

Epidemiologische Studien auf Basis von Sekundärdaten

Enno Swart und Peter Ihle

Die Versorgungsforschung beschäftigt sich vornehmlich mit der Frage nach dem Versorgungsbedarf umschriebener Bevölkerungen und damit, welche Leistungen welche Versicherten an welchem Ort und von wem bekommen, welche zeitlichen Trends der Inanspruchnahme sich zeigen und wie heterogen die Versorgung in regionaler und sozialer Hinsicht ist. Es liegt nahe, zur Beantwortung dieser und ähnlicher Fragen die bei der Gesetzlichen Krankenversicherung (GKV) vorhandenen Routinedaten zu nutzen. GKV-Routinedaten bieten sich für die Versorgungsforschung aus einer Vielzahl von Gründen an: Sie stehen im Prinzip jederzeit zur Nutzung bereit, sind daher vergleichsweise kostengünstig und (mehr oder weniger) aktuell verfügbar. Weitere herausragende Eigenschaften sind ihr eindeutiger Personen- bzw. Bevölkerungsbezug, ihre Vollständigkeit und weitgehende Verzerrungsfreiheit oder die Möglichkeit längsschnittlicher Analysen, um nur einige Vorteile zu nennen (Swart und Ihle, 2005). Die konkreten Nutzungsmöglichkeiten von GKV-Routinedaten sind vielfältig und können einschlägigen Publikationen entnommen werden (Swart et al., 2008; Grobe, 2008; Schubert et al., 2008).

Die Nutzer von GKV-Routinedaten sehen sich bei allen Chancen, die ihnen die Daten bieten, auch mit methodischen Herausforderungen konfrontiert. Dazu gehört zuerst die datenschutzrechtlich einwandfreie und bevölkerungsbezogene Erschließung vorhandener Datenquellen (Ihle, 2008). Konkrete, jedoch lösbare Probleme ergeben sich zusätzlich aus der Beschränkung auf den GKV-versicherten Personenkreis (ca. 85% der Bevöl-

kerung Deutschlands; BMG, 2011), und auf gegenüber der GKV abgerechnete Leistungen. Hinzu kommt das weitgehende Fehlen klinischer Informationen und die fragliche Validität bzgl. der Klassifikation von Diagnosen und Prozeduren bzw. Verzerrungen durch die Analyse singulärer Datenkörper einzelner Kostenträger (Hoffmann et al., 2008a,b, Schubert et al., 2010). In realistischer Abschätzung ihrer Grenzen werden auf diese Weise Chancen und Potenziale der GKV-Daten erkennbar. Dabei überwiegen insgesamt die Vorteile bei weitem, da die Informationen aus Sekundärdaten zur Inanspruchnahme medizinischer Leistungen aus Primärerhebungen entweder gar nicht oder nur mit hohem Aufwand bei gleichzeitig geringerer Qualität gewonnen werden können.

Sekundärdatenanalyse, also die Nutzung von Daten im Rahmen wissenschaftlicher und/oder praxisrelevanter Untersuchungen ohne direkten Bezug zum primären Erhebungsanlass, stellt sich demnach nicht mehr als eine wie früher häufig noch missverstandene „2.-Klasse-Wissenschaft“ dar, sondern als eigenständiges Forschungsfeld, das in einem spezifischen inhaltlichen und strukturellen Kontext spezielle Methoden bereit hält und gegenüber der (Primär-)Epidemiologie gleiche Wertigkeit besitzt. So werden Sekundärdatenanalysen im Methodenmemorandum des Deutschen Netzwerks Versorgungsforschung (Hoffmann et al., 2005) als wichtiger und gleichberechtigter methodischer Zweig der Versorgungsforschung anerkannt (Glaeske et al., 2009) und die Nutzung von Sekundärdaten auch vielfach in Forschungsbeschreibungen explizit empfohlen.

Der Vollständigkeit halber sei erwähnt, dass mit dem Begriff Sekundärdaten nicht allein Routinedaten der GKV gemeint sind, sondern beispielsweise auch Daten der Renten- oder Unfallversicherung. Diese werden durch die Forschungsdatenzentren der Rentenversicherung, der Bundesagentur für Arbeit und der statistischen Ämter ebenfalls fall- und zum Teil personenbezogen für die Forschung bereitgestellt (Swart und Ihle, 2005; nähere Informationen sind unter www.fdz-rv.de bzw. <http://fdz.iab.de> bzw. www.forschungsdatenzentrum.de) erhältlich. Die thematisierten Chancen und Grenzen der Nutzung von Sekundärdaten gelten grundsätzlich auch für diese Datenkörper.

