämie)	klinische Symptome	gepaarter t-test

(Fortsetzung)

Tabelle 7: Methodische Aspekte der Zulassungsstudien zu Orphan Drugs (Studienebene) (Fortsetzung)

Arzneimittel	Studien ^a	Studiendesign	Patienten [gesamt (IG / KG)]	Anzahl Zentren / Länder / Kontinente ^b	Rekrutierungs- / Studienzeit- raum ^c [Monate]	Primärer Endpunkt	Patientenrelevante sekundäre Endpunkte (ohne UE)	Statistische Methodik (primäre Analyse)
Cayston	2	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, national, Phase III)	246 ([82 + 82] / [41 + 41])	56 / 1 / 1	- / 20	Bedarf an zusätzlichen Anti-PA- Antibiotika	klinische Symptome; Hospitalisierung	Überlebenszeitanalyse (Log-Rank-Test)
		RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Phase III)	166 (83 / 83)	53 / 4 / 2	- / 23	Veränderung des CFQ-R (respiratori- sche Domäne)	klinische Symptome; Hospitalisierung	ANCOVA
Ceplene	1	RCT (unverblindet, aktive Kontrolle, multizentrisch, multinational, Phase III)	320 (160 / 160)	100 / 10 / 4	27 / 77	Zeit bis Rückfall oder Tod	Gesamtüberleben; Lebensqualität	Überlebenszeitanalyse (Log-Rank-Test, stratifiziert)
Cystadane	-	Literatur-Review: Fall-Berichte, retrospektiv, historische Kontrollen	140	-	- / -	-	-	-
Dacogen	1	RCT (unverblindet, aktive Kontrolle, multizentrisch, multinational, Phase III)	485 (243 / 242)	65 / ≥ 5 / 4	- / 46	Gesamt- überleben	-	Überlebenszeitanalyse (Hazard-Ratio, Log-Rank- Test, stratifiziert); Interim- Analysen
Defitelio	2	unverblindet, historische Kontrollen, multizentrisch, multinational, Phase III	134 (102 / 32)	35 / 3 / 2	- / 29	vollständiges Ansprechen	Gesamtüberleben	Risiko-Differenz, Chi- Quadrat-Test
		RCT (unverblindet, unbehandelte Kontrolle, multizentrisch, multinational, Phase III)	356 (180 / 173)	28 / 11 / 2	- / 43	Auftreten einer Lebervenen- Verschluss- Krankheit	Gesamtüberleben; Multi-Organversagen; GVHD; transplantationsbezo- gene Mikroangiopathie	Z-Test; Überlebenszeitanalyse (Log-Rank-Test); adaptive Interim-Analysen

(Fortsetzung)

Tabelle 7: Methodische Aspekte der Zulassungsstudien zu Orphan Drugs (Studienebene) (Fortsetzung)

Arzneimittel	Studien ^a	Studiendesign	Patienten [gesamt (IG / KG)]	Anzahl Zentren / Länder / Kontinente ^b	Rekrutierungs- / Studienzeit- raum ^c [Monate]	Primärer Endpunkt	Patientenrelevante sekundäre Endpunkte (ohne UE)	Statistische Methodik (primäre Analyse)
Diacomit	2	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, national)	42 (22 / 20)	≥ 2 / 1 / 1	20 / –	Krampfanfall- reduktion (Ansprechen)	Krampfanfälle	Chi-Quadrat-Test
		RCT (doppelblind, Placebo- Kontrolle, multizentrisch, national)	23 (12 / 11)	≥ 2 / 1 / 1	– / –	Krampfanfall- reduktion (Ansprechen)	Krampfanfälle	Chi-Quadrat-Test
Elaprase	1	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, Phase II/III)	96 ([32 + 32] / 32)	9 / ≥ 3 / 3	– / –	kombinierter Endpunkt: – prozentuale Veränderung des erwarteten FVC – 6-Minuten- Gehtest	–	Rang-ANCOVA
Esbriet	2	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Phase III)	435 ([174 + 87] / 174)	64 / 8 / 3	– / 28	absolute Veränderung des erwarteten FVC	Verschlechterung Lungenfibrose; 6-Minuten-Gehtest Symptome (UCSD SOBQ)	Rang-ANCOVA (Mantel- Haenszel mean score chi- square test)
		RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Phase III)	344 (171 + 173)	46 / 7 / 3	– / 31	absolute Veränderung des erwarteten FVC	Verschlechterung Lungenfibrose; 6-Minuten-Gehtest Symptome (UCSD SOBQ)	Rang-ANCOVA (Mantel- Haenszel mean score chi- square test)

(Fortsetzung)

Tabelle 7: Methodische Aspekte der Zulassungsstudien zu Orphan Drugs (Studienebene) (Fortsetzung)

Arzneimittel	Studien ^a	Studiendesign	Patienten [gesamt (IG / KG)]	Anzahl Zentren / Länder / Kontinente ^b	Rekrutierungs- / Studienzeit- raum ^c [Monate]	Primärer Endpunkt	Patientenrelevante sekundäre Endpunkte (ohne UE)	Statistische Methodik (primäre Analyse)
Evoltra	1	unkontrolliert, multizentrisch, national, Phase II	61	18 / 1 / 1	- / -	Remission	Gesamtüberleben	Häufigkeiten mit 95 %-KI
Exjade	1	RCT (unverblindet, aktive Kontrolle, multizentrisch, Phase III)	591 (297 / 294)	5 / 1 / 1	22 / -	Veränderung der Eisenkonzentration in der Leber	-	Risiko-Differenz, exakte KI; Test auf Nichtunterlegenheit
Fabrazyme	1	RCT (doppelblind, Placebo Kontrolle, multizentrisch, multinational, Phase III)	58 (29 / 29)	≥ 2 / ≥ 2 / 2	- / -	Verringerung der Anhäufung von GL- 3 aus dem kapillären Endothel der Niere	Schmerzreduktion (Short Form McGill Pain Questionnaire)	Signifikanztests
Firazyr	2	RCT (doppelblind, aktive Kontrolle, multizentrisch, multinational, Phase III)	74 (36 / 38)	31 / 12 / 2	- / -	Zeit bis Einsetzen von Symptom- linderung (VAS)	Symptomlinderung	mediane Zeit bis Ereignis, Signifikanztests
		RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Phase III)	56 (27 / 29)	26 / 4 / 3	- / -	Zeit bis Einsetzen von Symptom- linderung (VAS)	Symptomlinderung	mediane Zeit bis Ereignis, Signifikanztests
Firdapse	-	Literatur-Review: 2 RCT	38	- / - / -	- / -	-	-	-

(Fortsetzung)

Tabelle 7: Methodische Aspekte der Zulassungsstudien zu Orphan Drugs (Studienebene) (Fortsetzung)

Arzneimittel	Studien ^a	Studiendesign	Patienten [gesamt (IG / KG)]	Anzahl Zentren / Länder / Kontinente ^b	Rekrutierungs- / Studienzeit- raum ^c [Monate]	Primärer Endpunkt	Patientenrelevante sekundäre Endpunkte (ohne UE)	Statistische Methodik (primäre Analyse)
Gliolan	1	RCT (unverblindet, aktive Kontrolle, multizentrisch, multinational, Phase III)	415 (207 / 208)	≥ 2 / ≥ 2 / –	– / –	Anteil Patienten ohne definitiven residualen kontrast- steigernden Tumor bei frühem post- operativem MRI; progressions- freies Überleben	Gesamtüberleben	multiple Testprozedur (O'Brian-Fleming); Odds Ratios; Chi-Quadrat-Test
Glivec	3	unkontrolliert, historische Kontrollen, Phase II	532	– / – / –	– / –	Zytogeneti- sches / häma- tologisches Ansprechen	–	deskriptiv
		unkontrolliert, Phase II	–	– / – / –	– / –	zytogeneti- sches / häma- tologisches Ansprechen	–	deskriptiv, Landmark- Analyse
		unkontrolliert, Phase II	–	– / – / –	– / –	zytogeneti- sches / häma- tologisches Ansprechen	–	deskriptiv
			alle 3 Studien: 1057					

