Dr. rer. nat. Gerrit Müller
Dr. rer. nat. Lena Fanter
Dipl.-Ges.oec. Julia Knierim

Analyse von Orphan-Drug-Verfahren in der frühen Nutzenbewertung: RCTs versus bestverfügbare vergleichende Evidenz

Zusammenfassung

Für Arzneimittel zur Therapie seltener Erkrankungen (Orphan Drugs) stellen die Anforderungen der frühen Nutzenbewertung nach § 35a SGB V eine besondere Herausforderung dar. Das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) schlussfolgert im Arbeitspapier GA21-01 zu Orphan Drugs, die sich im Rahmen einer Re-Evaluation einer regulären Nutzenbewertung unterzogen haben, dass der Zusatznutzen der initialen Nutzenbewertung zum Marktzugang fiktiv gewesen sei. Deshalb sollten für Orphan Drugs grundsätzlich nur randomisierte kontrollierte Studien (RCT) mit geeignetem aktiven Kontrollarm für die Bewertung des Zusatznutzens herangezogen werden. Eine eigene Analyse auf Basis der gleichen Datengrundlage zeigt, dass zwischen initialer und regulärer Nutzenbewertung durchschnittlich 1.146 Tage liegen und sich die therapeutischen Ausgangssituationen unterscheiden. Während knapp 66% der Orphan Drugs (27/41) zum Zeitpunkt der initialen Bewertung über einen Solistenstatus verfügten, hatten circa 30% (8/27) diesen Status zum Zeitpunkt der Re-Evaluation verloren. Allein dieser Umstand reduzierte die Wahrscheinlichkeit für einen Zusatznutzen in regulären Verfahren deutlich. In 75% dieser Verfahren (6/8) wurde kein Zusatznutzen attestiert. Ebenso der Fakt, dass die Häufigkeit einer Erkrankung, gerade im Kontext der seltenen Leiden, für die Durchführbarkeit einer RCT von wesentlicher Bedeutung ist, wurde durch das IQWiG ignoriert. In den Bewertungsverfahren mit eingereichter RCT lag die durchschnittliche Größe der Zielpopulation bei 2.445 Patienten während diese in den Verfahren ohne RCT bei 471 Patienten lag. In den Orphan-Drug-Verfahren mit zugrunde gelegter RCT (30/41) entsprach in 10 Verfahren der Vergleichsarm nicht der vom Gemeinsamen Bundesausschuss (G-BA) festgelegten zweckmäßigen Vergleichstherapie, was ausschlaggebend dafür war, dass jeweils kein Zusatznutzen festgestellt werden konnte. Sofern der Vergleichsarm den G-BA Anforderungen entsprach (20/30), so wurde in 85% (17/20) ein Zusatznutzen zuerkannt. Als adäquate Umsetzung der zweckmäßigen Vergleichstherapie wäre mehrheitlich ein Placebo-Vergleichsarm als ausreichend bewertet worden (15/20). Von den 17 Orphan-Drug-Verfahren mit RCT und Zusatznutzen im regulären Bewertungsverfahren betrug die durchschnittliche Größe der Zielpopulation 1.650 Patienten (10/17), wenn der klinische Vergleich gegen Placebo geführt wurde, vs. 3.439 Patienten bei einem aktiven Komparator in der RCT (7/17). Die Analyse zeigt, dass RCT bei seltenen Erkrankungen durchgeführt werden, sofern die Prävalenz der Krankheit relativ hoch ist. Etwa 85% der seltenen Krankheiten haben jedoch eine Prävalenz von weniger als 1 zu 1.000.000, was bedeutet, dass weniger als 80 GKV-Patienten in Deutschland an diesen Krankheiten leiden. Die Forderung des IQWiG nach vergleichender RCT-generierter Evidenz für alle Orphan Drugs sollte daher kritisch hinterfragt werden (Nguengang Wakap, Lambert et al. 2020). Für diese niedrigen Patientenzahlen bei bestimmten seltenen Krankheiten sollte die bestverfügbare vergleichende Evidenz sorgfältig daraufhin geprüft werden, ob die Daten zur Beantwortung der wissenschaftlichen Fragestellung der Nutzenbewertung geeignet sind. Für die Zukunft halten die Autoren gemeinsame Anstrengungen zur Entwicklung einer adäquaten und akzeptablen Methodik unter Anerkennung einer höheren Ergebnisunsicherheit für notwendig.