Die Nutzung von Sekundärdaten bedarf dabei besonderer Methoden und stellt prinzipiell ebenso hohe wissenschaftliche Anforderungen wie bei Primärerhebungen. Diese Tatsache - verbunden mit dem Anspruch einer gleichwertigen Wissenschaft bzw. Forschungsmethodik - konfrontiert die Nutzer bei der Sekundärdatenanalyse mit allgemein anerkannten wissenschaftlichen Standards. Dazu gehört unter anderem die „Gute Epidemiologische Praxis“ (Hoffmann et al., 2008a), vor allem aber die „Gute Praxis Sekundärdatenanalyse“ (GPS), die im Jahr 2012 ihre zweite Revision durch die Arbeitsgemeinschaft *Erhebung und Nutzung von Sekundärdaten der Deutschen Gesellschaft für Sozialmedizin und Prävention* (DGSMPP) und der *Deutschen Gesellschaft für Epidemiologie* (DGEpi) erfuhr (AGENS, 2012). In der GPS wird auf die spezifischen Anforderungen und Rahmenbedingungen von Sekundärdatenanalysen eingegangen, sie ergänzt

und konkretisiert damit als eigenständige Leitlinie die in der GEP formulierten allgemeingültigen Qualitätsanforderungen in der Epidemiologie.

Eine Erweiterung der beschriebenen Nutzungsmöglichkeiten von Sekundärdaten in Richtung ‚traditioneller‘ Epidemiologie eröffnet sich im individuellen Datenlinkage mit Primärdaten aus epidemiologischen Studien. Dieser Ansatz wird inzwischen von vielen Epidemiologen als Königsweg gesehen, um die Begrenzungen von Primär- und Sekundärdaten gleichzeitig und komplementär zu überwinden

(Hoffmann et al., 2008b). Aktuelle Studien lassen erkennen, dass die Verknüpfung mit Primärdaten technisch möglich und rechtlich zulässig ist – wenngleich i.d.R. mit hohem logistischem Aufwand verbunden - und neuartige Erkenntnisgewinne verspricht (March et al., 2012; Swart et al., 2011a,b). Aufgrund dieser Erfahrungen ist in der ab 2013 geplanten Nationalen Kohorte ebenfalls ein individuelles Datenlinkage von Primär- und Sekundärdaten (GKV-Abrechnungsdaten, Krebsregisterdaten etc.) vorgesehen (Hoffmann et al., 2011).

Der korrespondierende Autor erklärt, dass kein Interessenkonflikt vorliegt.

Literatur siehe *Literatur zum Schwerpunktthema*.

<http://journals.elsevier.de/pubhef/literatur>

doi:10.1016/j.phf.2012.06.016

Dr. Enno Swart
Institut für Sozialmedizin und
Gesundheitsökonomie
Medizinische Fakultät
Otto-von-Guericke-Universität
Magdeburg
Leipziger Str. 44
39120 Magdeburg
enno.swart@med.ovgu.de

Einleitung

Die Nutzung von Routinedaten der gesetzlichen Krankenversicherung und anderer Sozialversicherungsträger (Sekundärdaten) hat in den letzten 20 Jahren innerhalb der Versorgungsforschung und –epidemiologie einen starken Aufschwung erfahren. In Kenntnis der Chancen und Limitationen der Nutzung dieser Daten, sowie der Etablierung spezifischer methodischer Standards, werden diese Sekundärdaten nun auch für epidemiologische Studien durch die Möglichkeiten der Verknüpfung von Primär- und Sekundärdaten interessant.

Summary

The importance of secondary data from statutory health insurances and other social insurance institutions for health services research has grown over the past 20 years. Knowledge about chances and limitations of these data as well as the establishment of specific methodological standards provides a well-grounded basis for epidemiological research which can benefit from the possibilities that the combined use of primary and secondary data offers.