(Fortsetzung)

Tabelle 7: Methodische Aspekte der Zulassungsstudien zu Orphan Drugs (Studienebene) (Fortsetzung)

Arzneimittel	Studien ^a	Studiendesign	Patienten [gesamt (IG / KG)]	Anzahl Zentren / Länder / Kontinente ^b	Rekrutierungs- / Studienzeit- raum ^c [Monate]	Primärer Endpunkt	Patientenrelevante sekundäre Endpunkte (ohne UE)	Statistische Methodik (primäre Analyse)
Glybera	3	unkontrolliert, monozentrisch	8	1 / 1 / 1	- / -	Reduktion Nüchtern- Plasma- Triglyzerid- Spiegel	-	deskriptiv
		unkontrolliert, monozentrisch	14	1 / 1 / 1	- / -	Reduktion Nüchtern- Plasma- Triglyzerid- Spiegel	-	deskriptiv
		unkontrolliert, multizentrisch, national	5	2 / 1 / 1	- / -	Reduktion Nüchtern- Plasma- Triglyzerid- Spiegel	Symptomreduktion	deskriptiv
Iclusig	1	unkontrolliert, multizentrisch, multinational, Phase II	449	68 / 12 / 4	14 / -	zytogeneti- sches / häma- tologisches Ansprechen	Gesamtüberleben	Häufigkeiten mit exakten 95 %-KI
Ilaris	1	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational)	31 (15 / 16)	12 / 6 / 3	- / -	Krankheits- schub	Symptome	Odds Ratio, Signifikanztests
Imnovid	1	RCT (unverblindet, aktive Kontrolle, multizentrisch, Phase III)	455 (302 / 153)	92 / 15 / 3	- / 19	progressions- freies Überleben	Gesamtüberleben; Knochenschmerzen; Lebensqualität (EORTC QLQ-C30)	Überlebenszeitanalyse (Log-Rank-Test, stratifiziert; Hazard Ratios); Interim-Analysen

(Fortsetzung)

Tabelle 7: Methodische Aspekte der Zulassungsstudien zu Orphan Drugs (Studienebene) (Fortsetzung)

Arzneimittel	Studien ^a	Studiendesign	Patienten [gesamt (IG / KG)]	Anzahl Zentren / Länder / Kontinente ^b	Rekrutierungs- / Studienzeit- raum ^c [Monate]	Primärer Endpunkt	Patientenrelevante sekundäre Endpunkte (ohne UE)	Statistische Methodik (primäre Analyse)
Increlex	4	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Cross-over, Phase III)	76 gesamt (8 in RCT)	gesamt: ≥ 22 / 22 / ≥ 2	- / -	lineare Wachstums- rate	Größenveränderung	gepaarter t-Test
		unkontrolliert				Wachstums- rate	Größenveränderung	
		unkontrolliert, multizentrisch, Phase III				jährliche Wachstums- rate	-	
		unkontrolliert				-	Größenveränderung	
Inovelon	1	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational)	139 (75 / 64)	43 / 9 / 3	- / -	prozentuale Veränderung von: - Krampfanfall- Frequenz (gesamt) - Krampfanfall- Frequenz (tonisch- atonisch) - Krampfanfall- Schweregrad	Reduktion von Krampfanfällen	Wilcoxon-Rangsummentest

(Fortsetzung)

Tabelle 7: Methodische Aspekte der Zulassungsstudien zu Orphan Drugs (Studienebene) (Fortsetzung)

Arzneimittel	Studien ^a	Studiendesign	Patienten [gesamt (IG / KG)]	Anzahl Zentren / Länder / Kontinente ^b	Rekrutierungs- / Studienzeit- raum ^c [Monate]	Primärer Endpunkt	Patientenrelevante sekundäre Endpunkte (ohne UE)	Statistische Methodik (primäre Analyse)
Jakavi	2	RCT (unverblindet, aktive Kontrolle, multizentrisch, multinational, Phase III)	219 (146 / 73)	56 / 9 / 1	- / 19	Milz- Volumen Ansprechrate	Gesamtüberleben; Symptome	(exakter) CMH-Test
		RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Phase III)	309 (155 / 154)	89 / 3 / 2	- / 16	Milz- Volumen Ansprechrate	Gesamtüberleben; Lebensqualität (EORTC QLQ-C30)	Fishers exakter-Test
Kalydeco	2	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Phase III)	167 (84 / 83)	65 / ≥ 3 / 3	- / 20	Veränderung FEV ₁	pulmonare Exazerbationen; Lebensqualität (CFQ-R)	MMRM
		RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Phase III)	52 (26 / 26)	24 / ≥ 3 / 3	- / 14	Veränderung FEV ₁	pulmonare Exazerbationen; Lebensqualität (CFQ-R)	MMRM
Kuvan	2	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Phase III)	88 (41 / 47)	27 / ≥ 2 / 2	- / -	Phenylalanin- Blutspiegel	-	Signifikanztests
		RCT (doppelblind, Placebo- Kontrolle, multizentrisch, Phase III)	45 (33 / 12)	11 / 1 / 1	- / -	Phenylalanin- Blutspiegel	-	Signifikanztests
Litak	1	unkontrolliert, multizentrisch, multinational	63	18 / 3 / 1	36 / -	Remission	-	deskriptiv

(Fortsetzung)

Tabelle 7: Methodische Aspekte der Zulassungsstudien zu Orphan Drugs (Studienebene) (Fortsetzung)

Arzneimittel	Studien ^a	Studiendesign	Patienten [gesamt (IG / KG)]	Anzahl Zentren / Länder / Kontinente ^b	Rekrutierungs- / Studienzeit- raum ^c [Monate]	Primärer Endpunkt	Patientenrelevante sekundäre Endpunkte (ohne UE)	Statistische Methodik (primäre Analyse)
Lysodren	–	Literatur-Review: Fall-Serien, retrospektiv, multizentrisch, multinational	1064	≥ 2 / ≥ 2 / ≥ 2	– / –	–	–	–
Mepact	1	RCT (unverblindet, aktive Kontrolle, multizentrisch, multinational, Phase III)	678 ([167+171] / [174+166])	178 / ≥ 2 / –	– / –	Krankheits- freies Überleben	Gesamtüberleben	Überlebenszeitanalyse (Log-Rank-Test, stratifiziert)
Mozobil	2	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, Phase III)	298 (150 / 148)	32 / 2 / 1	– / 24	Erreichen der angestrebten CD34+ Zellzahl	–	CMH-Test
		RCT (doppelblind, Placebo- Kontrolle, multizentrisch, Phase III)	302 (148 / 154)	39 / 3 / 2	– / 36	Erreichen der angestrebten CD34+ Zellzahl	–	CMH-Test
Myozyme	2	unverblindet, historische Kontrolle, multizentrisch, multinational, Phase II/III (ongoing)	60 (18 / 42)	7 / 5 / 3	– / –	Anteil lebender Patienten ohne invasive Beatmung	Größenveränderung Anteil Patienten ohne jegliche Beatmung	Überlebenszeitanalysen (Kaplan-Meier); Interim- Analysen
		unverblindet, historische Kontrolle, multizentrisch, multinational, Phase II (ongoing)	69 (21 / 48)	6 / 4 / 3	– / –	Gesamt- überleben	Anteil lebender Patienten ohne Beatmung; motorische Entwicklung	Überlebenszeitanalysen (Hazard-Ratio), Interim- Analysen

