

Kurzfassung

Der Gemeinsame Bundesausschuss (G-BA) hat am 28.01.2025 das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) mit einer wissenschaftlichen Ausarbeitung zur Methodik im Bereich der anwendungsbegleitenden Datenerhebung (AbD) mit einer Fokussierung auf eine Bewertung von Themenbereichen beauftragt.

Fragestellung

Ziel des vorliegenden Berichts ist eine wissenschaftliche Ausarbeitung zur Methodik (Generierung und Auswertung versorgungsnaher Daten zum Zwecke der Nutzenbewertung) im Bereich der AbD mit einer Fokussierung auf eine Bewertung der folgenden Themenbereiche, jeweils im Kontext nicht randomisierter vergleichender Studien:

- Confounderidentifikation und -auswahl
- Schätzung notwendiger Fallzahlen vor dem Hintergrund unzureichender Vorabinformationen
- Beobachtungsstart
- Beobachtungswechsel
- patientenberichtete Endpunkte (PROs)
- fehlende Werte
- Propensity-Score-Analysen in Anwendungsgebieten mit kleinen Patientenkollektiven

Dabei sollen eine orientierende Recherche nach geeigneter Literatur für die oben genannten Themenbereiche und eine Sichtung der in den bisherigen Prüfungen von Studienprotokoll und statistischem Analyseplan (SAP) identifizierten Mängel (ggf. mit Ergänzung der genannten Themenbereiche) erfolgen.

Für eine AbD eignen sich primär bereits bestehende Register als potenzielle Datenquelle. Die Ausarbeitung fokussiert daher auch auf Patientenregister als Instrumente zur Datenerhebung mit dem Ziel der Generierung versorgungsnaher Daten, da andere Datenquellen (z. B. GKV-Routinedaten) zum aktuellen Zeitpunkt nicht geeignet erscheinen, die Anforderungen an die Daten(qualität) im Rahmen der Nutzenbewertung zu erfüllen.

Methoden

Sichtung der bisherigen Prüfungen von Studienprotokoll und statistischem Analyseplan

Die Änderungsbedarfe, die in bisherigen Prüfungen der von den pUs eingereichten Studienprotokolle und SAPs adressiert wurden, wurden gesichtet. Auf Basis dieser Sichtung sollten

die beauftragten Themenbereiche, sofern erforderlich, um weitere Bereiche ergänzt werden. Da alle identifizierten Änderungsbedarfe einem der 7 Themenbereiche zugeordnet werden konnten, wurden im Rahmen des Auftrages keine weiteren Themenbereiche bearbeitet.

Erarbeiteter Vorschlag für die Anforderung an eine systematische Confounderidentifikation

Für den Themenbereich Confounderidentifikation und -auswahl wurde für den vorliegenden Auftrag auf das IQWiG-Arbeitspapier GA23-02 zurückgegriffen (veröffentlicht am 18.06.2025). Es fand somit keine erneute Erarbeitung der Methodik statt. Die Darstellung im vorliegenden Bericht beschränkt sich im Wesentlichen auf die abgeleiteten Handlungsempfehlungen zur systematischen Confounderidentifikation für nicht randomisierte vergleichende Studien. Für weiterführende Informationen wird auf das Arbeitspapier GA23-02 verwiesen.

Informationsbeschaffung

Für die weiteren 6 Themenbereiche umfasste die Informationsbeschaffung zu methodischen Dokumenten und Übersichtsarbeiten eine orientierende Recherche in PubMed, eine Website-Suche bei Zulassungsbehörden und anderen HTA-Agenturen sowie eine gezielte Websuche. Zusätzlich wurden ausgehend von den zuvor identifizierten methodischen Dokumenten und Übersichtsarbeiten Referenzlisten gesichtet sowie die „Similar Articles“-Funktion in PubMed und die „Cited-by“-Funktion in OpenAlex verwendet.

