

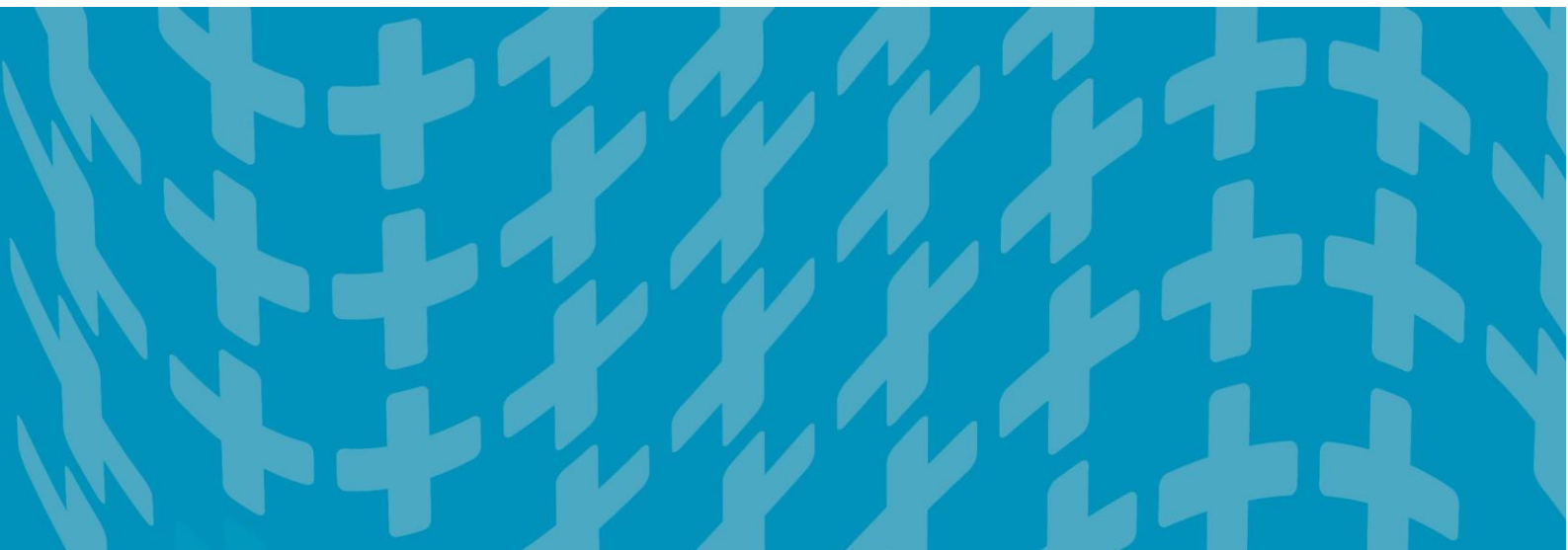
Anwendungsbegleitende Daten in der Nutzenbewertung: Empfehlungen zur Evidenzgenerierung und -auswertung

Hans-Holger Bleß

05. November 2019

Im Auftrag vom

Verband forschender Arzneimittelhersteller (vfa) e.V.



Inhaltsverzeichnis

1	Executive Summary	5
2	Hintergrund	8
3	Regulatorische Rahmenbedingungen	8
3.1	Arzneimittel zur Behandlung seltener Erkrankungen („Orphan Drugs“)	10
3.2	Bedingte Zulassung („Conditional Approval“).....	11
3.3	Zulassung unter außergewöhnlichen Umständen („Approval under exceptional Circumstances“)	11
4	Praktischer Umgang mit nicht-randomisierten Daten	12
4.1	Positionierung des G-BA	12
4.1.1	Stellungnahme des G-BA zum GSAV	12
4.1.2	Verwendung nicht-randomisierter Daten in Nutzenbewertungsverfahren 13	
4.2	Positionierung des IQWiG.....	14
4.2.1	Stellungnahme des IQWiG zum GSAV.....	15
4.2.2	Regelungen des Methodenpapiers 5.0.....	15
4.3	Positionierung der Europäischen Arzneimittel-Agentur (EMA).....	17
4.3.1	Post-authorisation efficacy studies (PAES).....	17
4.4	Positionierung anderer Institutionen.....	19
4.4.1	GRADE	19
4.4.2	EUnetHTA	20
4.4.3	Cochrane Collaboration.....	22
4.4.4	Deutsches Netzwerk für Versorgungsforschung (DNVF)	23
5	Mögliche Evidenzarten.....	24
5.1	Fall/Kontroll-Studien	25
5.2	Prospektive vergleichende Kohortenstudien.....	25
5.3	Einarmige Studien	26
5.4	Historische Vergleiche	27

5.5	Prä-Post-Vergleiche.....	27
6	Mögliche Auswertungsmethoden.....	28
6.1	Umgang mit der Irrtumswahrscheinlichkeit	28
6.2	Forderung nach „dramatischen Effekten“	29
6.3	Adjustierungsverfahren.	30
6.3.1	Propensity Score Matching (PSM).....	30
6.3.2	Matching-adjustierter indirekter Vergleich (MAIC)	31
6.3.3	Simulierter Therapievergleich (Simulated Treatment Comparison, STC) 31	
6.3.4	Bayes-Benchmarking Analysis (BBA)	31
7	Stufenmodell zur Evidenzgenerierung	31
7.1	Bewertung und Einbeziehung der erwartbaren Evidenz	32
7.2	Identifizierung des höchstmöglichen Evidenzlevels	33
8	Handlungsempfehlung.....	35
9	Literaturverzeichnis	40

Tabellenverzeichnis

Tabelle 1: Evidenz bei besonderen Zulassungsformen	10
Tabelle 2: Übersicht der Verfahren von Orphan Drugs mit Registerauflagen durch den G-BA	14
Tabelle 3: Zusammenfassung des GRADE-Ansatzes zur Evidenzbewertung	20

Abbildungsverzeichnis

Abbildung 1: „Evidenzpyramide“ aus Sicht des IQWiGs.....	17
Abbildung 2: REQueST	22
Abbildung 3: Prinzip einer retrospektiven Fall-Kontroll-Studie	25
Abbildung 4: Prinzip einer prospektiven Kohortenstudie.....	26
Abbildung 5: Studiendesign auferlegter Studien für CMA (N=77)	32
Abbildung 6: Bewertung erwartbarer und generierbarer Evidenz.....	34

1 EXECUTIVE SUMMARY

Mit dem GSAV wurde der G-BA für Orphan Drugs, Arzneimittel mit bedingter Zulassung sowie unter außergewöhnlichen Umständen erteilter Zulassung ermächtigt, den pharmazeutischen Unternehmer aufzufordern, zum Zweck der Nutzenbewertung anwendungsbegleitende Daten zu erheben und auszuwerten.

Nach dem Willen des Gesetzgebers sollen für diese Arzneimittel Studien niedrigerer Evidenzstufen als RCT (d.h. nicht-randomisierte Daten) für die Ableitung eines (quantifizierbaren) Zusatznutzens herangezogen werden.

Mit den Zulassungsentscheidungen für diese Gruppe von Arzneimitteln bestätigt die EMA ein positives Nutzen-Risiko-Verhältnis bei Erfüllung eines nicht gedeckten medizinischen Bedarfs in einer Situation, in der die für eine reguläre Zulassung erforderlichen Daten erst zu einem späteren Zeitpunkt erbracht werden können bzw. deren Erbringung unmöglich ist. Aufbauend auf diesen Zulassungsentscheidungen der EMA sieht nun der deutsche Gesetzgeber mit dem GSAV eine Berücksichtigung von nicht-randomisierten Daten zur Ableitung eines (quantifizierbaren) Zusatznutzens vor. Mit der Vorgabe, bei der Nutzenbewertung auch Evidenz mit geringerer Sicherheit in der Aussage zu akzeptieren, ermöglicht der Gesetzgeber die in der AM-NutzenV geforderte Kongruenz zwischen Nutzenbewertung und Feststellungen der Zulassungsbehörde.

Ziel der Datenerhebung ist die Verbesserung der Evidenzlage. Damit ist eine Datenerhebung nur dann vertretbar, wenn erwartet werden kann, dass diese Daten grundsätzlich für die Quantifizierung eines Zusatznutzens verwertbar sind.

Hieraus ergibt sich die Frage, welche (i) Arten der Evidenzgenerierung, (ii) welche Auswertungsmethoden bzw. (iii) welche Methoden der Evidenzbewertung geeignet sind, um diese Zielstellung zu erfüllen und (iv) welche begleitenden beratenden Maßnahmen hierfür notwendig sind.

Auf Grundlage einer Übersicht über den aktuellen Umgang mit nicht-randomisierten Daten seitens wichtiger Organisationen (EMA, G-BA, IQWiG, Cochrane Collaboration, GRADE working group, EUnetHTA, DNVF) sowie einer Übersicht zu relevanten methodischen Auswertungsverfahren werden Empfehlungen zu diesen Fragen formuliert.

Arten der Evidenzgenerierung

Die Identifizierung der geeignetsten Evidenzform sollte unter Berücksichtigung der vorhandenen Evidenz, der erwartbaren Evidenz sowie weiterer kontextueller und methodischer Voraussetzungen erfolgen:

- A) Liegen randomisierte vergleichende Studien (RCT) bereits vor oder sind erwartbar, werden anwendungsbegleitende Datenerhebungen als nicht zielführend erachtet, da die Nutzenbewertung auf Grundlage höherer Evidenzstufe durchgeführt werden kann.
- B) Sind keine RCT vorhanden oder erwartbar, werden drei Evidenzformen mit abgestufter Evidenzstärke für eine mögliche Datenerhebung als geeignet erachtet. Welche

Evidenzform die geeignetste ist, hängt dabei von den begleitenden kontextuellen bzw. methodischen Voraussetzungen ab:

- a. Die prospektiv vergleichende Kohortenstudie (z. B. Registerstudie) bietet die höchste Evidenzstärke und ist am geeignetsten, solange Verzerrungsaspekte über die Erfassung wesentlicher Störgrößen vertretbar kontrollierbar sind und Therapieentscheidungen in der Versorgung nicht zu einem unkontrollierbaren Selektionsbias führen. Dies kann z. B. der Fall sein, wenn ein Medikament in einen Markt eintritt, in dem keine sinnvollen Therapiealternativen zur Verfügung stehen und deshalb fast ausschließlich angewendet wird. Zusätzlich sollte sichergestellt sein, dass auch für die Kontrollgruppe hinreichend vollständige oder repräsentative Daten erhoben werden können.
- b. Der historische Vergleich (z. B. auf Basis von Registerdaten) bietet eine geminderte Evidenzstärke und ist am geeignetsten, wenn die Voraussetzungen für (a.) nicht erfüllt sind und historische Daten mit anwendungsbegleitenden Daten sinnvoll verglichen werden können. Voraussetzung hierfür ist die Verfügbarkeit von historischen Daten auf patientenindividueller Ebene für relevante und übereinstimmend operationalisierte Endpunkte auch in Hinblick auf wesentliche Störgrößen (z. B. prognostische, krankheitsspezifische, demografische Faktoren). Die historischen Daten könnten dann bspw. mit hinreichend vollständigen oder repräsentativen Registerdaten für die neue Therapieoption verglichen werden.
- c. Der Prä-Post-Vergleich (z. B. Registerstudie) bietet eine niedrigere Evidenzstärke und ist am geeignetsten, wenn die Voraussetzungen für (b.) nicht erfüllt sind und keine Spontanverbesserung für eine (fiktive) unbehandelte Kontrollgruppe erwartbar ist.

Sind keine der o.g. Voraussetzungen erfüllbar, sollte keine Datenerhebung beauftragt werden, da dann grundsätzlich nicht zu erwarten ist, dass die erhobenen Daten für die Quantifizierung eines Zusatznutzens verwendbar sind.

Bei der Auswahl der geeignetsten Evidenzform sind bereits bestehende bzw. absehbare Auflagen der EMA prioritär zu berücksichtigen. Dabei soll die geeignetste Evidenzform unter den Auflagen identifiziert werden und bei Notwendigkeit um weitere Aspekte der Datenerhebung ergänzt werden.

Auswertungsmethoden

Werden prospektiv vergleichende Kohortenstudien (a.) oder historische Vergleiche (b.) als geeignete Evidenzform ausgewählt, sollten wesentliche Störgrößen erfasst und durch Anwendung geeigneter Adjustierungsverfahren ausreichend kontrolliert werden. Mögliche Unterschiede z. B. zwischen prognostischen, krankheitsspezifischen oder demografischen Faktoren können auf diese Weise bei der Effektschätzung hinreichend berücksichtigt werden. Hierfür stehen international anerkannte Methoden zur Verfügung, wie z. B. Propensity Score Matching (PSM) oder der Matching-adjustierte indirekte Vergleich (MAIC). Die Auswahl der geeigneten Adjustierungsverfahren und der Störfaktoren sollte dabei prospektiv erfolgen.

Trotz einer Anwendung von geeigneten Adjustierungsverfahren ist in der Regel mit einer geringeren Ergebnissicherheit gegenüber randomisierten Studien zu rechnen, die jedoch nach dem Willen des Gesetzgebers zu akzeptieren ist. Die in regulären Nutzenbewertungsverfahren übliche Erforderlichkeit eines „dramatischen Effekts“ für die Ableitung eines Zusatznutzens auf Grundlage von nicht-randomisierten Daten würde der gesetzlichen Intention nicht gerecht werden. Zwar kann die Größe des Effektes die Sicherheit der zu treffenden Aussage zusätzlich erhöhen. Grundsätzlich ermöglicht aber eine adäquate Adjustierung von Störgrößen mit geeigneten statistischen Methoden in dieser Situation valide Aussagen zum Zusatznutzen ohne weitere Voraussetzungen an die zu erwartende Effektgröße.

Ein weiterer wichtiger Aspekt der Auswertung betrifft die Festlegung einer akzeptablen Irrtumswahrscheinlichkeit. Diese sollte die Besonderheiten der Therapiesituationen, insb. die Anzahl von Patienten in seltenen Erkrankungen hinreichend berücksichtigen. In Abhängigkeit der zu erwartenden Patientenzahlen kann es erforderlich sein, eine erhöhte Irrtumswahrscheinlichkeit (über 5 %) zu akzeptieren, um auf diese Weise ein Ergebnis innerhalb einer sinnvollen Studiendauer zu erzielen. Werden diese Besonderheiten nicht hinreichend berücksichtigt, besteht die Gefahr, dass die beauftragten Daten nicht innerhalb der vorgesehenen Zeit für die Quantifizierung eines Zusatznutzens verwertbar sind.

