Leibniz Institute
for Prevention Research and
Epidemiology – BIPS

Nutzung von Sekundärdaten für die pharmakoepidemiologische Forschung - machen wir das Beste draus!

Iris Pigeot, Bianca Kollhorst, Vanessa Didelez

DOI
10.1055/a-1633-3827

Published in
Das Gesundheitswesen

Document version
Accepted manuscript

This is the author's final accepted version. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

Online publication date
25 October 2021

Corresponding author
Iris Pigeot

Citation
Pigeot I, Kollhorst B, Didelez V. Nutzung von Sekundärdaten für die pharmakoepidemiologische Forschung - machen wir das Beste draus! Das Gesundheitswesen. 2021;83(S 02):S69-S76.

This is an accepted manuscript of an article published by Thieme Publishing Group in *Das Gesundheitswesen* on 25 October 2021, available online at https://www.thieme-connect.de/products/ejournals/abstract/10.1055/a-1633-3827.

Einleitung

Die Nutzung von Sekundärdaten für die Forschung hat zunehmend an Bedeutung gewonnen. Dies mag zum einen daran liegen, dass generell die Nachnutzung von Daten für die Forschung im Sinne der FAIR-Prinzipien [1] derzeit stark befördert wird [2, 3]. Zum anderen hat sich aber auch gezeigt, dass Sekundärdaten wahre „Datenschätze“ sein können [4]. Allerdings muss bei der Nutzung von Sekundärdaten wie z.B. den Abrechnungsdaten gesetzlicher Krankenkassen beachtet werden, dass diese nicht zu Forschungszwecken erhoben wurden, sondern wie in dem genannten Beispiel zu Abrechnungszwecken von Behandlungen. Es gibt also vielfache Gründe, bei der Nutzung von Sekundärdaten vorsichtig zu sein: So enthalten Abrechnungsdaten zwar sehr viele Informationen über das Behandlungsgeschehen von Versicherten und spiegeln die medizinische Alltagspraxis in der Bevölkerung inklusive vulnerablen Gruppen wider, sie sind aber nicht zur Beantwortung einer speziellen Forschungsfrage erhoben worden. Zudem fehlen häufig bestimmte Informationen, die zur Beantwortung einer Forschungsfrage essentiell wären. Da es aber in vielen Fällen keine „besseren“ Daten gibt, um schnell ohne eine zeit- und kostenaufwändige Primärdatenerhebung dringende Fragen zu beantworten, gilt es, das Potential von Sekundärdaten so gut wie möglich auszunutzen. Dabei wird man vor methodische Herausforderungen gestellt, die durch die Zeitabhängigkeit sowohl von Expositionen als auch Endpunkten sowie durch das notwendige Adjustieren für (zeitabhängiges) Confounding, aber auch durch ungemessenes Confounding entstehen.

In diesem Paper werden wir am Beispiel der pharmakoepidemiologischen Datenbank (GePaRD) [5] des Leibniz-Instituts für Präventionsforschung und Epidemiologie – BIPS verschiedene methodische Ansätze ansprechen, die sich mit den typischen Problemen von Sekundärdaten befassen. Dazu stellen wir zunächst GePaRD kurz vor, bevor wir auf Strategien eingehen, um verschiedene Biasquellen zu vermeiden und um den durch ungemessenes Confounding entstehenden Bias abzuschätzen. Anhand von Beispielen werden die verschiedenen Ansätze illustriert. Ein weiteres Kapitel befasst sich ebenfalls anhand von Beispielen mit den Möglichkeiten und den Limitationen von Record Linkage, um fehlende Informationen aus anderen Datenquellen oder zusätzlichen Primärdatenerhebungen zu ergänzen [6].

Datenlage

Am BIPS wurde 2004 eine pharmakoepidemiologische Forschungsdatenbank (GePaRD) aufgebaut, die kontinuierlich erweitert wird. Die Datenbank umfasst ca. 25 Millionen Versicherte von vier gesetzlichen Krankenversicherungen, und zwar von der AOK Bremen/Bremerhaven, Die Techniker (TK), DAK-Gesundheit und der hkk Krankenkasse. Sie deckt ungefähr 20% der lebenden deutschen Bevölkerung und alle Regionen Deutschlands ab. Jährliche pseudonymisierte Datenlieferungen werden an eine unabhängige Vertrauensstelle übermittelt, die erste Qualitätskontrollen durchführt, mögliche Rückfragen mit den Krankenversicherungen bearbeitet und ein neues internes Pseudonym vergibt, bevor der jeweilige Datensatz an die Datenaufbereitungsstelle des BIPS für weitere Qualitätsprüfungen übergeben wird. Der bereinigte Datensatz wird anschließend mit den Daten aus den Vorjahren verknüpft. Die Datenbank enthält zentrale demographische Informationen sowie Informationen zu Krankenhausaufenthalten, ambulanten Behandlungen und Verschreibungen. Da das BIPS nicht Eigner der Daten ist, steht es uns rechtlich nicht zu, Externen direkten Zugang zu GePaRD einzuräumen. In Übereinstimmung mit dem GePaRD-Datenschutzkonzept und den deutschen Datenschutzregulationen ist der Zugang zu den Daten nur den Mitarbeiter:innen des BIPS in den Räumen des Instituts und auch nur im Kontext bewilligter Forschungsprojekte erlaubt. In 2020 wurde GePaRD vom Rat für Sozial- und

Wirtschaftsdaten vorläufig als Forschungsdatenzentrum akkreditiert. Damit können externe Forschende im Rahmen von Kooperationsprojekten und nach Abschluss eines entsprechenden Gastwissenschaftler-Vertrags gemeinsam mit Forschenden am BIPS mit GePaRD arbeiten. Dabei erhalten Forschende am BIPS zur Beantwortung einer spezifischen Forschungsfrage nur Zugang zu einem speziell zu diesem Zweck erstellten Analysedatensatz. Typische Forschungsfragen betreffen das Monitoring der Anwendung von Arzneimitteln und Impfstoffen nach Zulassung sowie die Untersuchung von möglichen Fehlversorgungen bei bestimmten Indikationen, (seltenen) Arzneimittel- und Impfstoffrisiken, Arzneimittel-/Impfstoffrisiken mit einer langen Latenzzeit und von Arzneimittelinteraktionen.

