

Datennutzung im Gesundheitswesen aus Sicht der Versorgungsforschung

von Holger Gothe¹, Enno Swart² und Peter Ihle³

ABSTRACT

Daten im Gesundheitswesen zu Forschungs- und Entwicklungszwecken zu nutzen, wird zunehmend attraktiver. Prinzipiell bieten sich zahlreiche Datenquellen dafür an; der Fantasie und den Erwartungen an neue Erkenntnisse scheinen kaum Grenzen gesetzt. Und doch stellen sich wissbegierigen Datennutzern eine Reihe von Herausforderungen in den Weg. Der vorliegende Beitrag beschreibt die Chancen und Risiken der Datennutzung im Gesundheitswesen aus Sicht der Versorgungsforschung und schildert das Alternieren zwischen Faszination und Ernüchterung, dem diejenigen sich ausgesetzt sehen, die den Daten auf den Grund gehen wollen. Pointiert lässt sich diese Polarisierung in fünf Paradoxa fassen, die im Beitrag aufgezeigt werden.

Schlüsselwörter: Versorgungsforschung, Datentransparenz, Routinedaten, Sekundärdaten, Big Data

Using data in the health care system for research and development purposes is becoming increasingly attractive. In principle, there are numerous data sources available for this purpose; there seem to be no limits to the imagination and expectations of new findings. However, there are a number of challenges for data users who are eager to learn. This article points out the opportunities and risks of data use in the German health care system from the perspective of health services research and describes the alternation between fascination and disillusionment that those who want to get to the bottom of the data are exposed to. This polarization can be summarized in five paradoxes, which are formulated in the article.

Keywords: health services research, data transparency, routine data, secondary data, big data

1 Datenverfügbarkeit gestern, heute und morgen

Das Interesse an der Nutzung von Daten im Gesundheitswesen ist in den vergangenen Jahren sprunghaft gestiegen. Gründe dafür gibt es viele: politische Setzungen wie die zunehmende Bedeutung von bevölkerungsbezogenen Nutzen nachweisen für gesundheitsbezogene Interventionen, technische Rahmenbedingungen wie die fulminante Entwicklung der Informationstechnologie, die die Datenverarbeitung einem breiteren Publikum erschließt. Außerdem sind da die

jüngsten Impulse der Legislative: Mit dem „Gesetz für mehr Sicherheit in der Arzneimittelversorgung“ (GSAV) hat der Gesetzgeber verfügt, dass bestimmte Arzneimittel auch nach der Zulassung sukzessive ihren zusätzlichen Nutzen belegen sollen – und zwar anhand von Daten, die während des Arzneimitteleinsatzes gesammelt werden sollen, mittels sogenannter anwendungsbegleitender Datenerhebungen. Solche Entwicklungen in der Evidenzgenerierung werden nicht nur bestimmten Datensammlungs- und Dokumentationsformen wie beispielsweise Registern (Stausberg et al. 2020) Vorschub leisten, sie führen uns zugleich auch die damit einhergehenden Chancen und Risiken vor Augen.

¹Dr. med. Holger Gothe, Bereich Versorgungsforschung, IGES Institut GmbH · Friedrichstraße 180 · 10117 Berlin · Telefon: 030 230809 91
E-Mail: holger.gothe@iges.com

²PD Dr. rer. biol. hum. Enno Swart, Institut für Sozialmedizin und Gesundheitssystemforschung (ISMG), Medizinische Fakultät · Otto-von-Guericke-Universität Magdeburg · Leipziger Str. 44 · 39120 Magdeburg · Telefon: 0391 67 24306 · E-Mail: enno.swart@med.ovgu.de

³Peter Ihle, PMV forschungsguppe, Universität zu Köln · Herderstraße 52 · 50931 Köln · Telefon: 0221 478 85532 · E-Mail: peter.ihle@uk-koeln.de

Die Datenlandschaft, die sich heute beobachten lässt, ist bereits sehr facettenreich und birgt interessante Perspektiven, ist aber bei weitem nicht das Arkadien, zu dem sie sich bei kreativem Einsatz aller Möglichkeiten zum Vorteil der Versicherten und Patienten entwickeln könnte.

2 Die Paradoxa der Datenlandschaft

Waren es in den Anfängen der Analyse großer Datensätze für die Versorgungsforschung in erster Linie technische Limitationen, die in apparativen Voraussetzungen begründet waren, wie etwa der eingeschränkten Verfügbarkeit und Leistungsfähigkeit der nötigen Hardware, so sind es heute im Wesentlichen die Komponenten Datenschutz, Administration, kritisch reflektierende Zurückhaltung sowie fehlende Fantasie bezüglich der Möglichkeiten, die einer freizügigeren Nutzung von verfügbaren und erschließbaren Daten entgegenstehen. Hieraus ergeben sich mehrere Paradoxa, die die Versorgungsforschung darin behindern, das Höchstmaß an Erkenntnis aus ihrem noch immer zu wenig genutzten Substrat – nämlich den Versorgungsdaten – zu ziehen. Fünf dieser Paradoxa sollen im Folgenden formuliert und zur Diskussion gestellt werden.

