

Hans-Holger Bleß

Anwendungsbegleitende Daten in der Nutzenbewertung

Empfehlungen zur Evidenzgenerierung und -auswertung

Mit folgendem Konzept 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 relevanter Organisationen des Health Technology Assessments (HTA) sowie einer Übersicht zu in Frage kommenden methodischen Auswertungsverfahren werden Empfehlungen zu Evidenzgenerierung, Evidenzbewertung und erforderlichen G-BA-Vorgaben formuliert, die eine Berücksichtigung der Daten für die Quantifizierbarkeit eines Zusatznutzens ermöglichen.

>> 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.

Regulatorische Rahmenbedingungen

Ein Arzneimittel kann von der EMA den Orphan Drug Status zuerkannt bekommen, wenn es zum Einsatz bei lebensbedrohlichen, seltenen Erkrankungen eingesetzt wird, für die keine zufriedenstellende Therapieoption existiert. Bei vorhandenen Therapiemöglichkeiten muss das neue Arzneimittel einen erheblichen Nutzen aufweisen [2]. Der Begriff „erheblicher Nutzen“ ist dabei definiert als ein klinisch relevanter Vorteil oder ein bedeutender Beitrag zur Behandlung von Patienten [3]. 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].

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 insbesondere offene Fragen zu Qualität, Unbedenklichkeit und Wirksamkeit des Arzneimittels zu beantworten. Maßgeblich für diesen Zulassungsweg ist, dass der Nutzen der sofortigen Verfügbarkeit des Arzneimittels für die öffentliche Gesundheit die Risiken überwiegt, die sich aus der Tat-

Zusammenfassung

Hintergrund: Mit dem GSAV wird der G-BA ermächtigt für Orphan Drugs und Arzneimittel mit bedingter oder unter besonderen Bedingungen erteilter Zulassung anwendungsbegleitende Datenerhebungen und Auswertungen zum Zweck der Nutzenbewertung zu fordern. Hierbei sind nach Intention des Gesetzgebers auch nicht-randomisierte Studien und damit niedrigere Ergebnissicherheiten zu akzeptieren.

Methodik: Auf Grundlage einer Übersicht über den aktuellen Umgang mit nicht-randomisierten Studien seitens relevanter Organisationen sowie einer Übersicht zu methodischen Auswertungsverfahren werden Empfehlungen bezüglich Evidenzgenerierung und -auswertung formuliert, die der Zielsetzung dienen, verwertbare Daten für einen quantifizierbaren Zusatznutzen generieren zu können und damit die Evidenzlage zu verbessern.

Ergebnisse: In einem Stufenmodell zur Auswahl der geeigneten Art der Evidenzgenerierung werden prospektive vergleichende Kohortenstudien, historische Vergleiche oder Prä-Post-Vergleiche unter bestimmten Voraussetzungen für durchführbar und geeignet angesehen, wobei die von der EMA beauftragten Studien bei dieser Auswahl vorrangig genutzt werden sollten. International anerkannte Auswertungsverfahren der Confounder-Adjustierung werden als geeignet angesehen eine ausreichende Ergebnissicherheit für Aussagen zum Zusatznutzen liefern zu können, auch ohne weitere Anforderungen an die Effektgröße (insb. „dramatischer Effekt“). Bei statistischen Tests sollte zudem eine erhöhte Irrtumswahrscheinlichkeit akzeptiert werden, wenn mögliche Limitationen in der zu erwartenden Fallzahl in einer sinnvollen Studiendauer absehbar sind. International akzeptierte Methoden der Evidenzbewertung führender EbM-Akteure wie GRADE und Cochrane können einen geeigneten Rahmen für die Anerkennung der Daten bieten.

Konklusion: Die Intention des Gesetzgebers, im Rahmen der anwendungsbegleitenden Datenerhebung auch nicht-randomisierte Studien zur Quantifizierung eines Zusatznutzens heranzuziehen ist mit dem vorgeschlagenen Konzept umsetzbar. Es setzt einen möglichen Rahmen in dem Daten und Auswertung für einen quantifizierbaren Zusatznutzen verwertbar sind, um die Evidenzlage zu verbessern. Eine verbindliche Beratung bezüglich aller wesentlichen Aspekte der Datengenerierung und -auswertung ist jedoch erforderlich, damit sichergestellt werden kann, dass eine aufwändige Datenerhebung vermieden wird, die erwartbar nicht zu einem (quantifizierbaren) Zusatznutzen beitragen kann.