GePaRD bietet viele Vorteile: So sind Studien mit verschiedenen Designs vergleichsweise schnell und kostengünstig durchführbar. Die große Patientenzahl und die vielen Datenjahre ermöglichen die Untersuchung seltener, schwerwiegender und sich spät manifestierender Ereignisse. Zudem können auch besondere PatientInnengruppen wie z.B. alte Personen, Pflegebedürftige oder Schwangere in den Studien berücksichtigt werden. Die in der Datenbank enthaltenen Angaben werden direkt von den Krankenkassen übermittelt, so dass keine studienbedingten Verzerrungen wie z.B. Erinnerungsfehler auftreten. Da darüber hinaus das öffentliche Interesse an der Forschung das Geheimhaltungsinteresse des/der Versicherten bei Projekten basierend auf GePaRD erheblich überwiegt, also gemäß §75, Sozialgesetzbuch X die individuelle Einwilligung der Versicherten nicht erforderlich und nur ein geringer Loss-to-Follow-up aufgrund von Krankenkassenwechseln zu beobachten ist, weist GePaRD eine hohe Repräsentativität auf.

Allerdings sehen wir uns auch bei GePaRD mit den in der Einleitung angesprochenen Limitationen konfrontiert. So besteht die Möglichkeit von Über-, Unter- und Fehlkodierungen durch die Verwendung von Abrechnungsdaten. Zudem fehlen verschiedene Informationen: Das Krankheits- und Versorgungsgeschehen in der Zeit vor Einrichtung der Datenbank ist nicht erfasst. Es fehlen Angaben zu Arzneimitteln wie tatsächliche Arzneimitteleinnahme, Tagesdosis, Therapiedauer und Angaben zur Einnahme von OTC („Over-the-Counter“)-Präparaten, Laborparameter sowie Angaben zu Lebensstilfaktoren wie z.B. körperliche Bewegung, Gewicht und Rauchen.

Methoden und Ergebnisse

Strategien zur Vermeidung von Bias

Um Problemen möglichst schon in einem frühen Stadium einer Studie zu begegnen, sollten die Fragestellung, das Analysedesign sowie die statistische Auswertungsmethode so gewählt werden, dass zumindest die vermeidbaren Verzerrungen und Fehlerquellen tatsächlich vermieden werden [7]. Dies ist nicht selbstverständlich, da insbesondere die zeitliche Struktur von typischen Sekundärdaten eine Herausforderung darstellt, die es bei einer kontrollierten randomisierten Studie in dieser Form nicht gibt. So ist es in der Praxis eher die Ausnahme als die Regel, dass Behandlungen zu eindeutig definierten Zeitpunkten stattfinden, sondern sie werden zu mehr oder weniger klar bestimmten Zeitpunkten begonnen, abgebrochen, gewechselt oder dauern über längere oder kürzere Zeiträume an. Mit anderen Worten, Behandlungen sind in der Praxis zeitabhängig. Typische (vermeidbare) Fehlerquellen lassen sich oft darauf zurückführen, dass diese Zeitabhängigkeit nicht angemessen berücksichtigt wird.

Vermeidbare Verzerrungsquellen

Eine typische vermeidbare Fehlerquelle bei der statistischen Auswertung von Sekundärdaten ist als „Immortal-Time-Bias“ [8] zwar seit langem bekannt, findet sich aber immer wieder auch in aktuellen Arbeiten [9]. Da bei Sekundärdaten im Gegensatz zu randomisierten Studien keine von Anfang an klare Zuweisung von Personen zu Behandlungs- und Kontrollgruppe gegeben ist, muss diese Zuordnung anhand der vorhandenen Daten nachträglich vorgenommen werden. Das Phänomen der Immortal-Time tritt dann auf, wenn die Zuordnung zur Behandlungsgruppe implizit oder explizit derart ist, dass für eine Person in dieser Gruppe der Endpunkt (z.B. Tod) in einem bestimmten Zeitraum logisch nicht auftreten konnte, da sie sonst der Kontrollgruppe zugerechnet worden wäre. Betrachtet man das Mammographie-Screening als Beispiel, liegt es nahe, anhand von vorliegenden Registerdaten alle Frauen, die irgendwann an einem Screening teilgenommen haben, in die Screening-Gruppe einzuordnen und alle anderen in die Kontrollgruppe. Aber eine Frau, deren Daten vorliegen und die mit 63 Jahren zum ersten Mal beim Screening war, kann logischerweise nicht vorher verstorben sein. D.h. Personen in einer derart definierten Screening-Gruppe haben allein deshalb schon einen „Überlebensvorsprung“, weil sie lange genug überlebt haben, um an einem Screening teilzunehmen. Solche Überlegungen sind konkret für die laufende „ZEBra“-Studie [10] zur Evaluation der Brustkrebsmortalität im deutschen Mammographie-Screening-Programm wichtig, bei der Krebsregister- und Krankenkassendaten wie u.a. GePaRD verknüpft werden [11]. Das Immortal-Time Problem würde bei einer randomisierten Studie nicht auftreten, denn die Personen würden zu Beginn der Studie zufällig den Gruppen zugewiesen, so dass es auch in der Screening-Gruppe möglich wäre, dass eine Probandin verstirbt, bevor sie jemals beim Screening war. Eine ähnlich gelagerte vermeidbare Verzerrungsquelle ist der „Prevalent-User-Bias“ [12], worauf hier nicht weiter eingegangen wird.

Das Target Trial Prinzip

Die obigen Verzerrungsquellen lassen sich durch eine klare Fragestellung, sorgfältiges Design und eine systematische statistische Analyse vermeiden. Hierbei ist das Prinzip des „Target Trial“ [13, 14] nützlich. Für eine gegebene Forschungsfrage wird zunächst die ideale Studie in Form eines Studienprotokolls beschrieben – in der klinischen Epidemiologie ist dies oft eine randomisierte Studie, die aber aus praktischen, ethischen oder finanziellen Gründen nicht durchgeführt werden kann. Zu dem Protokoll gehören z. B. konkrete Fragestellung, Einschlusskriterien, zu vergleichende Behandlungen, Endpunkt, Zeitraum und Effektmaß. Diese ideale Studie wird dann so eng wie möglich mit den verfügbaren Daten nachgebildet („emuliert“): So sollten etwa ein gleicher Startpunkt („time zero“) und gleiche Einschlusskriterien verwendet werden. Wenn dies nicht möglich ist, weil z.B. die dazu notwendige Information nicht vorhanden ist, kann zumindest abgewogen werden, wie schwerwiegend diese Abweichung ist. Ggf. sollten geeignete Sensitivitätsanalysen durchgeführt werden, um den Einfluss dieser Abweichung auf das Ergebnis abzuschätzen.