Methoden der Evidenzbewertung

Führende Institutionen der evidenzbasierten Medizin stellen international akzeptierte Instrumente für die Bewertung der Aussagekraft von nicht-randomisierte Daten zur Verfügung (GRADE, Cochrane ROBINS-I). Diese erlauben eine standardisierte Bewertung des Verzerrungspotentials und weiterer Kriterien, welche die Evidenzstärke positiv oder negativ beeinflussen können. Diese Methoden der Evidenzbewertung sehen die Erreichbarkeit niedriger, moderater und sogar hoher Evidenzlevels für nicht-randomisierte Daten vor. Diese Methoden können somit einen geeigneten Rahmen für die Anerkennung von anwendungsbegleitenden Daten für die Nutzenbewertung bieten. Mit dem REQueST-Instrument (Registry Evaluation and Quality Standards Tool [EUnetHTA]) steht ein weiteres systematisches Bewertungstool für Register zur Verfügung, welches einheitliche Qualitätsanforderungen an Register stellt. Es ermöglicht sowohl die Bewertung bestehender Register als auch die Einhaltung von relevanten Qualitätskriterien beim Aufbau eines Registers.

Vorgaben der G-BA-Beratungen

Der G-BA wird durch das GSAV verpflichtet, Anforderungen an die Datenerhebung und Auswertung, insb. zur Methodik zu bestimmen. Mit diesen Vorgaben muss sichergestellt werden, dass eine Datenerhebung vermieden wird, die erwartbar nicht zur Quantifizierbarkeit eines Zusatznutzens beitragen kann. Eine detaillierte Beratung des G-BA bezüglich aller wesentlichen Aspekte der Datengenerierung und -auswertung, inkl. der Fragestellung (PICO-Schema), Evidenzform, Störfaktoren und Adjustierungsmethoden, Stichprobenumfang, Dauer und akzeptabler statistischer Sicherheit ist deshalb erforderlich. Diese Beratungen sollten frühestmöglich unter Nutzung des Dialogs mit der EMA erfolgen und verbindlichen Charakter haben. Für die medizinische Einschätzung der Umsetzbarkeit

der Anforderungen sollten auch die medizinischen Fachgesellschaften eingebunden werden. Vorgaben sollten zudem im Einvernehmen mit den Zulassungsbehörden erfolgen.

2 HINTERGRUND

Mit dem GSAV wird der G-BA ermächtigt anwendungsbegleitende Datenerhebungen und Auswertungen zum Zweck der Nutzenbewertung zu fordern für

- Arzneimittel für seltene Leiden (Orphan Drugs, OD),
- Arzneimittel mit bedingter Zulassung (Conditional Approval),
- Arzneimittel, die unter außergewöhnlichen Umständen (Approval under exceptional Circumstances) zugelassen wurden.

Zugleich wird der G-BA ermächtigt die Verordnungsfähigkeit von in der Nutzenbewertung befindlichen Arzneimitteln an die geforderte Datenerhebung zu koppeln wie auch indikationsbezogene Datenerhebungen zu fordern.

Auf Basis der anwendungsbegleitend erhobenen Daten erfolgt eine erneute Nutzenbewertung. Vorgaben zu Art, Umfang, Dauer und zur Methodik sowie zu patientenrelevanten Endpunkten und deren Erfassung müssen im Vorfeld vom G-BA festgelegt werden.

Die Intention des Gesetzgebers ist, dass diese Daten vom G-BA in der Nutzenbewertung herangezogen werden, um eine Quantifizierung des Zusatznutzens zu ermöglichen.

Diese Neuregelung beinhaltet ein Spannungspotential zwischen den Evidenzanforderungen, die sich für die Nutzenbewertung durch IQWiG und G-BA in der Spruchpraxis etabliert haben und der nun vom Gesetzgeber ausdrücklich geforderten Berücksichtigung nicht-randomisierte Studiendaten.

Die konkrete Ausgestaltung der gesetzlichen Vorgaben erfolgt durch den G-BA; das IQWiG wurde bereits vom G-BA beauftragt, eine „Wissenschaftliche Ausarbeitung von Konzepten zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V“ vorzunehmen.

Mit folgendem Bericht soll ein Beitrag zur Methodendiskussion um die Verwendung von anwendungsbegleitenden Daten in der Nutzenbewertung geleistet werden. Auf Grundlage einer Übersicht über den aktuellen Umgang mit nicht-randomisierten Daten seitens wichtiger Organisationen (EMA, G-BA, IQWiG, Cochrane Collaboration, GRADE working group, EUnetHTA, DNVF) sowie einer Übersicht zu relevanten methodischen Auswertungsverfahren sollen Empfehlungen bezüglich geeigneter Arten der Evidenzgenerierung, geeigneter Auswertungsmethoden bzw. Methoden der Evidenzbewertung und erforderlichen G-BA-Vorgaben formuliert werden, die eine Berücksichtigung der Daten für einen (quantifizierbaren) Zusatznutzen ermöglichen.

3 REGULATORISCHE RAHMENBEDINGUNGEN

Die Europäische Arzneimittelagentur (European Medicines Agency, EMA) sieht grundsätzlich („soweit möglich“) die Verwendung von RCT im Zulassungsverfahren vor. Ein nicht-

randomisiertes Design bei der Prüfung von Arzneimitteln ist grundsätzlich möglich, unterliegt allerdings einer besonderen Begründungspflicht [1].

Für bestimmte Arzneimittel in definierten Versorgungssituationen hat die EMA spezielle Zulassungswege etabliert, die (teilweise) einen Zugang zur Versorgung basierend auch auf geringerer Evidenz ermöglichen.

Tabelle 1: Evidenz bei besonderen Zulassungsformen

Zulassungsform	Voraussetzungen	Evidenzanforderungen
Arzneimittel zur Behandlung seltener Erkrankungen („Orphan Drugs“)	Fehlende Therapiemöglichkeit oder erheblicher Nutzen bei bestehender Therapiealternative	Evidenzanforderungen entsprechen regulärer Zulassung
Arzneimittel mit bedingter Zulassung („Conditional Approval“)	Positives Nutzen-Risiko-Verhältnis bei Erfüllung eines nicht gedeckten medizinischen Bedarfs bei schwerwiegenden Erkrankungen	Beibringung ergänzender Daten wahrscheinlich
Arzneimittel mit Zulassung unter außergewöhnlichen Umständen („Approval under exceptional Circumstances“)	Positives Nutzen-Risiko-Verhältnis bei nachgewiesener Unmöglichkeit umfassende Daten vorzulegen	Beibringung ergänzender Daten nur teilweise realistisch

Quelle 1: inav GmbH (2019).

3.1 Arzneimittel zur Behandlung seltener Erkrankungen („Orphan Drugs“)

Ein Arzneimittel kann von der EMA den Orphan Drug Status zuerkannt bekommen, wenn folgende Kriterien erfüllt sind:

- Es ist zur Behandlung, Vorbeugung oder Diagnose einer lebensbedrohlichen oder chronisch schwächenden Krankheit bestimmt.
- Die Prävalenz der Erkrankung in der Europäischen Union (EU) ist niedriger als 5 von 10.000, oder es ist unwahrscheinlich, dass die Vermarktung des Arzneimittels ausreichende Erträge bringt, um die für seine Entwicklung erforderlichen Investitionen zu rechtfertigen.
- Es existiert keine zufriedenstellende Methode zur Diagnose, Vorbeugung oder Behandlung des betreffenden Leidens oder falls eine solche Methode existiert, hat das neue Arzneimittel ein erhebliches Nutzen. [2]

Der Begriff „erheblicher Nutzen“ ist dabei definiert als ein klinisch relevanter Vorteil oder ein bedeutender Beitrag zur Behandlung von Patienten [3].

Die Anerkennung des Orphan Drug-Status erfolgt durch die EMA auf Empfehlung des bei ihr angesiedelten Committee for Orphan Medicinal Products (COMP). Die Empfehlung wird anhand der genannten Regelungen erteilt und beinhaltet somit, dass das Arzneimittel voraussichtlich einen erheblichen therapeutischen Nutzen im Vergleich zu bisher vorhandenen Behandlungsoptionen aufweist. Im Rahmen des Zulassungsverfahrens wird die Einhaltung dieser Regelungen durch das COMP erneut überprüft, so dass nach erteilter Zulassung der Zusatznutzen belegt ist.

In der Gesetzesbegründung des Arzneimittelmarktneuordnungsgesetzes (AMNOG) zur Ausnahmeregelung für Orphan Drugs heißt es daher: „Arzneimittel, die [...] als Arzneimittel für seltene Leiden ausgewiesen sind, erbringen durch diese Ausweisung bereits einen Beleg für ihren Zusatznutzen.“ [4].

3.2 Bedingte Zulassung („Conditional Approval“)

Insbesondere für Patienten mit lebensbedrohlichen Erkrankungen soll ein schneller Zugang zu neuen Therapien ermöglicht werden. So kann für betreffende Arzneimittel eine sogenannte „bedingte Zulassung“ erteilt werden, mit der ein Marktzugang vor Abschluss der vollständigen klinischen Prüfung möglich ist.

Bedingte Zulassungen gelten jeweils für ein Jahr und sind an Auflagen und eine engmaschige Kontrolle geknüpft, bis eine reguläre Zulassung erteilt wird. Im Rahmen dieser Auflagen sind ein positives Nutzen-Risiko-Verhältnis nachzuweisen sowie offene Fragen zu Qualität, Unbedenklichkeit und Wirksamkeit des Arzneimittels zu beantworten.

Für eine bedingte Zulassung kommen Arzneimittel in Frage, die für

- schwerwiegende oder lebensbedrohenden Krankheiten bestimmt sind,
- den Einsatz in Notfallsituationen als Reaktion auf Bedrohungen der öffentlichen Gesundheit vorgesehen sind,
- die Behandlung von seltenen Leiden eingesetzt werden (vgl. 3.1) [5].

Nachzuweisende Voraussetzungen für eine bedingte Zulassung sind:

- Positives Nutzen-Risiko-Verhältnis
 - Hierzu ist die Vorhandene Evidenz zum Nutzen abzuwägen gegen bekannte Risiken sowie das Risiko noch fehlender Daten
- Es ist wahrscheinlich, dass der Antragsteller umfassende Daten vorlegen kann
- Erfüllung eines nicht gedeckten medizinischen Bedarfs
- Der Nutzen der sofortigen Verfügbarkeit des Arzneimittels für die öffentliche Gesundheit überwiegt die Risiken, die sich aus der Tatsache ergeben, dass noch zusätzliche Daten erforderlich sind [5].

3.3 Zulassung unter außergewöhnlichen Umständen („Approval under exceptional Circumstances“)

Arzneimittel, für die der Antragsteller nachweisen kann, dass umfassende Daten aufgrund spezifischer gesetzlicher Gründe nicht bereitgestellt werden können, eignen sich möglicherweise für eine Zulassung unter außergewöhnlichen Umständen. Diese Art der Zulassung wird jährlich überprüft, um das Nutzen-Risiko-Verhältnis neu zu bewerten. Die Auflagen der Zulassung unter außergewöhnlichen Umständen zielt auf die Bereitstellung von Informationen über die sichere und wirksame Verwendung des Produkts ab und führt normalerweise nicht zum Abschluss eines vollständigen Zulassungsdossiers. Wenn hingegen absehbar ist, dass umfassende Daten erzielbar sind, wird das Verfahren der bedingten Zulassung gewählt (vgl. 3.2) [6].

Für eine Zulassung unter außergewöhnlichen Umständen muss eine der folgenden Bedingungen erfüllt sein:

- Die Indikationen, für die das betreffende Produkt bestimmt ist, sind derart selten, dass nach vernünftigem Ermessen keine umfassenden Nachweise erbracht werden können.
- Nach dem derzeitigen Stand der wissenschaftlichen Erkenntnisse liegen keine umfassenden Informationen vor.
- Es steht im Widerspruch zu den allgemein akzeptierten Grundsätzen der medizinischen Ethik, solche Informationen zu sammeln [6].

4 PRAKTISCHER UMGANG MIT NICHT-RANDOMISIER- TEN DATEN

4.1 Positionierung des G-BA

Bereits im Vorfeld der Gesetzgebung wurde über die Eignung von Registerdaten für die Nutzenbewertung diskutiert. Der G-BA stellte als Diskussionsbeitrag folgende Kriterien hierzu auf:

„Wesentliche Anforderungen an Registerdaten als Diskussionsgrundlage:

1. Generierung von vergleichender Evidenz muss gewährleistet sein
 - Mit indikationsspezifischen Registern werden auch Daten alternativ behandelte Patienten erfasst
 - Beginn der Datenerhebung vor dem geplanten Marktzugang (Kontrollgruppe)
2. Erhebung von Krankheitssymptomatik und Lebensqualität ist essentiell
3. Register sollen mindestens repräsentative Stichprobe der GKV-Population untersuchen, bei Orphan-Drugs in der Regel Vollerhebung
4. Kompatibilität mit bestehenden Registern (Vermeidung von Doppelerfassungen)“ [7]

4.1.1 Stellungnahme des G-BA zum GSAV

In seiner Stellungnahme verweist der G-BA darauf, dass die Verwendung nicht-randomisierter Studien für Arzneimittel in Indikationen gelte, „in denen es unmöglich oder unangemessen ist, Studien höchster Evidenzstufe durchzuführen oder zu fordern“ [8]. Namentlich zählt der G-BA Anwendungsbeobachtungen, Fall-Kontroll-Studien oder Registerstudien als vom GSAV umfasst auf und verweist darauf, dass auch kontrollierte Studien von den anwendungsbegleitenden Datenerhebungen mit umfasst sein sollten.

Dass der G-BA auch nicht-randomisierte Studien in besonderen Situationen für die Nutzenbewertung als geeignet ansieht, wird deutlich in seiner Stellungnahme zu den vorgesehenen Änderungen im Transfusionsgesetz und in der Transfusionsgesetz-Meldeverordnung.

Hier begrüßt der G-BA die vorgesehene Einführung von Melde- und Dokumentationspflicht für Faktorpräparate zur Behandlung der Hämophilie. Mit dieser Struktur könnten nach Auffassung des G-BA „auch Daten zu neuartigen Wirkstoffen [...] zentral gesammelt und ggf. für eine vergleichende Bewertung im Rahmen der Nutzenbewertung, [...] sowie Daten zur langfristigen Sicherheit und Wirksamkeit“ verwendet werden [8].

4.1.2 Verwendung nicht-randomisierter Daten in Nutzenbewertungsverfahren

In seiner Spruchpraxis hat der G-BA bereits in der Vergangenheit auch nicht-randomisierte Studien zur Anerkennung eines Zusatznutzens herangezogen. In mehreren Verfahren wurde die Befristung des Beschlusses mit der Generierung von Registerdaten beauftragt, die somit als grundsätzlich für die Nutzenbewertung für verwertbar erachtet werden können.

Auch zeigt die Spruchpraxis des G-BA, dass grundsätzlich Situationen existieren, in denen die Evidenz einer nicht-randomisierten Studie zur Quantifizierung eines Zusatznutzens ausreichend ist.