Schlüsselwörter

Anwendungsbegleitende Datenerhebung, Nutzenbewertung, GSAV, nicht-randomisierte Studien

Crossref/doi

doi: 10.24945/MVF.01.20.1866-0533.2202

sache ergeben, dass noch zusätzliche Daten erforderlich sind [5].

Arzneimittel, für die umfassende Daten aufgrund gesetzlicher oder ethischer Gründe nicht bereitgestellt werden können, kommen für eine Zulassung unter außergewöhnlichen Umständen in Frage. Diese Art der Zulassung wird mit Auflagen erteilt und das Nutzen-Risiko-Verhältnis jährlich neu bewertet, führt allerdings normalerweise

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 Spannungspotenzial 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-randomisierter Studiendaten.

se nicht zum Abschluss eines vollständigen Zulassungsdossiers [6].

Praktischer Umgang mit nicht-randomisierten Daten

EMA akzeptiert nicht-randomisierte Daten nach Zulassung unter Vorbedingungen

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.

In Rahmen von Wirksamkeitsstudien nach Markteintritt (Post-authorisation efficacy studies, PAES) 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 [7].

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.

Weiterhin sind auch Prä-Post-Vergleiche möglich, bei denen ein Patient seine eigene Kontrolle für Vergleiche vor und nach der Behandlung ist.

Für anwendungsbegleitende Register empfiehlt die EMA die Anbindung an bestehende Krankheitsregister, um „eine fortlaufende Bewertung der Krankheitsergebnisse und einen Vergleich verschiedener Behandlungsoptionen nach einer ähnlichen Methodik“ zu ermöglichen [7]. In seinem „Discussion paper registries“ verdeutlicht die EMA die Bevorzugung krankheitsbezogener gegenüber produktbezogenen Registern [8]. Als wesentlichen Punkt sieht die EMA die Notwendigkeit der Vermeidung und Kontrolle eines Selektionsbias. 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 [8].

Berücksichtigung von Non-RCT in besonderen Fällen entspricht G-BA Spruchpraxis

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:

I. Generierung von vergleichender Evidenz muss gewährleistet sein

- Mit indikationsspezifischen Registern werden auch Daten alternativer behandelter Patienten erfasst
- Beginn der Datenerhebung vor dem geplanten Marktzugang (Kontrollgruppe)

II. Erhebung von Krankheitssymptomatik und Lebensqualität ist essentiell

III. Register sollen mindestens repräsentative Stichprobe der GKV-Population untersuchen, bei Orphan-Drugs in der Regel Vollerhebung

IV. Kompatibilität mit bestehenden Registern (Vermeidung von Doppelerfassungen) [9]

In seiner Stellungnahme zum GSAV 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“ [10]. Zugleich fordert er, dass auch kontrollierte Studien von den anwendungsbegleitenden Datenerhebungen mit umfasst sein sollten.

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

IQWiG hält GSAV-Datenerhebung zur Bewertung des Zusatznutzens für ungeeignet

In der Stellungnahme des IQWiG zum GSAV-Referentenentwurf schreibt das Institut bereits in der Einleitung, dass es die anwendungsbegleitende Datenerhebung in der geplanten Form für ungeeignet hält [11]. Das IQWiG hebt hervor, dass zur Behebung von Erkenntnisdefiziten bei der Nutzenbewertung vorrangig direkt vergleichende, randomisierte, kontrollierte Studien Verwendung finden müssten. 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 Produktregistern werden „vergleichende Aussagen als Grundlage der Bewertung eines Zusatznutzens nicht möglich sein“ [11].

In seinem aktuellen Methodenpapier 5.0 führt das Institut aus, dass „andere Studientypen als RCTs [...] in der Regel für einen Kausalitätsnachweis nicht geeignet“ sind [12]. 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. besonderer Voraussetzungen und spezieller Qualitätsanforderungen“ [12]. 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 [12].

Erreichbarkeit hoher Evidenzlevel für Beobachtungsstudien im GRADE-Ansatz möglich

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. 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 [13].

Cochrane-Tool (ROBINS-I) eignet sich zur Bewertung von Non-RCT

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 [14]. 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 Verzerrungspotenzial zu bewerten. Zur Identifizierung potentieller Störgrößen wird die Zusammenarbeit mit Fachexperten und die Verwendung literaturbasierter Wissens empfohlen. 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 Verzerrungspotenzial in einer qualitativ hochwertigen randomisierten Studie.