Schlüsselwörter

AMNOG, Orphan Drugs, Arzneimittel, IQWiG, Nutzenbewertung

Crossref/doi

http://doi.org/10.24945/MVF.02.23.1866-0533.2483

Zur Beurteilung der Wirksamkeit und Sicherheit eines neuen Wirkstoffes ist für die europäische Zulassungsbehörde (European Medicines Agency, EMA) die randomisierte kontrollierte Studie (RCT) der Goldstandard. Mit der Einführung des AMNOG (Arzneimittelmarktneuordnungsgesetz) am 01.01.2011 rückten methodische Aspekte zur geeigneten Beantwortung der relevanten Fragestellung, nämlich der Nutzenbewertung innovativer Pharmakotherapien gemäß § 35a SGB V, in den Vordergrund. Im Grundsatz wird für die Evaluation der relativen Wirksamkeit eines neuen Arzneimittels gegenüber der zweckmäßigen Vergleichstherapie (zVT) klinische Evidenz aus RCT als Bewertungsgrundlage herangezogen. Für Arzneimittel zur Therapie seltener Erkrankungen (Orphan Drugs, OD) bringt die Evidenzgenerierung mittels RCT besondere Herausforderungen mit sich. Aufgrund dessen traten für Orphan Drugs im Jahr 2000 die EU-gemeinschaftsrechtlichen Besonderheiten in Kraft mit dem Ziel, die Erforschung, Entwicklung und das Inverkehrbringen geeigneter Arzneimittel durch die pharmazeutische Industrie zu fördern (Verordnung (EG) Nr. 141/2000). Vor demselben Hintergrund gelten auch im AMNOG-Verfahren für Orphan Drugs besondere Regelungen, die die Ziele der europäischen Verordnung auf nationaler Ebene implementieren (§ 35a Abs. 1 S. 11 f. SGB V.). So gilt der Zusatznutzen von Orphan Drugs mit der Zulassung als belegt. Am 23.12.2021 veröffentlichte das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) das Arbeitspapier GA21-01, im Rahmen dessen es sich kritisch mit dem Evidenzkörper und der Nutzenbewertung von Orphan Drugs auseinandersetzt (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) 2021). In dem Arbeitspapier nimmt das IQWiG einen Vergleich vor zwischen einer Nutzenbewertung der Orphan Drugs, die die Besonderheiten bei Markteintritt berücksichtigt (im Folgenden OD-Nutzenbewertung), und einer regulären Nutzenbewertung. Das IQWiG schlussfolgert aus seiner Analyse im Arbeitspapier GA21-01, dass auch für Orphan Drugs ausschließlich randomisiert kontrollierte Studien mit Vergleich gegenüber der zweckmäßigen Vergleichstherapie als Evidenzbasis für die Bewertung des Zusatznutzens berücksichtigt werden sollten.

Methodik

>> Die Autoren führten eine eigene Analyse der vom IQWiG extrahierten Datenbasis durch. Hierfür wurden aus den Beschlüssen des Gemeinsamen Bundesausschusses (G-BA) die Anzahl der Patienten bzw. der für die Behandlung infrage kommenden Patientengruppen zum Zeitpunkt der initialen OD-Nutzenbewertung, die Anzahl der Patienten in den jeweiligen bewertungsrelevanten Studien, die Art der Studie (RCT, ja oder nein), der Zusatznutzen (Zusatznutzen, ja oder nein) im jeweiligen Verfahren und der Solistenstatus (Solist, ja oder nein) zum Zeitpunkt des jeweiligen Verfahrens extrahiert. Aus den G-

— — —

Genderhinweis: Aus Gründen der besseren Lesbarkeit wird auf eine geschlechtsneutrale Differenzierung verzichtet. Entsprechende Begriffe gelten im Sinne der Gleichbehandlung grundsätzlich für alle Geschlechter. Die verkürzte Sprachform beinhaltet keine Wertung.

BA-Beschlussdaten wurde der Zeitraum zwischen Verfahren in Tagen berechnet. Für die regulären Nutzenbewertungen wurde zudem die Frage erhoben, ob die zweckmäßige Vergleichstherapie innerhalb des Verfahrens bedient wurde (zVT, ja oder nein). Nachfolgend wurden die durch das IQWiG postulierten Aussagen anhand der extrahierten Daten überprüft (Appendix). Die Ergebnisse dieser Analyse sind im Folgenden dargestellt.

Ergebnis

1) Orphan-Drug-Nutzenbewertung versus regulärer Nutzenbewertung – unterschiedliche Ausgangssituationen der Bewertungsverfahren

Sofern AMNOG-relevante Orphan Drugs eine Umsatzgrenze von 50 Mio. Euro pro Jahr überschreiten oder der Orphan-Drug-Status zurückgegeben wird, durchlaufen diese ein weiteres reguläres Nutzenbewertungsverfahren. Im Gegensatz zu dem initialen OD-Bewertungsprozess, der den regulatorischen Besonderheiten der Orphan-Drug-Zulassung Rechnung trägt (ein Zusatznutzen gilt mit Zulassung als belegt), muss der Zusatznutzen im Rahmen des regulären Bewertungsprozesses gegenüber der zweckmäßigen Vergleichstherapie gemäß den methodischen Anforderungen des G-BA demonstriert werden. In dem IQWiG-Bericht GA21-01 wird seitens der Autoren beanstandet, dass der erteilte Zusatznutzen zum Zeitpunkt des Marktzugangs von Orphan Drugs im Rahmen der OD-Bewertung häufig irreführend sei, da sich in späteren regulären Re-Evaluationen kein Zusatznutzen gegenüber der zweckmäßigen Vergleichstherapie ergeben würde. Eine gemäß IQWiG adäquate Abbildung des Therapiestandards hängt zentral von der Therapiesituation zum Zeitpunkt der Durchführung der regulären Zweitbewertung ab. Oftmals ändert sich in dem Zeitraum zwischen Erst- und Zweitbewertung der Behandlungsstandard, sodass die zulassungsbegründende Evidenz der Initialbewertung allein deshalb nicht mehr vom IQWiG berücksichtigt wird, weil diese nicht die geforderte als zweckmäßig eingeschätzte Vergleichstherapie abbildet. Die mediane Dauer zwischen initialem OD- und regulärem Nutzenbewertungsverfahren liegt bei 994 Tagen (P_{25} 581, P_{75} 1477) und durchschnittlich bei 1146 Tagen (SD 706,6) (Tab. 1, eigene Berechnung; IQWiG Arbeitspapier GA21-01). Das heißt, knapp 3 Jahre liegen durchschnittlich zwischen beiden Nutzenbewertungen.

Des Weiteren postuliert das IQWiG im Rahmen seines Arbeitspapiers, dass ein irreführender fiktiver Zusatznutzen der Orphan Drugs bereits vorhandene Therapieoptionen benachteiligt. Dabei sind Orphan Drugs dadurch gekennzeichnet, dass sie zum Zeitpunkt der Zulassung entweder über einen Solistenstatus verfügen oder einen „erheblichen Nutzen" zeigen im Vergleich zur aktuellen Versorgungssituation (Verordnung (EG) Nr. 141/2000). Das heißt, dass keine weitere zugelassene Therapie in der betreffenden Indikation zur Verfügung steht oder für das relevante Orphan Drug die vorgelegte Evidenz von der europäischen Zulassungsbehörde in Bezug auf das Nutzen-Risiko-Profil so bewertet wurde, dass von einem erheblichen Nutzen gegenüber bisher verfügbaren Therapien für die Patienten auszugehen ist.