4.1.2.1 Fallbeispiel: Nutzenbewertung von Pasireotid

Für die Nutzenbewertung von Pasireotid (Orphan Drug, Beschluss vom 06.12.201) zieht der G-BA eine Phase-III-Studie ohne Kontrollgruppe heran, in deren beiden Studienarmen lediglich unterschiedliche Dosierungen geprüft wurden. Maßgeblich für die Anerkennung eines Zusatznutzens war die im Prä-Post-Vergleich dargestellte Normalisierung eines pathologisch erhöhten Laborwertes [9].

4.1.2.2 Fallbeispiel: Nutzenbewertung von Cerliponase alfa:

Für die Nutzenbewertung von Cerliponase alfa wurden die Ergebnisse der unkontrollierten Zulassungsstudien mit einer historischen Kontrolle unbehandelter Patienten aus einem bereits existierenden Patientenregister verglichen. Hierzu wurden prospektiv Matching-Kriterien für die 1:1-Paarbildung definiert. Analysen, die auf anderen Matching-Methoden basieren wurden als Sensitivitätsanalysen herangezogen.

Die Zulassung von Cerliponase alfa erfolgte unter „außergewöhnlichen Umständen“ und war mit Auflagen verbunden. Mit einer nicht-interventionellen Post-Autorisation Sicherheitsstudie (PASS) soll anhand von Registerdaten die Langzeitsicherheit beurteilt werden. Anhand einer nicht-interventionellen PAES sollen Krankheitsprogression, sowie Sicherheit und Verträglichkeit beurteilt werden.

Der G-BA hielt es für notwendig, darüber hinaus Daten für die in Deutschland mit Cerliponase alfa behandelten Patienten in einem klinischen Register zu erfassen, um für den deutschen Versorgungskontext repräsentative sowie vollständige Daten zu patientenrelevanten Endpunkten zu generieren [10].

4.1.2.3 Fallbeispiel: Nutzenbewertung von Axicabtagen-Ciloleucel

Bei der Nutzenbewertung der CAR-T-Zell Therapie Axicabtagen-Ciloleucel anerkennt der G-BA die vom pharmazeutischen Unternehmer (pU) vorgelegten historischen Vergleiche als „hinreichend valide für die Bewertung des Ausmaßes des Zusatznutzens“ [11]. Auf-

grund weiterer Unsicherheiten wird in der Gesamtschau dennoch lediglich ein nicht quantifizierbaren Zusatznutzen anerkannt.

Die Befristung wird mit der Generierung weiterer prospektiver, vergleichender Evidenz (z. B. aus Beobachtungsstudien) sowie einem erneuten historischen Vergleich anhand eines noch zu erwartenden Datenschnittes beauftragt [11]. Bislang erfolgte jedoch noch in keinem Fall eine erneute Nutzenbewertung auf Basis der vorzulegenden Registerdaten.

Tabelle 2: Übersicht der Verfahren von Orphan Drugs mit Registereauflagen durch den G-BA

INN / Handelsname	Zulassungsart	Ausmaß Zusatznutzen
Sebelipase alfa (Kanuma®)	Orphan Drug	Nicht quantifizierbar
Asfotasealfa (Strensiq®)	Orphan Drug	Nicht quantifizierbar
Idebenon (Raxone®)	Orphan Drug	Nicht quantifizierbar
Afamelanotif (Scenesse®)	Orphan Drug	Nicht quantifizierbar
Cerliponase alfa (Brineura®)	Orphan Drug Approval under exceptional circumstances	Nicht quantifizierbar

Quelle 2: inav GmbH (2019)

4.2 Positionierung des IQWiG

Die Berücksichtigung versorgungsnaher Daten (sog. Real World Evidenz, RWE) zur Ergänzung der aus RCT gewonnenen Evidenz ist in den zurückliegenden Jahren bereits vielfach diskutiert worden. Das IQWiG hat in dieser Diskussion stets die Position vertreten, dass nicht-randomisierte Daten zu Zwecken der Bewertung des Zusatznutzens nur in sehr wenigen Ausnahmefällen Verwendung finden könnten, ansonsten aber für diese Zwecke ungeeignet sind.

Bereits im November 2015 fand das Herbst-Symposium des IQWiG zu dem Thema „Real World Data: Ein Gewinn für die Nutzenbewertung?“ statt. In der anschließenden Pressemitteilung resümiert die Institutsleitung, dass der Begriff „Real World Data“ verzichtbar sei. Es müsse vielmehr der in Wissenschaft und Politik zu beobachtenden Tendenz entgegengetreten werden, erreichte Standards aufzugeben und ohne Not erhöhte Unsicherheit in Kauf zu nehmen [12].

Die Position des IQWiG zur Bedeutung von versorgungsnahen Daten wird auch aus einer 2017 publiziertem Beitrag zum Stellenwert von Registerdaten hervor. Hier heißt es un-

missverständlich „Patientenregister [...] eignen sich regelhaft nicht für die Nutzenbewertung medizinischer Interventionen“ [13].

4.2.1 Stellungnahme des IQWiG zum GSAV

In der Stellungnahme des IQWiG zum GSAV-Referentenentwurf begrüßt das Institut einerseits die Ermächtigung des G-BA zur ergänzenden Datenanforderung und regt an, diese Kompetenz grundsätzlich auf sämtliche zu bewertenden Arzneimittel auszudehnen. Zur im Gesetzentwurf verankerten konkreten Ausgestaltung schreibt das IQWiG jedoch bereits in der Einleitung: „Die im Referentenentwurf beschriebene und im Begründungsteil weiter konkretisierte Art der Datenerhebung („anwendungsbegleitende Datenerhebung“) ist jedoch häufig für die Bewertung des Zusatznutzens unzureichend und daher in der derzeitigen Form ungeeignet“ [14].

Vom Gesetzgeber für geeignet gehaltene Studientypen, wie Anwendungsbeobachtungen oder Fall-Kontrollstudien streicht das IQWiG aus dem Gesetzentwurf – lediglich Registerstudien und klinische, vergleichende Studien verbleiben nach den Vorstellungen des Instituts. Ebenso wird der Wortlaut „anwendungsbegleitend“ konsequent aus dem Gesetzentwurf gestrichen.

Das IQWiG hebt hervor, dass Erkenntnisdefizite bei der Nutzenbewertung den Vergleich der Anwendung eines Arzneimittels mit seiner Nicht-Anwendung betreffen, wofür „vorrangig direkt vergleichende, randomisierte, kontrollierte Studien“ Verwendung finden müssten. Eigens wird zudem auf die Machbarkeit von RCTs auch bei Orphan Drugs nach erfolgter Zulassung hingewiesen.

Zu den vom Gesetzgeber vorgesehenen Studientypen äußert sich das IQWiG unmissverständlich ablehnend:

- „Anwendungsbeobachtungen sind per definitionem für die erforderlichen vergleichenden Aussagen ungeeignet“
- „Fall-Kontroll-Studien sind [...] „für die hier relevanten Fragestellungen [...] völlig unpassend“
- Register seien zeitaufwändig im Aufbau; mit nicht vergleichenden Produktregister werden „vergleichende Aussagen als Grundlage der Bewertung eines Zusatznutzens nicht möglich sein“ [14].

4.2.2 Regelungen des Methodenpapiers 5.0

In seinem aktuellen Methodenpapier 5.0 positioniert sich das IQWiG zu der Forderung „andere Studientypen ohne Randomisierung hinzuzunehmen, um Alltagsbedingungen besser einfließen zu lassen“. Hierzu führt das Institut aus, dass „andere Studientypen als RCTs [...] in der Regel für einen Kausalitätsnachweis nicht geeignet“ sind [15].

Nicht randomisierte vergleichende Studien wären grundsätzlich kein Garant für eine Strukturgleichheit der Gruppen und lieferten daher immer ein potenziell verzerrtes Ergebnis.

Die Verwendung nicht randomisierter Studien als Nachweis der Kausalität einer Intervention benötigt nach Auffassung des IQWiG „einer besonderen Begründung bzw. besonde-

rer Voraussetzungen und spezieller Qualitätsanforderungen“ [15]. Ausnahmen erkennt das Institut an, bei

- Vorliegen von „dramatischen Effekten“, womit eine Umkehr eines bislang nicht beeinflussbaren deterministischen Krankheitsverlaufs durch eine Intervention zu verstehen ist.
- extrem seltenen Erkrankungen oder extrem spezifischer Krankheitskonstellationen für die die Forderung nach (parallel) vergleichenden Studien unangemessen ist. Hier würden historische Vergleiche anwendbar sein.

Ansonsten stellt aus Sicht des IQWiGs „die randomisierte kontrollierte Studie den Goldstandard bei der Bewertung medikamentöser und nicht medikamentöser Interventionen dar“ [15].

Unterhalb der Evidenzanforderungen für eine Nutzenbewertung nimmt das IQWiG im Rahmen der Potentialbewertung ein detailliertes Ranking von nicht-randomisierten Studien vor:

- „geringe qualitative Ergebnissicherheit: Ergebnis einer höherwertigen nicht randomisiert vergleichenden Studie (beispielsweise quasirandomisierte kontrollierte Studien, nicht randomisierte kontrollierte Studien mit aktiver Zuteilung der Intervention nach vorab geplanter Regel, prospektive vergleichende Kohortenstudien mit passiver Zuteilung der Intervention) mit adäquater Kontrolle für Confounder,
- sehr geringe qualitative Ergebnissicherheit: Ergebnis einer höherwertigen nicht randomisiert vergleichenden Studie (siehe vorheriger Punkt), jedoch ohne adäquate Kontrolle für Confounder, oder Ergebnis einer sonstigen nicht randomisiert vergleichenden Studie (beispielsweise retrospektive vergleichende Kohortenstudien, historisch kontrollierte Studien, Fall-Kontroll-Studien),
- minimale qualitative Ergebnissicherheit: Ergebnis einer nicht vergleichenden Studie (beispielsweise 1-armige Kohortenstudien, Verlaufsbeobachtungen oder Fallserien, Querschnittsstudien oder sonstige nicht vergleichende Studien), das einen indirekten Vergleich mit den Ergebnissen anderer Studien (Literaturkontrollen) erlaubt“ [15].

Als entscheidend erkennt das IQWiG bei fehlender Randomisierung die ausreichende Kontrolle von Confoundern, insbesondere durch multifaktorielle statistische Methoden [15].

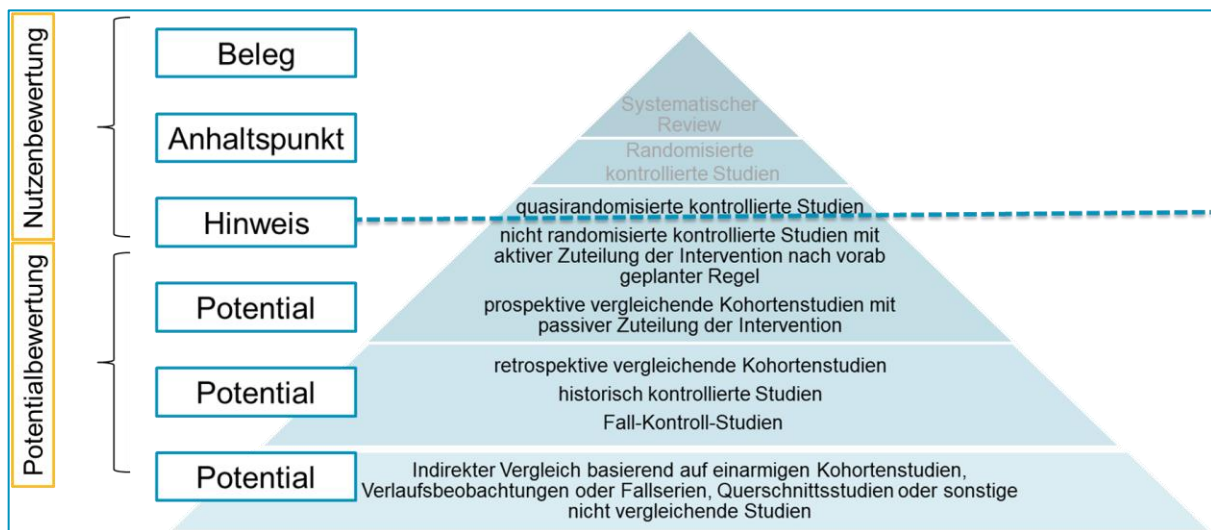


Abbildung 1: „Evidenzpyramide“ aus Sicht des IQWiG

Quelle 3: Darstellung der inav GmbH (2019) nach Allgemeinem Methodenpapier des IQWiG [15]

4.3 Positionierung der Europäischen Arzneimittel-Agentur (EMA)

Die EMA diskutiert an mehreren Stellen die Verwendung nicht-randomisierter Studien zur weiteren Evidenzgenerierung nach erteilter Zulassung. Die dort angestellten methodischen Überlegungen sind die Basis für etwaige Beauftragungen, die die EMA erteilt, um offengebliebene Fragen zur Wirksamkeit und Sicherheit von Arzneimitteln nach Markteintritt zu beantworten.

4.3.1 Post-authorisation efficacy studies (PAES)

4.3.1.1 Nicht randomisierte vergleichende Studien

In Rahmen von Wirksamkeitsstudien nach Markteintritt favorisiert die EMA vergleichende Studiendesigns mit hinsichtlich des Krankheitsverlaufs ähnlichen Patientengruppen. Zudem hält es die EMA für wünschenswert, in der Beobachtungsstudienpopulation eine Teilpopulation zu identifizieren, die derjenigen in vorhandenen RCT weitgehend ähnlich ist. Wenn für diese Populationen ähnliche Ergebnisse zu finden sind, würde das Vertrauen in die Gesamtstudienresultate erhöht werden. Sensitivitätsanalysen zum Testen der Robustheit von Studienergebnissen werden daher eine hohe Bedeutung zugemessen [16].

4.3.1.2 Studien mit historischem Vergleich

Der Vergleich mit historischen Datensätzen kommt für die EMA dann in Frage, wenn es nicht möglich ist, prospektive Daten zu gleichzeitigen Kontrollen zu erhalten. Als Herausforderung werden dabei mögliche Unterschiede zwischen den zu vergleichenden Populationen benannt, die die Behandlungsentscheidung und einen, womöglich im Zeitverlauf veränderten, klinischen Hintergrund betreffen.

Von Bedeutung ist hierbei eine gute Charakterisierung der historischen Daten hinsichtlich gemessener und dokumentierter Auswahlkriterien sowie bekannter Prognosevariablen und deren Abgleich mit der Beobachtungsstudie.

Darüber hinaus erwähnt die EMA die Möglichkeit, zeitgleiche und historische Ansätze gleichzeitig anzuwenden. So könnten z. B. Krankheitsrisikobewertungen auf der Grundlage von Gewichten aus historischen Daten berechnet werden und dann für gleichzeitige Kontrollen verwendet werden. Weiterhin sind auch Prä-Post-Vergleiche möglich, bei denen ein Patient seine eigene Kontrolle für Vergleiche vor und nach der Behandlung ist.

