Thieme

Sonderheft Methodische Aspekte der Sekundärdatenanalyse

Aufbereitung und Validierung verschiedener Sekundärdatenquellen

Autoren
Enno Swart[1], Holger Gothe[2, 3, 4], Falk Hoffmann[5], Peter Ihle[6], Ingrid Schubert[6], Christoph Stallmann[1], Stefanie March[1]

Institute
1 Institut für Sozialmedizin und Gesundheitssystemforschung, Medizinische Fakultät, Otto-von-Guericke-Universität Magdeburg
2 IGES Institut, Berlin
3 Medizinische Fakultät, Lehrstuhl Gesundheitswissenschaften / Public Health, TU Dresden, Dresden
4 Department für Public Health, Versorgungsforschung und Health Technology Assessment, UMIT, Hall in Tirol, Austria
5 Department für Versorgungsforschung, Fakultät Medizin und Gesundheitswissenschaften, Carl von Ossietzky Universität Oldenburg
6 PMV forschungsgruppe, Universität zu Köln

Bibliografie
DOI https://doi.org/10.1055/a-1083-5461
Gesundheitswesen 2020; 82: (Suppl. 1) S1–S3
© Georg Thieme Verlag KG Stuttgart · New York
ISSN 0949-7013

Korrespondenzadresse
PD Dr. rer. biol. hum. Enno Swart
Medizinische Fakultät
Institut für Sozialmedizin und Gesundheitssystemforschung (ISMG)
Otto-von-Guericke-Universität Magdeburg
Leipziger Straße 44
39120 Magdeburg
enno.swart@med.ovgu.de

Die wissenschaftliche Nutzung von Sekundärdaten, nicht allein von solchen der gesetzlichen Krankenversicherung (GKV), hat sich in den vergangenen Jahren fest in der Versorgungsforschung etabliert. Das zeigen das Memorandum Methoden der Versorgungsforschung [1] und Grundlagenwerke der Sekundärdatenanalyse [2]. Gleichwohl entwickelt sich eine hohe Dynamik in diesem Forschungszweig, einerseits durch die zahlreichen durch den Innovationsfonds des Gemeinsamen Bundesausschusses (G-BA) geförderten Projekte mit Nutzung von Sekundärdaten, andererseits durch die rasant zunehmenden wissenschaftlichen Studien mit einem individuellen Datenlinkage von Primär- und Sekundärdaten oder Sekundärdaten unterschiedlicher Dateneigner. Weiterhin geraten zunehmend Sekundärdaten jenseits der weithin bekannten GKV-Abrechnungsdaten in den Fokus der Wissenschaft. Daher erscheint es den Herausgebern sinnvoll, ergänzend zu den in früheren Jahren etablierten oder neu publizierten Standards der Sekundärdatenanalyse [3–5] einen breiten methodischen Überblick zu geben über die für die Sekundärdatenanalyse zugänglichen Datenquellen, über methodische Herausforderungen bei deren Erschließung, Validierung und Aufbereitung sowie neue methodische Ansätze bei der Analyse der Daten zu diskutieren.

Die beiden Sonderhefte zu Methoden der Sekundärdatenanalyse – Heft 1: Aufbereitung und Validierung verschiedener Sekundärdatenquellen, Heft 2: Datenlinkage und spezifische Methoden der Sekundärdatenanalyse – sind gemeinsam herausgegeben durch die Arbeitsgruppen Erhebung und Nutzung von Sekundärdaten (AGENS) der Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSMP) und der Deutschen Gesellschaft für Epidemiologie (DGEpi) sowie Validierung und Linkage von Sekundärdaten des Deutschen Netzwerks Versorgungsforschung (DNVF). Nach einem Call im Frühjahr 2018 gingen 40 Entwürfe für Beiträge ein, die durch jeweils zwei der sieben Herausgeberinnen und Herausgeber unter Berücksichtigung möglicher Interessenskonflikte unabhängig voneinander unter Verwendung einer standardisierten Matrix mit Blick auf die Zielsetzung der Sonderhefte bewertet wurden. Abschließend fand in einer finalen Abstimmung in dieser Gruppe die Auswahl von insgesamt 20 Beiträgen für die beiden Sonderhefte statt. Die fertigen Manuskripte wurden anschließend einem normalen Reviewprozess unterworfen, in dem zwei Beiträge von den Autorinnen und Autoren zurückgezogen wurden.