Zentrale Komponenten des Target Trials [15] angesichts von zeitabhängigen Behandlungen sind die explizite Formulierung der Fragestellung und der zu vergleichenden Behandlungsgruppen – diese sollten sich an klinisch relevanten Entscheidungsoptionen oder Interventionen orientieren. Wegen des zeitlichen Charakters spricht man auch oft von „Behandlungsstrategien“. Bei der Evaluation des Mammographie-Screenings, wie etwa in der ZEBra-Studie, kann man sich z.B. für einen strikten Kontrast zwischen den Strategien, (i) „niemals teilnehmen“ und (ii) „ab 50 regelmäßig alle zwei Jahre teilnehmen, es sein denn eine Krebsdiagnose tritt auf“, interessieren; oder man schwächt Strategie (ii)

zu (iii) „ab 50 mindestens einmal teilnehmen“ ab. Um Immortal-Time-Bias zu vermeiden, wird von dem definierten Startpunkt an und an jedem folgenden Zeitpunkt der gesamten Studiendauer jede Person jeder Behandlungsstrategie zugeordnet, mit der ihre bisherige tatsächliche Behandlungshistorie kompatibel ist; ansonsten wird sie „künstlich zensiert“. Dies kann dazu führen, dass eine Person gleichzeitig mehreren Vergleichsgruppen zugeordnet wird (man spricht hier manchmal von „Klonen“) [15]. Im obigen Screening-Beispiel wäre eine Frau, die mit 55 noch an keinem Screening teilgenommen hat, sowohl mit Strategie (i) als auch (iii) kompatibel; nimmt sie mit 56 Jahren zum ersten Mal an einem Screening teil, dann ist sie mit Strategie (iii) kompatibel, aber wird ab dann für Strategie (i) zensiert. So zählen die sechs Screening-freien Jahre zu beiden Vergleichsgruppen und werden nicht systematisch verzerrend nur einer zugerechnet – sie nur zu (i) zu zuordnen, würde Strategie (iii) nicht korrekt repräsentieren, da auch hier der Startpunkt beim Alter von 50 gewählt wird; sie nur Strategie (iii) zu zuordnen, würde Immortal Time bias induzieren. Obwohl das Target-Trial-Prinzip von Natur aus prospektiv ausgelegt ist, kann es auch mit Fall-Kontroll-Designs [16] kombiniert werden.

(Zeitabhängiges) Confounding

Die kritischste Verzerrungsquelle wurde bereits kurz angesprochen: Confounding. Für die meisten Fragestellungen, die anhand von Sekundärdaten beantwortet werden sollen, ist davon auszugehen, dass sich die Vergleichsgruppen systematisch unterscheiden, z.B. sind Screening-affine Personen ggf. insgesamt gesundheitsbewusster als andere. Allerdings enthalten Sekundärdaten oft sehr detaillierte Information, die zumindest manchmal als Proxy angesehen werden kann. Unter der Annahme, dass ausreichend Information in der Datenbasis vorhanden ist („no-unmeasured-confounding“, NUC), und wenn sich die Fragestellung als Vergleich von zeitunabhängigen Behandlungsgruppen formulieren lässt, können zur Adjustierung für das Confounding viele weitläufig bekannte Methoden verwendet werden [17]. Solche Vergleiche entsprechen oft einem „Intention-to-Treat“ Effekt, z.B. nach Diagnose „beginne mit Medikament A“ versus „beginne mit Medikament B“, wobei spätere Abweichungen von der anfänglichen Behandlung erlaubt sind. Zur Auswahl von Kovariablen, die zur Confounder-Adjustierung plausibel und ausreichend sind, gibt es zwar verschiedene statistische Methoden, diese müssen aber immer durch Expertenwissen ergänzt werden [18].

Zur Confounder-Adjustierung sind hier speziell Methoden zu nennen, die auf dem „Propensity-Score“ (PS) [19] beruhen, d.h. auf der Wahrscheinlichkeit für die Behandlungsgruppe bei gegebenen Baseline-Kovariablen. Der PS ist besonders dann nützlich, wenn inhaltliche Kenntnisse über die Verschreibungspraxis zu einem plausiblen Modell für die Behandlungsentscheidung genutzt werden können. Der PS kann dann statistisch im Rahmen von Matching, Gewichtung, Regression oder Stratifizierung eingesetzt werden und eignet sich auch besonders für den Fall, dass die Confoundermenge hochdimensional ist. Eine flexible Modellierung, für die auch immer häufiger maschinelle Lernverfahren eingesetzt werden, ist hier wünschenswert. Allerdings ist das Hauptkriterium nicht die Passung des PS-Modells, sondern die Balanciertheit der Kovariablen [20]. Jedoch kann Balanciertheit nicht immer erreicht werden, wenn sich Behandlungs- und Kontrollgruppe zu sehr unterscheiden. In einer Studie mit GePaRD Daten wurden z.B. Patienten nach einem elektivem Eingriff mit stationärer versus ambulanter Versorgung verglichen, wozu ein PS-Ansatz auf die hochdimensionale Menge an in GePaRD verfügbaren Informationen angewandt wurde. Dabei musste die Analyse aufgrund der eingeschränkten Vergleichbarkeit auf eine Teilmenge der Patienten begrenzt werden, was die Verallgemeinerbarkeit der Ergebnisse beeinträchtigt [21]. Es ist allerdings auch eine Stärke des PS-Ansatzes, dass solche Probleme in den Daten eher erkannt werden.