Neben der Verwendung von Primärdaten sieht die EMA die Nutzung von Sekundärdaten, wie z. B. elektronisch Patientenakten, als weiteren Hauptansatz für die Datenbeschaffung [16].

4.3.1.3 Verwendung von Registern

Anwendungsbegleitende Register können die Wirksamkeit und Sicherheit von Arzneimitteln bei der routinemäßigen Behandlung von Krankheiten nach der Zulassung untersuchen. Empfohlen wird dabei die Anbindung an bestehende Krankheitsregister, um „eine fortlaufende Bewertung der Krankheitsergebnisse und einen Vergleich verschiedener Behandlungsoptionen nach einer ähnlichen Methodik“ zu ermöglichen [16]. Die Daten aus bestehenden Registern können dabei durch zusätzliche Datenerfassung oder Verknüpfung mit externen Datenquellen ergänzt werden. Neu zu erstellende Register sollen auf den einschlägigen methodischen Standards basieren und in einem Protokoll beschrieben werden.

Mit den so gewonnen Registerdaten sind eine Vielzahl unterschiedlicher Beobachtungsstudien realisierbar:

- prospektive Kohortenstudien mit eingebetteter Fall-Kontroll-Analyse,
- Anfangskohortenstudien¹,
- retrospektive Kohortenstudien,
- Verlaufsstudien zur Untersuchung des natürlichen Krankheitsverlaufs und
- Kohortenstudien mit internen Komparatoren [16].

In seinem „Discussion paper registries“ verdeutlicht die EMA die Bevorzugung krankheitsbezogener gegenüber produktbezogenen Registern [17]. Als wesentlichen Punkt sieht die EMA die Notwendigkeit der Vermeidung und Kontrolle eines Selektionsbias und führt einige grundsätzliche Anforderungen für Register bzw. Registerstudien an:

- Sorgfältige und standardisierte Dokumentation des zeitlichen Verlaufs von Ereignissen.
- Indikationsbezogene Harmonisierung der Erhebung relevanter Daten mit Gültigkeit für sämtliche Register in der betreffenden Indikation zur Unterstützung nachgelagerter behördlicher Entscheidungen.

¹ Einschluss der Patienten zum Zeitpunkt der Erstmanifestation der Krankheit

- Vereinheitlichung unterschiedlich verwendeter Terminologien für Krankheiten, diagnostische Tests, Symptome, Arzneimittel, Wirkstoffe, unerwünschte Ereignisse und andere relevante Daten auf einen internationalen Standard.
- Qualitätsmanagement als fester Bestandteil des Registermanagements in Hinblick auf die Konsistenz, Vollständigkeit, Genauigkeit und Aktualität der erhobenen Daten.
- Einbeziehung des Registers in bestehende Pharmakovigilanzsysteme und Integration präspezifizierter unerwünschter Ereignisse von besonderem Interesse (adverse events of special interest, AESI) im Datenerfassungssystem.
- Vorbereitungen für die Verwendung eines Patientenregisters nach der Zulassung sollten vor dem Zeitpunkt des Scientific Advice oder in einem frühen Stadium des Genehmigungsantragsverfahrens beginnen.
- Vollständige Erhebung sämtlicher für den Zweck der Studie erforderlicher Daten, einschließlich der verfügbaren Daten zu potenziellen Störfaktoren sowie der für Sensitivitätsanalysen gemäß Studienprotokoll und statistischem Analyseplan (SAP) benötigten Daten.
- Auswahl des am besten geeigneten statistischen Verfahren von Fall zu Fall unter Berücksichtigung der konkreten wissenschaftlichen Fragestellung.

Unterstützend bietet die EMA die Bereitstellung von Tools zur Erleichterung der Anerkennung von Krankheitsregistern als Datenquelle für die Durchführung von Studien zu regulatorischen Zwecken an. Zugleich fördert die EMA den frühzeitigen Dialog mit den Aufsichtsbehörden und bietet eine wissenschaftliche Beratung für Studienprotokolle und die Qualifizierungsverfahren für Register an [17].

4.4 Positionierung anderer Institutionen

4.4.1 GRADE

Das Grading of Recommendations, Assessment, Development and Evaluation (GRADE) System ist ein international verbreitetes und akzeptiertes Konzept zur Evidenzbewertung, wie auch zur Entwicklung von Handlungsempfehlungen im Gesundheitswesen. Der GRADE-Ansatz dient zwar der Bewertung von Evidenzkörpern insgesamt, bedient sich dabei allerdings einer Bewertung des Evidenzlevels auch auf Studien- bzw. Endpunktebene.

Hierbei werden vier Evidenzlevel (hoch, mittel, niedrig, sehr niedrig) voneinander unterschieden. Randomisierte Studien werden dabei zunächst als Evidenz hoher Qualität angenommen, während Beobachtungsstudien initial als Evidenz niedriger Qualität eingestuft werden. Beide Studientypen können ausgehend von der initialen Einstufung je nach Beschaffenheit in ihrer Qualität herab- oder heraufgestuft werden.

Als herabstufende Kriterien gelten (i) Limitationen im Studiendesign, die auf eine hohe Wahrscheinlichkeit von Verzerrungen hinweisen, (ii) Indirektheit der Evidenz, (iii) unerklärte Heterogenität oder Inkonsistenz der Ergebnisse, (iiii), Ungenauigkeit der Ergebnisse (breite Konfidenzintervalle) oder (iiiii) eine hohe Wahrscheinlichkeit eines Publikationsbias. Faktoren, die Evidenz heraufstufen sind (i) große Effektmaße, (ii) sämtliche

Confounder verzerren zu Ungunsten des nachzuweisenden Effektes sowie (iii) eine vorhandene Dosis-Wirkungs-Beziehung [18]. Teilweise wird dabei zwischen einer Herab- bzw. Heraufstufung um ein oder zwei Kategorien unterschieden, in Abhängigkeit vom Ausmaß des betrachteten Effektes.

In Folge dieser Bewertung kann also eine zweifach hochgestufte Beobachtungsstudie dieselbe Qualitätsstufe erreichen, wie eine RCT. Umgekehrt sind auf RCT beruhende Evidenzkörper bei Vorliegen herabstufender Mängel nicht höher einzustufen als wenn sie auf hochwertigen Beobachtungsstudien basieren. Eine Übersicht ist aus Tabelle 3 zu entnehmen.

Tabelle 3: Zusammenfassung des GRADE-Ansatzes zur Evidenzbewertung

Studiendesign	Anfängliche Studienqualität	Heraufzustufen, wenn	Herabzustufen, wenn	Abschließende Studienqualität
Randomisierte kontrollierte Studie	Hoch (++++)	Großer Effekt +1: groß +2: sehr groß	Risiko für Bias -1: hoch -2: sehr hoch	Hoch (++++)
		Dosis-Wirkungsbeziehung +1: nachgewiesen	<u>Inkonsistenz</u> -1: hoch -2: sehr hoch	Mäßig (+++)
Beobachtungsstudie	Niedrig (++)	<u>Sämtliche Confounder</u> +1: würden einen nachgewiesenen Effekt reduzieren +2: würden einen falschen Effekt nahelegen, wenn kein Effekt nachweisbar ist	Ungenauigkeit -1: hoch -2: sehr hoch <u>Publikationsbias</u> -1: wahrscheinlich -2: sehr wahrscheinlich	Niedrig (++)
				Sehr niedrig (+)

Quelle 4: Darstellung der inav GmbH (2019) nach Bahlsem et al. (2018) [18]

4.4.2 EUnetHTA

Im Rahmen des europäischen Netzwerks für die Bewertung von Gesundheitstechnologien (EUnetHTA) wurde mit REQueST (Registry Evaluation and Quality Standards Tool) ein Tool entwickelt, das die europäischen Health Technology Assessment (HTA)-Agenturen bei der Errichtung und Auswertung von Registern unterstützen soll [19]. Das Tool ermöglicht eine konsistente Bewertung der Eignung von Registern für HTA und adressiert zugleich Bedenken hinsichtlich der Zuverlässigkeit der Registrierungsdaten für diese Zwecke.

Für das Assessment ist ein dreistufiges Vorgehen vorgesehen. Im ersten Schritt („Methodische Information“) werden Registern danach bewertet, ob die Daten und Methodik den Anforderungen der Fragestellung entsprechen. Hierzu werden Informationen zu sieben Punkten ausgewertet:

- Art des Registers
- Ziele und Forschungsfragen
- Geografisches und organisatorisches Umfeld (einschließlich Datenlieferanten)

- Dauer
- Größe
- Ein- und Ausschluss Kriterien
- Follow-Up

In einem zweiten Schritt („Grundlegende Standards“) werden Minimalanforderungen definiert, ohne deren Einhaltung die HTA-Agentur das Register nicht zu Zwecken der Evidenzbewertung verwenden sollte.

- Registerprotokoll
- Governance
- Qualitätssicherung
- Finanzierung
- Datenerfassung
- Mindestdatensatz
- Standarddefinitionen, Terminologie und Spezifikationen
- Konfounder
- Datenbereinigung
- Schutz, Sicherheit und Schutzmaßnahmen
- Einverständniserklärung.

Im dritten Schritt („Zusätzliche Anforderungen“) werden weitere Aspekte wie Interoperabilität, Datenquellen und Ethik adressiert.

Das Tool ist geeignet, Register anhand eines – unter Mitwirkung des IQWiG – entstandenen Konzeptes auf seine Eignung hin zu bewerten und stellt zugleich die Kriterien für ein neu aufbauendes, hochwertiges Register nachvollziehbar dar [[19](#)].

Introduction

Instructions for use

Methodological Information

Essential Standards

Additional Requirements

Output

Glossary and explanations

DRAFT - FOR PUBLIC CONSULTATION

EUnETHA WPSb standards in registries draft tool

Output

Instructions

Individual scores are automatically copied from the Methodological Information, Essential Standards and Additional Requirements worksheets. The area score is automatically generated by the tool.

Registry name: [Registry holder to complete]
 Date of assessment: [Registry holder to complete]

Key:

Satisfactory

 Needs development / clarification

 Unsure / pass

Abbildung 2: REQueST

Quelle 5: EUnethTA (Request tool and vision paper) (2019) [20]

4.4.3 Cochrane Collaboration

Die Cochrane Collaboration hat mit ROBINS-I ("Risk Of Bias In Non-randomised Studies - of Interventions") ein Tool zur reliablen Bewertung von Verzerrungen nicht-randomisierter Studien entwickelt [21]. Denkbare Verzerrungsquellen, die vor, während oder nach der Intervention vorhanden sein können, werden separat bewertet und führen zu einer Gesamtbewertung der Studie.

Vor der Intervention sind hierbei prognostische Confounder sowie das Verzerrungspotential zu bewerten. Zur Identifizierung potentieller Störgrößen wird die Zusammenarbeit mit Fachexperten und die Verwendung literaturbasierten Wissens empfohlen.

Während der Intervention ist die mögliche Verzerrung durch die Klassifizierung der Interventionen selbst zu bewerten. Die anderen potentiell verzerrenden Effekte liegen zeitlich nach Beginn der Intervention: Verzerrungen aufgrund von Abweichungen von den beabsichtigten Interventionen, fehlende Daten, Messung der Ergebnisse und Auswahl der berichteten Ergebnisse.

Im Ergebnis erfolgt eine Bewertung des Verzerrungspotenzials einer Studie in eine der Kategorien "Geringes Risiko", "Mäßiges Risiko", "Schweres Risiko" und "Kritisches Verzerrungspotenzial". Die Kategorie des „geringen Risikos“ entspricht dabei dem Verzerrungs-

potenzial in einer qualitativ hochwertigen randomisierten Studie. Diese Kategorie wird für eine nicht randomisierte Studie nur in Ausnahmefällen erreicht werden.

Die Notwendigkeit der Akzeptanz von höheren Verzerrungspotentialen wird mit den Regelungen des GSAV grundsätzlich notwendig. Ungeachtet dieser Notwendigkeit und einem möglichen Grenzwert für die Aussagesicherheit, sollte das Verzerrungspotential der im Kontext der anwendungsbegleitenden Datenerhebung erfolgenden Studien anhand eines einheitlichen Maßstabes bewertet werden.

4.4.4 Deutsches Netzwerk für Versorgungsforschung (DNVF)

Das DNVF hatte bereits 2010 ein Memorandum „Register für die Versorgungsforschung“ veröffentlicht, das 2018 mit einem Update 2018 eine Aktualisierung der Empfehlungen erfahren hat [22].

Ein Register wird vom DNVF als „möglichst aktive, prospektive, standardisierte Dokumentation von Beobachtungseinheiten“ definiert, die „zu vorab festgelegten, aber im Zeitverlauf erweiterbaren Fragestellungen, für die ein präziser Bezug zur Zielpopulation darstellbar ist“ Antworten liefert.

Als Maßgabe für eine möglichst hohe interne Validität sieht das DNVF „ein hohes Maß an Standardisierung aller verwendeten Methoden von der Erfassung über die Dateneingabe bis hin zur Auswertung“ sowie den „Einsatz von registereigenem oder eigens beauftragtem, qualifiziertem und geschultem Personal“ Um eine möglichst hohe externe Validität sicherzustellen muss nach Auffassung der Autoren die Quellpopulation eines Registers repräsentativ für die Zielpopulation sein. Hierzu sei die „Quellpopulation und Auswertungskollektiv [...] genau zu charakterisieren (z. B. Region der Stichprobe, Grunderkrankung, therapeutische Maßnahme, Alter, Geschlecht, Zeitraum, usw.)“ [22].

Das Netzwerk differenziert zwischen den Einsatzmöglichkeiten von RCT und Registern. Während RCTs den „Therapieeffekt[s] unter experimentellen Bedingungen (efficacy)“ mit „weitgehenden Schutz vor konfundierenden Störgrößen“ misst, während ein Register, die „Wirksamkeit der „Versorgung mit einem Wirkstoff/Produkt [...] im Kontext des Versorgungssystems dokumentiert, evaluiert und analysiert (effectiveness)“ [22].

Register haben nach Auffassung des DNVF gegenüber klinischen Studien eine unterstützende Funktion „durch die Beobachtung und die Evaluation von Wirksamkeit, Sicherheit und Effizienz in der Versorgungsroutine“ [22].

Unter Verweis auf eine kontroverse Diskussion zur Frage, ob Register zur Bewertung von Wirksamkeit bzw. des Zusatznutzens herangezogen werden können, sieht das Netzwerk durchaus die Möglichkeit, dass Register eine unterstützende Rolle in der Nutzenbewertung spielen können. Das DNVF erwähnt allerdings lediglich die Definition der Zweckmäßigen Vergleichstherapie bzw. epidemiologische Fragestellungen.