Alle Beiträge sind Open Access verfügbar und ein Großteil ist in englischer Sprache verfasst, um so einem möglichst breiten Publikum zugänglich zu sein und auch die internationale Diskussion voranzutreiben. Die Sonderhefte richten sich vorzugsweise an alle Kolleginnen und Kollegen, die wissenschaftlich mit Sekundärdaten arbeiten, sei es seit vielen Jahren, sei es erst seit kurzem im Zuge der Förderung zahlreicher Innovationsfonds-Projekte. Ein kurzer Überblick über die neun Beiträge des ersten Sonderhefts zeigt bereits die Vielfalt der verfügbaren Datenquellen und die Notwendigkeit ihrer fundierten Kenntnis, um das wissenschaftliche Potenzial der jeweiligen Datenquellen zu nutzen.

Viele auf GKV-Daten basierende Sekundärdatenanalysen nutzen Daten einer oder nur weniger Kassen. Auch in diesen Fällen kommt der Prüfung und der einheitlichen und transparenten Aufbereitung der von den Dateneignern übermittelten Rohdaten eine zentrale Rolle zu. Diese Aufgabe wird umso komplexer und wichtiger, je größer die Zahl der in eine Studie involvierten Kassen ist. In einem Werkstattbericht schildern March et al. die damit verbundenen Herausforderungen und Lösungsansätze in einer Evaluationsstudie unter Beteiligung von mehr als 70 datenliefernden Kassen. Es zeigt sich dabei, dass auch bei einer Konsentierung der Datensatzbeschreibung und der Datenübermittlung sowie einer weitgehenden Standardisierung der Einleseprozesse die Aufbereitung und Prüfung der Daten selbst nach mehreren Zyklen keines-

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

wegs automatisiert ablaufen kann, sondern vielfältige Rückkopplungsprozesse mit den Dateneignern erfordern. Diese sind begründet durch externe (z. B. Kassenfusionen, neue Kodiervorschriften) und interne Faktoren (z. B. Umstellungen der EDV), die auch in anderen mehrjährigen Studien einer kontinuierlichen Qualitätssicherung der Prozesse bedürfen.

ten Thoren et al. beschreiben die Nutzung des beim DIMDI vorgehaltenen kassenübergreifenden Datensatzes nach der Datentransparenzverordnung (DaTraV) zur Bestimmung der Diabetesmellitus-Typ-2-Population im Zuge der frühen Nutzenbewertung des Gemeinsamen Bundesauschusses. Zur Identifikation der Typ-2-Diabetiker wird ein Algorithmus für die kombinierte Nutzung der Diagnose- und Arzneimittelverordnungsdaten (ICD- bzw. ATC-kodiert) angewendet. Der auf ein Berichtsjahr bezogene Algorithmus erweist sich als geeignet unter der Annahme zeitlich konstanter morbiditätsbezogener Rahmenbedingungen. Gleichwohl zeigen sich Limitationen des DaTraV-Datensatzes durch den begrenzten Variablenbestand ohne vertragsärztliche oder stationäre Leistungsdaten und aufgrund der 2013 noch geltenden Nichtberücksichtigung der im Jahresverlauf durch Tod ausgeschiedenen Versicherten. Letztgenannter Nachteil ist inzwischen überwunden, nicht jedoch die inhaltliche Begrenzung. Daher kommt der mit der Neufassung des § 303ff SGB V geplanten Etablierung eines Forschungsdatenzentrums aus der Sicht der Herausgeberinnen und Herausgeber eine herausragende Bedeutung zu.

Ditscheid et al. thematisieren die Validität der zeitlichen Angaben zum Zeitpunkt der Leistungserbringung und der Leistungsabrechnung. Bei der Analyse von Leistungsinanspruchnahmen ist zu berücksichtigen, dass Leistungen auch deutlich nach Ende des Berichtszeitraums oder einem Todesdatum abgerechnet werden und dass das Abrechnungsdatum nur monats- (Arzneimittelverordnungen) oder quartalsgenau (vertragsärztliche Leistungen) vorliegt. Die Autoren konkretisieren die Verzögerungen bis zur Abrechnung einer Leistung und damit die Aufnahme des Vorgangs in eine wissenschaftlich nutzbare Datenbank. Dieser Tatbestand ist bei der zeitlichen Planung von Sekundärdatenanalysen und der vollständigen Verfügbarkeit der Leistungsdaten eines Berichtszeitraums oder eines verstorbenen Versicherten zu berücksichtigen.