Bewertung ausgewählter Aspekte zur Generierung versorgungsnaher Daten im Kontext nicht randomisierter vergleichender Studien

Confounderidentifikation und -auswahl

In dem Arbeitspapier GA23-02 des IQWiG wurde gemäß dem Ansatz von Pufulete 2022 eine systematische Confounderidentifikation in der Indikation schubförmig remittierende multiple Sklerose (RRMS) an einem geeigneten Vergleich zweier medikamentöser Therapien mit Hinblick auf wesentliche patientenrelevante Endpunkte durchgeführt. Hierfür wurde eine Arbeitsgruppe externer Sachverständiger beauftragt, die ihre Ergebnisse in einem Bericht dargestellt hat. Auf Basis dieses Ergebnisberichts wurden im Rahmen der Bearbeitung von GA23-02 anhand weiterführender Untersuchungen durch das IQWiG die folgenden Handlungsempfehlungen für eine systematische Confounderidentifikation für nicht randomisierte vergleichende Studien (und somit auch für die AbD) abgeleitet:

1) Identifikation

Potenzielle Confoundervariablen sollen aus einer systematischen Literaturrecherche und aus Interviews mit klinischen Expertinnen und Experten identifiziert werden.

2) Zusammenführung

Confoundervariablen, die ein inhaltlich gleiches Konstrukt messen oder sich inhaltlich überlappen, sollen zusammengeführt werden. Für jeden dieser potenziellen Confounder ist eine eindeutige Operationalisierung festzulegen.

3) Relevanzbewertung

Jeder potenzielle Confounder sollte über eine Einbindung von klinischen Expertinnen und Experten in seiner Bedeutung bewertet werden. Das Ergebnis kann für eine inhaltlich gestützte Selektion von Confoundervariablen verwendet werden.

Das Verfahren der systematischen Confounderidentifikation ist trotz der im Arbeitspapier GA23-02 beschriebenen Möglichkeiten zur Aufwandseinsparung ressourcenintensiv. Wichtig ist an dieser Stelle jedoch zwischen dem Aufwand im Vorfeld einer AbD (für die Identifikation und Verdichtung potenzieller Confounder) und dem Gesamtaufwand für die AbD (darunter fällt u. a. der Aufwand für die Datenerhebung und für die Dokumentation) zu differenzieren. Es kann zielführend sein, im Vorfeld einen höheren Aufwand zu investieren, um den Aufwand für die AbD in Gänze zu reduzieren. In einem nächsten Schritt soll im Rahmen der Erstellung des Arbeitspapiers GA25-02 geprüft werden, ob weitere Maßnahmen zur Aufwandsreduktion bei der systematischen Identifikation und Erhebung von Confoundern möglich sind. Dabei sind potenzielle Optionen zur Aufwandsreduktion dem damit einhergehenden Informationsverlust und dem potenziellen Verlust der Ergebnissicherheit gegenüberzustellen.

Schätzung notwendiger Fallzahlen vor dem Hintergrund unzureichender Vorabinformationen

Der G-BA beauftragt das IQWiG regelhaft mit einer wissenschaftlichen Ausarbeitung von Konzepten für AbDs und deren Auswertung. Die Konzepte sollen auch die Anforderungen an die Dauer und den Umfang der jeweiligen Datenerhebung beschreiben. Dabei geht es um die Beantwortung der Frage, ob die geplante AbD prinzipiell realisierbar und sinnvoll durchführbar ist. Zu diesem Zweck wird keine Fallzahlplanung durchgeführt, da die hierfür notwendige Informationsbasis nicht vorliegt. Die Konzepterstellung verfolgt vielmehr das Ziel, eine Einschätzung zur grundsätzlichen Machbarkeit einer AbD zu geben. Je nach Informationsbasis zum Zeitpunkt der Konzepterstellung wird dabei entweder auf eine orientierende Fallzahl-schätzung oder eine orientierende Betrachtung von Fallzahl-szenarien als Herangehensweise zurückgegriffen.

Parameter für die Fallzahlschätzung

Um im Rahmen der orientierenden Fallzahlschätzung die notwendige Fallzahl zu ermitteln, müssen Annahmen für mehrere Parameter getroffen werden. Zu diesen Parametern gehören u. a. die Effektgröße, das Signifikanzniveau (Irrtumswahrscheinlichkeit, auch Fehler 1. Art) und die Power (1 – Fehler 2. Art). Die Annahmen für die Effektgröße beruhen bei Fallzahl-planungen für randomisierte kontrollierte Studien (RCTs) in der Regel auf Informationen aus

Pilotstudien oder vergleichenden Phase-II-Studien, die eine relativ sichere Abschätzung dieses Parameters erlauben. Für AbDs, die durch den G-BA aufgrund fehlender direkt vergleichender Evidenz gefordert werden, liegt diese Informationsbasis als Ausgangspunkt für eine Fallzahlplanung nicht vor. Aus diesem Grund kann die Fallzahlschätzung im Rahmen einer Konzepterstellung für eine AbD (häufig auf Basis nicht vergleichender Zulassungsstudien) allenfalls einen orientierenden Charakter haben.