Wie schon erwähnt, ergeben sich die wesentlichen Schwierigkeiten aber genau dadurch, dass Behandlungen zeitabhängig sind (Abbildung 1), und dies gilt entsprechend auch für potenzielles Confounding. Um diesem entgegen zu wirken, sind einfache PS Methoden nicht geeignet. Zeitabhängiges Confounding [22] ist dadurch charakterisiert, dass manche Faktoren selbst von einer vergangenen Behandlungsentscheidung beeinflusst werden, dann aber wiederum zukünftige Behandlungen beeinflussen. Ein konkretes Beispiel ist eine Studie zum Effekt von Diabetesmedikation auf kardiovaskuläre Ereignisse basierend auf GePaRD [23]: In der Regel wird mit der Standardtherapie Metformin begonnen, dann aber bei erhöhtem HbA1c-Wert oder anderen diabetischen Komplikationen ein zusätzliches glukosesenkendes Medikament gegeben, d.h. der HbA1c-Wert ist ein potenzieller zeitabhängiger Confounder. Unter der Annahme, dass genug Information über zeitabhängige Confounder vorliegt, lässt sich dies durch geeignete statistische Verfahren adjustieren. Dazu gehören marginale Strukturmodelle (MSMs) [24], die durch sequenzielle Gewichtung mit der (zeitabhängigen) inversen Behandlungswahrscheinlichkeit angepasst werden. MSMs sind „marginal" über die zeitabhängigen Kovariablen, die nicht explizit, sondern nur durch die Gewichte mit modelliert werden. In obigem Beispiel konnte gezeigt werden [23], dass ein augenscheinlich positiver Effekt der zusätzlichen Diabetesmedikation auf kardiovaskuläre Ereignisse (in einer naiven Analyse) sich bei geeignetem Einbeziehen von zeitabhängigen Confoundern wie dem HbA1c-Wert in einem MSM als negativer Effekt herausstellte, was medizinisch plausibler war.

Eine Alternative zu MSMs beruht auf der (parametrischen) „g-Formel" [25], die den Gesamteffekt einer Behandlungsstrategie aus mehreren Modellen für die zeitliche Abfolge sequenziell zusammensetzt. *Nicht* geeignet sind Ansätze [26], bei denen ein einziges Regressionsmodell für den Endpunkt in Abhängigkeit von allen zeitabhängigen Behandlungsvariablen und den Baseline- und zeitabhängigen Confoundern spezifiziert wird; dazu zählen auch Hazardmodelle mit zeitabhängigen Variablen. Allerdings lassen sich diese Ansätze durch sequenzielles Modellieren und geeignetes „Klonen" (siehe oben) auch so anpassen, dass sie zeitabhängiges Confounding korrekt berücksichtigen [27]. Wenn bei einem Ereignisendpunkt nur wenige diskrete Behandlungsstrategien verglichen werden sollen, wie oben im Screening-Beispiel, dann kann der MSM-Ansatz durch das schon beschriebene künstliche Zensieren nahezu nicht-parametrisch umgesetzt werden. Da (zeitabhängiges) Confounding hier zur Folge hat, dass das künstliche Zensieren nicht unabhängig ist, muss dies durch inverse Gewichtung ausgeglichen werden („Inverse Probability of Censoring Weighting"). Bei anderen Endpunkten oder komplexeren Behandlungsstrategien sind parametrische Modellansätze, basierend z.B. auf MSMs oder der g-Formel, flexible Alternativen. Die parametrische g-Formel eignet sich insbesondere für dynamische Behandlungsstrategien, d.h. wenn eine Behandlung an den individuellen Krankheitsverlauf angepasst werden soll, z.B. „nach Diagnose beginne erst mit Medikament A und, wenn Unverträglichkeit auftritt, wechsle zu B". Die beiden Ansätze, MSM und g-Formel, sind zwei fundamentale Herangehensweisen, für die es inzwischen viele Variationen gibt [28]. Hierbei wird bei MSMs ein (zeitabhängiges) Modell für die Behandlungshistorie in Abhängigkeit von zeitlich vorhergehenden Kovariablen benötigt; bei der g-Formel ist dagegen ein (zeitabhängiges) Modell für die Kovariablen gegeben der zeitlich vorhergehenden Behandlungshistorie erforderlich. Im Sinne einer Sensitivitätsanalyse ist es sinnvoll, die Ergebnisse beider Ansätze zu vergleichen. In der Praxis werden aufgrund der Verschachtelung des Vorgehens und auch der „Klone" Standardfehler oft mittels Bootstrap berechnet [23].

Die genannten statistischen Methoden sollen (zeitabhängiges) Confounding auf geeignete Weise kontrollieren. Dies ist natürlich nur unter der NUC Annahme möglich, dass die Datenbasis die dafür benötigte Information überhaupt enthält – und dies ist eine kühne und gut zu begründende Annahme.

Wenn entscheidende Information fehlt, können das weder ein Target-Trial-Ansatz noch Methoden wie hochdimensionale PS-Algorithmen, MSMs oder g-Formeln ausgleichen.

Strategien zur Vermeidung von Bias durch ungemessenes Confounding

Wenn wichtige Informationen wie z.B. der Rauchstatus oder das Gewicht einer Person nicht in den Daten enthalten sind, ist die oben getroffene NUC Annahme nicht plausibel. Im Folgenden werden zunächst zwei Methoden vorgestellt, mit denen man entweder für ungemessenes Confounding in der Analyse kontrolliert, indem eine „natürliche Randomisierung" (instrumentelle Variable) verwendet wird, oder der Bias durch ungemessenes Confounding abgeschätzt werden kann (quantitative Bias-Analyse).

Instrumentelle Variablen

Eine beobachtbare Variable ist eine geeignete instrumentelle Variable (IV), wenn sie die folgenden drei Annahmen erfüllt (Abbildung 2): (1) die IV und die tatsächliche Behandlung sind assoziiert, (2) die IV und die ungemessenen Confounder sind unabhängig voneinander (Unabhängigkeitsannahme), (3) die IV und der Endpunkt sind bedingt unabhängig, gegeben dem ungemessenen Confounder und der Behandlung (Ausschlussrestriktion) [29, 30]. Die instrumentelle Variablenanalyse imitiert eine randomisierte Studie, indem sie die zufällige Variation in der Behandlungszuordnung ausnutzt, um eine Variable zu definieren, die die tatsächliche Behandlung beeinflusst, aber selbst keinen unabhängigen Effekt auf den Endpunkt hat. Indem man in der Analyse die IV anstatt der tatsächlichen Behandlung nutzt, wird der Patient mittels Pseudo-Randomisierung einer alternativen Behandlung zugeordnet. Unter der Bedingung, dass alle IV Annahmen erfüllt sind, sollten dann wie in einer randomisierten Studie alle gemessenen und ungemessenen Confounder in den Pseudeo-Behandlungsarmen gleich verteilt sein [29].