2.1 Das Vorhanden-aber-spät-zugänglich-Paradoxon

Sekundärdaten, insbesondere die Routinedaten der Krankenkassen, entstehen zum Zeitpunkt der Leistungserbringung, also in der Arztpraxis, in der Apotheke, im Labor, in den Krankenhäusern und so weiter. Damit sind sie – zumindest theoretisch – sofort für Auswertungen zugänglich. Einer zeitnahen Datenanalyse steht jedoch die Vielzahl der Leistungserbringer und damit auch die Vielzahl der Speicherorte entgegen. Erst die abrechnungstechnischen Prozesse führen die Daten personenbezogen bei den Krankenkassen zusammen. Diese Prozesse benötigen aber je nach betrachtetem Sektor unterschiedlich lange Übermittlungszeiten: Daten aus dem stationären Sektor sind heutzutage oft schon nach nur wenigen Tagen verfügbar, Verwaltungsdaten in der Regel innerhalb eines Monats, Daten aus dem ambulanten ärztlichen Sektor hingegen erst nach mehreren Monaten. Eine Versorgungsforschung, die alle Sektoren einschließt und damit das Konzept der Real World Data nahezu ideal bedient, hinkt zum jetzigen Zeitpunkt der aktuellen Versorgungserbringung um rund ein Jahr hinterher. Weitere administrative und qualitätssichernde Prozesse im Zuge der Datenerschließung verzögern die Analyse zusätzlich. Von einer Datennutzung on demand kann also nicht die Rede sein. Dieser Zeitverzug wirkt umso problematischer, als dass akute Notfallsituationen schnelle datenbasierte Entscheidungen erfordern. Dies wird uns aktuell durch die SARS-CoV-2-Pandemie deutlich vor Augen geführt. Die für solche Analysen und damit auch die Bewertung aktueller

Gesundheitslagen sowie das Monitoring und Management des Versorgungsgeschehens benötigten Daten sind bereits personenbeziehbar im System: direkter Erregernachweis im Labor, Liegezeiten im Krankenhaus mit Beatmungstunden, Komorbiditäten und Arzneimitteltherapien aus dem ambulanten Sektor; für die Versorgungsforschung sind sie jedoch erst frühestens zwölf Monate später verfügbar. So wird aus etwas prinzipiell Vorhandenem etwas spät Zugängliches.

2.2 Das Verlockend-aber-nicht-geheuer-Paradoxon

Durch ihre ursprüngliche Zweckbestimmung – im Falle von Krankenkassendaten zum Zwecke der Abrechnung von Versorgungsleistungen – sind die Daten in der Regel sehr gut geeignet, alle in der alltäglichen Versorgungspraxis anfallenden Leistungen abzubilden – zumindest, sofern es sich um Leistungen handelt, die den Krankenkassen zur Erstattung übermittelt werden. Dies ist insbesondere dann vorteilhaft, wenn ein Blick über diejenigen Sektoren hinweg angestrebt wird, die an einem komplexen Versorgungsgeschehen und insbesondere bei chronisch-progredienten oder chronisch-exazerbierenden Erkrankungen beteiligt sind. In solchen Konstellationen die Prozessdaten der Krankenversicherungen – ob ihrer über den primären Zweck hinausgehenden Nutzung oft Sekundärdaten genannt – analysieren zu können, ist aufgrund der Einblicke in die stationäre und ambulante Versorgung also sehr verlockend. Zugleich wird die Euphorie dadurch gedämpft, dass aufgrund der Zielsetzung, mit der diese Daten ursprünglich gesammelt wurden, durch abrechnungstriggernde Phänomene (beispielsweise Upgrading und Upcoding von Leistungspositionen im intramuralen Sektor) grundsätzlich ein Bias verbunden sein kann: Sekundärdaten werden von Kritikern zum Teil mit gewissem Recht verstanden als Daten minderer Qualität. Es ist verlockend, in sie einzutauchen, lässt aber manches Partikularergebnis als nicht ganz geheuer erscheinen.

2.3 Das Machbar-aber-nicht-sinnvoll-Paradoxon

Es ist unbestritten, dass große Datensätze, die nicht unter wissenschaftlichen Kautelen entstehen, wie dies klassischerweise bei den sogenannten Sekundärdaten der Fall ist, eine Reihe von eindrucksvollen Vorteilen, aber auch einige gravierende Nachteile haben, die es in lege artis durchgeführten Auswertungs- und Interpretationsprozessen zu berücksichtigen gilt.