Um konfundierenden Störgrößen oder effektmodifizierende Variablen auszugleichen, werden unterschiedliche statistische Kontroll- und/oder Adjustierungsverfahren angeführt:

- Stratifizierung/Subgruppenanalysen: Aufteilung des Registerkollektivs anhand potenziell konfundierender/effektmodifizierender Variable in hinsichtlich der Stratifizierungsvariablen homogener Strata/Subgruppen.
- Matching/Parallelisierung: Paarbildung mit hinsichtlich potenziell konfundierender/effektmodifizierender Variablen identischer bzw. vergleichbarer Gruppen oder Gruppenteilnehmer.
- Propensity-Score-Matching: Ermittlung des Propensity-Scores (PS) mit anschließendem Matching. Der PS ist dabei die Wahrscheinlichkeit, der zu prüfenden Therapie zu erhalten und wird mit einem logistischen Regressionsmodell ermittelt, in dem die zugeordnete Therapie die abhängige Variable darstellt.
- Standardisierung: Nachträgliche Angleichung an eine hinsichtlich potenziell konfundierender/effektmodifizierender Variablen stratifizierte Vergleichspopulation durch prozentuale Wichtung.
- Multivariate Modellierung: Statistische Effektadjustierung durch Integration potenziell konfundierender/effektmodifizierender Variablen als Kovariaten [22].

5 MÖGLICHE EVIDENZARTEN

Im Unterschied zu klinischen Studien handelt es sich bei der anwendungsbegleitenden Datenerhebung um nicht-interventionelle Prüfungen nach dem Arzneimittelgesetz (AMG). Dort ist in § 4 Abs. 23 definiert: „Eine nicht interventionelle Prüfung ist eine Untersuchung, in deren Rahmen Erkenntnisse aus der Behandlung von Personen mit Arzneimitteln gemäß den in der Zulassung festgelegten Angaben für deren Anwendung anhand epidemiologischer Methoden analysiert werden; dabei folgt die Behandlung einschließlich der Diagnose und Überwachung nicht einem vorab festgelegten Prüfplan, sondern ausschließlich der ärztlichen Praxis“ [23].

Zudem fokussiert der Gesetzgeber die Einbeziehung nicht-randomisierter Studien. Gegen deren Verwendung in der Nutzenbewertung wurde seitens IQWiG und G-BA bislang damit argumentiert, dass sie mangels Randomisierung nicht für die Darstellung kausaler Zusammenhänge geeignet seien, da unbekannte konfundierende Faktoren nicht kontrolliert würden.

Diese Argumentation ist zwar vom Grundsatz her nicht verkehrt, verkennt allerdings den Sachverhalt, dass jegliche vermeintlich kausale Aussage aus Studien sowohl eine Ergebnisunsicherheit als auch eine Fehlerwahrscheinlichkeit beinhaltet.

Entsprechend sieht die AM-NutzenV unterschiedliche Grade der Ergebnissicherheit vor, die auf Studienebene oder auch auf Ebene des jeweiligen Endpunktes zu bestimmen sind. Neben dem „Beleg“ werden auch „Hinweise“ oder „Anhaltspunkte“ zur Anerkennung eines Zusatznutzens anerkannt.

Die Verwendung nicht-randomisierter Studien für die Bewertung des Zusatznutzens geht demnach mit einer gegenüber RCT erniedrigten Ergebnissicherheit auf Studienebene einher. Dieser Sachverhalt spricht indes nicht grundsätzlich gegen die Verwendung derarti-

ger Studien. Vielmehr beinhaltet er die Aufgabe, den Fehler durch Anwendung spezieller Methoden möglichst gering zu halten und bestmöglich abzuschätzen, in welcher Größenordnung verbleibende, verzerrende Faktoren das Studienergebnis beeinflussen.

5.1 Fall/Kontroll-Studien

Bei einer Fall-Kontroll-Studie handelt es sich in der Regel um eine retrospektive Beobachtungsstudie. Üblicherweise werden sie für den Vergleich von Expositionsfaktoren herangezogen. Hierzu wird eine Population mit Erkrankung („Fall“) mit einer Population ohne Erkrankung („Kontrolle“) verglichen.

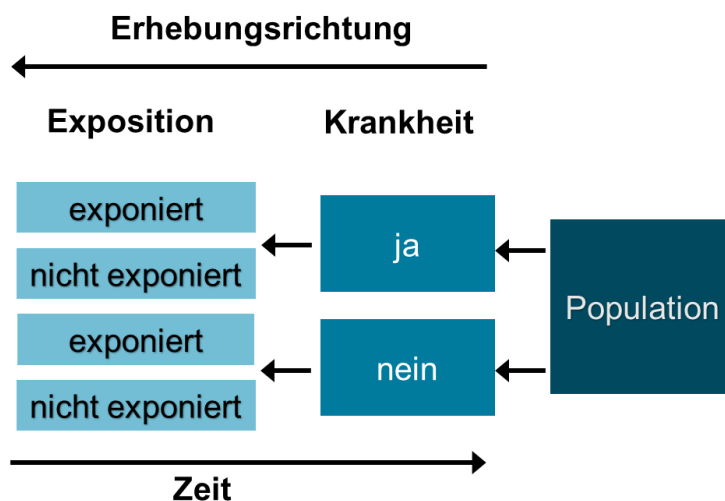


Abbildung 3: Prinzip einer retrospektiven Fall-Kontroll-Studie

Quelle 6: Darstellung der inav GmbH (2019) nach Rhöring et al. (2009) [24]

Die grundlegende Herausforderung bei der Anwendung von Fall-Kontroll-Studien ist der Sachverhalt, dass ein signifikanter Unterschied zwischen den beiden Studienarmen eine Korrelation, aber keine Kausalität zwischen Risikofaktor und Erkrankung aufzeigt.

Die, der Nutzenbewertung zugrunde liegende Fragestellung ist jedoch, ob ein patientenrelevanter Effekt kausal durch die zu bewertende Intervention in höherem Ausmaß positiv beeinflusst wird als durch die Vergleichstherapie.

5.2 Prospektive vergleichende Kohortenstudien

Eine Kohorte ist eine Gruppe von Personen, die über einen längeren Zeitraum beobachtet wird. In einer Kohortenstudie werden Gruppen verglichen, die verschiedenen Einflüssen ausgesetzt sind, wie z. B. der Anwendung einer neuen Therapieoption.

Ein Inter-Kohortenvergleich, bei dem die Mitglieder verschiedener Kohorten miteinander verglichen werden, stellt ein anwendbares Studiendesign für Nutzenvergleiche dar. Für die Einbeziehung neu auf dem Markt befindlicher Arzneimittel bietet sich ein prospektives Design an. Die Ereignisse, die gemessen werden sollen, wie auch mögliche Einflussgrößen müssen im Vorfeld exakt beschrieben werden.

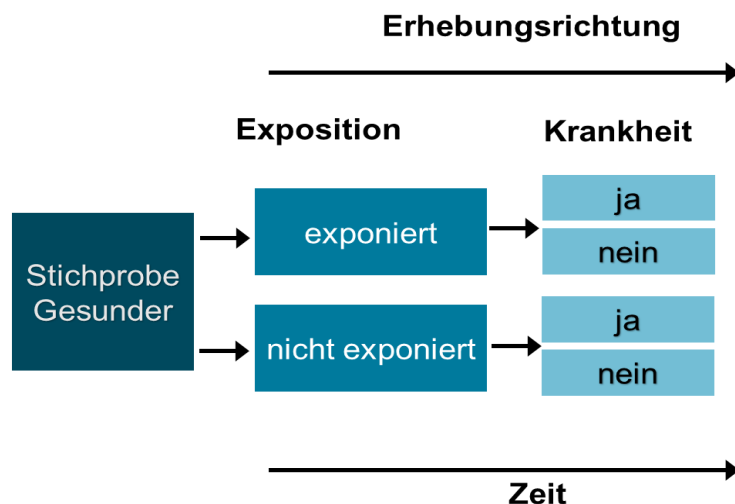


Abbildung 4: Prinzip einer prospektiven Kohortenstudie

Quelle 7: Darstellung der inav GmbH (2019) nach Rhöring et al. (2009) [23]

Eine Herausforderung bei der Verwendung von Kohorten-Studien für die Nutzenbewertung besteht darin, die Merkmalsgleichheit zwischen den beiden Studienarmen zu gewährleisten. Hierbei ist insbesondere die Kontrolle des Selektionsbias bedeutsam, da sich Patienten, bei denen eine neue Therapieform angewendet wird, womöglich bedeutsam von Patienten unterscheiden, denen die neue Therapieform nicht angeboten wird.

5.3 Einarmige Studien

Die EMA erteilt, in den von der neuen Gesetzgebung umfassten Sondersituationen auch Zulassungen, die auf einarmigen Studien basieren (z. B. Nivolumab oder Pembrolizumab in der Indikation Hodgkin Lymphom, Vismodegib in der Indikation Basalzellkarzinom und Crizotinib in der Indikation ROS1-positives, fortgeschrittenes nicht kleinzelliges Lungenkarzinom).

Die nach dem Methodenpapier des IQWiG, wie auch der Verfahrensordnung des G-BA grundsätzlich verwendbaren adjustierte indirekte Vergleiche über Brückenkomparatoren sind allerdings auf dieser Datenbasis nicht möglich.

Eine vergleichende Bewertung gegenüber der ZVT kann dann nur durch einen Vergleich gegenüber Armen anderer Studien oder einen historischen Vergleich vorgenommen werden.

Die Herausforderung bei derartigen Vergleichen besteht darin, Unterschiede in den Patientenkollektiven in Bezug auf potenziell verzerrende Faktoren, oder auch zeitliche Unterschiede bei der Studiendurchführung auszugleichen. Hierzu werden Verfahren benötigt, die eine Adjustierung nach solchen verzerrenden Faktoren ermöglichen.

Gängige Verfahren sind Matching-adjustierte indirekter Vergleiche (MAIC) oder das Propensity-Score-Matching, bei denen Patientencharakteristika des Interventionsarmes durch Gewichtung denen des Kontrollarmes angepasst werden [vgl. 5.5].

5.4 Historische Vergleiche

Die Vollständigkeit *produktbezogener* Register kann über die gesetzlich vorgesehene Verordnungseinschränkung weitgehend sichergestellt werden, auch wenn eine patienten- seitige Rücknahme des Einverständnisses zur Datenverwertung möglich ist. Die gesetzlich ebenfalls erwähnten *krankheitsbezogenen* Register werden hingegen keine Vollständigkeit gewährleisten können, da es für eine Verpflichtung zur Datenerhebung bei Therapie mit nicht von der Nutzenbewertung betroffenen Therapieoptionen keine gesetzliche Grundlage gibt. Zudem ist davon auszugehen, dass auch in einem vollständigen, krankheitsbezogenen Register ein schwer zu kontrollierender Selektionsbias existiert zwischen Patienten, denen eine neuartige Therapieoption angeboten wird und Patienten, die nach dem bisherigen Standard therapiert werden.

Die vom Gesetzesvorhaben umfassten Besonderheiten in der Zulassung betreffen zudem Situationen, in denen keine hinreichende Therapiealternative existiert. In Folge werden Situationen entstehen, in denen lediglich ein historischer Vergleich mit anderen Studienarmen bzw. ein Vergleich mit dem natürlichen Krankheitsverlauf vorgenommen werden kann.

Ein historischer Vergleich stellt grundsätzlich einen nicht adjustierten indirekten Vergleich dar. Für die Verwertbarkeit muss – mangels Randomisierung – eine hinreichende Ähnlichkeit der Studien insbesondere in Hinblick auf die Population sowie das Studiendesign gewährleistet sein.

Im Methodenpapier des IQWiG wird allerdings deutlich, dass in der regulären Nutzenbewertung - auch bei großer Ähnlichkeit, der für einen historischen Vergleich herangezogenen Studienarme - nur im Falle sogenannter dramatischer Effekte Akzeptanz finden [15].

Bei Vorliegen „extrem seltener Erkrankungen oder extrem spezifischer Krankheitskonstellationen“ hält das IQWiG hingegen eine Bewertung anhand historischer Vergleichsdaten für möglich [15]. Weiterhin erwähnt das Methodenpapier die Verwendung historischer Vergleiche im Rahmen der Potentialbewertung.

Die Forderung nach „dramatischen Effekten“ kann in Bezug auf die vom GSAV umfassten Arzneimittel entfallen, da die anwendungsbegleitende Datenerhebung auf den, von der EMA in Sonderzulassungen bereits getroffenen Entscheidungen aufbaut [vgl.6.2].

5.5 Prä-Post-Vergleiche

Für die Durchführbarkeit von historischen Vergleichen ist es grundsätzlich notwendig Daten aus der Zeit vor Einführung der zu bewertenden Arzneimittels heranzuziehen. Ob und in welcher Qualität derartige Daten verfügbar sind, sollte für den Beschluss des G-BA zur anwendungsbegleitenden Datenerhebung berücksichtigt werden, kann allerdings zu diesem Zeitpunkt nicht mehr beeinflusst werden.

Falls keine geeigneten Daten für einen historischen Vergleich zur Verfügung stehen, können für die Quantifizierung des Zusatznutzens Prä-Post-Vergleiche herangezogen werden. Die grundsätzliche Verwendbarkeit derartiger Vergleiche wurde in den Verfahren zu Pasireotid [vgl. 4.1.2.113] sowie zu Vestronidase alfa aufgezeigt, in welchem der G-BA einen

intra-individuellen Prä-Post-Vergleich für die mittlere prozentuale Veränderung eines relevanten Laborparameters für die Bewertung herangezogen hat [25].

Ein derartiger Vergleich bietet sich insbesondere für die Bewertung von Endpunkten aus den Kategorien Lebensqualität und Morbidität an. So kann mit intra-individuellen Prä-Post-Vergleichen bewertet werden, ob eine neue Therapieoption in der Lage ist, die Lebensqualität oder Parameter der Morbidität zu verbessern oder zumindest konstant zu halten.

Für derartige Eingruppen-Prä-Post-Untersuchungen existieren unterschiedliche Berechnungsmethoden für die Effektstärke-Indizes. Häufig verwendet wird das Cohens d: Hierzu wird die Differenz zwischen den Mittelwerten der Kohorte zu den Messzeitpunkten gebildet und durch die Standardabweichung dividiert. Für die Bestimmung der Standardabweichung existieren allerdings unterschiedliche Methoden. Eine Bewertung könnte sich dann an der z. B. mit Cohens d ermittelten Effektstärke nach Gignac et al. orientieren (klein/mittel/groß) [26].

Die Durchführung von Prä-Post Vergleichen sollte nur dann in Erwägung gezogen werden, wenn einerseits kein historischer Vergleich möglich ist und andererseits aufgrund von Kenntnissen des Krankheitsverlauf nicht zu erwarten ist, dass eine (fiktive) unbehandelte Kontrollgruppe eine Verbesserung des Gesundheitszustandes oder der Lebensqualität erfahren würde.