Um aus einer nicht randomisierten vergleichenden Studie eine Aussage zum Nutzen oder zum Schaden des zu bewertenden Arzneimittels ableiten zu können, ist ein ausreichend großer beobachteter Effekt erforderlich, von dem angenommen werden kann, dass er nicht allein durch systematische Verzerrung hervorgerufen wird. Dies soll in der AbD statistisch in Form eines Tests auf eine verschobene Nullhypothese (H_0) umgesetzt werden. Dabei muss das Konfidenzintervall des in der AbD beobachteten Effekts vollständig jenseits eines konkreten Schwellenwertes liegen. Einen wissenschaftlichen Standard, wie groß die Schwelle für die verschobene Nullhypothese sein sollte, gibt es derzeit nicht. Als Schwellenwert wurde durch das IQWiG im Rapid Report A19-43 ein Bereich von 2 bis 5 für das relative Risiko (bzw. 0,2 bis 0,5 z. B. bei mortalitätssenkenden Interventionen) bezogen auf die untere (bzw. die obere) Konfidenzintervallgrenze festgelegt. Aufgrund methodischer Weiterentwicklungen erscheint es jedoch angezeigt, sich in einem nächsten Schritt (außerhalb der vorliegenden Ausarbeitung) einer Überprüfung der festgelegten Schwelle anzunehmen.

Keine oder unzureichende Vorabinformationen für eine orientierende Schätzung zum Umfang der anwendungsbegleitenden Datenerhebung

Im Rahmen der Ausarbeitung eines AbD-Konzepts wird eine orientierende Fallzahlschätzung immer dann durchgeführt, wenn sie sinnvoll möglich ist. Zum Zeitpunkt der Konzepterstellung (meist vor der Zulassung) liegen die benötigten Informationen zur Schätzung der Effektgröße für ein neues Arzneimittel jedoch in der Regel nicht vor (z. B. liegen ggf. nur vorläufige, zum Teil widersprüchliche, Ergebnisse zu lediglich selektiv in Kongressabstracts oder Pressemitteilungen dargestellten Endpunkten vor). Damit sind die Voraussetzungen für eine orientierende Fallzahlschätzung nicht erfüllt.

In diesen Fällen kommt eine orientierende Betrachtung von Fallzahlszenarien zur Anwendung, um dennoch Aussagen zum Umfang der Datenerhebung treffen zu können. Diese Herangehensweise folgt im Wesentlichen den Grundsätzen der orientierenden Fallzahlschätzung und hat lediglich einen anderen Ausgangspunkt. In einem 1. Schritt wird die geschätzte Patientenzahl im zu bewertenden Anwendungsgebiet in Deutschland ermittelt. In einem 2. Schritt wird mittels Festlegung der Patientenzahl, des Signifikanzniveaus sowie der Power die Frage beantwortet, welche Effektgrößen auf Basis der im zu bewertenden Anwendungsgebiet grundsätzlich für eine AbD in Deutschland zur Verfügung stehenden Patientenzahl (Präzision der Schätzung) und unter Berücksichtigung der Gefahr systematischer Verzerrung

wegen des nicht randomisierten Vergleichs (verschobene Nullhypothese) aufgedeckt werden können.

Aufdeckbar bedeutet, dass das im Rahmen einer AbD geschätzte 95 %-Konfidenzintervall mit 80 % Power bei gegebener Patientenzahl vollständig oberhalb bzw. vollständig unterhalb der verschobenen Hypothesengrenze liegt. Neben der Patientenzahl wird, sofern verfügbar, das Basisrisiko entweder unter Behandlung mit der Intervention oder unter Behandlung mit den Komparatoren für einen patientenrelevanten Endpunkt als Anker herangezogen, um auf diese Weise ein Spektrum der im Anwendungsgebiet möglichen Szenarien aufzuzeigen.