Die Vorteile von Sekundärdaten liegen in erster Linie darin, dass sie nicht durch systematische Fehler einer Versuchsanordnung verzerrt sind, wie sie in Form von Ein- und Ausschlusskriterien für randomisierte kontrollierte Studien (RCTs) typisch sind. Sekundärdaten gestatten es zudem, große Populationen über lange Zeiträume zu analysieren. Aufgrund ihrer Herkunft aus der Versorgungsrealität ermöglichen sie eine Beobachtung von Interventionen unter den naturalistischen Bedingungen des medizinischen Alltags

ABBILDUNG 1

Positive Konnotationen des Begriffs Gesundheitsdaten im Internet



Die Wortwolke veranschaulicht, wie häufig welche Begriffe im Zusammenhang mit dem Begriff Gesundheitsdaten im Internet auftreten. Am häufigsten findet sich die Attribution „groß“ mit 205.000, gefolgt von „schnell“ mit 159.000 Nennungen. Schlusslicht ist „wirklichkeitstreu“ mit 25 Nennungen.

Quelle: eigene Darstellung; Grafik: G+G Wissenschaft 2020

und sind damit für die Nutzen- und Effizienzbewertung von zunehmender Relevanz (Gothe et al. 2013; Swart et al. 2014).

Diese Vorteile gehen mit dem Nachteil einher, dass Vergleiche von Populationen, die unterschiedlichen Interventionen ausgesetzt sind, dadurch verzerrt sein können, dass die Zuweisung zur Interventions- und Kontrollgruppe nicht a priori nach dem Zufallsprinzip erfolgen kann. Dies birgt die Gefahr eines *confoundings*, wodurch die Interpretierbarkeit und insbesondere das Ableiten kausaler Zusammenhänge zwischen Intervention und Outcome kompromittiert wird (Berger et al. 2009; Cox et al. 2009; Johnson et al. 2009) und avancierte Methoden der Auswertung angezeigt sein können (Hernán et al. 2008; Hernán und Robins 2017).

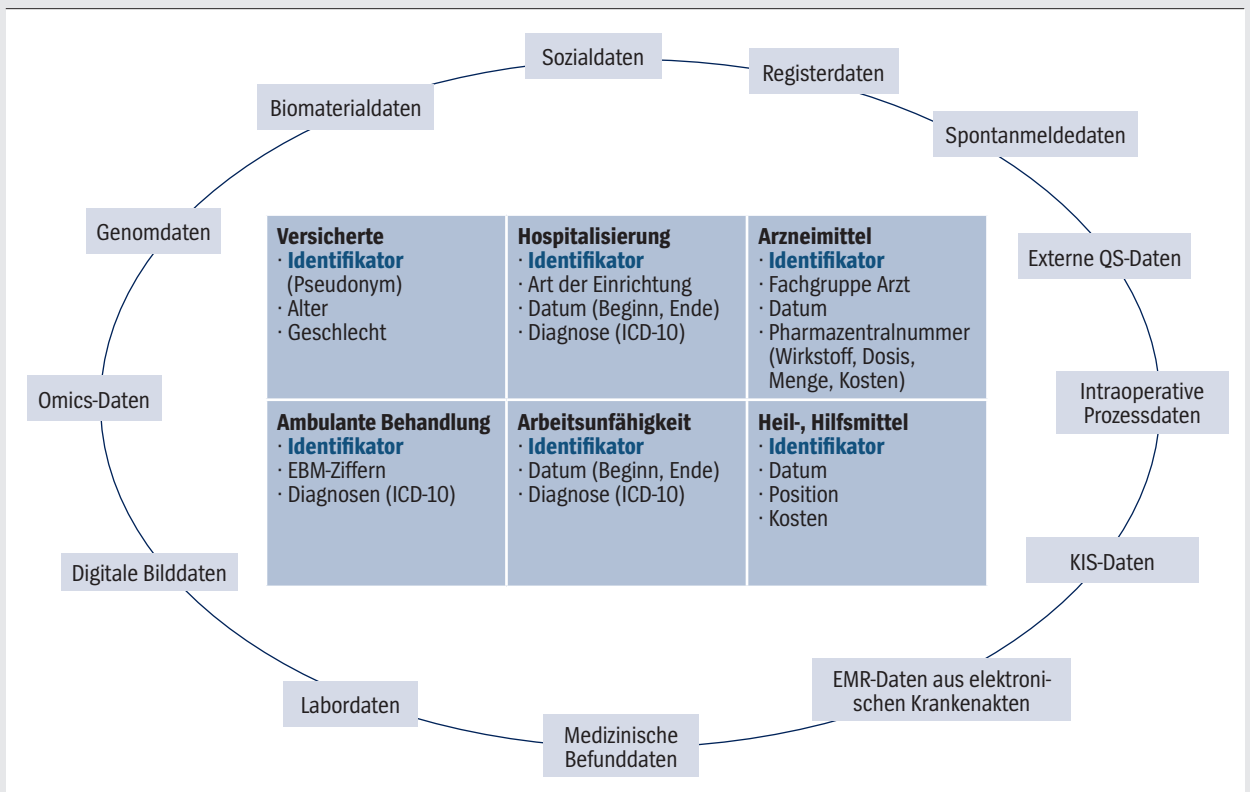
Hierin ist somit das dritte Paradoxon zu sehen, das die Arbeit mit Versorgungsdaten kennzeichnet. Gleichwohl sind Sekundärdaten, da sie genau dem Setting entstammen, für das gemäß Fünftem Buch des Sozialgesetzbuchs (SGB V) Nutzen-, Schaden- und Effizienzeffekte von Interventionen bewertet werden sollen, nämlich dem System der gesetzlichen Krankenversicherung (GKV), von unschätzbarem Wert für die Versorgungsforschung. Und aus diesem Grunde hat die Sekundärdatenforschung gerade in der zurückliegenden Dekade eine so beeindruckende Dynamik entfaltet.