(Fortsetzung)

Tabelle 7: Methodische Aspekte der Zulassungsstudien zu Orphan Drugs (Studienebene) (Fortsetzung)

Arzneimittel	Studien ^a	Studiendesign	Patienten [gesamt (IG / KG)]	Anzahl Zentren / Länder / Kontinente ^b	Rekrutierungs- / Studienzeit- raum ^c [Monate]	Primärer Endpunkt	Patientenrelevante sekundäre Endpunkte (ohne UE)	Statistische Methodik (primäre Analyse)
Naglazyme	1	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Phase III)	39 (19 / 20)	39 / 6 / 3	- / 10	12-Minuten- Gehtest	Treppensteigetest	Repeated measures linear analysis, adjustiert
Nexavar	1	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Phase III)	769 (384 / 385)	≥ 19 / 19 / ≥ 2	- / -	Gesamt- überleben	-	Überlebenszeitanalyse (Log-Rank-Test)
Nexobrid	1	RCT (unverblindet, aktive Kontrolle, multizentrisch, multinational, Phase III)	156 (75 / 81)	32 / ≥ 5 / 4	- / -	prozentualer Anteil der behandelten Wunde, der heraus- geschnitten oder dermabradiert wurde; prozentualer Anteil der behandelten Wunde, der autotransplan- tiert wurde	prozentualer Anteil aller Wunden, der herausgeschnitten oder dermabradiert wurde; Zeit bis zum kompletten Wundverschluss; Blutverlust	Chi-Quadrat-Test (Fishers exakter Test); sequenzielle Testprozedur, Interim- Analysen mit Stopping- Rule
NovoThirteen	1	unverblindet, historische Kontrollen, multizentrisch, multinational	41	29 / 11 / 2	- / 21	Blutungs- episoden	Blutungen	Poisson-Modell, adjustiert

(Fortsetzung)

Tabelle 7: Methodische Aspekte der Zulassungsstudien zu Orphan Drugs (Studienebene) (Fortsetzung)

Arzneimittel	Studien ^a	Studiendesign	Patienten [gesamt (IG / KG)]	Anzahl Zentren / Länder / Kontinente ^b	Rekrutierungs- / Studienzeit- raum ^c [Monate]	Primärer Endpunkt	Patientenrelevante sekundäre Endpunkte (ohne UE)	Statistische Methodik (primäre Analyse)
Nplate	2	RCT (doppelblind, Placebo- Kontrolle, Phase III)	63 (42 / 21)	35 / ≥ 2 / 2	- / -	Inzidenz des dauerhaften Thrombozy- ten- ansprechens	-	CMH-Test
		RCT (doppelblind, Placebo- Kontrolle, Phase III)	62 (41 / 21)	35 / ≥ 2 / 2	- / -	Inzidenz des dauerhaften Thrombozy- tenanspre- chens	-	CMH-Test
Onsenal	1	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Phase II)	83 ([34 + 32] / 17)	2 / 2 / 2	- / -	prozentuale Veränderung der Anzahl kolorektaler Polypen	-	Signifikanztests
Orfadin	1	unkontrolliert, multizentrisch, multinational	207	87 / 25 / -	78 / 78	Gesamt- überleben Überleben ohne Transplanta- tion Tod durch Leberversagen Transplanta- tion wegen Leberversagen hepatozellu- läres Karzinom	-	geschätzte Überlebenswahrschein- lichkeiten mit 95 %-KI

(Fortsetzung)

Tabelle 7: Methodische Aspekte der Zulassungsstudien zu Orphan Drugs (Studienebene) (Fortsetzung)

Arzneimittel	Studien ^a	Studiendesign	Patienten [gesamt (IG / KG)]	Anzahl Zentren / Länder / Kontinente ^b	Rekrutierungs- / Studienzeit- raum ^c [Monate]	Primärer Endpunkt	Patientenrelevante sekundäre Endpunkte (ohne UE)	Statistische Methodik (primäre Analyse)
Orphacol	–	Literatur-Review: Fall-Berichte: unverblindet, nicht randomisierte n-of-1- Studien, historische Kontrollen, retrospektiv, multizentrisch, multinational	49	–	– / –	–	–	–
Pedea	1	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, national)	131 (65 / 66)	11 / 1 / 1	– / 13	Anteil Patienten mit chirurgischer Ligatur der PDA	spezifische Komplikationen der PDA/Frühreife	deskriptiv
Peyona	1	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, national)	87 (46 / 41)	9 / 1 / 1	– / –	Reduktion von Apnoe- Episoden	–	Chi-Quadrat-Test
Photobarr	1	RCT (teilverblindet, Placebo- Kontrolle, multizentrisch)	208 (138 / 70)	≥ 2 / – / –	– / 47	vollständiges Ansprechen	Gesamtüberleben	Fishers exakter-Test
Plenadren	1	RCT (unverblindet, aktive Kontrolle, multizentrisch, national, Cross-over, Phase II/III)	64	5 / 1 / 1	– / 12	Serum- Cortisol AUC	tägliche Erschöpfung (VAS); Lebensqualität (SF-36, FIS, PGWB)	Signifikanztests

(Fortsetzung)

Tabelle 7: Methodische Aspekte der Zulassungsstudien zu Orphan Drugs (Studienebene) (Fortsetzung)

Arzneimittel	Studien ^a	Studiendesign	Patienten [gesamt (IG / KG)]	Anzahl Zentren / Länder / Kontinente ^b	Rekrutierungs- / Studienzeit- raum ^c [Monate]	Primärer Endpunkt	Patientenrelevante sekundäre Endpunkte (ohne UE)	Statistische Methodik (primäre Analyse)
Prialt	3	RCT (doppelblind, Placebo- Kontrolle, Phase II/III)	111 (71 / 40)	44 / 1 / 1	- / -	prozentuale Veränderung in VASPI	Schmerzlinderung (VASPI, CPRS, WBPI)	ANCOVA
		RCT (unverblindet, Placebo- Kontrolle, Phase II/III)	255 (169 / 86)	- / - / -	- / -	prozentuale Veränderung in VASPI	Schmerzlinderung (VASPI, CPRS, WBPI)	ANCOVA
		RCT (doppelblind, Placebo- Kontrolle, Phase II/III)	220 (112 / 108)	44 / 1 / 1	- / -	prozentuale Veränderung in VASPI	Schmerzlinderung (CPRS, WBPI, McGill Pain Questionnaire)	Signifikanztests
Procybsi	1	RCT (unverblindet, aktive Kontrolle, multizentrisch, Cross-over, Nichtunterlegenheitsstudie, Phase III)	43	8 / 3 / 2	- / -	Cystin-Gehalt der Leukozyten	-	Test auf Nichtunterlegenheit
Replagal	2	RCT (doppelblind, Placebo- Kontrolle, monozentrisch, Phase II)	26	1 / 1 / 1	- / -	Schmerz (BPI)	-	ANCOVA
		RCT (doppelblind, Placebo- Kontrolle, monozentrisch, Phase II)	15	1 / 1 / 1	- / -	Kardialer Gb3-Gehalt	-	ANCOVA
Revatio	1	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational)	278 ([69+68+71] / 70)	68 / 23 / 6	- / -	6-Minuten- Gehtest	Klinische Verschlechterung; Dyspnoe (Borg Dyspnoe Index)	t-Test; sequenzielle Testprozedur