Ein kritischer Punkt in der IV-Analyse ist die Identifizierung eines geeigneten Instruments [31] und dass Annahmen (2) und (3) nicht empirisch, sondern nur durch Hintergrund- oder Expertenwissen zu begründen sind. Ein bekanntes Beispiel für eine IV ist die Arztpräferenz. Angewendet in einer Studie zum Risiko gastrointestinaler Blutungen unter Therapie mit entweder traditionellen nicht-steroidalen Antirheumatika (tNSAID) oder selektiven COX-2-Hemmern konnte in GePaRD gezeigt werden, dass die Präferenz des Arztes für ein tNSAID im Vergleich zum COX-2-Hemmern alle Annahmen eines validen Instruments zu erfüllen scheint [32]. Allerdings war eine Limitation der Studie, dass die Stärke der Assoziation zwischen der wahren Therapie und der Arztpräferenz nur moderat war, was wiederum zu einer hohe Varianz in der Schätzung führte. Die Annahmen (2) und (3) sind hier zwar plausibel, eine Verletzung konnte aber letztlich nicht ausgeschlossen werden. Studien basierend auf Krankenversicherungsdaten in anderen Ländern wie UK oder USA konnten eine valide und relativ starke Arztpräferenz in ihren Daten identifizieren. Damit wird deutlich, dass eine Prüfung der IV-Annahmen in jedem Gesundheitssystem und für jede Fragestellung individuell erforderlich ist.

Quantitative Bias-Analysen

Das Prinzip der Bias-Analyse sieht vor, dass man eine Reihe von inhaltlich plausiblen Annahmen über einen ungemessenen Confounder trifft, um anschließend in verschiedenen Szenarien den daraus resultierenden Bias für den Schätzer des Behandlungseffekts quantitativ abzuschätzen [33]. Damit kann man bestimmen, wie sensitiv der geschätzte Behandlungseffekt in Bezug auf ungemessenes Confounding ist und unter welchen Bedingungen sich die Schlussfolgerungen, die man aus seiner

Studie zieht, entscheidend verändern. Insbesondere ist hier wichtig zu bewerten, ob die Annahmen, die man über den ungemessenen Confounder getroffen hat, realistisch erscheinen. Wenn beispielweise eine unrealistisch starke Assoziation zwischen Confounder und Endpunkt nötig ist, um den beobachteten Zusammenhang zwischen Behandlung und Endpunkt zu annullieren, dann lässt sich argumentieren, dass es unplausibel ist, dass das Studienergebnis durch ungemessenes Confounding beeinflusst wird [34].

Über die Jahre sind verschiedene Typen von Bias-Analysen vorgeschlagen worden, um den Effekt ungemessenen Confoundings zu quantifizieren [35-37]. Eine leicht umzusetzende Bias-Analyse, die von Lash et al. [37] vorgeschlagen wurde, sieht beispielsweise eine simple Post-hoc-Korrektur des Behandlungsschätzers vor. Unter der Annahme, dass der ungemessene Confounder binär ist und außerdem unabhängig von allen gemessenen Confoundern, müssen hierfür als Sensitivitätsparameter die Prävalenz des ungemessenen Confounders in beiden Behandlungsarmen und die Assoziation des ungemessenen Confounders mit dem Endpunkt spezifiziert werden. Bei der Wahl der Sensitivitätsparameter ist es empfehlenswert, sich an den Werten der gemessenen Confounder zu orientieren oder an Informationen aus anderen Datenquellen. Alle drei Sensitivitätsparameter können dann variiert werden und für jedes Szenario wird ein für das entsprechende ungemessene Confounding modifizierter Behandlungsschätzer berechnet. Anschließend sollte diskutiert werden, immer in Hinblick auf die jeweilige Fragestellung, ab wann eine beobachtete Assoziation zwischen Behandlung und Endpunkt durch den ungemessenen Confounder annulliert wird und inwieweit das Szenario realistisch erscheint.

Record Linkage als eine Lösung für ungemessenes Confounding

Während die oben vorgestellten Ansätze zur Berücksichtigung von ungemessenem Confounding hauptsächlich bei der statistischen Analyse zur Anwendung kommen, besteht die Idee des Record Linkage darin, den zu analysierenden Sekundärdatensatz um die fehlenden Variablen anzureichern. Dieser Ansatz erfordert in der Regel ein besonderes Datenschutzkonzept, das in Abhängigkeit von den zu verknüpfenden Variablen und deren Gewinnung z.B. die Einholung einer informierten Einwilligung und die Einrichtung einer Treuhandstelle vorsehen muss [6]. Ein prominentes Beispiel für dieses Vorgehen ist die NAKO Gesundheitsstudie, bei der die Teilnehmenden um ihr Einverständnis gebeten werden, ihre Studiendaten mit Sekundär- und Registerdaten in der NAKO-Treuhandstelle zu verknüpfen. Dabei erlauben z.B. die Abrechnungsdaten der gesetzlichen und privaten Krankenversicherungen u.a. die Ermittlung inzidenter Krankheitsfälle und eine umfassende Beurteilung der Krankheitsgeschichte ohne zusätzliche zeitliche Belastung der Teilnehmenden [38].