In dem zugrundeliegenden Datensatz vom IQWiG mit 41 Orphan-Drug-Verfahren haben 27 (65,9%) einen Solistenstatus zum Zeitpunkt der initialen OD-Nutzenbewertung. Das heißt, zu diesem Zeitpunkt lag keine zugelassene Therapie für die jeweilige seltene Erkrankung vor (eigene Berechnung; IQWiG Arbeitspapier GA21-01). In 8 von diesen 27 (29,6%) Verfahren mit Solistenstatus haben die Orphan Drugs zum Zeitpunkt der Re-Evaluation ihren Solistenstatus verloren (Tab. 2, eigene Berechnung; IQWiG Arbeitspapier GA21-01). Für reguläre Orphan-Drug-Verfahren, deren Solistenstatus nicht verloren ging, resultiert mit einer 2,5-fach höheren Wahrscheinlichkeit, dass ein Zusatznutzen ausgesprochen wird im Vergleich zu den Verfahren, in denen in der Zeit zwischen Erst- und Zweitbewertung eine weitere Therapieoption zugelassen wurde. Bei 6 der 8 (75,0%) regulären Nutzenbewertungen von Orphan Drugs, bei denen zwischen Erst- und Zweitbewertung eine weitere Therapieoption zugelassen wurde, wurde vom G-BA kein Zusatznutzen erteilt.

Es sollte hinterfragt werden, dass zum Zeitpunkt der nachfolgenden Vollbewertung (im Median nach 994 Tagen, durchschnittlich nach 1.146 Tagen) Evidenz vorliegen muss, die nicht nur, wie vom IQWiG eingefordert, auf einer direkt vergleichenden RCT beruht, sondern zusätzlich zum Zeitpunkt des Zweitverfahrens den dann aktuellen therapeutischen Standard im Sinne der zweckmäßigen Vergleichstherapie berücksichtigt. Eine direkt vergleichende Studie gegen einen in der Zwischenzeit neu benannten Therapiestandard (zweckmäßige Vergleichstherapie) ist in dieser Zeit faktisch nicht umsetzbar. Genau dieser Umstand führte bei 8 Nutzenbewertungen der zugrundeliegenden 41 Orphan-Drug-Verfahren dazu, dass bei 6 Verfahren kein Zusatznutzen belegt werden konnte, da zwischen beiden Evaluationsprozessen neue Therapien in der jeweiligen Indikation in den Verkehr gebracht wurden. Dieser Fakt wird vom IQWiG in seiner Bewertung nicht berücksichtigt. Hier einen fiktiven Zusatznutzen bei Marktzugang zu unterstellen, erscheint unangemessen. Es stellt sich zudem generell die Frage, wie bei den 27 Orphan Drugs mit Solistenstatus (65,9%) zum Zeitpunkt der Zulassung von einem fiktiven Zusatznutzen zu sprechen ist, da schlicht keine zugelassene Therapiealternative verfügbar ist. Die Darstellung des IQWiG, dass zur Markteinführung ein fiktiver Zusatznutzen vorlag, bedarf einer differenzierten, seriösen Betrachtung.

Zeitunterschied zwischen Orphan-Drug- und regulären Nutzenbewertungsverfahren

n	Mittelwert (SD)	Median (P_{25}, P_{75})
41	1.145,9 (706,6)	994 (581, 1.477)

Tab. 1: Dauer in Tagen zwischen Orphan-Drug- und regulärem Nutzenbewertungsverfahren. Legende: SD, Standardabweichung; P_{25}, 25%-iges Perzentil; P_{75}, 75%-iges Perzentil.

2) Prävalenz als Maßstab für die Evidenzgrundlage – wo liegt der Unterschied in der Umsetzbarkeit einer RCT?

Für eine erfolgreiche Durchführung einer RCT spielt die Epidemiologie der Erkrankung bzw. die Größe der Population der Erkrankten, für die eine Therapie infrage kommt, eine kritische Rolle. Das IQWiG konkludiert allgemein, dass RCT auch für Orphan Drugs möglich sind. Eine differenzierte Analyse der Datenbasis des IQWiG-Arbeitspapiers GA21-01 der insgesamt 41 Orphan-Drug-Verfahren zeigt, dass die Anzahl der für die Therapie in Frage kommenden Patienten gemäß den Angaben aus dem jeweiligen G-BA-Beschluss (Zielpopulation) auch mit der Größe der Studienpopulation assoziiert ist. Also je größer die Zielpopulation ist, desto größer ist auch die Studienpopulation (Korrelationskoeffizient nach nach Spearman 0.7471, p-Wert < 0,001) (Tab. 3, eigene Berechnung; IQWiG Arbeitspapier GA21-01).

Auch die Durchführbarkeit einer RCT ist mit der Größe der Population assoziiert. Je größer eine Population ist, die für eine Studie infrage kommt, desto einfacher ist die Umsetzbarkeit einer RCT. Es ist deshalb auch folgerichtig, dass sich die Größe der Populationen zwischen den

Zusatznutzen und Solistenstatus[1] von Orphan-Drug-Verfahren-Nutzenbewertungsverfahren		
	Solistenstatus verloren: nein [n (%)]	Solistenstatus verloren: ja [n (%)]
Zusatznutzen	12 (63,2)	2 (25,0)
Kein Zusatznutzen	7 (36,8)	6 (75,0)
Risk Ratio (0,95-KI), p-Wert nach Fisher	2,5 (0,72; 8,80), 0,103	

Tab. 2: Zusatznutzen und Solistenstatus[1] von Orphan-Drug-Verfahren (n = 27). Legende: [1]: Solistenstatus zum Zeitpunkt der initialen Zusatznutzenbewertung bei Marktzugang im Vergleich zum Zeitpunkt der regulären Zusatznutzenbewertung; KI, Konfidenzintervall.