2.4 Das Aufdringlich-aber-sensibel-Paradoxon

Die Auswertung von Daten steht oftmals im Spannungsfeld konkurrierender Interessen oder Vorgaben, umso mehr, als dass es sich bei den hier betrachteten Gesundheitsdaten um sensible und damit besonders schützenswerte Daten handelt (Amtsblatt der Europäischen Union: Verordnung (EU) 2016/679). Insbesondere das Spannungsfeld zwischen dem Recht der Betroffenen auf informationelle Selbstbestimmung auf der einen und dem Recht auf Freiheit der Forschung auf der anderen Seite muss in jedem datenbasierten Projekt aufgelöst werden, unter anderem durch ein zwingendes Datenschutzkonzept. Die Auswertung der Daten ist jedoch weniger Selbstzweck, sondern geeignet, rechtliche Vorgaben zu erfüllen, beispielsweise das Wirtschaftlichkeitsgebot nach § 12 SGB V (1): „Die Leistungen müssen ausreichend, zweckmäßig und wirtschaftlich sein.“ Aus dem Recht auf Forschung wird dadurch die Verpflichtung, Entscheidungen für die Übernahme einer Leistung durch die GKV datenbasiert zu fällen. In § 70 SGB V wird zudem festgeschrieben, dass „die Versorgung der Versicherten [...] in der fachlich gebotenen Qualität [...] erbracht werden“ muss. Strukturierte Behandlungsprogramme nach § 137 SGB V müssen hinsichtlich ihrer Auswirkung auf die Versorgung bewertet und evaluiert werden, Modellvorhaben müssen wissenschaftlich begleitet werden (§ 63 SGB V).

ABBILDUNG 2

Bereits erhobene Routinedaten und weitere, prinzipiell einbindbare Daten



Die Daten in der Mitte der Abbildung werden schon regelhaft erhoben und ausgewertet. Die Daten im Kreis darum herum ließen sich einbinden. KIS steht dabei für Krankenhausinformationssysteme, EMR für *electronical medical record*, QS für Qualitätssicherung und Omics für bioanalytische Hochdurchsatzverfahren.

Quelle: eigene Darstellung; Grafik: G-G Wissenschaft 2020

Dies sind rechtliche Vorgaben, die eine versichertenbezogene Datenbasis nahezu erzwingen. Und über allem steht das Prinzip der Solidargemeinschaft (§ 1 SGB V), das ein zusätzliches Spannungsfeld aus dem Recht auf informationelle Selbstbestimmung der Betroffenen und der gesetzlich verankerten Verpflichtung zum Nachweis des partikularen Gruppennutzens aufbaut. Diese Spannungsfelder sind im jeweiligen Forschungskontext immer wieder neu zu bewerten. Was sich aufdrängt, ist die Verpflichtung zur Forschung, auch wenn ihr Substrat sensible Daten sind.