Wie bereits erwähnt, sind in GePaRD keine klinischen Parameter abgebildet, da diese zu Abrechnungszwecken nicht benötigt werden. Diese ließen sich z.B. durch die Verknüpfung der Abrechnungsdaten mit Daten eines Krankenhausinformationssystems (KIS) „auffüllen“. Um dies zu testen, wurde in [39] am Beispiel von Krankenhauseinweisungen aufgrund einer Herzinsuffizienz versucht, Abrechnungsdaten aus GePaRD mit Daten des KIS des Klinikums Oldenburg zu verknüpfen. Dabei sollten das datenschutzrechtlich sehr aufwändige direkte Linkage (Goldstandard) über die pseudonymisierte Krankenversicherungsnummer mit einem indirekten Linkage-Verfahren verglichen werden. Das indirekte Linkage erfolgte anhand eines Abgleichs von Aufnahme- und Entlassungsdatum und einer Übereinstimmung eines bestimmten Anteils von KIS-Diagnosen mit GePaRD-Diagnosen. Dabei zeigte sich im indirekten Linkage eine durchgängig hohe Spezifität, während die Sensitivität vom zuvor festgelegten Grad der Übereinstimmung der Diagnosen abhing. Auch wenn man daraus

schließen kann, dass ein indirektes Linkage prinzipiell sinnvoll umgesetzt werden kann, wurde kritisch hinterfragt, ob der Aufwand in diesem Fall gerechtfertigt war, da bedingt durch einen hohen Anteil an fehlenden Angaben im KIS der gewünschte Informationsgewinn größtenteils nicht erzielt werden konnte.

In einem noch laufenden Projekt, finanziert durch die DFG, sollte die Möglichkeit eruiert werden, fehlende Daten zu Krebserkrankungen in GePaRD durch die Verlinkung mit Daten der epidemiologischen Krebsregister am Beispiel des kanzerogenen Potenzials von Antidiabetika zu ergänzen. Auch hier sollte ein indirektes Linkage-Verfahren mit einem direkten Linkage verglichen werden, da sich das direkte Linkage in der Praxis als extrem aufwändig herausstellt. Der datenschutzrechtlich erforderliche Datenfluss ist in Abbildung 3 dargestellt. Dieser ließ sich jedoch nicht jedem Einzelfall umsetzen, da aufgrund des föderalen Systems in Deutschland in jedem Bundesland das derzeit geltende Krebsregistergesetz anders umgesetzt wird, so dass z.B. in einem Bundesland weder das direkte noch das indirekte Linkage als mit dem Krebsregistergesetz kompatibel angesehen und dementsprechend auch nicht ermöglicht wurde. Das Projekt hat sich bedingt durch die große Anzahl von Bewilligungsprozessen erheblich verzögert.

In den beiden letzten Beispielen musste keine informierte Einverständniserklärung der Versicherten eingeholt werden, da die Datennutzung von GePaRD über §75 SGB X und die der anderen Datenquellen durch entsprechende Gesetze geregelt sind. Im folgenden Beispiel war dies allerdings erforderlich, da die fehlenden Informationen durch eine Primärbefragung eines Teils der Versicherten im Rahmen eines sogenannten Two-Phase-Designs [40] gewonnen werden sollten, wodurch das Datenschutzkonzept deutlich komplexer wurde. In dem Projekt sollte das Blutungsrisiko nach Einnahme von Phenprocoumon untersucht und zusätzlich die Frage beantwortet werden, ob die auf GePaRD basierende Schätzung des Blutungsrisikos valide ist oder ob zusätzliche Confounder-Information das geschätzte Risiko deutlich verändert [41]. Abgesehen von den hohen datenschutzrechtlichen Anforderungen ergab sich als weiteres Problem die für Primärbefragungen typische geringe Teilnahmebereitschaft: Von 4.000 angeschriebenen Personen konnten nur 3.280 Briefe zugestellt werden, von denen nur 505 (15.4%) zu einem Interview bereit waren. Die Befragungsdaten ergaben, dass Rauchen und ein hoher Body Mass Index zwar das Blutungsrisiko erhöhten, jedoch nicht zu einer deutlichen Änderung der Ergebnisse alleine basierend auf GePaRD führten, was für die Validität von Datenbankstudien spricht.

Fazit

In vielen Studien konnte der Nutzen von Sekundärdaten eindeutig belegt werden. Jedoch muss man sich der möglichen Verzerrungsquellen bei ihrer Nutzung bewusst sein, um durch geeignete statistische Methoden gegensteuern zu können. In diesem Beitrag haben wir einige dieser methodischen Probleme und mögliche Lösungen vorgestellt.

Wünschenswert wäre in diesem Zusammenhang, wenn grundsätzlich die Nachnutzung solcher Daten für die Forschung vereinfacht würde, insbesondere in Hinblick auf die Verknüpfung verschiedener Datenquellen. Hier zeigt sich, dass das Fehlen eines bundesweiten sogenannten Unique Identifiers eine schwer überwindbare Hürde darstellt, die je nach Datenquelle durch bundeslandspezifische gesetzliche Regelungen zur Nachnutzung noch einmal vergrößert wird. Zu hoffen bleibt, dass sich im Rahmen des Aufbaus einer Nationalen Forschungsdateninfrastruktur bundeseinheitliche Lösungen

finden lassen, die die Auffindbarkeit, den Zugang, die Interoperabilität und die Nachnutzung auch von Sekundärdaten und die Verknüpfung von verschiedenen Datenquellen deutlich erleichtern werden.