Nutzenbewertungsverfahren mit und ohne RCT unterscheidet. Es wurden hierzu die 41 Orphan-Drug-Verfahren aus dem IQWiG-Papier anhand der Anzahl der für die Therapie infrage kommenden Patienten gemäß G-BA-Beschluss untersucht. In 30 (73,2%) Bewertungsverfahren, die auf Basis einer RCT durchgeführt wurden, lag die mittlere Größe der Population bei 2.445 Patienten (SD 467,5) mit einem Median von 1.725 Patienten (P_{25} 250 Patienten, P75 3550 Patienten), gegenüber 471 Patienten (SD 193,1) und einem Median von 440 Patienten (P_{25} 17 Patienten, P_{75} 455 Patienten) in 11 (26,8%) Verfahren ohne RCT als Bewertungsgrundlage. Da die Daten keine Normalverteilung aufweisen (Shapiro-Wilk-Test, p-Wert < 0,001), wurde der Unterschied anhand des Rangsummentests bewertet. Hierbei zeigte sich eine statistische Signifikanz (p-Wert 0,0123) (Tab. 4, eigene Berechnung; IQWiG Arbeitspapier GA21-01). Das lässt die Schlussfolgerung zu, dass die tatsächliche Durchführung einer RCT statistisch signifikant mit der Größe die Zielpopulation assoziiert ist.

Bei seltenen Krankheiten mit relativ hoher Prävalenz werden RCT durchgeführt. Pharmazeutische Unternehmen schließen also keinesfalls grundsätzlich RCT für seltene Erkrankungen durchzuführen aus, dennoch ist die operative Durchführbarkeit bei geringen Patientenzahlen in multizentrischen, internationalen Studien eine große Herausforderung. Dies bestätigt abermals, dass ein differenzierterer Blick auf die jeweilige Epidemiologie der seltenen Erkrankung und der Umsetzung einer RCT angezeigt ist. Bei sehr seltenen Erkrankungen, bei der die Anzahl der potenziell verfügbaren und zu rekrutierenden Patienten für klinische Studien offensichtlich limitiert ist, sollten alternative Evidenzgrundlagen mit höherer Ergebnisunsicherheit für den Nutzenbewertungsprozess in Betracht gezogen werden können.

Gleiches gilt auch für Orphan-Drug-Solisten. Von 27 Verfahren mit Solistenstatus zum Zeitpunkt der initialen Orphan-Drug-Bewertung (Tab. 5) wiesen 17 (63,0%) eine RCT als Bewertungsbasis auf, während bei 10 (37,0%) Nutzenbewertungsverfahren keine RCT-Evidenz zugrunde lag. Die Anzahl der für die Therapie infrage kommenden Patienten gemäß den Angaben aus den jeweiligen G-BA-Beschlüssen beträgt bei den Orphan-Drug-Solisten ohne RCT-Evidenz durchschnittlich 288 (SD 216,3) und bei den Orphan-Drug-Solisten mit nutzenbewertungsrelevanter RCT 1.669 (SD 2.793,6) (Tab. 5, eigene Berechnung; IQWiG Arbeitspapier GA21-01).

Eine populationsfokussierte Analyse des IQWiG-Datensatzes zeigt, dass RCT auch bei Solisten durchgeführt werden. Bei den Orphan-Drug-Solisten ohne RCT ist die mittlere Zielpopulation fast um ein 6-Faches niedriger im Vergleich zu den Solisten mit RCT als Bewertungsbasis, sodass zunächst zu hinterfragen ist, ob eine RCT aufgrund dieser geringen Anzahl an Patienten mit Blick auf Studien- und Rekrutierungsdauer adäquat umsetzbar gewesen wäre. Falls solche Studien generell durchführbar erscheinen, ist weiterhin zu berücksichtigen, dass für ein kleines Studienkollektiv bei gleicher Effektstärke eine niedrigere statistische Power den Nachweis eines Zusatznutzens erheblich erschwert. Weiterhin stellt bei einem solch kleinen Studienkollektiv auch eine Randomisierung per se keine gleichmäßige Verteilung möglicher Störvariablen (sogenannter Confounder) auf die Interventionsarme sicher.

Größe der Zielpopulation mit und ohne RCT von Orphan-Drug-Verfahren			
	n (%)	Mittelwert (SD)	Median (P_{25}, P_{75})
RCT	30 (73,2)	2.445,1 (2.560,4)	1725 (250, 3.550)
Keine RCT	11 (26,8)	471,3 (640,3)	440 (17, 455)
Rangsummentest, p-Wert	0,0123		

Tab. 4: Größe der Zielpopulation mit und ohne RCT von Orphan-Drug-Verfahren (n = 41). Legende: SD, Standardabweichung; P_{25}, 25%-iges Perzentil; P_{75}, 75%-iges Perzentil.

Mediane Zielpopulation und Studienpopulation mit Korrelationskoeffizient von Orphan-Drug-Verfahren			
n	Zielpopulation Median (P_{25}, P_{75})	Studienpopulation Median (P_{25}, P_{75})	Spearmans rho (p-Wert)
41	1.057 (250; 2.300)	161 (75,5; 416,5)	0.7471 (< 0.001)

Tab. 3: Mediane Zielpopulation und Studienpopulation mit Korrelationskoeffizient von Orphan-Drug-Verfahren (n = 41). Legende: P_{25}, 25%-iges Perzentil; P_{75}, 75%-iges Perzentil.

3) Evidenzgrundlage RCT – Wie gelingt es, auch die zweckmäßige Vergleichstherapie adäquat umzusetzen?