2.5 Das Leitlinienkonform-aber-missbrauchsgefährdet-Paradoxon

Befürworten eines systematischen Ausbaus der Nutzung von Versorgungsdaten und ihrer systematischen Vernetzung wird in der öffentlichen Diskussion um Chancen und Risiken regelmäßig vorgeworfen, unreflektiert und geradezu

übergriffig die Nutzung von Daten zu empfehlen oder zu fordern, die ihrem Charakter nach eine wissenschaftliche Nutzung ausschließen. Individuelle Gesundheitsdaten werden jenseits von klinischer Forschung mit ihren strengen organisatorischen und rechtlichen Rahmenbedingungen von vielen Kritikern der Versorgungsdatenforschung als grundsätzlich ungeeignet angesehen; eine entsprechende Nutzung wird als unzulässig verstanden. Diese Einschätzung resultiert aus der Befürchtung, ein zu liberaler Umgang mit den Datenbeständen würde die Büchse der Pandora öffnen und einer unkontrollierten und unautorisierten Nutzung auch durch Unbefugte den Weg ebnen.

Derartig unberechtigten und mangels konkreter Verstöße in der Vergangenheit unbelegten Vorwürfen begegnet die Versorgungsforschung damit, dass sie selbst sich spezifische Regeln des wissenschaftlichen Arbeitens, der Reflexion über Potenziale und Limitationen der Nutzung dieser Daten

sowie der Transparenz über ihr wissenschaftliches Tun gegeben hat. Etablierte Leitlinien wie die Gute Praxis Sekundärdatenanalyse (GPS) (Swart *et al.* 2015), die Gute Praxis Datenlinkage (GPD) (March *et al.* 2019) und der Berichtsstandard STROSA (Swart *et al.* 2016) seien hier genannt. Zusammen mit den weitergehenden gesetzlichen Regelungen, etwas der Datenschutzgrundverordnung (DSGVO), ist damit bestmöglich Gewähr gegeben, dass jenseits eines begründeten Verdachts eben kein Missbrauch dieser Daten anzunehmen ist. Das sieht auch das Bundesverfassungsgericht so, das im März 2020 einen Eilantrag gegen die Revision der Datentransparenzverordnung nach Paragraph 303 a–f im Rahmen des Digitale-Versorgung-Gesetzes (DVG) abschlägig beschieden hat.

3 Datenbegehrlichkeiten der Versorgungsforschung

Bei allen in den fünf Paradoxa aufgegriffenen Vorzügen sind Sekundärdaten jedoch mit dem grundsätzlichen Nachteil behaftet, dass in ihnen für gewöhnlich keine Daten aus der Vermessung (also Körpergewicht, Größe etc.), keine klinischen Befundparameter, keine (patho)physiologischen Messwerte, Untersuchungs- oder Laborbefunde enthalten sind. Auch ist die Patientenperspektive in Form von *patient related outcomes* oder der gesundheitsbezogenen Lebensqualität nicht hinreichend abgebildet. Die alleinige Nutzung von Sekundärdaten greift daher an vielen Stellen zu kurz. Allerdings weisen auch eigens erhobene klinische oder patientenindividuelle Primärdaten allein in der Regel nicht den erforderlichen Informationsgehalt auf. Daher ist die Verknüpfung von Primär- und Sekundärdaten, zum Beispiel für die Evaluation von Therapieregimes chronischer Erkrankungen, bei denen Patienten zwischen verschiedenen Versorgungssektoren navigieren, die Methode der Wahl. Schließlich kann ein allumfassendes Bild langfristiger Behandlungseffekte sowie ihrer Kosten-Effektivität nur gezeichnet werden, wenn beide Datentypen miteinander verknüpft werden. Würden solche Optionen genutzt, ergäben sich günstige Perspektiven sowohl für die Versorgungsforschung als auch für die Versorgungsplanung und -steuerung. Auch könnten Leistungserbringer die Entwicklung von Parametern wie Qualitätsmaßen, Verlaufsinformationen und Befunden während des Behandlungsprozesses monitorieren, begleitend analysieren und gegebenenfalls steuernd in die medizinische Versorgung eingreifen (Paradebeispiel für ein System der „lernenden Versorgung“). Ergebnisse derartiger Studien könnten neben ihrer unmittelbaren Verwertbarkeit in der Nutzenbewertung auch Ausgangspunkt für gesundheitsökonomische Extrapolationen und Hochrechnungen auf zukünftige Versorgungspotenziale sein (Matusiewicz und Wasem 2012; Gothe *et al.* 2013).