(Fortsetzung)

Tabelle 7: Methodische Aspekte der Zulassungsstudien zu Orphan Drugs (Studienebene) (Fortsetzung)

Arzneimittel	Studien ^a	Studiendesign	Patienten [gesamt (IG / KG)]	Anzahl Zentren / Länder / Kontinente ^b	Rekrutierungs- / Studienzeit- raum ^c [Monate]	Primärer Endpunkt	Patientenrelevante sekundäre Endpunkte (ohne UE)	Statistische Methodik (primäre Analyse)
Revestive	2	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Phase III)	84 ([35+33] / 16)	32 / 9 / 2	30 / 39	Volumen der parenteralen Ernährung (Ansprechen)	–	Rang-ANCOVA; sequenzielle Testprozedur
		RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Phase III)	86 (43 / 43)	27 / 10 / 2	– / 27	Volumen der parenteralen Ernährung (Ansprechen)	–	CMH-Test
Revlimid	2	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Phase III)	353 (177 / 176)	48 / 2 / 1	– / 29	Zeit bis Progression	Gesamtüberleben; symptomatische skelettale Ereignisse	Überlebenszeitanalyse (Log-Rank-Test)
		RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Phase III)	351 (176 / 175)	50 / ≥ 3 / 3	– / 24	Zeit bis Progression	Gesamtüberleben; symptomatische skelettale Ereignisse	Überlebenszeitanalyse (Log-Rank-Test); Interim- Analysen
Revolade	2	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Phase III)	114 (76 / 38)	63 / 23 / 6	– / 12	Veränderung der Plättchen- anzahl	Auftreten und Schwere von Symptomen (WHO Bleeding Score, ITP Bleeding Score); Lebensqualität (SF-36)	logistisches Regressionsmodell, adjustiert
		RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Phase III)	197 (135 / 62)	75 / 23 / 6	– / 20	Veränderung der Plättchen- anzahl	Auftreten und Schwere von Symptomen (WHO Bleeding Score, ITP Bleeding Score); Lebensqualität (SF-36)	GEE, adjustiert; $\alpha = 0,01$, zweiseitig

(Fortsetzung)

Tabelle 7: Methodische Aspekte der Zulassungsstudien zu Orphan Drugs (Studienebene) (Fortsetzung)

Arzneimittel	Studien ^a	Studiendesign	Patienten [gesamt (IG / KG)]	Anzahl Zentren / Länder / Kontinente ^b	Rekrutierungs- / Studienzeit- raum ^c [Monate]	Primärer Endpunkt	Patientenrelevante sekundäre Endpunkte (ohne UE)	Statistische Methodik (primäre Analyse)
Rilonacept Regeneron	1	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, Phase III)	47 (23 / 24)	27 / 1 / 1	– / –	Symptom- Score (KSS)	Krankheitsschübe	ANCOVA
Savene	2	unkontrolliert, multizentrisch, national, Phase II	23	10 / 1 / 1	24 / –	chirurgischer Eingriff	Auftreten von Folgeerkrankungen (Sensibilitätsstörungen, Hautatrophie, Schmerzen, Bewegungseinschrän- kungen)	Häufigkeiten mit exakten Binomial-95 %-KI; Interim- Analysen
		unkontrolliert, multizentrisch, multinational, Phase II/III	57	24 / 4 / 1	38 / –	chirurgischer Eingriff	Nekrose Auftreten von Folgeerkrankungen (eingeschränkte Funktion der oberen Extremitäten, schwere Entstellung, Schmerzen, neurologische Ausfälle)	Häufigkeiten mit exakten Binomial-95 %-KI, exakter Binomial-Test; Interim- Analysen
Signifor	1	RCT (doppelblind, aktive Kontrolle, multizentrisch, multinational, Phase III)	165 (83 / 82)	53 / 18 / 4	– / –	Normalisie- rung des freien Corti- sols im Urin (Ansprechen)	Symptomverbesserung; Lebensqualität	Häufigkeiten mit 95 %-KI
Siklos	–	Register-Auswertungen Literatur-Review klinischer Studien: – Fall-Serien, retrospektiv – 1 RCT	–	–	– / –	–	–	–

(Fortsetzung)

Tabelle 7: Methodische Aspekte der Zulassungsstudien zu Orphan Drugs (Studienebene) (Fortsetzung)

Arzneimittel	Studien ^a	Studiendesign	Patienten [gesamt (IG / KG)]	Anzahl Zentren / Länder / Kontinente ^b	Rekrutierungs- / Studienzeit- raum ^c [Monate]	Primärer Endpunkt	Patientenrelevante sekundäre Endpunkte (ohne UE)	Statistische Methodik (primäre Analyse)
Soliris	1	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Phase III)	87 (43 / 44)	34 / ≥ 3 / 3	- / 17	Haemoglobin- Stabilisierung transfundierte PRBC- Einheiten	Vermeidung von Transfusionen; Symptome (FACIT Fatigue Sclae)	Wilcoxon-Rangsummentest
Somavert	2	RCT (doppelblind, Placebo- Kontrolle, Phase II)	46 ([16+15] / 15)	- / - / -	- / -	prozentuale Suppression der IGF-1- Konzentration	-	ANOVA, adjustiert; hierarchische Testprozedur
		RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Phase III)	111 ([26+26+28] / 31)	16 / ≥ 2 / 2	- / -	Veränderung des IGF-1	Symptom-Score	ANOVA; hierarchische Testprozedur
Sprycel	5	unkontrolliert, Phase II	186	- / - / -	- / 6	zytogeneti- sches Ansprechen	-	Häufigkeiten mit exakten 95 %-KI
		RCT (unverblindet, aktive Kontrolle, multizentrisch, Phase II)	150 (101 / 49)	58 / 23 / -	- / -	-	-	deskriptiv
		unkontrolliert, Phase II	107	- / - / -	- / 8	hämatologi- sches Ansprechen	-	Häufigkeiten mit exakten 95 %-KI
		unkontrolliert, multizentrisch, multinational, Phase II	74	35 / ≥ 2 / ≥ 2	- / 8	hämatologi- sches Ansprechen	-	Häufigkeiten mit exakten 95 %-KI
		unkontrolliert, Phase II	78	- / - / -	- / 7	hämatologi- sches Ansprechen	-	Häufigkeiten mit exakten 95 %-KI

(Fortsetzung)

Tabelle 7: Methodische Aspekte der Zulassungsstudien zu Orphan Drugs (Studienebene) (Fortsetzung)

Arzneimittel	Studien ^a	Studiendesign	Patienten [gesamt (IG / KG)]	Anzahl Zentren / Länder / Kontinente ^b	Rekrutierungs- / Studienzeit- raum ^c [Monate]	Primärer Endpunkt	Patientenrelevante sekundäre Endpunkte (ohne UE)	Statistische Methodik (primäre Analyse)
Sutent	1	RCT (doppelblind, Placebo- Kontrolle, Phase III)	312 (207 / 105)	61 / 11 / 4	14 / –	Zeit bis Progression	Gesamtüberleben	Überlebenszeitanalyse (Log-Rank-Test)
Tasigna	2	unkontrolliert, multizentrisch, multinational, Phase II	132	63 / 10 / 4	7 / 14	zytogeneti- sches Ansprechen	Gesamtüberleben	Häufigkeiten mit exakten 95 %-KI; Flemings single- stage, single-arm Testprozedur
		unkontrolliert, Phase II	64	– / – / –	9 / 13	hämatologi- sches Ansprechen	Gesamtüberleben	Häufigkeiten mit exakten 95 %-KI; Flemings single- stage, single-arm Testprozedur
Tepadina	–	Literatur-Review: Fall-Serien, retrospektiv	7174	–	– / –	–	–	–
Thalidomide Celgene	1	RCT (unverblindet, aktive Kontrolle, multizentrisch, multinational, Phase III)	447 (125 / [196+126])	73 / 3 / 1	– / 66	Gesamt- überleben	–	Überlebenszeitanalyse (Log-Rank-Test); Interim - Analysen

