

Kurzfassung

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

Fragestellung

Ziele der vorliegenden Ausarbeitung sind

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

Methoden

Informationsbeschaffung und -bewertung

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

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

Ermittlung von Qualitätskriterien für Patientenregister

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

Ergebnisse

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

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

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

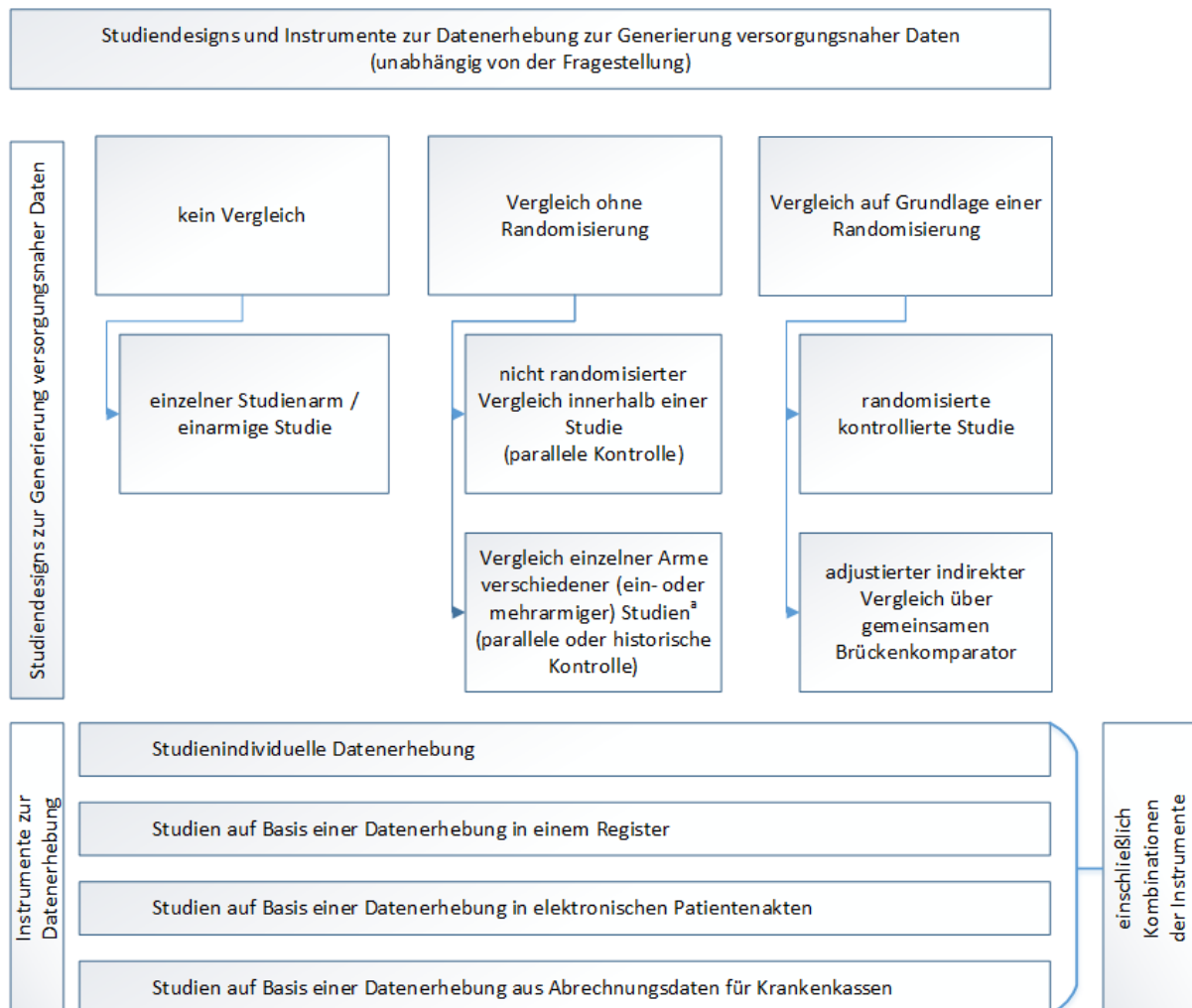
Die Definition versorgungsnaher Daten impliziert weder ein bestimmtes Studiendesign noch ein bestimmtes Instrument der Datenerhebung.

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

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

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

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



a: einschließlich Studien zum Spontanverlauf

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

Versorgungsnahe Daten in der Nutzenbewertung

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

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

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

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

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

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

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

Wahl des Designs und Studienplanung

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

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

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

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

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

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

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

Instrumente zur Datenerhebung

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

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

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

Studien auf Basis einer Datenerhebung in einem Register

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

Anforderungen an die Datenqualität

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

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

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

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

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

Interviews mit Registerbetreibern

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

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

Anforderungen an die Auswertung

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

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

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

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

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

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

Anforderungen an die Berichterstattung

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

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

Optimierte Studien für die Entscheidungsfindung im Gesundheitswesen

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

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

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

Definition der Fragestellung

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

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

Evidenzlücken in Nutzenbewertungen von Arzneimitteln (Orphan Drugs)

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

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

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

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

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

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

Mögliche Prozessschritte der anwendungsbegleitenden Datenerhebung

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

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

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

Fazit

Studiendesign und Datenerhebung

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

Versorgungsnahe vergleichende Studien ohne Randomisierung

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

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

Versorgungsnahe vergleichende Studien mit Randomisierung

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

Anwendungsbegleitende Erhebung versorgungsnaher Daten

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