Die NAKO-Gesundheitsstudie stellt ein herausragendes Beispiel für die Potenziale eines Datenlinkage dar. Sie ist die bislang größte in Deutschland etablierte epidemiologische Kohortenstudie. Ihre netto über 200.000 Teilnehmer werden regelmäßig im Abstand von vier bis fünf Jahren umfangreich zu aktuellen (chronischen) Erkrankungen, Risikofaktoren und vielfältigen Determinanten von Gesundheit und Krankheit befragt und einem vielfältigen Untersuchungsprogramm unterzogen inklusive der Sammlung von Biomaterialien. Ein erweitertes Programm sowie eine Magnetresonanztomografie (MRT) sind für eine Teilmenge der Teilnehmer vorgesehen (GNCC 2014; Wichmann *et al.* 2012). Ziel der Studie ist die Untersuchung der Ursachen für chronische Krankheiten, ihrer vorklinischen Stadien und ihres weiteren Verlaufs. Das Ergebnis der zunächst für zehn Jahre finanzierten, aber auf einen längeren Verlauf ausgelegten Studie ist der Aufbau und die breite Nutzung einer Forschungsdatenbank aus Primär- und ergänzenden Gesundheits- und Registerdaten. Zu denen gehören Routinedaten verschiedener Sozialversicherungsträger (unter anderem gesetzliche Kranken- und Rentenversiche-

Wer ist AGENS?

AGENS ist eine stetig wachsende Non-Profit-Arbeitsgruppe, die sich dem Themenfeld Sekundärdaten verschrieben hat. Zu AGENS gehören derzeit rund 450 Kollegen aus Forschung und Praxis, aus den Reihen der Sekundärdatennutzer wie auch der Dateneigner. Die Mitgliedschaft bei den Muttergesellschaften Deutsche Gesellschaft für Sozialmedizin und Prävention (DGSMP) und/oder Deutsche Gesellschaft für Epidemiologie (DGEpi) wird ausdrücklich empfohlen, ist aber nicht Voraussetzung für eine Mitarbeit.

Welche Ziele verfolgt AGENS?

AGENS wirkt darauf hin, den Zugang zu Sekundärdaten zu erleichtern und damit deren Nutzung für wissenschaftliche Forschung zu verbessern. Ein ganz wesentlicher Aspekt für das Engagement von AGENS sind dabei von jeher der Datenzugang und die Datenaufbereitung als essenzielle Voraussetzungen jeder nachfolgenden Datenanalyse. Dabei wirkt AGENS auch bei Politikberatung mit, etwa bei der Novellierung des § 303f SGB V oder der Ausgestaltung des im Digitale-Versorgung-Gesetz (DVG) angelegten Forschungsdatenzentrums. Wichtig dabei ist die Beachtung datenschutzrechtlicher Belange bei gleichzeitiger Förderung der (Versorgungs-)Forschung.

Wie funktioniert AGENS?

AGENS stellt ein offenes Forum für alle Interessenten an Sekundärdatenanalysen dar. Der Informationsaustausch erfolgt über den stetig wachsenden E-Mail-Verteiler, am Rande eigener Veranstaltungen und in speziellen Sessions auf den Jahrestagungen der Muttergesellschaften. Bei Interesse an Informationen über die Aktivitäten von AGENS genügt eine Mail an enno.swart@med.ovgu.de.

rung) und Daten der epidemiologischen und klinischen Krebsregister. Diese Daten dienen primär der passiven teilnehmerbezogenen Nachverfolgung und sind geeignet, dem häufig selektiven *loss to follow-up* zu begegnen, der Kohortenstudien immanent ist (Stallmann et al. 2015). Eine mit 90 Prozent sehr hohe Quote der aktiven Einwilligung der Teilnehmer zur ergänzenden Nutzung dieser Daten belegt die grundsätzlich in der Bevölkerung vorhandene Bereitschaft zur Nutzung dieser Daten für wissenschaftliche Zwecke (Schipf et al. 2020).

Das Sekundärdatenkonzept schließt von seinem Grundverständnis keine Datenquelle aus. Ganz im Gegenteil, es könnte sogar als Verpflichtung an die Forschenden verstanden werden, alle verfügbaren Daten daraufhin zu prüfen, ob sie „als Basis für Auswertungen im Sinne der wissenschaftlichen Fragestellung“ geeignet sind (GPS, Leitlinie 3, Empfehlung 3.1; siehe Swart et al. 2015). Vor dem Hintergrund dieser Hypothese öffnet sich ein Datenhorizont, der schier unerschöpflich zu sein scheint: Wearables generieren sekundengenaue Tagesaktivitätsprofile von Personen; Gesundheits-Apps drängen auf den Markt und dokumentieren das Krankheitsgeschehen; in unterschiedlichen Plattformen der sozialen Medien werden gesundheitsrelevante Daten im Tagebuchformat gesammelt. Dass diese Informationen grundsätzlich als Datenbasis für Sekundärdatenanalysen dienen können, kann nicht bezweifelt werden. Welche Voraussetzungen aber müssen sie erfüllen, um auch den Standards guter wissenschaftlicher Praxis zu genügen? Diese Frage lässt sich nicht allein theoretisch beantworten. Auch die Routinedaten haben erst durch ihre aktive Nutzung für Forschungsprojekte den Beweis ihrer Eignung erbracht. Spezifische Analysemethoden mussten entwickelt und validiert werden, Fallstricke bei Analyse und Ergebnisinterpretation mussten erkannt und diskutiert werden. Diesen datenspezifischen Evolutionsprozess, zuweilen als „digitale Transformation im Gesundheitswesen“ (Matusiewicz et al. 2017) bezeichnet, werden auch neue Datenkörper durchlaufen. Welche Daten am Ende in empirische Prozesse einbezogen werden, wird nicht zuletzt von der fantasievollen Ausgestaltung und Durchführung innovativer Versorgungsforschungsprojekte abhängen.