Entscheidungsfindung für oder gegen eine orientierende Fallzahlschätzung

Bei der Entscheidung für oder gegen die orientierende Fallzahlschätzung im Rahmen von AbD-Verfahren werden keine festen Grenzen angelegt. Ob eine orientierende Fallzahlschätzung auf Basis der jeweils vorliegenden Datenkonstellation als sinnvoll durchführbar erachtet wird, ist im Einzelfall zu prüfen.

Unabhängig vom gewählten methodischen Vorgehen für die Konzepterstellung (orientierende Fallzahlschätzung oder orientierende Betrachtung von Fallzahlszenarien) sollte der pU des zu bewertenden Arzneimittels die (finale) Fallzahlplanung der AbD zum Zeitpunkt der Zwischenanalysen auf Basis der dann vorhandenen vorläufigen Ergebnisse vornehmen. Die (finale) Fallzahlplanung unter Berücksichtigung der Ergebnisse von Zwischenanalysen vorzunehmen, wird als grundsätzlich geeignet beurteilt, da die Schätzung der notwendigen Fallzahl bei versorgungsnahen Datenerhebungen (in Abgrenzung zu klinischen Studien) nicht mit Hinblick auf eine aktive Rekrutierung erfolgt. Es werden stattdessen möglichst alle Patientinnen und Patienten einer Registerpopulation, die die Einschlusskriterien für die AbD erfüllen, nach ihrer Einwilligung in die AbD eingeschlossen. Dies erfolgt auch vor dem Hintergrund, dass, je nach Propensity-Score-Verfahren, ggf. ein relevanter Anteil der eingeschlossenen Patientinnen und Patienten nicht in der Auswertung berücksichtigt werden kann.

Beobachtungsstart und Behandlungswechsel

Beobachtungsstart

Eine Herausforderung bei versorgungsnahen Studien ohne Randomisierung ist eine adäquate Festlegung des Startzeitpunkts der Beobachtung (Index-Datum $[t_0]$), da es im Gegensatz zu RCTs (Zeitpunkt der Randomisierung) keinen festen Startzeitpunkt für die Beobachtung gibt. Eine Verzerrung der Ergebnisse (z. B. durch Immortal-Time-Bias oder Selection Bias) entsteht, wenn der Beginn der Beobachtung (t_0), die Prüfung der Erfüllung der Einschlusskriterien und die Behandlungszuteilung nicht übereinstimmen.

Beim Active-Comparator-New-User(ACNU)-Design wird für beide Behandlungsgruppen der Tag der Therapieentscheidung (oder eine bestmögliche Näherung daran) als t_0 festgesetzt. Es

werden ausschließlich jene Patientinnen und Patienten in die Studie eingeschlossen, die eine Indikation für eine Umstellung ihrer Therapie haben. Zum Umgang mit Situationen, in denen nicht in beiden Behandlungsgruppen zu Beginn der Beobachtung mit einer neuen Therapie begonnen wird, eignet sich das Prevalent-New-User(PNU)-Design. Ziel ist dabei, potenzielle Verzerrungen durch eine falsche Wahl des Beobachtungsstarts für Patientinnen und Patienten in der Kontrollgruppe zu reduzieren. Die Anwendung dieses Designs ist daran gebunden, dass geeignete Verfahren zur Berücksichtigung erhaltener Vorthherapie(n) zum Einsatz kommen. Die Robustheit der Ergebnisse sollte mittels Sensitivitätsanalysen überprüft werden.

Behandlungswechsel

In vergleichenden Studien (mit oder ohne Randomisierung) haben Patientinnen und Patienten die Möglichkeit, bedarfsgesteuert von einer Behandlung zu einer anderen zu wechseln. Dieser Vorgang wird als „Treatment Switching“ bezeichnet. Im Rahmen des Rapid Reports wird der Begriff auf einen Wechsel von der Intervention zum Komparator oder auf einen Wechsel vom Komparator zur Intervention begrenzt.

Ein Wechsel von der Intervention zum Komparator entspricht in einer AbD im Regelfall einem Wechsel zu einer allgemein anerkannten, dem medizinischen Standard entsprechenden Behandlung. Der Wechsel ist somit Teil einer Behandlungsstrategie, die der normalen klinischen Praxis entspricht. In einer solchen Situation ist eine unverzerrte Schätzung des Therapieeffekts möglich. Ein Wechsel vom Komparator zur Intervention ist dann unproblematisch, wenn die Intervention eine adäquate Folgetherapie im Rahmen der Behandlungsstrategie darstellt (dies trifft z. B. dann zu, wenn für die Intervention bereits zu einem früheren Zeitpunkt eine Zulassung in einer späteren Therapielinie erfolgt ist). Diese Konstellation ist im Rahmen einer AbD jedoch eher nicht zu erwarten. In allen anderen Fällen kann ein Wechsel vom Komparator zur Intervention zu verzerrten Ergebnissen führen und es ist nachzuweisen, dass die Ergebnisse auf die Fragestellung der AbD übertragbar sind.