(Fortsetzung)

Tabelle 7: Methodische Aspekte der Zulassungsstudien zu Orphan Drugs (Studienebene) (Fortsetzung)

Arzneimittel	Studien ^a	Studiendesign	Patienten [gesamt (IG / KG)]	Anzahl Zentren / Länder / Kontinente ^b	Rekrutierungs- / Studienzeit- raum ^c [Monate]	Primärer Endpunkt	Patientenrelevante sekundäre Endpunkte (ohne UE)	Statistische Methodik (primäre Analyse)
Thelin	3	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Phase III)	178 ([55+63] / 60)	23 / 2 / 1	- / -	Veränderung der maximalen Sauerstoff- aufnahme- kapazität	6-Minuten-Gehtest; klinische Verschlechterung; NYHA-Klassifikation	ANCOVA
		RCT (doppelblind, aktive und Placebo-Kontrolle, multizentrisch, multinational, Phase III)	247 ([61+62] / [62+62])	55 / 4 / 4	- / -	6-Minuten- Gehtest	klinische Verschlechterung; Dyspnoe (BORG Dyspnoe Index); WHO Funktionsklasse 6-Minuten-Gehtest Hospitalisierung	nicht parametrische ANCOVA
		RCT (doppelblind, Placebo- Kontrolle multizentrisch, multinational,)	98 ([32+32] / 34)	11 / 5 / 3	- / -	6-Minuten- Gehtest	klinische Verschlechterung; WHO Funktionsklasse	t-Test
Tobi Podhaler	2	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Phase III)	102 (48 / 54)	33 / 8 / 3	- / 18	Veränderung FEV ₁	Exazerbationen; Hospitalisierungen	ANCOVA, adjustiert; Interim-Analysen
		RCT (unverblindet, Placebo- Kontrolle, multizentrisch, multinational, Phase III)	553 (329 / 224)	127 / 16 / 5	- / 38	Veränderung FEV ₁	Exazerbationen; Hospitalisierungen	ANCOVA; Nichtunterlegenheits- Analyse

(Fortsetzung)

Tabelle 7: Methodische Aspekte der Zulassungsstudien zu Orphan Drugs (Studienebene) (Fortsetzung)

Arzneimittel	Studien ^a	Studiendesign	Patienten [gesamt (IG / KG)]	Anzahl Zentren / Länder / Kontinente ^b	Rekrutierungs- / Studienzeit- raum ^c [Monate]	Primärer Endpunkt	Patientenrelevante sekundäre Endpunkte (ohne UE)	Statistische Methodik (primäre Analyse)
Torisel	1	RCT (unverblindet, aktive Kontrolle, multizentrisch, multinational, Phase III)	626 ([209+210] / 207)	148 / 23 / 6	23 / 34	Gesamt- überleben	–	Überlebenszeitanalyse (Log-Rank-Test, stratifiziert)
Tracleer	2	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, Phase III)	32 (21 / 11)	6 / 2 / 2	– / –	6-Minuten- Gehtest	klinische Verschlechterung; Dyspnoe (BORG Dyspnoe Index) WHO-Funktionsklasse Raynaud-Syndrom	t-Test (einseitig, $\alpha = 0,05$)
		RCT (doppelblind, Placebo- Kontrolle, multizentrisch, Phase III)	213 ([74+70] / 69)	27 / ≥ 2 / 4	– / –	6-Minuten- Gehtest	klinische Verschlechterung; Dyspnoe (BORG Dyspnoe Index); WHO-Funktionsklasse	Mann-Whitney-U-Test
Trisenox	2	unkontrolliert, monozentrisch, Phase I/II	12	1 / 1 / 1	– / 8	–	Gesamtüberleben	Häufigkeiten mit 95 %-KI
		unkontrolliert, multizentrisch, national, Phase II	40	9 / 1 / 1	– / 16	vollständige Remission	Gesamtüberleben	Häufigkeiten mit 95 %-KI
Ventavis	1	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, Phase III)	203 (101 / 102)	≥ 2 / – / –	– / –	kombinierter Endpunkt: Gehtest, NYHA- Klassifikation, pulmonare Hypertension, Tod	Mortalität; 6-Minuten-Gehtest; Dyspnoe (BORG / Mahler Dyspnoe Index); NYHA-Klassifikation; Notwendigkeit einer Transplantation Lebensqualität	Mantel-Haenszel-Test, Fishers exakter Test

(Fortsetzung)

Tabelle 7: Methodische Aspekte der Zulassungsstudien zu Orphan Drugs (Studienebene) (Fortsetzung)

Arzneimittel	Studien ^a	Studiendesign	Patienten [gesamt (IG / KG)]	Anzahl Zentren / Länder / Kontinente ^b	Rekrutierungs- / Studienzeit- raum ^c [Monate]	Primärer Endpunkt	Patientenrelevante sekundäre Endpunkte (ohne UE)	Statistische Methodik (primäre Analyse)
Vidaza	1	RCT (unverblindet, aktive Kontrolle, multizentrisch, multinational, Phase III)	358 (179 / 179)	79 / 15 / 3	30 / 44	Gesamt- überleben	Zeit bis zur Bildung einer AML (Blasten im Knochenmark > 30 %)	Überlebenszeitanalyse (Log-Rank-Test, stratifiziert)
Volibris	2	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Phase III)	201 ([67+67] / 67)	≥ 2 / ≥ 2 / ≥ 2	27 / -	6-Minuten- Gehtest	klinische Verschlechterung; SF-36 körperliche Funktion; Dyspnoe (BORG Dyspnoe Index)	Wilcoxon- Rangsummentest, stratifiziert; hierarchische Testprozedur
		RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Phase III)	192 ([64+63] / 65)	≥ 2 / ≥ 2 / ≥ 2	23 / -	6-Minuten- Gehtest	klinische Verschlechterung; SF-36 körperliche Funktion; Dyspnoe (BORG Dyspnoe Index)	Wilcoxon- Rangsummentest, stratifiziert; hierarchische Testprozedur
Votubia	1	unkontrolliert, monozentrisch, Phase II	28	1 / 1 / 1	24 / 36	Volumen der primären SEGA-Läsion	Krampfanfall-Frequenz; Lebensqualität (QOLCE); neuropsychologische Funktion (WPPSI-III u. a.); Lebensqualität (QOLCE)	Wilcoxon-Vorzeichen- Rang-Test
Vpriv	1	RCT (doppelblind, aktive Kontrolle, multizentrisch, multinational, Phase III)	35 (17 / 18)	11 / 9 / 5	- / 17	Veränderung Hämoglobin	-	97,5 %-KI, einseitig; Nichtunterlegenheitstest

(Fortsetzung)

Tabelle 7: Methodische Aspekte der Zulassungsstudien zu Orphan Drugs (Studienebene) (Fortsetzung)