6 MÖGLICHE AUSWERTUNGSMETHODEN

6.1 Umgang mit der Irrtumswahrscheinlichkeit

Die Höhe der akzeptierten Irrtumswahrscheinlichkeit sollte grundsätzlich in die Betrachtung geeigneter Auswertungsmethoden einbezogen werden. Die übliche Verwendung eines 95 %igen Konfidenzintervalls beinhaltet die Akzeptanz eines Fehlerniveaus bei 5 % (Wahrscheinlichkeit, eine zutreffende Nullhypothese zu Unrecht abzulehnen). Streng genommen gilt dies nur für den primären Endpunkt einer Studie, während weitere Endpunkte den α -Fehler nicht mehr auf diesem Signifikanzniveau kontrollieren. Allerdings werden auch diese Endpunkte ohne kritische Diskussion in der Nutzenbewertung herangezogen. Die von IQWiG und G-BA vorgenommene Nutzenbewertung befindet sich somit außerhalb des confirmatorischen Settings.

Für die vom GSAV umfassten Arzneimittel ist davon auszugehen, dass die für eine anwendungsbegleitende Datenerhebung zur Verfügung stehende Anzahl von Patienten limitiert ist. Zugleich scheint es aus regulatorischer Perspektive sinnvoll, die Dauer der veranlassten Datengenerierung zu limitieren, um innerhalb einer angemessenen Zeitspanne die darauf aufbauende Nutzenbewertung vollziehen zu können. Somit stehen mit Fallzahl und Studiendauer die beiden bedeutsamsten Parameter bereits fest, die üblicherweise bei einer Studienplanung herangezogen werden, um eine Aussage mit einem Fehlerniveaus von 5 % ableiten zu können. Vor diesem Hintergrund ist es ratsam – in Abhängigkeit von der zu erwartenden Effektgröße – das geforderte Fehlerniveau entsprechend den Erfordernissen des Einzelfalls anzupassen.

Ein vergleichbares Vorgehen – allerdings bezogen auf randomisierte Studien – ist auch dem IQWiG nicht fremd: „Bei sehr seltenen Erkrankungen könnte [...] erwogen werden, ein größeres statistisches Irrtumsniveau für regulative Entscheidungen zuzulassen [...] 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.“ [27].

In Situationen, in denen der Nachweis eines patientenrelevanten Effektes nicht innerhalb eines akzeptablen Zeitraums möglich erscheint (z. B. aufgrund geringer Patientenzahlen), geht die EMA einen alternativen Weg. Durch Akzeptanz von Surrogatendpunkten wird ermöglicht auch bei geringen Patientenzahlen in kürzerer Zeit Aussagen zum Nutzen einer neuen Therapieoption zu machen. Die Akzeptanz niedrigerer Anforderungen an die Validität von Surrogatendpunkten geht mit einer niedrigeren Ergebnissicherheit einher, könnte allerdings vor dem Hintergrund der gesetzlichen Intention in Bezug auf die vom GSAV umfassten Arzneimittel in die Diskussion einbezogen werden.

6.2 Forderung nach „dramatischen Effekten“

Die Bewertung von Arzneimitteln anhand anwendungsbegleitend erhobener Daten baut auf den von der EMA in Sonderzulassungen bereits getroffenen Entscheidungen auf. Die Miteinbeziehung dieser Entscheidungen ermöglicht die in der AM-NutzenV geforderte Kongruenz zwischen Nutzenbewertung und Feststellungen der Zulassungsbehörde. Ein Festhalten an der in regulären Nutzenbewertungsverfahren üblichen Erforderlichkeit eines „dramatischen Effekts“ würde der gesetzlichen Intention nicht gerecht werden und sollte in Bezug auf die vom GSAV umfassten Arzneimittel entfallen.

Die Kritik des G-BA an indirekten Vergleichen, die nicht über einen gemeinsame Brückenkomparator durchgeführt wurden, geht oftmals in die Richtung, dass trotz Adjustierung die zu vergleichenden Therapieeffekte nicht in einer Größenordnung lägen, mit der ausgeschlossen werden kann, dass die beobachteten Effekte allein auf systematischer Verzerrung oder Zufallsbefunden beruhen.

Der Anspruch, eine Fehleinschätzung mit absoluter Sicherheit ausschließen zu wollen, steht jedoch der auch in regulären Bewertungsverfahren vorgenommenen Einschätzung der Ergebnissicherheit und Irrtumswahrscheinlichkeit entgegen. Bereits die übliche Akzeptanz von Evidenz mit einer Ergebnissicherheit von „Anhaltspunkt“ oder „Hinweis“ beinhaltet eine verbleibende Unsicherheit. Letztlich ist selbst bei einem „Beleg“ unter Anwendung eines 95 %igen Konfidenzintervalls ein Irrtum niemals ausgeschlossen. Der explizite Auftrag des Gesetzgebers beinhaltet die Akzeptanz von Evidenz mit geringerer Aussagekraft als RCT sie entfalten. Die Beschränkung der Akzeptanz eines historischen Vergleiches auf dramatische Effekte ist vor diesem Hintergrund nicht zielführend. Sinnvollerweise sollte sich die grundsätzliche Akzeptanz historischer Vergleiche künftig auf die vom GSAV umfassten Besonderheiten der Zulassung erstrecken.

Die Anforderungen an die Verwendbarkeit historischer Vergleiche benennt der G-BA im Bewertungsverfahren von Elosulfase alfa: „Aufgrund der per se geringen Ergebnissicher-

heit eines historischen Vergleichs sind sowohl Vollständigkeit als auch eine möglichst gleiche Erfassungsmethode der zugrundeliegenden Daten, vor allem hinsichtlich prognostischer Faktoren, Repräsentativität und Selektion der möglichen Kontrollen, zentrale Voraussetzungen“ [28].

Ein Beispiel für die Anerkennung eines historischen Vergleichs im Rahmen der Nutzenbewertung ist das Verfahren Vismodegib [29], bei dem die Ansprechrate gegenüber Best-Supportive-Care (BSC) zur Anerkennung eines Zusatznutzens führte. Hierbei handelte es sich um ein Verfahren, das von der neuen Regelung des GSAV betroffen gewesen wäre, da die Zulassung von Vismodegib unter „exceptional circumstances“ erfolgte.

Weitere Verfahren mit Anerkennung historischer Vergleiche betreffen die Indikationen Chronische Hepatitis C und HIV-Infektion.

6.3 Adjustierungsverfahren.

Um eine Adjustierung nach unterschiedlichen Patientencharakteristika zu erreichen existieren unterschiedliche Verfahren.

Grundlegende Voraussetzung für diese Verfahren ist die Kenntnis potentiell verzerrender Faktoren wie prognostische Faktoren, Krankheitscharakteristika oder demografische Faktoren. Anders als bei einer Randomisierung können unbekannte Störgrößen über eine Adjustierung nicht erfasst werden.

Es gibt eine Vielzahl unterschiedlicher Adjustierungsmethoden, von denen einige im Folgenden kurz beschrieben werden.

6.3.1 Propensity Score Matching (PSM)

Hierbei handelt es sich um ein zweistufiges Verfahren, in dem zunächst der Propensity Score (PS) bestimmt wird. Hierbei handelt es sich um die Wahrscheinlichkeit, die zu prüfende Therapie zu erhalten. Sie wird anhand der bekannten Faktoren, die für eine Therapieentscheidung relevant sind mit einer logistischen Regression ermittelt.

Wichtig bei der Ermittlung des PS ist es, keine Faktoren einzubeziehen, die Einfluss auf den Therapieerfolg haben können.

In einem zweiten Schritt wird nun jedem Patienten ein oder mehrere nicht behandelte Patienten mit dem gleichen Propensity Score zugeordnet und so eine Kontrollgruppe gebildet. Nach diesem Matching kann der Vergleich zwischen behandelten und nicht behandelten Patienten erfolgen und der Therapieeffekt geschätzt werden.

Mit diesem Verfahren gelingt eine Adjustierung auf bekannte und gemessene Störgrößen. Unbekannte Störgrößen, die bei einer Randomisierung gleichmäßig verteilt werden würden, können so nicht berücksichtigt werden. Allerdings sind auch die nicht bekannten Störgrößen häufig mit den bekannten assoziiert, so dass sie durch das PSM ebenfalls erfasst werden können. Als Nachteil des Verfahrens kann genannt werden, dass nicht für jeden Patienten ein Matchingpartner gefunden werden kann und somit die Fallzahl und in Folge auch die statistische Power sinken kann [30].

6.3.2 Matching-adjustierter indirekter Vergleich (MAIC)

Der MAIC ist ein weiteres Verfahren für einen adjustierten indirekten Vergleichs, der es ermöglicht, Studienarme aus unterschiedlichen Studien auch ohne Brückenkomparator zu vergleichen. Dieser Ansatz stellt eine Art von Propensity Score Wichtung dar, bei dem die Patientencharakteristika durch deren Gewichtung angeglichen werden. Hierzu wird für eine Studie mit bekannten Patientencharakteristika die Wahrscheinlichkeit ermittelt, mit der Patienten in der Behandlungsgruppe sind. Deren reziproker Wert wird dann im Vergleich zur Behandlungsgruppe der anderen Studie gewichtet [31].

6.3.3 Simulierter Therapievergleich (Simulated Treatment Comparison, STC)

Bei dem simulierten Therapievergleich (STC) handelt es sich um ein relativ neues Verfahren, das 2010 von (dem ehemaligen IQWiG-Mitarbeiter) Caro et.al. 2010 [32] beschrieben wurde und im ersten Bewertungsverfahren von Axitinib erstmalig in der Nutzenbewertung vorgenommen wurde.

Bei diesem Verfahren werden aus einer Index-Studie potentielle prädiktive Faktoren identifiziert und mit Regressionsverfahren ein Adjustierungskoeffizient errechnet. Mit diesem Koeffizienten kann dann ein anderer Studienarm so kalibriert werden, dass eine vergleichende Studie simuliert wird. Durch diese Adjustierung ist gegenüber einem naiven indirekten Vergleich vorteilhaft, da nach identifizierten Effektmodifikatoren adjustiert wird.

Als bedeutsamen Nachteil der STC hebt das IQWiG im Verfahren zu Axitinib hervor, dass sie „keine Konfidenzintervalle für Überlebenszeit-Analysen [liefert], so dass unklar bleibt, ob ein statistisch signifikanter Effekt vorliegt“ [33].

6.3.4 Bayes-Benchmarking Analysis (BBA)

Ziel der BBA ist es, für einen gegenüber einer Vergleichstherapie zu bewertendem Endpunkt einen Grenzwert zu identifizieren, dessen Überschreitung einer statistisch signifikanten Überlegenheit der neuen Therapie entspricht. Das Verfahren wurde in der Nutzenbewertung von Daclatasvir verwendet, vom IQWiG aber aufgrund von Unvollständigkeit der verwendeten Daten nicht näher bewertet [34].

7 STUFENMODELL ZUR EVIDENZGENERIERUNG

Es ist nicht davon auszugehen, dass die von der neuen Gesetzgebung umfassten Arzneimittel mit einem generischen Konzept durch RWE-Daten bewertbar werden. Vielmehr wird für verschiedene Situationen spezifisch nach einem geeigneten Vorgehen gesucht werden müssen, die erforderliche Evidenz im Rahmen des Möglichen zu generieren.

Hierzu bietet sich ein zweistufiges Verfahren an, bei dem in einem ersten Schritt die noch zu erwartende Evidenz in Hinblick auf ihre Eignung für das Nutzenbewertungsverfahren bewertet wird. In einem zweiten Schritt gilt es dann den höchstmöglichen Evidenzlevel zu identifizieren und anhand der zu erwartenden Erkenntnisse über eine anwendungsbegleitende Datenerhebung und deren Ausgestaltung zu entscheiden.

Grundsätzlich sollte bei vorhandener oder erwartbarer RCT keine anwendungsbegleitende Datenerhebung erfolgen, da die Nutzenbewertung auf Grundlage höherer Evidenzstufe durchgeführt werden kann. In diesem Fall kann bei offengebliebenen Fragestellungen der Nutzenbewertung die Befristung und Neubewertung zum Zeitpunkt der Vorlage der Studienergebnisse erfolgen.

7.1 Bewertung und Einbeziehung der erwartbaren Evidenz

Für einen Großteil der vom GSAV umfassten Arzneimittel ist aus dem Zulassungsprozess heraus auch nach dem Markteintritt noch weitere Evidenz zu erwarten. So sind CMA grundsätzlich mit Auflagen seitens der EMA versehen. Hierbei kann es sich um unterschiedliche Studientypen handeln, in den meisten Fällen (77/107) [35] ist jedoch Evidenz aus klinischen Studien zu erwarten, wobei sich die Auflage mehrheitlich (48/77) [35] auf die Vorlage der endgültigen Ergebnisse aus bereits laufenden klinischen Studien bezieht. Bei über 50 % der angeforderten klinischen Studien (40/77) handelt es sich um mehrarmige, randomisierte Studien [35] [vgl. Abb. 5].