In der Subgruppe der 30 Orphan-Drug-Verfahren (Tabelle 6), denen eine RCT zugrunde lag, wurde in keinem Verfahren ein Zusatznutzen anerkannt, in dem die vom G-BA geforderte zweckmäßige Vergleichstherapie nicht in der RCT abgebildet wurde (10 von 30 Verfahren). Wurde hingegen die zweckmäßige Vergleichstherapie im Rahmen einer RCT bedient (20 von 30 Verfahren, 66,7%), so wurde in 17 der 20 (85,0%) Verfahren ein Zusatznutzen erteilt. Bei 13 von 30 (43,3%) Orphan-Drug-Verfahren wurde somit trotz RCT-generierter Evidenzbasis kein Zusatznutzen ausgesprochen. Folglich kann konstatiert werden, dass neben der RCT-Evidenzbasis das entscheidende methodische Kriterium für die Ableitung eines Zusatznutzens in regulären Nutzenbewertungen ein direkter klinischer Vergleich mit einem aktiven Vergleichsarm ist, der vom G-BA als adäquate Umsetzung der zweckmäßigen Vergleichstherapie akzeptiert wird. Bei über der Hälfte der Orphan-Drug-Verfahren mit RCT als Bewertungsgrundlage (20 von 30 Verfahren, 66,7%) wurden diese Anforderungen zum Zeitpunkt der Re-Evaluierung erfüllt.

Eine genauere Betrachtung jener 20 Orphan-Drug-Verfahren (Tab. 7), in denen eine RCT vorlag, die auch die vom G-BA definierte zVT adäquat abbildete, zeigt, dass es sich hierbei größtenteils um Verfahren handelte, für die ein Placebo-Vergleichsarm eine akzeptable Um-

Größe der Zielpopulation mit und ohne RCT von Orphan-Drug-Verfahren			
	n (%)	Mittelwert (SD)	Median (P_{25}, P_{75})
RCT	17 (63,0)	1669,3 (2793,6)	400 (95, 1680)
Keine RCT	10 (37,0)	288,4 (216,3)	405 (17, 455)
Rangsummentest, p-Wert	0,22783		

Tab. 5: Größe der Zielpopulation bei Solisten mit und ohne RCT in Orphan-Drug-Verfahren (n = 27). Legende: SD, Standardabweichung; P_{25}, 25%-iges Perzentil; P_{75}, 75%-iges Perzentil.

setzung der vom G-BA festgelegten zweckmäßigen Vergleichstherapie darstellte. Dies war operationalisiert über „Best Supportive Care" oder „beobachtendes Abwarten" (13/20 Verfahren, 65%) bzw. wenn ein Placebo als Vergleichsintervention eine gleichermaßen zweckmäßige Therapieoption repräsentierte („Idelalisib oder BSC" und „Pirfenidon oder BSC"; 2/20 Verfahren, 10%). Es handelte sich folglich um Verfahren, in denen die pivotale placebokontrollierte Studie mangels Therapiealternativen im Rahmen der Nutzenbewertung für den G-BA die geeignete Evidenzgrundlage darstellte. Hätte der G-BA in diesen 15 RCT-basierten Verfahren jedoch einen spezifischen Wirkstoff als zweckmäßige Vergleichstherapie festgelegt, wäre grundsätzlich kein Zusatznutzen feststellbar gewesen.

Die 17 Orphan-Drug-Verfahren mit attestierten Zusatznutzen basierend auf einer RCT wurden weiterhin dahingehend näher untersucht, inwiefern sich die Patientenzahlen der Zielpopulation bzw. die Größe der Studienpopulationen zwischen placebokontrollierten und pharmakologisch aktivkontrollierten Studien voneinander unterscheiden. In 10 der 17 (58,8%) Bewertungsverfahren lag die mittlere Größe der Population mit Placebo-Vergleichsarm bei 1.650 Patienten (SD 3.163,97) und einem Median von 271,5 Patienten (P_{25} 44 Patienten, P_{75} 1.680 Patienten), während die mittlere Größe der Population mit aktivem Komparator in der RCT 3.439 Patienten (SD 2.677) mit einem Median von 2.300 Patienten (P_{25} 1.200 Patienten, P_{75} 5.850 Patienten) in 7 (41,2%) Verfahren betrug (Tab. 8).

Ein vergleichbarer Trend zeigt sich auch bei der Untersuchung der Studienpopulation. Bei identischer Vorgehensweise entspricht die mittlere Größe der Studienpopulation mit Placebo-Kontrollarm 197 Patienten (SD 173,33) und einem Median von 140,5 Patienten (P_{25} 52 Patienten, P_{75} 265 Patienten) im Vergleich zu einer mittleren Studienpopulation von 472 Patienten (SD 336,33) und einem Median von 455 Patienten (P_{25} 122 Patienten, P_{75} 792 Patienten) in den Orphan-Drug-Verfahren mit aktiver Vergleichsintervention (Tab. 9). Der Unterschied der Populationsgrößen wurde jeweils anhand des Rangsummentests statistisch ausgewertet (Größe der Patientenpopulation p-Wert 0,0508; Größe der Studienpopulation p-Wert 0,0708) (Tab. 8 und 9, eigene Berechnung; IQWiG Arbeitspapier GA21-01). Auf Basis dieser Auswertung lässt sich erneut unterstreichen, dass neben der tatsächlichen Durchführung einer RCT auch die Inkorporation einer aktiven Vergleichsintervention in das RCT-Design mit der Größe der Zielpopulation und damit der Häufigkeit der Erkrankung assoziiert ist.