Unabhängig von einem Behandlungswechsel werden zum Zwecke der Nutzenbewertung von Arzneimitteln primär Auswertungen entsprechend der Treatment-Policy-Strategie (d. h. eine Schätzung des Effekts für die gesamte Behandlungsstrategie) und des Intention-to-treat(ITT)-Prinzips benötigt. Um einen potenziell verzerrenden Einfluss eines Behandlungswechsels auf die ITT-Auswertung unter Anwendung von Verfahren für Überlebenszeiten zu untersuchen, können Sensitivitätsanalysen durchgeführt werden, in denen Patientinnen und Patienten mit Wechsel vom Komparator zur Intervention zum Zeitpunkt ihres Wechsels zensiert werden (Per-Protocol-Analysen) oder in denen die Anwendung komplexer Methoden zur Berücksichtigung von Treatment Switching (wie z. B. Inverse Probability of Censoring Weighting [IPCW]) erfolgt.

Besondere Situation bei der Auswertung versorgungsnah erhobener Daten

Da der G-BA AbDs häufig in Anwendungsgebieten von Arzneimitteln für seltene Erkrankungen (Orphan Drugs) fordert, ergibt sich per se eine niedrige für eine AbD zur Verfügung stehende Patientenzahl. Je nach zu bewertenden Therapien ergibt sich zusätzlich die Schwierigkeit, eine ausreichende Patientenzahl rekrutieren zu können. Um vor diesem Hintergrund dennoch eine möglichst hohe Patientenzahl unter Behandlung mit einer neuen Therapie in der Analyse zu berücksichtigen, besteht im Sinne eines pragmatischen Vorgehens die Option, Patientinnen und Patienten mit Wechsel vom Komparator auf die Intervention den Behandlungsgruppen in Abhängigkeit ihrer Beobachtungszeit unter Therapie mit dem Komparator zuzuordnen (d. h., dass Patientinnen und Patienten, die nach kürzerer Zeit auf die Intervention wechseln und für die noch eine angemessene Beobachtungszeit unter Therapie mit der Intervention zu erwarten ist, in der Interventionsgruppe ausgewertet werden können).

Je nach interessierendem Vergleich (neuer Wirkstoff vs. Komparator) kann es Fälle geben, in denen eine Therapieoption nach der Therapieentscheidung nicht sofort zur Verfügung steht oder nicht sofort wirksam ist. Es kann erforderlich sein, eine Brückentherapie einzusetzen, die als Teil des Therapiekonzepts zu verstehen ist. Entscheidend für einen adäquaten Umgang mit Brückentherapien ist, dass der Tag der Entscheidung für die eigentlich geplante Therapie (z. B. Zeitpunkt der Entscheidung für eine CAR-T-Zelltherapie im Tumorboard) als Startzeitpunkt der Beobachtung (Index-Datum $[t_0]$) gemäß Target-Trial-Konzept festgesetzt wird.

Patientenberichtete Endpunkte

Um bestehende Evidenzlücken zu neuen Arzneimitteln zu schließen, fordert der G-BA für AbDs regelhaft die Erhebung von Endpunkten zur Symptomatik und gesundheitsbezogenen Lebensqualität. Diese Endpunkte werden üblicherweise mit PROs erhoben. Hierfür sind zum Zwecke der Generierung aussagekräftiger Daten über den Verlauf der Erkrankung unter Therapie für beide Behandlungsgruppen einheitliche Erhebungszeitpunkte zu wählen. Die Erhebung sollte mehrfach im Jahr und in standardisierten Intervallen zu definierten Zeitpunkten erfolgen, mit zu Beginn engeren und später breiteren Intervallen. Dabei sollten für die Erhebungszeitpunkte angemessene Toleranzfenster präspezifiziert werden, die nicht aneinandergrenzen und eine sinnvolle Zuordnung zu einem Erhebungszeitpunkt möglich machen. Im Kontext einer AbD werden Daten in der Versorgungspraxis erhoben, sodass sich die Bedingungen, unter denen PRO-Daten erfasst werden, deutlich von den Bedingungen in einer klinischen Studie unterscheiden. Mittels digitaler Befragungen (z. B. per App und / oder per Patientenportal) ist eine niedrigschwellige Erhebung unabhängig von Arztbesuchen möglich. Digitale Tools werden in Registern und in der Routineversorgung bereits erfolgreich zur Datenerhebung eingesetzt.