Wer über neue Daten für die Versorgungsforschung spricht, kommt nicht umhin, auch über neue Auswertungskonzepte und neue Datenhaltungskonzepte zu sprechen. Unter dem Schlagwort „künstliche Intelligenz“ durchforsten neue Auswertungstools die Daten, oft vergesellschaftet mit dem Begriff „Big Data“. Vielversprechende Ansätze sind gemacht, eine abschließende Beurteilung mit der Brille der Versorgungsforschung scheint aber verfrüht. Gleiches gilt auch für die Beurteilung neuer Datenhaltungskonzepte. Verteilte Datenbanken sollen die aus datenschutzrechtlicher Sicht problematische Datenkonzentration vermeiden helfen, intelligente Analyse-Wizards durchforsten heterogene Datenbanken, verknüpfen und analysieren die benötigten Daten und stellen lediglich

das Ergebnis zur Verfügung. Das Kopieren von sensiblen Datenbankinhalten wird dadurch vermieden. Eng verbunden mit dem Begriff der verteilten Datenbanken sind auch *smart contracts* auf Basis des Block-Chain-Konzepts. Personen können damit ihre Daten beispielsweise für Forschungsprojekte bereitstellen ohne Einschaltung eines Dritten (zum Beispiel Notar oder Vertrauensstelle) und können die Erlaubnis zur Datennutzung jederzeit auch wieder revidieren. Auch hierfür existieren vielversprechende Ansätze, die sich in der Praxis aus Sicht der Versorgungsforschung zu bewähren haben.

4 Fazit

Die Notwendigkeit einer intensiveren wissenschaftlichen Nutzung verfügbarer Daten zeigt der durch Bundesgesundheitsminister Jens Spahn aktiv vorangetriebene Prozess der Überarbeitung der Datentransparenzverordnung nach § 303 ff. SGB V, in dem zunächst die Daten der gesetzlichen Krankenversicherung und perspektivisch weitere Datenkörper einer umfangreicheren Nutzung durch Dritte zugeführt werden sollen. Im aktuellen Referentenentwurf werden die Limitationen der bisherigen über das (vormals) DIMDI zugänglichen Daten des morbiditätsorientierten Risikostrukturvergleichs überwunden, und es wird – bei allen verbliebenen offenen Fragen der konkreten Umsetzung – gezeigt, wie eine neue Datenbereitstellungs- und Nutzungskultur zukünftig aussehen könnte, mit dem Ziel, die aufgezeigten Paradoxa zumindest teilweise zu überwinden.

Literatur

- Amtsblatt der Europäischen Union (2016):** Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung); eur-lex.europa.eu/legal-content/DE/TXT/ → Suchwort: Verordnung (EU) 2016/679
- Berger ML, Mamdani M, Atkins D, Johnson ML (2009):** Good Research Practices for Comparative Effectiveness Research: Defining, Reporting and Interpreting Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report – Part I. *Value Health*, Vol. 12, No. 8, 1044–1052
- Cox E, Martin BC, Van Staa T et al. (2009):** Good Research Practices for Comparative Effectiveness Research: Approaches to Mitigate Bias and Confounding in the Design of Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report – Part II. *Value Health*, Vol. 12, No. 8, 1053–1061

German National Cohort Consortium (2014): The German National Cohort: Aims, Study Design and Organization. *European Journal of Epidemiology*, Vol. 29, No. 5, 371–382

Gothe H, Siebert U, Kugler J (2013): Sekundärdaten in der gesundheitsökonomischen Evaluierung von Telemedizinischen Anwendungen. In: Ammenwerth E, Hörbst A, Hayn D, Schreier G (Hrsg.): Tagungsband der eHealth 2013 „Health Informatics meets eHealth – von der Wissenschaft zur Anwendung und zurück“, 23.–24. Mai 2013. Wien: OCG, 255–260

Hernán MA et al. (2008): Observational Studies Analyzed like Randomized Experiments: An Application to Postmenopausal Hormone Therapy and Coronary Heart Disease. *Epidemiology*, Vol. 19, No. 6, 766–779

Hernán MA, Robins JM (2017): Per-Protocol Analyses of Pragmatic Trials. *New England Journal of Medicine*, Vol. 377, No. 14, 1391–1398