Graduiert je nach Prävalenz stellen sich die methodischen Fragen nach dem Studientyp (einarmige oder zweiarmige Studie) und dem Studiendesign hinsichtlich der Vergleichsintervention (Placebo oder aktiver Komparator, sofern eine weitere Therapieoption zur Verfügung steht). Es wird abermals deutlich, dass die Prävalenz auch innerhalb der Orphan Drugs differenziert zu betrachten und gemäß ihrem Potential der tatsächlichen Evidenzgenese zu bewerten ist.

Das IQWiG fordert in dem Arbeitspapier GA21-01 grundsätzlich, dass die postulierten methodischen Anforderungen auch bei der Initialbewertung aller Orphan-Drug-Verfahren anzuwenden seien. Seltene Erkrankungen sind in ihrer Prävalenz und Krankheitsbiologie äußerst heterogen. Eine Generalforderung nach RCT und zwingender Inkorporation einer pharmakotherapeutischen Vergleichssubstanz für Orphan Drugs erscheint nach Revision des IQWiG-Datensatzes methodisch diskutabel. Darüber hinaus sei an diesem Punkt daran erinnert, dass die Vergabe eines Orphan-Drug-Status durch die EMA voraussetzt, dass für das relevante Orphan Drug im Anwendungsgebiet keine alternative Therapieoption vorliegt oder dass der relevante Wirkstoff einen „erheblichen Nutzen" im Vergleich zum Therapiestandard darstellt. Ungeachtet dessen sieht das IQWiG es grundsätzlich als notwendig an, einen Placebo-Kontrollarm oder je nach therapeutischer Situation die nach Bewertung der EMA unterlegene Therapie in das Studiendesign zu implementieren, um so den definierten Therapiestandard für die versorgungsrelevante Fragestellung abzubilden.

Insgesamt bleibt in der Betrachtung des IQWiG unberücksichtigt,

Zweckmäßige Vergleichstherapie bei RCT-generierter Evidenz von Orphan Drug-Verfahren	
Zweckmäßige Vergleichstherapie	**n (%)**
Best Supportive Care (BSC)	12 (60)
Beobachtendes Abwarten	1 (5)
Bortezomib in Kombination mit pegyliertem, liposomalen Doxorubicin oder Bortezomib in Kombination mit Dexamethason oder Lenalidomid in Kombination mit Dexamethason oder Elotuzumab in Kombination mit Lenalidomid und Dexamethason	2 (10)
Dexamethason, hochdosiert (die patientenindividuell geeignete Therapieoption)	1 (5)
Idelalisib oder BSC	1 (5)
Pirfenidon oder BSC	1 (5)
Temsirolimus (die patientenindividuell geeignete Therapieoption)	1 (5)
Therapie nach Maßgabe des Arztes	1 (5)

Tab. 7: Zweckmäßige Vergleichstherapie bei RCT-generierter Evidenz von Orphan-Drug-Verfahren (n = 20).

Zusatznutzen bei RCT-generierter Evidenz von Orphan-Drug-Verfahren		
	Vergleichsarm entspricht zweckmäßiger Vergleichstherapie: ja [n (%)]	Vergleichsarm entspricht zweckmäßiger Vergleichstherapie: nein [n (%)]
Zusatznutzen	17 (85,0)	0 (0)
Kein Zusatznutzen	3 (15,0)	10 (100)
Risikodifferenz* (0,95-KI), p-Wert nach Fisher	85,0% (69,4%; 100%), < 0,0001	

Tab. 6: Zusatznutzen bei RCT-generierter Evidenz von Orphan-Drug-Verfahren (n = 30). Legende: * Risk Ratio nicht berechenbar; KI, Konfidenzintervall.

dass sich die Fallzahl des benötigten Studienkollektivs zur Durchführung einer RCT und die Studiendauer dadurch bedeutend erhöht. Folglich stehen Orphan Drugs den Patienten mit seltenen Leiden deutlich später in der Routineversorgung zur Verfügung. Exemplarisch verweisen wir hier auf die Fallzahlschätzung und Studiendauer der anwendungsbegleitenden Datenerhebung von Onasemnogen-Abeparvovec (Zolgensma®), indiziert für die Behandlung der spinalen Muskelatrophie (SMA), einer hereditären Erkrankung mit einer Prävalenz von 4 zu 100.000 (AveXis EU Ltd. 2020). Hier wurde eine Anzahl von 500 Patienten berechnet, um einen Zusatznutzen quantifizieren zu können. Daraus resultierend ergibt sich eine Studiendauer von bis zu 10 Jahren (Beschluss des Gemeinsamen Bundesausschusses über die Änderung der Arzneimittel-Richtlinie (AM-RL): Onasemnogen-Abeparvovec (spinale Muskelatrophie; Forderung einer anwendungsbegleitenden Datenerhebung und von Auswertungen) (Gemeinsame Bundesausschuss (G-BA) 2021). Inwieweit dieses Instrument als adäquat zur Schließung von versorgungsrelevanten Evidenzlücken heranzuziehen ist, bleibt fragwürdig, zumal auch hier nicht-randomisierte Evidenz erhoben wird. Auch das IQWiG erkennt dieses Problem der limitierten Evidenzgrundlage anwendungsbegleitender Datenerhebungen und regt die Durchführung pragmatischer RCT an. Dennoch eignen sich solche Konzepte nur, sofern bereits zur Verfügung stehende Therapieoptionen existieren, was in dem zugrunde liegenden Datensatz des IQWiG in 2/3 (27 von 41, 65,9%) der Verfahren zum Zeitpunkt des Marktzugangs jener Orphan Drugs nicht der Fall war.