Fehlende Werte

Maßnahmen zur Vermeidung fehlender Werte

In Studien lassen sich fehlende Werte in der Regel nicht vollständig vermeiden. Um dennoch hinreichend sichere Aussagen zum Nutzen und Schaden eines Arzneimittels sowohl unter den Bedingungen der Routineversorgung als auch unter Berücksichtigung der Fragestellung der Nutzenbewertung machen zu können, sind Maßnahmen sinnvoll und notwendig, mit denen der Anteil fehlender Werte so gering wie möglich gehalten wird. Um hochwertige Daten aus dem Versorgungsalltag praktikabel und ohne große Hürden vorhalten und bereitstellen zu können, muss eine dauerhaft verfügbare und fortgeführte Dateninfrastruktur und Dokumentation personell und finanziell in den Registern aufgebaut werden. Sowohl der Aufbau als auch der Unterhalt dieser Infrastruktur könnten durch die pUs, die auf die Daten zur Durchführung registerbasierter Studien zurückgreifen wollen, unterstützt werden. Des Weiteren wurden als förderliche Faktoren u. a.

- eine Beschränkung der Erhebung auf jene Daten, die für die Beantwortung der zu untersuchenden Fragestellung relevant sind,
- die Implementierung einer geeigneten Monitoringstrategie,
- ein regelmäßiges Schulungsangebot für das Erhebungspersonal zur Dateneingabe und
- das Zurückspielen der im Register gespeicherten Daten an die Patientinnen und Patienten (inklusive Aufklärung über den Nutzen der Erfassung patientenberichteter Daten) identifiziert.

Methodischer Umgang mit fehlenden Werten im Kontext von Propensity-Score-Analysen

Modelle zur Schätzung von Propensity Scores mittels logistischer Regression erfordern für alle potenziell relevanten Confounder vollständige Daten, d. h., dass Patientinnen und Patienten mit (mindestens) 1 fehlendem Wert in der Analyse nicht berücksichtigt werden. Wenn jedoch ausschließlich Patientinnen und Patienten mit vollständigen Daten in die Analysen eingehen (Complete-Case-Analysen), führt dies zu einer deutlichen Verzerrung des Ergebnisses, wenn die Annahme Missing completely at Random (MCAR) nicht erfüllt ist. Davon ist in Studien auf Basis versorgungsnaher Daten in aller Regel auszugehen. Demnach sind bei Propensity-Score-Analysen statistische Verfahren zum Umgang mit fehlenden Werten notwendig. Grundsätzlich werden Verfahren der multiplen Imputation als angemessen angesehen. Es steht eine Reihe von statistischen Verfahren der multiplen Imputation zur Verfügung, von denen die multiple Imputation mittels Chained Equations (MICE) als etablierte Methode zum Umgang mit fehlenden Werten anerkannt ist. Zur Überprüfung der Robustheit der so berechneten Ergebnisse kommen Sensitivitätsanalysen mit modellbasierten Methoden in Betracht, denen jeweils die MAR- oder Missing-not-at-Random(MNAR)-Annahme zugrunde liegt.