Arzneimittel	Studien ^a	Studiendesign	Patienten [gesamt (IG / KG)]	Anzahl Zentren / Länder / Kontinente ^b	Rekrutierungs- / Studienzeit- raum ^c [Monate]	Primärer Endpunkt	Patientenrelevante sekundäre Endpunkte (ohne UE)	Statistische Methodik (primäre Analyse)
Vyndaqel	1	RCT (doppelblind, Placebo- Kontrolle, multizentrisch, multinational, Phase II/III)	128 (65 / 63)	8 / 7 / 2	12 / 29	NIS-LL-Score (Ansprechen) Lebensqualität (QOL-DN)	Veränderung NIS-LL- Score	Chi-Quadrat-Test ANCOVA
Wilzin	1	unkontrolliert, monozentrisch	148	1 / 1 / 1	- / -	-	Sprach- / neurologische Funktion	deskriptiv
Xagrid	4	unkontrolliert, multizentrisch	35	≥ 2 / - / -	- / -	vollständiges Ansprechen partielles Ansprechen	Symptome	deskriptiv
		unkontrolliert, multizentrisch	274	≥ 2 / - / -	- / -	vollständiges Ansprechen partielles Ansprechen	Symptome	deskriptiv
		unkontrolliert, multizentrisch	242	≥ 2 / - / -	- / -	vollständiges Ansprechen partielles Ansprechen	Symptome	deskriptiv
		unkontrolliert, multizentrisch	2410	≥ 2 / - / -	- / -	vollständiges Ansprechen	-	deskriptiv
Xaluprine	1	RCT (unverblindet, aktive Kontrolle, Cross-over, Bioäquivalenzstudie)	60	1 / 1 / 1	- / -	pharmako- kinetische Variablen (AUC, Cmax)	-	ANOVA, 90 %-KI für Test/Referenz-Ratio, Äquivalenztest

(Fortsetzung)

Tabelle 7: Methodische Aspekte der Zulassungsstudien zu Orphan Drugs (Studienebene) (Fortsetzung)

Arzneimittel	Studien ^a	Studiendesign	Patienten [gesamt (IG / KG)]	Anzahl Zentren / Länder / Kontinente ^b	Rekrutierungs- / Studienzeit- raum ^c [Monate]	Primärer Endpunkt	Patientenrelevante sekundäre Endpunkte (ohne UE)	Statistische Methodik (primäre Analyse)
Xyrem	2	RCT (doppelblind, Placebo- Kontrolle, multizentrisch)	136 ([34+33+35] / 34)	18 / - / -	- / -	Kataplexie- Anfälle	vollständige / partielle Kataplexie-Anfälle; hypnagoge Halluzination; Schlaf-Paralyse- Episoden; starke Schläfrigkeit bei Tage (Epworth Sleepiness Scale); Schlaf-Attacken / unbeabsichtigter kurzer Schlaf; nächtliches Aufwachen; gesamte Schlafmenge	ANOVA / ANCOVA
		RCT (doppelblind, Placebo- Kontrolle, multizentrisch)	55 (26 / 29)	14 / - / -	- / -	Kataplexie- Anfälle	vollständige / partielle Kataplexie-Anfälle; hypnagoge Halluzination; Schlaf-Paralyse- Episoden; starke Schläfrigkeit bei Tage (Epworth Sleepiness Scale); Schlaf-Attacken / unbeabsichtigter kurzer Schlaf; nächtliches Aufwachen; gesamte Schlafdauer	ANCOVA

(Fortsetzung)

Tabelle 7: Methodische Aspekte der Zulassungsstudien zu Orphan Drugs (Studienebene) (Fortsetzung)

Arzneimittel	Studien ^a	Studiendesign	Patienten [gesamt (IG / KG)]	Anzahl Zentren / Länder / Kontinente ^b	Rekrutierungs- / Studienzeit- raum ^c [Monate]	Primärer Endpunkt	Patientenrelevante sekundäre Endpunkte (ohne UE)	Statistische Methodik (primäre Analyse)
Yondelis	1	RCT (unverblindet, aktive Kontrolle, multizentrisch, multinational, Phase II)	266 (134 / 132)	≥ 8 / 8 / 4	25 / –	Zeit bis Progression	Gesamtüberleben	Überlebenszeitanalyse (Log-Rank-Test); Interim- Analysen
Zavesca	1	unkontrolliert, multizentrisch, Phase I/II	28	≥ 2 / 1 / 1	– / –	Organ- volumen hämatolo- gisches und biochemisches Ansprechen Gesamt- Ansprechen	–	deskriptiv

–: keine, unklare oder nicht relevante Angaben

a: Anzahl zulassungsrelevanter Studien („main studies“ im EPAR)

b: Kontinente: Antarktika, Nordamerika, Südamerika, Asien, Europa, Afrika, Australien

c: Studiendauer (Beginn 1. Patient bis letzte Messung letzter Patient)

d: Zeit bis zur zulassungsrelevanten Interim-Analyse

AHI: Apnoe/Hypopnoe Index; ANCOVA: Kovarianz-Analyse; ANOVA: Varianzanalyse; AUC: area under the curve; BPI: Brief Pain Inventory; CD: Cluster of Differentiation; CFQ-R: Cystic Fibrosis Questionnaire-Revised; CHAQ/HAQ: Childhood Health Assessment Questionnaire / Health Assessment Questionnaire; CMH-Test: Cochran-Mantel-Haenszel-Test; CPRS: komplexes regionales Schmerzsyndrom; EORTC QLQ-C30: European Organization for Research and Treatment of Cancer – Core Quality of Life questionnaire; EPAR: European Public Assessment Report; FACIT: Functional Assessment of Chronic Illness Therapy; FEV₁: forcierte Einsekundenkapazität; FIS: Fragebogen zu interpersonellen Schuldgefühlen; FKSI-DRS: Functional Assessment of Cancer Therapy – Kidney Symptom Index – Disease-Related Symptoms; FVC: forcierte Vitalkapazität; Gb3: Globotriaosylceramide; GEE: generalized estimating equation; GL-3: Globotriaosylceramid; GVHD: Graft-versus-Host Disease; IG: Interventionsgruppe; IGF-1: insulinähnlicher Wachstumsfaktor; ITP: Immune thrombocytopenic purpura; KG: Kontrollgruppe; KI: Konfidenzintervall; KSS: Key Symptom Score; LQ: Lebensqualität; MMRM: Mixed Model Repeated Measures; MRI: Magnetresonanztomografie; NIS-LL: Neurologic Impairment Score-Lower Limb; NYHA: New York Heart Association; PA: Pseudomonas aeruginosa; PDA: persistierender Ductus arteriosus; PGWB: Psychological General Well-Being; PRBC: packed red blood cells; PRO: patientenberichteter Endpunkt; QOLCE: Quality of Life Childhood Epilepsy; QOL-DN: Norfolk Quality of Life Questionnaire-Diabetic Neuropathy; RCT: randomisierte, kontrollierte Studie; SEGA: subependymal giant cell astrocytoma; SF-36: Short Form (36) Gesundheitsfragebogen; UCSD-SOBQ: University of California, San Diego – Shortness of Breath Questionnaire; UE: unerwünschte Ereignisse; UFC: urinary free cortisol; VAS: visuelle Analog-Skala; VASPI: Visual Analogue Scale of Pain Intensity; WBPI: Wisconsin Brief Pain Inventory; WHO: World Health Organization; WPPSI-III: Wechsler Preschool and Primary Scale of Intelligence

Tabelle 8: Übersicht der Zulassungsstudien zu Orphan Drugs (Arzneimittlebene)