Literaturverzeichnis

1. Wilkinson MD, Dumontier M, Aalbersberg IJ et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016; 3: 160018.
2. Nationale Forschungsdateninfrastruktur (NFDI) e. V. Nationale Forschungsdateninfrastruktur (2021). Im Internet: https://www.nfdi.de; Stand: 30.03.2021
3. EOSC Association. European Open Science Cloud (2021). Im Internet: https://eosc.eu/; Stand: 30.03.2021
4. Jacobs S, Stallmann C, Pigeot I. Verknüpfung großer Sekundär- und Registerdatenquellen mit Daten aus Kohortenstudien. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2015; 58: 822-828.
5. Pigeot I, Ahrens W. Establishment of a pharmacoepidemiological database in Germany: methodological potential, scientific value and practical limitations. Pharmacoepidemiol Drug Saf 2008; 17: 215-223.
6. March S, Andrich S, Drepper J et al. Gute Praxis Datenlinkage (GPD). Gesundheitswesen 2019; 81: 636-650.
7. Hernán MA, Sauer BC, Hernández-Díaz S et al. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. J Clin Epidemiol 2016; 79: 70-75.
8. Suissa S. Immortal time bias in pharmacoepidemiology. Am J Epidemiol 2008; 167: 492-499.
9. Haque R, Shi J, Schottinger JE et al. Tamoxifen and antidepressant drug interaction in a cohort of 16,887 breast cancer survivors. J Natl Cancer Inst 2016; 108.
10. Institut für Epidemiologie und Sozialmedizin der Universität Münster. ZEBra-MSP Evaluation der Brustkrebsmortalität im deutschen Mammographie-Screening-Programm (2021). Im Internet: https://www.medizin.uni-muenster.de/epi/forschung/projekte/zebra-msp.html; Stand: 13.04.2021
11. Langner I, Riedel O, Czwikla J et al. Linkage of routine data to other data sources in Germany: a practical example illustrating challenges and solutions. Gesundheitswesen 2020; 82: S117-S121.
12. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. Am J Epidemiol 2003; 158: 915-920.
13. García-Albéniz X, Hsu J, Hernán MA. The value of explicitly emulating a target trial when using real world evidence: an application to colorectal cancer screening. Eur J Epidemiol 2017; 32: 495-500.
14. Didelez V. Commentary: Should the analysis of observational data always be preceded by specifying a target experimental trial? Int J Epidemiol 2016; 45: 2049-2051.
15. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol 2016; 183: 758-764.
16. Dickerman BA, García-Albéniz X, Logan RW et al. Emulating a target trial in case-control designs: an application to statins and colorectal cancer. Int J Epidemiol 2020; 49: 1637-1646.
17. Hernán MA, Robins JM. Causal inference: what if. Boca Raton: Chapman & Hall/CRC; 2020
18. Witte J, Didelez V. Covariate selection strategies for causal inference: classification and comparison. Biom J 2019; 61: 1270-1289.
19. Jackson JW, Schmid I, Stuart EA. Propensity scores in pharmacoepidemiology: beyond the horizon. Curr Epidemiol Rep 2017; 4: 271-280.
20. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat Med 2009; 28: 3083-3107.
21. Enders D, Ohlmeier C, Garbe E. The potential of high-dimensional propensity scores in health services research: an exemplary study on the quality of care for elective percutaneous coronary interventions. Health Serv Res 2018; 53: 197-213.
22. Daniel RM, Cousens SN, De Stavola BL et al. Methods for dealing with time-dependent confounding. Stat Med 2013; 32: 1584-1618.

23. Enders D, Engel S, Linder R et al. Robust versus consistent variance estimators in marginal structural Cox models. Stat Med 2018; 37: 3455-3470.
24. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000; 11: 550-560.
25. Young JG, Cain LE, Robins JM et al. Comparative effectiveness of dynamic treatment regimes: an application of the parametric g-formula. Stat Biosci 2011; 3: 119-143.
26. Li X, Young JG, Toh S. Estimating effects of dynamic treatment strategies in pharmacoepidemiologic studies with time-varying confounding: a primer. Curr Epidemiol Rep 2017; 4: 288-297.
27. Gran JM, Røysland K, Wolbers M et al. A sequential Cox approach for estimating the causal effect of treatment in the presence of time-dependent confounding applied to data from the Swiss HIV Cohort Study. Stat Med 2010; 29: 2757-2768.
28. Petersen M, Schwab J, Gruber S et al. Targeted maximum likelihood estimation for dynamic and static longitudinal marginal structural working models. J Causal Inference 2014; 2: 147-185.
29. Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol 2018; 47: 358.
30. Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. Stat Methods Med Res 2007; 16: 309-330.
31. Martens EP, Pestman WR, de Boer A et al. Instrumental variables: application and limitations. Epidemiology 2006; 17: 260-267.
32. Kollhorst B, Abrahamowicz M, Pigeot I. The proportion of all previous patients was a potential instrument for patients' actual prescriptions of nonsteroidal anti-inflammatory drugs. J Clin Epidemiol 2016; 69: 96-106.
33. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. Pharmacoepidemiol Drug Saf 2006; 15: 291-303.
34. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. Ann Intern Med 2017; 167: 268-274.
35. Greenland S. Multiple-bias modelling for analysis of observational data. J R Stat Soc Ser A Stat Soc 2005; 168: 267-306.
36. Dorie V, Harada M, Carnegie NB et al. A flexible, interpretable framework for assessing sensitivity to unmeasured confounding. Stat Med 2016; 35: 3453-3470.
37. Lash TL, Fox MP, Fink AK. Applying quantitative bias analysis to epidemiologic data. 1. Aufl. New York: Springer; 2009. doi:10.1007/978-0-387-87959-8
38. Ahrens W, Greiser KH, Linseisen J et al. Erforschung von Erkrankungen in der NAKO Gesundheitsstudie. Die wichtigsten gesundheitlichen Endpunkte und ihre Erfassung. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2020; 63: 376-384.
39. Ohlmeier C, Hoffmann F, Giersiepen K et al. Verknüpfung von Routinedaten der Gesetzlichen Krankenversicherung mit Daten eines Krankenhausinformationssystems: machbar, aber auch „nützlich“? Gesundheitswesen 2015; 77: e8-e14.
40. Cain KC, Breslow NE. Logistic regression analysis and efficient design for two-stage studies. Am J Epidemiol 1988; 128: 1198-1206.
41. Behr S, Schill W, Pigeot I. Does additional confounder information alter the estimated risk of bleeding associated with phenprocoumon use-results of a two-phase study. Pharmacoepidemiol Drug Saf 2012; 21: 535-545.