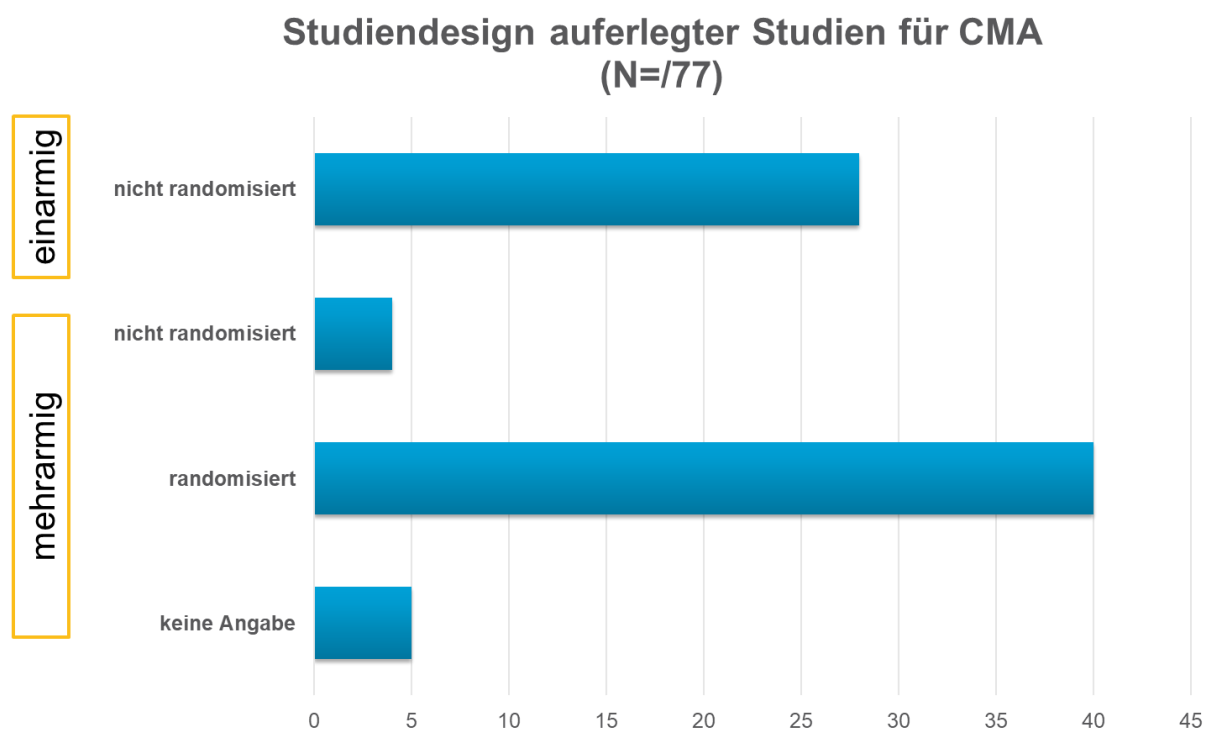


Abbildung 5: Studiendesign auferlegter Studien für CMA (N=77)

Quelle 8: Darstellung der inav GmbH nach EMA (2017) [35]

Grundsätzlich wird empfohlen frühzeitig den Dialog mit der Zulassungsbehörde zu suchen und produktbezogen eine Harmonisierung der Anforderungen anzustreben. Auf diese Weise können z. B. Daten aus einem europäischen Register umfassend verwendet werden, während sich eine Auflage des G-BA ausschließlich auf den deutschen Versorgungskontext wirksam umsetzen ließe. Über die Einbeziehung von Registerdaten

7.2 Identifizierung des höchstmöglichen Evidenzlevels

Wenn keine Evidenz erwartbar ist, die ganz oder nach ergänzenden Auflagen geeignet ist einen Zusatznutzen zu quantifizieren, muss geprüft werden, auf welchem Evidenzlevel anwendungsbegleitend erhobene Daten zur Quantifizierung eines Zusatznutzens erhoben werden können.

Im ersten Schritt ist hierbei zu prüfen, ob eine prospektiv vergleichende Kohortenstudie in Betracht gezogen werden kann. Dies ist insbesondere dann der Fall, wenn

1. sichergestellt werden kann, dass auch für die Kontrollgruppe hinreichend vollständige oder repräsentative Daten erhoben werden können.
2. die Kriterien der Therapieentscheidung nicht zu einem unkontrollierbaren Selektionsbias führen.

Zur Bewertung der Therapieentscheidungskriterien sollten die wissenschaftlichen Fachgesellschaften einbezogen werden.

Ist eine prospektiv vergleichende Kohortenstudie nicht sinnvoll durchführbar, stellen ein-armige Studien mit einem historischen Vergleich das bestmögliche Studiendesign dar [vgl.: Abbildung 6: Bewertung erwartbarer und generierbarer Evidenz].

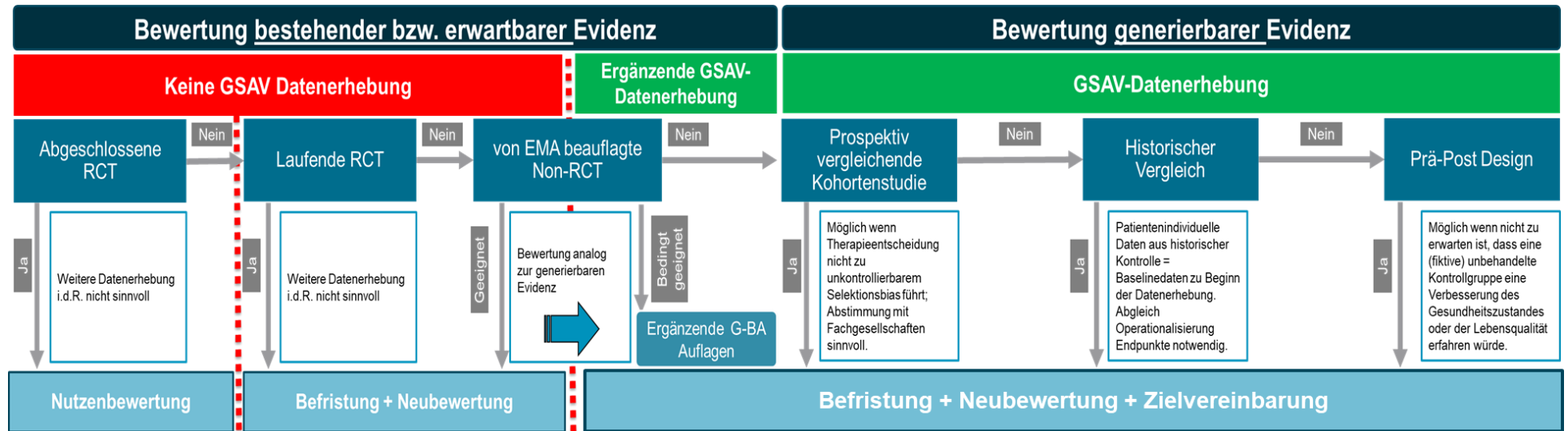


Abbildung 6: Bewertung erwartbarer und generierbarer Evidenz

Quelle 9. inav GmbH (2019)

Maßgeblich für einen belastbaren historischen Vergleich ist das Vorliegen von Daten, die den natürlichen Krankheitsverlauf bzw. den Krankheitsverlauf unter Standard of Care-Bedingungen abbilden. Hierzu ist im Vorfeld der Beauftragung ein Abgleich mit bestehenden Registern bzw. eine Studienrecherche vorzunehmen.

Weiterhin ist es notwendig, den historischen Daten patientenindividuelle Daten entnehmen zu können, die für den Krankheitsverlauf prognostische Relevanz haben können. Sind diese Daten in ausreichendem Umfang verfügbar, kann mit einem geeigneten Adjustierungsverfahren eine Merkmalsähnlichkeit zwischen Interventionsarm und historischem Kontrollarm hergestellt werden. Die maßgeblichen patientenindividuellen Daten müssen dann als Baselinedaten zu Beginn der anwendungsbegleitenden Datenerhebung erfasst werden.

Zugleich muss eine vergleichbare Operationalisierung der Endpunkte zwischen der historischen Kontrollgruppe und der Interventionsgruppe sichergestellt werden.

Denkbar ist auch - unter Verwendung bzw. Erfassung der Ein- und Ausschlusskriterien der klinischen Studie - ein Register vor Erteilung der Zulassung zu beginnen, um auf diese Weise einen idealen historischen Vergleichsarm zu erzielen. Voraussetzung hierfür ist eine frühzeitige Entscheidung, ob eine anwendungsbegleitende Datenerhebung für das betreffende Arzneimittel vorgesehen ist. Limitiert ist dieser Ansatz dadurch, dass gerade bei seltenen Erkrankungen, oftmals ein Großteil der in Frage kommenden Patienten bereits im Rahmen klinischer Studien behandelt wird.

Ist ein historischer Vergleich nicht sinnvoll durchführbar soll geprüft werden, ob anhand eines Prä-Post Designs die benötigte Evidenz generierbar ist.

8 HANDLUNGSEMPFEHLUNG

Mit dem GSAV wurde der G-BA für Orphan-Drugs und Arzneimittel mit bedingter Zulassung bzw. mit Zulassung unter ungewöhnlichen Umständen ermächtigt, den pharmazeutischen Unternehmer aufzufordern zum Zweck der Nutzenbewertung anwendungsbegleitende Daten zu erheben und auszuwerten.

Nach dem Willen des Gesetzgebers ist für diese Arzneimittel im Rahmen der anwendungsbegleitenden Datenerhebung die Berücksichtigung von nicht-randomisierten Daten zur Quantifizierung eines Zusatznutzens vorgesehen. Unter Berücksichtigung dieser Zielsetzung muss die Datenerhebung zu einer Verbesserung der Evidenzlage beitragen können. Damit ist eine Datenerhebung nur dann als vertretbar anzusehen, wenn erwartet werden kann, dass diese Daten grundsätzlich für die Quantifizierung eines Zusatznutzens herangezogen werden können.

Die von der Gesetzgebung umfassten Arzneimittel mit „besonderen“ Zulassungen eint, dass die üblicherweise im Nutzenbewertungsverfahren verlangte Evidenz aus RCT nicht oder nur eingeschränkt zur Verfügung steht und / oder mit der Zulassung bereits ein Zusatznutzen gegenüber bisherigen Therapieoptionen festgestellt wurde.

Basierend auf diesen bereits erfolgten Feststellungen der Zulassungsbehörde ermöglicht die vom GSAV geforderte Berücksichtigung nicht-randomisierter Daten zur Quantifizierung eines Zusatznutzens die nach der AM-NutzenV erforderliche Kongruenz zwischen Nutzenbewertung und Feststellungen der Zulassungsbehörde. Die Akzeptanz der damit einhergehenden gegenüber RCT niedrigeren Ergebnissicherheit auf Studienebene entspricht der Intention des Gesetzgebers und spricht demnach nicht grundsätzlich gegen die Verwendung derartiger Studien. Mögliche Fehlerquellen sind durch Anwendung spezieller Methoden möglichst gering zu halten. Zugleich ist bestmöglich abzuschätzen, in welcher Größenordnung verbleibende verzerrende Faktoren das Studienergebnis beeinflussen.

Vor diesem Hintergrund wird für die Generierung von Evidenz auf dem höchstmöglichen Niveau ein Stufenmodell unter Berücksichtigung der vorliegenden und erwartbaren Evidenz vorgeschlagen. Weiterhin werden die Eignung von Auswertungsmethoden sowie der Methoden der Evidenzbewertung diskutiert. Abschließend werden Empfehlungen zu notwendigen begleitenden beratenden Maßnahmen gegeben.

Identifizierung geeigneter Arten der Evidenzgenerierung

Liegen randomisierte vergleichende Studien (RCT) bereits vor oder ist im Rahmen der Zulassung bereits eine Berücksichtigung einer noch laufenden RCT vorgesehen, sollte diese - wie bereits in der Vergangenheit praktiziert - auch im Rahmen der Nutzenbewertung herangezogen werden. Eine anwendungsbegleitende Datenerhebung ist in diesen Fällen nicht sinnvoll, da die Nutzenbewertung auf Grundlage höherer Evidenzstufe durchgeführt werden kann.

Sind keine RCT vorhanden oder erwartbar, werden drei Evidenzformen mit abgestufter Evidenzstärke für eine mögliche Datenerhebung als grundsätzlich geeignet erachtet. Bei der Auswahl der geeignetsten Evidenzform sind begleitende kontextuelle bzw. methodische Voraussetzungen zu berücksichtigen.

a. Prospektiv vergleichende Kohortenstudien

Die Berücksichtigung von Studiendaten niedrigerer Evidenzstufen ist bereits in der AM-NutzenV vorgesehen, wenn es „unmöglich oder unangemessen ist, Studien höchster Evidenzstufe durchzuführen oder zu fordern“.

In der dort definierten Reihenfolge heranzuziehender Evidenzstufen sind unterhalb der RCT prospektiv vergleichende Kohortenstudien vorgesehen. Derartige Studien wären beispielsweise aufbauend auf einem indikationsbezogenen Register realisierbar.

Die sinnhafte Umsetzbarkeit derartiger Studien muss allerdings für die vom Gesetzesvorhaben umfassten Arzneimittel aus zwei Gründen besonders geprüft werden.

Zum einen hat der G-BA ausschließlich für wirkstoffbezogene Register die Möglichkeit, eine umfassende Umsetzung über die Kopplung der Datenerhebung an die Verordnungsfähigkeit zu gewährleisten. Krankheitsbezogene Register können vom G-BA zwar gefordert werden, die Vollständigkeit des Registers allerdings möglicherweise nicht sichergestellt werden.

Zum anderen schließen die betroffenen Arzneimittel bestehende Versorgungslücken bzw. weisen gegenüber bestehenden Therapieoptionen einen erheblichen Nutzen auf. Vor diesem Hintergrund wäre in einer prospektiv vergleichenden Kohortenstudie ein bedeutsamer Selektionsbias zu erwarten. So werden sich Patienten, die die neue Therapie erhalten und Patienten, bei denen keine oder die bislang verfügbaren Therapieoptionen angewendet werden, z. B. im Schweregrad der Erkrankung voneinander unterscheiden. Nur wenn dieser Selektionsbias kontrollierbar erscheint, kann auf Basis dieses Studiendesigns ein sinnhafter Vergleich vorgenommen werden.

Demnach bieten prospektiv vergleichende Kohortenstudien die höchste Evidenzstärke und sind dann zu wählen, wenn Verzerrungsaspekte vertretbar kontrollierbar sind und Therapieentscheidungen in der Versorgung nicht zu einem unkontrollierbaren Selektionsbias führen.

b. Historische Vergleiche

Treffen die Voraussetzungen für eine prospektiv vergleichende Kohortenstudie nicht zu, sollte geprüft werden, ob für die anwendungsbegleitende Datenerhebung ein hinreichend vollständiges oder repräsentatives, produktbezogenes Register verwendet werden kann, das mit einer historischen Kontrollgruppe verglichen wird. Hierzu sollte im Vorfeld des Studienbeginns mit systematischen Recherchen nach in Frage kommenden historische Kontrollgruppen recherchiert werden. Voraussetzung für die Eignung ist die Verfügbarkeit von historischen Daten auf patientenindividueller Ebene für relevante und übereinstimmend operationalisierte Endpunkte in Hinblick auf maßgebliche Störgrößen (z. B. prognostische, krankheitsspezifische, demografische Faktoren). Durch Erfassung dieser patientenindividuellen Daten ist bei der Auswertung eine adäquate Adjustierung für Störgrößen möglich. Sämtliche verfügbaren, für ein Matchingverfahren geeignete patientenindividuellen Daten sollten zugleich als Baseline-daten zu Beginn des Registers erfasst werden. Auf diese Weise ist bei der Auswertung die bestmögliche Adjustierung zu erzielen.

c. Prä-Post-Vergleiche

Sind die Voraussetzungen für einen historischen Vergleich nicht gegeben, sollte die Möglichkeit eines Prä-Post-Vergleiches geprüft werden. Ein solcher Vergleich ist dann sinnvoll, wenn für eine (fiktive) unbehandelte Kontrollgruppe keine Spontanverbesserung erwartbar ist. Wenn keine belastbare Kenntnis über den Krankheitsverlauf ohne die neue Therapieoption besteht, sollte hierzu durch Fachgesellschaften oder spezialisierte Behandler eine Einschätzung eingeholt werden.

Sind keine der o. g. Voraussetzungen erfüllt, sollte keine Datenerhebung beauftragt werden, da dann grundsätzlich nicht zu erwarten ist, dass die erhobenen Daten für die Quantifizierung eines Zusatznutzens verwendbar sind.