Johnson ML et al. (2009): Good Research Practices for Comparative Effectiveness Research: Analytic Methods to Improve Causal Inference from Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report – Part III. *Value Health*, Vol. 12, No. 8, 1062–1073

March S et al. (2019): Gute Praxis Datenlinkage (GPD). *Das Gesundheitswesen*, Vol. 81, No. 08/09, 636–650

Matusiewicz D, Pittelkau C, Elmer A (Hrsg.) (2017): Die Digitale Transformation im Gesundheitswesen: Transformation, Innovation, Disruption. Berlin: MWV Medizinisch Wissenschaftliche Verlagsgesellschaft

Matusiewicz D, Wasem J (2012): Gesundheitsökonomie in Deutschland – ein Blick zurück nach vorn. In: Mühlbauer BH, Kellerhoff F,

Matusiewicz D (Hrsg.): *Zukunftsperspektiven der Gesundheitswirtschaft*. Münster: LIT-Verlag Dr. W. Hopf, 420–439

Schiff S (2020): Die Basiserhebung der NAKO Gesundheitsstudie: Teilnahme an den Untersuchungsmodulen, Qualitätssicherung und Nutzung von Sekundärdaten. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz*, Vol. 63, No. 4, 254–266

Stallmann C et al. (2015): Individual Linkage of Primary Data with Secondary and Registry Data within Large Cohort Studies – Capabilities and Procedural Proposals. *Das Gesundheitswesen*, Vol. 77, No. 2, e37–e42

Stausberg J et al. (2020): Memorandum Register für die Versorgungsforschung: Update 2019. *Das Gesundheitswesen*, Vol. 82, No. 3, e39–e66

Swart E et al. (2015): Gute Praxis Sekundärdatenanalyse (GPS): Leitlinien und Empfehlungen. 3. Fassung; Version 2012/2014. *Das Gesundheitswesen*, Vol. 77, No. 2, 120–126

Swart E et al. (2016): Standardisierte BerichtsRoutine für Sekundärdaten Analysen (STROSA) – ein konsentierter Berichtsstandard für Deutschland, Version 2. *Das Gesundheitswesen*, Vol. 78, No. S01, e145–e160

Swart E, Ihle P, Gothe H, Matusiewicz D (Hrsg.) (2014): Routine-daten im Gesundheitswesen – Handbuch Sekundärdatenanalyse: Grundlage, Methoden und Perspektiven, 2., vollständig überarbeitete Auflage. Bern: Verlag Hans Huber

Wichmann HE et al. (2012): Die Nationale Kohorte. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz*, Vol. 55, No. 6–7, 781–789

(letzter Zugriff auf alle Internetquellen: 17. Juni 2020)

DIE AUTOREN



Dr. med. Holger Gothe,

Jahrgang 1961, ist Arzt und Kommunikationswissenschaftler. Als Versorgungsforscher widmet er sich seit über 30 Jahren dem analytischen Potenzial von GKV-Routinedaten. Seit 2000, unterbrochen durch eine mehrjährige Tätigkeit an der UMIT – Private Universität für Gesundheitswissenschaften, Medizinische Informatik und Technik in Hall in Tirol, leitet er den Bereich Versorgungsforschung des IGES Instituts in Berlin. Mit seinen beiden Koautoren ist er Sprecher der Arbeitsgruppe Erhebung und Nutzung von Sekundärdaten (AGENS) der Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSMP) und der Deutschen Gesellschaft für Epidemiologie (DGEpi).



PD Dr. rer. biol. hum. Enno Swart,

Jahrgang 1962, ist Statistiker, Epidemiologe und Versorgungsforscher. Er beschäftigt sich seit 25 Jahren mit der Erschließung und wissenschaftlichen Nutzung von Sekundärdaten und ist Gründungsmitglied und zusammen mit seinen Koautoren Sprecher der Arbeitsgruppe Erhebung und Nutzung von Sekundärdaten (AGENS) der Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSMP) und der Deutschen Gesellschaft für Epidemiologie (DGEpi).



Peter Ihle,

Jahrgang 1961, ist Arzt und seit 1991 wissenschaftlicher Mitarbeiter der PMV forschungsgruppe. Sein Hauptaufgabengebiet ist die Aufbereitung und Analyse von GKV-Routinedaten sowie das Thema Datenschutz- und Datensicherheit mit dem Schwerpunkt Pseudonymisierung. Er ist Gründungsmitglied und zusammen mit seinen Koautoren Sprecher der Arbeitsgruppe Erhebung und Nutzung von Sekundärdaten (AGENS) der Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSMP) und der Deutschen Gesellschaft für Epidemiologie (DGEpi).