EUnetHTA definiert einheitliche Anforderungen an Register und ermöglicht deren Bewertung

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 [15]. Das Tool ermöglicht eine konsistente Bewertung der Eignung von Registern für HTA und adressiert zugleich Bedenken hinsichtlich der Zuverlässigkeit der Registrierungsdaten für diese Zwecke. Das Tool ist geeignet, Register auf seine Eignung hin zu bewerten und stellt zugleich die Kriterien für ein neu aufbauendes, hochwertiges Register nachvollziehbar dar [15].

Mögliche Evidenzarten

Gegen die mit dem GSAV geforderte Berücksichtigung nicht-randomisierter Studien 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 derartiger 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.

Prospektive vergleichende Kohortenstudien grundsätzlich für Nutzenbewertung geeignet

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, ermöglicht grundsätzlich Aussagen zu kausalen Zusammenhängen und stellt daher 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.

Die Vollständigkeit produktbezogener Register kann über die gesetzlich vorgesehene Verordnungseinschränkung weitgehend sichergestellt werden, auch wenn eine patientenseitige 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.

Einbeziehung auch nicht vergleichender Studien notwendig

Die EMA erteilt, in den von der neuen Gesetzgebung umfassten Sondersituationen auch Zulassungen, die auf einarmigen Studien basieren. Die nach dem Methodenpapier des IQWiG, wie auch der Verfahrensordnung des G-BA grundsätzlich verwendbaren adjustierten indirekten Vergleiche über Brückenkompaktoren sind auf dieser Datenbasis nicht möglich. Zudem betreffen die vom Gesetzesvorhaben umfassten Besonderheiten in der Zulassung 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.

Vergleiche auf Basis einarmiger Studien möglich

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. 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 der Operationalisierung der relevanten Endpunkte gewährleistet sein.

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 Spruchpraxis des G-BA zeigt die grundsätzliche Verwendbarkeit derartiger Vergleiche [16].

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.

Mögliche Auswertungsmethoden

Aussagen zum Zusatznutzen sind mit geeigneten Adjustierungsverfahren möglich

Für die Gewährleistung der Merkmalsähnlichkeit der Vergleichsgruppen bei einer prospektiven Kohortenstudie oder einem historischen Vergleich werden Verfahren benötigt, die eine Adjustierung nach verzerrenden Faktoren ermöglichen. 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.

Es stehen mehrere grundsätzlich geeignete Methoden zur Verfügung, mit denen eine ausreichende Merkmalsähnlichkeit über eine Adjustierung nach unterschiedlichen Patientencharakteristika erreicht werden kann. Gängige Verfahren sind Matching-adjustierte indirekte Vergleiche (MAIC) oder das Propensity-Score-Matching (PSM), bei denen Patientencharakteristika des Interventionsarmes durch Gewichtung denen des Kontrollarms angepasst werden. Seltener verwendet werden Verfahren wie der Simulierte Therapievergleich (Simulated Treatment Comparison, STC) oder die Bayes-Benchmarking Analysis (BBA). Trotz Anwendung eines geeigneten Adjustierungsverfahren ergibt sich objektiv eine geringere Ergebnissicherheit gegenüber einer randomisierten Studie. Deren Akzeptanz ist jedoch mit den Festlegungen des Gesetzgebers zur Berück-

Literatur

- [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] European Medicines Agency. (2015). Scientific guidance on post-authorisation efficacy studies. (EMA/PDCO/CAT/CMDh/PRAC/CHMP/261500/2015). London; 2015.
- [8] European Medicines Agency (2018) Discussion paper: Use of patient disease registries for regulatory purposes – methodological and operational considerations. London; 2018.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden. Version 5.0. vom 10.07.2017. Köln; 2017.
- [13] 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. Journal of clinical epidemiology, 64(4), 401-406.
- [14] 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.
- [15] 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. Kopenhagen; 2019.
- [16] 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.
- [17] Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. (2014). IQWiG-Berichte – Nr. 241. Bewertung und Auswertung von Studien bei seltenen Erkrankungen. Rapidreport. Köln; 2014.
- [18] 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 - Elosulfate alfa (Neubewertung aufgrund neuer wissenschaftlicher Erkenntnisse). Berlin; 2018.
- [19] 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.

sichtigung von nicht-randomisierten Daten intendiert und darin begründet, dass die vorzunehmenden Nutzenbewertungen auf den Zulassungsentscheidungen der EMA für diese besondere Gruppe von Arzneimitteln aufbauen. Grundsätzlich ermöglicht eine adäquate Adjustierung in dieser Situation valide Aussagen zum Zusatznutzen.

Anpassung der Irrtumswahrscheinlichkeit kann bei geringen Patientenzahlen notwendig werden

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.