Schlüsselwörter:

Sekundärdaten = claims data, Routinedaten = administrative data, Versorgungsepidemiologie = health services research, Datenverknüpfung = data linkage

Literaturverzeichnis

- Arbeitsgruppe Erhebung und Nutzung von Sekundärdaten. Gute Praxis Sekundärdatenanalyse. 2. Revision, 2012 [abrufbar unter www.dgepi.de].
- Bundesministerium für Gesundheit. Daten des Gesundheitswesens 2011. Berlin [abrufbar unter www.bmg.bund.de].
- Glaeske G, Augustin M, Abholz HH, Banik N, Brüggjenjürgen B, Hasford J, et al. Epidemiologische Methoden für die Versorgungsforschung. Memorandum des Deutschen Netzwerks Versorgungsforschung Das Gesundheitswesen 2009;71:685–93.
- Grobe TG. Arbeiten mit Daten der Gmünder Ersatzkasse. Ein exemplarischer Überblick - Bundesgesundheitsbl - Gesundheitsforsch -Gesundheitsschutz 2008;51:1105–17.
- Hoffmann F, Andersohn F, Giersiepen K, Scharnetzky E, Garbe E. Validierung von Sekundärdaten. Grenzen und Möglichkeiten. Bundesgesundheitsbl - Gesundheitsforsch -Gesundheitsschutz 2008a;51:1118–27.
- Hoffmann W, Bobrowski C, Fendrich K. Sekundärdatenanalyse in der Versorgungsepidemiologie. Bundesgesundheitsbl - Gesundheitsforsch -Gesundheitsschutz 2008b;51:1193–201.
- Hoffmann W, Jöckel K-H, Kaaks R, Wichmann HE. The National Cohort. A prospective epidemiologic study resource for health and disease research in Germany 2011 [abrufbar unter www.nationale-kohorte.de].
- Hoffmann W, Latza U, Terschüren C. Leitlinien und Empfehlungen zur Sicherung Guter Epidemiologischer Praxis (GEP) – überarbeitete Fassung nach Revision. Das Gesundheitswesen 2005;65:217–25.
- Ihle P. Datenschutzrechtliche und methodische Aspekte beim Aufbau einer Routinedatenbasis aus der Gesetzlichen Krankenversicherung zu Forschungszwecken. Bundesgesundheitsbl - Gesundheitsforsch -Gesundheitsschutz 2008;51:1127–34.
- March S, Rauch A, Thomas D, Bender S, Swart E. Datenschutzrechtliche Vorgehensweise bei der Verknüpfung von Primär- und Sekundärdaten in einer Kohortenstudie: die lidA-Studie. Das Gesundheitswesen (e pub first 01.02.2012).
- Schubert I, Ihle P, Köster I. Interne Validierung von Diagnosen in GKV-Routinedaten: Konzeption mit Beispielen und Falldefinition. Das Gesundheitswesen 2010;72:316–22.
- Schubert I, Köster I, Küpper-Nybelen J, Ihle P. Versorgungsforschung mit GKV-Routinedaten. Nutzungsmöglichkeiten versichertenbezogener Krankenkassendaten für Fragestellungen der Versorgungsforschung. Bundesgesundheitsbl - Gesundheitsforsch -Gesundheitsschutz 2008; 51:1095–105.
- Swart E, Ihle P (Hrsg.) Routinedaten im Gesundheitswesen. Handbuch Sekundärdatenanalyse. Bern: Hans Huber Verlag, 2005.
- Swart E, Deh U, Robra B-P. Die Nutzung der GKV-Daten für die kleinräumige Analyse und Steuerung der stationären Versorgung. Bundesgesundheitsbl - Gesundheitsforsch -Gesundheitsschutz 2008;51. 1183-1192.
- Swart E, March S, Thomas D, Salomon T, von dem Knesebeck O. Die Eignung von Sekundärdaten zur Evaluation eines Interventionsprojekts – Erfahrungen aus der AGil-Studie. Prävention und Gesundheitsförderung 2011a; 6:305–11.
- Swart E, March S, Thomas D, Salomon T, von dem Knesebeck O. Erfahrungen mit der Datenverknüpfung von Primär- und Sekundärdaten in einer Interventionsstudie. Das Gesundheitswesen 2011b;73:e126–32. und 2011; 73:860–1.