Diskussion

Die seltenen Leiden sind trotz ihres unter anderem definierten Charakteristikums von einer Prävalenz von weniger als 5 Betroffenen je 10.000 Menschen in der EU in ihrer Häufigkeit äußerst heterogen. Viele Fragestellungen im methodischen Studiendesign sind assoziiert mit der Häufigkeit der seltenen Erkrankung. Die vorliegende Analyse zeigt, dass die Durchführung einer Studie im RCT- oder Nicht-RCT-Design mit der Krankheitsprävalenz assoziiert ist. Auch innerhalb der RCT-generierten Evidenz der Orphan-Drug-Verfahren hinsichtlich der Durchführung als placebokontrollierte oder aktivkontrollierte Studie offenbart die Analyse die Abhängigkeit der Krankheitshäufigkeit. Die zitierten IQWiG-Verweise, die die prinzipielle Durchführbarkeit von RCT bei seltenen und sehr seltenen Erkrankungen belegen sollen, befassen sich nicht dezidiert mit der Fragestellung des Komparators im Kontrollarm (Naumann-Winter, Kaiser et al. 2022). Vielmehr wird auch hier dargelegt, dass die mittlere Studienpopulation bei RCT mit Vergleich gegen die Standardtherapie doppelt so groß wie die mittlere Studienpopulation bei RCT mit Placebo-Vergleich war (Hee, Willis et al. 2017). Vor dem Hintergrund, dass ca. 85% der seltenen Erkrankungen eine Prävalenz von weniger als 1 zu 1.000.000 aufweisen, was umgerechnet für Deutschland maximal eine Anzahl von 80 Erkrankten in der GKV bedeutet, für die eine jeweilige Therapie infrage käme, ist die vom IQWiG formulierte pauschale Forderung nach RCT-generierter Evidenz zu hinterfragen (Nguengang Wakap, Lambert et al. 2020). Für diese geringe Anzahl an Patienten sollte die bestverfügbare Evidenz unter Anerkennung höherer Ergebnisunsicherheit akzeptiert werden. Beispielsweise könnten die Evidenzanforderungen nach Prävalenzkategorien differenziert werden, um so eine Bewertung unter Berücksichtigung der Epidemiologie zu ermöglichen. Verschiedene Prävalenzgraduierungen könnten definiert werden, die die Akzeptanz von nicht randomisierten Studien, von RCT mit Placebo-Kontrollarm und RCT mit Vergleich zum verfügbaren Therapiestandard, sofern dieser zum Zeitpunkt der Studiendurchführung verfügbar ist, erlauben. Auch die Frage des Solistenstatus bei Markteintritt sollte hinreichend gewürdigt werden. In zwei Drittel der Orphan-Drug-Verfahren stand zur

Größe der Zielpopulation bei RCT mit Zusatznutzen und korrekter Umsetzung der zweckmäßigen Vergleichstherapie

	n (%)	Mittelwert (SD)	Median (P_{25}, P_{75})
Placebo-Kontrollarm	10 (58,8)	1649,95 (3163,97)	271,5 (44, 1680)
Kein Placebo-Kontrollarm (aktiver Komparator)	7 (41,2)	3439,29 (2677,76)	2300 (1200, 5850)
Rangsummentest, p-Wert	0,0508		

Tab. 8: Größe der Zielpopulation bei RCT mit Zusatznutzen und korrekter Umsetzung der zweckmäßigen Vergleichstherapie mit und ohne Placebo-Kontrollarm in Orphan-Drug-Verfahren (n = 17). Legende: SD, Standardabweichung; P_{25}, 25%-iges Perzentil; P_{75}, 75%-iges Perzentil.

Größe der Studienpopulation bei RCT mit Zusatznutzen und korrekter Umsetzung der zweckmäßigen Vergleichstherapie

	n (%)	Mittelwert (SD)	Median (P_{25}, P_{75})
Placebo-Kontrollarm	10 (58,8)	196,8 (173,33)	140,5 (52, 265)
Kein Placebo-Kontrollarm (aktiver Komparator)	7 (41,2)	471, 71 (336,33)	455 (122, 792)
Rangsummentest, p-Wert	0,0708		

Tab. 9: Größe der Studienpopulation bei RCT mit Zusatznutzen und korrekter Umsetzung der zweckmäßigen Vergleichstherapie mit und ohne Placebo-Kontrollarm in Orphan-Drug-Verfahren (n = 17). Legende: SD, Standardabweichung; P_{25}, 25%-iges Perzentil; P_{75}, 75%-iges Perzentil.

Literatur

Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e. V. (AWMF) (2022). „Stellungnahme zum GKV-Finanzstabilisierungsgesetz."

AveXis EU Ltd. (2020). „Dossier zur Nutzenbewertung gemäß § 35a SGB V Onasemnogen-Abeparvovec (ZOLGENSMA®), spinale Muskelatrophie, Modul 3A."

Gemeinsame Bundesausschuss (G-BA) (2021). „Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V Onasemnogen-Abeparvovec (spinale Muskelatrophie); Forderung einer anwendungsbegleitenden Datenerhebung und von Auswertungen." BAnz AT 19.04.2021 B3.

Hee, S. W., A. Willis, C. Tudur Smith, S. Day, F. Miller, J. Madan, M. Posch, S. Zohar and N. Stallard (2017). „Does the low prevalence affect the sample size of interventional clinical trials of rare diseases? An analysis of data from the aggregate analysis of clinicaltrials.gov." Orphanet J Rare Dis 12(1): 44.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2021). „Evidenz zu Orphan Drugs." Arbeitspapier GA21-01.

Naumann-Winter, F., T. Kaiser and A. Behring (2022). „Evidence-based health care with pharmaceuticals for rare diseases: the role of digitalisation." Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 65(11): 1170-1177.

Nguengang Wakap, S., D. M. Lambert, A. Olry, C. Rodwell, C. Gueydan, V. Lanneau, D. Murphy, Y. Le Cam and A. Rath (2020). „Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database." Eur J Hum Genet 28(2): 165-173.

Markteinführung keine weitere zugelassene Therapie in der betreffenden Indikation zur Verfügung.