Mit dem Stichwort Sekundärdaten werden vielfach nur Leistungsdaten der gesetzlichen Krankenversicherung verstanden. Gleichwohl existieren weitere wichtige Datenquellen für die Versorgungsforschung, wie etwa die fallbezogenen Krankenhausabrechnungsdaten. Nimptsch et al. thematisieren Herausforderungen bei der Operationalisierung von Analyseeinheiten und Zielgrößen zum Zwecke der Minimierung von klassifizierungs- und anderweitig begründeten Verzerrungen. Die von der Autorengruppe als zielführend angesehenen Lösungsansätze ergeben sich aus der Beantwortung geeigneter Fragen, bspw. „Wie kann ich die Untersuchungsgruppen definieren?“, „Wie kann ich meine Zielgrößen messen?“ oder „Wie kann ich eine Verzerrung bei der Risikoadjustierung vermeiden?“. Bei Beachtung der in diesem Beitrag dargelegten und an Beispielen erläuterten Prinzipien steht einer breiten Nutzung dieser über die Forschungsdatenzentren der statistischen Ämter des Bundes und der Länder bereitgestellten Daten nichts im Wege.

Sozialdaten enthalten nur wenige Informationen zum sozioökonomischen Status. Dazu gehört in den Daten der gesetzlichen Kranken- wie der Rentenversicherung für erwerbstätig Versicherte die 2011 letztmalig überarbeitete Klassifikation der beruflichen Tätigkeit, über den die Berufstätigkeit, Angaben zur Schul- und Berufsausbildung und zur Vertragsform (Teilzeit-/Vollzeit bzw. Befristung des Vertrags) standardisiert verschlüsselt sind. Dieser Tätigkeitsschlüssel kann verwendet werden, um in arbeitsepidemiologischen Untersuchungen näherungsweise über eine Job-Exposure-Matrix (JEM) individuelle Arbeitsbelastungen zu bestimmen und diese auf Zusammenhänge mit gesundheitsbezogenen Outcomes zu untersuchen. Das zeigen Brünger et al. an zwei Beispielen mit scientific-use-files der Deutschen Rentenversicherung. Zwar liegen für die große Mehrheit der sozialversicherungspflichtig Beschäftigten die Angaben zum Tätigkeitsschlüssel vor, gleichwohl bedarf die Anwendung des Schlüssels bei jeder Studie einer eigenständigen Prüfung. Bei GKV-Daten ist zu beachten, dass diese Angaben nur für Mitglieder, jedoch nicht für Familienversicherte oder Rentnerinnen und Rentner vorliegen.

Das Projekt Qualitätssicherung mit Routinedaten (QSR)[1] nutzt seit mehreren Jahren stationäre GKV-Abrechnungsdaten und Daten anderer Versorgungssektoren zur Herstellung von Qualitätstransparenz im Krankenhaus. Dieser Ansatz wird von Behrendt et al. auf die Qualität der stationären Langzeitpflege übertragen. Beispielhaft wird über eine individuelle patientenbezogene Verknüpfung von Leistungsdaten der gesetzlichen Kranken- und Pflegeversicherung unter Nutzung einrichtungsbezogener Pflegetransparenzberichte (PTB) die Häufigkeit inzidenter Dekubitusdiagnosen und deren interinstitutionelle Variabilität untersucht. Dabei zeigen sich deutlich höhere Inzidenzen als in den PTB berichtet. Wenn es gelingt, die Validität pflegebezogener Versorgungsaspekte, die sich in Routinedaten nur schwer abbilden lassen, zu erhöhen, wird hierfür eine vergleichende Qualitätsbewertung von Pflegeeinrichtungen ermöglicht.

Bei Diskussionen über Sekundärdaten werden mitunter Daten jenseits der Sozialversicherungsträger ausgeblendet. Arndt et al. erweitern die Perspektive um die der bevölkerungsweiten Krebsregister, die derzeit durch die Verschmelzung epidemiologischer und klinischer Krebsregister, und damit einer Verbreiterung ihrer Inhalte, eine Aufwertung ihres Verwertungspotenzials erfahren. Die Autoren beschreiben die formale Struktur, Dateninhalte und Zugangsmöglichkeiten zu den Landeskrebsregistern, mit einer Referenz zu weiteren international verfügbaren Krebsregisterdaten. Ein Fokus der Darstellung liegt auf den individuellen Verlinkungsmöglichkeiten der Krebsregister mit Primärdaten aus epidemiologischen Studien und/oder aus anderen Sekundärdatenquellen (vgl. auch die Beiträge von Langer et al. und Bartholomäus et al. aus dem zweiten Schwerpunktheft). Breiten Raum nimmt die Beschreibung eines möglichen Bias bei der Nutzung der Registerdaten und von Strategien zu dessen Vermeidung oder Minimierung ein.