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 Fehlerniveau 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.“ [17].

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.

Erforderlichkeit eines „dramatischen Effekts“ entspricht nicht der Intention des Gesetzgebers

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 gemeinsamen 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.

Zu den Anforderungen an die Verwendbarkeit historischer Vergleiche führt der G-BA aus: „Aufgrund der per se geringen Ergebnissicherheit 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“ [18]. Beispiele für die Anerkennung historischer Vergleiche finden sich in der Spruchpraxis des G-BA und bestätigen deren grundsätzliche Eignung [19].

Handlungsempfehlungen zur Evidenzgenerierung und Auswertung

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. Es ist nicht davon auszugehen, dass die von der neuen Gesetzgebung umfassten Arzneimittel mit einem generischen Konzept bewertbar

werden. Vielmehr wird für verschiedene Situationen spezifisch nach einem geeigneten Vorgehen gesucht werden müssen, die erforderliche Evidenz im Rahmen des Möglichen zu generieren.

Evidenzgenerierung

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.

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, wobei es sich um unterschiedliche Studientypen handeln kann.

Es 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 innerhalb des deutschen Versorgungskontextes wirksam umsetzen ließe.

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. 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 Koppelung 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 Ba-

sis dieses Studiendesigns ein sinnhafter Vergleich vorgenommen werden.

Ist eine prospektiv vergleichende Kohortenstudie nicht sinnvoll durchführbar, stellen einarmige Studien mit einem historischen Vergleich das bestmögliche Studiendesign dar.

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 Baseline-Daten 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.

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.

Unter Berücksichtigung der Zielsetzung des GSAV 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. 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.

Evidenzbewertung und Auswertungsmethoden

Für die vom GSAV umfassten Arzneimittel ist davon auszugehen, dass die für eine anwendungsbegleitende Datenerhebung zur Verfügung stehende Anzahl von Patienten limitiert ist. Zugleich