Die Erkenntnis des IQWiG, dass es in mehr als der Hälfte der Fälle bei Markteinführung unklar sei, wie ein neues Orphan Drug im Vergleich der zur Verfügung stehenden Therapieoptionen abschneidet, kann demnach aus dem zugrunde liegenden Datensatz nicht abgeleitet werden. Die Diskussion für neue Bewertungsansätze sollte neben Methodikern auch mit Klinikern und weiteren Fachexperten stattfinden und nicht ideologisch getrieben sein. Von medizinischen Fachgesellschaften wird bereits eine fokussierte Förderung für die Entwicklung von Arzneimitteln für seltene Erkrankungen auf Indikationen mit niedriger Inzidenz und Prävalenz diskutiert (Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e. V. (AWMF) 2022). Die Autoren plädieren deshalb für einen Dialog mit allen relevanten Akteuren, der eine differenzierte Betrachtungsweise auf Orphan-Drug-Verfahren erlaubt und kompromissorientierte Lösungsansätze fördert, die auch in der praktischen Umsetzung für die frühe Nutzenbewertung Berücksichtigung finden können. <<

Autorenerklärung

Die Autor:innen sind angestellt bei Takeda Pharma Vertrieb GmbH & Co. KG. Die Studie wurde durch die Takeda Pharma Vertrieb GmbH & Co. KG konzipiert und durchgeführt.

Zitationshinweis

Müller, G., Fanter, L., Knierim, J.: „Analyse von Orphan-Drug-Verfahren in der frühen Nutzenbewertung: RCTs und bestverfügbare vergleichende Evidenz“, in: „Monitor Versorgungsforschung“ (02/23), S. 39–44. http://doi.org/10.24945/MVF.02.23.1866-0533.2483

More neutrality in the analysis of orphan drug procedures in the German early benefit assessments

For drugs intended to use in rare diseases (orphan drugs), the requirements of the early benefit assessment according to § 35a SGB V (Social Security Code, book 5) pose a particular challenge. The Institute for Quality and Efficiency in Health Care (IQWiG) concludes in his working paper GA21-01 on orphan drugs that have undergone a regular benefit assessment as part of a re-evaluation that the additional benefit of the initial benefit assessment for market access was fictitious. Therefore, in IQWiG's view, only randomized controlled trials (RCTs) with an appropriate active control arm should be applied for the assessment of the added benefit of orphan drugs. The authors conducted an own analysis based on the database used by IQWiG on 41 orphan drugs. That analysis showed an average of 1,146 days between the initial and regular benefit assessment and that the respective therapeutic settings differ: While almost 66% of the orphan drugs (27/41) had a soloist status at the time of the initial evaluation, approximately 30% (8/27) had lost this status at the time of the re-evaluation. This circumstance alone already reduced the probability of an additional benefit in regular assessments significantly. In 75% of these procedures (6/8), no additional benefit was given. IQWiG also disregarded the fact that the frequency of a disease, especially in the context of orphan diseases, is essential for the feasibility of an RCT. In those benefit assessment procedures where the company had submitted an RCT, the average size of the target population was 2,445 patients, while in those procedures without presented RCT data, it was 471 patients. Of the orphan drug procedures with presented RCT data (30/41), in 10 procedures the comparator arm did not correspond to the appropriate comparator therapy defined by the Federal Joint Committee (G-BA), which was the decisive factor for the outcome of no additional benefit being identified. However, if the comparator matched the G-BA requirement (20/30), an additional benefit was granted in 85% (17/20). In most of these procedures (15/20), a placebo comparator arm was deemed sufficient by G-BA as an adequate implementation of the appropriate comparator therapy. Of note, in the 17 orphan drug assessments with presented RCT data and achieved an added benefit rating in the regular assessment process, the average target population size in the respective indication was 1,650 patients (10/17) when the clinical comparison was conducted against placebo vs. 3,439 patients in those indications where RCT data were generated against an active comparator (7/17).

The analysis showed that RCTs are being conducted also in rare diseases when the disease prevalence is relatively high. However, approximately 8% of rare diseases have a prevalence of less than 1 in 1,000,000, meaning less than 80 SHI patients in Germany suffering from those diseases. IQWiG's claim for comparative RCT-generated evidence for all orphan drugs therefore needs to be put into question (Nguengang Wakap, Lambert et al. 2020).

For those small patient numbers in specific rare diseases, the best available comparative evidence should be evaluated carefully for data being eligible to answer the scientific question of the benefit assessment. For the future, joint efforts aiming for developing adequate and acceptable methodologies while acknowledging a higher uncertainty of evidence are deemed necessary by the authors.

Keywords

AMNOG, orphan drugs, medicinal products, IQWiG, benefit assessment

Dr. rer. nat. Gerrit Müller ORCID: 0000-0002-5585-0067



ist Health Technology Assessment (HTA) Manager bei Takeda. Zuvor war er als Fachreferent für Arzneimittel im Referat AMNOG, Abteilung Arznei- und Heilmittel beim GKV-Spitzenverband tätig. Er studierte Pharmazie an der FU Berlin und promovierte dort im Anschluss im Fachbereich Pharmakologie.
Kontakt: gerrit.mueller@takeda.com

Dr. rer. nat. Lena Fanter ORCID: 0000-0002-9631-739X



ist Fachapothekerin für Arzneimittelinformation und leitet das Evidence Generation & Value Demonstration Team bei Takeda. Sie studierte Pharmazie an der WWU Münster und promovierte dort im Anschluss am Institut für Pharmazeutische und Medizinische Chemie.
Kontakt: lena.fanter@takeda.com

Dipl.-Ges.oec. Julia Knierim



ist Gesundheitsökonomin und leitet den Bereich Patient Value & Access bei Takeda. Ihr Studium mit Schwerpunkt auf evidenzbasierte Medizin schloss sie am Institut für Gesundheitsökonomie und Klinische Epidemiologie der Universität Köln ab.
Kontakt: julia.knierim@takeda.com