Propensity-Score-Analysen in Anwendungsgebieten mit kleinen Patientenkollektiven

Bei Studien ohne Randomisierung ist die für einen fairen Vergleich notwendige Strukturgleichheit der zu vergleichenden Gruppen in der Regel nicht gegeben. Demnach müssen Gruppenunterschiede in möglichen Confoundern bei Schätzung von Effekten durch Adjustierung für diese Confounder berücksichtigt werden. Zur Confounderadjustierung im Kontext einer AbD spielt primär die Propensity-Score-Methode eine Rolle. Propensity-Score-Verfahren wurden ursprünglich für große Datensätze (epidemiologische Fragestellungen) entwickelt. Im Rahmen der vorliegenden Ausarbeitung wurden Simulationsstudien zu Propensity-Score-Analysen bei kleinen Fallzahlen identifiziert. Diese Studien zeigen, dass Propensity-Score-Verfahren auch in Anwendungsgebieten mit kleinen Patientenpopulationen unter bestimmten Bedingungen anwendbar sind und zu interpretierbaren Ergebnissen führen können. Ein relevantes Problem besteht bei dieser Methode allerdings in der Gefahr der fehlenden Konvergenz der Modelle für die Effektschätzung. Dem Problem der fehlenden Konvergenz durch eine zu hohe Anzahl von Confoundern kann bis zu einem gewissen Grad dadurch begegnet werden, dass weniger wichtige Confounder schrittweise aus dem Modell entfernt werden. Jenes Vorgehen erfordert eine präspezifizierte Festlegung einer Reihenfolge der Confounder nach ihrer Wichtigkeit und erfolgt unter Inkaufnahme einer erhöhten Unsicherheit. Insgesamt bilden die in den Simulationsstudien untersuchten Szenarien die Situationen, wie sie in den laufenden AbD-Verfahren absehbar auftreten werden, jedoch nur in Teilen ab, da diese vorwiegend andere Effektmaße (wie Odds Ratio oder Risikodifferenz) und anteilig eine geringe Anzahl potenzieller Confounder betrachten. Zum jetzigen Zeitpunkt bleibt aufgrund von fehlender Evidenz daher unklar, in welchen Fällen (und unter welchen Voraussetzungen) Propensity-Score-Analysen in einer AbD zu interpretierbaren Ergebnissen führen werden.

Fazit

Für die Nutzenbewertung von Arzneimitteln bedarf es Daten zum Vergleich mit dem therapeutischen Standard. Da die Zulassung von Orphan Drugs häufig auf Basis nicht vergleichender Daten erfolgt, wurde das Verfahren der AbD eingeführt, mit dem Ziel, bestehende Evidenzlücken zu schließen und so eine bessere Datenbasis für die Nutzenbewertung zu erhalten. Die Datenerhebungen sind dabei als vergleichende Studien ohne Randomisierung durchzuführen. Unter Einhaltung bestimmter Qualitätsanforderungen können Studien auf Basis von Registerdaten diese Erkenntnislücke schließen.

- Bei der Planung von vergleichenden Studien ohne Randomisierung auf Basis versorgungsnaher Daten ist die Target-Trial-Emulation ein empfohlener Ansatz, um systematische (vermeidbare) Verzerrungen zu minimieren. Voraussetzung für eine optimale Emulation einer hypothetischen RCT mittels Beobachtungsdaten ist, dass die hierfür erforderlichen Daten im Registerdatensatz mit der notwendigen Vollständigkeit und Datentiefe enthalten sind. Eine hohe Datenqualität kann in der Breite nur dann erreicht werden, wenn die Datengenerierung und die Nutzbarmachung von

Registerdaten praktikabel und mit möglichst wenig Aufwand möglich sind. Als förderlich wird der Aufbau und der Unterhalt einer dauerhaft verfügbaren (funktionsfähigen) Dateninfrastruktur eingeschätzt. Dies könnte durch pUs, die auf die Register zur Durchführung registerbasierter Studien zurückgreifen wollen, unterstützt werden.