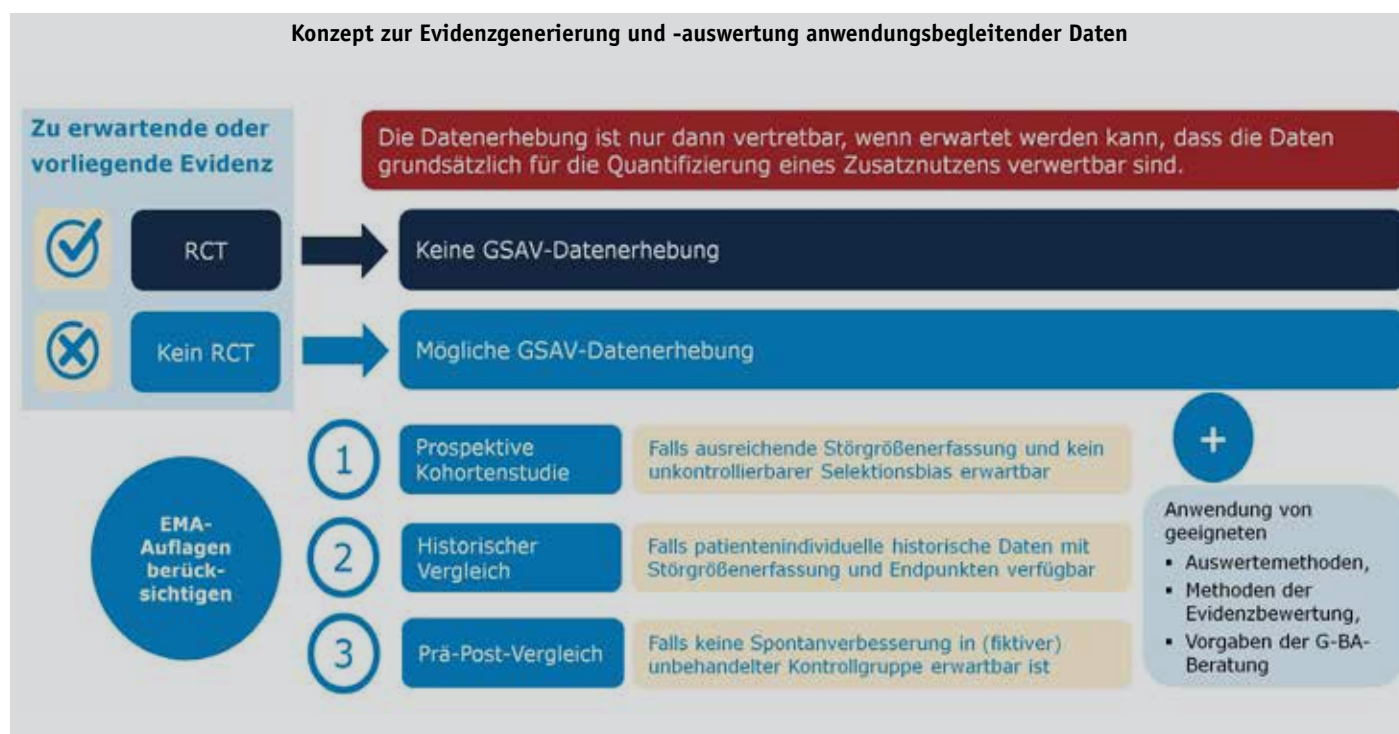


Abb. 1: Konzept zur Evidenzgenerierung und -auswertung anwendungsbegleitender Daten. Quelle: inav (2019).

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 Zusatznutzens 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 indirekte Vergleich (MAIC). Grundlegende Voraussetzung für derartige Verfahren ist die Kenntnis und Berücksichtigung wesentlicher verzerrender Faktoren. Anders als bei einer Randomisierung können mit diesen Adjustierungsverfahren unbekannte Störgrößen jedoch nicht erfasst werden. Daraus resultiert eine geringere Ergebnissicherheit, deren Akzeptanz jedoch mit den Festlegungen des Gesetzgebers zur Berücksichtigung von nicht-randomisierten Daten, die auf den Zulas-

sungsentscheidungen der EMA für diese besondere Gruppe von Arzneimitteln 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.

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. <<

Using application-accompanying data for benefit assessments – Recommendations for evidence generation and evaluation

Background: The law for higher safety in the supply of medicines (GSAV) authorizes the G-BA to request application-accompanying data collection and evaluations for the purpose of benefit assessments for orphan drugs, products with conditional authorization and drugs authorized under exceptional circumstances. In accordance with the intention of the legislator, non-randomized studies and thus lower certainty of results has to be accepted.

Methodology: Based on an overview of the current handling of non-randomized studies by relevant organizations and an overview of methodological evaluation methods, recommendations are formulated with regard to the generation and evaluation of evidence, which serve the purpose of being able to generate usable data for a quantifiable additional benefit and thus the evidence base improve.

Results: In a step-by-step model for selecting the appropriate type of evidence generation, prospective comparative cohort studies, historical comparisons or pre-post comparisons are considered feasible and suitable under certain conditions, whereby the studies commissioned by the EMA should be given priority in this selection. Internationally recognized evaluation methods of confounder adjustment are considered suitable to be able to provide a sufficient certainty of results for statements about the additional benefit, even without further requirements regarding the effect size (esp. „dramatic effect“). In the case of statistical tests, an increased probability of error should also be accepted if possible limitations in the expected number of cases can be foreseen within a reasonable period of study. Internationally accepted methods of evidence assessment by leading EbM players such as GRADE and Cochrane can provide a suitable framework for the recognition of the data.

Conclusion: The intention of the legislator to use non-randomized studies to quantify an additional benefit in the context of the application-related data collection can be implemented with the proposed concept. It sets a possible framework in which data and evaluation can be used for a quantifiable additional benefit in order to improve the evidence base. Binding advice regarding all essential aspects of data generation and evaluation is required, however, so that it can be ensured that time-consuming data collection is avoided that cannot be expected to contribute to a (quantifiable) additional benefit.

Keywords

application-accompanying data collection, benefit assessment, GSAV, non-randomized studies

Autorenerklärung

Hans-Holger Bleß ist Mitarbeiter des Unternehmens inav GmbH. Die Analyse erfolgte im Auftrag und mit finanzieller Unterstützung des vfa.

Zitationshinweis

Bleß, H.: „Anwendungsbegleitende Daten in der Nutzenbewertung – Empfehlungen zur Evidenzgenerierung und -auswertung“, in: „Monitor Versorgungsforschung“ (01/20), S. 47-54, doi: 10.24945/MVF.01.20.1866-0533.2202

Hans-Holger Bleß

ist seit Januar 2019 Mitinhaber und Geschäftsführer des privaten Instituts für angewandte Versorgungsforschung (inav GmbH). Seine Arbeitsschwerpunkte bestehen in der strategischen Planung für den Markteintritt neuer Arzneimittel, der wissenschaftlichen und gesundheitspolitischen Begleitung der Nutzenbewertung. Er studierte Pharmazie an der Freien Universität Berlin.
Kontakt: bless@inav-berlin.de