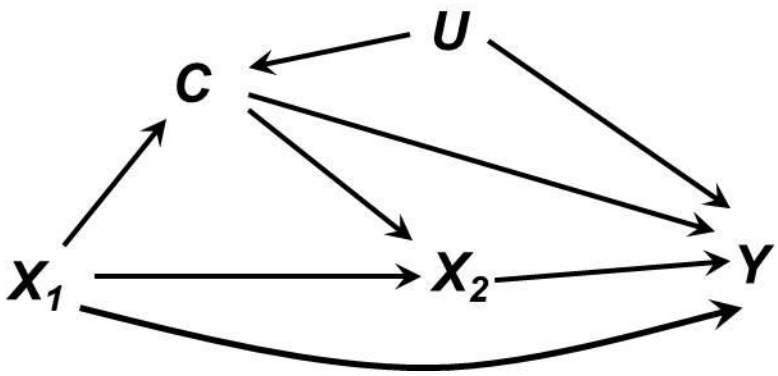

Abbildung 1: Kausaler Graph zur Illustration von zeitabhängigem Confounding. Hier ist Y der interessierende Endpunkt, X1 eine erste Behandlungsentscheidung; C ist eine mögliche Folge (z.B. Unverträglichkeit), die dann zu einer weiteren Behandlungsentscheidung X2 führt. Hierbei ist C auch oft Ausdruck eines zugrunde liegenden unbeobachteten gesundheitlichen Problems U. Baseline Confounding ist hier nicht dargestellt. In diesen Fall benötigen die g-Formel oder MSMs Daten über C aber nicht über U.

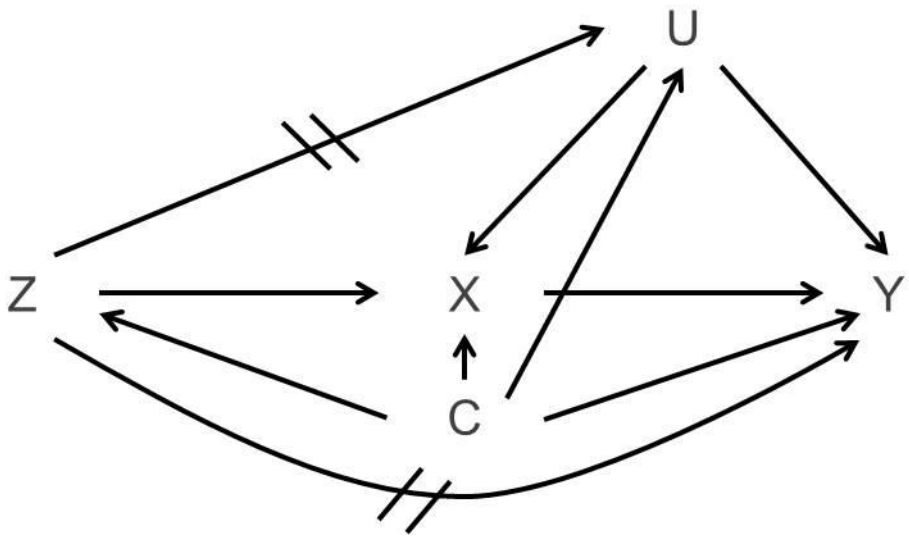

Abbildung 2: Graph zur Darstellung der drei IV Annahmen. Hier ist Y der interessierende Endpunkt, X die wahre Behandlung, Z die instrumentelle Variable, C gemessene Confounder und U der ungemessene Confounder. In Rot sind die drei IV Annahmen dargestellt.

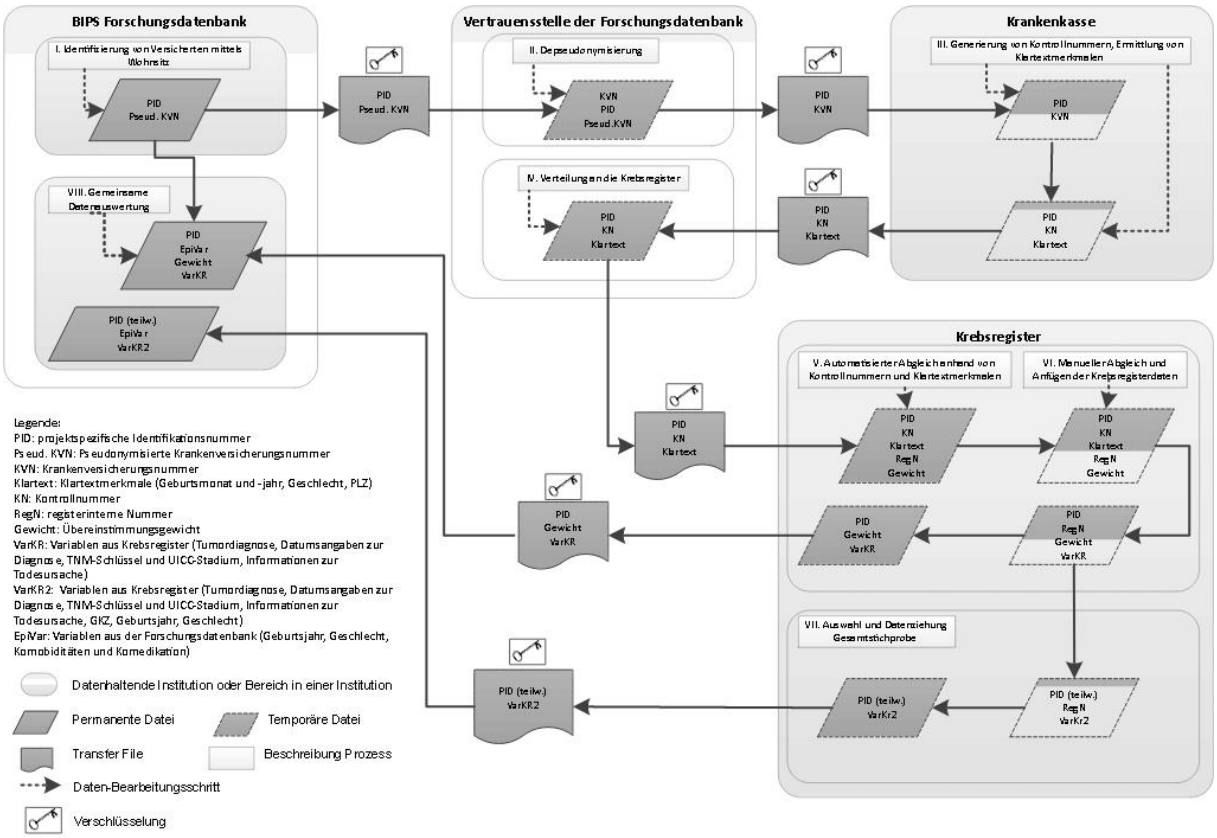

Abbildung 3: Datenfluss zur Verknüpfung von Krebsregisterdaten mit GePaRD