- In vergleichenden Studien ohne Randomisierung zum Zweck des Vergleichs von Therapieeffekten ist für eine adäquate Confounderkontrolle eine systematische Identifikation der relevanten Confounder sowie eine Berücksichtigung dieser Confounder in der Auswertung notwendig. Die Confounderidentifikation nach dem Ansatz von Pufulete 2022 (über eine systematische Literaturrecherche und Einbindung von Klinikerinnen und Klinikern) wird als machbar eingeschätzt und stellt ein grundsätzlich sinnvolles Vorgehen dar. Vor der Relevanzbewertung der Confounder durch klinische Expertinnen und Experten ist zu empfehlen, eine intensive Verdichtung der identifizierten Confounder vorzunehmen. Grundsätzlich kann es zielführend sein, im Vorfeld einer AbD einen höheren Aufwand zur Reduktion der Anzahl zu erhebender Confounder zu investieren, um den Gesamtaufwand für die AbD durch Aufwandseinsparungen bei der Datenerhebung und -auswertung zu reduzieren.
- In den Konzepten für eine AbD wird abgeschätzt, ob innerhalb eines akzeptablen Zeitraums ausreichend Patientinnen und Patienten eingeschlossen werden können, um mit der AbD grundsätzlich aussagekräftige Ergebnisse für eine Nutzenbewertung generieren zu können. In der Regel liegen nur unsichere Informationen für diese Abschätzung vor. Wenn hinreichende Informationen zur Intervention und dem Komparator vorliegen, wird zu diesem Zwecke eine orientierende Fallzahlschätzung durchgeführt. Fehlen notwendige Informationen für die Annahmen, die einer orientierenden Fallzahlschätzung zugrunde liegen, erfolgt eine orientierende Betrachtung von Fallzahlszenarien, um mögliche aufdeckbare Effekte zu zeigen. Beide Herangehensweisen folgen dabei dem gleichen Grundsatz (sie fußen auf identischen Parametern) und sie unterscheiden sich lediglich in dem Parameter, der mithilfe der übrigen Parameter geschätzt werden soll.
- Für eine langfristige Datenerhebung (und Begleitung von Patientinnen und Patienten) in der Routineversorgung sind Anreize erforderlich, die den Aufwand der Datengenerierung kompensieren und sowohl die Zentren als auch die Patientinnen und Patienten zur möglichst vollständigen Datenerhebung motivieren.
- Zur Sicherstellung, dass Aufwand und Nutzen bei PRO-Erhebungen in einem angemessenen Verhältnis stehen, werden digitale Befragungen (z. B. per App oder per Patientenportal) empfohlen. Auf diese Weise ist eine niedrigschwellige PRO-Erhebung unabhängig von Arztbesuchen möglich. Digitale Tools werden in Registern und in der Routineversorgung bereits erfolgreich zur Datenerhebung eingesetzt und ermöglichen,

dass PRO-Daten für Forschungszwecke barrierearm zur Verfügung gestellt werden können.

- Eine Herausforderung bei der Auswertung versorgungsnaher Daten ohne Randomisierung stellt die Festlegung des Startzeitpunkts der Beobachtung (Index-Datum [t_0]) dar. Wird in beiden Behandlungsgruppen zum Beobachtungsstart (t_0) mit einer neuen Therapie begonnen (Umstellungsindikation), kann in einer Studie das ACNU-Design verwendet werden. Dabei entspricht t_0 idealerweise dem Tag der Therapieentscheidung. In einer Therapiesituation, in der Patientinnen und Patienten in der Kontrollgruppe die Behandlung mit einem etablierten therapeutischen Standard fortführen, stellt das PNU-Design eine geeignete Alternative dar.
- Unabhängig von der Wahl des Designs und der Auswertungsstrategien sollte grundsätzlich immer primär das Ergebnis der ITT-Analyse (entsprechend der Treatment-Policy-Strategie) gezeigt werden. Dies erfordert, dass alle Patientinnen und Patienten (unabhängig von einem Behandlungswechsel im Verlauf der Beobachtung) gemäß ihrer ursprünglichen Gruppenzuteilung ausgewertet werden. Dem verzerrenden Einfluss von Behandlungswechseln im Verlauf der AbD kann mit Sensitivitätsanalysen begegnet werden.
- Eine häufig angewendete Methode zur Berücksichtigung von Confoundern in vergleichenden Studien ohne Randomisierung auf Basis von Registern stellt die Auswertung auf Basis von Propensity Scores dar. Da Modelle zur Schätzung von Propensity Scores mittels logistischer Regression für alle potenziell relevanten Confounder vollständige Daten erfordern, sind statistische Verfahren zum Umgang mit fehlenden Werten notwendig. Als etablierte Methode ist die multiple Imputation mittels Chained Equations (MICE) anerkannt. Propensity-Score-Verfahren sind auch in kleinen Patientenkollektiven unter bestimmten Bedingungen anwendbar und können zu interpretierbaren Ergebnissen führen. Die in den identifizierten Simulationsstudien untersuchten Szenarien bilden die Situationen, wie sie in den laufenden AbD-Verfahren absehbar auftreten werden, in Teilen jedoch nicht ab. Zum jetzigen Zeitpunkt bleibt daher unklar, in welchen Fällen (und unter welchen Voraussetzungen) Propensity-Score-Analysen in einer AbD zu interpretierbaren Ergebnissen führen werden.