Arzneimittel	Erkrankung sehr selten	RCT ^a	RCT: doppel- blind	RCT: aktive Kontrolle	RCT: Nichtunterlegen- heit / Äquivalenz	Patientenrelevante Endpunkte		Multizentrisch / -national / -kontinental ^a	Spezielle statistische Methodik
						primär	sekundär		
						Mort. / Morb. / LQ	Mort. / Morb. / LQ		
Adcetris	○	○				○ / ○ / ○	● / ○ / ○	● / ● / ●	
Afinitor	○	●	●	○	○	○ / ○ / ○	● / ● / ●	● / ● / ●	● ^b
Aldurazyme	●	●	●	○	○	○ / ● / ○	○ / ● / ○	● / ● / ●	○
Arzerra	○	○				○ / ○ / ○	● / ● / ●	● / ● / -	
Atriance	○	○				○ / ○ / ○	● / ○ / ○	● / ● / ●	
Bosulif	○	●	○	●	○	○ / ○ / ○	● / ○ / ○	● / ● / ●	● ^c
Bronchitol	○	●	●	●	○	○ / ○ / ○	○ / ● / ●	● / ● / ●	● ^c
Busilvex	●	○				○ / ○ / ○	● / ○ / ○	● / ● / -	
Carbaglu	○	○				○ / ○ / ○	○ / ● / ○	● / ● / ○	
Cayston	○	●	●	○	○	○ / ● / ○	○ / ● / ○	● / ● / ●	● ^c
Ceplene	○	●	○	●	○	○ / - / ○	● / - / ●	● / ● / ●	● ^c
Cystadane	○								
Dacogen	○	●	○	●	○	● / ○ / ○	○ / ○ / ○	● / ● / ●	● ^{b,c}
Defitelio	○	●	○	○	○	○ / - / ○	● / - / ○	● / ● / ●	● ^b
Diacomit	●	●	●	○	○	○ / ● / ○	○ / ● / ○	● / ○ / ○	○
Elaprase	●	●	●	○	○	○ / ● / ○	○ / ○ / ○	● / ● / ●	● ^c
Esbriet	○	●	●	○	○	○ / ○ / ○	○ / ● / ○	● / ● / ●	● ^c

(Fortsetzung)

Tabelle 8: Übersicht der Zulassungsstudien zu Orphan Drugs (Arzneimittlebene) (Fortsetzung)

Arzneimittel	Erkrankung sehr selten	RCT ^a	RCT: doppel- blind	RCT: aktive Kontrolle	RCT: Nichtunterlegen- heit / Äquivalenz	Patientenrelevante Endpunkte		Multizentrisch / -national / -kontinental ^a	Spezielle statistische Methodik
						primär	sekundär		
						Mort. / Morb. / LQ	Mort. / Morb. / LQ		
Evoltra	○	○				○ / ○ / ○	● / ○ / ○	● / ○ / ○	
Exjade	●	●	○	●	●	○ / ○ / ○	○ / ○ / ○	● / ○ / ○	○
Fabrazyme	●	●	●	○	○	○ / ○ / ○	○ / ● / ○	● / ● / ●	○
Firazyr	●	●	●	●	○	○ / ● / ○	○ / ● / ○	● / ● / ●	○
Firdapse	●								
Gliolan	●	●	○	●	○	○ / ○ / ○	● / ○ / ○	● / ● / -	○
Glivec	○	○				○ / ○ / ○	○ / ○ / ○	- / - / -	
Glybera	●	○				○ / ○ / ○	○ / ● / ○	● / ○ / ○	
Iclusig	○	○				○ / ○ / ○	● / ○ / ○	● / ● / ●	
Ilaris	○	●	●	○	○	○ / - / ○	○ / ● / ○	● / ● / ●	○
Imnovid	○	●	○	●	○	○ / ○ / ○	● / ● / ●	● / ● / ●	● ^{b,c}
Increlex	○	●	●	○	○	○ / ● / ○	○ / ● / ○	● / ● / ●	● ^d
Inovelon	○	●	●	○	○	○ / ● / ○	○ / ● / ○	● / ● / ●	○
Jakavi	○	●	●	●	○	○ / ○ / ○	● / ● / ●	● / ● / ●	● ^c
Kalydeco	○	●	●	○	○	○ / ○ / ○	○ / ● / ●	● / ● / ●	● ^c
Kuvan	○	●	●	○	○	○ / ○ / ○	○ / ○ / ○	● / ● / ●	○
Litak	○	○				○ / ○ / ○	○ / ○ / ○	● / ● / ○	

(Fortsetzung)

Tabelle 8: Übersicht der Zulassungsstudien zu Orphan Drugs (Arzneimittlebene) (Fortsetzung)

Arzneimittel	Erkrankung sehr selten	RCT ^a	RCT: doppel- blind	RCT: aktive Kontrolle	RCT: Nichtunterlegen- heit / Äquivalenz	Patientenrelevante Endpunkte		Multizentrisch / -national / -kontinental ^a	Spezielle statistische Methodik
						primär	sekundär		
						Mort. / Morb. / LQ	Mort. / Morb. / LQ		
Lysodren	●								
Mepact	○	●	○	●	○	○ / - / ○	● / ○ / ○	● / ● / -	● ^c
Mozobil	○	●	●	○	○	○ / ○ / ○	○ / ○ / ○	● / ● / ●	● ^c
Myozyme	●	○				● / ● / ○	○ / ● / ○	● / ● / ●	
Naglazyme	●	●	●	○	○	○ / ● / ○	○ / ● / ○	● / ● / ●	● ^c
Nexavar	○	●	●	○	○	● / ○ / ○	○ / ○ / ○	● / ● / ●	● ^c
Nexobrid	○	●	○	●	○	○ / ● / ○	○ / ● / ○	● / ● / ●	● ^b
NovoThirteen	●	○				○ / ● / ○	○ / ● / ○	● / ● / ●	
Nplate	○	●	●	○	○	○ / - / ○	○ / - / ○	● / ● / ●	● ^c
Onsenal	○	●	●	○	○	○ / ● / ○	○ / ○ / ○	● / ● / ●	○
Orfadin	●	○				● / ● / ○	○ / ○ / ○	● / ● / -	
Orphacol	●								
Pedea	○	●	●	○	○	○ / ● / ○	○ / ● / ○	● / ○ / ○	○
Peyona	○	●	●	○	○	○ / ● / ○	○ / ○ / ○	● / ○ / ○	○
Photobarr	○	●	○	○	○	○ / - / ○	● / - / ○	● / - / -	○
Plenadren	○	●	○	●	○	○ / ○ / ○	○ / ● / ●	● / ○ / ○	● ^d
Prialt	○	●	●	○	○	○ / ● / ○	○ / ● / ○	● / - / -	● ^c

(Fortsetzung)

Tabelle 8: Übersicht der Zulassungsstudien zu Orphan Drugs (Arzneimittlebene) (Fortsetzung)

Arzneimittel	Erkrankung sehr selten	RCT ^a	RCT: doppel- blind	RCT: aktive Kontrolle	RCT: Nichtunterlegen- heit / Äquivalenz	Patientenrelevante Endpunkte		Multizentrisch / -national / -kontinental ^a	Spezielle statistische Methodik
						primär	sekundär		
						Mort. / Morb. / LQ	Mort. / Morb. / LQ		
Procysbi	●	●	○	●	●	○ / ○ / ○	○ / ○ / ○	● / ● / ●	● ^d
Replagal	●	●	●	○	○	○ / ● / ○	○ / ○ / ○	○ / ○ / ○	● ^c
Revatio	●	●	●	○	○	○ / ● / ○	○ / ● / ○	● / ● / ●	○
Revestive	○	●	●	○	○	○ / - / ○	○ / - / ○	● / ● / ●	● ^c
Revlimid	○	●	●	○	○	○ / ○ / ○	● / ● / ○	● / ● / ●	● ^b
Revolade	○	●	●	○	○	○ / ○ / ○	○ / ● / ●	● / ● / ●	● ^c
Rilonacept Regeneron	○	●	●	○	○	○ / ● / ○	○ / ● / ○	● / ● / ○	● ^c
Savene	●	○				○ / ● / ○	○ / ● / ○	● / ● / ○	
Signifor	○	●	●	●	○	○ / ○ / ○	○ / ● / ●	● / ● / ●	○
Siklos	○								
Soliris	●	●	●	○	○	○ / - / ○	○ / ● / ○	● / ● / ●	○
Somavert	○	●	●	○	○	○ / ○ / ○	○ / ○ / ○	● / ● / ●	● ^c
Sprycel	○	●	○	●	○	○ / ○ / ○	○ / ○ / ○	● / ● / ●	○
Sutent	○	●	●	○	○	○ / ○ / ○	● / ○ / ○	● / ● / ●	○
Tasigna	○	○				○ / ○ / ○	● / - / ○	● / ● / ●	
Tepadina	●								
Thalidomide Celgene	○	●	○	●	○	● / ○ / ○	○ / ○ / ○	● / ● / ○	● ^b