Bei der Auswahl der geeignetsten Evidenzform sind bereits bestehende Auflagen der EMA prioritär zu berücksichtigen. Dabei soll die geeignetste Evidenzform unter den Auflagen identifiziert werden und bei Notwendigkeit um weitere Aspekte der Datenerhebung ergänzt werden.

Auswertung anwendungsbegleitend erhobener Daten

Für die vom GSAV umfassten Arzneimittel ist davon auszugehen, dass die für eine anwendungsbegleitende Datenerhebung zur Verfügung stehende Anzahl von Patienten limitiert ist. Zugleich scheint es aus regulatorischer Perspektive sinnvoll, die Dauer der veranlassten Datengenerierung zu limitieren, um innerhalb einer angemessenen Zeitspanne die darauf aufbauende Nutzenbewertung vollziehen zu können. Somit stehen mit Fallzahl und Studiendauer die beiden bedeutsamsten Parameter bereits fest, die bei einer Studienplanung herangezogen werden, um eine Aussage mit dem üblicherweise akzeptiertem Fehlerniveau von 5 % ableiten zu können. Die Beibehaltung der Forderung nach einem 95 %igen Konfidenzintervall würde demnach bei begrenzten Patientenzahlen die gesetzlich intendierte Nachweismöglichkeit eines quantifizierbaren Zusatznutzen entweder konterkarieren oder diese nur innerhalb einer unerwünscht langen Zeitspanne ermöglichen. Vor diesem Hintergrund sollte das geforderte Fehlerniveau grundsätzlich in die Betrachtung geeigneter Auswertungsmethoden einbezogen werden und – in Abhängigkeit von der zu erwartenden Effektgröße –entsprechend den Erfordernissen des Einzelfalls angepasst werden.

Mit der im GSAV vorgesehenen Berücksichtigung von nicht-randomisierten Daten entfällt zugleich die Randomisierung, welche die Merkmalsähnlichkeit zwischen zu vergleichenden Studienarmen gewährleisten kann. Es stehen jedoch mehrere grundsätzlich geeignete Methoden zur Verfügung, mit denen eine ausreichende Merkmalsähnlichkeit über eine Adjustierung nach unterschiedlichen Patientencharakteristika erreicht werden kann. Gängige und international anerkannte Verfahren sind z. B. das Propensity Score Matching (PSM) oder der Matching-adjustierte indirekter Vergleich (MAIC). Grundlegende Voraussetzung für derartige Verfahren ist die Kenntnis und Berücksichtigung wesentlicher verzerrender Faktoren wie prognostische Faktoren, Krankheitscharakteristika oder demografische Faktoren. Anders als bei einer Randomisierung können mit diesen Adjustierungsverfahren unbekannte Störgrößen jedoch nicht erfasst werden. Damit ergibt sich objektiv eine geringere Ergebnissicherheit, deren Akzeptanz jedoch mit den Festlegungen des Gesetzgebers zur Berücksichtigung von nicht-randomisierten Daten, die auf den Zulassungsentscheidungen der EMA für diese besondere Gruppe von Arzneimittel aufbauen, intendiert ist. Grundsätzlich ermöglicht eine adäquate Adjustierung in dieser Situation valide Aussagen zum Zusatznutzen. Dabei sind auch keine weiteren Voraussetzungen an die zu erwartende Effektgröße erforderlich. Zwar kann die Größe des Effektes die Sicherheit der zu treffenden Aussage zusätzlich erhöhen. Das Vorliegen eines „dramatischen Effektes“ zur Ableitung eines Zusatznutzens auf Basis von nicht-randomisierten Daten wie in regulären Nutzenbewertungsverfahren ist jedoch nicht erforderlich und würde der gesetzlichen Intention der GSAV-Datenerhebung nicht gerecht werden.

Die Auswahl der geeigneten Adjustierungsverfahren sollte prospektiv erfolgen. Die Auswahl der Störfaktoren sollte auch unter Einbindung der wissenschaftlichen Fachgesellschaften oder spezialisierte Behandler bestimmt werden.

Evidenzbewertung

Für die Bewertung der anwendungsbegleitend erhobenen Evidenz bieten international führende Institutionen der evidenzbasierten Medizin geeignete Instrumente an. So stellt der GRADE-Ansatz ein international akzeptiertes Bewertungskonzept für Evidenzkörper dar, das die zugrunde liegenden Studien anhand definierte Kriterien auf- oder abwertet

und so die Qualität der Evidenz reliabel bewertet. Dabei ist die Erreichbarkeit niedriger, moderater und sogar hoher Evidenzlevels für nicht-randomisierte Daten vorgesehen. Bemerkenswert erscheint die Tatsache, dass der Beitrag einer aufgewerteten Beobachtungsstudie zum Evidenzkörper die gleiche Qualitätsstufe wie eine RCT erreichen kann. Das Instrument kann somit einen geeigneten Rahmen für die Anerkennung von anwendungsbegleitenden Daten für die Nutzenbewertung bieten.

Mit REQueST wurde im Rahmen von EUnetHTA ein Tool zur konsistenten Bewertung der Eignung von Registern für HTA entwickelt. Mit diesem Instrument lässt sich die Qualität bestehender Register bewerten. Zugleich stellt es Kriterien für ein neu aufbauendes, hochwertiges Register nachvollziehbar dar. Weitere Instrumente, wie das von der Cochrane Collaboration entwickelte ROBINS-I zur Bewertung von Verzerrungen nicht-randomisierter Studien können bei Bedarf ergänzend genutzt werden.

Vorgaben der G-BA-Beratungen

Der G-BA hat nach den Vorgaben des GSAV Anforderungen an Dauer, Art und Umfang von Datenerhebung und Auswertung festzulegen und dabei insbesondere Vorgaben zur Methodik sowie zu patientenrelevanten Endpunkten und deren Erfassung zu bestimmen. Für eine zielführende anwendungsbegleitende Datenerhebung sollte in diesem Rahmen sichergestellt werden, dass eine aufwändige Datenerhebung vermieden wird, die erwartbar nicht zur Quantifizierbarkeit eines Zusatznutzens beitragen kann.

Unabhängig von der gewählten Studienart sollten vor dem Hintergrund unterschiedlicher denkbarer Auswertungsmöglichkeiten die methodischen Prinzipien der Analyse der Daten im Vorfeld zwischen G BA und pharmazeutischem Unternehmer erörtert werden. Die wesentlichen Aspekte der Datengenerierung und -auswertung, inkl. der Fragestellung (PICO-Schema), Evidenzform, Störfaktoren und Adjustierungsmethoden, Stichprobenumfang, Dauer und akzeptabler statistischer Sicherheit sollten hierzu vorab festgelegt werden.

Diese Beratungen sollten unter frühestmöglicher Nutzung des Dialogs mit der EMA erfolgen und verbindlichen Charakter haben. Der frühzeitige Dialog mit der Zulassungsbehörde sollte zugleich genutzt werden, eine Harmonisierung der Auflagen von Zulassung und G-BA anzustreben. Für die medizinische Einschätzung der Umsetzbarkeit der Anforderungen sollten auch die medizinischen Fachgesellschaften eingebunden werden und Vorgaben sollten im Einvernehmen mit den Zulassungsbehörden erfolgen.

9 LITERATURVERZEICHNIS

[1]	Die Kommission der europäischen Gemeinschaften. (2003). Richtlinie 2003/63/EG der Kommission vom 25.Juni 2003 zur Änderung der Richtlinie 2001/83/EG des Europäischen Parlaments und des Rates zur Schaffung eines Gemeinschaftskodexes für Humanarzneimittel. Brüssel; 2003.
[2]	European Parliament and the Council of the European Union. (2000). Regulation (EC) No 141/2000 of the European Parliament and of the Council of 16 December 1999 on orphan medicinal products. Official J Eur Communities, 43, L18.
[3]	European Medicines Agency. (2010). Recommendation on elements required to support the medical plausibility and assumption of significant benefit for an orphan designation. (EMA/COMP/15893/2009). London; 2010.
[4]	Deutscher Bundestag (2010). 17. Wahlperiode. Gesetzentwurf der Bundesregierung - Entwurf eines Gesetzes zur Neuordnung des Arzneimittelmarktes in der gesetzlichen Krankenversicherung (Arzneimittelmarktneuordnungsgesetz – AMNOG). Drucksache 17/3116. Berlin; 2010.
[5]	European Medicines Agency. (2006). Guideline on the Scientific Application and the Practical Arrangements Necessary to Implement Commission Regulation (EC) No 507/2006 on the Conditional Marketing Authorisation for Medicinal Products for Human Use Falling Within the Scope of Regulation (EC) No 726/2004 (EMA/509951/2006). London; 2006.
[6]	European Medicines Agency. (2005). Guideline on Procedures for The Granting of a Marketing Authorisation under Exceptional Circumstances, Pursuant to Article 14 (8) of Regulation (Ec) No 726/2004 (EMA/357981/2005) London; 2005.
[7]	Behring, A. (2018). Bewertung im Gesundheitswesen – Fluch oder Segen. Arzneimittel-Bewertungen aus Sicht des G-BA. Vortrag im Rahmen des Barmer Versorgungs-Forschungskongresses 2018. Berlin; 2018.
[8]	Gemeinsamer Bundesausschuss. (2018). Stellungnahme der unparteiischen Mitglieder des Gemeinsamen Bundesausschusses (G-BA) vom 14.12.2018 zur Anhörung des Bundesministeriums für Gesundheit zum Referentenentwurf für ein Gesetz für mehr Sicherheit in der Arzneimittelversorgung (GSAV). Berlin; 2018.
[9]	Gemeinsamer Bundesausschuss. Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V – Pasireotid. Berlin; 2012.
[10]	Gemeinsamer Bundesausschuss. (2017). Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Cerliponase

	alfa. Berlin; 2017.
[11]	Gemeinsamer Bundesausschuss. (2019). Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Axicabtagen-Ciloleucel. Berlin; 2019.
[12]	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. (2015). Pressemitteilung vom 18.12.2015. Herbst-Symposium 2015: Kontroverse Positionen zu "Real World Data" in Nutzenbewertungen. Online verfügbar unter: https://www.iqwig.de/de/presse/pressemitteilungen/2015/herbst-symposium-2015-vortraege-jetzt-online.7104.html [zuletzt abgerufen am 30.10.2019].
[13]	Windeler, J., Lauterberg, J., Wieseler, B., Sauerland, S., & Lange, S. (2017). Patientenregister für die Nutzenbewertung: Kein Ersatz für randomisierte Studien. <i>Deutsches Ärzteblatt</i> , 114, A783-6.
[14]	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. (2015). Stellungnahme des IQWiG zum Referentenentwurf des Bundesministeriums für Gesundheit. Entwurf eines Gesetzes für mehr Sicherheit in der Arzneimittelversorgung (GSAV) vom 14. November 2018. Köln; 2018.
[15]	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden. Version 5.0. vom 10.07.2017. Köln; 2017.
[16]	European Medicines Agency. (2015). Scientific guidance on post-authorisation efficacy studies. (EMA/PDCO/CAT/CMDh/PRAC/CHMP/261500/2015). London; 2015.
[17]	European Medicines Agency (2018) Discussion paper: Use of patient disease registries for regulatory purposes – methodological and operational considerations. London; 2018.
[18]	Balshem, H., Helfand, M., Schünemann, H. J., Oxman, A. D., Kunz, R., Brozek, J., ... & Guyatt, G. H. (2011). GRADE guidelines: 3. Rating the quality of evidence. <i>Journal of clinical epidemiology</i> , 64(4), 401-406.
[19]	European Network for Health Technology Assessment (2019). Vision paper on the sustainable availability of the proposed Registry Evaluation and Quality Standards Tool (REQueST). EUnetHTA JA3 WP5B2: Vision paper. Copenhagen; 2019.
[20]	European Network for Health Technology Assessment (2019). Request tool and vision paper. Online verfügbar unter: https://www.eunetha.eu/request-tool-and-its-vision-paper-are-now-available-for-public-consultation/ [zuletzt abgerufen am 16.09.2019].
[21]	Jüni, P., Loke, Y. K., Pigott, T. D., Ramsay, C. R., Regidor, D., Rothstein, H. R., ... & Shrier, I. (2016). Risk of bias in non-randomized studies of interventions (ROBINS-I): detailed guidance.
[22]	Stausber, J.; Maier, B.; Bestehorn, K. et al. (2019). Memorandum Register für die Versorgungsforschung:Update 2018.
[23]	Bundesministerium für Justiz und Verbraucherschutz. Bundesamt für Justiz. Gesetz über den Verkehr mit Arzneimitteln (Arzneimittelgesetz - AMG)

	§ 4 Sonstige Begriffsbestimmungen. Abs 23. Online verfügbar unter: https://www.gesetze-im-internet.de/amg_1976/__4.html [zuletzt abgerufen am 04.11.2019].
[24]	Röhrig, B., du Prel, J. B., Wachtlin, D., & Blettner, M. (2009). Studientypen in der medizinischen Forschung. <i>Deutsches Ärzteblatt</i> , 106(15), 262-268
[25]	Gemeinsamer Bundesausschuss. (2019). Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V Vestronidase alfa. Berlin; 2019.
[26]	Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. <i>Personality and individual differences</i> , 102, 74-78.
[27]	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. (2014). IQWiG-Berichte – Nr. 241. Bewertung und Auswertung von Studien bei seltenen Erkrankungen. Rapidreport. Köln; 2014.
[28]	Gemeinsamer Bundesausschuss. (2018). Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Elosulfase alfa (Neubewertung aufgrund neuer Wissenschaftlicher Erkenntnisse). Berlin; 2018.
[29]	Gemeinsamer Bundesausschuss. (2016). Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V- Vismodegib. Berlin; 2016.
[30]	Kuss, O., Blettner, M., & Börgermann, J. (2016). Propensity Score–eine alternative Methode zur Analyse von Therapieeffekten. <i>Dtsch Arztebl Int</i> , 113, 597-603.
[31]	Signorovitch, J. E., Sikirica, V., Erder, M. H., Xie, J., Lu, M., Hodgkins, P. S., ... & Wu, E. Q. (2012). Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. <i>Value in Health</i> , 15(6), 940-947.
[32]	Caro, J. J., & Ishak, K. J. (2010). No head-to-head trial? Simulate the missing arms. <i>Pharmacoeconomics</i> , 28(10), 957-967.
[33]	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. (2012). IQWiG-Berichte – Nr. 149. Axitinib –Nutzenbewertung gemäß § 35a SGB V. Dossierbewertung. Köln; 2012.
[34]	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. (2014). IQWiG-Berichte – Nr. 261. Daclatasvir –Nutzenbewertung gemäß § 35a SGB V. Dossierbewertung. Köln; 2014.
[35]	European Medicines Agency. (2017). Conditional marketing authorisation–report on ten years of experience at the European Medicines Agency. London; 2017.