Das Notfallgeschehen ist einer der Versorgungssektoren mit einer besonders ausgeprägten Dynamik. Von vielen Akteuren werden die Zunahme der Fälle in den Notfallaufnahmen der Krankenhäuser und die steigende Arbeitsbelastung des Personals als eine der größten Herausforderungen für das Gesundheitswesen betrachtet. Für die Erarbeitung von Lösungsansätzen zur Verbesse-

[1] https://www.qualitaetssicherung-mit-routinedaten.de/

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

rung der Versorgungssituation ist ein empirisch gestütztes Verständnis des Leistungsgeschehens essentiell. Greiner et al. beschäftigen sich mit der Problematik, dass je nach stationärer oder ambulanter Versorgung der Notfälle unterschiedliche Dokumentations- und Abrechnungsvorschriften der verschiedenen Dateneigner gelten und eine Identifizierung der Notfälle oftmals nur indirekt möglich ist. Ansatzpunkte für eine verbesserte Dokumentation und eine eindeutige Identifizierung der Notfälle sind die Einführung eines Fachabteilungsschlüssels für Notaufnahmen sowie ein einheitlicher Dokumentationsstandard in den angestrebten integrierten Notfallzentren, inkl. der Erfassung der Triage und des CEDIS-Katalogs für Vorstellungsgründe.

Schon innerhalb eines Landes stellt sich oftmals die Frage nach der Vergleichbarkeit von Datenquellen und damit der aus ihnen abgeleiteten Kennzahlen und Ergebnisse. Das gilt umso mehr bei der Nutzung formal vergleichbarer Datenquellen aus mehreren Ländern, etwa beim Vergleich der Strukturen und Prozesse verschiedener Gesundheitssysteme. Voraussetzung für den Vergleich verschiedener Datenquellen ist ein Kriterienkatalog als Grundlage für deren Gegenüberstellung. Am Beispiel der deutschen DaTraV-Daten und der niederländischen VECTIS-Datenbank, die ebenfalls der Risikoadjustierung innerhalb des gesetzlichen Krankenversicherungssystems dient, fokussieren Hrzic et al. den Vergleich auf die jeweilige Bezugspopulation, die Datenursprünge und Datenverarbeitung bis hin zu den Inhalten der für Forscher zugänglichen Datenbank und den Nutzungsvoraussetzungen. Die Autoren kommen zu dem Schluss, dass trotz vieler Gemeinsamkeiten von einer Vergleichbarkeit keineswegs ausgegangen werden kann, und benennen Ansätze zu einer Harmonisierung der wissenschaftlichen Nutzung unterschiedlicher Datenquellen.

Wir wünschen nun allen Leserinnen und Lesern einen Erkenntnisgewinn bzgl. der Planung und Durchführung eigener Sekundärdatenanalysen oder bei der kritischen Reflexion von Forschungsberichten und wissenschaftlichen Publikationen auf der Basis von Sekundärdatenanalysen. Ein zweites Sonderheft mit weiteren neun Beiträgen thematisiert das Datenlinkage und spezifische Methoden der Sekundärdatenanalyse.

Interessenkonflikt

Die Autorinnen/Autoren geben an, dass kein Interessenkonflikt besteht.

Literatur

[1] Pfaff H, Glaeske G, Neugebauer EAM et al. Memorandum III: Methoden für die Versorgungsforschung (Teil I). Gesundheitswesen 2009; 71: 505–510. doi:10.1055/s-0029-1234066

[2] Swart E, Ihle P, Gothe H et al., Hrsg. Routinedaten im Gesundheitswesen: Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und Perspektiven. 2. vollst. überarb. Aufl. Bern: Hans Huber; 2014

[3] Swart E, Gothe H, Geyer S et al. Gute Praxis Sekundärdatenanalyse (GPS): Leitlinien und Empfehlung, 3. Revision, Fassung 2012/2014. Gesundheitswesen 2015; 77: 120–126. doi:10.1055/s-0034-1396815

[4] Swart E, Bitzer EM, Gothe H et al. STandardisierte BerichtsROutine für SekundärdatenAnalysen (STROSA) – ein konsentierter Berichtsstandard für Deutschland, Version 2. Gesundheitswesen 2016; 78 (suppl 1): e145–e160. doi:10.1055/s-0042-108647

[5] March S, Andrich S, Drepper J et al. Gute Praxis Datenlinkage (GPD). Gesundheitswesen 2019; 81: 636–650. doi:10.1055/a-0962-9933

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.