(Fortsetzung)

Tabelle 8: Übersicht der Zulassungsstudien zu Orphan Drugs (Arzneimittlebene) (Fortsetzung)

Arzneimittel	Erkrankung sehr selten	RCT ^a	RCT: doppel- blind	RCT: aktive Kontrolle	RCT: Nichtunterlegen- heit / Äquivalenz	Patientenrelevante Endpunkte		Multizentrisch / -national / -kontinental ^a	Spezielle statistische Methodik
						primär	sekundär		
						Mort. / Morb. / LQ	Mort. / Morb. / LQ		
Thelin	●	●	●	●	○	○ / ● / ○	○ / ● / ○	● / ● / ●	● ^c
Tobi Podhaler	○	●	●	○	●	○ / ○ / ○	○ / ● / ○	● / ● / ●	● ^{b,c}
Torisel	○	●	○	●	○	● / ○ / ○	○ / ○ / ○	● / ● / ●	● ^c
Tracleer	○	●	●	○	○	○ / ● / ○	○ / ● / ○	● / ● / ●	● ^e
Trisenox	○	○				○ / ○ / ○	● / ○ / ○	● / ○ / ○	
Ventavis	●	●	●	○	○	○ / ● / ○	● / ● / ●	● / - / -	● ^c
Vidaza	○	●	○	●	○	● / ○ / ○	○ / ● / ○	● / ● / ●	● ^c
Volibris	●	●	●	○	○	○ / ● / ○	○ / ● / ○	● / ● / ●	● ^c
Votubia	○	○				○ / ○ / ○	○ / ● / ●	○ / ○ / ○	
Vpriv	●	●	●	●	●	○ / ○ / ○	○ / ○ / ○	● / ● / ●	○
Vyndaqel	●	●	●	○	○	○ / ● / ●	○ / ● / ○	● / ● / ●	● ^c
Wilzin	○	○				○ / ○ / ○	○ / ● / ○	○ / ○ / ○	
Xagrid	○	○				○ / ○ / ○	○ / ● / ○	● / - / -	
Xaluprine	○	●	○	●	●	○ / ○ / ○	○ / ○ / ○	○ / ○ / ○	● ^d
Xyrem	○	●	●	○	○	○ / ● / ○	○ / ● / ○	● / - / -	● ^c
Yondelis	○	●	○	●	○	○ / ○ / ○	● / ○ / ○	● / ● / ●	● ^b
Zavesca	○	○				○ / ○ / ○	○ / ○ / ○	● / ○ / ○	

(Fortsetzung)

Tabelle 8: Übersicht der Zulassungsstudien zu Orphan Drugs (Arzneimittlebene) (Fortsetzung)

●: ja; ○: nein; -: keine oder unklare Angaben

a: Bei mehreren Studien zu einem Arzneimittel erfolgt der Eintrag „ja“, falls mindestens 1 Studie die entsprechende Eigenschaft aufweist.

b: Adaptive Verfahren / sequenzielle Verfahren / Interim-Analysen

c: Kontrolle von Störgrößen (Regressionsmodell / Stratifizierung mit Berücksichtigung von Messungen im Zeitverlauf und/oder prognostischen Faktoren und Baseline-Werten)

d: Cross-over-Studie

e: Erhöhung des Irrtumsniveaus

LQ: Lebensqualität; Morb.: Morbidität; Mort.: Mortalität; RCT: randomisierte kontrollierte Studie

Anhang B – Offenlegung potenzieller Interessenkonflikte des externen Reviewers

Im Folgenden sind die potenziellen Interessenkonflikte des externen Reviewers dargestellt. Alle Informationen beruhen auf Selbstangaben der einzelnen Personen anhand des „Formblatts zur Offenlegung potenzieller Interessenkonflikte“ mit Stand 12/2011. Das Formblatt ist unter <http://www.iqwig.de> abrufbar. Die in diesem Formblatt aufgeführten Fragen finden sich im Anschluss an diese Zusammenfassung.

Externer Reviewer

Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6
Norbert Benda (BfArM)	ja	nein	nein	nein	nein	nein

Im „Formblatt zur Offenlegung potenzieller Interessenkonflikte“ wurden folgende 6 Fragen gestellt (Version 12/2011):

Frage 1: Sind oder waren Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor angestellt bei einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere bei einem pharmazeutischen Unternehmen, einem Hersteller von Medizinprodukten oder einem industriellen Interessenverband?

Frage 2: Beraten Sie oder haben Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor ein Unternehmen, eine Institution oder einen Interessenverband im Gesundheitswesen, insbesondere ein pharmazeutisches Unternehmen, einen Hersteller von Medizinprodukten oder einen industriellen Interessenverband, direkt oder indirekt beraten?

Frage 3: Haben Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor direkt oder indirekt von einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere einem pharmazeutischen Unternehmen, einem Hersteller von Medizinprodukten oder einem industriellen Interessenverband, Honorare erhalten für Vorträge, Stellungnahmen oder Artikel?

Frage 4: Haben Sie und / oder hat die Einrichtung⁵, die Sie vertreten, abseits einer Anstellung oder Beratungstätigkeit innerhalb des laufenden Jahres und der 3 Kalenderjahre davor von einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere einem pharmazeutischen Unternehmen, einem Hersteller von Medizinprodukten oder einem industriellen Interessenverband, finanzielle Unterstützung für Forschungsaktivitäten, andere wissenschaftliche Leistungen oder Patentanmeldungen erhalten?

Frage 5: Haben Sie und / oder hat die Einrichtung⁵, bei der Sie angestellt sind bzw. die Sie vertreten, innerhalb des laufenden Jahres und der 3 Kalenderjahre davor sonstige finanzielle oder geldwerte Zuwendungen (z. B. Ausrüstung, Personal, Unterstützung bei der Ausrichtung einer Veranstaltung, Übernahme von Reisekosten oder Teilnahmegebühren ohne wissenschaftliche Gegenleistung) erhalten von einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere von einem pharmazeutischen Unternehmen, einem Hersteller von Medizinprodukten oder einem industriellen Interessenverband?

Frage 6: Besitzen Sie Aktien, Optionsscheine oder sonstige Geschäftsanteile eines Unternehmens oder einer anderweitigen Institution, insbesondere von einem pharmazeutischen Unternehmen oder einem Hersteller von Medizinprodukten? Besitzen Sie Anteile eines „Branchenfonds“, der auf pharmazeutische Unternehmen oder Hersteller von Medizinprodukten ausgerichtet ist?

⁵ Sofern Sie in einer ausgedehnten Institution tätig sind, genügen Angaben zu Ihrer Arbeitseinheit, zum Beispiel Klinikabteilung, Forschungsgruppe etc.